

Package ‘PHInfiniteEstimates’

November 24, 2023

Type Package

Title Tools for Inference in the Presence of a Monotone Likelihood

Version 2.9.5

Date 2023-11-23

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. Supported by grant NSF DMS 1712839.

License GPL-3

Depends survival, lpSolve, coxphf, nph

Suggests KMsurv

Imports Rdpack (>= 0.7)

RdMacros Rdpack

Encoding UTF-8

RoxygenNote 7.2.3

NeedsCompilation no

Repository CRAN

Date/Publication 2023-11-24 04:30:02 UTC

R topics documented:

aalenjohansen	2
andersenplot	3
arjasplot	3
bestbeta	4
checkcensor	5
checkresults	5
compareplot	6

compete.simulation	6
convertbaselineltolr	7
convertmtol	8
convertstoml	9
drawdiagram	10
fixcoxph	11
gehan.wilcoxon.test	12
heinzschemper	13
inference	14
lrapproximations	15
network	16
newllk	17
PHInfiniteEstimates	18
pllk	19
reduceLR	20
simcode	21
simultaneouscoverage	23
summarizefits	24
summarizetable	25
survregpredict	25
testcox	26
voter.ml	26

Index 27

aalenjohansen	<i>Calculate the Aalen-Johansen (1978) estimate in the Competing risk context. See Aalen, Odd O., and Søren Johansen. "An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations." Scandinavian Journal of Statistics 5, no. 3 (1978): 141-50. Accessed January 15, 2021. http://www.jstor.org/stable/4615704.</i>
---------------	--

Description

Calculate the Aalen-Johansen (1978) estimate in the Competing risk context. See Aalen, Odd O., and Søren Johansen. "An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations." *Scandinavian Journal of Statistics* 5, no. 3 (1978): 141-50. Accessed January 15, 2021. <http://www.jstor.org/stable/4615704>.

Usage

```
aalenjohansen(times, causes)
```

Arguments

times	Event times.
causes	Causes, with 0 coded as censored, 1 as cause of interest, other for competing.

Value

a list with components

- times Unique times
- surv Aalen-Johansen estimator for cause 1.

andersenplot	<i>Plot hazards for two strata for each time. At times with an event in one but not the other group, the fitted hazard remains constant, and so the plot is a step function. If hazards are proportional between strata, then the plot should be close to a straight line.</i>
--------------	--

Description

Plot hazards for two strata for each time. At times with an event in one but not the other group, the fitted hazard remains constant, and so the plot is a step function. If hazards are proportional between strata, then the plot should be close to a straight line.

Usage

```
andersenplot(fit)
```

Arguments

fit	A coxph fit with a stratification term.
-----	---

arjasplot	<i>Examine the potential role of treatment in treatment in a model already including sex. Straight lines that are not 45 degrees indicate the appropriateness of new variable as a linear effect.</i>
-----------	---

Description

Examine the potential role of treatment in treatment in a model already including sex. Straight lines that are not 45 degrees indicate the appropriateness of new variable as a linear effect.

Usage

```
arjasplot(formulastring, time, stratifier, status, mydata)
```

Arguments

formulastring	A formula for a coxph fit.
time	The name of the time variable
stratifier	The name of the stratifier variable
status	The name of the status variable
mydata	The data frame.

`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, usecc = FALSE)
```

Arguments

<code>fit</code>	Output from a Cox PH regression, with <code>x=TRUE</code> and <code>y=TRUE</code>
<code>exclude</code>	data set with stratum and patient number to exclude.
<code>start</code>	Starting value
<code>touse</code>	columns of the design matrix to use.
<code>usecc</code>	Logical variable indicating whether to use a continuity correction, or numerical variable representing the continuity correction.

Value

Fitted survival analysis regression parameter of class `coxph`

References

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

Examples

```
bfit<-coxph(Surv(TIME, CENS)~T+N+CD, data=breast, x=TRUE)
noccfitt<-bestbeta(bfit)
bestbeta(bfit, usecc=TRUE, start=noccfitt$start)
```

checkcensor	<i>Check how censoring impacts sampling properties of KM fit and log rank test.</i>
-------------	---

Description

Check how censoring impacts sampling properties of KM fit and log rank test.

Usage

```
checkcensor(nsamp = 1000, nobs = 1000)
```

Arguments

nsamp	Number of MC samples
nobs	Number of observations

Value

biases of fits.

checkresults	<i>Produce a graphical assessment of Monte Carlo experiment on fidelity of proportional hazards regression to the uniform ideal.</i>
--------------	--

Description

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

Usage

```
checkresults(regnsimulation, frac = 0.1)
```

Arguments

regnsimulation	A structure with a component out, matrix with columns representing definitions of p-values and as many rows as there MC samples.
frac	Proportion for bottom of distribution to be assessed.

Value

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

compareplot *Plot results of simcode*

Description

Plot results of simcode

Usage

```
compareplot(simresults)
```

Arguments

simresults the result of simcode

Value

nothing.

compete.simulation *Simulate from a competing risk model with correlated log normal errors, and plot various estimates.*

Description

Simulate from a competing risk model with correlated log normal errors, and plot various estimates.

Usage

```
compete.simulation(ncr = 4, sig = 0.8, ns = 1000)
```

Arguments

ncr Number of competing risks.
sig correlation among competing risks.
ns number of observations.

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, Winters K (2007). "The British Election Study." <https://www.britishelectionstudy.com>.

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

Kolassa JE (2016). "Inference in the Presence of Likelihood Monotonicity for Polytomous and Logistic Regression." *Advances in Pure Mathematics*, **6**, 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)
# The package mlogit is scheduled for archiving. If it is available, the
# next two lines fit the model using mlogit.
# 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 JE (1997). “Infinite Parameter Estimates in Logistic Regression.” *Scandinavian Journal of Statistics*, **24**, 523–530. doi:10.1111/14679469.00078.

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

Examples

```
#Uses data set breast from package coxphf.
data(breast)
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	<i>Convert a proportional hazards regression to a multinomial regression.</i>
--------------	---

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 `convertmto1`, and then to `reducelR` of (Kolassa 1997). See examples in the `convertmto1` documentation.

Value

a data set on which to apply conditional multinomial regression, corresponding to the proportional hazards regression analysis. In order to run the line commented out below, you would need this: `# @importFrom mlogit mlogit.data`

References

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

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

Kolassa JE, Zhang J (2019). https://higherlogicdownload.s3.amazonaws.com/AMSTAT/fa4dd52c-8429-41d0-abdf-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	<i>Remove observations from a proportional hazards regression, and return the fit of the reduced model.</i>
----------	---

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 regression. 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, Zhang J (2019). https://higherlogicdownload.s3.amazonaws.com/AMSTAT/fa4dd52c-8429-41d0-abdf-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)

testdat2 <- data.frame(Time=c(4,3,1,1,2,2,3),
  Cen=c(1,1,1,0,0,0,0), Primary=c(0,2,1,1,1,0,0), Sex=c(0,0,0,0,1,1,1))
(bcfit<-coxph(Surv(Time,Cen)~Primary + Sex, testdat2, x=TRUE, ties="breslow"))
fixcoxph(bcfit,bcfit$x,"Primary")
```

gehan.wilcoxon.test *Perform Gehan's application to the Wilcoxon test for multiple samples, testing for equivalence of survival curve. See Klein and Moeschberger (1997) Survival Analysis (7.3.3) and pp. 193-194.*

Description

Perform Gehan's application to the Wilcoxon test for multiple samples, testing for equivalence of survival curve. See Klein and Moeschberger (1997) Survival Analysis (7.3.3) and pp. 193-194.

Usage

```
gehan.wilcoxon.test(  
  myformula,  
  data,  
  gehan = TRUE,  
  plot = FALSE,  
  alpha = 0.05,  
  subset = NULL  
)
```

Arguments

myformula	Proportional hazards formula appropriate for survfit
data	the data set
gehan	logical flag triggering the Wilcoxon test (gehan=TRUE), with weights equal to total at risk, or the log rank test (gehan=FALSE) with weights all 1.
plot	logical flag triggering plotting.
alpha	Nominal test level for plotting on graph
subset	Apply to a subset of the data.

Value

An htest-like object with the chi-square version of the test.

Examples

```
data(breast)#From package coxphf  
gehan.wilcoxon.test(Surv(TIME,CENS)~G,data=breast)
```

heinzschemper	<i>Simulate operating characteristics of repaired Cox regression and competitors.</i>
---------------	---

Description

This function is intended to verify the operating characteristics of the approximate conditional inferential approach of Kolassa and Zhang (2019) to proportional hazards regression. An exponential 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.

Usage

```
heinzschemper(
  nobs = 50,
  k = 5,
  B = 1,
  c = 0,
  nsamp = 1000,
  beta = NULL,
  add = NULL,
  half = NULL,
  verbose = FALSE,
  smoothfirst = FALSE
)
```

Arguments

nobs	number of observations in simulated data set.
k	number of covariates in simulated data set. Each covariate is dichotomous.
B	odds of 1 vs. 0 in dichotomous variables.
c	censoring proportion.
nsamp	number of samples.
beta	regression parameters, all zeros if null, and all the same value if a scalar.
add	partial simulation results to be added to, or NULL if de novo.
half	does nothing; provided for compatibility with simcode.
verbose	Triggers verbose messages.
smoothfirst	Triggers normal rather than dichotomous interest covariate.

Value

a list with components

- out matrix with columns corresponding to p-values.

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(
  netout,
  alpha = 0.05,
  rng = c(-5, 5),
  alternative = c("two.sided", "less", "greater")
)
```

Arguments

netout	List of the sort provided by network.
alpha	Test level, or 1- confidence level.
rng	Range of possible parameter values.
alternative	String indicating two- or one-sided alternative, and, if one-sided, direction.

Value

List of outputs, including

- ospv Observed one-sided p values
- tspv Observed two-sided p value.
- ci confidence interval.
- estimate Maximum conditional likelihood estimator.
- null.value Value of parameter under null hypothesis.
- data.name Name of data set
- method Method used to generate test.
- statistic sufficient statistic value for inference variable.
- p.value p.value
- conf.int confidence interval.
- alternative String indicating two- or one-sided alternative, and, if one-sided, direction.

and including standard stats:::orint.htest components, and of class htest.

References

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

Goorin AM, Perez–Atayde A, Gebhardt M, 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))

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

Irapproximations	<i>Assess the accuracy of the log rank statistic approximation to the true value, in the case without censoring. Provides plots of statistics, and empirical test level.</i>
------------------	--

Description

Assess the accuracy of the log rank statistic approximation to the true value, in the case without censoring. Provides plots of statistics, and empirical test level.

Usage

```
Irapproximations(nobs = 10, ratio = 1, nsamp = 1000)
```

Arguments

nobs	number of observations in each group. This currently supports only equal group size data sets.
ratio	Ratio of group means; use 1 for null.
nsamp	Monte Carlo sample size.

Value

a vector of empirical test sizes.

Examples

```
lrapproximations(nsamp=100)
```

network	<i>This function enumerates conditional sample spaces associated with logistic regression,</i>
---------	--

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,  
  data.name = "Test data"  
)
```

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.
data.name	Name of the data set.

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

Goorin AM, Perez–Atayde A, Gebhardt M, 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,
  cc1 = 0
)
```

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 likelihood
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.
cc1	Continuity correction for first component of the score.

Value

a list with components

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

References

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

Description

The PHInfiniteEstimates package Proportional hazards estimation in the presence of partial likelihood monotonicity 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 JE (1997). “Infinite Parameter Estimates in Logistic Regression.” *Scandinavian Journal of Statistics*, **24**, 523–530. doi:10.1111/14679469.00078.

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

 pllk

Partial likelihood for proportional hazards

Description

Partial likelihood for proportional hazards

Usage

```
pllk(beta, xmat, ind, cc = NULL)
```

Arguments

beta	parameter vector
xmat	regression matrix
ind	censoring indicator, 1 for event and any other value otherwise.
cc	Continuity correction for sum of x vectors with multiple occurrences in risk set. For binary covariates is half. Default a vector of zeros.

Value

a list with components

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

Examples

```
#Uses data set breast from package coxphf.
data(breast)
xmat<-as.matrix(breast)[order(breast$TIME),c("T","N")]
ind<-breast$CENS[order(breast$TIME)]
short<-coxph(Surv(TIME,CENS)~ T+ N,data=breast)
p1lk(as.vector(coef(short)),xmat,ind)
```

reduceLR	<i>Reduce a logistic regression with monotone likelihood to a conditional regression with double descending likelihood.</i>
----------	---

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, verybig = 1e+07)
```

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
verybig	threshold for condition number to declare colinearity.

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 IpSolve. 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 indicators 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, Patel NR (1987). “Computing Distributions for Exact Logistic Regression.” *Journal of the American Statistical Association*, **82**(400), pp. 1110-1117. ISSN 01621459, doi:10.2307/2289388.

Goorin AM, Perez-Atayde A, Gebhardt M, 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 JE (1997). “Infinite Parameter Estimates in Logistic Regression.” *Scandinavian Journal of Statistics*, **24**, 523–530. doi:10.1111/14679469.00078.

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

Kolassa JE, Zhang J (2019). https://higherlogicdownload.s3.amazonaws.com/AMSTAT/fa4dd52c-8429-41d0-abdf-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")
```

simcode

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

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

Examples

```
data(breast)

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

simultaneouscoverage *Calculate simultaneous coverage of pointwise confidence intervals.*

Description

Simulate exponential event times with expectation 1. Simulate censoring times with expectation 2. Calculate confidence intervals and check simultaneous coverage.

Usage

```
simultaneouscoverage(nsamp, nobs)
```

Arguments

nsamp	Number of Monte Carlo samples.
nobs	Number of observations per sample.

Value

Simultaneous coverage proportion.

Examples

```
simultaneouscoverage(1000,20)
```

summarizefits

Summarize proportional hazards fits

Description

Summarize proportional hazards fits

Usage

```
summarizefits(
  repairedfit,
  penalizedout,
  penalizedoutsmaller,
  iv,
  verbose = TRUE
)
```

Arguments

repairedfit	coxph fit
penalizedout	coxphf fit
penalizedoutsmaller	smaller coxphf fit
iv	name of the variable of interest, as a character string
verbose	logical flag triggering intermediate messaging.

Value

a vector with components

- Wald p-value from the Cox regression fit.
- partial likelihood ratio p-value from Cox regression fit.
- parameter estimate from the Cox regression fit.
- standard error from the Cox regression fit.
- Conditional Skovgaard standard error from the Cox regression fit.
- Signed root of the partial likelihood ratio statistic from Cox regression fit.
- partial likelihood ratio statistic p-value from coxphf
- Wald p-value from coxphf
- parameter estimate from coxphf
- standard error from coxphf
- number of parameters in reduced fit.

References

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

summarizetable	<i>Summarize the results of simulations investigating operating conditions for the data reduction method to avoid monotone likelihood. Files are of form "hsxxx", for xxx numerals.</i>
----------------	---

Description

Summarize the results of simulations investigating operating conditions for the data reduction method to avoid monotone likelihood. Files are of form "hsxxx", for xxx numerals.

Usage

```
summarizetable()
```

survregpredict	<i>Fit survival probabilities from a survreg object.</i>
----------------	--

Description

Fit survival probabilities from a survreg object.

Usage

```
survregpredict(fit, newdata, time)
```

Arguments

fit	a survreg object. This should not contain strata(). It also must use the log transformation.
newdata	a new data set with covariates from the fit.
time	a time value (on the original, and not log, scale).

Examples

```
#Fit the survival probability for an individual with extent 1 and
#differentiation 2 at 700 days from a Weibull regression using the
#colon cancer data set distributed as part of the survival package.
fit<-survreg(Surv(time,status)~factor(extent)+differ,data=colon)
survregpredict(fit,data.frame(extent=1,differ=2),700)
```

testcox	<i>Test size of asymptotic Cox tests.</i>
---------	---

Description

Test size of asymptotic Cox tests.

Usage

```
testcox(nsamp = 1000, nobs = 50, ncov = 5, randomcov = TRUE)
```

Arguments

nsamp	Number of MC samples
nobs	Number of observations
ncov	Number of covariates
randomcov	Indicator of whether to draw random covariates.

Value

level of two-sided test of nominal size 0.05.

voter.ml	<i>Subset of British elections data used in (Kolassa 2016). Data are from (Sanders et al. 2007).</i>
----------	--

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, Winters K (2007). “The British Election Study.” <https://www.britishelectionstudy.com>.

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

Examples

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

Index

aalenjohansen, [2](#)
andersenplot, [3](#)
arjasplot, [3](#)

bestbeta, [4](#)

checkcensor, [5](#)
checkresults, [5](#)
compareplot, [6](#)
compete.simulation, [6](#)
convertbaselineltolr, [7](#)
convertmtol, [8](#)
convertstoml, [9](#)

drawdiagram, [10](#)

fixcoxph, [11](#)

gehan.wilcoxon.test, [12](#)

heinzschemper, [13](#)

inference, [14](#)

lrapproximations, [15](#)

network, [16](#)
newllk, [17](#)

PHInfiniteEstimates, [18](#)
pllk, [19](#)

reduceLR, [20](#)

simcode, [21](#)
simultaneouscoverage, [23](#)
summarizefits, [24](#)
summarizetable, [25](#)
survregpredict, [25](#)

testcox, [26](#)

voter.ml, [26](#)