# Package 'multipleNCC'

February 12, 2020

**Type** Package

**Title** Weighted Cox-Regression for Nested Case-Control Data

**Version** 1.2-2

**Date** 2020-02-06

**Author** Nathalie C. Stoer, Sven Ove Samuelsen

**Maintainer** Nathalie C. Stoer <nathalie.c.stoer@kreftregisteret.no>

**Description** Fit Cox proportional hazard models with a weighted partial likelihood. It handles one or multiple endpoints, additional matching and makes it possible to reuse controls for other endpoints.

**Depends** survival, mgcv

**License** GPL-2

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2020-02-12 10:30:03 UTC

## R topics documented:

---

`multipleNCC-package` *Weighted partial likelihood for nested case-control data*

---

## Description

Fits Cox proportional hazards models with a weigthed partial likelihood. It handles competing risks (with one endpoint being a special situation). It uses cases and controls from other endpoints as additional controls for each endpoint. See `wpl` for help.

Four weight estimators are implemented; Kaplan-Meier type `KMprob`, GAM (`GAMprob`), GLM (`GLMprob`) and local averaging (`Chenprob`)

## Details

| | |
|---|---|
| Package: | multipleNCC |
| Type: | Package |
| Version: | 1.2-1 |
| Date: | 2016-04-16 |
| License: | GPL-2 |
| LazyLoad: | yes |

## Author(s)

Nathalie C. Stoer

Maintainer: <nathalie.c.stoer@kreftregisteret.no>

## References

Samuelsen, SO. (1997) A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika* **84(2)**, 379-394
Samuelsen, SO., et al. (2007) Stratified case-cohort analysis of general cohort sampling designs. *Scand J Stat* **34(1)**, 103-119
Chen, KN. (2001) Generalized case-cohort sampling. *J Roy Stat Soc Ser B* **63(4)**, 791 - 809
Stoer NC and Samuelsen SO (2012): Comparison of estimators in nested case-control studies with multiple outcomes. Lifetime Data Analysis, 18(3), 261-283.

## See Also

`wpl`, `coxph`, `Chenprob`, `GLMprob`, `GAMprob`, `KMprob`

Chenprob *Sampling probabilities estimated with local averaging.*

## Description

Estimates sampling probabilities with local averaging (Chen, 2001). The weights included in the Cox-regressions (wpl) and which could be used for other procedures are inverse sampling probabilities i.e. the inverse of these probabilities. The probabilties are estimated for all subjects in the cohort.

## Usage

```
Chenprob(survtime, samplestat, no.intervals = 10, left.time = 0,
no.intervals.left = c(3,4))
```

## Arguments

| | |
|---|---|
| `survtime` | Follow-up time for all cohort subjects |
| `samplestat` | A vector containing sampling and status information: 0 represents non-sampled subjects in the cohort, 1: sampled controls, 2,3,... indicate different events. Cohort dimension. |
| `no.intervals` | Number of intervals for censoring times for Chen-weights with only right censoring |
| `left.time` | Entry time if the survival times are left-truncated. Cohort dimension. |
| `no.intervals.left` | Number of intervals for Chen-weights with left-truncation. A vector on the form [number of intervals for left truncated time, number of intervals for survival time]. |

## Value

A vector of cohort dimension of sampling probabilities.

## Author(s)

Nathalie C. Stoer

## References

Chen KN (2001) Generalized case-cohort sampling. J Roy Stat Soc Ser B 63(4):791-809
Stoer NC and Samuelsen SO (2012): Comparison of estimators in nested case-control studies with multiple outcomes. Lifetime Data Analysis, 18(3), 261-283.

## See Also

wpl, coxph, GAMprob, GLMprob, KMprob

## Examples

```
data(CVD_Accidents)
attach(CVD_Accidents)
Chenprob(agestop,samplestat,left.time=agestart)
Chenprob(agestop,samplestat,left.time=agestart,no.intervals.left=c(3,4))

function (survtime, samplestat, no.intervals, left.time = 0, no.intervals.left = 0)
{
    n.cohort = length(survtime)
    status = rep(0, n.cohort)
    status[samplestat > 1] = 1
    samplestat[samplestat > 1] = 1
    ind.no = 1:length(samplestat)
    p = pChen(status, survtime, samplestat, ind.no, n.cohort,
        no.intervals, left.time, no.intervals.left)
    p[status == 1] = 1
    p
  }
```

---

# CVD_Accidents *Causes of death in three counties in Norway in 1974-2000*

---

## Description

Causes of death from 1974-2000 for all men and women participating in a cardiovascular health screening in 1974-1978 in three counties in Norway. All variables are know for all cohort members and it is thus a synthetic nested case-control stidy. One control per case is sampled for cardiovascular disease cases and subjects who died from alcohol abuse, liver disease, and accidents and violence. The controls are matched sex and BMI plus/minus 2 in addition to being alive at the time the case died.

## Usage

```
data(CVD_Accidents)
```

## Format

A data frame with 3933 observations on the following 23 variables.

`agestart` Age at health survey, inclusion time

`agestop` Age at censoring

`dead` Indicator for death from any cause (0=censored, 1=dead)

`dead1` Indicator for cancer death (0=censored or dead from other cause than cancer, 1=dead from cancer)

`dead2` Indicator for death from cardiovascular disease, including sudden death (0=censored or dead from other causes than cardiovascular diseas, 1=dead from cardiovascular diseas)

`dead3` Indicator for death from other medical causes (0=censored or dead from cancer, cardiovascular diseas, alcohol abuse, liver disease, violence or accidents, 1=dead from other medical causes)

`dead4` Indicator for death from alcohol abuse, liver disease, violence and accidents (0=censored or death from other medical causes than alcohol abuse, liver disase, violence or accidents, 1=death from alcohol abuse, liver disease, violence and accidents)

`sex` sex (1=male, 2=female)

`county` county in Norway (5=Oppland, 14=Sogn og Fjordane, 20=Finnmark)

`sbp` Systolic blood pressure at health screening

`bmi` Body mass index at helth screening

`smkstart` Age started smoking

`smkgr` Smoking group (1=never smoked, 2=former smoker, 3=1-9 cigaretts per day, 4=10-19 cigaretts per day, 5=20+ cigaretts per day, 6=pipe or cigar)

`smoking3gr` Smoking 3 groups (1=never smoked, 2=former smoker, 3=smoker)

`samplestat` Indicator for sampling and events (0=non-sampled subjects in the cohort, 1=sampled controls, 2=dead from cardiovascular disease, 3=dead from alcohol abuse, liver disease, violence or accidents

`dead24` Indicator for death from either cardiovascular disease or alcohol abuse, liver disease, violence or accidents (0=censored or dead from other causes than cardiovascular disease, alcohol abuse, liver disease, violence or accidents, 1=death from cardiovascular disease, alcohol abuse, liver disease,violence or accidents)

## Source

http://folk.uio.no/borgan/abg-2008/data/data.html

---

`GAMprob` *Sampling probabilities estimated with generalized additive models.*

---

## Description

Estimates sampling probabilities with generalized additive models. The weights included in the Cox-regressions (wpl) and which could be used for other procedures are inverse sampling probabilities i.e. the inverse of these probabilities. The probabilties are estimated for all subjects in the cohort.

`survtime`, `left.time` and continuous matching variables will be smoothed on while categorical matching variables are taken as factors.

## Usage

```
GAMprob(survtime, samplestat, left.time = 0, match.var = 0, match.int = 0)
```

## Arguments

| | |
|---|---|
| survtime | Follow-up time for all cohort subjects |
| samplestat | A vector containing sampling and status information: 0 represents non-sampled subjects in the cohort, 1: sampled controls, 2,3,... indicate different events. Cohort dimension. |
| left.time | Entry time if the survival times are left-truncated. Cohort dimension. |
| match.var | If the controls are matched to the cases (on other variables than time), match.var is the vector of matching variables. Cohort dimension. |
| match.int | A vector of length 2*number of matching variables. For caliper matching (matched on value pluss/minus epsilon) match.int should consist of c(-epsilon,epsilon). For exact matching match.int should consist of c(0,0). |

## Value

A vector of cohort dimension of sampling probabilities.

## Author(s)

Nathalie C. Stoer

## References

Stoer NC and Samuelsen SO (2013): Inverse probability weighting in nested case-control studies with additional matching - a simulation study. Statistics in Medicine, 32(30), 5328-5339.

## See Also

wpl, coxph, Chenprob, GLMprob, KMprob, gam

## Examples

```
data(CVD_Accidents)
attach(CVD_Accidents)
GAMprob(agestop,samplestat,agestart)
GAMprob(agestop,samplestat,agestop,match.var=cbind(sex,bmi),match.int=c(0,0,-2,2))

## The function is currently defined as
function (survtime, samplestat, left.time = 0, match.var = 0, match.int = 0)
{
    n.cohort = length(survtime)
    status = rep(0, n.cohort)
    status[samplestat > 1] = 1
    samplestat[samplestat > 1] = 1
    pgam = pGAM(status, survtime, samplestat, n.cohort, left.time)
    p = rep(1, n.cohort)
    p[status == 0] = pgam
    p
  }
```

GLMprob *Sampling probabilities estimated with logistic regression.*

## Description

Estimates sampling probabilities with logistic regression. The weights included in the Cox-regressions (wpl) and which could be used for other procedures are inverse sampling probabilities i.e. the inverse of these probabilities. The probabilties are estimated for all subjects in the cohort.

`survtime`, `left.time` and continuous matching variables are included in the logistic regression as continuous variables while categorical matching variables are taken as factors.

## Usage

```
GLMprob(survtime, samplestat, left.time = 0, match.var = 0, match.int = 0)
```

## Arguments

| | |
|---|---|
| `survtime` | Follow-up time for all cohort subjects |
| `samplestat` | A vector containing sampling and status information: 0 represents non-sampled subjects in the cohort, 1: sampled controls, 2,3,... indicate different events. Cohort dimension. |
| `left.time` | Entry time if the survival times are left-truncated. Cohort dimension. |
| `match.var` | If the controls are matched to the cases (on other variables than time), match.var is the vector of matching variables. Cohort dimension. |
| `match.int` | A vector of length 2*number of matching variables. For caliper matching (matched on value pluss/minus epsilon) match.int should consist of c(-epsilon,epsilon). For exact matching match.int should consist of c(0,0). |

## Value

A vector of cohort dimension of sampling probabilities.

## Author(s)

Nathalie C. Stoer

## References

Stoer NC and Samuelsen SO (2013): Inverse probability weighting in nested case-control studies with additional matching - a simulation study. Statistics in Medicine, 32(30), 5328-5339.

## See Also

`wpl`, `coxph`, `Chenprob`, `GAMprob`, `KMprob`

## Examples

```
data(CVD_Accidents)
attach(CVD_Accidents)
GLMprob(agestop,samplestat,agestart)
GLMprob(agestop,samplestat,agestart,match.var=cbind(sex,bmi),match.int=c(0,0,-2,2))

## The function is currently defined as
function (survtime, samplestat, left.time = 0, match.var = 0,
    match.int = 0)
{
    n.cohort = length(survtime)
    status = rep(0, n.cohort)
    status[samplestat > 1] = 1
    samplestat[samplestat > 1] = 1
    pglm = pGLM(status, survtime, samplestat, n.cohort, left.time,
        match.var, match.int)
    p = rep(1, n.cohort)
    p[status == 0] = pglm
    p
  }
```

---

KMprob *Sampling probabilities estimated with a Kaplan-Meier type formula*

---

## Description

Estimates sampling probabilities with a Kaplan-Meier type formula. The weights included in the Cox-regressions (wpl) and which could be used for other procedures are inverse sampling probabilities i.e. the inverse of these probabilities. The probabilties are estimated for all subjects in the cohort.

## Usage

```
KMprob(survtime, samplestat, m, left.time = 0, match.var = 0, match.int = 0)
```

## Arguments

| | |
|---|---|
| survtime | Follow-up time for all cohort subjects |
| samplestat | A vector containing sampling and status information: 0 represents non-sampled subjects in the cohort, 1: sampled controls, 2,3,... indicate different events. Cohort dimension. |
| m | Number of sampled controls. A scalar if equal number of controls for all case. If unequal number of controls per case: A vector of length number of cases. The vector must be in the same order as the cases in the samplestat-vector. |
| left.time | Entry time if the survival times are left-truncated. Cohort dimension. |
| match.var | If the controls are matched to the cases (on other variables than time), match.var is the vector of matching variables. Cohort dimension. |

`match.int` A vector of length 2*number of matching variables. For caliper matching (matched on value pluss/minus epsilon) match.int should consist of c(-epsilon,epsilon). For exact matching match.int should consist of c(0,0).

## Value

A vector of cohort dimension of sampling probabilities.

## Author(s)

Nathalie C. Stoer

## References

Samuelsen SO. A pseudolikelihood approach to analysis of nested case-control studies. Biometrika, 84(2):379-394, 1997.
Stoer NC and Samuelsen SO (2013): Inverse probability weighting in nested case-control studies with additional matching - a simulation study. Statistics in Medicine, 32(30), 5328-5339.

## See Also

wpl, coxph, Chenprob, GLMprob, GAMprob

## Examples

```
data(CVD_Accidents)
attach(CVD_Accidents)
KMprob(agestop,samplestat,m=1,agestart)
KMprob(agestop,samplestat,m=1,agestart,match.var=cbind(bmi),match.int=c(-2,2))

## The function is currently defined as
function (survtime, samplestat, m,  left.time = 0, match.var = 0, match.int = 0)
{
    n.cohort = length(survtime)
    status = rep(0, n.cohort)
    status[samplestat > 1] = 1
    o = order(survtime)
    status = status[o]
    survtime = survtime[o]
    if (length(left.time) == n.cohort) {
        left.time = left.time[o]
    }
    if (length(match.var) == n.cohort) {
        match.var = match.var[o]
    }
    if (length(match.var) > n.cohort) {
        match.var = match.var[o, ]
    }
    tilbakestill = (1:n.cohort)[o]
    p = pKM(status, survtime, m, n.cohort, left.time, match.var,
        match.int)
    p[status > 0] = 1
```

```
  p = p[order(tilbakestill)]
  p
}
```

---

`ModelbasedVar` *Modelbased variance using Kaplan-Meier weights*

---

## Description

For internal use only

## Author(s)

Nathalie C. Stoer and Sven Ove Samuelsen

---

`multipleNCC-internal` *Internal function*

---

## Description

Internal function

## Author(s)

Nathalie C. Stoer

---

`pChen` *Chen-weights*

---

## Description

Estimates Chen-weights. For internal use only. Users should use the wrapper `Chenprob`.

## Author(s)

Nathalie C. Stoer

## See Also

`wpl`, `Chenprob`

pGAM *Generlaized additive model weights*

## Description

Estimates GAM-weights. For internal use only. Users should use the wrapper GAMprob.

## Author(s)

Nathalie C. Stoer

## See Also

wpl, GAMprob, gam

pGLM *Logistic regression weights*

## Description

Estimates GLM-weights. For internal use only. Users should use the wrapper GLMprob.

## Author(s)

Nathalie C. Stoer

## See Also

wpl, GLMprob, glm

pKM *Kaplan-Meier weights*

## Description

Estimates Kaplan-Meier weights. For internal use only. Users should use the wrapper KMprob.

## Author(s)

Nathalie C. Stoer

## See Also

wpl, KMprob

## PoststratVar *Modelbased variance using Chen-weights*

### Description

For internal use only.

### Author(s)

Nathalie C. Stoer and Sven Ove Samuelsen

## print.wpl *Print a wpl object*

### Description

Prints the fit of (each) weighted Cox-regression

### Usage

```
  ## S3 method for class 'wpl'
print(x,...)
```

### Arguments

| | |
|---|---|
| x | The result of a call to wpl |
| ... | For future methods |

### Author(s)

Nathalie C. Stoer

### See Also

wpl

summary.wpl *Summary method for wpl*

## Description

produces a summary of a fitted wpl object

## Usage

```
  ## S3 method for class 'wpl'
summary(object,...)
```

## Arguments

| | |
|---|---|
| object | the result of a wpl fit |
| ... | for future methods |

## Author(s)

Nathalie C. Stoer

## See Also

wpl

wpl *Weighted partial likelihood for nested case-control data*

## Description

Fits Cox proportional hazards models for nested case-control data with a weigthed partial likelihood. Matching between cases and controls is broken which enables the controls to be reused for other endpoints. It handles competing risks (with simple survival data with one endpoint being a special case) and cases and controls from one endpoint are being used as additional controls for another endpoint. There are four choices of weights; Samuelsen (1997) KM, estimated with logistic regression (glm), logistic genralized additive model (gam) and local averaging (Chen, 2001) (Chen). KM, glm and gam handle additional matching, while all of them handle left-truncation.

## Usage

```
wpl(x, data, samplestat, m = 1, weight.method = "KM", no.intervals = 10,
variance = "robust", no.intervals.left = c(3, 4), match.var = 0, match.int = 0)
```

## Arguments

| | |
|---|---|
| x | A formula object, with the response on the left of a ~ operator, and the terms on the right. The response must be a survival object as returned by the Surv function. The status variable going in to Surv is not actually used but should have 1 for cases and zero for controls and non-sampled subjetcs. All elements going into the formula should have lenght equal to the number of subjects in the cohort. Generally some of the covariates are not known for all subjects in the cohort (due to the NCC-sampling). The covariate values for those subjects should just be given some value e.g. 0 (not NA). Which value choosen is not important as the values are never used. |
| data | data.frame in which to interpret the variables named in the formula. |
| samplestat | A vector containing sampling and status information: 0 represents non-sampled subjects in the cohort, 1: sampled controls, 2,3,... indicate different events. Cohort dimension. |
| m | Number of sampled controls. A scalar if equal number of controls for all cases. If unequal number of controls per case: A vector of length number of cases. The vector must be in the same order as the cases in the samplestat-vector. |
| weight.method | Which weigths should be used, possibilities "KM", "gam", "glm", "Chen" |
| no.intervals | Number of intervals for censoring times for Chen-weights with only right censoring |
| variance | Default is robust variances, but model based variance (only for KM-weights), "Modelbased" and variance based on stratified case-cohort "Poststrat" (only for Chen-weights) is also possible. Pseudo-variance and Strat-variance will appear under "est.se(coef)" in the output. |
| no.intervals.left | Number of intervals for Chen-weights with left-truncation. A vector on the form [number of intervals for left truncated time, number of intervals for survival time]. |
| match.var | If the controls are matched to the cases (on other variables than time), match.var is the vector or matrix of matching variables. Cohort dimension. |
| match.int | A vector of length 2*number of matching variables. For caliper matching (matched on value pluss/minus epsilon) match.int should consist of c(-epsilon,epsilon). For exact matching match.int should consist of c(0,0). |

## Value

An object of class wpl representing the fit. Objects of this class have methods for the functions print and summary. The wpl-object consists of the following elements which are repeated for each endpoint. Unfortunately only the values for the first endpoint can be reached by $-operator(ex. fit$coefficients only return the coefficients for the first endpoint)

| | |
|---|---|
| coefficients | The vector of coefficients. |
| var | Robust or estimated variance |
| weighted.loglik | A vector of length 2 containing the log-likelihood with the initial values and with the final values of the coefficients. |

| | |
|---|---|
| `iter` | Number of iterations used |
| `linear.predictors` | The vector of linear predictors, one per subject. Note that this vector has been centered, see predict.coxph for more details |
| `residuals` | The martingale residuals |
| `means` | Vector of column means of the X matrix |
| `method` | The computation method used |
| `n` | The number of observations used in the fit |
| `nevent` | The number of events (usually deaths) used in the fit |
| `naive.var` | naive.var |
| `rscore` | The robust log-rank statistic |
| `wald.test` | The Wald test of whether the final coefficients differ from the initial values |
| `y` | Inclusion time and event/censoring time |
| `weights` | The vector of weights, which are inverse sampling probabilities |
| `est.var` | Estimated variance (T) or robust variance (F) |

.
.
.

## Author(s)

Nathalie C. Stoer

## References

Samuelsen SO. A pseudolikelihood approach to analysis of nested case-control studies. Biometrika, 84(2):379-394, 1997.
Stoer NC and Samuelsen SO (2013): Inverse probability weighting in nested case-control studies with additional matching - a simulation study. Statistics in Medicine, 32(30), 5328-5339.

## See Also

`coxph`, `Chenprob`, `GLMprob`, `GAMprob`

## Examples

```
data(CVD_Accidents)
wpl(Surv(agestart,agestop,dead24)~factor(smoking3gr)+bmi+factor(sex),data=CVD_Accidents,
samplestat=CVD_Accidents$samplestat,weight.method="gam")

wpl(Surv(agestart,agestop,dead24)~factor(smoking3gr)+bmi+factor(sex),data=CVD_Accidents,
samplestat=CVD_Accidents$samplestat,m=1,match.var=cbind(CVD_Accidents$sex,
CVD_Accidents$bmi),match.int=c(0,0,-2,2),weight.method="glm")
```

```
## The function is currently defined as
function (x, data, samplestat, m = 1, weight.method = "KM", no.intervals = 10,
    variance = "robust", no.intervals.left = c(3, 4), match.var = 0,
    match.int = 0)
{
    UseMethod("wpl")
  }
```

# Index