

Description of the help functions for estimation of the parametrical survival model

November 11, 2020

1. Preliminaries

To start computations execute the following commands:

- Define working directory `WD="PathToWorkingDirectory"`;
- Put script `"HelpFunctionsParamAnalysisNew7.R"`;
- Set working directory using operator `"setwd(WD)"`;
- Install and open packages `"xtable"`, `"ucminf"`, `"MASS"`, `"survival"`;
- execute command `"source("HelpFunctionsParamAnalysisNew7.R")"`;
- execute command `"data(lung)"`.

Structure of the data

The data set includes the following fields:

- Time-to-failure and censoring in the case without left truncation or time-of-start , time-of-failure, and censoring in the case with left truncation at the time of begin. Censoring must be either 0 (no event) or 1 (event);
- Covariates (continuous or categorical) used in a study (can be empty set).

2. Description of the help functions

Function *NamFact*.

Description

This function returns the list of the factor names in the study after converting categorical variables in binary ones.

Usage

```
NamFact(data, formula.scale, formula.shape);
```

Arguments

data: The data set;

formula.scale: The formula object defining the fields for time-to-failure (or time-of-start and time-to-failure) and for covariates influencing the proportional hazard term;

formula.shape: The object defining the fields for covariates influencing the shape.

Value

List of the factor names.

Function *msurv*.

Description

This function calculates the marginal survival.

Usage

```
msurv(ID);
```

Arguments

ID: Number of the object in the data set.

Value

Marginal survival for the object number *ID*.

Function *cumhazard*.

Description

This function calculates cumulative survival.

Usage

cumhazard(*ID*);

Arguments

ID: Number of the object in the data set.

Value

Cumulative hazard for the object number *ID*.

Function *hazard*.

Description

This function calculates instant hazard.

Usage

hazard(*ID*);

Arguments

ID: Number of the object in the data set.

Value

Instant hazard for the object number *ID*.

Function *LikGenNPH*.

Description

This function calculates the neglikelihood for parametrical survival model.

Usage

LikGenNPH(θ , *D*, *nf*, *nk*, *ncl*, *dist*);

Arguments

θ : Vector of parameters in the form

$$\theta = (\log a, \log b, \beta_{shape}, \beta_{scale}, \log \sigma^2),$$

where *a* and *b* are slope and shape parameters defining the hazard functions (Weibull or Gompertz, see details below), β_{shape} and β_{scale} are the Cox-regression parameters for shape and scale, respectively, and σ^2 is the variance of frailty. This vector must

include at least two parameters ($\log a$ and $\log b$);

D: The data set as described in previous section;

nf: The number of continuous and binary factors in the data set *D* corresponding to the covariates used in the Cox-regression for proportional hazard term;

nk: The number of continuous and binary factors in the data set *D* corresponding to the covariates used in the Cox-regression for shape *b*;

ncl: The number of clusters in the data set *D* corresponding to the covariate defining the shared frailty (is equal to 0 for the fixed-effect model);

dist: distribution of the time-to-failure ('Weibull' or 'Gompertz').

Value

Negloglikelihood.

Function *GrGenNPH*.

Description

This function calculates the neggradient of the likelihood for parametrical survival model.

Usage

GrGenNPH($\theta, D, nf, nk, ncl, dist$);

Arguments

θ : Vector of parameters in the form

$$\theta = (\log a, \log b, \beta_{shape}, \beta_{scale}, \log \sigma^2),$$

where *a* and *b* are slope and shape parameters defining the hazard functions (Weibull or Gompertz, see details below), β_{shape} and β_{scale} are the Cox-regression parameters for shape and scale, respectively, and σ^2 is the variance of frailty. This vector must include at least two parameters ($\log a$ and $\log b$);

D: The data set as described in previous section;

nf: The number of continuous and binary factors in the data set *D* corresponding

to the covariates used in the Cox-regression for proportional hazard term;
nk: The number of continuous and binary factors in the data set *D* corresponding to the covariates used in the Cox-regression for shape *b*;
ncl: The number of clusters in the data set *D* corresponding to the covariate defining the shared frailty (is equal to 0 for the fixed-effect model);
dist: distribution of the time-to-failure ('Weibull' or 'Gompertz').

Value

Neggradient of the loglikelihood.

Function *ParNPHCox*.

Description

This function calculates the estimates of unknown parameters, their standard errors, and other attributes.

Usage

```
c(par, se, LogLik, Tab, Names, Conc, pval, p.contrast, pstrata) :=  

ParNPHCox(formula.scale, formula.shape, cluster, dist, data, expr, strata);
```

Arguments

formula.scale: The formula object defining the fields for time-to-failure (or time-of-start and time-to-failure) and for covariates influencing the proportional hazard term;

formula.shape: The object defining the fields for covariates influencing the shape;

cluster: The name of the covariate defining the random effect (is equal to NULL for the fixed-effect model);

dist: distribution of the time-to-failure ('Weibull' or 'Gompertz');

Data: Data set.

expr: vector of expressions for contrasts. NULL, otherwise;

strata: Full list of the factors used in the study and their levels in the form:

strata=list(factor1=level of the factor1,...,last factor=level of the last factor).

NULL, otherwise.

Value

par: parameter estimates for vector $(\log a, \log b, \beta_{shape}, \beta_{scale}, \log \sigma^2)$;

se: standard errors for vector *par*;

LogLik: the value of the loglikelihood;

Tabs: the table of parameter estimates, their standard errors, and *p*-values in the LaTeX format;

Names: names of estimated parameters.

Conc: concordance and its standard error.

pval: vector of *p*-values for parameter estimates. For null hypothesis the values of a, b, σ^2 and Cox-regression coefficients are equal to zero.

p.contrast: data frame for means, CIs, and *p*-values for contrasts;

pstrata: data frame for times, means, and 95% CIs of the marginal survivals, cumulative hazards and hazards. The column names are:

"Time", "Mean.survival", "Low.survival", "Upper.survival", "Mean.cumulative.hazard",
"Low.cumulative.hazard", "Upper.cumulative.hazard", "Mean.hazard", "Low.hazard",
"Upper.hazard".

Details

Two kinds of the time-to-failure distribution are used in this function:

- Weibull with baseline cumulative hazard function

$$H_{base}(t; a, b) = (t/a)^b;$$

- Gompertz with baseline cumulative hazard function

$$H_{base}(t; a, b) = (a/b)(\exp(bt) - 1).$$

The cumulative hazard function is defined by

$$H(t; a, b, \beta_{shape}, \beta_{scale}, \mathbf{u}) = e^{\beta_{scale} \mathbf{u}} H_{base}(t; a, b e^{\beta_{shape} \mathbf{u}})$$

for the covariate vector \mathbf{u} . Here $e^{\beta_{scale}\mathbf{u}}$ is the Cox-regression term for proportional hazard and $e^{\beta_{shape}\mathbf{u}}$ is the Cox-regression term for shape.

The univariate survival function is defined by

$$\begin{aligned} S(t; a, b, \beta_{shape}, \beta_{scale}, \sigma^2, \mathbf{u}) &= \mathbb{E} \exp(-ZH(t; a, b, \beta_{shape}, \beta_{scale}, \mathbf{u})) \\ &= (1 + \sigma^2 H(t; a, b, \beta_{shape}, \beta_{scale}, \mathbf{u}))^{-1/\sigma^2} \end{aligned}$$

for gamma-distributed frailty Z with mean 1 and variance σ^2 .

The 'formula.scale' and 'formula.shape' are formula objects used in the R-package survival and have the form $Surv(time, Cens) \sim factor_1 + \dots + factor_k$ or $Surv(start, stop, Cens) \sim factor_1 + \dots + factor_k$, where $factor_i$, $i = 1, \dots, k$, is the name of factor used in the Cox-like regression. Interactions between factors is allowed. If no factors are used in the Cox regression the value of 1 stands on the right-hand of the formula. The records with NA values for factors used in both formulas are excluded from the analysis.

Remark 1. The concordance is evaluated using the function 'survConcordance' from the R-package 'survival'.

Remark 2. The mean contrasts, their CIs, and their p -values are calculated on the basis of the empirical distribution for contrasts constructed using 10^6 generations of the vector of parameters. It is assumed that this vector is normally distributed with known mean and covariance matrix calculated by parameter estimation.

Remark 3. In some cases (flat likelihood function, multimodality, etc.) the hessian cannot be correctly estimated and the function issues the following error message:

Error in ParNPHCox(formula.scale, formula.shape, cluster, dist, data = ...) :
hessian cannot be correctly calculated.

Change the model and try again.

It can occur if, for example, the data set is not large enough and the shape and scale parameters compete in searching the maximum likelihood function. In this case it is recommended to simplify the model excluding some factors from the formulas for

shape or scale Cox-regression and try to calculate the estimates again.

Remark 4. The number of factors can increase after conversion of any categorical factor in binary one. For example, if factor "type" has three levels "A", "B", and "C" we will get after conversion two binary factors - "typeB" and "typeC" (factor "typeA" is the baseline one and does not appear in the list of the binary factors). The command "strata=list(...)" must include at most one non-zero level for each non-baseline level of the converted categorical factor. For example, it is correct to write "strata=list(typeB=1,typeC=0,...)" or briefly "strata=list(typeB=1,...)" and it is not correct to write "strata=list(typeB=1,typeC=1,...)".

Remark 5. If the covariate is a numerical one but we want to consider it as a categorical one we can convert it in categorical variable using the function "as.factor(variable)".

Example 1

```
library(ucminf)
library(MASS)
library(xtable)
library(survival)
data("lung",package="survival")
lung$sex=lung$sex-1
lung$status=lung$status-1
lung$ph.ecog=as.factor(lung$ph.ecog)
formula.scale=as.formula('Surv(time, status) ~ age + sex*ph.ecog')
formula.shape=as.formula('Surv(time, status) ~ 1')
cluster='ph.karno'
dist='Weibull'
expr=expression(a -500* b, log(msurv(19)/msurv(14))-1, age.scale,sex.scale)
NamFact(lung,formula.scale,formula.shape)
[1] "age" "ph.ecog1" "ph.ecog2" "ph.ecog3" "sex" "sex:ph.ecog1" "sex:ph.ecog2"
"sex:ph.ecog3"
strata=list(age=0,sex=1,ph.ecog1=1)
c(par,se,LogLik,Tab,Names,Conc,pval,p.contrast,pstrata):=
ParNPHCox(formula.scale,formula.shape,cluster,dist,data=lung,expr,strata)
```

	Estimates	CI	p-value
Sample size	226		
Number of non-censored	163		
a	719.069	291.56 - 1771.594	0
b	1.372	1.215 - 1.549	0
exp(age.scale)	1.009	0.99 - 1.028	0.342
exp(sex.scale)	0.524	0.247 - 1.111	0.0922
exp(ph.ecog1.scale)	1.457	0.921 - 2.304	0.1075
exp(ph.ecog2.scale)	2.282	1.342 - 3.881	0.0024
exp(ph.ecog3.scale)	6.298	0.843 - 47.295	0.0728
exp(sex:ph.ecog1.scale)	1.097	0.456 - 2.641	0.8363
exp(sex:ph.ecog2.scale)	1.249	0.463 - 3.373	0.6605
exp(sex:ph.ecog3.scale)	1	1 - 1	0.9983
Sigma2	0	0 - 0	0
Concordance (se)	0.846 (0.026)		
Loglik	-1126.33		
AIC	2274.65		

Table 1: Parameter estimates. Weibull model.

	contrast	CI	p-value
a - 500 * b	111.8866	-410.8308-1091.1863	0.92
log(msurv(19)/msurv(14)) - 1	0.6556	0.2013-1.2754	0.0016
age.scale	0.0091	-0.0097-0.028	0.342
sex.scale	-0.6464	-1.3992-0.1054	0.0922
exp(age.scale)	1.0092	0.9903-1.0284	0

Table 2: Table of contrasts. Weibull model.

Example 2

```
formula.scale=as.formula('Surv(time, status) ~ sex')
formula.shape=as.formula('Surv(time, status) ~ sex')
cluster='pat.karno'
dist='Gompertz'
expr=expression(a - b, msurv(19) - msurv(14), sex.scale + sex.shape)
NamFact(lung,formula.scale,formula.shape)
[1] "sex"
strata=list(sex=1)
c(par,se,LogLik,Tab,Names,Conc,pval,p.contrast,pstrata):=
ParNPHCox(formula.scale,formula.shape,cluster,dist,data=lung,expr,strata)
```

	Estimates	CI	p-value
Sample size	225		
Number of non-censored	162		
1000a	2.353	1.609 - 3.444	0
100b	0.112	0.052 - 0.242	0
exp(sex.shape)	2.199	0.886 - 5.468	0.089
exp(sex.scale)	0.411	0.232 - 0.726	0.0021
σ^2	0.062	0.006 - 0.594	0
Concordance (se)	0.9 (0.026)		
Loglik	-1128.5		
AIC	2267		

Table 3: Parameter estimates. Gompertz model.

	contrast	CI	p-value
a - b	0.0012	-6e-04-0.0027	0.1585
msurv(19) - msurv(14)	0.097	0.0475-0.1531	2e-04
sex.scale + sex.shape	-0.1016	-0.7357-0.5336	0.7548

Table 4: Table of contrasts. Gompertz model.