Package ‘frailtyHL’

October 13, 2022

Type Package
Title Frailty Models via Hierarchical Likelihood
Version 2.3
Date 2019-11-19
Author Il Do Ha, Maengseok Noh, Jiwoong Kim, Youngjo Lee
Maintainer Maengseok Noh <msnoh@pknu.ac.kr>
Description Implements the h-likelihood estimation procedures for general frailty models including competing-risk models and joint models.
Depends R (>= 3.5.0), methods, Matrix, survival, cmprsk
License Unlimited
NeedsCompilation no
Repository CRAN
Date/Publication 2019-11-22 09:00:05 UTC

R topics documented:

frailtyHL-package .................................................. 2
bladder .............................................................. 3
bladder0 ........................................................... 4
cgd ................................................................. 4
CmpRsk ............................................................ 5
frailty.vs ........................................................... 6
frailtyHL ........................................................... 7
hlike.frailty ......................................................... 9
jmfit ............................................................... 9
jointmodeling ..................................................... 10
kidney .............................................................. 11
mlmfit ............................................................ 12
rats ................................................................. 12
ren ................................................................. 13
renal .............................................................. 14
test ............................................................... 15
Description

The frailtyHL package fits frailty models which are Cox’s proportional hazards models incorporating random effects. The function implements the h-likelihood estimation procedures. For the frailty distribution lognormal and gamma are allowed. The h-likelihood uses the Laplace approximation when the numerical integration is intractable, giving a statistically efficient estimation in frailty models. (Ha, Lee and Song, 2001; Ha and Lee, 2003, 2005; Lee, Nelder and Pawitan, 2017; Ha, Jeong and Lee, 2017). This package handles various random-effect survival models such as time-dependent frailties, competing-risk frailty models, AFT random-effect models, and joint modelling of linear mixed models and frailty models. It also provides penalized variable-selection procedures (LASSO, SCAD and HL).

Details

<table>
<thead>
<tr>
<th>Package:</th>
<th>frailtyHL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type:</td>
<td>Package</td>
</tr>
<tr>
<td>Version:</td>
<td>2.1</td>
</tr>
<tr>
<td>Date:</td>
<td>2016-09-19</td>
</tr>
<tr>
<td>License:</td>
<td>Unlimited</td>
</tr>
<tr>
<td>LazyLoad:</td>
<td>yes</td>
</tr>
</tbody>
</table>

This is version 2.2 of the frailtyHL package.

Author(s)

Il Do Ha, Maengseok Noh, Jiwoong Kim, Youngjo Lee
Maintainer: Maengseok Noh <msnoh@pknu.ac.kr>

References


**Examples**

```r
data(kidney)
kidney_g12 <- frailtyHL(Surv(time, status) ~ sex + age + (1|id), kidney)
```

---

### Bladder Cancer Data

#### Description

Bladder is an extension of Bladder0 to competing risks with 396 patients with bladder cancer from 21 centers, focusing on two competing endpoints, i.e., time to first bladder recurrence (an event of interest; Type 1 event) and time to death prior to recurrence (competing event; Type 2 event).

#### Usage

```r
data("bladder")
```

#### Format

A data frame with 396 observations on the following 13 variables.

- **OBS**: Observation number
- **center**: Institution number of 24 centers
- **surtime**: Time to event
- **status**: Event indicator (1 = recurrence, 2 = death before recurrence, 0 = no event)
- **CHEMO**: Treatment indicator representing chemotherapy (0 = No, 1 = Yes)
- **AGE**: Age (0 = <= 65 years; 1 = > 65 years)
- **SEX**: Sex (0 = male, 1 = female)
- **PRIORREC**: Prior recurrent rate (0 = primary; 1, <= 1/yr; 2, > 1/yr)
- **NOTUM**: Number of tumors (0 = single; 1, 2-7 tumors; 2, >= 8 tumors)
- **TUM3CM**: Tumor size (0 = < 3cm; 1, >= cm)
- **TLOCC**: T category (0 = Ta, 1 = T1)
- **CIS**: Carcinoma in situ (0 = No, 1 = Yes)
- **GLOCAL**: G grade (0 = G1, 1 = G2, 2 = G3)

#### References


### bladder0

**Bladder cancer data**

**Description**

Bladder0 is a subset of 410 patients from a full data set with bladder cancer from 21 centers that participated in the EORTC trial (Sylvester et al., 2006). Time to event is the duration of the disease free interval (DFI), which is defined as time from randomization to the date of the first recurrence.

**Usage**

```r
data("bladder0")
```

**Format**

A data frame with 410 observations on the following 5 variables.

- **Center**  Institution number of 24 centers
- **Surtime** Time to the first recurrence from randomization
- **Status**  Censoring indicator (1=recurrence, 0=no event)
- **Chemo** Treatment indicator representing chemotherapy (0=No, 1=Yes)
- **Tustat** Indicator representing prior recurrent rate (0=Primary, 1=Recurrent)

**References**


### cgd

**Chronic Granulomatous Disease (CGD) Infection Data**

**Description**

The CGD data set in Fleming and Harrington (1991) is from a placebo-controlled randomized trial of gamma interferon in chronic granulomatous disease. In total, 128 patients from 13 hospitals were followed for about 1 year. The number of patients per hospital ranged from 4 to 26. Each patient may experience more than one infection. The survival times (times-to-event) are the times between recurrent CGD infections on each patient (i.e. gap times). Censoring occurred at the last observation for all patients, except one, who experienced a serious infection on the date he left the study.
CmpRsk

Usage

data("cgd")

Format

A data frame with 203 observations on the following 16 variables.

- **id**: Patient number for 128 patients
- **center**: Enrolling center number for 13 hospitals
- **random**: Date of randomization
- **treat**: Gamma-interferon treatment (rIFN-g) or placebo (Placebo)
- **sex**: Sex of each patient (male, female)
- **age**: Age of each patient at study entry, in years
- **height**: Height of each patient at study entry, in cm
- **weight**: Weight of each patient at study entry, in kg
- **inherit**: Pattern of inheritance (autosomal recessive, X-linked)
- **steroids**: Using corticosteroids at times of study entry (1=Yes, 0=No)
- **propylac**: Using prophylactic antibiotics at time of study entry (1=Yes, 0=No)
- **hos.cat**: A categorization of the hospital region into 4 groups
- **tstart**: Start of each time interval
- **enum**: Sequence number. For each patient, the infection records are in sequence number order
- **tstop**: End of each time interval
- **status**: Censoring indicator (1=uncensored, 0=censored)

References


---

CmpRsk

*Model Formula of Competing Risk*

Description

A CmpRsk object is used as the response variable in the model formula. It is created using the function CmpRsk(time, index), where time is the event time and index is an event indicator.

Usage

CmpRsk(time, index)
Arguments

time    the event time
index   the event indicator; values of index must be sequential whole numbers where 0 denotes right censoring and positive numbers refer to different event types.

Description

frailty.vs is variable-selection procedures (LASSO, SCAD and HL) of fixed effects in frailty models.

Usage

frailty.vs(formula, model, penalty, data, B = NULL, v = NULL, alpha = NULL, tun1 = NULL, tun2 = NULL, varfixed = FALSE, varinit = 0.1)

Arguments

formula   A formula object, with the response on the left of a ~ operator, and the terms for the fixed and random effects on the right. e.g. formula=Surv(time,status)~x+(1|id), time : survival time, status : censoring indicator having 1 (0) for uncensored (censored) observation, x : fixed covariate, id : random effect.
model     Log-normal frailty models ("lognorm")
penalty   Penalty functions ("LASSO" or "SCAD" or "HL")
data      Dataframe used
B          Initial values of fixed effects
v          Initial values of random effects. Zeros are default
alpha     Initial value of variance of random effects.
tun1      Tuning parameter gamma for LASSO, SCAD and HL
tun2      Tuning parameter omega for HL
varfixed  Logical value: if TRUE (FALSE), the value of one or more of the variance terms for the frailties is fixed (estimated).
varinit   Starting values for frailties, the default is 0.1.
Description

frailtyHL is used to fit frailty models using h-likelihood estimation procedures. For the frailty distribution, lognormal and gamma are allowed. In particular, nested (multilevel) frailty models allow survival studies for hierarchically clustered data by including two iid normal random effects. The h-likelihood uses the Laplace approximation when the numerical integration is intractable, giving a statistically efficient estimation in frailty models (Ha, Lee and Song, 2001; Ha and Lee, 2003, 2005; Lee, Nelder and Pawitan, 2017).

Usage

frailtyHL(formula, data, weights, subset, na.action, RandDist = "Normal", mord = 0, dord = 1, Maxiter = 200, convergence = 10^-6, varfixed = FALSE, varinit = c(0.163), varnonneg = FALSE)

Arguments

formula A formula object, with the response on the left of a ~ operator, and the terms for the fixed and random effects on the right. e.g. formula=Surv(time,status)~x+(1|id), time : survival time, status : censoring indicator having 1 (0) for uncensored (censored) observation, x : fixed covariate, id : random effect.
data Dataframe for formulaMain.
weights Vector of case weights.
subset Expression indicating which subset of the rows of data should be used in the fit. All observations are included by default.
na.action A missing-data filter function.
RandDist Distribution for random effect ("Normal" or "Gamma").
mord The order of Laplace approximation to fit the mean parameters (0 or 1); default=0.
dord The order of Laplace approximation to fit the dispersion components (1 or 2); default=1.
Maxiter The maximum number of iterations; default=200.
convergence Specify the convergence criterion, the default is 1e-6.
varfixed Logical value: if TRUE (FALSE), the value of one or more of the variance terms for the frailties is fixed (estimated).
varinit Starting values for frailties, the default is 0.1.
varnonneg Logical value: if TRUE (FALSE), gives zero (NaN) SE for random effects when they are estimated by zeros.
Details

frailtyHL package produces estimates of fixed effects and frailty parameters as well as their standard errors. Also, frailtyHL makes it possible to fit models where the frailty distribution is normal and gamma and estimate variance components when frailty structure is allowed to be shared or nested.

References


Examples

#### Analysis of kidney data
data(kidney)

#### Normal frailty model using order = 0, 1 for the mean and dispersion
kidney_ln01<-frailtyHL(Surv(time,status)-sex+age+(1|id),kidney,
RandDist="Normal",mord=0,dord=1)

#### Gamma frailty model using order = 1, 2 for the mean and dispersion
#kidney_g12<-frailtyHL(Surv(time,status)-sex+age+(1|id),kidney,
#RandDist="Gamma",mord=1,dord=2)

#### Analysis of rats data
data(rats)

#### Cox model
rat_cox<-frailtyHL(Surv(time,status)-rx+(1|litter),rats,
varfixed=TRUE,varinit=c(0))

#### Normal frailty model using order = 1, 1 for the mean and dispersion
#rat_ln11<-frailtyHL(Surv(time,status)-rx+(1|litter),rats,
#RandDist="Normal",mord=1,dord=1)

#### Gamma frailty model using order = 1, 2 for the mean and dispersion
#rat_g12<-frailtyHL(Surv(time,status)-rx+(1|litter),rats,
#RandDist="Gamma",mord=1,dord=2)

#### Analysis of CGD data
data(cgd)

#### Multilevel normal frailty model using order = 1, 1 for the mean and dispersion
#cgd_ln11<-frailtyHL(Surv(tstop-tstart,status)-treat+(1|center)+(1|id),cgd,
#RandDist="Normal",mord=1,dord=1,convergence=10^-4, varinit=c(0.03,1.0))


**hlike.frailty**

*Competing Risk Frailty Models using H-Likelihood*

**Description**

Perform hierarchical likelihood estimation of the univariate frailty model, cause-specific frailty model and subhazard frailty model. Assuming either a univariate normal or multivariate normal distribution for the random effects $V$, where different covariance structures can be assumed for the multivariate normal distribution.

**Usage**

```r
hlike.frailty(formula, data, inits, order = 1, frailty.cov = "none", subHazard = FALSE, alpha = 0.05, MAX.ITER = 100, TOL = 1e-06)
```

**Arguments**

- `formula`: left-hand side is a CmpRsk object (see details), right-hand side is predictors (currently limited to numeric main effects), must include a cluster term that identifies the cluster variable.
- `data`: dataframe containing the variables used in the formula.
- `inits`: list of initial values, three named components: beta, v and theta.
- `order`: numeric, order of the Laplace approximation, 0=no order, 1=first-order, 2=second-order; second-order only applies to models with a univariate normal distribution.
- `frailty.cov`: character string "none", "independent" or "unstructured" specifying the covariance structure for a multivariate normal distribution; "none" indicates univariate normal distribution.
- `subHazard`: logical, if TRUE fits the subhazard frailty model.
- `alpha`: numeric, 100(1-alpha) percent confidence intervals.
- `MAX.ITER`: numeric, maximum number of iterations.
- `TOL`: numeric, tolerance limit.

---

**jmfit**

*Joint Modelling of Longitudinal and Time-to-Event Data*

**Description**

jmfit is used to fit joint modelling of longitudinal and time-to-event data by using h-likelihood. The response of interest would involve repeated measurements over time on the same subject as well as time to an event of interest with or without competing risks.

**Usage**

```r
jmfit(jm, data, jm2 = NULL, data2 = NULL, Maxiter)
```
jointmodeling

Arguments

jm list of jointmodeling objects which specify the first responses of interest.
data list of dataframes containing the variables used in the jm.
jm2 list of jointmodeling object which specifies the second responses.
data2 dataframes containing the variables used in the jm2.
Maxiter numeric, maximum number of iterations

Description

The jointmodeling specifies jointly both the hazard model in the frailty model and the mean model in the linear mixed model.

Usage

jointmodeling(Model = "mean", RespDist = "gaussian", Link = NULL, LinPred = "constant", RandDist = NULL, Offset = NULL)

Arguments

Model This option specifies the mean model when Model="mean" (default).
RespDist This option specifies the distribution of response variables (linear mixed model: "gaussian" or accelerated failure time model: "AFT" or frailty model: "FM")
Link The link function for the linear predictor is specified by the option Link. For "AFT" or "FM" (or "gaussian") in RespDist, it is specified by "log" (or "identity").
LinPred The option LinPred specifies the fixed and random terms for the linear predictor.
RandDist The option RandDist specifies the distributions of the random terms represented in the option LinPred.
Offset The option Offset can be used to specify a known component to be included in the linear predictor specified by LinPred during fitting.
Kidney Infection Data

Description

The data presented by McGilchrist and Aisbett (1991) consist of times to the first and second recurrences of infection in 38 kidney patients using a portable dialysis machine. Infections can occur at the location of insertion of the catheter. The catheter is later removed if infection occurs and can be removed for other reasons, in which case the observation is censored.

Usage

data("kidney")

Format

A data frame with 76 observations on the following 10 variables.

- id  Patient number for 38 patients
- time  Time to infection since insertion of the catheter
- status  Censoring indicator (1=uncensored, 0=censored)
- age  Age of each patient, in years
- sex  Sex of each patient (1=male, 2=female)
- disease  Disease type (GN, AN, PKD, Other)
- frail  Frailty estimate from original paper
- GN  Indicator for disease type GN
- AN  Indicator for disease type AN
- PKD  Indicator for disease type PKD

References


mlmfit

**Accelerated Failure Time (AFT) Models with Random Effects**

**Description**
mlmfit is used to fit linear mixed models with censoring by using h-likelihood.

**Usage**

```
mlmfit(jm1, data, weights, subset, na.action, Maxiter = 200)
```

**Arguments**

- **jm1**: This option requires jointmodeling object which specifies the AFT random-effect model.
- **data**: dataframe containing the variables used in the jm1
- **weights**: Vector of case weights.
- **subset**: Expression indicating which subset of the rows of data should be used in the fit. All observations are included by default.
- **na.action**: A missing-data filter function.
- **Maxiter**: numeric, maximum number of iterations

---

**Rats data**

**Description**
Rats data set presented by Mantel et al. (1977) is based on a tumorigenesis study of 50 litters of female rats. For each litter, one rat was selected to receive the drug and the other two rats were placebo-treated controls. The survival time is the time to the development of tumor, measured in weeks. Death before occurrence of tumor yields a right-censored observation; 40 rats developed a tumor, leading to censoring of about 73 percent.

**Usage**

```
data("rats")
```

**Format**
A data frame with 150 observations on the following 4 variables.

- **litter**: Litter number for 50 female rats
- **rx**: Treatment(1=drug, 0=placebo)
- **time**: Time to the development of tumor in weeks
- **status**: Censoring indicator(1=uncensored, 0=censored)
Mammary tumor data

Description
The data set by presented Gail et al. (1980) is based on multiple occurrences of mammary tumors for 48 female rats. The primary outcome of interest was time to development of a mammary tumor for 23 female rats in the treatment group and 25 female rats in the control group. Initially, 76 rats were injected with a carcinogen for mammary cancer at day zero, and then all rats were given retinyl acetate to prevent cancer for 60 days. After 60 days, forty-eight rats which remained tumor-free were randomly assigned to continue being treated with retinoid prophylaxis (treatment group) or to the control group receiving no further retinoid prophylaxis. Rats were palpated for tumors twice weekly and observation ended 182 days after the initial carcinogen injection. In some cases, there were multiple tumors detected by the same day. The number of tumors ranges from 0 to 13.

Usage

data("ren")

Format
A data frame with 254 observations on the following 6 variables.

rat  Rat id
time1  Start time
time2  Stop time
del  Censoring indicator(1=tumor, 0=censored)
gp  Treatment indicator(1=drug, 0=control)
time  time2-time1 (time=time+0.01 if there are ties)

References
Renal transplant data

Description

This is a data set from a clinical study to investigate the chronic renal allograft dysfunction in renal transplants (Ha et al., 2017). Data were available from 87 male and 25 female renal transplanted patients who survived more than 4 years after transplant. For each patient, both repeated-measure outcomes (serum creatinine levels) at several time points and a terminating event time (graft-loss time) were observed.

Usage

data("renal")

Format

A data frame with 1395 observations on the following 9 variables.

- id: Patient id
- month: Time points (month) at which the measurements of sCr were recorded
- cr: Serum creatinine (sCr) level
- sex: Sex(1=male, 0=female)
- age: Age(years)
- icr: Reciprocal of sCr(=1/sCr)
- sur_time: Time to graft loss
- status: Censoring indicator(1=graft loss, 0=no event)
- first: The first survival time (time to graft loss) of each patient

References


Simulated data with clustered competing risks

Description

A data set for the cause-specific hazard frailty model assuming a bivariate normal distribution is generated using a technique similar to Beyersmann et al. (2009) and Christian et al. (2016). Let there be two event types, Types 1 and 2, as well as independent censoring. Consider a sample size \( n = 100 \) with \( (q, n_i) = (50, 3) \). Here, \( q \) is the number of clusters and \( n_i \) is the cluster size. The random effects (log-frailties) are from bivariate normal with mean vector \((0, 0)\) and variance-covariance matrix having \((1,1,-0.5)\). Data are generated from the conditional cause-specific hazard rates for each event type given the random effects. Here, for Type 1 event the two true regression parameters are \((0.6, -0.4)\) with a constant baseline hazard 2 and for Type 2 event the true parameters are \((-0.3, 0.7)\) with a constant baseline hazard 0.5, respectively. The covariates \( x_1 \) and \( x_2 \) are generated from a standard normal distribution and a Bernoulli distribution with probability 0.5, respectively. Censoring times are generated from a Uniform(0, 1.3) distribution. Under this scenario, with 25.2% censoring, the proportions of Type 1 and Type 2 events are 53.2% and 21.6%, respectively.

Usage

```R
data(“test”)
```

Format

A data frame with 250 observations on the following 6 variables.

- **obs**: Observation number
- **id**: Id number
- **time**: Time to event
- **status**: Event indicator(2=Type 2 event, 1=Type 1 event, 0=censored)
- **x1**: A covariate from standard normal distribution
- **x2**: A covariate from Bernoulli normal distribution

References


Index

bladder, 3
bladder0, 4

cgd, 4
CmpRsk, 5

frailty.vs, 6
frailtyHL, 7
frailtyHL-package, 2

hlike.frailty, 9

jmfit, 9
jointmodeling, 10

kidney, 11

mlmfit, 12

rats, 12
ren, 13
renal, 14

test, 15