Package ‘frailtypack’

January 16, 2020

Version 3.1.0

Title General Frailty Models: Shared, Joint and Nested Frailty Models with Prediction; Evaluation of Failure-Time Surrogate Endpoints

Author Virginie Rondeau, Juan R. Gonzalez, Yassin Mazroui, Audrey Mauguen, Amadou Diakite, Alexandre Laurent, Myriam Lopez, Agnieszka Krol, Casimir L. Sofeu, Julien Dumerc, Denis Rustand

Maintainer Virginie Rondeau <Virginie.Rondeau@inserm.fr>

Depends R (>= 2.10), survival, boot, MASS, survC1, doBy

Imports statmod, nlme, shiny,
        shinyjs, shinyBS, shinydashboard, rhandsontable, shinythemes, jsonlite

LazyLoad no

Description The following several classes of frailty models using a penalized likelihood estimation on the hazard function but also a parametric estimation can be fit using this R package:
1) A shared frailty model (with gamma or log-normal frailty distribution) and Cox proportional hazard model. Clustered and recurrent survival times can be studied.
2) Additive frailty models for proportional hazard models with two correlated random effects (intercept random effect with random slope).
3) Nested frailty models for hierarchically clustered data (with 2 levels of clustering) by including two iid gamma random effects.
4) Joint frailty models in the context of the joint modelling for recurrent events with terminal event data for clustered data or not. A joint frailty model for two semi-competing risks and clustered data is also proposed.
5) Joint general frailty models in the context of the joint modelling for recurrent events with terminal event data with two independent frailty terms.
6) Joint Nested frailty models in the context of the joint modelling for recurrent events with terminal event, for hierarchically clustered data (with two levels of clustering) by including two iid gamma random effects.
7) Multivariate joint frailty models for two types of recurrent events and a terminal event.
8) Joint models for longitudinal data and a terminal event.
9) Trivariate joint models for longitudinal data, recurrent events and a terminal event.
10) Joint frailty models for the validation of surrogate endpoints in multiple randomized clinical trials with failure-time endpoints
11) Two-part joint model for longitudinal semicontinuous data and a terminal event.
Prediction values are available (for a terminal event or for a new recurrent event). Left-truncated (not for Joint model), right-censored data, interval-censored data (only for Cox proportional hazard and shared frailty model) and strata are allowed. In each model, the random effects have the gamma or normal distribution. Now, you can also consider time-varying covariates effects in Cox, shared and joint frailty models (1-5). The package includes concordance measures for Cox proportional hazards models and for shared frailty models. Moreover, the package can be used with its shiny application, in a local mode or by following the link below.

License  GPL (>= 2.0)

URL  https://virginierondeau.wixsite.com/virginierondeau/software-frailtypack
     https://frailtypack-pkg.shinyapps.io/shiny_frailtypack

Suggests  knitr, rmarkdown, testthat

VignetteBuilder  knitr

Repository  CRAN

Date/Publication  2020-01-16 13:20:13 UTC

NeedsCompilation  yes

RoxygenNote  6.1.1
R topics documented:

- multivPenal
- num.id
- plot.additivePenal
- plot.Diffepoce
- plot.epoce
- plot.frailtyPenal
- plot.jointNestedPenal
- plot.jointPenal
- plot.jointSurroPenal
- plot.longiPenal
- plot.multivPenal
- plot.nestedPenal
- plot.predFrailty
- plot.predJoint
- plot.predLongi
- plot.trivPenal
- plot.trivPenalNL
- predict.jointSurroPenal
- prediction
- print.additivePenal
- print.Cmeasures
- print.frailtyPenal
- print.jointNestedPenal
- print.jointPenal
- print.longiPenal
- print.multivPenal
- print.nestedPenal
- print.prediction
- print.trivPenal
- print.trivPenalNL
- readmission
- runShiny
- slope
- ste
- subcluster
- summary.additivePenal
- summary.frailtyPenal
- summary.jointNestedPenal
- summary.jointPenal
- summary.jointSurroPenal
- summary.jointSurroPenalSimul
- summary.longiPenal
- summary.multivPenal
- summary.nestedPenal
- summary.trivPenal
- summary.trivPenalNL
- survDat
- SurvIC

Page dimensions: 612.0x792.0
General Frailty models: shared, joint and nested frailty models with prediction; Evaluation of Failure-Time Surrogate Endpoints

Description

Frailtypack fits several classes of frailty models using a penalized likelihood estimation on the hazard function but also a parametric estimation. 1) A shared frailty model and Cox proportional hazard model. Clustered and recurrent survival times can be studied. 2) Additive frailty models for proportional hazard models with two correlated random effects (intercept random effect with random slope). 3) Nested frailty models for hierarchically clustered data (with 2 levels of clustering) by including two iid gamma random effects. 4) Joint frailty models in the context of joint modelling for recurrent events with terminal event for clustered data or not. A joint frailty model for two semi-competing risks for clustered data is also proposed. 5) Joint General frailty models in the context of a joint modelling for recurrent events with terminal event data with two independent frailty terms. 6) Joint Nested frailty models in the context of joint modelling for recurrent events with terminal event, for hierarchically clustered data (with two levels of clustering) by including two iid gamma random effects. 7) Multivariate joint frailty models for two types of recurrent events and a terminal event. 8) Joint models for longitudinal data and a terminal event. 9) Trivariate joint models for longitudinal data, recurrent events and a terminal event. Prediction values are available. Left truncated (not for the joint models), right-censored data, interval-censored data (only for Cox proportional hazard and shared frailty model) and strata are allowed. In each model, the random effects have the gamma or normal distribution. Now, you can also consider time-varying effect covariates in Cox, shared and joint frailty models. The package includes concordance measures for Cox proportional hazards models and for shared frailty models. 10) Joint frailty models for the validation of surrogate endpoints in multiple randomized clinical trials with failure-time endpoints. This model includes a shared individual-level random effect, a shared trial random-effect associated with the hazard risks and a correlated random effects-by-trial interaction.

Details

<table>
<thead>
<tr>
<th>Package</th>
<th>frailtypack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Package</td>
</tr>
<tr>
<td>Version</td>
<td>3.0.3.3</td>
</tr>
<tr>
<td>Date</td>
<td>2019-08-31</td>
</tr>
<tr>
<td>License</td>
<td>GPL (&gt;= 2.0)</td>
</tr>
<tr>
<td>LazyLoad</td>
<td>no</td>
</tr>
</tbody>
</table>
Author(s)

Virginie Rondeau, Juan R. Gonzalez, Yassin Mazroui, Audrey Mauguen, Amadou Diakite, Alexandre Laurent, Myriam Lopez, Agnieszka Krol and Casimir L. Sofeu

References


Examples

```r
## Not run:
###--- Additive model with 1 covariate ---###

data(dataAdditive)
modAdd <- additivePenal(Surv(t1,t2,event)~
cluster(group)+var1+slope(var1),
correlation=TRUE,data=dataAdditive,
n.knots=8,kappa=10000,hazard="Splines")

###--- Joint model (recurrent and terminal events) with 2 covariates ---###

data(readmission)
```
###--- General Joint model (recurrent and terminal events) with 2 covariates ---###

```r
modJoint.gap <- frailtyPenal(Surv(time,event)~
cluster(id)+sex+dukes+charlson+terminal(death),
formula.terminalEvent=~sex+dukes+charlson,
data=readmission,n.knots=10,kappa=c(100,100),
recurrantAG=FALSE,hazard="Splines")
```

###--- Nested model (or hierarchical model) with 2 covariates ---###

```r
data(readmission)
modJoint.general <- frailtyPenal(Surv(time,event) ~ cluster(id) + dukes +
charlson + sex + chemo + terminal(death),
formula.terminalEvent = ~ dukes + charlson + sex + chemo,
data = readmission, jointGeneral = TRUE, n.knots = 8,
kappa = c(2.11e+08, 9.53e+11))
```

###--- Joint Nested Frailty model ---###

```r
#-- here is generated cluster (30 clusters)
readmissionNested <- transform(readmission,group=id%%30+1)

modJointNested_Splines <- frailtyPenal(formula = Surv(t.start, t.stop, event)
~ subcluster(id) + cluster(group) + dukes + terminal(death),
formula.terminalEvent = ~ dukes, data = readmissionNested, recurrentAG = TRUE,
n.knots = 8, kappa = c(9.55e+9, 1.41e+12), initialize = TRUE)

modJointNested_Weib <- frailtyPenal(Surv(t.start,t.stop,event)~subcluster(id)
+cluster(group)+dukes+ terminal(death),formula.terminalEvent=~dukes,
hazard = ('Weibull'), data=readmissionNested,recurrentAG=TRUE, initialize = FALSE)

Joint-GapSpline <- frailtyPenal(formula = Surv(time, event)
~ subcluster(id) + cluster(group) + dukes + terminal(death),
formula.terminalEvent = ~ dukes, data = readmissionNested, recurrentAG = FALSE,
init.Alpha = 1.091, Ksi = "None")
```

###--- Semiparametric Shared model ---###

```r
data(readmission)
sha.sp <- frailtyPenal(Surv(t.start,t.stop,event)~
sex+dukes+charlson+cluster(id),data=readmission,
n.knots=6,kappa=5000,recurrentAG=TRUE, 
cross.validation=TRUE,hazard="Splines")
```

###--- Parametric Shared model ---###

```r
data(readmission)
sha.p <- frailtyPenal(Surv(t.start,t.stop,event)~
```
cluster(id)+sex+dukes+charlson,
data=readmission,recurrentAG=TRUE,
hazard="Piecewise-per",nb.int=6)

###--- Joint model for longitudinal ---###
###--- data and a terminal event ---###
data(colorectal)
data(colorectallongi)

# Survival data preparation - only terminal events
colorectalSurv <- subset(colorectal, new.lesions == 0)

model.weib.RE <- longiPenal(Surv(time1, state) ~ age + treatment + who.PS + prev.resection, tumor.size ~ year * treatment + age + who.PS, colorectalSurv,data.Longi = colorectallongi, random = c("1", "year"), id = "id", link = "Random-effects", left.censoring = -3.33, hazard = "Weibull")

###--- Trivariate joint model for longitudinal ---###
###--- data, recurrent and terminal events ---###
data(colorectal)
data(colorectallongi)

# (computation takes around 40 minutes)
model.spli.RE.cal <- trivPenal(Surv(time0, time1, new.lesions) ~ cluster(id) + age + treatment + who.PS + terminal(state), formula.terminalEvent =~ age + treatment + who.PS + prev.resection, tumor.size ~ year * treatment + age + who.PS, data = colorectal, data.Longi = colorectallongi, random = c("1", "year"), id = "id", link = "Random-effects", left.censoring = -3.33, recurrentAG = TRUE, n.knots = 6, kappa=c(0.01, 2), method.GH="Pseudo-adaptive", n.nodes=7, init.B = c(-0.07, -0.13, -0.16, -0.17, 0.42, #recurrent events covariates -0.23, -0.1, -0.09, -0.12, 0.8, -0.23, #terminal event covariates 3.02, -0.30, 0.05, -0.63, -0.02, -0.29, 0.11, 0.74)) #biomarker covariates

##---Surrogacy evaluation based on generations with a combination
##of Monte Carlo and classical Gaussian Hermite integration.
## (Computation takes around 5 minutes)
# Generation of data to use
data.sim <- jointSurrSimul(n.obs=600, n.trial = 30,cens.adm=549.24,
alpha = 1.5, theta = 3.5, gamma = 2.5, zeta = 1, sigma.s = 0.7, sigma.t = 0.7, rsqrt = 0.8, betas = -1.25, betat = -1.25, full.data = 0, random.generator = 1, seed = 0, nb.reject.data = 0)

# Joint surrogate model estimation
joint.surro.sim.MCGH <- jointSurroPenal(data = data.sim, int.method = 2, nb.mc = 300, nb.gh = 20)
additivePenal

Fit an Additive Frailty model using a semiparametric penalized likelihood estimation or a parametric estimation

Description

Fit an additive frailty model using a semiparametric penalized likelihood estimation or a parametric estimation. The main issue in a meta-analysis study is how to take into account the heterogeneity between trials and between the treatment effects across trials. Additive models are proportional hazard model with two correlated random trial effects that act either multiplicatively on the hazard function or in interaction with the treatment, which allows studying for instance meta-analysis or multicentric datasets. Right-censored data are allowed, but not the left-truncated data. A stratified analysis is possible (maximum number of strata = 2). This approach is different from the shared frailty models.

In an additive model, the hazard function for the \( j^{th} \) subject in the \( i^{th} \) trial with random trial effect \( u_i \) as well as the random treatment-by-trial interaction \( v_i \) is:

\[
\begin{align*}
\lambda_{ij}(t | u_i, v_i) &= \lambda_0(t) \exp(u_i + v_i X_{ij1} + \sum_{k=1}^{p} \beta_k X_{ijk}) \\
\text{cov}(u_i, v_i) &= \rho \sigma \tau \\
u_i &\sim \mathcal{N}(0, \sigma^2), v_i \sim \mathcal{N}(0, \tau^2)
\end{align*}
\]

where \( \lambda_0(t) \) is the baseline hazard function, \( \beta_k \) the fixed effect associated to the covariate \( X_{ijk} \) (k=1,...,p), \( \beta_1 \) is the treatment effect and \( X_{ij1} \) the treatment variable. \( \rho \) is the corresponding correlation coefficient for the two frailty terms.

Usage

```r
additivePenal(formula, data, correlation = FALSE, recurrentAG = FALSE, cross.validation = FALSE, n.knots, kappa, maxit = 350, hazard = "Splines", nb.int, LIMparam = 1e-4, LIMlogl = 1e-4, LIMderiv = 1e-3, print.times = TRUE)
```

Arguments

- `formula`: a formula object, with the response on the left of a \( \sim \) operator, and the terms on the right. The response must be a survival object as returned by the ‘Surv’ function like in survival package. The slope() function is required. Interactions are possible using * or :.
- `data`: a ’data.frame’ with the variables used in ’formula’.
- `correlation`: Logical value. Are the random effects correlated? If so, the correlation coefficient is estimated. The default is FALSE.
- `recurrentAG`: Always FALSE for additive models (left-truncated data are not allowed).
cross.validation
Logical value. Is cross validation procedure used for estimating smoothing parameter in the penalized likelihood estimation? If so a search of the smoothing parameter using cross validation is done, with kappa as the seed. The cross validation is not implemented for two strata. The default is FALSE.

n.knots
integer giving the number of knots to use. Value required in the penalized likelihood estimation. It corresponds to the (n.knots+2) splines functions for the approximation of the hazard or the survival functions. Number of knots must be between 4 and 20. (See Note)

kappa
positive smoothing parameter in the penalized likelihood estimation. In a stratified additive model, this argument must be a vector with kappas for both strata. The coefficient kappa of the integral of the squared second derivative of hazard function in the fit. To obtain an initial value for kappa, a solution is to fit the corresponding shared frailty model using cross validation (see cross.validation). We advise the user to identify several possible tuning parameters, note their defaults and look at the sensitivity of the results to varying them. Value required. (See Note)

maxit
maximum number of iterations for the Marquardt algorithm. Default is 350

hazard
Type of hazard functions: "Splines" for semiparametric hazard functions with the penalized likelihood estimation, "Piecewise-per" for piecewise constant hazards functions using percentile, "Piecewise-equ" for piecewise constant hazard functions using equidistant intervals, "Weibull" for parametric Weibull functions. Default is "Splines".

nb.int
Number of intervals (between 1 and 20) for the parametric hazard functions ("Piecewise-per", "Piecewise-equ").

LIMparam
Convergence threshold of the Marquardt algorithm for the parameters (see Details), $10^{-4}$ by default.

LIMlogl
Convergence threshold of the Marquardt algorithm for the log-likelihood (see Details), $10^{-4}$ by default.

LIMderiv
Convergence threshold of the Marquardt algorithm for the gradient (see Details), $10^{-3}$ by default.

print.times
a logical parameter to print iteration process. Default is TRUE.

Details
The estimated parameter are obtained by maximizing the penalized log-likelihood or by a simple log-likelihood (in the parametric case) using the robust Marquardt algorithm (Marquardt,1963). The parameters are initialized with values obtained with Cox proportional hazard model. The iterations are stopped when the difference between two consecutive loglikelihoods was small ($< 10^{-4}$), the estimated coefficients were stable (consecutive values ($< 10^{-4}$), and the gradient small enough ($< 10^{-3}$). To be sure of having a positive function at all stages of the algorithm, the spline coefficients were reparametrized to be positive at each stage. The variance space of the two random effects is reduced, so the variances are positive, and the correlation coefficient values are constrained to be between -1 and 1. The marginal log-likelihood depends on integrations that are approximated by using the Laplace integration technique with a first order approximation. The smoothing parameter can be fixed or estimated by maximizing likelihood cross-validation criterion. The usual
squared Wald statistic was modified to a mixture of two $\chi^2$ distribution to get significance test for the variance of the random effects.

**INITIAL VALUES**

The splines and the regression coefficients are initialized to 0.1. An adjusted Cox model is fitted, it provides new initial values for the splines coefficients and the regression coefficients. The variances of the frailties are initialized to 0.1. Then an additive frailty model with independent frailties is fitted. At last, an additive frailty model with correlated frailties is fitted.

**Value**

An additive model or more generally an object of class 'additivePenal'. Methods defined for 'additivePenal' objects are provided for print, plot and summary.

- **b** sequence of the corresponding estimation of the splines coefficients, the random effects variances and the regression coefficients.
- **call** The code used for fitting the model.
- **coef** the regression coefficients.
- **cov** covariance between the two frailty terms ($\text{cov}(u_i, v_i)$)
- **cross.Val** Logical value. Is cross validation procedure used for estimating the smoothing parameters in the penalized likelihood estimation?
- **correlation** Logical value. Are the random effects correlated?
- **DoF** degrees of freedom associated with the "kappa".
- **formula** the formula part of the code used for the model.
- **groups** the maximum number of groups used in the fit.
- **kappa** A vector with the smoothing parameters in the penalized likelihood estimation corresponding to each baseline function as components.
- **loglikPenal** the complete marginal penalized log-likelihood in the semiparametric case.
- **loglik** the marginal log-likelihood in the parametric case.
- **n** the number of observations used in the fit.
- **n.events** the number of events observed in the fit.
- **n.iter** number of iterations needed to converge.
- **n.knots** number of knots for estimating the baseline functions.
- **n.strat** number of stratum.
- **rho** the corresponding correlation coefficient for the two frailty terms.
- **sigma2** Variance for the random intercept (the random effect associated to the baseline hazard functions).
- **tau2** Variance for the random slope (the random effect associated to the treatment effect across trials).
- **varH** the variance matrix of all parameters before positivity constraint transformation (Sigma2, Tau2, the regression coefficients and the spline coefficients). Then after, the delta method is needed to obtain the estimated variance parameters.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>varHIH</td>
<td>The robust estimation of the variance matrix of all parameters (Sigma2, Tau2, the regression coefficients and the spline coefficients).</td>
</tr>
<tr>
<td>varSigma2</td>
<td>The variance of the estimates of &quot;sigma2&quot;.</td>
</tr>
<tr>
<td>varTau2</td>
<td>The variance of the estimates of &quot;tau2&quot;.</td>
</tr>
<tr>
<td>varcov</td>
<td>Variance of the estimates of &quot;cov&quot;.</td>
</tr>
<tr>
<td>x</td>
<td>matrix of times where both survival and hazard functions are estimated. By default seq(0,max(time),length=99), where time is the vector of survival times.</td>
</tr>
<tr>
<td>lam</td>
<td>array (dim=3) of hazard estimates and confidence bands.</td>
</tr>
<tr>
<td>surv</td>
<td>array (dim=3) of baseline survival estimates and confidence bands.</td>
</tr>
<tr>
<td>median</td>
<td>The value of the median survival and its confidence bands. If there are two stratas or more, the first value corresponds to the value for the first strata, etc.</td>
</tr>
<tr>
<td>type.of.hazard</td>
<td>Type of hazard functions (0:&quot;Splines&quot;, &quot;1:Piecewise&quot;, &quot;2:Weibull&quot;).</td>
</tr>
<tr>
<td>type.of.Piecewise</td>
<td>Type of Piecewise hazard functions (1:&quot;percentile&quot;, 0:&quot;equidistant&quot;).</td>
</tr>
<tr>
<td>nbintervR</td>
<td>Number of intervals (between 1 and 20) for the parametric hazard functions (&quot;Piecewise-per&quot;, &quot;Piecewise-equi&quot;).</td>
</tr>
<tr>
<td>npar</td>
<td>number of parameters.</td>
</tr>
<tr>
<td>nvar</td>
<td>number of explanatory variables.</td>
</tr>
<tr>
<td>noVar</td>
<td>indicator of explanatory variable.</td>
</tr>
<tr>
<td>LCV</td>
<td>the approximated likelihood cross-validation criterion in the semiparametric case (with H minus the converged Hessian matrix, and l(.) the full log-likelihood).</td>
</tr>
<tr>
<td>AIC</td>
<td>the Akaike information Criterion for the parametric case.</td>
</tr>
<tr>
<td>n.knots.temp</td>
<td>initial value for the number of knots.</td>
</tr>
<tr>
<td>shape.weib</td>
<td>shape parameter for the Weibull hazard function.</td>
</tr>
<tr>
<td>scale.weib</td>
<td>scale parameter for the Weibull hazard function.</td>
</tr>
<tr>
<td>martingale.res</td>
<td>martingale residuals for each cluster.</td>
</tr>
<tr>
<td>frailty.pred</td>
<td>empirical Bayes prediction of the first frailty term.</td>
</tr>
<tr>
<td>frailty.pred2</td>
<td>empirical Bayes prediction of the second frailty term.</td>
</tr>
<tr>
<td>linear.pred</td>
<td>linear predictor: uses simply &quot;Beta’X + u_i + v_i * X_1&quot; in the additive Frailty models.</td>
</tr>
<tr>
<td>global_chisq</td>
<td>a vector with the values of each multivariate Wald test.</td>
</tr>
<tr>
<td>dof_chisq</td>
<td>a vector with the degree of freedom for each multivariate Wald test.</td>
</tr>
<tr>
<td>global_chisq.test</td>
<td>a binary variable equals to 0 when no multivariate Wald is given, 1 otherwise.</td>
</tr>
<tr>
<td>p.global_chisq</td>
<td>a vector with the p_values for each global multivariate Wald test.</td>
</tr>
<tr>
<td>names.factor</td>
<td>Names of the “as.factor” variables.</td>
</tr>
<tr>
<td>Xlevels</td>
<td>vector of the values that factor might have taken.</td>
</tr>
<tr>
<td>contrasts</td>
<td>type of contrast for factor variable.</td>
</tr>
<tr>
<td>beta_p.value</td>
<td>p-values of the Wald test for the estimated regression coefficients.</td>
</tr>
</tbody>
</table>
Note

"kappa" and "n.knots" are the arguments that the user have to change if the fitted model does not converge. "n.knots" takes integer values between 4 and 20. But with n.knots=20, the model would take a long time to converge. So, usually, begin first with n.knots=7, and increase it step by step until it converges. "kappa" only takes positive values. So, choose a value for kappa (for instance 10000), and if it does not converge, multiply or divide this value by 10 or 5 until it converges.

References


See Also

slope

Examples

```r
## Not run:
###--- Additive model with 1 covariate ---###
data(dataAdditive)
modAdd <- additivePenal(Surv(t1,t2,event)~cluster(group)+
var1+slope(var1),correlation=TRUE,data=dataAdditive,
n.knots=8,kappa=10000)

#-- Var1 is boolean as a treatment variable

## End(Not run)
```
Description
The often used data set for interval-censored data, described and given in full in Finkelstein and Wolfe (1985). It involves 94 breast cancer patients who were randomized to either radiation therapy with chemotherapy or radiation therapy alone. The outcome is time until the onset of breast retraction which is interval-censored between the last clinic visit before the event was observed and the first visit when the event was observed. Patients without breast retraction were right-censored.

Usage
data(bcos)

Format
A data frame with 94 observations and 3 variables:
- left: left end point of the breast retraction interval
- right: right end point of the breast retraction interval
- treatment: type of treatment received

Source

---

cluster

Identify clusters

Description
This is a special function used in the context of the models for grouped data. It identifies correlated groups of observations defined by using 'cluster' function, and is used of 'frailtyPenal' formula for fitting univariate and joint models.

Usage
cluster(x)

Arguments
x A character, factor, or numeric variable which is supposed to indicate the variable group

Value
x A variable identified as a cluster
Cmeasures

Concordance measures in shared frailty and Cox proportional hazard models

Description

Compute concordance probability estimation for Cox proportional hazard or shared frailty models in case of grouped data (Mauguen et al. 2012). Concordance is given at different levels of comparison, taking into account the cluster membership: between-groups, within-groups and an overall measure, being a weighted average of the previous two. Can also compute the c-index (Harrell et al. 1996) at these three levels. It is possible to exclude tied pairs from concordance estimation (otherwise, account for 1/2).

Usage

Cmeasures(fitc, ties = 1, marginal = 0, cindex = 0, Nboot = 0, tau = 0, data.val)

Arguments

fitc A frailtyPenal object, for a shared frailty model. If the fit is a Cox model, no clustering membership is taken into account and only marginal concordance probability estimation is provided. Only an overall measure is given, where all patients are compared two by two. If a counting process formulation is used to performed the fit, with ’t.start’ and ’t.stop’, the gap-times (t.stop-t.start) are used in the concordance estimation.

See Also

frailtyPenal

Examples

## Not run:
data(readmission)
modSha <- frailtyPenal(Surv(time,event)~as.factor(dukes)+cluster(id),
n.knots=10,kappa=10000,data=readmission,hazard="Splines")

print(modSha)

## End(Not run)
| **ties** | Indicates if the tied pairs on prediction value must be included (ties=1) or excluded (ties=0) from the concordance estimation. Default is ties=1. When included, tied pairs account for 1/2 in the concordance. |
| **marginal** | Indicates if the concordance based on marginal predictions must be given (marginal=1) in addition to conditional ones or not (marginal=0). Marginal predictions do not include the frailty estimation in the linear predictor computation: uses "Beta'X" instead of "Beta'X + log z_i". Default is marginal=0. |
| **cindex** | Indicates if the c-index (Harrell et al. 1996) must be computed (cindex=1) in addition to the concordance probability estimation or not (cindex=0). C-index is also given at the three comparison levels (between, within and overall). Default is cindex=0. |
| **Nboot** | Number of bootstrap resamplings to compute standard-error of the concordances measures, as well as a percentile 95% confidence interval. Nboot=0 indicates no bootstrap procedure. Maximum admitted is 1000. Minimum admitted is 2. Default is 0. Resampling is done at the group level. If Cox model is used, resampling is done at individual level. |
| **tau** | Time used to limit the interval on which the concordance is estimated. Note that the survival function for the underlying censoring time distribution needs to be positive at tau. If tau=0, the maximum of the observed event times is used. Default is tau=0. |
| **data.val** | A dataframe. It is possible to specify a different dataset than the one used in the model input in the argument 'fitc'. This new dataset will be a validation population and the function will compute new concordance measures from the parameters estimated on the development population. In this case for conditional measures, the frailties are a posteriori predicted. The two datasets must have the same covariates with the same coding without missing data. |

**Value**

| **call** | The shared frailty model evaluated. |
| **Frailty** | Logical value. Was model with frailties fitted. |
| **frequencies** | Numbers of patients, events and groups used to fit the model. |
| **Npairs** | Number of pairs of subjects, between-groups, within-groups and over all the population. If cindex=1, number of comparable (useable) pairs also available. |
| **Nboot** | Number of bootstrap resamplings required. |
| **ties** | A binary, indicating if the tied pairs on prediction were used to compute the concordance. |
| **CPEcond** | Values of Gonen & Heller’s measure (conditional). If Nboot>0, give SE, the standard-error of the parameters evaluated by bootstrap, IC.low and IC.high, the lower and upper bounds of the percentile confidence interval evaluated by bootstrap (2.5% and 97.5% percentiles). |
| **Cunocond** | Values of Uno’s measure (conditional). If Nboot>0, give SE, the standard-error of the parameters evaluated by bootstrap, IC.low and IC.high, the lower and upper bounds of the percentile confidence interval evaluated by bootstrap (2.5% and 97.5% percentiles). |
Cmeasures

marginal A binary, indicating if the marginal values were computed.

CPEmarg Values of Gonen & Heller’s measure (marginal), if marginal=1. If Nboot>0, give SE, the standard-error of the parameters evaluated by bootstrap, IC.low and IC.high, the lower and upper bounds of the percentile confidence interval evaluated by bootstrap (2.5% and 97.5% percentiles).

Cunomarg Values of Uno’s measure (marginal), if marginal=1. If Nboot>0, give SE, the standard-error of the parameters evaluated by bootstrap, IC.low and IC.high, the lower and upper bounds of the percentile confidence interval evaluated by bootstrap (2.5% and 97.5% percentiles).

cindex A binary, indicating if the c-indexes were computed.

cindexcond Values of the C-index of Harrell (conditional). If Nboot>0, give SE, the standard-error of the parameters evaluated by bootstrap, IC.low and IC.high, the lower and upper bounds of the percentile confidence interval evaluated by bootstrap (2.5% and 97.5% percentiles).

cindexmarg Values of the C-index of Harrell (marginal), if marginal=1. If Nboot>0, give SE, the standard-error of the parameters evaluated by bootstrap, IC.low and IC.high, the lower and upper bounds of the percentile confidence interval evaluated by bootstrap (2.5% and 97.5% percentiles).

References


See Also

print.Cmeasures, frailtyPenal

Examples

```r
## Not run:

#-- load data
data(readmission)

#-- a frailtypenal fit
fit <- frailtyPenal(Surv(time,event)~cluster(id)+dukes+charlson+chemo,data=readmission,cross.validation=FALSE, n.knots=10,kappa=1,hazard="Splines")
```
Follow-up of metastatic colorectal cancer patients: times of new lesions appearance and death

Description

Randomly chosen 150 patients from the follow-up of the FFCD 2000-05 multicenter phase III clinical trial originally including 410 patients with metastatic colorectal cancer randomized into two therapeutic strategies: combination and sequential. The dataset contains times of observed appearances of new lesions censored by a terminal event (death or right-censoring) with baseline characteristics (treatment arm, age, WHO performance status and previous resection).

Usage

data(colorectal)

Format

This data frame contains the following columns:

- **id**: identification of each subject. Repeated for each recurrence
- **time0**: start of interval (0 or previous recurrence time)
- **time1**: recurrence or censoring time
- **new.lesions**: Appearance of new lesions status. 0: censored or no event, 1: new lesions
- **treatment**: To which treatment arm a patient was allocated? 1: sequential (S); 2: combination (C)
- **age**: Age at baseline: 1: <50 years, 2: 50-69 years, 3: >69 years
- **who.PS**: WHO performance status at baseline: 1: status 0, 2: status 1, 3: status 2
- **prev.resection**: Previous resection of the primate tumor? 0: No, 1: Yes
- **state**: death indicator. 0: alive, 1: dead
- **gap.time**: interoccurrence time or censoring time
Note

We thank the Federation Francophone de Cancerologie Digestive and Gustave Roussy for sharing the data of the FFCD 2000-05 trial supported by an unrestricted Grant from Sanofi.

References


**colorectalLongi**  
*Follow-up of metastatic colorectal cancer patients: longitudinal measurements of tumor size*

**Description**

Randomly chosen 150 patients from the follow-up of the FFCD 2000-05 multicenter phase III clinical trial originally including 410 patients with metastatic colorectal cancer randomized into two therapeutic strategies: combination and sequential. The dataset contains measurements of tumor size (left-censored sums of the longest diameters of target lesions; transformed using Box-Cox) with baseline characteristics (treatment arm, age, WHO performance status and previous resection).

**Usage**

data(colorectalLongi)

**Format**

This data frame contains the following columns:

- **id** identification of each subject. Repeated for each recurrence
- **year** time of visit counted in years from baseline
- **tumor.size** Individual longitudinal measurement of transformed (Box-Cox with parameter 0.3) sums of the longest diameters, left-censored due to a detection limit (threshold $s = -3.33$).
- **treatment** To which treatment arm a patient was allocated? 1: sequential (S); 2: combination (C)
- **age** Age at baseline: 1: <50 years, 2: 50-69 years, 3: >69 years
- **who.PS** WHO performance status at baseline: 1: status 0, 2: status 1, 3: status 2
- **prev.resection** Previous resection of the primate tumor? 0: No, 1: Yes

**Note**

We thank the Federation Francophone de Cancerologie Digestive and Gustave Roussy for sharing the data of the FFCD 2000-05 trial supported by an unrestricted Grant from Sanofi.
Simulated data as a gathering of clinical trials databases

Description

This contains simulated samples of 100 clusters with 100 subjects in each cluster, like a gathering of clinical trials databases. Two correlated centred gaussian random effects are generated with the same variance fixed at 0.3 and the covariance at -0.2. The regression coefficient $\beta$ is fixed at -0.11. The percentage of right-censored data is around 30 percent which are generated from a uniform distribution on [1,150]. The baseline hazard function is considered as a simple Weibull.

Usage

data(dataAdditive)

Format

This data frame contains the following columns:

- **group**: identification variable
- **t1**: start of interval (=0, because left-truncated data are not allowed)
- **t2**: end of interval (death or censoring time)
- **event**: censoring status (0:alive, 1:death, as a censoring indicator)
- **var1**: dichotomous covariate (=0 or 1, as a treatment variable)
- **var2**: dichotomous covariate (=0 or 1, as a treatment variable)

Source

**dataMultiv**

Simulated data for two types of recurrent events and a terminal event

**Description**

This contains a simulated sample of 800 subjects and 1652 observations. This dataset can be used to illustrate how to fit a joint multivariate frailty model. Two gaussian correlated random effects were generated with mean 0, variances 0.5 and a correlation coefficient equals to 0.5. The coefficients $\alpha_1$ and $\alpha_2$ were fixed to 1. The three baseline hazard functions followed a Weibull distribution and right censoring was fixed at 5.

**Usage**

```r
data(dataMultiv)
```

**Format**

This data frame contains the following columns:

- **PATIENT** identification of patient
- **obs** number of observation for a patient
- **TIME0** start of interval
- **TIME1** end of interval (death or censoring time)
- **INDICREC** recurrent of type 1 status (0:no, 1:yes)
- **INDICMETA** recurrent of type 2 status (0:no, 1:yes)
- **INDICDEATH** censoring status (0:alive, 1:death)
- **v1** dichotomous covariate (0,1)
- **v2** dichotomous covariate (0,1)
- **v3** dichotomous covariate (0,1)
- **TIMEGAP** time to event

**dataNCC**

Simulated data for recurrent events and a terminal event with weights using nested case-control design

**Description**

This contains a simulated sample of 819 subjects and 1510 observations. This dataset can be used to illustrate how to fit a joint frailty model for data from nested case-control studies.

**Usage**

```r
data(dataNCC)
```
dataNested

Format

This data frame contains the following columns:

- **id**: identification of patient
- **cov1**: dichotomous covariate (0,1)
- **cov2**: dichotomous covariate (0,1)
- **t.start**: start of interval
- **t.stop**: end of interval (death or censoring time)
- **gapttime**: time to event
- **event**: recurrent event status (0: no, 1: yes)
- **deathdays**: time of terminal event (death or right-censoring)
- **death**: censoring status (0: alive, 1: death)
- **ncc.wts**: weights for NCC design

---

Simulated data with two levels of grouping

Description

This contains a simulated sample of 400 observations which allow establishing 20 clusters with 4 subgroups and 5 subjects in each subgroup, in order to obtain two levels of grouping. This data set is useful to illustrate how to fit a nested model. Two independent gamma frailty parameters with a variance fixed at 0.1 for the cluster effect and at 0.5 for the subcluster effect were generated. Independent survival times were generated from a simple Weibull baseline risk function. The percentage of censoring data was around 30 per cent. The right-censoring variables were generated from a uniform distribution on [1,36] and a left-truncating variable was generated with a uniform distribution on [0,10]. Observations were included only if the survival time is greater than the truncated time.

Usage

```r
data(dataNested)
```

Format

This data frame contains the following columns:

- **group**: group identification variable
- **subgroup**: subgroup identification variable
- **t1**: start of interval (0 or truncated time)
- **t2**: end of interval (death or censoring time)
- **event**: censoring status (0: alive, 1: death)
- **cov1**: dichotomous covariate (0,1)
- **cov2**: dichotomous covariate (0,1)
Source


---

**dataOvarian**  
**Advanced Ovarian Cancer dataset**

Description

This dataset combines the data that were collected in four double-blind randomized clinical trials in advanced ovarian cancer. In these trials, the objective was to examine the efficacy of cyclophosphamide plus cisplatin (CP) versus cyclophosphamide plus adriamycin plus cisplatin (CAP) to treat advanced ovarian cancer. The candidate surrogate endpoint $S$ is progression-free survival time, defined as the time (in years) from randomization to clinical progression of the disease or death. The true endpoint $T$ is survival time, defined as the time (in years) from randomization to death of any cause.

Usage

`data(dataOvarian)`

Format

This data frame contains the following columns:

- `patientID` The identification number of a patient
- `trialID` The center in which a patient was treated
- `trt` The treatment indicator, coded as 0 = cyclophosphamide plus cisplatin (CP) and 1 = cyclophosphamide plus adriamycin plus cisplatin (CAP)
- `timeS` The candidate surrogate (progression-free survival)
- `statusS` Censoring indicator for Progression-free survival
- `timeT` The true endpoint (survival time)
- `statusT` Censoring indicator for survival time

Source

Diffepoce

Difference of Expected Prognostic Observed Cross-Entropy (EPOCE) estimators and its 95% tracking interval between two joint models.

Description

This function computes the difference of two EPOCE estimates (CVPOL and MPOL) and its 95% tracking interval between two joint models estimated using frailtyPenal, longiPenal or trivPenal. Difference in CVPOL is computed when the EPOCE was previously estimated on the same dataset as used for estimation (using an approximated cross-validation), and difference in MPOL is computed when the EPOCE was previously estimated on an external dataset.

Usage

Diffepoce(epoce1, epoce2)

Arguments

epoce1 a first object inheriting from class epoce.
epoce2 a second object inheriting from class epoce.

Details

From the EPOCE estimates and the individual contributions to the prognostic observed log-likelihood obtained with epoce function on the same dataset from two different estimated joint models, the difference of CVPOL (or MPOL) and its 95% tracking interval is computed. The 95% tracking interval is: Delta(MPOL) +/- qnorm(0.975)*sqrt(VARIANCE) for an external dataset Delta(CVPOL) +/- qnorm(0.975)*sqrt(VARIANCE) for the dataset used in frailtyPenal, longiPenal or trivPenal

where Delta(CVPOL) (or Delta(MPOL)) is the difference of CVPOL (or MPOL) of the two joint models, and VARIANCE is the empirical variance of the difference of individuals contributions to the prognostic observed log-likelihoods of the two joint models.

The estimators of EPOCE from arguments epoce1 and epoce2 must have been computed on the same dataset and with the pred.times.

Value

new.data a boolean which is FALSE if computation is done on the same data as for estimation, and TRUE otherwise
pred.times time or vector of times used in the function
DEPOCE the difference between the two MPOL or CVPOL for each time
TIinf lower confidence band for the difference
TIsup upper confidence band for the difference
References


Examples

```r
## Not run:

# Example for joint frailty models
data(readmission)

# first joint frailty model
joint1 <- frailtyPenal(Surv(t.start,t.stop,event) ~ cluster(id) +
                      dukes + charlson + sex + chemo + terminal(death),
                      formula.terminalEvent = ~ dukes + charlson + sex + chemo ,
                      data = readmission, n.knots = 8, kappa = c(2.11e+08,9.53e+11),
                      recurrentAG=TRUE)

# second joint frailty model without dukes nor charlson as covariates
joint2 <- frailtyPenal(Surv(t.start,t.stop,event) ~ cluster(id) +
                      sex + chemo + terminal(death),
                      formula.terminalEvent = ~ sex + chemo ,
                      data = readmission, n.knots = 8, kappa = c(2.11e+08,9.53e+11),
                      recurrentAG=TRUE)

temps <- c(200,500,800,1100)

# computation of estimators of EPOCE for the two models
epoce1 <- epoce(joint1,temps)
epoce2 <- epoce(joint2,temps)

# computation of the difference
diff <- Diffepoce(epoce1,epoce2)

print(diff)
plot(diff)

# Example for joint models with a biomarker
data(colorectal)
data(colorectalLongi)

# Survival data preparation - only terminal events
colorectalSurv <- subset(colorectal, new.lesions == 0)

# first joint model for a biomarker and a terminal event
modLongi <- longiPenal(Surv(time0, time1, state) ~ age +
treatment + who.PS, tumor.size ~ year*treatment + age +
who.PS, colorectalSurv, data.Longi = colorectalLongi,
```
random = c("1", "year"), id = "id", link = "Random-effects",
left.censoring = -3.33, hazard = "Weibull",
method.GH = "Pseudo-adaptive")

# second joint model for a biomarker, recurrent events and a terminal event
# (computation takes around 30 minutes)
modTriv <- model.weib.RE.gap <- trivPenal(Surv(gap.time, new.lesions)
~ cluster(id) + age + treatment + who.PS + prev.resection + terminal(state),
formula.terminalEvent =~ age + treatment + who.PS + prev.resection,
tumor.size ~ year * treatment + age + who.PS, data = colorectal,
data.Longi = colorectalLongi, random = c("1", "year"), id = "id",
link = "Random-effects", left.censoring = -3.33, recurrentAG = FALSE,
hazard = "Weibull", method.GH="Pseudo-adaptive", n.nodes=7)

time <- c(1, 1.5, 2, 2.5)

# computation of estimators of EPOCE for the two models
epoce1 <- epoce(modLongi, time)
# (computation takes around 10 minutes)
epoce2 <- epoce(modTriv, time)

diff <- Diffepoce(epoce1, epoce2)
print(diff)
plot(diff)

## End(Not run)

---

**epoce**

Estimators of the Expected Prognostic Observed Cross-Entropy (EPOCE) for evaluating predictive accuracy of joint models.

**Description**

This function computes estimators of the Expected Prognostic Observed Cross-Entropy (EPOCE) for evaluating the predictive accuracy of joint models using frailtyPenal, longiPenal, trivPenal or trivPenalNL. On the same data as used for estimation of the joint model, this function computes both the Mean Prognosis Observed Loss (MPOL) and the Cross-Validated Prognosis Observed Loss (CVPOL), two estimators of EPOCE. The latter corrects the MPOL estimate for over-optimism by approximated cross-validation. On external, this function only computes MPOL.

**Usage**

epoce(fit, pred.times, newdata = NULL, newdata.Longi = NULL)
Arguments

- **fit**: A jointPenal, longiPenal, trivPenal or trivPenalNL object.
- **pred.times**: Time or vector of times to compute epoke.
- **newdata**: Optional. In case of joint models obtained with frailtyPenal, trivPenal or trivPenalNL. For models inheriting from trivPenal or trivPenalNL class, if newdata is given, newdata.Longi must be given as well. When missing, the data used for estimating the fit are used, and CVPOL and MPOL are computed (internal validation). When newdata is specified, only MPOL is computed on this new dataset (external validation). The new dataset and the dataset used in the estimation must have the same covariates with the same coding without missing data.
- **newdata_LONGI**: Optional. In case of joint models obtained with longiPenal, trivPenal or trivPenalNL. For models inheriting from longiPenal, if the newdata_LONGI is given, newdata must be NULL, but for models from trivPenal or trivPenalNL class, if newdata_LONGI is given, newdata must be provided as well. The two datasets newdata and newdata_LONGI must include the information concerning the same patients with the same characteristics and the appropriate data on follow up (recurrences for newdata and longitudinal measurements for newdata_LONGI).

Value

- **data**: name of the data used to compute epoke
- **new.data**: a boolean which is FALSE if computation is done on the same data as for estimation, and TRUE otherwise
- **pred.times**: time or vector of times used in the function
- **mpol**: values of MPOL for each pred.times
- **cvpol**: values of CVPOL for each pred.times
- **IndivContrib**: all the contributions to the log-likelihood for each pred.times
- **AtRisk**: number of subject still at risk for each pred.times

References


Examples

```r
## Not run:

########################################
#### EPOCE on a joint frailty model ####
########################################

data(readmission)
```
modJoint.gap <- frailtyPenal(Surv(t.start,t.stop,event)~ cluster(id) +
dukes + charlson + sex + chemo + terminal(death),
formula.terminalEvent = ~ dukes + charlson + sex + chemo ,
data = readmission, n.knots = 8, kappa =c(2.11e+08,9.53e+11),
recurrentAG=TRUE)

# computation on the same dataset
temps <- c(200,500,800,1100)
epoce <- epoce(modJoint.gap,temps)

print(epoce)
plot(epoce,type = "cvpol")

# computation on a new dataset
# here a sample of readmission with the first 50 subjects
s <- readmission[1:100,]
epoce <- epoce(modJoint.gap,temps,newdata=s)

print(epoce)
plot(epoce,type = "cvpol")

#################################################
#### EPOCE on a joint model for a biomarker ####
######### and a terminal event ###############
#################################################
data(colorectal)
data(colorectalLongi)

# Survival data preparation - only terminal events
colorectalSurv <- subset(colorectal, new.lesions == 0)

modLongi <- longiPenal(Surv(time0, time1, state) ~ age +
treatment + who.PS, tumor.size ~ year*treatment + age +
who.PS, colorectalSurv, data.Longi =colorectalLongi,
random = c("1", "year"), id = "id", link = "Random-effects",
left.censoring = -3.33, hazard = "Weibull",
method.GH = "Pseudo-adaptive")

# computation on the same dataset
time <- c(1, 1.5, 2, 2.5)
epoce <- epoce(modLongi, time)

print(epoce)
plot(epoce, type = "cvpol")

# computation on a new dataset
# here a sample of colorectal data with the first 50 subjects
s <- subset(colorectal, new.lesions == 0 & id%in%1:50)
s.Longi <- subset(colorectalLongi, id%in%1:50)
epoce <- epoce(modLongi, time, newdata = s, newdata.Longi = s.Longi)
print(epoce)
plot(epoce, type = "cvpol")

###################################################
#### EPOCE on a joint model for a biomarker, ######
#### recurrent events and a terminal event  ######
###################################################

data(colorectal)
data(colorectalLongi)

# Linear model for the biomarker
# (computation takes around 30 minutes)
model.trivPenalNL <- trivPenal(Surv(gap.time, new.lesions) ~ cluster(id) + age + treatment + who.PS + prev.resection + terminal(state),
formula.terminalEvent = age + treatment + who.PS + prev.resection, tumor.size ~ year * treatment + age + who.PS, data = colorectal,
data.Longi = colorectalLongi, random = c("1", "year"), id = "id",
link = "Random-effects", left.censoring = -3.33, recurrentAG = FALSE,
hazard = "Weibull", method.GH="Pseudo-adaptive", n.nodes=7)

# computation on the same dataset
time <- c(1, 1.5, 2, 2.5)
epoce <- epoce(model.trivPenalNL, time)
print(epoce)
plot(epoce, type = "cvpol")

# computation on a new dataset
# here a sample of colorectal data with the first 100 subjects
s <- subset(colorectal, id%in%1:100)
s.Longi <- subset(colorectalLongi, id%in%1:100)
# (computation takes around 10 minutes)
epoce <- epoce(model.trivPenalNL, time, newdata = s, newdata.Longi = s.Longi)

print(epoce)
plot(epoce, type = "cvpol")

# Non-linear model for the biomarker

# No information on dose - creation of a dummy variable
colorectalLongi$dose <- 1

# (computation can take around 40 minutes)
model.trivPenalNL <- trivPenalNL(Surv(time0, time1, new.lesions) ~ cluster(id) + age + treatment + terminal(state),
formula.terminalEvent = age + treatment, biomarker = "tumor.size", formula.KG ~ 1, formula.KD ~ treatment, dose = "dose", time.biomarker = "year",
data = colorectal, data.Longi = colorectalLongi, random = c("y0", "KG"), id = "id",
init.B = c(-0.22, -0.16, -0.35, -0.19, 0.04, -0.41, 0.23), init.Alpha = 1.86,
init.Eta = c(0.5, 0.57, 0.5, 2.34), init.Biomarker = c(1.24, 0.81, 1.07, -1.53),
recurrentAG = TRUE, n.knots = 5, kappa = c(0.01, 2), method.GH = "Pseudo-adaptive")

# computation on the same dataset
time <- c(1, 1.5, 2, 2.5)
epoke <- epoce(model.trivPenalNL, time)

## End(Not run)

---

### event2

**Identify event2 indicator**

#### Description

This is a special function used in the context of multivariate frailty model with two types of recurrent events and a terminal event (e.g., censoring variable related to both recurrent events). It contains the indicator of the recurrent event of type 2, normally 0=no event, 1=event, and is used on the right hand side of a formula of a 'multivPenal' object. Using `event2()` in a formula implies that a multivariate frailty model for two types of recurrent events and a terminal event is fitted.

#### Usage

`event2(x)`

#### Arguments

- `x` A numeric variable but should be a boolean which equals 1 if the subject has experienced an event of type 2 and 0 if not.

#### Value

- `x` an indicator for an event of type 2

#### See Also

`multivPenal`
Description

Shared Frailty model

Fit a shared gamma or log-normal frailty model using a semiparametric Penalized Likelihood estimation or parametric estimation on the hazard function. Left-truncated, right-censored data, interval-censored data and strata (up to 6 levels) are allowed. It allows to obtain a non-parametric smooth hazard of survival function. This approach is different from the partial penalized likelihood approach of Therneau et al.

The hazard function, conditional on the frailty term \( \omega_i \), of a shared gamma frailty model for the \( j^{th} \) subject in the \( i^{th} \) group:

\[
\lambda_{ij}(t|\omega_i) = \lambda_0(t) \omega_i \exp(\beta' Z_{ij})
\]

\( \omega_i \sim \Gamma \left( \frac{1}{\theta}, \frac{1}{\theta} \right) \quad \mathbb{E}(\omega_i) = 1 \quad \text{Var}(\omega_i) = \theta \)

where \( \lambda_0(t) \) is the baseline hazard function, \( \beta \) the vector of the regression coefficient associated to the covariate vector \( Z_{ij} \) for the \( j^{th} \) individual in the \( i^{th} \) group.

Otherwise, in case of a shared log-normal frailty model, we have for the \( j^{th} \) subject in the \( i^{th} \) group:

\[
\lambda_{ij}(t|\eta_i) = \lambda_0(t) \exp(\eta_i + \beta' Z_{ij})
\]

\( \eta_i \sim N(0, \sigma^2) \)

From now on, you can also consider time-varying effects covariates in your model, see timedep function for more details.

Joint Frailty model

Fit a joint either with gamma or log-normal frailty model for recurrent and terminal events using a penalized likelihood estimation on the hazard function or a parametric estimation. Right-censored data and strata (up to 6 levels) for the recurrent event part are allowed. Left-truncated data is not possible. Joint frailty models allow studying, jointly, survival processes of recurrent and terminal events, by considering the terminal event as an informative censoring.

There is two kinds of joint frailty models that can be fitted with \texttt{frailtyPenal}:

- The first one (Rondeau et al. 2007) includes a common frailty term to the individuals \( \omega_i \) for the two rates which will take into account the heterogeneity in the data, associated with unobserved covariates. The frailty term acts differently for the two rates \( \omega_i \) for the recurrent rate and \( \omega_\alpha^i \) for the death rate. The covariates could be different for the recurrent rate and death rate.

For the \( j^{th} \) recurrence \( (j = 1, \ldots, n_i) \) and the \( i^{th} \) subject \( (i = 1, \ldots, G) \), the joint gamma frailty model for recurrent event hazard function \( r_{ij}(.) \) and death rate \( \lambda_i(.) \) is:
\[
\begin{align*}
\{ r_{ij}(t|\omega_i) &= \omega_i r_0(t) \exp(\beta_1 Z_i(t)) \quad \text{(Recurrent)} \\
\lambda_i(t|\omega_i) &= \omega_i^\alpha \lambda_0(t) \exp(\beta_2 Z_i(t)) \quad \text{(Death)}
\end{align*}
\]

where \( r_0(t) \) (resp. \( \lambda_0(t) \)) is the recurrent (resp. terminal) event baseline hazard function, \( \beta_1 \) (resp. \( \beta_2 \)) the regression coefficient vector, \( Z_i(t) \) the covariate vector. The random effects of frailties \( \omega_i \sim \Gamma(\frac{1}{3}, \frac{1}{3}) \) and are iid.

The joint log-normal frailty model will be :
\[
\begin{align*}
\{ r_{ij}(t|\eta_i) &= r_0(t) \exp(\eta_i + \beta_1 Z_{ij}(t)) \quad \text{(Recurrent)} \\
\lambda_{ij}(t|\eta_i) &= \lambda_0(t) \exp(\alpha \eta_i + \beta_2 Z_{ij}(t)) \quad \text{(Death)}
\end{align*}
\]

where \( \eta_i \sim N(0, \sigma^2) \)

- The second one (Rondeau et al. 2011) is quite similar but the frailty term is common to the individuals from a same group. This model is useful for the joint modelling two clustered survival outcomes. This joint models have been developed for clustered semi-competing events. The follow-up of each of the two competing outcomes stops when the event occurs. In this case, \( j \) is for the subject and \( i \) for the cluster.

\[
\begin{align*}
\{ r_{ij}(t|u_i) &= u_i r_0(t) \exp(\beta_1 Z_{ij}(t)) \quad \text{(Time to event)} \\
\lambda_{ij}(t|u_i) &= u_i^\alpha \lambda_0(t) \exp(\beta_2 Z_{ij}(t)) \quad \text{(Death)}
\end{align*}
\]

It should be noted that in these models it is not recommended to include \( \alpha \) parameter as there is not enough information to estimate it and thus there might be convergence problems.

In case of a log-normal distribution of the frailties, we will have :
\[
\begin{align*}
\{ r_{ij}(t|v_i) &= r_0(t) \exp(v_i + \beta_1 Z_{ij}(t)) \quad \text{(Time to event)} \\
\lambda_{ij}(t|v_i) &= \lambda_0(t) \exp(\alpha v_i + \beta_2 Z_{ij}(t)) \quad \text{(Death)}
\end{align*}
\]

where \( v_i \sim N(0, \sigma^2) \)

This joint frailty model can also be applied to clustered recurrent events and a terminal event (example on "readmission" data below).

From now on, you can also consider time-varying effects covariates in your model, see timedep function for more details.

There is a possibility to use a weighted penalized maximum likelihood approach for nested case-control design, in which risk set sampling is performed based on a single outcome (Jazic et al., Submitted).

General Joint Frailty model Fit a general joint frailty model for recurrent and terminal events considering two independent frailty terms. The frailty term \( u_i \) represents the unobserved association between recurrences and death. The frailty term \( v_i \) is specific to the recurrent event rate. Thus, the general joint frailty model is:
Frailty Penal

\[
\begin{aligned}
&\left\{ \begin{array}{l}
    r_{ij}(t|u_i, v_i) = u_i v_i r_0(t) \exp(\beta_i t Z_i(t)) = u_i v_i r_{ij}(t) \\
    \lambda_i(t|u_i) = u_i \lambda_0(t) \exp(\beta_i t Z_i(t)) = u_i \lambda_i(t)
\end{array} \right. \\
&\text{where the iid random effects } u_i \sim \Gamma\left(\frac{1}{\theta}, \frac{1}{\theta}\right) \text{ and the iid random effects } v_i \sim \Gamma\left(\frac{1}{\eta}, \frac{1}{\eta}\right) \text{ are independent from each other. The joint model is fitted using a penalized likelihood estimation on the hazard. Right-censored data and time-varying covariates } Z_i(t) \text{ are allowed.}
\end{aligned}
\]

**Nested Frailty model**

*Data should be ordered according to cluster and subcluster*

Fit a joint nested frailty model using a Penalized Likelihood on the hazard function or using a parametric estimation. Nested frailty models allow survival studies for hierarchically clustered data by including two iid gamma random effects. Left-truncated and right-censored data are allowed. Stratification analysis is allowed (maximum of strata = 2).

The hazard function conditional on the two frailties \( v_i \) and \( w_{ij} \) for the \( k^{th} \) individual of the \( j^{th} \) subgroup of the \( i^{th} \) group is:

\[
\begin{aligned}
&\left\{ \begin{array}{l}
    \lambda_{ijk}(t|v_i, w_{ij}) = v_i w_{ij} \lambda_0(t) \exp(\beta' X_{ijk}) \\
    v_i \sim \Gamma\left(\frac{1}{\alpha}, \frac{1}{\alpha}\right) \text{ i.i.d. } E(v_i) = 1 \ Var(v_i) = \alpha \\
    w_{ij} \sim \Gamma\left(\frac{1}{\eta}, \frac{1}{\eta}\right) \text{ i.i.d. } E(w_{ij}) = 1 \ Var(w_{ij}) = \eta
\end{array} \right.
\end{aligned}
\]

where \( \lambda_0(t) \) is the baseline hazard function, \( X_{ijk} \) denotes the covariate vector and \( \beta \) the corresponding vector of regression parameters.

**Joint Nested Frailty Model**

Fit a joint model for recurrent and terminal events using a penalized likelihood on the hazard functions or a parametric estimation. Right-censored data are allowed but left-truncated data and stratified analysis are not allowed.

Joint nested frailty models allow studying, jointly, survival processes of recurrent and terminal events for hierarchically clustered data, by including the terminal event as an informative censoring and by including two iid gamma random effects.

The joint nested frailty model includes two shared frailty terms, one for the subgroup \( (u_{fj}) \) and one for the group \( (w_f) \) into the hazard functions. This random effects account the heterogeneity in the data, associated with unobserved covariates. The frailty terms act differently for the two rates \( (u_{fj}, w_f^{\times} \text{ for the recurrent rate and } u_{fj}, w_f \text{ for the terminal event rate}) \). The covariates could be different for the recurrent rate and death rate.

For the \( j^{th} \) recurrence (\( j = 1, ..., n_i \)) of the \( i^{th} \) individual (\( i = 1, ..., m_f \)) of the \( f^{th} \) group (\( f = 1, ..., n \)), the joint nested gamma frailty model for recurrent event hazard function \( r_{fij}(\cdot) \) and for terminal event hazard function \( \lambda_{fj} \) is:

\[
\begin{aligned}
&\left\{ \begin{array}{l}
    r_{fij}(t|\omega_f, u_{fj}, X_{fij}) = r_0(t) u_{fj} \omega_f \exp(\beta' X_{fij}) \\
    \lambda_{fj}(t|\omega_f, u_{fj}, X_{fij}) = \lambda_0(t) u_{fj} \omega_f \exp(\gamma' X_{fij})
\end{array} \right. \\
&\text{where } r_0(t) (\text{resp. } \lambda_0(t)) \text{ is the recurrent (resp. terminal) event baseline hazard function, } \beta (\text{resp. } \gamma) \text{ the regression coefficient vector, } X_{fij}(t) \text{ the covariates vector. The random effects are}
\end{aligned}
\]

\[
\begin{aligned}
&\omega_f \sim \Gamma\left(\frac{1}{\eta}, \frac{1}{\eta}\right) \\
&u_{fj} \sim \Gamma\left(\frac{1}{\theta}, \frac{1}{\theta}\right)
\end{aligned}
\]
Usage


Arguments

formula a formula object, with the response on the left of a \( \sim \) operator, and the terms on the right. The response must be a survival object as returned by the 'Surv' function like in survival package. In case of interval-censored data, the response must be an object as returned by the 'SurvIC' function from this package. Interactions are possible using * or :

formula.terminalEvent only for joint and joint nested frailty models : a formula object, only requires terms on the right to indicate which variables are modelling the terminal event. Interactions are possible using * or :

data a 'data.frame' with the variables used in 'formula'.

recurrentAG Logical value. Is Andersen-Gill model fitted? If so indicates that recurrent event times with the counting process approach of Andersen and Gill is used. This formulation can be used for dealing with time-dependent covariates. The default is FALSE.

cross.validation Logical value. Is cross validation procedure used for estimating smoothing parameter in the penalized likelihood estimation? If so a search of the smoothing parameter using cross validation is done, with kappa as the seed. The cross validation is not implemented for several strata, neither for interval-censored data. The cross validation has been implemented for a Cox proportional hazard model, with no covariates. The default is FALSE.

jointGeneral Logical value. Does the model include two independent random effects? If so, this will fit a general joint frailty model with an association between the recurrent events and a terminal event (explained by the variance \( \theta \)) and an association amongst the recurrent events (explained by the variance \( \eta \)).

n.knots integer giving the number of knots to use. Value required in the penalized likelihood estimation. It corresponds to the \( (n.knots+2) \) splines functions for the approximation of the hazard or the survival functions. We estimate I or M-splines of order 4. When the user set a number of knots equals to \( k \) \( (n.knots=k) \) then the number of interior knots is \( (k-2) \) and the number of splines is \( (k-2)+order \). Number of knots must be between 4 and 20. (See Note)

kappa positive smoothing parameter in the penalized likelihood estimation. In a stratified shared model, this argument must be a vector with kappas for both strata. In a stratified joint model, this argument must be a vector with kappas for both strata for recurrent events plus one kappa for terminal event. The coefficient kappa of the integral of the squared second derivative of hazard function in
the fit (penalized log likelihood). To obtain an initial value for kappa, a solution is to fit the corresponding shared frailty model using cross validation (See cross.validation). We advise the user to identify several possible tuning parameters, note their defaults and look at the sensitivity of the results to varying them. Value required. (See Note).

maxit
maximum number of iterations for the Marquardt algorithm. Default is 300

hazard
Type of hazard functions: "Splines" for semiparametric hazard functions using equidistant intervals or "Splines-per" using percentile with the penalized likelihood estimation, "Piecewise-per" for piecewise constant hazard function using percentile (not available for interval-censored data), "Piecewise-equ" for piecewise constant hazard function using equidistant intervals, "Weibull" for parametric Weibull functions. Default is "Splines". In case of jointGeneral = TRUE or if a joint nested frailty model is fitted, only hazard = "Splines" can be chosen.

nb.int
Number of time intervals (between 1 and 20) for the parametric hazard functions ("Piecewise-per", "Piecewise-equ"). In a joint model, you need to specify a number of time interval for both recurrent hazard function and the death hazard function (vector of length 2).

RandDist
Type of random effect distribution: "Gamma" for a gamma distribution, "LogN" for a log-normal distribution. Default is "Gamma". Not implemented for nested model. If jointGeneral = TRUE or if a joint nested frailty model is fitted, the log-normal distribution cannot be chosen.

nb.gh
Number of nodes for the Gaussian-Hermite quadrature. It can be chosen among 5, 7, 9, 12, 15, 20 and 32. The default is 20 if hazard = "Splines", 32 otherwise.

nb.gl
Number of nodes for the Gaussian-Laguerre quadrature. It can be chosen between 20 and 32. The default is 20 if hazard = "Splines", 32 otherwise.

betaknots
Number of inner knots used for the estimation of B-splines. Default is 1. See ’timedep’ function for more details. Not implemented for nested and joint nested frailty models.

betaorder
Order of the B-splines. Default is cubic B-splines (order = 3). See ’timedep’ function for more details. Not implemented for nested and joint nested frailty models.

initialize
Logical value, only for joint nested frailty models. Option TRUE indicates fitting an appropriate standard joint frailty model (without group effect, only the subgroup effect) to provide initial values for the joint nested model. Default is TRUE.

init.B
A vector of initial values for regression coefficients. This vector should be of the same size as the whole vector of covariates with the first elements for the covariates related to the recurrent events and then to the terminal event (interaction in the end of each component). Default is 0.1 for each (for Cox and shared model) or 0.5 (for joint and joint nested frailty models).

init.Theta
Initial value for variance of the frailties.

init.Alpha
Only for joint and joint nested frailty models : initial value for parameter alpha.

Alpha
Only for joint and joint nested frailty model : input "None" so as to fit a joint model without the parameter alpha.
**init.Ksi**
Only for joint nested frailty model: initial value for parameter $\xi$.

**Ksi**
Only for joint nested frailty model: input "None" indicates a joint nested frailty model without the parameter $\xi$.

**init.Eta**
Only for general joint and joint nested frailty models: initial value for the variance $\eta$ of the frailty $v_i$ (general joint model) and of the frailty $\omega_i$ (joint nested frailty model).

**LIMparam**
Convergence threshold of the Marquardt algorithm for the parameters (see Details), $10^{-3}$ by default.

**LIMlogl**
Convergence threshold of the Marquardt algorithm for the log-likelihood (see Details), $10^{-3}$ by default.

**LIMderiv**
Convergence threshold of the Marquardt algorithm for the gradient (see Details), $10^{-3}$ by default.

**print.times**
a logical parameter to print iteration process. Default is TRUE.

**Details**
Typical usages are for a Cox model

```r
frailtyPenal(Surv(time,event)~var1+var2, data, ...)
```
for a shared model

```r
frailtyPenal(Surv(time,event)~cluster(group)+var1+var2, data, ...)
```
for a joint model

```r
frailtyPenal(Surv(time,event)~cluster(group)+var1+var2+ var3+terminal(death), formula.terminalEvent=~ var1+var4, data, ...)
```
for a joint model for clustered data

```r
frailtyPenal(Surv(time,event)~cluster(group)+num.id(group2)+ var1+var2+var3+terminal(death), formula.terminalEvent=~var1+var4, data, ...)
```
for a joint model for data from nested case-control studies

```r
frailtyPenal(Surv(time,event)~cluster(group)+num.id(group2)+ var1+var2+var3+terminal(death)+wts(wts.ncc), formula.terminalEvent=~var1+var4, data, ...)
```
for a nested model

```r
frailtyPenal(Surv(time,event)~cluster(group)+subcluster(sbgroup)+ var1+var2, data, ...)
```
for a joint nested frailty model
frailtyPenal(Surv(time,event)~cluster(group)+subcluster(sbgroup)+
var1+var2+terminal(death), formula.terminalEvent=~var1+var4, data, ...)

The estimated parameter are obtained using the robust Marquardt algorithm (Marquardt, 1963) which is a combination between a Newton-Raphson algorithm and a steepest descent algorithm. The iterations are stopped when the difference between two consecutive log-likelihoods was small (< $10^{-3}$), the estimated coefficients were stable (consecutive values (< $10^{-3}$), and the gradient small enough (< $10^{-3}$). When frailty parameter is small, numerical problems may arise. To solve this problem, an alternative formula of the penalized log-likelihood is used (see Rondeau, 2003 for further details). Cubic M-splines of order 4 are used for the hazard function, and I-splines (integrated M-splines) are used for the cumulative hazard function.

The inverse of the Hessian matrix is the variance estimator and to deal with the positivity constraint of the variance component and the spline coefficients, a squared transformation is used and the standard errors are computed by the Δ-method (Knight & Xekalaki, 2000). The smooth parameter can be chosen by maximizing a likelihood cross validation criterion (Joly and other, 1998). The integrations in the full log likelihood were evaluated using Gaussian quadrature. Laguerre polynomials with 20 points were used to treat the integrations on $[0, \infty]$

**INITIAL VALUES**

The splines and the regression coefficients are initialized to 0.1. In case of shared model, the program fits, firstly, an adjusted Cox model to give new initial values for the splines and the regression coefficients. The variance of the frailty term $\theta$ is initialized to 0.1. Then, a shared frailty model is fitted.

In case of a joint frailty model, the splines and the regression coefficients are initialized to 0.5. The program fits an adjusted Cox model to have new initial values for the regression and the splines coefficients. The variance of the frailty term $\theta$ and the coefficient $\alpha$ associated in the death hazard function are initialized to 1. Then, it fits a joint frailty model.

In case of a general joint frailty model we need to initialize the jointGeneral logical value to TRUE.

In case of a nested model, the program fits an adjusted Cox model to provide new initial values for the regression and the splines coefficients. The variances of the frailties are initialized to 0.1. Then, a shared frailty model with covariates with only subgroup frailty is fitted to give a new initial value for the variance of the subgroup frailty term. Then, a shared frailty model with covariates and only group frailty terms is fitted to give a new initial value for the variance of the group frailties. In a last step, a nested frailty model is fitted.

In case of a joint nested model, the splines and the regression coefficients are initialized to 0.5 and the variances of the frailty terms $\eta$ and $\xi$ are initialized to 1. If the option 'initialize' is TRUE, the program fits a joint frailty model to provide initial values for the splines, covariates coefficients, variance $\theta$ of the frailty terms and $\alpha$. The variances of the second frailty term ($\eta$) and the second coefficient $\xi$ are initialized to 1. Then, a joint nested frailty model is fitted.

**NCC DESIGN**

It is possible to fit a joint frailty model for data from nested case-control studies using the approach of weighted penalized maximum likelihood. For this model, only splines can be used for baseline hazards and no time-varying effects of covariates can be included. To accommodate the nested case-control design, the formula for the recurrent events should simply include the special term wts(wts.ncc), where wts.ncc refers to a column of prespecified weights in the data set for every observation. For details, see Jazic et al., Submitted (available on request from the package authors).
Value

The following components are included in a 'frailtyPenal' object for each model.

- **b**: sequence of the corresponding estimation of the coefficients for the hazard functions (parametric or semiparametric), the random effects variances and the regression coefficients.
- **call**: The code used for the model.
- **formula**: the formula part of the code used for the model.
- **coef**: the regression coefficients.
- **cross.Val**: Logical value. Is cross validation procedure used for estimating the smoothing parameters in the penalized likelihood estimation?
- **DoF**: Degrees of freedom associated with the "kappa".
- **groups**: the maximum number of groups used in the fit.
- **kappa**: A vector with the smoothing parameters in the penalized likelihood estimation corresponding to each baseline function as components.
- **loglikPenal**: the complete marginal penalized log-likelihood in the semiparametric case.
- **loglik**: the marginal log-likelihood in the parametric case.
- **n**: the number of observations used in the fit.
- **n.events**: the number of events observed in the fit.
- **n.iter**: number of iterations needed to converge.
- **n.knots**: number of knots for estimating the baseline functions in the penalized likelihood estimation.
- **n.strat**: number of stratum.
- **varH**: the variance matrix of all parameters before positivity constraint transformation. Then, the delta method is needed to obtain the estimated variance parameters. That is why some variances don’t match with the printed values at the end of the model.
- **varHIH**: the robust estimation of the variance matrix of all parameters.
- **x**: matrix of times where both survival and hazard function are estimated. By default seq(0,max(time),length=99), where time is the vector of survival times.
- **lam**: array (dim=3) of hazard estimates and confidence bands.
- **surv**: array (dim=3) of baseline survival estimates and confidence bands.
- **median**: The value of the median survival and its confidence bands. If there are two stratas or more, the first value corresponds to the value for the first strata, etc.
- **nbintervR**: Number of intervals (between 1 and 20) for the parametric hazard functions ("Piecewise-per", "Piecewise-equi").
- **npar**: number of parameters.
- **nvar**: number of explanatory variables.
- **LCV**: the approximated likelihood cross-validation criterion in the semiparametric case (with H minus the converged Hessian matrix, and l(.) the full log-likelihood).

\[
LCV = \frac{1}{n} \left( \text{trace}(H^{-1} H) - l(.) \right)
\]
AIC the Akaike information Criterion for the parametric case.

\[ AIC = \frac{1}{n}(np - l(.) ) \]

n.knots.temp initial value for the number of knots.
shape.weib shape parameter for the Weibull hazard function.
scale.weib scale parameter for the Weibull hazard function.
martingale.res martingale residuals for each cluster.
martingaleCox martingale residuals for observation in the Cox model.
Frailty Logical value. Was model with frailties fitted?
frailty.pred empirical Bayes prediction of the frailty term (ie, using conditional posterior distributions).
frailty.var variance of the empirical Bayes prediction of the frailty term (only for gamma frailty models).
frailty.sd standard error of the frailty empirical Bayes prediction (only for gamma frailty models).
global_chisq a vector with the values of each multivariate Wald test.
dof_chisq a vector with the degree of freedom for each multivariate Wald test.
global_chisq.test a binary variable equals to 0 when no multivariate Wald is given, 1 otherwise.
p.global_chisq a vector with the p_values for each global multivariate Wald test.
names.factor Names of the "as.factor" variables.
Xlevels vector of the values that factor might have taken.
contrasts type of contrast for factor variable.
beta_p.value p-values of the Wald test for the estimated regression coefficients.

The following components are specific to shared models.
equidistant Indicator for the intervals used the estimation of baseline hazard functions (for splines or piecewise-constant functions) : 1 for equidistant intervals ; 0 for intervals using percentile (note: equidistant = 2 in case of parametric estimation using Weibull distribution).
intcens Logical value. Indicator if a joint frailty model with interval-censored data was fitted)
theta variance of the gamma frailty parameter \( Var(\omega_i) \)
sigma2 variance of the log-normal frailty parameter \( Var(\eta_i) \)
linear.pred linear predictor: uses simply "Beta'X" in the cox proportional hazard model or "Beta'X + log w_i" in the shared gamma frailty models, otherwise uses "Beta'X + w_i" for log-normal frailty distribution.
BetaTpsMat matrix of time varying-effects and confidence bands (the first column used for abscissa of times)
theta_p.value p-value of the Wald test for the estimated variance of the gamma frailty.
sigma2_p.value  p-value of the Wald test for the estimated variance of the log-normal frailty.

The following components are specific to joint models.

intcens  Logical value. Indicator if a joint frailty model with interval-censored data was fitted
theta  variance of the gamma frailty parameter \( \text{Var}(\omega_i) \) or \( \text{Var}(u_i) \)
sigma2  variance of the log-normal frailty parameter \( \text{Var}(\eta_i) \) or \( \text{Var}(v_i) \)
eta  variance of the second gamma frailty parameter in general joint frailty models \( \text{Var}(v_i) \)
indic_alpha  indicator if a joint frailty model with \( \alpha \) parameter was fitted
alpha  the coefficient \( \alpha \) associated with the frailty parameter in the terminal hazard function.

nbintervR  Number of intervals (between 1 and 20) for the recurrent parametric hazard functions ("Piecewise-per", "Piecewise-equi").
nbintervDC  Number of intervals (between 1 and 20) for the death parametric hazard functions ("Piecewise-per", "Piecewise-equi").
nvar  A vector with the number of covariates of each type of hazard function as components.
nvarRec  number of recurrent explanatory variables.
nvarEnd  number of death explanatory variables.
nVar1  indicator of recurrent explanatory variables.
nVar2  indicator of death explanatory variables.
xR  matrix of times where both survival and hazard function are estimated for the recurrent event. By default \text{seq}(0,\text{max(time)},\text{length}=99), where time is the vector of survival times.
xD  matrix of times for the terminal event.
lamR  array (dim=3) of hazard estimates and confidence bands for recurrent event.
lamD  the same value as lamR for the terminal event.
survR  array (dim=3) of baseline survival estimates and confidence bands for recurrent event.
survD  the same value as survR for the terminal event.
martingale.res  martingale residuals for each cluster (recurrent).
martingaleddeath.res  martingale residuals for each cluster (death).
linear.pred  linear predictor: uses "Beta’X + log w_i" in the gamma frailty model, otherwise uses "Beta’X + eta_i" for log-normal frailty distribution
lineardeath.pred  linear predictor for the terminal part: "Beta’X + alpha.log w_i" for gamma, "Beta’X + alpha.eta_i" for log-normal frailty distribution
Xlevels  vector of the values that factor might have taken for the recurrent part.
contrasts  type of contrast for factor variable for the recurrent part.
Xlevels2 vector of the values that factor might have taken for the death part.
contrasts2 type of contrast for factor variable for the death part.
BetaTpsMat matrix of time varying-effects and confidence bands for recurrent event (the first column used for abscissa of times of recurrence)
BetaTpsMatDc matrix of time varying-effects and confidence bands for terminal event (the first column used for abscissa of times of death)
alpha_p.value p-value of the Wald test for the estimated $\alpha$.
ncc Logical value whether nested case-control design with weights was used for the joint model.

The following components are specific to nested models.

alpha variance of the cluster effect ($\text{Var}(v_i)$)
etta variance of the subcluster effect ($\text{Var}(w_{ij})$)
subgroups the maximum number of subgroups used in the fit.
frailty.pred.group empirical Bayes prediction of the frailty term by group.
frailty.pred.subgroup empirical Bayes prediction of the frailty term by subgroup.
linear.pred linear predictor: uses "Beta'X + log v_i.w_ij".
subgbyg subgroup by group.
n.strat A vector with the number of covariates of each type of hazard function as components.
alpha_p.value p-value of the Wald test for the estimated variance of the cluster effect.
etta_p.value p-value of the Wald test for the estimated variance of the subcluster effect.

The following components are specific to joint nested frailty models.

theta variance of the subcluster effect ($\text{Var}(u_{fi})$)
etta variance of the cluster effect ($\text{Var}(\omega_f)$)
alpha the power coefficient $\alpha$ associated with the frailty parameter ($u_{fi}$) in the terminal event hazard function.
ksi the power coefficient $\xi$ associated with the frailty parameter ($\omega_f$) in the recurrent event hazard function.
indic_alpha indicator if a joint frailty model with $\alpha$ parameter was fitted or not.
indic_ksi indicator if a joint frailty model with $\xi$ parameter was fitted or not.
frailty.fam.pred empirical Bayes prediction of the frailty term by family.
eta_p.value p-value of the Wald test for the estimated variance of the cluster effect.
alpha_p.value p-value of the Wald test for the estimated power coefficient $\alpha$.
ksi_p.value p-value of the Wald test for the estimated power coefficient $\xi$. 
Note

From a prediction aim, we recommend you to input a data sorted by the group variable with numerical numbers from 1 to n (number of groups). In case of a nested model, we recommend you to input a data sorted by the group variable then sorted by the subgroup variable both with numerical numbers from 1 to n (number of groups) and from 1 to m (number or subgroups). "kappa" and "n.knots" are the arguments that the user have to change if the fitted model does not converge. "n.knots" takes integer values between 4 and 20. But with n.knots=20, the model would take a long time to converge. So, usually, begin first with n.knots=7, and increase it step by step until it converges. "kappa" only takes positive values. So, choose a value for kappa (for instance 10000), and if it does not converge, multiply or divide this value by 10 or 5 until it converges.

References


See Also

SurvIC, cluster, subcluster, terminal, num.id, timedep
## Examples

### Not run:

#### COX proportional hazard model (SHARED without frailties) ----###
#### estimated with penalized likelihood ----###

```r
data(kidney)
frailtyPenal(Surv(time,status)~sex+age,
n.knots=12,kappa=10000,data=kidney)
```

#### Shared Frailty model ----###

```r
frailtyPenal(Surv(time,status)~cluster(id)+sex+age,
n.knots=12,kappa=10000,data=kidney)
```

---

#### with an initialisation of regression coefficients

```r
frailtyPenal(Surv(time,status)~cluster(id)+sex+age,
n.knots=12,kappa=10000,data=kidney,init.B=c(-1.44,0))
```

---

#### with truncated data

```r
data(dataNested)
frailtyPenal(Surv(t1,t2,event) ~ cluster(group),
data=dataNested,n.knots=10,kappa=10000,
cross.validation=TRUE,recurrentAG=FALSE)
```

---

#### stratified analysis

```r
data(readmission)
frailtyPenal(Surv(time,event)~cluster(id)+dukes+strata(sex),
n.knots=10,kappa=c(10000,10000),data=readmission)
```

---

#### recurrentAG=TRUE

```r
frailtyPenal(Surv(t.start,t.stop,event)~cluster(id)+sex+dukes+charlson,data=readmission,n.knots=6,kappa=1e5,recurrentAG=TRUE)
```

---

#### cross.validation=TRUE

```r
frailtyPenal(Surv(t.start,t.stop,event)~cluster(id)+sex+dukes+charlson,data=readmission,n.knots=6,kappa=5000,recurrentAG=TRUE,cross.validation=TRUE)
```

---

#### log-normal distribution

```r
frailtyPenal(Surv(t.start,t.stop,event)~cluster(id)+sex+dukes+charlson,data=readmission,n.knots=6,kappa=5000,recurrentAG=TRUE,RandDist="LogN")
```
###--- Joint Frailty model (recurrent and terminal events) ---###

data(readmission)

--- Gap-time

modJoint.gap <- frailtyPenal(Surv(time,event)~cluster(id)+sex+dukes+charlson+terminal(death),formula.terminalEvent=~sex+dukes+charlson, data=readmission,n.knots=14,kappa=c(9.55e+9,1.41e+12), recurrentAG=FALSE)

--- Calendar time

modJoint.calendar <- frailtyPenal(Surv(t.start,t.stop,event)~cluster(id)+sex+dukes+charlson+terminal(death),formula.terminalEvent=~sex+dukes+charlson, data=readmission,n.knots=10,kappa=c(9.55e9,1.41e12), recurrentAG=TRUE)

--- without alpha parameter

modJoint.gap <- frailtyPenal(Surv(time,event)~cluster(id)+sex+dukes+charlson+terminal(death),formula.terminalEvent=~sex+dukes+charlson, data=readmission,n.knots=10,kappa=c(9.55e9,1.41e12), recurrentAG=FALSE,Alpha="None")

--- log-normal distribution

modJoint.log <- frailtyPenal(Surv(t.start,t.stop,event)~cluster(id)+sex+dukes+charlson+terminal(death),formula.terminalEvent=~sex+dukes+charlson, data=readmission,n.knots=10,kappa=c(9.55e9,1.41e12), recurrentAG=TRUE,RandDist="LogN")

###--- Joint frailty model for NCC data ---###

data(dataNCC)

modJoint.ncc <- frailtyPenal(Surv(t.start,t.stop,event)~cluster(id)+cov1+cov2+terminal(death)+wts(ncc.wts), formula.terminalEvent=~cov1+cov2, data=dataNCC,n.knots=8,kappa=c(1.6e+10,5.0e+03),recurrentAG=TRUE,RandDist="LogN")

###--- Joint Frailty model for clustered data ---###

--- here is generated cluster (5 clusters)

readmission <- transform(readmission,group=id%%5+1)

--- exclusion all recurrent events --

--- to obtain framework of semi-competing risks --

readmission2 <- subset(readmission, (t.start == 0 & event == 1) | event == 0)

joi.clus.gap <- frailtyPenal(Surv(time,event)~cluster(group)+num.id(id)+dukes+charlson+sex+chemo+terminal(death), formula.terminalEvent=~dukes+charlson+sex+chemo, data=readmission2,recurrentAG=FALSE, n.knots=8, kappa=c(1.e+10,1.e+10),Alpha="None")

###--- General Joint model (recurrent and terminal events) ---

---
with 2 covariates ---###

data(readmission)
modJoint.general <- frailtyPenal(Surv(time,event) ~ cluster(id) + dukes + charlson + sex + chemo + terminal(death),
formula.terminalEvent = ~ dukes + charlson + sex + chemo,
data = readmission, jointGeneral = TRUE, n.knots = 8,
kappa = c(2.11e+08, 9.53e+11))

###--- Nested Frailty model ---###

# Data should be ordered according to cluster and subcluster

data(dataNested)
modClu <- frailtyPenal(Surv(t1,t2,event)~cluster(group)+
subcluster(subgroup)+cov1+ cov2, data = dataNested,
n.knots = 8, kappa = 50000)

modClu.str <- frailtyPenal(Surv(t1,t2,event)~cluster(group)+
subcluster(subgroup)+cov1+strata(cov2), data = dataNested,
n.knots = 8, kappa = c(50000, 50000))

###--- Joint Nested Frailty model ---###

#-- here is generated cluster (30 clusters)
readmissionNested <- transform(readmission, group = id%%30+1)

modJointNested_Splines <- frailtyPenal(formula = Surv(t.start, t.stop, event)
~ subcluster(id) + cluster(group) + dukes + terminal(death),
formula.terminalEvent = ~ dukes, data = readmissionNested, recurrentAG = TRUE,
n.knots = 8, kappa = c(9.55e+9, 1.41e+12), initialize = TRUE)

modJointNested_Weib <- frailtyPenal(Surv(t.start, t.stop, event)~subcluster(id)
+ cluster(group) + dukes + terminal(death), formula.terminalEvent = ~ dukes,
hazard = c("Weibull"), data = readmissionNested, recurrentAG = TRUE, initialize = FALSE)

JoInes_GapSpline <- frailtyPenal(formula = Surv(time, event)
~ subcluster(id) + cluster(group) + dukes + terminal(death),
formula.terminalEvent = ~ dukes, data = readmissionNested,
recurrentAG = FALSE, n.knots = 8, kappa = c(9.55e+9, 1.41e+12),
initialize = TRUE, init.Alpha = 1.091, Ksi = "None")

# End(Not run)
Description

This meta-analysis was carried out by the GASTRIC (Global Advanced/Adjuvant Stomach Tumor Research international Collaboration) group, using individual data on patients with curatively resected gastric cancer. Data from all published randomized trials, with a patient recruitment end date before 2004, and comparing adjuvant chemotherapy with surgery alone for resectable gastric cancers, were searched electronically. The candidate surrogate endpoint \( S \) was Disease-free survival time, defined as the time (in days) to relapse, second cancer or dead from any cause. The true endpoint \( T \) was the overall survival time, defined as the time (in days) from randomization to death of any cause or to the last follow-up.

Usage

data(gastadj)

Format

This data frame contains the following columns:

- **trialID**: The trial in which the patient was treated
- **patientID**: The identification number of a patient
- **trt**: The treatment indicator, coded as 0 = Control and 1 = Experimental
- **timeS**: The candidate surrogate (progression-free survival in days)
- **statusS**: Censoring indicator for progression-free survival (0 = alive and progression-free, 1 = with progression or dead)
- **timeT**: The true endpoint (overall survival time in days)
- **statusT**: Censoring indicator for survival time (0 = alive, 1 = dead)

Source


---

**hazard**

_Hazard function_.

Description

Let \( t \) be a continuous variable, we determine the value of the hazard function to \( t \) after run fit.

Usage

hazard(t, ObjFrailty)
jointSurroPenal

Arguments

- `t`: time for hazard function.
- `ObjFrailty`: an object from the frailtypack fit.

Value

return the value of hazard function in `t`.

Examples

```r
## Not run:

#-- a fit Shared
data(readmission)
fit.shared <- frailtyPenal(Surv(time,event)~dukes+cluster(id)+
strata(sex),n.knots=10,kappa=c(10000,10000),data=readmission)

#-- calling survival
hazard(20,fit.shared)

## End(Not run)
```

---

### jointSurroPenal

**Fit the one-step Joint surrogate model for the evaluation of a candidate surrogate endpoint**

---

**Description**

**Joint Frailty Surrogate model definition**

Fit the one-step Joint surrogate model for the evaluation of a candidate surrogate endpoint, with different integration methods on the random effects, using a semiparametric penalized likelihood estimation. This approach extends that of Burzykowski et al. (2001) by including in the same joint frailty model the individual-level and the trial-level random effects.

For the `j`th subject (`j=1,\ldots,n_i`) of the `i`th trial (`i=1,\ldots,G`), the joint surrogate model is defined as follows:

\[
\begin{align*}
\lambda_{S,ij}(t|\omega_{ij},u_i,v_{S,i},Z_{ij1}) &= \lambda_{0S}(t) \exp(\omega_{ij} + u_i + v_{S,i}Z_{ij1} + \beta_SZ_{ij1}) \\
\lambda_{T,ij}(t|\omega_{ij},u_i,v_{T,i},Z_{ij1}) &= \lambda_{0T}(t) \exp(\zeta\omega_{ij} + \alpha u_i + v_{T,i}Z_{ij1} + \beta_TZ_{ij1})
\end{align*}
\]

where,

\[
\omega_{ij} \sim N(0,\theta), u_i \sim N(0,\gamma), \omega_{ij} \perp u_i, u_i \perp v_{S,i}, u_i \perp v_{T,i}
\]
and \((v_S, v_T)^T \sim N(0, \Sigma_v)\), with

\[
\Sigma_v = \begin{pmatrix}
\sigma_{v_S}^2 & \sigma_{v_ST} \\
\sigma_{v_ST} & \sigma_{v_T}^2
\end{pmatrix}
\]

In this model, \(\lambda_{0S}(t)\) is the baseline hazard function associated with the surrogate endpoint and \(\beta_S\) the fixed treatment effect (or log-hazard ratio); \(\lambda_{0T}(t)\) is the baseline hazard function associated with the true endpoint and \(\beta_T\) the fixed treatment effect. \(\omega_{ij}\) is a shared individual-level frailty that serve to take into account the heterogeneity in the data at the individual level; \(u_i\) is a shared frailty effect associated with the baseline hazard function that serve to take into account the heterogeneity between trials of the baseline hazard function, associated with the fact that we have several trials in this meta-analytical design. The power parameters \(\zeta\) and \(\alpha\) distinguish both individual and trial-level heterogeneities between the surrogate and the true endpoint. \(v_S, v_T\) are two correlated random effects treatment-by-trial interactions. \(Z_{ij}\) represents the treatment arm to which the patient has been randomized.

**Surrogacy evaluation**

We proposed new definitions of Kendall’s \(\tau\) and coefficient of determination as individual-level and trial-level association measurements, to evaluate a candidate surrogate endpoint (Sofeu et al., 2018). The formulations are given below.

**Individual-level surrogacy**

To measure the strength of association between \(S_{ij}\) and \(T_{ij}\) after adjusting the marginal distributions for the trial and the treatment effects, as show in Sofeu et al.(2018), we use the Kendall’s \(\tau\) define by:

\[
\tau = 2 \int_{u_i} \int_{\omega_{ij}} \int_{u_i'} \int_{\omega_{ij}'} \left\{ \frac{\exp(\omega_{ij} + u_i + \alpha u_i) + \exp(\omega_{ij}' + u_i' + \alpha u_i')}{(\exp(\omega_{ij} + u_i) + \exp(\omega_{ij}'))(\exp(\omega_{ij}' + u_i') + \exp(\omega_{ij} + u_i))} \right\} \frac{1}{\sqrt{2\pi\theta}} \exp \left[ -\frac{1}{2} \frac{\omega_{ij}^2}{\theta} \right] d\omega_{ij} d\omega_{ij}' du_i du_i' - 1
\]

where \(\theta, \zeta, \alpha, \gamma\) are estimated using the joint surrogate model defined previously. Kendall’s \(\tau\) is the difference between the probability of concordance and the probability of discordance of two realizations of \(S_{ij}\) and \(T_{ij}\). It belongs to the interval [-1,1] and assumes a zero value when \(S_{ij}\) and \(T_{ij}\) are independent. We estimate Kendall’s \(\tau\) using Monte-Carlo or Gaussian Hermite quadrature integration methods. Its confidence interval is estimated using parametric bootstrap.

**Trial-level surrogacy**

The key motivation for validating a surrogate endpoint is to be able to predict the effect of treatment on the true endpoint, based on the observed effect of treatment on the surrogate endpoint. As shown by Buyse et al. (2000), the coefficient of determination obtains from the covariance matrix \(\Sigma_v\) of the random effects treatment-by-trial interaction can be used to evaluate underlined prediction, and therefore as surrogacy evaluation measurement at trial-level. It is defined by:

\[
R_{trial}^2 = \frac{\sigma_{v_ST}^2}{\sigma_{v_S}^2 \sigma_{v_T}^2}
\]
The SEs of $R^2_{trial}$ is calculated using the Delta-method. We also propose $R^2_{trial}$ and 95% CI computed using the parametric bootstrap. The use of delta-method can lead to confidence limits violating the [0,1], as noted by (Burzykowski et al., 2001). However, using other methods would not significantly alter the findings of the surrogacy assessment.

Usage

```
jointSurroPenal(data, maxit=40, indicator.zeta = 1, indicator.alpha = 1, frail.base = 1, n.knots = 6, LIMparam = 0.001, LIMlogl = 0.001, LIMderiv = 0.001, nb.mc = 300, nb.gh = 32, nb.gh2 = 20, adaptatif = 0, int.method = 2, nb.iterPGH = 5, nb.MC.kendall = 10000, nboot.kendall = 1000, true.init.val = 0, theta.init = 1, sigma.ss.init = 0.5, sigma.tt.init = 0.5, sigma.st.init = 0.48, gamma.init = 0.5, alpha.init = 1, zeta.init = 1, betas.init = 0.5, betat.init = 0.5, scale = 1, random.generator = 1, kappa.use = 4, random = 0, random.nb.sim = 0, seed = 0, init.kappa = NULL, nb.decimal = 4, print.times = TRUE, print.iter=FALSE)
```

Arguments

- **data**: A data.frame containing at least 7 variables intitled:
  - `patienID`: A numeric, that represents the patient’s identifier, must be unique;
  - `trialID`: A numeric, that represents the trial in which each patient was randomized;
  - `timeS`: The follow up time associated with the surrogate endpoint;
  - `statusS`: The event indicator associated with the surrogate endpoint. Normally 0 = no event, 1 = event;
  - `timeT`: The follow up time associated with the true endpoint;
  - `statusT`: The event indicator associated with the true endpoint. Normally 0 = no event, 1 = event;
  - `trt`: The treatment indicator for each patient, with 1 = treated, 0 = untreated.
- **maxit**: maximum number of iterations for the Marquardt algorithm. Default is 40.
- **indicator.zeta**: A binary, indicates whether the power’s parameter $\zeta$ should be estimated (1) or not (0). If 0, $\zeta$ will be set to 1 during estimation. The default is 1. This parameter can be seted to 0 in case of convergence and identification issues.
- **indicator.alpha**: A binary, indicates whether the power’s parameter $\alpha$ should be estimated (1) or not (0). If 0, $\alpha$ will be set to 1 during estimation. The default is 1.
- **frail.base**: Considered the heterogeneity between trial on the baseline risk (1), using the shared cluster specific frailties ($u_i$), or not (0). The default is 1.
- **n.knots**: integer giving the number of knots to use. Value required in the penalized likelihood estimation. It corresponds to the (n.knots+2) splines functions for the approximation of the hazard or the survival functions. We estimate I or M-splines
of order 4. When the user set a number of knots equals to \( k \) (\text{n.knots}=k) then the number of interior knots is \((k-2)\) and the number of splines is \((k-2)+\text{order}\). Number of knots must be between 4 and 20. (See \text{frailtyPenal} for more details).

\text{LIMparam} \quad \text{Convergence threshold of the Marquardt algorithm for the parameters, } 10^{-3} \text{ by default (See } \text{frailtyPenal} \text{ for more details).}

\text{LIMlogl} \quad \text{Convergence threshold of the Marquardt algorithm for the log-likelihood, } 10^{-3} \text{ by default (See } \text{frailtyPenal} \text{ for more details).}

\text{LIMderiv} \quad \text{Convergence threshold of the Marquardt algorithm for the gradient, } 10^{-3} \text{ by default (See } \text{frailtyPenal} \text{ for more details).}

\text{nb.mc} \quad \text{Number of samples considered in the Monte-Carlo integration. Required in case int.method is equals to } \emptyset, 2 \text{ or } 4. \text{ A value between 100 and 300 most often gives good results. However, beyond 300, the program takes a lot of time to estimate the parameters. The default is 300.}

\text{nb.gh} \quad \text{Number of nodes for the Gaussian-Hermite quadrature. It can be chosen among 5, 7, 9, 12, 15, 20 and 32. The default is 32.}

\text{nb.gh2} \quad \text{Number of nodes for the Gauss-Hermite quadrature used to re-estimate the model, in case of non-convergence, defined as previously. The default is 20.}

\text{adaptatif} \quad \text{A binary, indicates whether the pseudo adaptive Gaussian-Hermite quadrature (1) or the classical Gaussian-Hermite quadrature (0) is used. The default is 0.}

\text{int.method} \quad \text{A numeric, indicates the integration method: } \emptyset \text{ for Monte carlo, 1 for Gaussian-Hermite quadrature, 2 for a combination of both Gaussian-Hermite quadrature to integrate over the individual-level random effects and Monte carlo to integrate over the trial-level random effects, 4 for a combination of both Monte carlo to integrate over the individual-level random effects and Gaussian-Hermite quadrature to integrate over the trial-level random effects. The default is 2.}

\text{nb.iterPGH} \quad \text{Number of iterations before the re-estimation of the posterior random effects, in case of the two-steps pseudo-adaptive Gaussian-hermite quadrature. If set to } \emptyset \text{ there is no re-estimation". The default is 5.}

\text{nb.MC.kendall} \quad \text{Number of generated points used with the Monte-Carlo to estimate integrals in the Kendall’s } \tau \text{ formulation. Beter to use at least 4000 points for stable results. The default is 10000.}

\text{nboot.kendall} \quad \text{Number of samples considered in the parametric bootstrap to estimate the confidence interval of the Kendall’s } \tau. \text{ The default is 1000.}

\text{true.init.val} \quad \text{Numerical value. Indicates if the given initial values to parameters (0) should be considered. If set to 2, } \alpha \text{ and } \gamma \text{ are initialised using two separated shared frailty model (see } \text{frailtyPenal} \text{ for more details); } \sigma^{2}_{v_S}, \sigma^{2}_{v_T} \text{ and } \sigma^{2}_{vST} \text{ are fixed by the user or the default values; } \zeta, \theta, \beta_S \text{ and } \beta_T \text{ are initialized using a classical joint frailty model, considering individual level random effects. If the joint frailty model is faced to convergence issues, } \beta_S \text{ and } \beta_T \text{ are initialized using two shared frailty models. In all others scenarios, if the simplified model does not converge, default given parameters values are used. Initial values for spline’s associated parameters are fixed to 0. 5. The default for this argument is 0.}

\text{theta.init} \quad \text{Initial values for } \theta, \text{ required if true.init.val is set to } \emptyset \text{ or 2. The default is 1.}

\text{sigma.ss.init} \quad \text{Initial values for } \sigma^{2}_{vS}, \text{ required if true.init.val is set to } \emptyset \text{ or 2. The default is 0.5.}
sigma.tt.init Initial values for $\sigma_{vp}^2$, required if true.init.val is set to 0 or 2. The default is 0.5.
sigma.st.init Initial values for $\sigma_{vST}$, required if true.init.val is set to 0 or 2. The default is 0.48.
gamma.init Initial values for $\gamma$, required if true.init.val is set to 0 or 2. The default is 0.5.
alpha.init Initial values for $\alpha$, required if true.init.val is set to 0 or 2. The default is 1.
zeta.init Initial values for $\zeta$, required if true.init.val is set to 0 or 2. The default is 1.
betas.init Initial values for $\beta_S$, required if true.init.val is set to 0 or 2. The default is 0.5.
betat.init Initial values for $\beta_T$, required if true.init.val is set to 0 or 2. The default is 0.5.
scale A numeric that allows to rescale the survival times, to avoid numerical problems in case of some convergence issues. If no change is need the argument is set to 1, the default value. eg: 365 aims to convert days to years.
random.generator Random number generator to use by the Fortran compiler, 1 for the intrinsec subroutine Random_number and 2 for the subroutine uniran(). The default is 1. In case of convergence problem with int.method set to 0, 2 or 4, that requires integration by Monte-Carlo, user could change the random numbers generator.
kappa.use A numeric, that indicates how to manage the smoothing parameters $k_1$ and $k_2$ in case of convergence issues. If it is set to 1, the given smoothing parameters or those obtained by cross-validation are used. If it is set to 3, the associated smoothing parameters are successively divided by 10, in case of convergence issues until 5 times. If it is set to 4, the management of the smoothing parameter is as in case 1, follows by the successive division as described in case 3 and preceded by the changing of the number of nodes for the Gauss-Hermite quadrature. The default is 4.
random A binary that says if we reset the random number generation with a different environment at each call (1) or not (0). If it is set to 1, we use the computer clock as seed. In the last case, it is not possible to reproduce the generated datasets. The default is 0. Required if random.generator is set to 1.
random.nb.sim If random is set to 1, a binary that indicates the number of generations that will be made.
seed The seed to use for data (or samples) generation. required if random is set to 0. The default is 0.
init.kappa smoothing parameter used to penalized the log-likelihood. By default (init.kappa = NULL) the values used are obtain by cross-validation.
nb.decimal Number of decimal required for results presentation.
print.times a logical parameter to print estimation time. Default is TRUE.
print.iter a logical parameter to print iteration process. Default is FALSE.
Details

The estimated parameter are obtained using the robust Marquardt algorithm (Marquardt, 1963) which is a combination between a Newton-Raphson algorithm and a steepest descent algorithm. The iterations are stopped when the difference between two consecutive log-likelihoods was small \((< 10^{-3})\), the estimated coefficients were stable (consecutive values \((< 10^{-3})\), and the gradient small enough \((< 10^{-3})\), by default. Cubic M-splines of order 4 are used for the hazard function, and I-splines (integrated M-splines) are used for the cumulative hazard function.

The inverse of the Hessian matrix is the variance estimator and to deal with the positivity constraint of the variance component and the spline coefficients, a squared transformation is used and the standard errors are computed by the \(\Delta\)-method (Knight & Xekalaki, 2000). The smooth parameter can be chosen by maximizing a likelihood cross validation criterion (Joly and other, 1998).

We proposed based on the joint surrogate model a new definition of the Kendall’s \(\tau\). Moreover, distinct numerical integration methods are available to approximate the integrals in the marginal log-likelihood.

Non-convergence case management procedure

Special attention must be given to initializing model parameters, the choice of the number of spline knots, the smoothing parameters and the number of quadrature points to solve convergence issues. We first initialized parameters using the user’s desired strategy, as specified by the option \texttt{true.init.val}. When numerical or convergence problems are encountered, with kappa.use set to 4, the model is fitted again using a combination of the following strategies: vary the number of quadrature point (nb. gh to nb. gh2 or nb. gh2 to nb. gh) in case of the use of the Gaussian Hermite quadrature integration (see \texttt{int.method}); divided or multiplied the smoothing parameters \((k_1, k_2)\) by 10 or 100 according to their preceding values, or used parameter vectors obtained during the last iteration (with a modification of the number of quadrature points and smoothing parameters). Using this strategy, we usually obtained during simulation the rejection rate less than 3%. A sensitivity analysis was conducted without this strategy, and similar results were obtained on the converged samples, with about a 23% rejection rate.

Value

This function return an object of class jointSurroPenal with elements:

- **EPS** A vector containing the obtained convergence thresholds with the Marquardt algorithm, for the parameters, the log-likelihood and for the gradient;
- **b** A vector containing estimates for the spline’s parameter’s, the power’s parameter \(\zeta\) (if \texttt{indicator.zeta} is set to 1), the standard error of the shared individual-level frailty \(\omega_{ij}(\theta)\), elements of the lower triangular matrix (L) from the Cholesky decomposition such that \(\Sigma = LL^T\), with \(\Sigma\) the covariances of the random effects \((v_S, v_T)\), the coefficient \(\alpha\) (if \texttt{indicator.alpha} is set to 1), the standard error of the random effect \(u_i\) and the regression coefficients \(\beta_S\) and \(\beta_T\);
- **varH** The variance matrix of all parameters in \(b\) (before positivity constraint transformation for the variance of the measurement error, for which the delta method is used);
- **varH2H** The robust estimation of the variance matrix of all parameters in \(b\);
- **loglikPenal** The complete marginal penalized log-likelihood;
LCV the approximated likelihood cross-validation criterion in the semiparametric case (with $H$ minus the converged Hessian matrix, and $l(.)$ the full log-likelihood).

\[ LCV = \frac{1}{n} \left( \text{trace}(H^{-1}H) - l(.) \right) \]

xS vector of times for surrogate endpoint where both survival and hazard function are estimated. By default seq(0, max(time), length=99), where time is the vector of survival times;

lamS array (dim = 3) of hazard estimates and confidence bands, for surrogate endpoint;

survS array (dim = 3) of baseline survival estimates and confidence bands, for surrogate endpoint;

xT vector of times for true endpoint where both survival and hazard function are estimated. By default seq(0, max(time), length = 99), where time is the vector of survival times;

lamT array (dim = 3) of hazard estimates and confidence bands, for true endpoint;

survT array (dim = 3) of baseline survival estimates and confidence bands, for true endpoint;

n.iter number of iterations needed to converge;

theta Estimate for $\theta$;

gamma Estimate for $\gamma$;

alpha Estimate for $\alpha$;

zeta Estimate for $\zeta$;

sigma.s Estimate for $\sigma_S$;

sigma.t Estimate for $\sigma_T$;

sigma.st Estimate for $\sigma_{ST}$;

beta.s Estimate for $\beta_S$;

beta.t Estimate for $\beta_T$;

ui A binary, that indicates if the heterogeneity between trial on the baseline risk has been considered (1), using the shared cluster specific frailties ($u_i$), or not (0);

ktau The Kendall’s $\tau$ with the correspondant 95% CI computed using the parametric bootstrap;

R2.boot The $R^2_{trial}$ with the correspondant 95% CI computed using the parametric bootstrap;

Coefficients The estimates with the corresponding standard errors and the 95% CI

kappa Positive smoothing parameters used for convergence. These values could be different to initial values if kappa.use is set to 3 or 4;

scale The value used to rescale the survival times

data The dataset used in the model

varcov.Sigma covariance matrix of ($\hat{\sigma}_S, \hat{\sigma}_T, \hat{\sigma}_{ST}$) obtained from the delta-method

parameter list of all arguments used in the model
Author(s)
Casimir Ledoux Sofeu <casimir.sofeu@u-bordeaux.fr>, <scl.ledoux@gmail.com> and Virginie Rondeau <virginie.rondeau@inserm.fr>

References

See Also
jointSurrSimul, summary.jointSurroPenal, jointSurroPenalSimul

Examples

# Generation of data to use
data.sim <- jointSurrSimul(n.obs=600, n.trial = 30, cens.adm = 549.24,
                        alpha = 1.5, theta = 3.5, gamma = 2.5, zeta = 1, sigma.s = 0.7,
                        sigma.t = 0.7, rsqrt = 0.8, betas = -1.25, betat = -1.25,
                        full.data = 0, random.generator = 1, seed = 0, nb.reject.data = 0)

## Not run:
# Surrogacy evaluation based on generated data with a combination of Monte Carlo
# and classical Gaussian Hermite integration.*
# (Computation takes around 5 minutes)
joint.surro.sim.MCGH <- jointSurroPenal(data = data.sim, int.method = 2,
                         nb.mc = 300, nb.gh = 20)

# Surrogacy evaluation based on generated data with a combination of Monte Carlo
# and Pseudo-adaptive Gaussian Hermite integration.
# (Computation takes around 4 minutes)
joint.surro.sim.MCPGH <- jointSurroPenal(data = data.sim, int.method = 2,
                         nb.mc = 300, nb.gh = 20, adaptatif = 1)

# Results
summary(joint.surro.sim.MCGH)
summary(joint.surro.sim.MCPGH)

# Data from the advanced ovarian cancer randomized clinical trials.
# Joint surrogate model with $\zeta$ fixed to 1, 8 nodes spline
# and the rescaled survival time.
jointSurroPenalSimul

Simulation studies based on the one-step Joint surrogate model for the evaluation of a candidate surrogate endpoint

Description

This function aims to allow simulation studies, based on the joint frailty surrogate model, described in jointSurroPenal

Usage

jointSurroPenalSimul(maxit=40, indicator.zeta = 1, indicator.alpha = 1, frail.base = 1, n.knots = 6, nb.dataset = 1, nbSubSimul=1000, ntrialSimul=30, LIMparam = 0.001, LIMlog1 = 0.001, LIMderiv = 0.001, nb.mc = 300, nb.gh = 32, nb.gh2 = 20, adaptatif = 0, int.method = 2, nb.iterPGH = 5, nb.MC.kendall = 10000, nboot.kendall = 1000, true.init.val = 0, theta.init = 1, sigma.ss.init = 0.5, sigma.tt.init = 0.5, sigma.st.init = 0.48, gamma.init = 0.5, alpha.init = 1, zeta.init = 1, betas.init = 0.5, betat.init = 0.5, random.generator = 1, equi.subj.trial = 1, prop.subj.trial = NULL,
equi.subj.trt = 1, prop.subj.trt = NULL,
theta2 = 3.5, zeta = 1, gamma.ui = 2.5, alpha.ui = 1,
betas = -1.25, betat = -1.25, lambdas = 1.8, nus = 0.0045,
lambdat = 3, nut = 0.0025, time.cens = 549, R2 = 0.81,
sigma.s = 0.7, sigma.t = 0.7, kappa.use = 4, random = 0,
random.nb.sim = 0, seed = 0, init.kappa = NULL,
nb.decimal = 4, print.times = TRUE, print.iter=FALSE)

Arguments

maxit maximum number of iterations for the Marquardt algorithm. Default is 40.
indicator.zeta A binary, indicates whether the power’s parameter $\zeta$ should be estimated (1) or not (0). If 0, $\zeta$ will be set to 1 during estimation. The default is 1. This parameter can be seted to 0 in case of identification issues.
indicator.alpha A binary, indicates whether the power’s parameter $\alpha$ should be estimated (1) or not (0). If 0, $\alpha$ will be set to 1 during estimation. The default is 1.
frail.base Considered the heterogeneity between trial on the baseline risk (1), using the shared cluster specific frailties ($u_i$), or not (0). The default is 1.
n.knots integer giving the number of knots to use. Value required in the penalized likelihood estimation. It corresponds to the (n.knots+2) splines functions for the approximation of the hazard or the survival functions. We estimate I or M-splines of order 4. When the user set a number of knots equals to k (n.knots=k) then the number of interior knots is (k-2) and the number of splines is (k-2)+order. Number of knots must be between 4 and 20. (See frailtyPenal for more details).
nb.dataset Number of dataset to analyze. The default is 1.
nbSubSimul Number of subjects.
ntrialSimul Number of trials.
LIMparam Convergence threshold of the Marquardt algorithm for the parameters, $10^{-3}$ by default (See frailtyPenal for more details).
LIMlogl Convergence threshold of the Marquardt algorithm for the log-likelihood, $10^{-3}$ by default (See frailtyPenal for more details).
LIMderiv Convergence threshold of the Marquardt algorithm for the gradient, $10^{-3}$ by default (See frailtyPenal for more details).
nb.mc Number of samples considered in the Monte-Carlo integration. Required in case int.method is equals to 0, 2 or 4. A value between 100 and 300 most often gives good results. However, beyond 300, the program takes a lot of time to estimate the parameters. The default is 300.
nb.gh Number of nodes for the Gaussian-Hermite quadrature. It can be chosen among 5, 7, 9, 12, 15, 20 and 32. The default is 32.
nb.gh2 Number of nodes for the Gauss-Hermite quadrature used to re-estimate the model, in case of non-convergence, defined as previously. The default is 20.
adaptatif A binary, indicates whether the pseudo adaptive Gaussian-Hermite quadrature (1) or the classical Gaussian-Hermite quadrature (0) is used. The default is 0.
jointSurroPenalSimul

int.method  A numeric, indicates the integration method: 0 for Monte carlo, 1 for Gaussian-Hermite quadrature, 2 for a combination of both Gaussian-Hermite quadrature to integrate over the individual-level random effects and Monte carlo to integrate over the trial-level random effects, 4 for a combination of both Monte carlo to integrate over the individual-level random effects and Gaussian-Hermite quadrature to integrate over the trial-level random effects. The default is 2.

nb.iterPGH  Number of iterations before the re-estimation of the posterior random effects, in case of the two-steps pseudo-adaptive Gaussian-hermite quadrature. If set to 0 there is no re-estimation”. The default is 5.

nb.MC.kendall  Number of generated points used with the Monte-Carlo to estimate integrals in the Kendall’s τ formulation. Better to use at least 4000 points for stable results. The default is 10000.

nboot.kendall  Number of samples considered in the parametric bootstrap to estimate the confidence interval of the Kendall’s τ. The default is 1000.

true.init.val  Numerical value. Indicates if the real parameter values (1), or the given initial values to parameters (0) should be considered. If set to 2, α and γ are initialised using two separated shared frailty model (see frailtyPenal for more details): $\sigma^2_{V_S}$, $\sigma^2_{V_T}$ and $\sigma^2_{V_{ST}}$ are fixed using the default initial values given by the user; $\zeta$, $\theta$, $\beta_S$ and $\beta_T$ are initialized using a classical joint frailty model, considering individual level random effects. If the joint frailty model is faced to convergence issues, $\beta_S$ and $\beta_T$ are initialized using two shared frailty models. In all others scenarios, if the simplified model does not converge, default given parameters values are used. Initial values for spline’s associated parameters are fixed to 0.5. The default for this argument is 0.

theta.init  Initial values for $\theta$, required if true.init.val is set to 0 or 2. The default is 1.

sigma.ss.init  Initial values for $\sigma^2_{V_S}$, required if true.init.val is set to 0 or 2. The default is 0.5.

sigma.tt.init  Initial values for $\sigma^2_{V_T}$, required if true.init.val is set to 0 or 2. The default is 0.5.

sigma.st.init  Initial values for $\sigma^2_{V_{ST}}$, required if true.init.val is set to 0 or 2. The default is 0.48.

gamma.init  Initial values for $\gamma$, required if true.init.val is set to 0 or 2. The default is 0.5.

alpha.init  Initial values for $\alpha$, required if true.init.val is set to 0 or 2. The default is 1.

zeta.init  Initial values for $\zeta$, required if true.init.val is set to 0 or 2. The default is 1.

betas.init  Initial values for $\beta_S$, required if true.init.val is set to 0 or 2. The default is 0.5.

betat.init  Initial values for $\beta_T$, required if true.init.val is set to 0 or 2. The default is 0.5.

random.generator  Random number generator to use by the Fortran compiler, 1 for the intrinsec subroutine Random_number and 2 for the subroutine un1ran(). The default is 1.
equi.subj.trial
A binary, that indicates if the same proportion of subjects per trial should be considered in the process of data generation (1) or not (0). In case of different trial sizes, fill in prop.subj.trial the proportions of subjects to be considered per trial. The default is 1.

prop.subj.trial
Vector of the proportions of subjects to consider per trial. Requires if the argument equi.subj.trial is different to 1. The size of this vector is equal to the number of trials.

equi.subj.trt
Indicates if the same proportion of treated subjects per trial should be considered (1) or not (0). If 0, fill in prop.subj.trt the proportions of treated subjects to be considered per trial. The default is 1.

prop.subj.trt
Vector of the proportions of treated subjects to consider per trial. Requires if the argument equi.subj.trt is different to 0.5. The size of this vector is equal to the number of trials.

theta2
True value for $\theta$. The default is 3.5.

zeta
True value for $\zeta$ in case of simulation. The default is 1.

gamma.ui
True value for $\gamma$ in case of simulation. The default is 2.5.

alpha.ui
True value for $\alpha$ in case of simulation. The default is 1.

betas
True value for $\beta_S$ in case of simulation. The default is -1.25.

betat
True value for $\beta_T$ in case of simulation. The default is -1.25.

lambdas
Desired scale parameter for the Weibull distribution associated with the Surrogate endpoint. The default is 1.8.

nus
Desired shape parameter for the Weibull distribution associated with the Surrogate endpoint. The default is 0.0045.

lambdat
Desired scale parameter for the Weibull distribution associated with the True endpoint. The default is 3.

nut
Desired shape parameter for the Weibull distribution associated with the True endpoint. The default is 0.0025.

time.cens
Censorship time. The default is 549, for about 40% of censored subjects.

R2
Desired $R^2_{trial}$. The default is 0.81.

sigma.s
True value for $\sigma^2_S$. The default is 0.7.

sigma.t
True value for $\sigma^2_T$. The default is 0.7.

kappa.use
A numeric, that indicates how to manage the smoothing parameters $k_1$ and $k_2$ in case of convergence issues. If it is set to 0, the first smoothing parameters that allowed convergence on the first dataset is used for all simulations. if it is set to 1, a smoothing parameter is estimated by cross-validation for each dataset generated. If it is set to 2, the same process for choosing kappas as in case 1 is used, but in case of convergence issue, the first smoothing parameters that allowed convergence among the three previous that have worked is used. If it is set to 3, the associated smoothing parameters are successively divided by 10, in case of convergence issues until 5 times. If it is set to 4, the management of the smoothing parameters is as in case 2, preceded by the successive division described in case 3 and by the changing of the number of nodes for the Gauss-Hermite quadrature. The default is 4.
A binary that says if we reset the random number generation with a different environment at each call (1) or not (0). If it is set to 1, we use the computer clock as seed. In the last case, it is not possible to reproduce the generated datasets. The default is 0. Required if random.generator is set to 1.

If random is set to 1, a binary that indicates the number of generations that will be made, equal to nb.dataset in this case.

The seed to use for data generation. Required if random is set to 0. The default is 0.

smoothing parameter used to penalized the log-likelihood. By default (init.kappa = NULL) the values used are obtain by cross-validation.

Number of decimal required for results presentation.

a logical parameter to print estimation time. Default is TRUE.

a logical parameter to print iteration process. Default is FALSE.

The estimated parameter are obtained using the robust Marquardt algorithm (Marquardt, 1963) which is a combination between a Newton-Raphson algorithm and a steepest descent algorithm. The iterations are stopped when the difference between two consecutive log-likelihoods was small (< 10^{-3}), the estimated coefficients were stable (consecutive values (< 10^{-3}), and the gradient small enough (< 10^{-3}), by default. Cubic M-splines of order 4 are used for the hazard function, and I-splines (integrated M-splines) are used for the cumulative hazard function.

The inverse of the Hessian matrix is the variance estimator and to deal with the positivity constraint of the variance component and the spline coefficients, a squared transformation is used and the standard errors are computed by the Δ-method (Knight & Xekalaki, 2000). The smooth parameter can be chosen by maximizing a likelihood cross validation criterion (Joly and other, 1998).

We proposed based on the joint surrogate model a new definition of the Kendall’s τ. Moreover, distinct numerical integration methods are available to approximate the integrals in the marginal log-likelihood.

Special attention must be given to initializing model parameters, the choice of the number of spline knots, the smoothing parameters and the number of quadrature points to solve convergence issues. We first initialized parameters using the user’s desired strategy, as specified by the option true.init.val. When numerical or convergence problems are encountered, with kappa.use set to 4, the model is fitted again using a combination of the following strategies: vary the number of quadrature point (nb. gh to nb. gh2 or nb. gh2 to nb. gh) in case of the use of the Gaussian Hermite quadrature integration (see int.method); divided or multiplied the smoothing parameters (k_1, k_2) by 10 or 100 according to their preceding values, or used parameter vectors obtained during the last iteration (with a modification of the number of quadrature points and smoothing parameters). Using this strategy, we usually obtained during simulation the rejection rate less than 3%. A sensitivity analysis was conducted without this strategy, and similar results were obtained on the converged samples, with about a 23% rejection rate.

This function return an object of class jointSurroPenalSimul with elements:
theta2  True value for θ;
zeta    true value for ζ;
gamma.ui true value for γ;
alpha.ui true value for α;
sigma.s  true value for σ_S;
sigma.t  true value for σ_T;
sigma.st true value for σ_{ST};
betas   true value for β_S;
betat   true value for β_T;
R2      true value for R^2_{trial};

nb.subject total number of subjects used;
nb.trials   total number of trials used;
nb.simul   number of simulated datasets;
nb.gh      number of nodes for the Gaussian-Hermite quadrature;
nb.gh2     number of nodes for the Gauss-Hermite quadrature used to re-estimate the model, in case of non-convergence;
nb.mc      number of samples considered in the Monte-Carlo integration;
kappa.use  a numeric, that indicates how to manage the smoothing parameters k_1 and k_2 in case of convergence issues;
n.knots    number of knots used for splines;
int.method integration method used;
n.iter     mean number of iterations needed to converge;
dataTkendall a matrix with nb.dataset line(s) and three columns, of the estimates of Kendall’s τ and theirs confidence intervals using the parametric bootstrap. All non-convergence cases are represented by a line of 0;
dataR2boot  a matrix with nb.dataset line(s) and three columns, of the estimates of R^2_{trial} and theirs confidence intervals using the parametric bootstrap. All non-convergence cases are represented by a line of 0.
dataParamEstim a dataframe including all estimates with the associated standard errors, for all simulation. All non-convergence cases are represented by a line of 0;
dataHessian  Dataframe of the variance-Covariance matrices of the estimates for all simulations
dataHessianIH Dataframe of the robust estimation of the variance matrices of the estimates for all simulations
dataab      Dataframe of the estimates for all simulations which rich convergence

Author(s)
Casimir Ledoux Sofeu <casimir.sofeu@u-bordeaux.fr>, <scl.ledoux@gmail.com> and Virginie Rondeau <virginie.rondeau@inserm.fr>
References


See Also

jointSurroPenal, summary.jointSurroPenalSimul, jointSurrSimul

Examples

```r
## Not run:
# Surroacy model evaluation performance study based on 10 generated data
# (Computation takes around 20 minutes using a processor including 40
# cores and a read only memory of 378 Go)
# To realize a simulation study on 100 samples or more (as required), use
# nb.dataset = 100

joint.simul &lt;- jointSurroPenalSimul(nb.dataset = 10, nbSubSimul=600,
ntrialSimul=30, LIMparam = 0.001, LIMlogl = 0.001,
LIMderiv = 0.001, nb.mc = 200, nb.gh = 20,
nb.gh2 = 32, true.init.val = 1, print.iter=F)

# results
summary(joint.simul, d = 3, R2boot = 1) # bootstrap
summary(joint.simul, d = 3, R2boot = 0) # Delta-method

## End(Not run)
```

jointSurroTKendall  Kendall's $\tau$ estimation using numerical integration methods

Description

This function estimate the Kendall’s $\tau$ based on the joint surrogate model described in jointSurroPenal (Sofeu et al., 2018), for the evaluation of a candidate surrogate endpoints, at the individual-level. We used the Monte-carlo and the gaussian Hermite quadrature methods for numerical integration. In case of Gaussian Hermite quadrature, it is better to choose at least 20 quadrature nodes for better results. The actual value of nodes used is the maximum between 20 and nb.gh

Usage

```
jointSurroTKendall(object = NULL, theta, gamma, alpha = 1, zeta = 1,
sigma.v = matrix(rep(0,4),2,2), int.method = 0,
nb.MC.kendall = 10000, nb.gh = 32,
random.generator = 1, random = 0,
random.nb.sim = 0, seed = 0, ui = 1)
```
Arguments

- **object**: An object inheriting from `jointSurroPenal` class. The default is `NULL`.
- **theta**: Variance of the individual-level random effect, \( \omega_{ij} \). Required if `object` is set to `NULL`.
- **gamma**: Variance of the trial-level random effect associated with the baseline risk, \( u_i \). Required if `object` is set to `NULL`. The default is `3.5`.
- **alpha**: Power parameter associated with \( u_i \). Required if `object` is set to `NULL`. The default is `1`.
- **zeta**: Power parameter associated with \( \omega_{ij} \). Required if `object` is set to `NULL`. The default is `1`.
- **sigma.v**: Covariance matrix of the random effects treatment-by-trial interaction \( (v_{S}, v_{T}) \).
- **int.method**: A numeric, indicates the integration method: `0` for Monte Carlo and `1` for Gaussian-Hermite quadrature. The default is `0`.
- **nb.MC.kendall**: Number of generated points used with the Monte-Carlo to estimate integrals in the Kendall’s \( \tau \) formulation. Better to use at least 4000 points for stable results. The default is `10000`.
- **nb.gh**: Number of nodes for the Gaussian-Hermite quadrature. The default is `32`.
- **random.generator**: Random number generator to use by the Fortran compiler, `1` for the intrinsic subroutine `Random_number` and `2` for the subroutine `uniran()`. The default is `1`.
- **random**: A binary that says if we reset the random number generation with a different environment at each call (1) or not (0). If it is set to 1, we use the computer clock as a seed. In the last case, it is not possible to reproduce the generated datasets. The default is `0`.
- **random.nb.sim**: If `random` is set to 1, a binary that indicates the number of generations that will be made.
- **seed**: The seed to use for data (or samples) generation. Required if `random` is set to 0. The default is `0`.
- **ui**: A binary, indicates whether one considered trial random effect associated with the baseline risk (1) or not (0). The default is `1`.

Value

This function return the estimated Kendall’s \( \tau \)

Author(s)

Casimir Ledoux Sofeu <casimir.sofeu@u-bordeaux.fr>, <scl.ledoux@gmail.com> and Virginie Rondeau <virginie.rondeau@inserm.fr>

References

See Also  
jointSurrSimul, summary.jointSurroPenal

Examples  
Ktau1 <- jointSurroTKendall(theta = 3.5, gamma = 2.5, nb.gh = 32)  
Ktau2 <- jointSurroTKendall(theta = 1, gamma = 0.8, alpha = 1, zeta = 1, nb.gh = 32)  

###---Kendall’s \( \tau \) from a joint surrogate model ---###  
data.sim <- jointSurrSimul(n.obs = 400, n.trial = 20, cens.adm = 549,  
alpha = 1.5, theta = 3.5, gamma = 2.5, zeta = 1, sigma.s = 0.7, sigma.t = 0.7, rsqrt = 0.8, betas = -1.25, betat = -1.25, full.data = 0, random.generator = 1, seed = 0, nb.reject.data = 0)  

## Not run:  
###---Estimation---###  
joint.surrogate <- jointSurroPenal(data = data.sim, nb.mc = 300, nb.gh = 20, indicator.alpha = 1, n.knots = 6)  
Ktau3 <- jointSurroTKendall(joint.surrogate)  
Ktau4 <- jointSurroTKendall(joint.surrogate, nb.MC.kendall = 4000, seed = 1)

## End(Not run)

---

jointSurrSimul  
Generate survival times for two endpoints using the joint frailty surrogate model

Description  
Date are generated from the one-step joint surrogate model (see jointSurroPenal for more details)

Usage  
jointSurrSimul(n.obs = 600, n.trial = 30, cens.adm = 549.24,  
alpha = 1.5, theta = 3.5, gamma = 2.5, zeta = 1, sigma.s = 0.7, sigma.t = 0.7, rsqrt = 0.8, betas = -1.25, betat = -1.25, frailt.base = 1, lambda.S = 1.8, nu.S = 0.0045, lambda.T = 3, nu.T = 0.0025, ver = 1, typeof = 1, equi.subj.trial = 1, equi.subj.trt = 1, prop.subj.trial = NULL, prop.subj.trt = NULL, full.data = 0, random.generator = 1, random = 0, random.nb.sim = 0, seed = 0, nb.reject.data = 0)
Arguments

- **n.obs**: Number of considered subjects. The default is 600.
- **n.trial**: Number of considered trials. The default is 30.
- **cens.adm**: Censorship time. The default is 549, for about 40% of censored subjects.
- **alpha**: Fixed value for $\alpha$. The default is 1.5.
- **theta**: Fixed value for $\theta$. The default is 3.5.
- **gamma**: Fixed value for $\gamma$. The default is 2.5.
- **zeta**: Fixed value for $\zeta$. The default is 1.
- **sigma.s**: Fixed value for $\sigma_s^2$. The default is 0.7.
- **sigma.t**: Fixed value for $\sigma_T^2$. The default is 0.7.
- **rsqrt**: Desired level of correlation between $v_{S_i}$ and $v_{T_i}$. $R^2_{trial} = rsqrt^2$. The default is 0.8.
- **betas**: Fixed value for $\beta_s$. The default is -1.25.
- **betat**: Fixed value for $\beta_T$. The default is -1.25.
- **frailt.base**: Considered the heterogeneity on the baseline risk (1) or not (0). The default is 1.
- **lambda.S**: Desired scale parameter for the Weibull distribution associated with the Surrogate endpoint. The default is 1.8.
- **nu.S**: Desired shape parameter for the Weibull distribution associated with the Surrogate endpoint. The default is 0.0045.
- **lambda.T**: Desired scale parameter for the Weibull distribution associated with the True endpoint. The default is 3.
- **nu.T**: Desired shape parameter for the Weibull distribution associated with the True endpoint. The default is 0.0025.
- **ver**: Number of covariates. For surrogate evaluation, we just considered one covariate, the treatment arm.
- **typeOf**: Type of joint model used for data generation: 0 = classical joint model with a shared individual frailty effect (Rondeau, 2007), 1 = joint surrogate model with shared frailty effects $u_i$ and $\omega_{ij}$, and two correlated random effects treatment-by-trial interaction $(v_{S_i}, v_{T_i})$ as described in Sofeu et al. (2018).
- **equi.subj.trial**: A binary variable that indicates if the same proportion of subjects should be included per trial (1) or not (0). If 0, the proportions of subject per trial are required in parameter **prop.subj.trial**.
- **equi.subj.trt**: A binary variable that indicates if the same proportion of subjects is randomized per trial (1) or not (0). If 0, the proportions of subject per trial are required in parameter **prop.subj.trt**.
- **prop.subj.trial**: The proportions of subjects per trial. Requires if **equi.subj.trial**=0.
- **prop.subj.trt**: The proportions of randomized subject per trial. Requires if **equi.subj.trt**=0.
jointSurSimul

full.data Specified if you wan the function to return the full dataset (1), including the random effects, or the restrictive dataset (0) with 7 columns required for the function jointSurroPenal.

random.generator Random number generator to use by the Fortran compiler, 1 for the intrinsic subroutine Random_number and 2 for the subroutine uniran(). The default is 1.

random A binary that says if we reset the random number generation with a different environment at each call (1) or not (0). If it is set to 1, we use the computer clock as seed. In the last case, it is not possible to reproduce the generated datasets". The default is 0. Required if random.generator is set to 1.

random.nb.sim required if random.generator is set to 1, and if random is set to 1.

seed The seed to use for data (or samples) generation. Required if the argument random.generator is set to 1. Must be a positive value. If negative, the program do not account for seed. The default is 0.

nb.reject.data Number of generation to reject before the considered dataset. this parameter is required when data generation is for simulation. With a fixed parameter and random.generator set to 1, all generated data are the same. By varying this parameter, different datasets are obtained during data generation. The default value is 0, in case of one dataset.

Details

We just considered in this generation, the Gaussian random effects. If the parameter full.data is set to 1, this function return a list containing several parameters, including the generated random effects. the desired individual level correlation (Kendall’s $\tau$) depend on the values of $\alpha$, $\theta$, $\gamma$ and $\zeta$.

Value

This function return if the parameter full.data is set to 0, a data.frame with columns:

- patientID A numeric, that represents the patient’s identifier, must be unique;
- trialID A numeric, that represents the trial in which each patient was randomized;
- trt The treatment indicator for each patient, with 1 = treated, 0 = untreated;
- timeS The follow up time associated with the surrogate endpoint;
- statusS The event indicator associated with the surrogate endpoint. Normally 0 = no event, 1 = event;
- timeT The follow up time associated with the true endpoint;
- statusT The event indicator associated with the true endpoint. Normally 0 = no event, 1 = event;

If the argument full.data is set to 1, additional columns corresponding to random effects $\omega_{ij}$, $u_i$, $v_{S_i}$ and $v_{T_i}$ are returned. Note that $u_i$, $v_{S_i}$ and $v_{T_i}$ are returned if typeOf is set to 1.

Author(s)

Casimir Ledoux Sofeu <casimir.sofeu@u-bordeaux.fr>, <scl.ledoux@gmail.com> and Virginie Rondeau <virginie.rondeau@inserm.fr>
longDat

References


See Also

jointSurrSimul

Examples

data.sim <- jointSurrSimul(n.obs=600, n.trial = 30, cens.adm=549.24, alpha = 1.5, theta = 3.5, gamma = 2.5, sigma.s = 0.7, zeta = 1, sigma.t = 0.7, rsqrt = 0.8, betas = -1.25, betat = -1.25, full.data = 0, random.generator = 1, seed = 0, nb.reject.data = 0)

longDat

Longitudinal semicontinuous biomarker dataset (TPJM)

Description

This is a simulated dataset used to illustrate the two-part joint model included in the longiPenal function.

Usage

data(longDat)

Format

This data frame contains the following columns:

- **id**: The identification number of a patient
- **timej**: The measurement times of the biomarker
- **trtY**: Treatment covariate
- **Y**: Biomarker value
**longiPenal**

*Fit a Joint Model for Longitudinal Data and a Terminal Event*

**Description**

Fit a joint model for longitudinal data and a terminal event using a semiparametric penalized likelihood estimation or a parametric estimation on the hazard function.

The longitudinal outcomes $y_i(t_{ik}) (k = 1, \ldots, n_i, i = 1, \ldots, N)$ for $N$ subjects are described by a linear mixed model and the risk of the terminal event is represented by a proportional hazard risk model. The joint model is constructed assuming that the processes are linked via a latent structure (Wulfsohn and Tsiatis 1997):

\[
\begin{align*}
\{ y_i(t_{ik}) &= X_{Li}(t_{ik})^\top \beta_L + Z_i(t_{ik})^\top b_i + \epsilon_i(t_{ik}) \quad \text{(Longitudinal)} \\
\lambda_i(t|b_i) &= \lambda_0(t) \exp(X_{Ti}(t)\beta_T + h(b_i, \beta_L, Z_i(t), X_{Li}(t))^\top \eta_T) \quad \text{(Terminal)}
\end{align*}
\]

where $X_{Li}(t)$ and $X_{Ti}$ are vectors of fixed effects covariates and $\beta_L$ and $\beta_T$ are the associated coefficients. Measurements errors $\epsilon_i(t_{ik})$ are iid normally distributed with mean 0 and variance $\sigma^2$. The random effects $b_i = (b_{0i}, \ldots, b_{qi})^\top \sim \mathcal{N}(0, B_1)$ are associated to covariates $Z_i(t)$ and independent from the measurement error. The relationship between the two processes is explained via $h(b_i, \beta_L, Z_i(t), X_{Li}(t))$ with coefficients $\eta_T$. Two forms of the function $h(\cdot)$ are available: the random effects $b_i$ and the current biomarker level $m_i(t) = X_{Li}(t_{ik})^\top \beta_L + Z_i(t_{ik})^\top b_i$.

We consider that the longitudinal outcome can be a subject to a quantification limit, i.e. some observations, below a level of detection $s$ cannot be quantified (left-censoring).

**Usage**

```r
longiPenal(formula, formula.LongitudinalData, data, data.Longi, 
formula.Binary=FALSE, random, random.Binary=FALSE, id, intercept = TRUE, 
link = "Random-effects", timevar=FALSE, left.censoring = 
FALSE, n.knots, kappa, maxit = 350, hazard = "Splines", init.B, 
init.Random, init.Eta, method.GH = "Standard", seed.MC=FALSE, n.nodes, LIMparam = 1e-3, 
LIMlogl = 1e-3, LIMderiv = 1e-3, print.times = TRUE)
```

**Arguments**

- `formula`: 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 like in survival package. Interactions are possible using `*` or `:`.

- `formula.LongitudinalData`: a formula object, only requires terms on the right to indicate which variables are modelling the longitudinal outcome. It must follow the standard form used for linear mixed-effects models. Interactions are possible using `*` or `:`.

- `data`: a 'data.frame' with the variables used in `formula`.

- `data.Longi`: a 'data.frame' with the variables used in `formula.LongitudinalData`. 
formula.Binary  a formula object, only requires terms on the right to indicate which variables are modelling the binary part of the two-part model fitting the longitudinal semicontinuous outcome. It must follow the standard form used for linear mixed-effects models. Interactions are possible using * or :.

random  Names of variables for the random effects of the longitudinal outcome. Maximum 3 random effects are possible at the moment. The random intercept is chosen using "1".

random.Binary  Names of variables for the random effects of the binary part of the two-part model fitting the longitudinal semicontinuous outcome. The random intercept is chosen using "1".

id  Name of the variable representing the individuals.

intercept  Logical value. Is the fixed intercept of the biomarker included in the mixed-effects model? The default is TRUE.

link  Type of link function for the dependence between the biomarker and death: "Random-effects" for the association directly via the random effects of the biomarker, "Current-level" for the association via the true current level of the biomarker. The option "Current-level" can be chosen only if the biomarker random effects are associated with the intercept and time (following this order). "Two-part", this structure is only applicable with two-part models, the effect of the current probability of positive value and the effect of the expected value among positive values on the risk of event is evaluated separately. The default is "Random-effects".

timevar  Indicates the time varying variables to take into account this evolution over time in the link with the survival model (useful with 'Current-level' and 'Two-part' links)

left.censoring  Is the biomarker left-censored below a threshold s? The default is FALSE, ie. no left-censoring. In case of a left-censored biomarker, this argument must be equal to the threshold s.

n.knots  Integer giving the number of knots to use. Value required in the penalized likelihood estimation. It corresponds to the (n.knots+2) splines functions for the approximation of the hazard or the survival functions. We estimate I or M-splines of order 4. When the user set a number of knots equals to k (n.knots=k) then the number of interior knots is (k-2) and the number of splines is (k-2)+order. Number of knots must be between 4 and 20. (See Note in frailtyPenal function)

kappa  Positive smoothing parameter in the penalized likelihood estimation. The coefficient kappa of the integral of the squared second derivative of hazard function in the fit (penalized log likelihood). To obtain an initial value for kappa, a solution is to fit the corresponding Cox model using cross validation (See cross.validation in function frailtyPenal). We advise the user to identify several possible tuning parameters, note their defaults and look at the sensitivity of the results to varying them.

maxit  Maximum number of iterations for the Marquardt algorithm. The default is 350.

hazard  Type of hazard functions: "Splines" for semiparametric hazard functions using equidistant intervals or "Splines-per" using percentile with the penalized likelihood estimation, "Weibull" for the parametric Weibull functions. The default is "Splines".
init.B Vector of initial values for regression coefficients. This vector should be of the same size as the whole vector of covariates with the first elements for the covariates related to the terminal event and then for the covariates related to the biomarker (interactions in the end of each component). Default is 0.5 for each.

init.Random Initial value for variance of the elements of the matrix of the distribution of the random effects. Default is 0.5 for each element.

init.Eta Initial values for regression coefficients for the link function. Default is 0.5 for each.


seed.MC Monte-carlo integration points selection (1=fixed, 0=random)

n.nodes Number of nodes for the Gauss-Hermite quadrature or the Monte-carlo method. They can be chosen among 5, 7, 9, 12, 15, 20 and 32 for the GH quadrature and any number for the Monte-carlo method. The default is 9.

LIMparam Convergence threshold of the Marquardt algorithm for the parameters (see Details of frailtyPenal function), $10^{-3}$ by default.

LIMlogl Convergence threshold of the Marquardt algorithm for the log-likelihood (see Details of frailtyPenal function), $10^{-3}$ by default.

LIMderiv Convergence threshold of the Marquardt algorithm for the gradient (see Details of frailtyPenal function), $10^{-3}$ by default.

print.times a logical parameter to print iteration process. The default is TRUE.

Details

Typical usage for the joint model

longiPenal(Surv(time,event)~var1+var2, biomarker ~ var1+var2, data, data.Longi, ...)

The method of the Gauss-Hermite quadrature for approximations of the multidimensional integrals, i.e. length of random is 2, can be chosen among the standard, non-adaptive, pseudo-adaptive in which the quadrature points are transformed using the information from the fitted mixed-effects model for the biomarker (Rizopoulos 2012) or multivariate non-adaptive procedure proposed by Genz et al. 1996 and implemented in FORTRAN subroutine HRMSYM. The choice of the method is important for estimations. The standard non-adaptive Gauss-Hermite quadrature ("Standard") with a specific number of points gives accurate results but can be time consuming. The non-adaptive procedure ("HRMSYM") offers advantageous computational time but in case of datasets in which some individuals have few repeated observations (biomarker measures or recurrent events), this method may be moderately unstable. The pseudo-adaptive quadrature uses transformed quadrature points to center and scale the integrand by utilizing estimates of the random effects from an appropriate linear mixed-effects model. This method enables using less quadrature points while preserving the estimation accuracy and thus lead to a better computational time. The Monte-Carlo method is also proposed for approximations of the multidimensional integrals.

NOTE. Data frames data and data.Longi must be consistent. Names and types of corresponding covariates must be the same, as well as the number and identification of individuals.
Value

The following components are included in a ‘longiPenal’ object for each model:

- `b` The sequence of the corresponding estimation of the coefficients for the hazard functions (parametric or semiparametric), the random effects variances and the regression coefficients.
- `call` The code used for the model.
- `formula` The formula part of the code used for the terminal event part of the model.
- `formula.LongitudinalData` The formula part of the code used for the longitudinal part of the model.
- `formula.Binary` The formula part of the code used for the binary part of the two-part model.
- `coef` The regression coefficients (first for the terminal event and then for the biomarker.
- `groups` The number of groups used in the fit.
- `kappa` The value of the smoothing parameter in the penalized likelihood estimation corresponding to the baseline hazard function for the terminal event.
- `logLikPenal` The complete marginal penalized log-likelihood in the semiparametric case.
- `logLik` The marginal log-likelihood in the parametric case.
- `n.measurements` The number of biomarker observations used in the fit.
- `max_rep` The maximal number of repeated measurements per individual.
- `n.deaths` The number of events observed in the fit.
- `n.iter` The number of iterations needed to converge.
- `n.knots` The number of knots for estimating the baseline hazard function in the penalized likelihood estimation.
- `n.strat` The number of stratum.
- `varH` The variance matrix of all parameters (before positivity constraint transformation for the variance of the measurement error, for which the delta method is used).
- `varHIH` The robust estimation of the variance matrix of all parameters.
- `xD` The vector of times where both survival and hazard function of the terminal event are estimated. By default seq(0,max(time),length=99), where time is the vector of survival times.
- `lamD` The array (dim=3) of baseline hazard estimates and confidence bands (terminal event).
- `survD` The array (dim=3) of baseline survival estimates and confidence bands (terminal event).
- `median` The value of the median survival and its confidence bands.
- `typeof` The type of the baseline hazard functions (0:“Splines”, “2:Weibull”).
- `npar` The number of parameters.
- `nvar` The vector of number of explanatory variables for the terminal event and biomarker.
- `nvarEnd` The number of explanatory variables for the terminal event.
The number of explanatory variables for the biomarker.

noVarEnd The indicator of absence of the explanatory variables for the terminal event.

noVarY The indicator of absence of the explanatory variables for the biomarker.

LCV The approximated likelihood cross-validation criterion in the semiparametric case (with H minus the converged Hessian matrix, and l(.) the full log-likelihood).

\[
LCV = \frac{1}{n} \text{trace}(H^{-1}H) - l(.)
\]

AIC The Akaike information Criterion for the parametric case.

\[
AIC = \frac{1}{n}(np - l(.)
\]

n.knots.temp The initial value for the number of knots.

shape.weib The shape parameter for the Weibull hazard function.

scale.weib The scale parameter for the Weibull hazard function.

martingaledeath.res The martingale residuals for each individual.

conditional.res The conditional residuals for the biomarker (subject-specific): \( R_i^{(m)} = y_i - X_{Li}^T \hat{\beta}_L - Z_i^T \hat{b}_i \).

marginal.res The marginal residuals for the biomarker (population averaged): \( R_i^{(c)} = y_i - X_{Li}^T \hat{\beta}_L \).

marginal.chol.res The Cholesky marginal residuals for the biomarker: \( R_i^{(m)} = U_i^{(m)} R_i^{(m)} \), where \( U_i^{(m)} \) is an upper-triangular matrix obtained by the Cholesky decomposition of the variance matrix \( V_{R_i^{(m)}} = \hat{V}_i - X_{Li} (\sum_{i=1}^{N} X_{Li} \hat{V}_i - 1) X_{Li}^{-1} X_{Li} \).

conditional.st.res The standardized conditional residuals for the biomarker.

marginal.st.res The standardized marginal residuals for the biomarker.

random.effects.pred The empirical Bayes predictions of the random effects (ie. using conditional posterior distributions).

pred.y.marg The marginal predictions of the longitudinal outcome.

pred.y.cond The conditional (given the random effects) predictions of the longitudinal outcome.

linear.death.pred The linear predictor for the terminal part.

global.chisq_d The vector with values of each multivariate Wald test for the terminal part.

dof.chisq_d The vector with degrees of freedom for each multivariate Wald test for the terminal part.
global_chisq.test_d
   The binary variable equals to 0 when no multivariate Wald is given, 1 otherwise (for the terminal part).

p.global_chisq_d
   The vector with the p_values for each global multivariate Wald test for the terminal part.

global_chisq
   The vector with values of each multivariate Wald test for the longitudinal part.

dof_chisq
   The vector with degrees of freedom for each multivariate Wald test for the longitudinal part.

global_chisq.test
   The binary variable equals to 0 when no multivariate Wald is given, 1 otherwise (for the longitudinal part).

p.global_chisq
   The vector with the p_values for each global multivariate Wald test for the longitudinal part.

names.factordc
   The names of the "as.factor" variables for the terminal part.

names.factor
   The names of the "as.factor" variables for the longitudinal part.

intercept
   The logical value. Is the fixed intercept included in the linear mixed-effects model?

B1
   The variance matrix of the random effects for the longitudinal outcome.

ResidualSE
   The standard deviation of the measurement error.

eta
   The regression coefficients for the link function.

ne_re
   The number of random effects used in the fit.

names.re
   The names of variables for the random effects.

link
   The name of the type of the link function.

eta_p.value
   p-values of the Wald test for the estimated regression coefficients for the link function.

beta_p.value
   p-values of the Wald test for the estimated regression coefficients.

leftCensoring
   The logical value. Is the longitudinal outcome left-censored?

leftCensoring.threshold
   For the left-censored biomarker, the value of the left-censoring threshold used for the fit.

prop.censored
   The fraction of observations subjected to the left-censoring.

methodGH
   The method used for approximations of the multidimensional integrals.

n.nodes
   The number of integration points.

References


D. Rustand, L. Briollais, C. Tournigand and V. Rondeau. Two-part joint model for a longitudinal semicontinuous marker and a terminal event with application to metastatic colorectal cancer data. *Under revision*.

See Also

plot.longiPenal, print.longiPenal, summary.longiPenal

Examples

```r
## Not run:

###--- Joint model for longitudinal data and a terminal event ---###

data(colorectal)
data(colorectalLongi)

# Survival data preparation - only terminal events
colorectalSurv <- subset(colorectal, new.lesions == 0)

# Baseline hazard function approximated with splines
# Random effects as the link function
model.spli.RE <- longiPenal(Surv(time1, state) ~ age + treatment + who.PS + prev.resection, tumor.size ~ year * treatment + age + who.PS , data=colorectalSurv, data.Longi = colorectalLongi, random = c("1", "year"), id = "id", link = "Random-effects", left.censoring = -3.33, n.knots = 7, kappa = 2)

# Weibull baseline hazard function
# Current level of the biomarker as the link function
model.weib.CL <- longiPenal(Surv(time1, state) ~ age + treatment + who.PS + prev.resection, tumor.size ~ year * treatment + age + who.PS , timevar="year", data=colorectalSurv, data.Longi = colorectalLongi, random = c("1", "year"), id = "id", link = "Current-level", left.censoring = -3.33, hazard = "Weibull")
```
loocv

### Two-part Joint model for semicontinuous longitudinal data and a terminal event ---

data(colorectal)
data(colorectalLongi)
colorectalSurv <- subset(colorectal, new.lesions == 0)

# Box-cox back transformation (lambda=0.3) and apply logarithm (with a 1 unit shift)
colorectalLongi$Yo <- (colorectalLongi$tumor.size*0.3+1)^(1/0.3)
colorectalLongi$Y <- log(colorectalLongi$Y+1) # log transformation with shift=1

# Two-part joint model - random-effects association structure (~15min)
TwoPartJoint_re <- longiPenal(Surv(time1, state)~age + treatment + who.PS+ prev.resection, Y~year*treatment, formula.Binary=Y~year*treatment, data = colorectalSurv, data.Longi = colorectalLongi, random = c("1"), random.Binary=c("1"), id = "id", link ="Random-effects", left.censoring = F, n.knots = 7, kappa = 2, hazard="Splines-per")
print(TwoPartJoint_re)

# Two-part joint model - current-level association structure (~15min)
# Simulated dataset (github.com/DenisRustand/TPJM_sim)
data(longDat)
data(survDat)
tte <- frailtyPenal(Surv(deathTimes, d)~trt, n.knots=5, kappa=0, data=survDat, cross.validation = T)
kap <- round(tte$kappa,2);kap # smoothing parameter
TPJM <- longiPenal(Surv(deathTimes, d)~trt, Y~timej*trtY, data=survDat, data.Longi = longDat, random = c("1","timej"), formula.Binary=Y~timej*trtY, random.Binary=c("1"), timevar="timej", id = "id", link = "Current-level", n.knots = 5, kappa = kap, hazard="Splines-per", method.GH="Monte-carlo", n.nodes=500, seed.MC=1)
print(TPJM)

## End(Not run)

---

### loocv

Leave-one-out crossvalidation for the one-step Joint surrogate model for the evaluation of a candidate surrogate endpoint.

#### Description

Leave-one-out crossvalidation for the evaluation of the joint surrogate model

#### Usage

loocv(object, unusedtrial, var.used = "error.estim", alpha = 0.05, dec = 3, print.times = TRUE)
Arguments

object  An object inheriting from jointSurroPenal class (output from calling the function jointSurroPenal).

unusedtrial  A list of trial not to be taken into account in the cross-validation. This parameter is useful when after excluding some trials, the model is facing convergence problem.

var.used  This argument takes two values. The first one is "error.estim" and indicates if the prediction variance takes into account the estimation errors from the estimates of the parameters. If estimates are supposed to be known or if the dataset includes a high number of trials with a high number of subject per trial, value "No.error" can be used. The default is error.estim.

alpha. The confidence level for the prediction interval. The default is 0.05

dec  The desired number of digits after the decimal point for parameters and confidence intervals. Default of 3 digits is used.

print.times  a logical parameter to print estimation time. Default is TRUE.

Value

Returns an object of class jointSurroPenallooocv containing a dataframe (result) including for each trial the number of included subjects, the observed treatment effect on the surrogate endpoint, the observed treatment effect on the true endpoint and the predicted treatment effect on the true endpoint with the associated prediction intervals. If the observed treatment effect on the true endpoint is included into the prediction interval, the last columns contains "*".

Author(s)

Casimir Ledoux Sofeu <casimir.sofeu@u-bordeaux.fr>, <scl.ledoux@gmail.com> and Virginie Rondeau <virginie.rondeau@inserm.fr>

References


See Also

groupSurroPenal

Examples

## Not run:
# Generation of data to use
data.sim <- jointSurrSimul(n.obs=600, n.trial = 30,cens.adm=549.24, 
alpha = 1.5, theta = 3.5, gamma = 2.5, zeta = 1, sigma.s = 0.7, 
sigma.t = 0.7, rsqrt = 0.8, betas = -1.25, betat = -1.25, 
full.data = 0, random.generator = 1, seed = 0,
multivPenal

Fit a multivariate frailty model for two types of recurrent events and a terminal event.

Description

Fit a multivariate frailty model for two types of recurrent events with a terminal event using a penalized likelihood estimation on the hazard function or a parametric estimation. Right-censored data are allowed. Left-truncated data and stratified analysis are not possible. Multivariate frailty models allow studying, with a joint model, three survival dependent processes for two types of recurrent events and a terminal event. Multivariate joint frailty models are applicable in mainly two settings. First, when focus is on the terminal event and we wish to account for the effect of previous endogenous recurrent event. Second, when focus is on a recurrent event and we wish to correct for informative censoring.

The multivariate frailty model for two types of recurrent events with a terminal event is (in the calendar or time-to-event timescale):

\[
\begin{align*}
    r_i^{(1)}(t|u_i, v_i) &= r_0^{(1)}(t) \exp(\beta_1' Z_i(t) + u_i) \quad \text{(rec. of type 1)} \\
    r_i^{(2)}(t|u_i, v_i) &= r_0^{(2)}(t) \exp(\beta_2' Z_i(t) + v_i) \quad \text{(rec. of type 2)} \\
    \lambda_i(t|u_i, v_i) &= \lambda_0(t) \exp(\beta_3' Z_i(t) + \alpha_1 u_i + \alpha_2 v_i) \quad \text{(death)}
\end{align*}
\]

where \( r_0^{(l)}(t), l \in 1, 2 \) and \( \lambda_0(t) \) are respectively the recurrent and terminal event baseline hazard functions, and \( \beta_1, \beta_2, \beta_3 \) the regression coefficient vectors associated with \( Z_i(t) \) the covariate vector. The covariates could be different for the different event hazard functions and may be time-dependent. We consider that death stops new occurrences of recurrent events of any type, hence given \( t > D \), \( dN^{R(l)}(t), l \in 1, 2 \) takes the value 0. Thus, the terminal and the two recurrent event processes are not independent or even conditional upon frailties and covariates. We consider the hazard functions of recurrent events among individuals still alive. components in the above multivariate frailty model are linked together by two Gaussian and correlated random effects \( u_i, v_i \):

\[
(u_i, v_i)^T \sim \mathcal{N}(0, \Sigma_{uv}),
\]

with

\[
\Sigma_{uv} = \begin{pmatrix}
    \theta_1 & \rho \sqrt{\theta_1 \theta_2} \\
    \rho \sqrt{\theta_1 \theta_2} & \theta_2
\end{pmatrix}
\]
Dependencies between these three types of event are taken into account by two correlated random effects and parameters \( \theta_1, \theta_2 \) the variance of the random effects and \( \alpha_1, \alpha_2 \) the coefficients for these random effects into the terminal event part. If \( \alpha_1 \) and \( \theta_1 \) are both significantly different from 0, then the recurrent events of type 1 and death are significantly associated (the sign of the association is the sign of \( \alpha_1 \)). If \( \alpha_2 \) and \( \theta_2 \) are both significantly different from 0, then the recurrent events of type 2 and death are significantly associated (the sign of the association is the sign of \( \alpha_2 \)). If \( \rho \), the correlation between the two random effects, is significantly different from 0, then the recurrent events of type 1 and the recurrent events of type 2 are significantly associated (the sign of the association is the sign of \( \rho \)).

**Usage**

```r
multivPenal(formula, formula.Event2, formula.terminalEvent, data, initialize = TRUE, recurrentAG = FALSE, n.knots, kappa, maxit = 350, hazard = "Splines", nb.int, print.times = TRUE)
```

**Arguments**

- `formula`: a formula object, with the response for the first recurrent event 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 like in survival package. Interactions are possible using * or :.

- `formula.Event2`: a formula object, with the response for the second recurrent event 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 like in survival package. Interactions are possible using * or :.

- `formula.terminalEvent`: a formula object, with the response for the terminal event 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 like in survival package.

- `data`: a 'data.frame' with the variables used in 'formula', 'formula.Event2' and 'formula.terminalEvent'.

- `initialize`: Logical value to initialize regression coefficients and baseline hazard functions parameters. When the estimation is semi-parametric with splines, this initialization produces also values for smoothing parameters (by cross validation). When initialization is requested, the program first fit two shared frailty models (for the two types of recurrent events) and a Cox proportional hazards model (for the terminal event). Default is TRUE.

- `recurrentAG`: Logical value. Is Andersen-Gill model fitted? If so indicates that recurrent event times with the counting process approach of Andersen and Gill is used. This formulation can be used for dealing with time-dependent covariates. The default is FALSE.

- `n.knots`: integer vector of length 3 (for the three outcomes) giving the number of knots to use. First is for the recurrent of type 1, second is for the recurrent of type 2 and third is for the terminal event hazard function. Value required in the penalized likelihood estimation. It corresponds to the (n.knots+2) splines functions for the approximation of the hazard or the survival functions. Number of knots must be between 4 and 20. (See Note)
kappa

Vector of length 3 (for the three outcomes) for positive smoothing parameters in the penalized likelihood estimation. First is for the recurrent of type 1, second is for the recurrent of type 2 and third is for the terminal event hazard function. The coefficient kappa of the integral of the squared second derivative of hazard function in the fit (penalized log likelihood). Initial values for the kappas can be obtained with the option "initialize=TRUE". We advise the user to identify several possible tuning parameters, note their defaults and look at the sensitivity of the results to varying them. Value required. (See Note)

maxit

Maximum number of iterations for the Marquardt algorithm. Default is 350.

hazard

Type of hazard functions: "Splines" for semi-parametric hazard functions with the penalized likelihood estimation, "Piecewise-per" for piecewise constant hazard function using percentile, "Piecewise-equ" for piecewise constant hazard function using equidistant intervals, "Weibull" for parametric Weibull function. Default is "Splines".

nb.int

An integer vector of length 3 (for the three outcomes). First is the number of intervals (between 1 and 20) for the recurrent of type 1 parametric hazard functions ("Piecewise-per", "Piecewise-equ"). Second is the number of intervals (between 1 and 20) for the recurrent of type 2 parametric hazard functions ("Piecewise-per", "Piecewise-equ"). Third is number of intervals (between 1 and 20) for the death parametric hazard functions ("Piecewise-per", "Piecewise-equ")

print.times

A logical parameter to print iteration process. Default is TRUE.

**Value**

Parameters estimates of a multivariate joint frailty model, more generally a 'multivPenal' object. Methods defined for 'multivPenal' objects are provided for print, plot and summary. The following components are included in a 'multivPenal' object for multivariate Joint frailty models.

b

Sequence of the corresponding estimation of the splines coefficients, the random effects variances, the coefficients of the frailties and the regression coefficients.

call

The code used for fitting the model.

n

The number of observations used in the fit.

groups

The number of subjects used in the fit.

n.events

The number of recurrent events of type 1 observed in the fit.

n.events2

The number of recurrent events of type 2 observed in the fit.

n.deaths

The number of deaths observed in the fit.

loglikPenal

The complete marginal penalized log-likelihood in the semi-parametric case.

loglik

The marginal log-likelihood in the parametric case.

LCV

The approximated likelihood cross-validation criterion in the semi-parametric case (with H minus the converged Hessian matrix, and l(.) the full log-likelihood.

\[
LCV = \frac{1}{n} (\text{trace}(H^{-1}pl H) - l(.))
\]
AIC \quad \text{the Akaike information Criterion for the parametric case.}

\[ AIC = \frac{1}{n} (np - l(\cdot)) \]

theta1 \quad \text{variance of the frailty parameter for recurrences of type 1 (Var(u_i))}

theta2 \quad \text{variance of the frailty parameter for recurrences of type 2 (Var(v_i))}

alpha1 \quad \text{the coefficient associated with the frailty parameter } u_i \text{ in the terminal hazard function.}

alpha2 \quad \text{the coefficient associated with the frailty parameter } v_i \text{ in the terminal hazard function.}

rho \quad \text{the correlation coefficient between } u_i \text{ and } v_i

npar \quad \text{number of parameters.}

coeff \quad \text{the regression coefficients.}

nvar \quad \text{A vector with the number of covariates of each type of hazard function as components.}

varH \quad \text{the variance matrix of all parameters before positivity constraint transformation (theta, the regression coefficients and the spline coefficients). Then, the delta method is needed to obtain the estimated variance parameters.}

varHIIH \quad \text{the robust estimation of the variance matrix of all parameters (theta, the regression coefficients and the spline coefficients).}

formula \quad \text{the formula part of the code used for the model for the recurrent event.}

formula.Event2 \quad \text{the formula part of the code used for the model for the second recurrent event.}

formula.terminalEvent \quad \text{the formula part of the code used for the model for the terminal event.}

x1 \quad \text{vector of times for hazard functions of the recurrent events of type 1 are estimated. By default seq(0,max(time),length=99), where time is the vector of survival times.}

lam1 \quad \text{matrix of hazard estimates and confidence bands for recurrent events of type 1.}

xSu1 \quad \text{vector of times for the survival function of the recurrent event of type 1.}

surv1 \quad \text{matrix of baseline survival estimates and confidence bands for recurrent events of type 1.}

x2 \quad \text{vector of times for the recurrent event of type 2 (see x1 value).}

lam2 \quad \text{the same value as lam1 for the recurrent event of type 2.}

xSu2 \quad \text{vector of times for the survival function of the recurrent event of type 2}

surv2 \quad \text{the same value as surv1 for the recurrent event of type 2.}

xEnd \quad \text{vector of times for the terminal event (see x1 value).}

lamEnd \quad \text{the same value as lam1 for the terminal event.}

xSuEnd \quad \text{vector of times for the survival function of the terminal event}

survEnd \quad \text{the same value as surv1 for the terminal event.}

median1 \quad \text{The value of the median survival and its confidence bands for the recurrent event of type 1.}
median2  The value of the median survival and its confidence bands for the recurrent event of type 2.
medianEnd The value of the median survival and its confidence bands for the terminal event.
type.of.Piecewise Type of Piecewise hazard functions (1:"percentile", 0:"equidistant").
n.iter number of iterations needed to converge.
type.of.hazard Type of hazard functions (0:"Splines", "1:Piecewise", "2:Weibull").
n.knots a vector with number of knots for estimating the baseline functions.
kappa a vector with the smoothing parameters in the penalized likelihood estimation corresponding to each baseline function as components.
n.knots.temp initial value for the number of knots.
zi splines knots.
time knots for Piecewise hazard function for the recurrent event of type 1.
timedc knots for Piecewise hazard function for the terminal event.
time2 knots for Piecewise hazard function for the recurrent event of type 2.
noVar indicator vector for recurrent, death and recurrent 2 explanatory variables.
nvarRec number of the recurrent of type 1 explanatory variables.
nvarEnd number of death explanatory variables.
nvarRec2 number of the recurrent of type 2 explanatory variables.
nbintervR Number of intervals (between 1 and 20) for the the recurrent of type 1 parametric hazard functions ("Piecewise-per", "Piecewise-equi").
nbintervDC Number of intervals (between 1 and 20) for the death parametric hazard functions ("Piecewise-per", "Piecewise-equi").
nbintervR2 Number of intervals (between 1 and 20) for the the recurrent of type 2 parametric hazard functions ("Piecewise-per", "Piecewise-equi").
istop Vector of the convergence criteria.
shape.weib shape parameters for the Weibull hazard function.
scale.weib scale parameters for the Weibull hazard function.
martingale.res martingale residuals for each cluster (recurrent of type 1).
martingale2.res martingale residuals for each cluster (recurrent of type 2).
martingaledeath.res martingale residuals for each cluster (death).
frailty.pred empirical Bayes prediction of the first frailty term.
frailty2.pred empirical Bayes prediction of the second frailty term.
frailty.var variance of the empirical Bayes prediction of the first frailty term.
frailty2.var variance of the empirical Bayes prediction of the second frailty term.
frailty.corr Correlation between the empirical Bayes prediction of the two frailty.
linear.pred linear predictor: uses Beta'X + ui in the multivariate frailty models.
linear2.pred  linear predictor: uses Beta'X + vi in the multivariate frailty models.
lineardeath.pred linear predictor for the terminal part form the multivariate frailty models: Beta'X + alpha1 ui + alpha2 vi

global_chisq  Recurrent event of type 1: a vector with the values of each multivariate Wald test.
dof_chisq    Recurrent event of type 1: a vector with the degree of freedom for each multivariate Wald test.
global_chisq.test Recurrent event of type 1: a binary variable equals to 0 when no multivariate Wald is given, 1 otherwise.
p.global_chisq Recurrent event of type 1: a vector with the p-values for each global multivariate Wald test.
names.factor Recurrent event of type 1: Names of the "as.factor" variables.
global_chisq2 Recurrent event of type 2: a vector with the values of each multivariate Wald test.
dof_chisq2 Recurrent event of type 2: a vector with the degree of freedom for each multivariate Wald test.
global_chisq.test2 Recurrent event of type 2: a binary variable equals to 0 when no multivariate Wald is given, 1 otherwise.
p.global_chisq2 Recurrent event of type 2: a vector with the p-values for each global multivariate Wald test.
names.factor2 Recurrent event of type 2: Names of the "as.factor" variables.
global_chisq_d Terminal event: a vector with the values of each multivariate Wald test.
dof_chisq_d Terminal event: a vector with the degree of freedom for each multivariate Wald test.
global_chisq.test_d Terminal event: a binary variable equals to 0 when no multivariate Wald is given, 1 otherwise.
p.global_chisq_d Terminal event: a vector with the p-values for each global multivariate Wald test.
names.factordc Terminal event: Names of the "as.factor" variables.

Note

"kappa" (kappa[1], kappa[2] and kappa[3]) and "n.knots" (n.knots[1], n.knots[2] and n.knots[3]) are the arguments that the user has to change if the fitted model does not converge. "n.knots" takes integer values between 4 and 20. But with n.knots=20, the model will take a long time to converge. So, usually, begin first with n.knots=7, and increase it step by step until it converges. "kappa" only takes positive values. So, choose a value for kappa (for instance 10000), and if it does not converge, multiply or divide this value by 10 or 5 until it converges. Moreover, it may be useful to change the value of the initialize argument.
References


See Also

terminal, event2, print.multivPenal, summary.multivPenal, plot.multivPenal

Examples

```r
## Not run:
###--- Multivariate Frailty model ---###
data(dataMultiv)
# (computation takes around 60 minutes)
modMultiv.spli <- multivPenal(Surv(TIMEGAP,INDICREC)-cluster(PATIENT)+v1+v2+
     event2(INDICMETA)+terminal(INDICDEATH),formula.Event2=~v1+v2+v3,
     formula.terminalEvent=~v1,data=dataMultiv,n.knots=c(8,8,8),
     kappa=c(1,1,1),initialize=FALSE)
print(modMultiv.spli)

modMultiv.weib <- multivPenal(Surv(TIMEGAP,INDICREC)-cluster(PATIENT)+v1+v2+
     event2(INDICMETA)+terminal(INDICDEATH),formula.Event2=~v1+v2+v3,
     formula.terminalEvent=~v1,data=dataMultiv,hazard="Weibull")
print(modMultiv.weib)

modMultiv.cpm <- multivPenal(Surv(TIMEGAP,INDICREC)-cluster(PATIENT)+v1+v2+
     event2(INDICMETA)+terminal(INDICDEATH),formula.Event2=~v1+v2+v3,
     formula.terminalEvent=~v1,data=dataMultiv,hazard="Piecewise-per",
     nb.int=c(6,6,6))
print(modMultiv.cpm)

## End(Not run)
```

Description

This is a special function used in addition to the `cluster()` function in the context of survival joint models for clustered data. This function identifies subject index. It is used on the right hand side of a 'frailtyPenal' formula. Using `num.id()` in a formula implies that a joint frailty model for clustered data is fitted (Rondeau et al. 2011).

Usage

`num.id(x)`

Arguments

`x` A character or numeric variable which is supposed to indicate the variable identifying individuals

References


See Also

`frailtyPenal`

Examples

```r
## Not run:
data(readmission)
#-- here is generated cluster (5 clusters)
readmission <- transform(readmission,group=id%%5+1)

#-- exclusion all recurrent events ---#
#-- to obtain framework of semi-competing risks ---#
readmission2 <- subset(readmission, (t.start == 0 & event == 1) | event == 0)

joi.clus.gap <- frailtyPenal(Surv(time,event)~cluster(group)+
num.id(id)+dukes+charlson+sex+chemo+terminal(death),
formula.terminalEvent=-dukes+charlson+sex+chemo,
data=readmission2,recurrentAG=FALSE, n.knots=8,
kappa=c(1.e+10,1.e+10) ,Alpha="None")

## End(Not run)
```
**plot.additivePenal**  
*Plot Method for an Additive frailty model.*

**Description**
Plots estimated baseline survival and hazard functions (output from an object of class 'additivePenal' object for additive frailty model). Confidence bands are allowed.

**Usage**
```r
## S3 method for class 'additivePenal'
plot(x, type.plot="Hazard", conf.bands=TRUE,
     pos.legend="topright", cex.legend=0.7, main, color=2, median=TRUE, Xlab = "Time", Ylab = "Hazard function", ...)
```

**Arguments**
- `x`: An object of a fitted additive frailty model (output from calling `additivePenal`).
- `type.plot`: A character string specifying the type of curve. Possible values are "Hazard", or "Survival". The default is "Hazard". Only the first words are required, e.g "Haz", "Su"
- `conf.bands`: Logical value. Determines whether confidence bands will be plotted. The default is to do so.
- `pos.legend`: The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright".
- `cex.legend`: Character expansion factor *relative* to current `par("cex")`. Default is 0.7
- `main`: Plot title
- `color`: Curve color (integer)
- `median`: Logical value. Determines whether survival median will be plotted. Default is TRUE.
- `Xlab`: Label of x-axis. Default is "Time"
- `Ylab`: Label of y-axis. Default is "Hazard function"
- `...`: Other graphical parameters like those in `plot.frailtyPenal`

**Value**
Print a plot of the baseline survival or hazard functions with the confidence bands or not (conf.bands argument)

**See Also**
`additivePenal`
Examples

```r
## Not run:
data(dataAdditive)
modAdd <- additivePenal(Surv(t1,t2,event)~cluster(group)+var1+slope(var1),
correlation=TRUE,data=dataAdditive,n.knots=8,kappa=862,hazard="Splines")
#-- 'var1' is boolean as a treatment variable
plot(modAdd)

## End(Not run)
```

Description

Plots values of the difference of two Cross-Validated Prognosis Observed Loss (CVPOL) computed with two joint frailty models. Confidence intervals are allowed.

Usage

```r
## S3 method for class 'Diffepoce'
plot(x, conf.bands=TRUE, Xlab = "Time", Ylab = "EPOCE difference", ...)
```

Arguments

- `x` An object inheriting from `Diffepoce` class.
- `conf.bands` Logical value. Determines whether confidence intervals will be plotted. The default is `FALSE`.
- `Xlab` Label of x-axis. Default is "Time".
- `Ylab` Label of y-axis. Default is "EPOCE difference".
- `...` Other unused arguments.

Value

Print one plot with one curve and its confidence interval.

See Also

`Diffepoce`
Description

Plots values of estimators MPOL and CVPOL for evaluating EPOCE. No confidence interval.

Usage

```r
## S3 method for class 'epoce'
plot(x, type, pos.legend="topright", cex.legend=0.7,
     Xlab="Time", Ylab="Epoce", ...)
```

Arguments

- `x` An object inheriting from `epoce` class
- `type` Type of estimator to plot. If new dataset was used only mpol can be plotted ("mpol"), otherwise mpol and cvpol can be plotted ("mpol" and "cvpol", default is "cvpol").
- `pos.legend` The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright".
- `cex.legend` size of the legend. Default is 0.7.
- `Xlab` Label of x-axis. Default is "Time"
- `Ylab` Label of y-axis. Default is "Epoce"
- `...` Other unused arguments.

Value

Print a curve of the estimator of EPOCE using time points defined in `epoce`.

See Also

- `epoce`
Plot Method for a Shared frailty model.

Description

Plots estimated baseline survival and hazard functions from an object of class 'frailtyPenal'. Confidence bands are allowed.

Usage

```r
## S3 method for class 'frailtyPenal'
plot(x, type.plot = "Hazard", conf.bands=TRUE,
     pos.legend = "topleft", cex.legend=0.7, main, color=2, median=TRUE, Xlab = "Time", Ylab = "Hazard function", ...)
```

Arguments

- `x`: A shared frailty model, i.e. a `frailtyPenal` class object (output from calling `frailtyPenal` function).
- `type.plot`: a character string specifying the type of curve. Possible value are "Hazard", or "Survival". The default is "Hazard". Only the first letters are required, e.g "Haz", "Su"
- `conf.bands`: Logical value. Determines whether confidence bands will be plotted. The default is to do so.
- `pos.legend`: The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topleft"
- `cex.legend`: character expansion factor relative to current `par("cex")`. Default is 0.7
- `main`: title of plot
- `color`: color of the curve (integer)
- `median`: Logical value. Determines whether survival median will be plotted. Default is TRUE.
- `Xlab`: Label of x-axis. Default is "Time"
- `Ylab`: Label of y-axis. Default is "Hazard function"
- `...`: other unused arguments

Value

Print a plot of a shared frailty model.

See Also

`frailtyPenal`
## Examples

```r
## Not run:
data(readmission)

###--- Shared frailty model ---###
modSha <- frailtyPenal(Surv(time,event)~as.factor(dukes)+cluster(id),
n.knots=10,kappa=10000,data=readmission,hazard="Splines")
plot(modSha,type="surv",conf=FALSE)

###--- Cox proportional hazard model ---###
modCox <- frailtyPenal(Surv(time,event)~as.factor(dukes),n.knots=10,
kappa=10000,data=readmission,hazard="Splines")
plot(modCox)

#-- no confidence bands
plot(modSha,conf.bands=FALSE)
plot(modCox,conf.bands=FALSE)

## End(Not run)
```

---

### plot.jointNestedPenal

Plot method for a joint nested frailty model.

#### Description

Plots estimated baseline survival and hazard functions of a joint nested frailty model (output from an object of class 'jointNestedPenal' for joint nested frailty models) for each type of event (terminal or recurrent). Confidence bands are allowed.

#### Usage

```r
## S3 method for class 'jointNestedPenal'
plot(x, event = "Both", type.plot = "Hazard", conf.bands = FALSE, pos.legend="topright", cex.legend = 0.7, ylim, main, color = 2, median=TRUE, Xlab = "Time", Ylab = "Hazard function", ...)
```
Arguments

x A joint nested model, i.e. an object of class jointNestedPenal for joint nested frailty model (output from calling frailtyPenal function).

event a character string specifying the type of curve. Possible value are "Terminal", "Recurrent", or "Both". The default is "Both".

type.plot a character string specifying the type of curve. Possible value are "Hazard", or "Survival". The default is "Hazard". Only the first letters are required, e.g "Haz", "Su"

conf.bands logical value. Determines whether confidence bands will be plotted. The default is to do so.

pos.legend The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright"

cex.legend character expansion factor *relative* to current 'par("cex")'. Default is 0.7

ylim y-axis limits

main plot title

color curve color (integer)

median Logical value. Determines whether survival median will be plotted. Default is TRUE.

Xlab Label of x-axis. Default is "Time"

Ylab Label of y-axis. Default is "Hazard function"

... other unused arguments

Value

Print a plot of the baseline survival or hazard functions for each type of event or both with the confidence bands or not (conf.bands argument)

See Also

frailtyPenal

Examples

## Not run:

#-- here is generated cluster (30 clusters)
readmissionNested <- transform(readmission,group=id%%30+1)

# Baseline hazard function approximated with splines with calendar-timescale

model.spli.AG <- frailtyPenal(formula = Surv(t.start, t.stop, event) ~ subcluster(id) + cluster(group) + dukes + terminal(death),
formula.terminalEvent = ~dukes, data = readmissionNested, recurrentAG = TRUE,
n.knots = 8, kappa = c(9.55e+9, 1.41e+12), initialize = TRUE)

# Plot the estimated baseline hazard function with the confidence intervals
plot(model.spli.AG)

# Plot the estimated baseline hazard function with the confidence intervals
plot(model.spli.RE, type = "Survival")

## End(Not run)

---

**plot.jointPenal**

*Plot Method for a Joint frailty model.*

**Description**

Plots estimated baseline survival and hazard functions of a joint frailty model (output from an object of class 'JointPenal' for joint frailty models) for each type of event (terminal or recurrent). Confidence bands are allowed.

**Usage**

```r
## S3 method for class 'jointPenal'
plot(x, event = "Both", type.plot = "Hazard", conf.bands = FALSE, pos.legend = "topright", cex.legend = 0.7, ylim, main, color = 2, median = TRUE, Xlab = "Time", Ylab = "Hazard function", ...)
```

**Arguments**

- `x`: A joint model, i.e. an object of class `frailtyPenal` for Joint frailty model (output from calling `frailtyPenal` function).
- `event`: a character string specifying the type of curve. Possible value are "Terminal", "Recurrent", or "Both". The default is "Both".
- `type.plot`: a character string specifying the type of curve. Possible value are "Hazard" or "Survival". The default is "Hazard". Only the first letters are required, e.g "Haz", "Su"
- `conf.bands`: logical value. Determines whether confidence bands will be plotted. The default is to do so.
- `pos.legend`: The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright"
- `cex.legend`: character expansion factor *relative* to current 'par("cex")'. Default is 0.7
- `ylim`: y-axis limits
- `main`: plot title
plot.jointPenal

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>color</code></td>
<td>curve color (integer)</td>
</tr>
<tr>
<td><code>median</code></td>
<td>Logical value. Determines whether survival median will be plotted. Default is TRUE.</td>
</tr>
<tr>
<td><code>Xlab</code></td>
<td>Label of x-axis. Default is &quot;Time&quot;.</td>
</tr>
<tr>
<td><code>Ylab</code></td>
<td>Label of y-axis. Default is &quot;Hazard function&quot;.</td>
</tr>
<tr>
<td><code>...</code></td>
<td>other unused arguments</td>
</tr>
</tbody>
</table>

**Value**

Print a plot of the baseline survival or hazard functions for each type of event or both with the confidence bands or not (conf.bands argument)

**See Also**

`frailtyPenal`

**Examples**

```r
## Not run:

data(readmission)

#-- Gap-time
modJoint.gap <- frailtyPenal(Surv(time,event)~cluster(id)+sex+dukes+charlson+terminal(death),formula.terminalEvent=~sex+dukes+charlson,
data=readmission,n.knots=14,kappa=c(100,100))

#-- It takes around 1 minute to converge --#

plot(modJoint.gap,type.plot="Haz",event="recurrent",conf.bands=TRUE)
plot(modJoint.gap,type.plot="Haz",event="terminal",conf.bands=TRUE)
plot(modJoint.gap,type.plot="Haz",event="both",conf.bands=TRUE)

plot(modJoint.gap,type.plot="Su",event="recurrent",conf.bands=TRUE)
plot(modJoint.gap,type.plot="Su",event="terminal",conf.bands=TRUE)
plot(modJoint.gap,type.plot="Su",event="both",conf.bands=TRUE)

## End(Not run)
```
Description

Plots estimated baseline survival and hazard functions for the surrogate endpoint and the true endpoint from an object of class 'jointSurroPenal'. Confidence bands are allowed.

Usage

```r
## S3 method for class 'jointSurroPenal'
plot(x, type.plot = "Hazard", conf.bands=TRUE,
     pos.legend = "topright", cex.legend=0.7, main, Xlab = "Time",
     Ylab = "Baseline hazard function", median = TRUE, x min = 0, x max = NULL,
     ylim = c(0,1), endpoint = 2, scale = 1, ...) 
```

Arguments

- `x` An object inheriting from jointSurroPenal class (output from calling the function jointSurroPenal).
- `type.plot` A character string specifying the type of curve. Possible value are "Hazard", or "Survival". The default is "Hazard". Only the first letters are required, e.g "Haz","Su".
- `conf.bands` Logical value. Determines whether confidence bands will be plotted. The default is to do so.
- `pos.legend` The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright".
- `cex.legend` Character expansion factor *relative* to current `par("cex")`. Default is 0.7.
- `main` Title of plot.
- `Xlab` Label of x-axis. Default is "Time".
- `Ylab` Label of y-axis. Default is "Baseline hazard function".
- `median` Logical value. Determines whether survival median will be plotted. Default is TRUE.
- `xmin` Minimum value for x-axis, the default is 0.
- `xmax` Maximum value for x-axis, the default is NULL.
- `ylim` Range of y-axis. Default is from 0 to 1.
- `endpoint` A binary that indicates the endpoint to represent. 0 for the surrogate endpoint, 1 for the true endpoint, and 2 for both surrogate endpoint and true endpoint. The default is 2.
- `scale` A numeric that allows to rescale the survival times. If no change is needed the argument is set to 1, the default value. eg: 1/365 aims to convert days to years.
- `...` other unused arguments.
plot.longiPenal

Value

Print a plot of the baseline survival or hazard functions for each type of event or both with the confidence bands or not (conf.bands argument)

See Also

jointSurroPenal

Examples

## Not run:

###--- Joint surrogate model ---###
###---evaluation of surrogate endpoints---###

data(dataOvarian)
joint.surro.ovar <- jointSurroPenal(data = dataOvarian, n.knots = 8,
  init.kappa = c(2000,1000), indicator.alpha = 0,
  nb.mc = 200, scale = 1/365)

# Baseline Hazards functions for both the surrogate endpoint
# and the true endpoint
plot(joint.surro.ovar,endpoint = 2,type.plot = "Haz", conf.bands = T)

# Baseline survival functions for both the surrogate endpoint
# and the true endpoint
plot(joint.surro.ovar,endpoint = 2,type.plot = "Su", conf.bands = T)

## End(Not run)

Plot Method for a joint model for longitudinal data and a terminal event.

Description

Plots estimated baseline survival and hazard functions for a terminal outcome from an object of class ‘longiPenal’. Confidence bands are allowed.
plot.longiPenal

Usage

## S3 method for class 'longiPenal'
plot(x, type.plot = "Hazard", conf.bands=TRUE,
     pos.legend= "topright", cex.legend=0.7, main, color, median=TRUE, Xlab = "Time", Ylab = "Hazard function", ...)

Arguments

x A joint model for longitudinal outcome and a terminal event, i.e. a longiPenal class object (output from calling longiPenal function).

type.plot a character string specifying the type of curve for the terminal event. Possible value are "Hazard", or "Survival". The default is "Hazard". Only the first words are required, e.g "Haz", "Su"

conf.bands Logical value. Determines whether confidence bands will be plotted. The default is to do so.

pos.legend The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright"

cex.legend character expansion factor *relative* to current `par("cex")`. Default is 0.7

main title of plot

color color of the curve (integer)

median Logical value. Determines whether survival median will be plotted. Default is TRUE.

Xlab Label of x-axis. Default is "Time"

Ylab Label of y-axis. Default is "Hazard function"

... other unused arguments

Value

Print a plot for the terminal event of the joint model for a longitudinal and survival data.

See Also

longiPenal

Examples

## Not run:
###--- Joint model for longitudinal data and a terminal event ---###
data(colorectal)
data(colorectallongi)

# Survival data preparation - only terminal events
colorectalSurv <- subset(colorectal, new.lesions == 0)

# Baseline hazard function approximated with splines
# Random effects as the link function

model.spli.RE <- longiPenal(Surv(time1, state) ~ age + treatment + who.PS
+ prev.resection, tumor.size ~ year * treatment + age + who.PS ,
colorectalSurv, data.Longi = colorectalLongi, random = c("1", "year"),
id = "id", link = "Random-effects", left.censoring = -3.33,
n.knots = 7, kappa = 2)
pdf(file = "/home/agareb1/etudiants/al10/newpack/test/plot_longi.pdf")

# Plot the estimated baseline hazard function with the confidence intervals
plot(model.spli.RE)

# Plot the estimated baseline hazard function with the confidence intervals
plot(model.spli.RE, type = "Survival")

## End(Not run)

---

**plot.multivPenal**

Plot Method for a multivariate frailty model.

### Description

Plots of estimated baseline survival and hazard functions of a multivariate frailty model (output from an object of class 'multivPenal' for multivariate frailty models) for each type of event (recurrent, terminal and second recurrent). Confidence intervals are allowed.

### Usage

```r
## S3 method for class 'multivPenal'
plot(x, event = "Both", type.plot = "Hazard", conf.bands = FALSE, pos.legend = "topright", cex.legend = 0.7, ylim, main, color1="red", color2="blue", colorEnd="green", median=TRUE, Xlab = "Time", Ylab = "Hazard function", ...)
```

### Arguments

- **x**: A joint multivariate model, i.e. an object of class multivPenal (output from calling multivPenal function).
- **event**: a character string specifying the type of outcome. Possible values are "Terminal", "Recurrent", "Recurrent2", or "Both". The default is "Both".
- **type.plot**: a character string specifying the type of curve. Possible values are "Hazard", or "Survival". The default is "Hazard". Only the first words are required, e.g. "Haz", "Su"
plot.nestedPenal

conf.bands logical value. Determines whether confidence intervals will be plotted. The
default is to do so.

pos.legend The location of the legend can be specified by setting this argument to a sin-
gle keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright".

cex.legend character expansion factor *relative* to current `par("cex")`. Default is 0.7

ylim y-axis limits

main plot title

color1 curve color for recurrent event of type 1 (integer or color name in quotation
marks)

color2 curve color for recurrent event of type 2 (integer or color name in quotation
marks)

colorEnd curve color for terminal event (integer or color name in quotation marks)

median Logical value. Determines whether survival median will be plotted. Default is
TRUE.

Xlab Label of x-axis. Default is "Time"

Ylab Label of y-axis. Default is "Hazard function"

... Other graphical parameters

Value

Print a plot of the baseline survival or hazard functions for each type of event or both with the confidence intervals or not (conf.bands argument)

See Also

multivPenal

plot.nestedPenal  Plot Method for a Nested frailty model.

Description

Plots estimated baseline survival and hazard functions (output from an object of class 'NestedPenal' for nested frailty models). Confidence bands are allowed.

Usage

```r
## S3 method for class 'nestedPenal'
plot(x, type.plot="Hazard", conf.bands=TRUE,
pos.legend="topright", cex.legend=0.7, main, color=2, median=TRUE, Xlab = "Time", Ylab = "Hazard function", ...)
```
Arguments

- **x**: A nested model, i.e. an object of class `frailtyPenal` for Nested frailty models (output from calling `frailtyPenal` function).
- **type.plot**: a character string specifying the type of curve. Possible value are "Hazard", or "Survival". The default is "Hazard". Only the first words are required, e.g "Haz", "Su"
- **conf.bands**: logical value. Determines whether confidence bands will be plotted. The default is to do so.
- **pos.legend**: The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright"
- **cex.legend**: character expansion factor relative to current `par("cex")`. Default is 0.7
- **main**: plot title
- **color**: curve color (integer)
- **median**: Logical value. Determines whether survival median will be plotted. Default is TRUE.
- **Xlab**: Label of x-axis. Default is "Time"
- **Ylab**: Label of y-axis. Default is "Hazard function"
- **...**: Other graphical parameters like those in `plot.frailtyPenal`

Value

Print a plot of the baseline survival or hazard functions with the confidence bands or not (conf.bands argument)

See Also

`frailtyPenal`

Examples

```R
## Not run:

data(dataNested)
modNested <- frailtyPenal(Surv(t1,t2,event)~cluster(group)+
  subcluster(subgroup)+cov1+cov2,data=dataNested,n.knots=8,
  kappa=50000,hazard="Splines")

plot(modNested,conf.bands=FALSE)

## End(Not run)
```
plot.predFrailty  
Plot predictions using a Cox or a shared frailty model.

Description

Plots predicted probabilities of event. Confidence intervals are allowed.

Usage

## S3 method for class 'predFrailty'
plot(x, conf.bands=FALSE, pos.legend="topright",
    cex.legend=0.7, ylim=c(0,1), Xlab = "Time t", Ylab, ...)

Arguments

x  
An object from the 'prediction' function, i.e. a predFrailty class object.
conf.bands  
Logical value. Determines whether confidence intervals will be plotted. The default is FALSE.
pos.legend  
The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright".
cex.legend  
size of the legend. Default is 0.7.
ylim  
range of y-axis. Default is from 0 to 1.
Xlab  
Label of x-axis. Default is "Time t".
Ylab  
Label of y-axis.
...  
Other unused arguments.

Value

Print one plot with as many curves as the number of profiles.

plot.predJoint  
Plot predictions using a joint frailty model.

Description

Plots predicted probabilities of terminal event. Confidence intervals are allowed.

Usage

## S3 method for class 'predJoint'
plot(x, conf.bands=FALSE,
    relapses=TRUE, pos.legend="topright", cex.legend=0.7, ylim=c(0,1), Xlab = "Time t", Ylab = "Prediction probability of event", ...)

Arguments

x  
An object from the 'prediction' function, i.e. a predJoint class object.
conf.bands  
Logical value. Determines whether confidence intervals will be plotted. The default is FALSE.
relapses  
Logical value. When TRUE will also plot relapsing predictions. Default is TRUE.
pos.legend  
The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright".
cex.legend  
size of the legend. Default is 0.7.
ylim  
range of y-axis. Default is from 0 to 1.
Xlab  
Label of x-axis. Default is "Time t".
Ylab  
Label of y-axis.
...  
Other unused arguments.

Value

Print one plot with as many curves as the number of profiles.
Arguments

x
An object from the 'prediction' function, more generally a predJoint class object.

conf.bands
Logical value. Determines whether confidence intervals will be plotted. The default is FALSE.

relapses
Logical value. Determines whether observed recurrent events will be plotted. The default is TRUE.

pos.legend
The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright".

cex.legend
Size of the legend. Default is 0.7.

ylim
Range of y-axis. Default is from 0 to 1.

Xlab
Label of x-axis. Default is "Time t".

Ylab
Label of y-axis. Default is "Prediction probability of event".

... Other unused arguments

Value

Print as many plots as the number of subjects.

plot.predLongi

Plot predictions using a joint model for longitudinal data and a terminal event or a trivariate joint model for longitudinal data, recurrent events and a terminal event.

Description

Plots predicted probabilities of the event. Confidence intervals are allowed.

Usage

```r
## S3 method for class 'predLongi'
plot(x, conf.bands=FALSE, pos.legend="topright",
cex.legend=0.7, ylim=c(0,1), Xlab = "Time t", Ylab, ...)
```

Arguments

x
An object inheriting from predLongi.

conf.bands
Logical value. Determines whether confidence intervals will be plotted. The default is FALSE.

pos.legend
The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright".

cex.legend
Size of the legend. Default is 0.7.
Plotting estimated baseline survival and hazard functions of a joint model (output from an object of class `trivPenal`) for each type of event (terminal or recurrent). Confidence bands are allowed.

### Usage

```r
## S3 method for class 'trivPenal'
plot(x, event = "Both", type.plot = "Hazard", conf.bands =
    FALSE, pos.legend="topright", cex.legend = 0.7, ylim, main, color = 2, median=TRUE, Xlab
    = "Time", Ylab = "Hazard function", ...)```

### Arguments

- **x**: A joint model, an object of class `trivPenal`.
- **event**: A character string specifying the type of curve. Possible value are "Terminal", "Recurrent", or "Both". The default is "Both".
- **type.plot**: A character string specifying the type of curve. Possible value are "Hazard", or "Survival". The default is "Hazard". Only the first words are required, e.g "Haz", "Su"
- **conf.bands**: Logical value. Determines whether confidence bands will be plotted. The default is to do so.
- **pos.legend**: The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright"
- **cex.legend**: Character expansion factor *relative* to current `par("cex")`. Default is 0.7
- **ylim**: y-axis limits
- **main**: Plot title
- **color**: Curve color (integer)
- **median**: Logical value. Determines whether survival median will be plotted. Default is TRUE.
Xlab

Label of x-axis. Default is "Time"

Ylab

Label of y-axis. Default is "Hazard function"

... other unused arguments

Value

Print a plot of the baseline survival or hazard functions for each type of event or both with the confidence bands or not (conf.bands argument)

See Also

trivPenal

Examples

```r
## Not run:
###--- Trivariate joint model for longitudinal data, ---###
###--- recurrent events and a terminal event ---###

data(colorectal)
data(colorectalLongi)

# Weibull baseline hazard function
# Random effects as the link function, Gap timescale
# (computation takes around 30 minutes)
model.weib.RE.gap <- trivPenal(Surv(gap.time, new.lesions) ~ cluster(id) + age + treatment + who.PS + prev.resection + terminal(state), formula.terminalEvent =~ age + treatment + who.PS + prev.resection, tumor.size ~ year * treatment + age + who.PS, data = colorectal, data.Longi = colorectalLongi, random = c("1", "year"), id = "id", link = "Random-effects", left.censoring = -3.33, recurrentAG = FALSE, hazard = "Weibull", method.GH="Pseudo-adaptive", n.nodes = 7)

plot(model.weib.RE.gap)
plot(model.weib.RE.gap, type = "survival")
```

## End(Not run)

plot.trivPenalNL

Plot Method for a Non-Linear Trivariate Joint Model for Recurrent Events and a Terminal Event with a Biomarker Described with an ODE.

Description

Plots estimated baseline survival and hazard functions of a joint model (output from an object of class `trivPenalNL`) for each type of event (terminal or recurrent). Confidence bands are allowed.
Usage

## S3 method for class 'trivPenalNL'
plot(x, event = "Both", type.plot = "Hazard", conf.bands = FALSE, pos.legend = "topright", cex.legend = 0.7, ylim, main, color = 2, median = TRUE, Xlab = "Time", Ylab = "Hazard function", ...)

Arguments

- **x**: A joint model, an object of class `trivPenalNL`.
- **event**: a character string specifying the type of curve. Possible values are "terminal", "recurrrent", or "both". The default is "both".
- **type.plot**: a character string specifying the type of curve. Possible values are "Hazard", or "survival". The default is "hazard". Only the first words are required, e.g "haz", "su"
- **conf.bands**: logical value. Determines whether confidence bands will be plotted. The default is to do so.
- **pos.legend**: The location of the legend can be specified by setting this argument to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The default is "topright".
- **cex.legend**: character expansion factor *relative* to current `par("cex")`. Default is 0.7
- **ylim**: y-axis limits
- **main**: plot title
- **color**: curve color (integer)
- **median**: Logical value. Determines whether survival median will be plotted. Default is TRUE.
- **Xlab**: Label of x-axis. Default is "Time"
- **Ylab**: Label of y-axis. Default is "Hazard function"
- **...**: other unused arguments

Value

Print a plot of the baseline survival or hazard functions for each type of event or both with the confidence bands or not (conf.bands argument)

See Also

- `trivPenalNL`

Examples

```r
## Not run:
###--- Trivariate joint model for longitudinal data, ---###
###---- recurrent events and a terminal event ----#

data(colorectal)
```
data(colorectalLongi)

# Weibull baseline hazard function
# Random effects as the link function, Gap timescale
# (computation takes around 30 minutes)
model.weib.RE.gap <- trivPenal(Surv(gap.time, new.lesions) ~ cluster(id) + age + treatment + who.PS + prev.resection + terminal(state),
  formula.terminalEvent =~ age + treatment + who.PS + prev.resection,
  tumor.size ~ year * treatment + age + who.PS, data = colorectal,
  data.Longi = colorectalLongi, random = c("1", "year"), id = "id",
  link = "Random-effects", left.censoring = -3.33, recurrentAG = FALSE,
  hazard = "Weibull", method.GH="Pseudo-adaptive", n.nodes = 7)

plot(model.weib.RE.gap)
plot(model.weib.RE.gap, type = "survival")

## End(Not run)

---

**predict.jointSurroPenal**

*Predict Method for the one-step Joint surrogate model for the evaluation of a candidate surrogate endpoint.*

---

### Description

Predict the treatment effect on the true endpoint \( (\beta_T) \), based on the treatment effect observed on the surrogate endpoint \( (\beta_S) \).

### Usage

```r
## S3 method for class 'jointSurroPenal'
predict(object, datapred = NULL, var.used = "error.estim",
        alpha. = 0.05, dec = 3, ...)
```

### Arguments

- **object**: An object inheriting from `jointSurroPenal` class (output from calling the function `jointSurroPenal`).
- **datapred**: Dataset to use for the prediction. If this argument is specified, the data structure must be the same as the parameter `data` in the function `jointSurroPenal`. However, if observation on te true endpoint are not available, columns `timeT` and `statusT` can be absent.
- **var.used**: This argument can take two values. The first one is "error.estim" and indicates if the prediction error take into account the estimation error of the estimates of the parameters. If the estimates are supposed to be known or if the dataset includes a high number of trials with a high number of subject per trial, value `No.error` can be used. The default is `error.estim`.
- **alpha.**: The confidence level for the prediction interval. The default is 0.05.
predict.jointSurroPenal

dec

The desired number of digits after the decimal point for parameters and confidence intervals. Default of 3 digits is used.

...

other unused arguments.

Details

Prediction is based on the formulas described in (Burzikwosky et al., 2006). We do not consider the case of prediction which suppose estimation error on the estimate of the treatment effect on the surrogate endpoint in the new trial.

Value

Return and display a dataframe including for each trial the number of included subjects, the observed treatment effect on the surrogate endpoint, the observed treatment effect on the true endpoint (if available) and the predicted treatment effect on the true endpoint with the associated prediction intervals. If the observed treatment effect on the true endpoint (if available) is included into the prediction interval, the last columns contains "*".

Author(s)

Casimir Ledoux Sofeu <casimir.sofeu@u-bordeaux.fr>, <scl.ledoux@gmail.com> and Virginie Rondeau <virginie.rondeau@inserm.fr>

References


See Also

jointSurroPenal

Examples

## Not run:

#### Joint surrogate model ####
#### evaluation of surrogate endpoints ####

data(dataOvarian)
joint.surro.ovar <- jointSurroPenal(data = dataOvarian, n.knots = 8, init.kappa = c(2000,1000), indicator.alpha = 0, nb.mc = 200, scale = 1/365)

# prediction of the treatment effects on the true endpoint in each trial of the dataOvarian dataset
# predict(joint.surro.ovar)
Description

For Cox proportional hazard model

A predictive probability of event between $t$ and horizon time $t+w$, with $w$ the window of prediction.

\[
P(t, t + w) = \frac{S_i(t) - S_i(t + w)}{S_i(t)} = 1 - \left( \frac{S_0(t + w)}{S_0(t)} \right)^{\exp(\beta' Z_i)}
\]

For Gamma Shared Frailty model for clustered (not recurrent) events

Two kinds of predictive probabilities can be calculated:

- a conditional predictive probability of event between $t$ and horizon time $t+w$, i.e. given a specific group

\[
P_{\text{cond}}(t, t + w) = \frac{S_{ij}(t|u_i) - S_{ij}(t + w|u_i)}{S_{ij}(t|u_i)} = 1 - \left( \frac{S_0(t + w)}{S_0(t)} \right)^{u_i \exp(\beta' Z_{ij})}
\]

- a marginal predictive probability of event between $t$ and horizon time $t+w$, i.e. averaged over the population

\[
P_{\text{marg}}(t, t + w) = 1 - \left( \frac{1 + \theta H_0(t) \exp(\beta' Z_{ij})}{1 + \theta H_0(t + w) \exp(\beta' Z_{ij})} \right)^{1/\theta}
\]

For Gaussian Shared Frailty model for clustered (not recurrent) events

Two kinds of predictive probabilities can be calculated:

- a conditional predictive probability of event between $t$ and horizon time $t+w$, i.e. given a specific group and given a specific Gaussian random effect $\eta$

\[
P_{\text{cond}}(t, t + w) = \frac{S_{ij}(t|\eta_i) - S_{ij}(t + w|\eta_i)}{S_{ij}(t|\eta_i)} = 1 - \left( \frac{S_0(t + w)}{S_0(t)} \right)^{\exp(\eta_i + \beta' Z_{ij})}
\]

- a marginal predictive probability of event between $t$ and horizon time $t+w$, i.e. averaged over the population
\[
P_{\text{marg}}(t, t+w) = \frac{\int_{-\infty}^{+\infty} (S_{ij}(t|\eta_i) - S_{ij}(t+w|\eta_i))g(\eta)d\eta}{\int_{-\infty}^{+\infty} S_{ij}(t)g(\eta)d\eta}
\]

For Gamma Shared Frailty model for recurrent events

Two kinds of predictive probabilities can be calculated:

- A marginal predictive probability of event between \( t \) and horizon time \( t+w \), i.e. averaged over the population.

\[
P_{\text{marg}}(t, t+w) = \frac{\int_{0}^{+\infty} (S_{ij}(t|u_i) - S_{ij}(t+w|u_i)) \cdot (u_i)^{J} S_{ij}(X_{i,J}|u_i)g(u)du}{\int_{0}^{+\infty} S_{ij}(t|u_i) (u_i)^{J} S_{ij}(X_{i,J}|u_i)g(u)du}
\]

- a conditional predictive probability of event between \( t \) and horizon time \( t+w \), i.e. given a specific individual.

This prediction method is the same as the conditional gamma prediction method applied for clustered events (see formula \( P_{\text{cond}} \) before).

For Gaussian Shared Frailty model for recurrent events

Two kinds of predictive probabilities can be calculated:

- A marginal predictive probability of event between \( t \) and horizon time \( t+w \), i.e. averaged over the population.

\[
P_{\text{marg}}(t, t+w) = \frac{\int_{0}^{+\infty} (S_{i(J+1)}(t|\eta_i) - S_{i(J+1)}(t+w|\eta_i)) \cdot \exp(J\eta_i) S_{i(J+1)}(X_{i,J}|\eta_i)g(\eta)d\eta}{\int_{0}^{+\infty} S_{i(J+1)}(t|\eta_i) \exp(J\eta_i) S_{i(J+1)}(X_{i,J}|\eta_i)g(\eta)d\eta}
\]

- a conditional predictive probability of event between \( t \) and horizon time \( t+w \), i.e. given a specific individual.

This prediction method is the same as the conditional Gaussian prediction method applied for clustered events (see formula \( P_{\text{cond}} \) before).

It is possible to compute all these predictions in two ways on a scale of times: - either you want a cumulative probability of developing the event between \( t \) and \( t+w \) (with \( t \) fixed, but with a varying window of prediction \( w \)); - either you want at a specific time the probability to develop the event in the next \( w \) (ie, for a varying prediction time \( t \), but for a fixed window of prediction). See Details.

For Joint Frailty model

Prediction for two types of event can be calculated: for a terminal event or for a new recurrent event, knowing patient’s characteristics.

- Prediction of death knowing patients’ characteristics:

It is to predict the probability of death in a specific time window given the history of patient \( i \) before the time of prediction \( t \). The history \( H_{i}^{J,l}, (l = 1, 2) \) is the information on covariates before time \( t \),
but also the number of recurrences and the time of occurrences. Three types of marginal probabilities are computed:

- a prediction of death between \( t \) and \( t+w \) given that the patient had exactly \( J \) recurrences (\( H_{i,t}^{J,1} \)) before \( t \)

\[
P^1(t, t+w) = P(D_i \leq t + w | D_i > t, H_{i,t}^{J,1}) = \frac{\int_0^\infty [S_i(t) - S_i(t+w)](u_i)^J S_{i(j+1)}^R(t) g(u) du_i}{\int_0^\infty S_i(t)(u_i)^J S_{i(j+1)}^R(t) g(u) du_i}
\]

- a prediction of death between \( t \) and \( t+w \) given that the patient had at least \( J \) recurrences (\( H_{i,t}^{J,2} \)) before \( t \)

\[
P^2(t, t+w) = P(D_i \leq t + w | D_i > t, H_{i,t}^{J,2}) = \frac{\int_0^\infty [S_i(t) - S_i(t+w)](u_i)^J S_{i(j+1)}^R(X_{ij}) g(u) du_i}{\int_0^\infty S_i(t)(u_i)^J S_{i(j+1)}^R(X_{ij}) g(u) du_i}
\]

- a prediction of death between \( t \) and \( t+w \) considering the recurrence history only in the parameters estimation. It corresponds to the average probability of death between \( t \) and \( t+w \) for a patient with these given characteristics.

\[
P^3(t, t+w) = P(D_i \leq t + w | D_i > t) = \frac{\int_0^\infty [S_i(t) - S_i(t+w)] g(u) du_i}{\int_0^\infty S_i(t) g(u) du_i}
\]

**Prediction of risk of a new recurrent event knowing patients’ characteristics**:

It is to predict the probability of a new recurrent event in a specific time window given the history of patient \( i \) before the time of prediction \( t \). The history \( H_{i,t}^{J} \) is the information on covariates before time \( t \), but also the number of recurrences and the time of occurrences. The marginal probability computed is a prediction of a new recurrent event between \( t \) and \( t+w \) given that the patient had exactly \( J \) recurrences (\( H_{i,t}^{J} \)) before \( t \):

\[
P^R(t, t + w) = P(X_{i(j+1)} \leq t + w | X_{i(j+1)} > t, D_i > t, H_{i,t}^{J}) =
\]

\[
\frac{\int_0^\infty [S_{i(j+1)}^R(t) - S_{i(j+1)}^R(t+w)](u_i)^J S_{i(j+1)}^R(t) S_{i(j+1)}^R(X_{ij}) g(u) du_i}{\int_0^\infty S_{i(j+1)}^R(t) S_{i(j+1)}^R(X_{ij}) g(u) du_i}
\]

It is possible to compute all these predictions in two ways: - either you want a cumulative probability of developing the event between \( t \) and \( t+w \) (with \( t \) fixed, but with a varying window of prediction \( w \)); - either you want at a specific time the probability to develop the event in the next \( w \) (ie, for a varying prediction time \( t \), but for a fixed window of prediction). See Details.

With Gaussian frailties (\( \eta \)), the same expressions are used but with \( u_i^J \) replaced by \( \exp(J \eta_i) \) and \( g(\eta) \) corresponds to the Gaussian distribution.

**For Joint Nested Frailty models**

Prediction of the probability of developing a terminal event between \( t \) and \( t+w \) for subject \( i \) who survived by time \( t \) based on the visiting and disease histories of their own and other family members observed by time \( t \).
Let \( Y^R_{fi}(t) \) be the history of subject \( i \) in family \( f \), before time \( t \), which includes all the recurrent events and covariate information. For disease history, let \( T^D_{fi}(t) = \min(T_{fi}, t) \) be the observed time to an event before \( t \); \( \delta^D_{fi}(t) \) the disease indicator by time \( t \) and \( X^D_{fi}(t) \) the covariate information observed up to time \( t \). We define the family history of subject \( i \) in family \( f \) by

\[
H_{f(-i)}(t) = \{Y^R_{fi}(t), T^D_{fi}(t), \delta^D_{fi}(t), X^D_{fi}(t), \forall l \in \{1, \ldots, i-1, i+1, \ldots, m_f\}\}
\]

which includes the visiting and disease history of all subjects except for subject \( i \) in family \( f \) as well as their covariate information by time \( t \).

The prediction probability can be written as :

\[
P(T^D_{fi} > t + s | T^D_{fi} > t, Y_{fi}(t), H_{(f-i)}(t)) = \\
\frac{\int \int P(t < T^D_{fi} < t + s | X^D_{fi}, \omega_{fi})P(Y_{fi}(t)|X^D_{fi}(t), \omega_i)P(H_{f(-i)}(t)|X^D_{fi}(t), \omega_{fi})g_{ui}g_{\omega f}}{\int \int P(T^D_{fi} > t | X^D_{fi}, \omega_{fi})P(Y_{fi}(t)|X^D_{fi}(t), \omega_i)P(H_{f(-i)}(t)|X^D_{fi}(t), \omega_{fi})g_{ui}g_{\omega f}}
\]

For Joint models for longitudinal data and a terminal event

The predicted probabilities are calculated in a specific time window given the history of biomarker measurements before the time of prediction \( Y_{fi}(t) \). The probabilities are conditional also on covariates before time \( t \) and that the subject was at risk at \( t \). The marginal predicted probability of the terminal event is

\[
P(t, t+w) = P(D_i \leq t+w | D_i > t, Y_{fi}(t)) = \frac{\int_0^\infty [S^D_i(t) - S^D_i(t+w)] f(Y_{fi}(t)|X_{Li}, b_i) f(b_i) db_i}{\int_0^\infty S^D_i(t) f(Y_{fi}(t)|X_{Li}, b_i) f(b_i) db_i}
\]

These probabilities can be calculated in several time points with fixed time of prediction \( t \) and varying window \( w \) or with fixed window \( w \) and varying time of prediction \( t \). See Details for an example of how to construct time windows.

For Trivariate joint models for longitudinal data, recurrent events and a terminal event

The predicted probabilities are calculated in a specific time window given the history of biomarker measurements \( Y_{fi}(t) \) and recurrences \( H^{J1}_i \) (complete history of recurrences with known \( J \) number of observed events) before the time of prediction \( t \). The probabilities are conditional also on covariates before time \( t \) and that the subject was at risk at \( t \). The marginal predicted probability of the terminal event is

\[
P(t, t+w) = P(D_i \leq t+w | D_i > t, H^{J1}_i, Y_{fi}(t)) = \\
\frac{\int_0^\infty [S^D_i(t) - S^D_i(t+w)] \exp(J(v_c+g(t)^\top \eta_{R})) S^R_i(t) f(Y_{fi}(t)|X_{Li}, b_i) f(u_i) du_i}{\int_0^\infty S^D_i(t) \exp(J(v_c+g(t)^\top \eta_{R})) S^R_i(t) f(Y_{fi}(t)|X_{Li}, b_i) f(u_i) du_i}
\]

The biomarker history can be represented using a linear (trivPenal) or non-linear mixed-effects model (trivPenalNL).

These probabilities can be calculated in several time points with fixed time of prediction \( t \) and varying window \( w \) or with fixed window \( w \) and varying time of prediction \( t \). See Details for an example of how to construct time windows.
Usage

prediction(fit, data, data.Longi, t, window, event="Both", conditional = FALSE, MC.sample=0, individual)

Arguments

fit A frailtyPenal, jointPenal, longiPenal, trivPenal or trivPenalNL object.
data Data frame for the prediction. See Details.
data.Longi Data frame for the prediction used for joint models with longitudinal data. See Details.
t Time or vector of times for prediction.
window Window or vector of windows for prediction.
event Only for joint and shared models. The type of event you want to predict: "Terminal" for a terminal event, "Recurrent" for a recurrent event or "Both". Default value is "Both". For joint nested model, only 'Terminal' is allowed. In a shared model, if you want to predict a new recurrent event then the argument "Recurrent" should be use. If you want to predict a new event from clustered data, do not use this option.
conditional Only for prediction method applied on shared models. Provides distinction between the conditional and marginal prediction methods. Default is FALSE.
MC.sample Number of samples used to calculate confidence bands with a Monte-Carlo method (with a maximum of 1000 samples). If MC.sample=0 (default value), no confidence intervals are calculated.
individual Only for joint nested model. Vector of individuals (of the same family) you want to make prediction.

Details

To compute predictions with a prediction time t fixed and a variable window:

prediction(fit, datapred, t=10, window=seq(1,10,by=1))

Otherwise, you can have a variable prediction time and a fixed window.

prediction(fit, datapred, t=seq(10,20,by=1), window=5)

Or fix both prediction time t and window.

prediction(fit, datapred, t=10, window=5)

The data frame building is an important step. It will contain profiles of patient on which you want to do predictions. To make predictions on a Cox proportional hazard or a shared frailty model, only covariates need to be included. You have to distinguish between numerical and categorical variables (factors). If we fit a shared frailty model with two covariates sex (factor) and age (numeric), here is the associated data frame for three profiles of prediction.
prediction

```r
datapred <- data.frame(sex=0, age=0)
datapred$sex <- as.factor(datapred$sex)
levels(datapred$sex) <- c(1, 2)
datapred[1,] <- c(1,40) # man, 40 years old
datapred[2,] <- c(2,45) # woman, 45 years old
datapred[3,] <- c(1,60) # man, 60 years old
```

**Time-dependent covariates:** In the context of time-dependent covariate, the last previous value of the covariate is used before the time \( t \) of prediction.

It should be noted, that in a data frame for both marginal and conditional prediction on a shared frailty model for clustered data, the group must be specified. In the case of marginal predictions this can be any number as it does not influence predictions. However, for conditional predictions, the group must be also included in the data set used for the model fitting. The conditional predictions apply the empirical Bayes estimate of the frailty from the specified cluster. Here, three individuals belong to group 5.

```r
datapred <- data.frame(group=0, sex=0, age=0)
datapred$sex <- as.factor(datapred$sex)
levels(datapred$sex) <- c(1, 2)
datapred[1,] <- c(5,1,40) # man, 40 years old (cluster 5)
datapred[2,] <- c(5,2,45) # woman, 45 years old (cluster 5)
datapred[3,] <- c(5,1,60) # man, 60 years old (cluster 5)
```

To use the prediction function on joint frailty models and trivariate joint models, the construction will be a little bit different. In these cases, the prediction for the terminal event takes into account covariates but also history of recurrent event times for a patient. You have to create a data frame with the relapse times, the indicator of event, the cluster variable and the covariates. Relapses occurring after the prediction time may be included but will be ignored for the prediction. A joint model with calendar-timescale need to be fitted with `Surv(start, stop, event)`, relapse times correspond to the "stop" variable and indicators of event correspond to the "event" variable (if event=0, the relapse will not be taken into account). For patients without relapses, all the values of "event" variable should be set to 0. Finally, the same cluster variable name needs to be in the joint model and in the data frame for predictions ("id" in the following example). For instance, we observe relapses of a disease and fit a joint model adjusted for two covariates sex (1:male 2:female) and chemo (treatment by chemotherapy 1:no 2:yes). We describe 3 different profiles of prediction all treated by chemotherapy: 1) a man with four relapses at 100, 200, 300 and 400 days, 2) a man with only one relapse at 1000 days, 3) a woman without relapse.

```r
datapred <- data.frame(time=0, event=0, id=0, sex=0, chemo=0)
datapred$sex <- as.factor(datapred$sex)
levels(datapred$sex) <- c(1, 2)
datapred$chemo <- as.factor(datapred$chemo)
levels(datapred$chemo) <- c(1, 2)
datapred[1,] <- c(100, 1, 1, 2) # first relapse of the patient 1
datapred[2,] <- c(200, 1, 1, 2) # second relapse of the patient 1
datapred[3,] <- c(300, 1, 1, 2) # third relapse of the patient 1
datapred[4,] <- c(400, 1, 1, 2) # fourth relapse of the patient 1
datapred[5,] <- c(1000, 1, 2, 2) # one relapse at 1000 days for patient 2
datapred[6,] <- c(100, 0, 3, 2, 2) # patient 3 did not relapse
```

The data can also be the dataset used to fit the joint model. In this case, you will obtain as many prediction rows as patients.
Finally, for the predictions using joint models for longitudinal data and a terminal event and trivariate joint models, a data frame with the history of the biomarker measurements must be provided. It must include data on measurements (values and time points), cluster variable and covariates. Measurements taken after the prediction time may be included but will be ignored for the prediction. The same cluster variable name must be in the data frame, in the data frame used for the joint model and in the data frame with the recurrent event and terminal event times. For instance, we observe two patients and each one had 5 tumor size measurements (patient 1 had an increasing tumor size and patient 2, decreasing). The joint model used for the predictions was adjusted on sex (1: male, 2: female), treatment (1: sequential arm, 2: combined arm), WHO baseline performance status (1: 0 status, 2: 1 status, 3: 2 status) and previous resection of the primate tumor (0: no, 1: yes). The data frame for the biomarker measurements can be:

```r
datapredj_longi <- data.frame(id = 0, year = 0, tumor.size = 0, treatment = 0, age = 0, who.PS = 0, prev.resection = 0)
datapredj_longi$treatment <- as.factor(datapredj_longi$treatment)
levels(datapredj_longi$treatment) <- 1:2
datapredj_longi$age <- as.factor(datapredj_longi$age)
levels(datapredj_longi$age) <- 1:3
datapredj_longi$who.PS <- as.factor(datapredj_longi$who.PS)
levels(datapredj_longi$who.PS) <- 1:3
datapredj_longi$prev.resection <- as.factor(datapredj_longi$prev.resection)
levels(datapredj_longi$prev.resection) <- 1:2
# patient 1: increasing tumor size
datapredj_longi[1,] <- c(1, 0, 1.2, 2, 1, 1, 1) datapredj_longi[2,] <- c(1, 0.3, 1.4, 2, 1, 1, 1) datapredj_longi[3,] <- c(1, 0.6, 1.9, 2, 1, 1, 1) datapredj_longi[4,] <- c(1, 0.9, 2.5, 2, 1, 1, 1) datapredj_longi[5,] <- c(1, 1.5, 3.9, 2, 1, 1, 1)
# patient 2: decreasing tumor size
ndatapredj_longi[6,] <- c(2, 0, 1.2, 2, 1, 1, 1) datapredj_longi[7,] <- c(2, 0.3, 0.7, 2, 1, 1, 1) datapredj_longi[8,] <- c(2, 0.5, 0.3, 2, 1, 1, 1) datapredj_longi[9,] <- c(2, 0.7, 0.1, 2, 1, 1, 1) datapredj_longi[10,] <- c(2, 0.9, 0.1, 2, 1, 1, 1)
```

Value

The following components are included in a 'predFrailty' object obtained by using prediction function for Cox proportional hazard and shared frailty model.

- `nnpred` Number of individual predictions
- `x.time` A vector of prediction times of interest (used for plotting predictions): vector of prediction times t if fixed window. Otherwise vector of prediction times t+w
- `window` Prediction window or vector of prediction windows
- `pred` Predictions estimated for each profile
- `icproba` Logical value. Were confidence intervals estimated?
- `predLow` Lower limit of Monte-Carlo confidence interval for each prediction
- `predHigh` Upper limit of Monte-Carlo confidence interval for each prediction
- `type` Type of prediction probability (marginal or conditional)
- `group` For conditional probability, the list of group on which you make predictions
The following components are included in a ’predJoint’ object obtained by using prediction function for joint frailty model.

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>npred</td>
<td>Number of individual predictions</td>
</tr>
<tr>
<td>x.time</td>
<td>A vector of prediction times of interest (used for plotting predictions): vector of prediction times t if fixed window. Otherwise vector of prediction times t+w</td>
</tr>
<tr>
<td>window</td>
<td>Prediction window or vector of prediction windows</td>
</tr>
<tr>
<td>group</td>
<td>Id of each patient</td>
</tr>
<tr>
<td>pred1</td>
<td>Estimation of probability of type 1: exactly j recurrences</td>
</tr>
<tr>
<td>pred2</td>
<td>Estimation of probability of type 2: at least j recurrences</td>
</tr>
<tr>
<td>pred3</td>
<td>Estimation of probability of type 3</td>
</tr>
<tr>
<td>pred1_rec</td>
<td>Estimation of prediction of relapse</td>
</tr>
<tr>
<td>icproba</td>
<td>Logical value. Were confidence intervals estimated?</td>
</tr>
<tr>
<td>predLow1</td>
<td>Lower limit of Monte-Carlo confidence interval for probability of type 1</td>
</tr>
<tr>
<td>predHigh1</td>
<td>Upper limit of Monte-Carlo confidence interval for probability of type 1</td>
</tr>
<tr>
<td>predLow2</td>
<td>Lower limit of Monte-Carlo confidence interval for probability of type 2</td>
</tr>
<tr>
<td>predHigh2</td>
<td>Upper limit of Monte-Carlo confidence interval for probability of type 2</td>
</tr>
<tr>
<td>predLow3</td>
<td>Lower limit of Monte-Carlo confidence interval for probability of type 3</td>
</tr>
<tr>
<td>predHigh3</td>
<td>Upper limit of Monte-Carlo confidence interval for probability of type 3</td>
</tr>
<tr>
<td>predHigh1_rec</td>
<td>Upper limit of Monte-Carlo confidence interval for prediction of relapse</td>
</tr>
<tr>
<td>predLow1_rec</td>
<td>Lower limit of Monte-Carlo confidence interval for prediction of relapse</td>
</tr>
</tbody>
</table>

The following components are included in a ’predLongi’ object obtained by using prediction function for joint models with longitudinal data.

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>npred</td>
<td>Number of individual predictions</td>
</tr>
<tr>
<td>x.time</td>
<td>A vector of prediction times of interest (used for plotting predictions): vector of prediction times t if fixed window. Otherwise vector of prediction times t+w</td>
</tr>
<tr>
<td>window</td>
<td>Prediction window or vector of prediction windows</td>
</tr>
<tr>
<td>group</td>
<td>Id of each patient</td>
</tr>
<tr>
<td>pred</td>
<td>Estimation of probability</td>
</tr>
<tr>
<td>icproba</td>
<td>Logical value. Were confidence intervals estimated?</td>
</tr>
<tr>
<td>predLow</td>
<td>Lower limit of Monte-Carlo confidence intervals</td>
</tr>
<tr>
<td>predHigh</td>
<td>Upper limit of Monte-Carlo confidence intervals</td>
</tr>
<tr>
<td>trivariate</td>
<td>Logical value. Are the prediction calculated from the trivariate model?</td>
</tr>
</tbody>
</table>
## Not run:

### prediction on a COX or SHARED frailty model ###

```r
data(readmission)
#-- here is a generated cluster (31 clusters of 13 subjects)
readmission <- transform(readmission, group=id%%31+1)
#-- we compute predictions of death
#-- we extract last row of each subject for the time of death
readmission <- aggregate(readmission, by=list(readmission$id),
                         FUN=function(x){x[length(x)]})[, -1]

#-- predictions on a Cox proportional hazard model --##
cox <- frailtyPenal(Surv(t.stop, death) ~ sex + dukes, n.knots=10, kappa=10000, data=readmission)

#-- construction of the data frame for predictions
datapred <- data.frame(sex=0, dukes=0)
datapred$sex <- as.factor(datapred$sex)
levels(datapred$sex)<- c(1,2)
datapred$dukes <- as.factor(datapred$dukes)
levels(datapred$dukes)<- c(1,2,3)
datapred[1,] <- c(1,2) # man, dukes 2
datapred[2,] <- c(2,3) # woman, dukes 3

#-- prediction of death for two patients between 100 and 100+w,
#-- with w in (50, 100, ..., 1900)
pred.cox <- prediction(cox, datapred, t=100, window=seq(50, 1900, 50))
plot(pred.cox)

#-- prediction of death for two patients between t and t+400,
#-- with t in (100, 150, ..., 1500)
```
pred.cox2 <- prediction(cox, datapred, t=seq(100,1500,50), window=400)
plot(pred.cox2)

##-- predictions on a shared frailty model for clustered data --##
sha <- frailtyPenal(Surv(t.stop,death)~cluster(group)+sex+dukes,
n.knots=10,kappa=10000,data=readmission)

##-- marginal prediction
# a group must be specified but it does not influence the results
# in the marginal predictions setting
datapred$group[1:2] <- 1
pred.sha.marg <- prediction(sha, datapred, t=100, window=seq(50,1900,50))
plot(pred.sha.marg)

##-- conditional prediction, given a specific cluster (group=5)
datapred$group[1:2] <- 5
pred.sha.cond <- prediction(sha, datapred, t=100, window=seq(50,1900,50),
conditional = TRUE)
plot(pred.sha.cond)

##-- marginal prediction of a recurrent event, on a shared frailty model
data(readmission)
datapred <- data.frame(t.stop=0,event=0,id=0,sex=0,dukes=0)
datapred$sex <- as.factor(datapred$sex)
levels(datapred$sex)<- c(1,2)
datapred$dukes <- as.factor(datapred$dukes)
levels(datapred$dukes)<- c(1,2,3)
datapred[1,] <- c(100,1,1,1,2) #man, dukes 2, 3 recurrent events
datapred[2,] <- c(200,1,1,1,2)
datapred[3,] <- c(300,1,1,1,2)
datapred[4,] <- c(350,0,2,1,2) #man, dukes 2 0 recurrent event

##-- Shared frailty model with gamma distribution
sha <- frailtyPenal(Surv(t.stop,event)~cluster(id)+sex+dukes,n.knots=10,
kappa=100000,data=readmission)
pred.sha.rec.marg <- prediction(sha, datapred, t=200, window=seq(50,1900,50),
event='Recurrent',MC.sample=100)
plot(pred.sha.rec.marg,conf.bands=TRUE)

##-- conditional prediction of a recurrent event, on a shared frailty model
pred.sha.rec.cond <- prediction(sha, datapred, t=200, window=seq(50,1900,50),
event='Recurrent',conditional = TRUE,MC.sample=100)
plot(pred.sha.rec.cond,conf.bands=TRUE)

###########################################################################################
### prediction on a JOINT frailty model ###################################################
###########################################################################################
data(readmission)
prediction

```r
##-- predictions of death on a joint model --##
joi <- frailtyPenal(Surv(t.start,t.stop,event)~cluster(id)
+sex+dukes+terminal(death),formula.terminalEvent=~sex
+dukes,data=readmission,n.knots=10,kappa=c(100,100),recurrantAG=TRUE)

##-- construction of the data frame for predictions
datapredj <- data.frame(t.stop=0,event=0,id=0,sex=0,dukes=0)
datapredj$sex <- as.factor(datapredj$sex)
levels(datapredj$sex) <- c(1,2)
datapredj$dukes <- as.factor(datapredj$dukes)
levels(datapredj$dukes) <- c(1,2,3)
datapredj[,1] <- c(100,1,1,1,2)
datapredj[,2] <- c(200,1,1,1,2)
datapredj[,3] <- c(300,1,1,1,2)
datapredj[,4] <- c(400,1,1,1,2)
datapredj[,5] <- c(380,1,2,1,2)

#-- prediction of death between 100 and 100+500 given relapses
pred.joint0 <- prediction(joi,datapredj,t=100,window=500,event = "Terminal")
print(pred.joint0)

#-- prediction of death between 100 and 100+w given relapses
# (with confidence intervals)
pred.joint <- prediction(joi,datapredj,t=100,window=seq(50,1500,50),
event = "Terminal",MC.sample=100)
plot(pred.joint,conf.bands=TRUE)
# each y-value of the plot corresponds to the prediction between [100,x]

#-- prediction of death between t and t+500 given relapses
pred.joint2 <- prediction(joi,datapredj,t=seq(100,1000,50),
window=500,event = "Terminal")
plot(pred.joint2)
# each y-value of the plot corresponds to the prediction between [x,x+500],
# or in the next 500

#-- prediction of relapse between 100 and 100+w given relapses
# (with confidence intervals)
pred.joint <- prediction(joi,datapredj,t=100,window=seq(50,1500,50),
event = "Recurrent",MC.sample=100)
plot(pred.joint,conf.bands=TRUE)
# each y-value of the plot corresponds to the prediction between [100,x]

#-- prediction of relapse and death between 100 and 100+w given relapses
# (with confidence intervals)
pred.joint <- prediction(joi,datapredj,t=100,window=seq(50,1500,50),
event = "Both",MC.sample=100)
plot(pred.joint,conf.bands=TRUE)
# each y-value of the plot corresponds to the prediction between [100,x]
```
data(colorectal)
data(colorectalLongi)

# Survival data preparation - only terminal events
colorectalSurv <- subset(colorectal, new.lesions == 0)

#-- construction of the data-frame for predictions
#-- biomarker observations
datapredj_longi <- data.frame(id = 0, year = 0, tumor.size = 0, treatment = 0,
                                  age = 0, who.PS = 0, prev.resection = 0)
datapredj_longi$treatment <- as.factor(datapredj_longi$treatment)
levels(datapredj_longi$treatment) <- 1:2
datapredj_longi$age <- as.factor(datapredj_longi$age)
levels(datapredj_longi$age) <- 1:3
datapredj_longi$who.PS <- as.factor(datapredj_longi$who.PS)
levels(datapredj_longi$who.PS) <- 1:3
datapredj_longi$prev.resection <- as.factor(datapredj_longi$prev.resection)
levels(datapredj_longi$prev.resection) <- 1:2

# patient 1: increasing tumor size
 datapredj_longi[,1] <- c(1, 0, 1.2, 2, 1, 1, 1)
datapredj_longi[,2] <- c(1, 0.3, 1.4, 2, 1, 1, 1)
datapredj_longi[,3] <- c(1, 0.6, 1.9, 2, 1, 1, 1)
datapredj_longi[,4] <- c(1, 0.9, 2.5, 2, 1, 1, 1)
datapredj_longi[,5] <- c(1, 1.5, 3.9, 2, 1, 1, 1)

# patient 2: decreasing tumor size
 datapredj_longi[,6] <- c(2, 0, 1.2, 2, 1, 1, 1)
datapredj_longi[,7] <- c(2, 0.3, 0.7, 2, 1, 1, 1)
datapredj_longi[,8] <- c(2, 0.5, 0.3, 2, 1, 1, 1)
datapredj_longi[,9] <- c(2, 0.7, 0.1, 2, 1, 1, 1)
datapredj_longi[,10] <- c(2, 0.9, 0.1, 2, 1, 1, 1)

#-- terminal event
 datapredj <- data.frame(id = 0, treatment = 0, age = 0, who.PS = 0,
                          prev.resection = 0)
datapredj$treatment <- as.factor(datapredj$treatment)
levels(datapredj$treatment) <- 1:2
datapredj$age <- as.factor(datapredj$age)
levels(datapredj$age) <- 1:3
datapredj$who.PS <- as.factor(datapredj$who.PS)
datapredj$prev.resection <- as.factor(datapredj$prev.resection)
levels(datapredj$prev.resection) <- 1:2
levels(datapredj$who.PS) <- 1:3
datapredj[,1] <- c(1, 2, 1, 1, 1)
datapredj[,2] <- c(2, 2, 1, 1, 1)

model.spli.CL <- longiPenal(Surv(time1, state) ~ age + treatment + who.PS + prev.resection, tumor.size ~ year * treatment + age + who.PS, colorectalSurv, data.Longi = colorectalLongi, random = c("1", "year"), id = "id", link = "Current-level", left.censoring = -3.33, n.knots = 6, kappa = 1)
#-- prediction of death between 1 year and 1+2 given history of the biomarker
pred.jointLongi0 <- prediction(model.spli.CL, datapredj, datapredj_longi,
t = 1, window = 2)
print(pred.jointLongi0)

#-- prediction of death between 1 year and 1+w given history of the biomarker
pred.jointLongi <- prediction(model.spli.CL, datapredj, datapredj_longi,
t = 1, window = seq(0.5, 2.5, 0.2), MC.sample = 100)
plot(pred.jointLongi, conf.bands = TRUE)
# each y-value of the plot corresponds to the prediction between [1,x]

#-- prediction of death between t and t+0.5 given history of the biomarker
pred.jointLongi2 <- prediction(model.spli.CL, datapredj, datapredj_longi,
t = seq(1, 2.5, 0.5), window = 0.5, MC.sample = 100)
plot(pred.jointLongi2, conf.bands = TRUE)
# each y-value of the plot corresponds to the prediction between [x,x+0.5],
or in the next 0.5

#############################################################################
##### marginal prediction on a JOINT NESTED model for a terminal event ######
#############################################################################
#*--Warning! You can compute this prediction method with ONLY ONE family
#*--by dataset of prediction.
#*--Please make sure your data frame contains a column for individuals AND a
#*--column for the reference number of the family chosen.

data(readmission)
readmissionNested <- transform(readmission, group=id%%30+1)

#-- construction of the data frame for predictions :
#-- family 5 was selected for the prediction
DataPred <- readmissionNested[which(readmissionNested$group==5),]

#-- Fitting the model
modJointNested_Splines <-
  frailtyPenal(formula = Surv(t.start, t.stop, event)~subcluster(id)+
    cluster(group) + dukes + terminal(death),formula.terminalEvent
  =dukes, data = readmissionNested, recurrentAG = TRUE,n.knots = 8,
  kappa = c(9.55e+9, 1.41e+12), initialize = TRUE)

#-- Compute prediction over the individuals 274 and 4 of the family 5
predRead <- prediction(modJointNested_Splines, data=DataPred,t=500,
  window=seq(100,1500,200), conditional=FALSE, individual = c(274,4))

########################################################################
##### prediction on TRIVARIATE JOINT model (linear and non-linear) ######
########################################################################
data(colorectal)
data(colorectallongi)
#-- construction of the data frame for predictions
#-- history of recurrences and terminal event
datapredj <- data.frame(time0 = 0, time1 = 0, new.lesions = 0, id = 0,
treatment = 0, age = 0, who.PS = 0, prev.resection = 0)
datapredj$treatment <- as.factor(datapredj$treatment)
levels(datapredj$treatment) <- 1:2
datapredj$age <- as.factor(datapredj$age)
levels(datapredj$age) <- 1:3
datapredj$who.PS <- as.factor(datapredj$who.PS)
levels(datapredj$who.PS) <- 1:3
datapredj$prev.resection <- as.factor(datapredj$prev.resection)
levels(datapredj$prev.resection) <- 1:2

datapredj[1,] <- c(0, 0.4, 1, 1, 2, 1, 1, 1)
datapredj[2,] <- c(0.4, 1.2, 1, 1, 2, 1, 1, 1)
datapredj[3,] <- c(0, 0.5, 1, 2, 2, 1, 1, 1)

# Linear trivariate joint model
# (computation takes around 40 minutes)
model.trivPenal <- trivPenal(Surv(time0, time1, new.lesions) ~ cluster(id) + age + treatment + who.PS + terminal(state),
formula.terminalEvent =~ age + treatment + who.PS + prev.resection,
formula.size =~ year * treatment + age + who.PS, data = colorectal,
data.Longi = colorectalLongi, random = c("1", "year"), id = "id",
link = "Random-effects", left.censoring = -3.33, recurrentAG = TRUE,
n.knots = 6, kappa=c(0.01, 2), method.GH="Pseudo-adaptive",
n.nodes = 7, init.B = c(-0.07, -0.13, -0.16, -0.17, 0.42, # recurrent events covarates
-0.23, -0.1, -0.09, -0.12, 0.8, -0.23, # terminal event covarates
3.02, -0.3, 0.05, -0.63, -0.02, -0.29, 0.11, 0.74)) # biomarker covariates

#-- prediction of death between 1 year and 1+2
pred.jointTri0 <- prediction(model.trivPenal, datapredj,
datapredj_longi, t = 1, window = 2)
print(pred.jointTri0)

#-- prediction of death between 1 year and 1+w
pred.jointTri <- prediction(model.trivPenal, datapredj,
datapredj_longi, t = 1, window = seq(0.5, 2.5, 0.2), MC.sample = 100)
plot(pred.jointTri, conf.bands = TRUE)

#-- prediction of death between t and t+0.5
pred.jointTri2 <- prediction(model.trivPenal, datapredj,
datapredj_longi, t = seq(1, 2.5, 0.5), window = 0.5, MC.sample = 100)
plot(pred.jointTri2, conf.bands = TRUE)

# No information on dose - creation of a dummy variable
colorectalLongi$dose <- 1

# (computation can take around 40 minutes)
model.trivPenalNL <- trivPenalNL(Surv(time0, time1, new.lesions) ~ cluster(id) + age + treatment + terminal(state), formula.terminalEvent =~ age + treatment, biomarker = "tumor.size", formula.KG ~ 1, formula.KD ~ treatment, dose = "dose", time.biomarker = "year", data = colorectal, data.Longi = colorectalLongi, random = c("y0", "KG"), id = "id", init.B = c(-0.22, -0.16, -0.35, -0.19, 0.04, -0.41, 0.23), init.Alpha = 1.86, init.Eta = c(0.5, 0.57, 0.5, 2.34), init.Biomarker = c(1.24, 0.81, 1.07, -1.53), recurrentAG = TRUE, n.knots = 5, kappa = c(0.01, 2), method.GH = "Pseudo-adaptive")

#-- prediction of death between 1 year and 1+2
pred.jointTriNL0 <- prediction(model.trivPenalNL, datapredj, datapredj_longi, t = 1, window = 2)
print(pred.jointTriNL0)

#-- prediction of death between 1 year and 1+w
pred.jointTriNL <- prediction(model.trivPenalNL, datapredj, datapredj_longi, t = 1, window = seq(0.5, 2.5, 0.2), MC.sample = 100)
plot(pred.jointTriNL, conf.bands = TRUE)

#-- prediction of death between t and t+0.5
pred.jointTriNL2 <- prediction(model.trivPenalNL, datapredj, datapredj_longi, t = seq(2, 3, 0.2), window = 0.5, MC.sample = 100)
plot(pred.jointTriNL2, conf.bands = TRUE)

## End(Not run)

print.additivePenal

Print a Short Summary of parameter estimates of an additive frailty model

Description

Prints a short summary of the parameter estimates of an additive frailty model or more generally of an ‘additivePenal’ object

Usage

## S3 method for class 'additivePenal'
print(x, digits = max(options()$digits - 4, 6), ...)

Arguments

x the result of a call to the additivePenal function
digits number of digits to print
... other unused arguments
**Value**

Print the parameter estimates of the survival or hazard functions.

**See Also**

`additivePenal`

---

### print.Cmeasures

`print.Cmeasures` *Print a short summary of results of `Cmeasure` function.*

**Description**

Print a short summary of results of the concordance measure estimated by the `Cmeasure` function.

**Usage**

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

**Arguments**

- `x` a `Cmeasures` object.
- `...` Other unused arguments

**Value**

Print concordance measures estimated.

**See Also**

`Cmeasures`

---

### print.frailtyPenal

`print.frailtyPenal` *Print a Short Summary of parameter estimates of a shared frailty model.*

**Description**

Prints a short summary of parameter estimates of a `frailtyPenal` object.

**Usage**

```r
## S3 method for class 'frailtyPenal'
print(x, digits = max(options()$digits - 4, 6), ...
```

**Arguments**

- `x` a `frailtyPenal` object.
- `digits` maximum number of digits to be printed.
- `...` Other unused arguments

**Value**

Print parameter estimates.
Arguments

- `x`: the result of a call to the `frailtyPenal` function.
- `digits`: number of digits to print.
- `...`: other unused arguments.

Value

Print the parameter estimates of the survival or hazard functions.

See Also

`frailtyPenal`
print.jointPenal  

Print a Short Summary of parameter estimates of a joint frailty model

Description

Prints a short summary of parameter estimates of a joint frailty model, or more generally an object of class 'frailtyPenal' for joint frailty models.

Usage

## S3 method for class 'jointPenal'
print(x, digits = max(options()$digits - 4, 6), ...)

Arguments

x the result of a call to the jointPenal function
digits number of digits to print
... other unused arguments

Value

Print, separately for each type of event (recurrent and terminal), the parameter estimates of the survival or hazard functions.

See Also

frailtyPenal

print.longiPenal  

Print a Summary of parameter estimates of a joint model for longitudinal data and a terminal event

Description

Prints a short summary of parameter estimates of a joint model for longitudinal data and a terminal event, an object inheriting from class 'longiPenal'.

Usage

## S3 method for class 'longiPenal'
print(x, digits = max(options()$digits - 4, 6), ...)

Arguments

x the result of a call to the longiPenal function
digits number of digits to print
... other unused arguments

Value

Print, separately for each type of event (recurrent and terminal), the parameter estimates of the survival or hazard functions.
print.multivPenal

Arguments

  x  an object inheriting from longiPenal class
  digits  number of digits to print
  ...  other unused arguments

Value

Print, separately for each part of the model (longitudinal and terminal) the parameter estimates and details on the estimation.

See Also

longiPenal

print.multivPenal  

Print a Short Summary of parameter estimates of a multivariate frailty model

Description

Prints a short summary of parameter estimates of a multivariate frailty model, or more generally an object of class 'multivPenal'.

Usage

## S3 method for class 'multivPenal'
print(x, digits = max(options()$digits - 4, 6),
  ...)

Arguments

  x  the result of a call to the multivPenal function
  digits  number of digits to print
  ...  other unused arguments

Value

Print, separately for each type of event (recurrent1, recurrent2 and terminal), the parameter estimates of the survival or hazard functions.

See Also

multivPenal
print.nestedPenal

**print.nestedPenal**

Print a Short Summary of parameter estimates of a nested frailty model

## Description
Prints a short summary of parameter estimates of a nested frailty model

## Usage
```r
## S3 method for class 'nestedPenal'
print(x, digits = max(options()$digits - 4, 6),
      ...)  
```

## Arguments
- **x**: the result of a call to the frailtyPenal function for nested frailty models
- **digits**: number of digits to print
- **...**: other unused arguments

## Value
- **n**: the number of observations used in the fit.
- **n.groups**: the maximum number of groups used in the fit
- **n.events**: the number of events observed in the fit
- **eta**: variance of the subcluster effect (\(Var(w_{ij})\))
- **theta**: variance of the cluster effect (\(Var(v_i)\))
- **coef**: the coefficients of the linear predictor, which multiply the columns of the model matrix.
- **SE(H)**: the standard error of the estimates deduced from the variance matrix of theta and of the coefficients.
- **SE(HIH)**: the standard error of the estimates deduced from the robust estimation of the variance matrix of theta and of the coefficients.
- **p**: p-value

## See Also
- frailtyPenal
print.prediction  
*Print a short summary of results of prediction function.*

**Description**

Print a short summary of results of prediction function.

**Usage**

```r
## S3 method for class 'predFrailty'
print(x, digits = 3, ...)
## S3 method for class 'predJoint'
print(x, digits = 3, ...)
## S3 method for class 'predLongi'
print(x, digits = 3, ...)
```

**Arguments**

- `x`: An object from the 'prediction' function, objects inheriting from `predFrailty`, `predJoint` and `predLongi` classes.
- `digits`: Number of digits to print
- `...`: Other unused arguments

**Value**

Print the probabilities estimated.

**See Also**

`prediction`

---

print.trivPenal  
*Print a Summary of parameter estimates of a joint model for longitudinal data, recurrent events and a terminal event*

**Description**

Prints a short summary of parameter estimates of a joint model for longitudinal data, recurrent events and a terminal event, an object inheriting from class 'trivPenal'.

**Usage**

```r
## S3 method for class 'trivPenal'
print(x, digits = max(options()$digits - 4, 6), ...)
```
Arguments

x an object inheriting from trivPenal class
digits number of digits to print
... other unused arguments

Value

Print, separately for each part of the model (longitudinal, recurrent and terminal) the parameter estimates and details on the estimation.

See Also

trivPenal

print.trivPenalNL

Print a Summary of parameter estimates of a non-linear trivariate joint model for longitudinal data, recurrent events and a terminal event

Description

Prints a short summary of parameter estimates of a non-linear trivariate joint model for longitudinal data, recurrent events and a terminal event, an object inheriting from class 'trivPenalNL'.

Usage

## S3 method for class 'trivPenalNL'
print(x, digits = max(options()$digits - 4, 6), ...)

Arguments

x an object inheriting from trivPenalNL class
digits number of digits to print
... other unused arguments

Value

Print, separately for each part of the model (biomarker growth, biomarker decline, recurrent events and terminal event) the parameter estimates and details on the estimation.

See Also

trivPenalNL
Description

This contains rehospitalization times after surgery in patients diagnosed with colorectal cancer

Usage

data(readmission)

Format

This data frame contains the following columns:

- **id**: identification of each subject. Repeated for each recurrence
- **enum**: which readmission
- **t.start**: start of interval (0 or previous recurrence time)
- **t.stop**: recurrence or censoring time
- **time**: interoccurrence or censoring time
- **event**: rehospitalization status. All event are 1 for each subject excepting last one that it is 0
- **chemo**: Did patient receive chemotherapy? 1: No; 2: Yes
- **sex**: gender: 1:Males 2:Females
- **dukes**: Dukes’ tumoral stage: 1:A-B; 2:C 3:D
- **charlson**: Comorbidity Charlson’s index. Time-dependent covariate. 0: Index 0; 1: Index 1-2; 3: Index >=3
- **death**: death indicator. 1:dead and 0:alive

Source

runShiny

Shiny application for modelisation and prediction of frailty models

Description

This function loads the shiny package and runs the application for modelisation and prediction of several frailty models using package frailtypack.

Usage

runShiny()

Value

No value returned.

References

Rizopoulos D. (2016)

Examples

## Not run:
runShiny()

## End(Not run)

slope

Identify variable associated with the random slope

Description

This is a special function used in the context of survival additive models. It identifies the variable which is in interaction with the random slope ($v_i$). Generally, this variable is the treatment variable. Using interaction() in a formula implies that an additive frailty model is fitted.

Usage

slope(x)

Arguments

x A factor, a character or a numerical variable
Value

  x  The variable in interaction with the random slope

Note

It is necessary to specify which variable is in interaction with the random slope, even if only one explanatory variable is included in the model.

See Also

  additivePenal

Examples

````
## Not run:

data(dataAdditive)

##-- Additive with one covariate --##
modAdd1cov <- additivePenal(Surv(t1,t2,event)~cluster(group)+var1+ slope(var1),data=dataAdditive,n.knots=8,kappa=10000,hazard="Splines")

##-- Additive with two covariates --##
set.seed(1234)
dataAdditive$var2 <- rbinom(nrow(dataAdditive),1,0.5)
modAdd2cov <- additivePenal(Surv(t1,t2,event)~cluster(group)+var1+ var2+slope(var1),data=dataAdditive,n.knots=8,kappa=10000, hazard="Splines")

##-- Additive with 2 covariates and stratification --##
dataAdditive$var2 <- rbinom(nrow(dataAdditive),1,0.5)
modAddstrat <- additivePenal(Surv(t1,t2,event)~cluster(group)+ strata(var2)+var1+slope(var1),data=dataAdditive,n.knots=8, kappa=c(10000,10000),hazard="Splines")

## End(Not run)
````
**ste**  
*Surrogate threshold effect for the one-step Joint surrogate model for the evaluation of a candidate surrogate endpoint.*

### Description

This function computes the surrogate threshold effect (STE) from the one-step joint surrogate model. The STE is defined as the minimum treatment effect on the surrogate necessary to predict a non-zero effect on the true endpoint (Burzykowski et al., 2006).

### Usage

```r
ste(object, var.used = "error.estim", alpha. = 0.05, 
pred.int.use = "up")
```

### Arguments

- **object**: An object inheriting from `jointSurroPenal` class (output from calling the function `jointSurroPenal`).
- **var.used**: This argument takes two values. The first one is "error.estim" and indicates if the prediction error take into account the estimation error of the estimates of the parameters. If the estimates are supposed to be known or if the dataset includes a high number of trials with a high number of subject per trial, value `No.error` can be used. The default is `error.estim`.
- **alpha.**: The confidence level for the prediction interval. The default is `0.05`.
- **pred.int.use**: A character string that indicates the bound of the prediction interval to use to compute the STE. Possible values are `up` for the upper bound (the default) or `lw` for the lower bound.

### Details

The STE is obtained by solving the equation \( l(\alpha_0) = 0 \) (resp. \( u(\alpha_0) = 0 \)), where \( \alpha_0 \) represents the corresponding STE, and \( l(\alpha_0) \) (resp. \( u(\alpha_0) \)) is the lower (resp. upper) bound of the prediction interval of the treatment effect on the true endpoint \((\beta + b_0)\). Thereby,

\[
l(\alpha_0) \equiv E(\beta + b_0|\alpha_0, \vartheta) - Z_{1-(\gamma/2)} \sqrt{Var(\beta + b_0|\alpha_0, \vartheta)}
\]

and

\[
u(\alpha_0) \equiv E(\beta + b_0|\alpha_0, \vartheta) + Z_{1-(\gamma/2)} \sqrt{Var(\beta + b_0|\alpha_0, \vartheta)}
\]

where \( \vartheta \) represents the set of estimates for the fixed-effects and the variance-covariance parameters of the random effects obtained from the joint surrogate model (Sofeu et al., 2018). Given that negative values of treatment effect indicate a reduction of the risk of failure and are considered beneficial, STE is recommended to be computed from the upper prediction limit \( u(\alpha_0) \). The details on the computation of STE is described in Burzykowski et al. (2006).
Value

Returns and displays the STE.

Author(s)

Casimir Ledoux Sofeu <casimir.sofeu@u-bordeaux.fr>, <scl.ledoux@gmail.com> and Virginie Rondeau <virginie.rondeau@inserm.fr>

References


See Also

jointSurroPenal predict

Examples

## Not run:

#### Joint surrogate model ####
#### evaluation of surrogate endpoints ####

data(dataOvarian)
joint.surro.ovar <- jointSurroPenal(data = dataOvarian, n.knots = 8,
           init.kappa = c(2000,1000), indicator.alpha = 0,
           nb.mc = 200, scale = 1/365)

# =====STE=====
# ste(joint.surro.ovar, var.used = "error.estim")
# Assuming no errors on the estimates
# ste(joint.surro.ovar, var.used = "No.error", pred.int.use = "up")

## End(Not run)
subcluster

Identify subclusters

Description

This is a special function used in the context of survival nested or joint nested models. It identifies correlated groups of observations within other groups defined by using 'cluster' function from 'survival' package, and is used on the right hand side of 'frailtyPenal' formula for fitting a nested or joint nested model. Using subcluster() in a formula implies that a nested or a joint nested frailty model is estimated.

Usage

subcluster(x)

Arguments

x
A character, factor, or numeric variable which is supposed to indicate the variable subgroup

Value

x
A variable identified as a subcluster

See Also

frailtyPenal

Examples

## Not run:

data(dataNested)
modClu <- frailtyPenal(Surv(t1,t2,event)~cluster(group)+subcluster(subgroup)+cov1+cov2,data=dataNested, n.knots=8,kappa=c(50000,50000),hazard="Splines")
print(modClu)

#-- here is generated cluster (30 clusters)
readmissionNested <- transform(readmission,group=id%%30+1)

modJointNested_Splines <- frailtyPenal(formula = Surv(t.start, t.stop, event) ~ subcluster(id) + cluster(group) + dukes + terminal(death), formula.terminalEvent = ~dukes, data = readmissionNested, recurrentAG = TRUE, n.knots = 8, kappa = c(9.55e+9, 1.41e+12), initialize = TRUE)
summary.additivePenal  summary of parameter estimates of an additive frailty model

Description
This function returns hazard ratios (HR) and its confidence intervals.

Usage
## S3 method for class 'additivePenal'
summary(object, level = 0.95, len = 6, d = 2,
  lab="hr", ...)

Arguments
- **object**: output from a call to additivePenal.
- **level**: significance level of confidence interval. Default is 95%.
- **len**: the total field width. Default is 6.
- **d**: the desired number of digits after the decimal point. Default of 6 digits is used.
- **lab**: label of printed results.
- **...**: other unused arguments.

Value
Prints HR and its confidence intervals for each covariate. Confidence level is allowed (level argument).

See Also
- additivePenal

Examples

## Not run:
data(dataAdditive)

modAdd <- additivePenal(Surv(t1,t2,event)~cluster(group)+var1+slope(var1),
correlation=TRUE,data=dataAdditive,n.knots=8,kappa=862,hazard="Splines")
summary.frailtyPenal

summary of parameter estimates of a shared frailty model

Description

This function returns hazard ratios (HR) and its confidence intervals.

Usage

## S3 method for class 'frailtyPenal'
summary(object, level = 0.95, len = 6, d = 2,
lab="hr", ...)  

Arguments

- `object`: output from a call to frailtyPenal.
- `level`: significance level of confidence interval. Default is 95%.
- `len`: the total field width. Default is 6.
- `d`: the desired number of digits after the decimal point. Default of 6 digits is used.
- `lab`: label of printed results.
- `...`: other unused arguments.

Value

Prints HR and its confidence intervals. Confidence level is allowed (level argument).

See Also

frailtyPenal
Examples

```r
## Not run:
data(kidney)
##-- Shared frailty model --##
modSha <- frailtyPenal(Surv(time,status)-age+sex+cluster(id),
n.knots=8,kappa=10000,data=kidney,hazard="Splines")
##-- Cox proportional hazard model --##
modCox <- frailtyPenal(Surv(time,status)-age+sex,
n.knots=8,kappa=10000,data=kidney,hazard="Splines")
#-- confidence interval at 95% level (default)
summary(modSha)
summary(modCox)
#-- confidence interval at 99% level
summary(modSha,level=0.99)
summary(modCox,level=0.99)
## End(Not run)
```

### Description

This function returns hazard ratios (HR) and its confidence intervals.

### Usage

```r
## S3 method for class 'jointNestedPenal'
summary(object, level = 0.95, len = 6, d = 2, lab="hr", ...)
```
summary.jointPenal

**Arguments**

- **object**: output from a call to frailtyPenal for joint nested models
- **level**: significance level of confidence interval. Default is 95%.
- **len**: the total field width. Default is 6.
- **d**: the desired number of digits after the decimal point. Default of 6 digits is used.
- **lab**: label of printed results.
- ... other unused arguments.

**Value**

Prints HR and its confidence intervals for each covariate. Confidence level is allowed (level argument).

**See Also**

frailtyPenal

**Examples**

```r
## Not run:

#-- here is generated cluster (30 clusters)
readmissionNested <- transform(readmission, group=id%%30+1)

# Baseline hazard function approximated with splines with calendar-timescale

model.spli.AG <- frailtyPenal(formula = Surv(t.start, t.stop, event) ~ subcluster(id) + cluster(group) + dukes + terminal(death),
 formula.terminalEvent = ~dukes, data = readmissionNested, recurrentAG = TRUE, n.knots = 8, kappa = c(9.55e+9, 1.41e+12),
 initialize = TRUE)

summary(model.spli.AG)

## End(Not run)
```

### Description

This function returns hazard ratios (HR) and its confidence intervals.
Usage

## S3 method for class 'jointPenal'
summary(object, level = 0.95, len = 6, d = 2,
lab="hr", ...)

Arguments

object    output from a call to frailtyPenal for joint models
level     significance level of confidence interval. Default is 95%.
len       the total field width. Default is 6.
d         the desired number of digits after the decimal point. Default of 6 digits is used.
lab       label of printed results.
...       other unused arguments.

Value

Prints HR and its confidence intervals for each covariate. Confidence level is allowed (level argument).

See Also

frailtyPenal

Examples

## Not run:
data(readmission)

#-- gap-time
modJoint.gap <- frailtyPenal(Surv(time,event)~cluster(id)+sex+dukes+
charlson+terminal(death),formula.terminalEvent=~sex+dukes+charlson,
data=readmission,n.knots=14,kappa=c(9.55e+9,1.41e+12))

#-- calendar time
modJoint.calendar <- frailtyPenal(Surv(t.start,t.stop,event)~cluster(id)+
sex+dukes+charlson+terminal(death),formula.terminalEvent=~sex+dukes+charlson,
data=readmission,n.knots=10,kappa=c(9.55e+9,1.41e+12),recurrrentAG=TRUE)

#-- It takes around 1 minute to converge
summary(modJoint.gap)
summary(modJoint.calendar)

## End(Not run)
**summary.jointSurroPenal**

*Short summary of the random effects parameters, the fixed treatment effects, and the surrogacy evaluation criteria estimated from a joint surrogate model*

**Description**

This function returns the estimate of the coefficients and their standard error with p-values of the Wald test for the joint surrogate model, also hazard ratios (HR) and their confidence intervals for the fixed treatment effects, and finally an estimate of the surrogacy evaluation criterion (Kendall’s $\tau$ and $R^2_{trial}$).

**Usage**

```r
## S3 method for class 'jointSurroPenal'
summary(object, d = 4, len = 3, int.method.kt = 0, nb.gh = 32, ...)
```

**Arguments**

- `object`: an object inheriting from `jointSurroPenal` class.
- `d`: The desired number of digits after the decimal point for parameters. The maximum of 4 digits is required for the estimates. Default of 3 digits is used.
- `len`: The desired number of digits after the decimal point for p-value and convergence criteria. Default of 4 digits is used.
- `int.method.kt`: A binary, indicates the integration method for Kendall’s $\tau$ estimation: 0 for Monte carlo, and 1 for Gaussian Hermite quadrature. The default is 0.
- `nb.gh`: Number of nodes for the Gaussian-Hermite quadrature. The default is 32 1 for Gaussian-Hermite quadrature.
- `...`: other unused arguments.

**Value**

For the variances parameters of the random effects, it prints the estimate of the coefficients with their standard error, Z-statistics and p-values of the Wald test. For the fixed treatment effects, it also prints HR and its confidence intervals for each covariate. For the surrogacy evaluation criteria, it prints the estimated Kendall’s $\tau$ with its 95% Confidence interval obtained by the parametric bootstrap, the estimated $R^2_{trial}$ with standard error and the 95% Confidence interval obtained by Delta-method (Dowd et al., 2014), $R^2_{trial}$ and its 95% Confidence interval obtained by the parametric bootstrap. We notice that, using the bootstrap, the standard error of the point estimate is not available. We propose a classification of $R^2_{trial}$ according to a modification to surrogate criteria proposed by the Institute of Quality and Efficiency in Health Care (Prasad et al., 2015). We also display the surrogate threshold effect (STE) with the associated hazard risk. The rest of parameters concerns the convergence characteristics and included: the penalized marginal log-likelihood, number of iterations, the LCV and the Convergence criteria.
**Author(s)**

Casimir Ledoux Sofeu <casimir.sofeu@u-bordeaux.fr>, <scl.ledoux@gmail.com> and Virginie Rondeau <virginie.rondeau@inserm.fr>

**References**


**See Also**

`jointSurroPenal` `jointSurroTKendall`

**Examples**

```r
###---Data generation---###
data.sim <- jointSurrSimul(n.obs=400, n.trial = 20, cens.adm=549,
alpha = 1.5, theta = 3.5, gamma = 2.5, zeta = 1,
sigma.s = 0.7, sigma.t = 0.7, rsqrt = 0.8, betas = -1.25,
betat = -1.25, full.data = 0, random.generator = 1,
seed = 0, nb.reject.data = 0)

## Not run:
###---Estimation---###
joint.surrogate <- jointSurroPenal(data = data.sim, nb.mc = 300,
nb.gh = 20, indicator.alpha = 1, n.knots = 6)

summary(joint.surrogate)
summary(joint.surrogate, d = 4, len = 3, int.method.kt = 1, nb.gh = 25)

## End(Not run)
```

---

**Description**

This function returns the true value, the mean of the estimates, the empirical standard error, the mean of the estimated standard errors (Mean SE), and the coverage probability for model parameters.
Usage

```r
## S3 method for class 'jointSurroPenalSimul'
summary(object, d = 3, R2boot = 0, ...)
```

Arguments

- `object`: an object inheriting from `jointSurroPenalSimul` class.
- `d`: The desired number of digits after the decimal point. Default of 3.
- `R2boot`: A binary that specifies whether the confidence interval of $R^2_{\text{trial}}$ should be computed using parametric bootstrap (1) or Delta-method (0). The default is 0.
- `...`: other unused arguments.

Value

For each parameter of the joint surrogate model, we print the true simulation value, the empirical standard error (empirical SE), the mean of the estimated standard errors (Mean SE), and the coverage probability (CP) for each model parameters. For the Kendall’s $\tau$, the 95% Confidence interval is obtained by the parametric bootstrap. For $R^2_{\text{trial}}$($R^2_{\text{trial}}$), the standard error is obtained by Delta-method and the 95% Confidence interval could be obtained directly or by parametric bootstrap. We also display the total number of non convergence case with the associated percentage ($R:n(\%)$), the mean number of iterations to reach convergence, and other estimation and simulation parameters.

Author(s)

Casimir Ledoux Sofeu `<casimir.sofeu@u-bordeaux.fr>`, `<scl.ledoux@gmail.com>` and Virginie Rondeau `<virginie.rondeau@inserm.fr>`

See Also

- `jointSurroPenalSimul`

Examples

```r
# Studies simulation

## Not run:
# (Computation takes around 45 minutes using a processor including 40
# cores and a read only memory of 378 Go)
joint.simul <- jointSurroPenalSimul(nb.dataset = 10, nbSubSimul=600,
                                   ntrialSimul=30, LIMparam = 0.001, LIMlogl = 0.001,
                                   LIMderiv = 0.001, nb.mc = 200, nb.gh = 20,
                                   nb.gh2 = 32, true.init.val = 1, print.iter=F)

# results
summary(joint.simul, d = 3, R2boot = 1) # bootstrap
summary(joint.simul, d = 3, R2boot = 0) # Delta-method

## End(Not run)
```
**summary.longiPenal**  
*Short summary of fixed covariates estimates of a joint model for longitudinal data and a terminal event*

**Description**

This function returns coefficients estimates and their standard error with p-values of the Wald test for the longitudinal outcome and hazard ratios (HR) and their confidence intervals for the terminal event.

**Usage**

```r
## S3 method for class 'longiPenal'
summary(object, level = 0.95, len = 6, d = 2,
lab=c("coef","hr"), ...)  
```

**Arguments**

- `object` an object inheriting from longiPenal class
- `level` significance level of confidence interval. Default is 95%.
- `len` the total field width for the terminal part. Default is 6.
- `d` the desired number of digits after the decimal point. Default of 6 digits is used.
- `lab` labels of printed results for the longitudinal outcome and the terminal event respectively.
- `...` other unused arguments.

**Value**

For the longitudinal outcome it prints the estimates of coefficients of the fixed covariates with their standard error and p-values of the Wald test. For the terminal event it prints HR and its confidence intervals for each covariate. Confidence level is allowed (level argument).

**See Also**

- `longiPenal`

**Examples**

```r
## Not run:
#### Joint model for longitudinal data and a terminal event ####

data(colorectal)
data(colorectallLongi)
```
# Survival data preparation - only terminal events
colorectalSurv <- subset(colorectal, new.lesions == 0)

# Baseline hazard function approximated with splines
# Random effects as the link function
model.spli.RE <- longiPenal(Surv(time1, state) ~ age + treatment + who.PS
+ prev.resection, tumor.size ~ year * treatment + age + who.PS
, colorectalSurv, data.Longi = colorectalLongi, random = c("1", "year"),
  id = "id", link = "Random-effects", left.censoring = -3.33,
  n.knots = 7, kappa = 2)

# Weibull baseline hazard function
# Current level of the biomarker as the link function
model.weib.CL <- longiPenal(Surv(time1, state) ~ age + treatment + who.PS
+ prev.resection, tumor.size ~ year * treatment + age + who.PS
, colorectalSurv, data.Longi = colorectalLongi, random = c("1", "year"),
  id = "id", link = "Current-level", left.censoring = -3.33, hazard = "Weibull")

summary(model.spli.RE)
summary(model.weib.CL)

## End(Not run)

---

**summary.multivPenal**

**summary of parameter estimates of a multivariate frailty model.**

**Description**

This function returns hazard ratio (HR) and its confidence intervals.

**Usage**

```r
## S3 method for class 'multivPenal'
summary(object, level = 0.95, len = 6, d = 2, lab = "hr", ...)
```

**Arguments**

- **object**: output from a call to multivPenal for joint multivariate models
- **level**: significance level of confidence interval. Default is 95%.
- **len**: the total field width. Default is 6.
- **d**: the desired number of digits after the decimal point. Default of 6 digits is used.
- **lab**: label of printed results.
- **...**: other unused arguments.
Value

Prints HR and its confidence intervals for each covariate. Confidence level is allowed (level argument).

See Also

multivPenal

Description

This function returns hazard ratios (HR) and its confidence intervals for each regression coefficient.

Usage

```r
## S3 method for class 'nestedPenal'
summary(object, level = 0.95, len = 6, d = 2,
lab="hr", ...)
```

Arguments

- `object`: output from a call to nestedPenal.
- `level`: significance level of confidence interval. Default is 95%.
- `len`: the total field width. Default is 6.
- `d`: the desired number of digits after the decimal point. Default of 6 digits is used.
- `lab`: label of printed results.
- `...`: other unused arguments.

Value

Prints HR and its confidence intervals for each regression coefficient. Confidence level is allowed (level argument).

See Also

frailtyPenal

summary.nestedPenal summary of regression coefficient estimates of a nested frailty model
Examples

## Not run:

data(dataNested)

modNested <- frailtyPenal(Surv(t1,t2,event)~cluster(group)+
subcluster(subgroup)+cov1+cov2,data=dataNested,
n.knots=8,kappa=c(50000,50000),hazard="Splines")

#- It takes 90 minutes to converge (depends on processor)

summary(modNested)

## End(Not run)

summary.trivPenal  

Short summary of fixed covariates estimates of a joint model for longitudinal data, recurrent events and a terminal event

Description

This function returns coefficients estimates and their standard error with p-values of the Wald test for the longitudinal outcome and hazard ratios (HR) and their confidence intervals for the terminal event.

Usage

## S3 method for class 'trivPenal'

summary(object, level = 0.95, len = 6, d = 2,
lab=c("coef","hr"), ...)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>object</td>
<td>an object inheriting from trivPenal class</td>
</tr>
<tr>
<td>level</td>
<td>significance level of confidence interval. Default is 95%.</td>
</tr>
<tr>
<td>len</td>
<td>the total field width for the terminal part. Default is 6.</td>
</tr>
<tr>
<td>d</td>
<td>the desired number of digits after the decimal point. Default of 6 digits is used.</td>
</tr>
<tr>
<td>lab</td>
<td>labels of printed results for the longitudinal outcome and the terminal event respectively.</td>
</tr>
<tr>
<td>...</td>
<td>other unused arguments.</td>
</tr>
</tbody>
</table>
Value

For the longitudinal outcome it prints the estimates of coefficients of the fixed covariates with their standard error and p-values of the Wald test. For the terminal event it prints HR and its confidence intervals for each covariate. Confidence level is allowed (level argument).

See Also

trivPenal

Examples

```r
## Not run:
###--- Trivariate joint model for longitudinal data, ---###
###--- recurrent events and a terminal event ---###
data(colorectal)
data(colorectalLongi)

# Weibull baseline hazard function
# Random effects as the link function, Gap timescale
# (computation takes around 30 minutes)
model.weib.RE.gap <- trivPenal(Surv(gap.time, new.lesions) ~ cluster(id) + age + treatment + who.PS + prev.resection + terminal(state),
                              formula.terminalEvent =~ age + treatment + who.PS + prev.resection,
                              tumor.size ~ year * treatment + age + who.PS, data = colorectal,
                              data.Longi = colorectalLongi, random = c("1", "year"), id = "id",
                              link = "Random-effects", left.censoring = -3.33, recurrentAG = FALSE,
                              hazard = "Weibull", method.GH="Pseudo-adaptive", n.nodes = 7)

summary(model.weib.RE.gap)

## End(Not run)
```

Description

This function returns coefficients estimates and their standard error with p-values of the Wald test for the biomarker growth (KG) and decline (KD) and hazard ratios and their confidence intervals for the terminal event.
Usage

## S3 method for class 'trivPenalNL'
summary(object, level = 0.95, len = 6, d = 2,
lab=c("coef","hr"), ...)

Arguments

- **object**: an object inheriting from trivPenal class
- **level**: significance level of confidence interval. Default is 95%.
- **len**: the total field width for the terminal part. Default is 6.
- **d**: the desired number of digits after the decimal point. Default of 6 digits is used.
- **lab**: labels of printed results for the longitudinal outcome and the terminal event respectively.
- **...**: other unused arguments.

Value

For the longitudinal outcome it prints the estimates of coefficients of the fixed covariates with their standard error and p-values of the Wald test (separately for the biomarker growth and decline). For the terminal event it prints HR and its confidence intervals for each covariate. Confidence level is allowed (level argument).

See Also

- trivPenalNL

Examples

## Not run:
###--- Trivariate joint model for longitudinal data, ---###
###--- recurrent events and a terminal event ---###
data(colorectal)
data(colorectalLongi)

# Weibull baseline hazard function
# Random effects as the link function, Gap timescale
# (computation takes around 30 minutes)
model.weib.RE.gap <-trivPenal(Surv(gap.time, new.lesions) ~ cluster(id) + age + treatment + who.PS + prev.resection + terminal(state),
formula.terminalEvent =~ age + treatment + who.PS + prev.resection,
tumor.size ~ year * treatment + age + who.PS, data = colorectal,
data.Longi = colorectalLongi, random = c("1", "year"), id = "id",
link = "Random-effects", left.censoring = -3.33, recurrentAG = FALSE,
hazard = "Weibull", method.GH="Pseudo-adaptive", n.nodes = 7)
summary(model.weib.RE.gap)

## End(Not run)

**survDat**  
*Survival dataset (TPJM)*

---

**Description**

This is a simulated dataset used to illustrate the two-part joint model included in the longiPenal function.

**Usage**

data(survDat)

**Format**

This data frame contains the following columns:

- **id**: The identification number of a patient
- **deathTimes**: The event times (death or censoring)
- **d**: Censoring indicator
- **trt**: Treatment covariate

---

**SurvIC**  
*Create a survival object for interval censoring and possibly left truncated data*

---

**Description**

This is a function used in case of interval-censoring as a response variable in a model formula only for Cox proportional hazard or shared frailty model. Sometimes, an unobserved event might occur in a time interval [L,U]. RecurrentAG argument gets invalid with the use of SurvIC. Note that this function used a Kronecker product which can suffer from computation issue when the number of subjects in each cluster is high. Time dependent variables are not allowed.

**Usage**

SurvIC(t0, lower, upper, event)
**survival**

**Survival function**

**Description**

Let t be a continuous variable, we determine the value of the survival function to t after run fit.

**Usage**

```r
survival(t, ObjFrailty)
```
terminal

Arguments

  t    time for survival function.

ObjFrailty   an object from the frailtypack fit.

Value

return the value of survival function in t.

Examples

  ## Not run:
  #-- a fit Shared
  data(readmission)

  fit.shared <- frailtyPenal(Surv(time,event)~dukes+cluster(id)+
     strata(sex),n.knots=10,kappa=c(10000,10000),data=readmission)

  #-- calling survival
  survival(20,fit.shared)

  ## End(Not run)

terminal   Identify terminal indicator

Description

This is a special function used in the context of recurrent event models with terminal event (e.g.,
censoring variable related to recurrent events). It contains the status indicator, normally 0=alive,
1=dead, and is used on the right hand side of a formula of a ‘frailtyPenal’, ‘longiPenal’ and ‘trivPen-
al’ functions. Using terminal() in a formula implies that a joint frailty model for recurrent events
and terminal events is fitted.

Usage

  terminal(x)

Arguments

  x    A numeric variable but should be a Boolean which equals 1 if the subject is dead
       and 0 if he is alive or censored, as a death indicator.
Value

x a death indicator

See Also

frailtyPenal

timedep Identify time-varying effects

Description

This is a special function used in the context of Cox models and shared and joint frailty models. It identifies time-varying effects of covariates in the model. It is used in ‘frailtyPenal’ on the right hand side of formula or of formula.terminalEvent.

When considering time-varying effects in a survival model, regression coefficients can be modeled with a linear combination of B-splines \( B(t) \) with coefficients \( \zeta \) of order \( q \) with \( m \) interior knots:

\[
\beta(t) = \sum_{j=-q+1}^{m} \zeta_j B_{j,q}(t)
\]

You can notice that a linear combination of B-splines of order 1 without any interior knots (0 interior knot) is the same as a model without time-varying effect (or with constant effect over time).

Statistical tests (likelihood ratio tests) can be done in order to know whether the time-dependent coefficients are significantly different from zero or to test whether a covariate has a time-dependent effect significantly different from zero or not. These tests are correct only with a parametric approach yet.

- Proportional Hazard assumption ?
Time-dependency of a covariate effect can be tested. We need to estimate \( m + q \) parameters \( \zeta_j \) for \( j = -q + 1, \ldots, m \) for a time-varying coefficient. Only one \((q = 1, m = 0)\) parameter is estimated for a constant effect. A global test is done.

\[
H_0 : \beta(t) = \beta
\]

The corresponding LR statistic has a \( \chi^2 \) distribution of degree \( m + q - 1 \).

- Significant association ?
We can also use a LR test to test whether a covariate has a significant effect on the hazard function. The null hypothesis is:

\[
H_0 : \beta(t) = 0
\]

For that we fit a model considering the covariate with a regression coefficient modeled using B-splines and a model without the covariate. Hence, the LR statistic has a \( \chi^2 \) distribution of degree \( m + q \).
Usage

timedep(x)

Arguments

x  A numerical or a factor variable that would have a time-varying effect on the event

Value

x  A variable identified with a time-varying effect

References


Examples

```r
## Not run:
data(readmission)

###--- Shared Frailty model with time-varying effect ---###
sha.time <- frailtyPenal(Surv(time,event)~cluster(id)+dukes+charlson+timedep(sex)+chemo,data=readmission,n.knots=8,kappa=1,betaknots=3,betaorder=3)

#-- print results of the fit and the associated curves for the
time-dependent effects
print(sha.time)

###--- Joint Frailty model with time-varying effect ---###
joi.time <- frailtyPenal(Surv(time,event)~cluster(id)+timedep(sex)+chemo+terminal(death),formula.terminalEvent=~timedep(sex)+chemo,
data=readmission,n.knots=8,kappa=c(1,1),betaknots=3,betaorder=3)

print(joi.time)

## End(Not run)
```
**Description**

Fit a trivariate joint model for longitudinal data, recurrent events and a terminal event using a semi-latent structure (Krol et al. 2015):

\[
\begin{align*}
    y_{ik}(t_{ik}) &= X_{Li}(t_{ik})^T \beta_L + Z_i(t_{ik})^T b_i + e_i(t_{ik}) \\
    r_{ij}(t|b_i) &= r_0(t) \exp(v_i + X_{Rij}(t) \beta_R + g(b_i, \beta_L, Z_i(t), X_{Li}(t))^T \eta_R) \\
    \lambda_i(t|b_i) &= \lambda_0(t) \exp(\alpha v_i + X_{Ti}(t) \beta_T + h(b_i, \beta_L, Z_i(t), X_{Li}(t))^T \eta_T)
\end{align*}
\]

where \( X_{Li}(t), X_{Rij}(t) \) and \( X_{Ti}(t) \) are vectors of fixed effects covariates and \( \beta_L, \beta_R \) and \( \beta_T \) are the associated coefficients. Measurements errors \( e_i(t_{ik}) \) are iid normally distributed with mean 0 and variance \( \sigma_e^2 \). The random effects \( b_i = (b_{i1}, \ldots, b_{iq})^T \sim N(0, B_1) \) are associated to covariates \( Z_i(t) \) and independent from the measurement error. The relationship between the biomarker and recurrent events is explained via \( g(b_i, \beta_L, Z_i(t), X_{Li}(t)) \) with coefficients \( \eta_R \) and between the biomarker and terminal event is explained via \( h(b_i, \beta_L, Z_i(t), X_{Li}(t)) \) with coefficients \( \eta_T \). Two forms of the functions \( g(\cdot) \) and \( h(\cdot) \) are available: the random effects \( b_i \) and the current biomarker level \( m_i(t) = X_{Li}(t_{ik})^T \beta_L + Z_i(t_{ik})^T b_i \). The frailty term \( v_i \) is gaussian with mean 0 and variance \( \sigma_v \). Together with \( b_i \) constitutes the random effects of the model:

\[
u_i = \begin{pmatrix} b_i \\ v_i \end{pmatrix} \sim N(0, \begin{pmatrix} B_1 & 0 \\ 0 & \sigma_v^2 \end{pmatrix})
\]

We consider that the longitudinal outcome can be a subject to a quantification limit, i.e. some observations, below a level of detection \( s \) cannot be quantified (left-censoring).

**Usage**

```r
```

**Arguments**

- `formula` 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 like in survival package. Interactions are possible using * or :.
formula.terminalEvent

A formula object, only requires terms on the right to indicate which variables are modelling the terminal event. Interactions are possible using * or :.

formula.LongitudinalData

A formula object, only requires terms on the right to indicate which variables are modelling the longitudinal outcome. It must follow the standard form used for linear mixed-effects models. Interactions are possible using * or :.

data

A 'data.frame' with the variables used in formula.

data.Longi

A 'data.frame' with the variables used in formula.LongitudinalData.

random

Names of variables for the random effects of the longitudinal outcome. Maximum 3 random effects are possible at the moment. The random intercept is chosen using "1".

id

Name of the variable representing the individuals.

intercept

Logical value. Is the fixed intercept of the biomarker included in the mixed-effects model? The default is TRUE.

link

Type of link functions for the dependence between the biomarker and death and between the biomarker and the recurrent events: "Random-effects" for the association directly via the random effects of the biomarker, "Current-level" for the association via the true current level of the biomarker. The option "Current-level" can be chosen only if the biomarker random effects are associated with the intercept and time (following this order). The default is "Random-effects".

left.censoring

Is the biomarker left-censored below a threshold $s$? If there is no left-censoring, the argument must be equal to FALSE, otherwise the value of the threshold must be given.

recurrentAG

Logical value. Is Andersen-Gill model fitted? If so indicates that recurrent event times with the counting process approach of Andersen and Gill is used. This formulation can be used for dealing with time-dependent covariates. The default is FALSE.

n.knots

Integer giving the number of knots to use. Value required in the penalized likelihood estimation. It corresponds to the (n.knots+2) splines functions for the approximation of the hazard or the survival functions. We estimate I or M-splines of order 4. When the user set a number of knots equals to k (n.knots=k) then the number of interior knots is (k-2) and the number of splines is (k-2)+order. Number of knots must be between 4 and 20. (See Note in frailtyPenal function)

kappa

Positive smoothing parameters in the penalized likelihood estimation. The coefficient kappa of the integral of the squared second derivative of hazard function in the fit (penalized log likelihood). To obtain an initial value for kappa, a solution is to fit the corresponding Cox model using cross validation (See cross.validation in function frailtyPenal). We advise the user to identify several possible tuning parameters, note their defaults and look at the sensitivity of the results to varying them.

maxit

Maximum number of iterations for the Marquardt algorithm. Default is 300

hazard

Type of hazard functions: "Splines" for semiparametric hazard functions using equidistant intervals or "Splines-per" using percentile with the penalized likelihood estimation, "Weibull" for the parametric Weibull functions. The default is "Splines".
init.B  Vector of initial values for regression coefficients. This vector should be of the same size as the whole vector of covariates with the first elements for the covariates related to the recurrent events, then to the terminal event and then to the biomarker (interactions in the end of each component). Default is 0.5 for each.

init.Random  Initial value for variance of the elements of the matrix of the distribution of the random effects.

init.Eta  Initial values for regression coefficients for the link functions, first for the recurrent events ($\eta_R$) and for the terminal event ($\eta_T$).


n.nodes  Number of nodes for the Gauss-Hermite quadrature. They can be chosen among 5, 7, 9, 12, 15, 20 and 32. The default is 9.

LIMparam  Convergence threshold of the Marquardt algorithm for the parameters (see Details), $10^{-3}$ by default.

LIMlogl  Convergence threshold of the Marquardt algorithm for the log-likelihood (see Details), $10^{-3}$ by default.

LIMderiv  Convergence threshold of the Marquardt algorithm for the gradient (see Details), $10^{-3}$ by default.

print.times  a logical parameter to print iteration process. Default is TRUE.

**Details**

Typical usage for the joint model

```r
trivPenal(Surv(time,event)~cluster(id) + var1 + var2 +
terminal(death), formula.terminalEvent =~ var1 + var3, biomarker ~
var1+var2, data, data.Longi, ...)
```

The method of the Gauss-Hermite quadrature for approximations of the multidimensional integrals, i.e. length of random is 2, can be chosen among the standard, non-adaptive, pseudo-adaptive in which the quadrature points are transformed using the information from the fitted mixed-effects model for the biomarker (Rizopoulos 2012) or multivariate non-adaptive procedure proposed by Genz et al. 1996 and implemented in FORTRAN subroutine HRMSYM. The choice of the method is important for estimations. The standard non-adaptive Gauss-Hermite quadrature ("Standard") with a specific number of points gives accurate results but can be time consuming. The non-adaptive procedure ("HRMSYM") offers advantageous computational time but in case of datasets in which some individuals have few repeated observations (biomarker measures or recurrent events), this method may be moderately unstable. The pseudo-adaptive quadrature uses transformed quadrature points to center and scale the integrand by utilizing estimates of the random effects from an appropriate linear mixed-effects model (this transformation does not include the frailty in the trivariate model, for which the standard method is used). This method enables using less quadrature points while preserving the estimation accuracy and thus lead to a better computational time.

**NOTE.** Data frames `data` and `data.Longi` must be consistent. Names and types of corresponding covariates must be the same, as well as the number and identification of individuals.
The following components are included in a 'trivPenal' object for each model:

- **b**: The sequence of the corresponding estimation of the coefficients for the hazard functions (parametric or semiparametric), the random effects variances and the regression coefficients.
- **call**: The code used for the model.
- **formula**: The formula part of the code used for the recurrent event part of the model.
- **formula.terminalEvent**: The formula part of the code used for the terminal event part of the model.
- **formula.LongitudinalData**: The formula part of the code used for the longitudinal part of the model.
- **coef**: The regression coefficients (first for the recurrent events, then for the terminal event and then for the biomarker).
- **groups**: The number of groups used in the fit.
- **kappa**: The values of the smoothing parameters in the penalized likelihood estimation corresponding to the baseline hazard functions for the recurrent and terminal events.
- **logLikPenal**: The complete marginal penalized log-likelihood in the semiparametric case.
- **logLik**: The marginal log-likelihood in the parametric case.
- **n.measurements**: The number of biomarker observations used in the fit.
- **max_rep**: The maximal number of repeated measurements per individual.
- **n**: The number of observations in 'data' (recurrent and terminal events) used in the fit.
- **n.events**: The number of recurrent events observed in the fit.
- **n.deaths**: The number of terminal events observed in the fit.
- **n.iter**: The number of iterations needed to converge.
- **n.knots**: The number of knots for estimating the baseline hazard function in the penalized likelihood estimation.
- **n.strat**: The number of stratum.
- **varH**: The variance matrix of all parameters (before positivity constraint transformation for the variance of the measurement error, for which the delta method is used).
- **varHIH**: The robust estimation of the variance matrix of all parameters.
- **xR**: The vector of times where both survival and hazard function of the recurrent events are estimated. By default seq(0,max(time),length=99), where time is the vector of survival times.
- **lamR**: The array (dim=3) of baseline hazard estimates and confidence bands (recurrent events).
- **survR**: The array (dim=3) of baseline survival estimates and confidence bands (recurrent events).
The vector of times where both survival and hazard function of the terminal event are estimated. By default seq(0,max(time),length=99), where time is the vector of survival times.

The array (dim=3) of baseline hazard estimates and confidence bands.

The array (dim=3) of baseline survival estimates and confidence bands.

The value of the median survival and its confidence bands for the recurrent event.

The value of the median survival and its confidence bands for the terminal event.

The type of the baseline hazard function (0:"Splines", "2:Weibull").

The number of parameters.

The vector of number of explanatory variables for the recurrent events, terminal event and biomarker.

The number of explanatory variables for the recurrent events.

The number of explanatory variables for the terminal event.

The number of explanatory variables for the biomarker.

The indicator of absence of the explanatory variables for the recurrent events.

The indicator of absence of the explanatory variables for the terminal event.

The indicator of absence of the explanatory variables for the biomarker.

The approximated likelihood cross-validation criterion in the semiparametric case (with H minus the converged Hessian matrix, and l(.) the full log-likelihood).

$$LCV = \frac{1}{n}(trace(H^{-1}H) - l(.))$$

The Akaike information Criterion for the parametric case.

$$AIC = \frac{1}{n}(np - l(.))$$

The initial value for the number of knots.

The shape parameter for the Weibull hazard functions (the first element for the recurrences and the second one for the terminal event).

The scale parameter for the Weibull hazard functions (the first element for the recurrences and the second one for the terminal event).

The martingale residuals related to the recurrences for each individual.

The martingale residuals related to the terminal event for each individual.

The conditional residuals for the biomarker (subject-specific): \( R_{i}^{(m)} = y_i - X_i^{L_i} \hat{\beta}_L - Z_i^{L_i} \hat{b}_i \).

The marginal residuals for the biomarker (population averaged): \( R_{i}^{(c)} = y_i - X_i^{L_i} \hat{\beta}_L \).
marginal_chol.res

The Cholesky marginal residuals for the biomarker: $R_i^{(m)} = \hat{U}_i^{(m)} R_i^{(m)}$, where $\hat{U}_i^{(m)}$ is an upper-triangular matrix obtained by the Cholesky decomposition of the variance matrix $V_{R_i^{(m)}} = \hat{V}_i - X_{Li} \left( \sum_{i=1}^{N} X_{Li} \hat{V}_i^{-1} X_{Li} \right)^{-1} X_{Li}^T$.

conditional_st.res

The standardized conditional residuals for the biomarker.

marginal_st.res

The standardized marginal residuals for the biomarker.

random.effects.pred

The empirical Bayes predictions of the random effects (ie. using conditional posterior distributions).

frailty.pred

The empirical Bayes predictions of the frailty term (ie. using conditional posterior distributions).

pred.y.marg

The marginal predictions of the longitudinal outcome.

pred.y.cond

The conditional (given the random effects) predictions of the longitudinal outcome.

linear.pred

The linear predictor for the recurrent events part.

lineardeath.pred

The linear predictor for the terminal event part.

global_chisqR

The vector with values of each multivariate Wald test for the recurrent part.

dof_chisqR

The vector with degrees of freedom for each multivariate Wald test for the recurrent part.

global_chisq.testR

The binary variable equals to 0 when no multivariate Wald is given, 1 otherwise (for the recurrent part).

p.global_chisqR

The vector with the p_values for each global multivariate Wald test for the recurrent part.

global_chisqT

The vector with values of each multivariate Wald test for the terminal part.

dof_chisqT

The vector with degrees of freedom for each multivariate Wald test for the terminal part.

global_chisq.testT

The binary variable equals to 0 when no multivariate Wald is given, 1 otherwise (for the terminal part).

p.global_chisqT

The vector with the p_values for each global multivariate Wald test for the terminal part.

global_chisqY

The vector with values of each multivariate Wald test for the longitudinal part.

dof_chisqY

The vector with degrees of freedom for each multivariate Wald test for the longitudinal part.

global_chisq.testY

The binary variable equals to 0 when no multivariate Wald is given, 1 otherwise (for the longitudinal part).
p.global_chisqY  
The vector with the p_values for each global multivariate Wald test for the longitudinal part.

names.factorR  
The names of the "as.factor" variables for the recurrent part.

names.factorT  
The names of the "as.factor" variables for the terminal part.

names.factorY  
The names of the "as.factor" variables for the longitudinal part.

AG  
The logical value. Is Andersen-Gill model fitted?

intercept  
The logical value. Is the fixed intercept included in the linear mixed-effects model?

B1  
The variance matrix of the random effects for the longitudinal outcome.

sigma2  
The variance of the frailty term ($\sigma_v$).

alpha  
The coefficient $\alpha$ associated with the frailty parameter in the terminal hazard function.

ResidualSE  
The variance of the measurement error.

etaR  
The regression coefficients for the link function $g(\cdot)$.

etaT  
The regression coefficients for the link function $h(\cdot)$.

ne_re  
The number of random effects $b$ used in the fit.

names.re  
The names of variables for the random effects $b_i$.

link  
The name of the type of the link functions.

leftCensoring  
The logical value. Is the longitudinal outcome left-censored?

leftCensoring.threshold  
For the left-censored biomarker, the value of the left-censoring threshold used for the fit.

prop.censored  
The fraction of observations subjected to the left-censoring.

methodGH  
The Gaussian quadrature method used in the fit.

n.nodes  
The number of nodes used for the Gaussian quadrature in the fit.

alpha_p.value  
p-value of the Wald test for the estimated coefficient $\alpha$.

sigma2_p.value  
p-value of the Wald test for the estimated variance of the frailty term ($\sigma_v$).

etaR_p.value  
p-values of the Wald test for the estimated regression coefficients for the link function $g(\cdot)$. 

etaT_p.value  
p-values of the Wald test for the estimated regression coefficients for the link function $h(\cdot)$.

beta_p.value  
p-values of the Wald test for the estimated regression coefficients.

Note  
It is recommended to initialize the parameter values using the results from the reduced models (for example, longiPenal for the longitudinal and terminal part and frailtyPenal for the recurrent part. See example.
References


See Also

`plot.trivPenal`, `print.trivPenal`, `summary.trivPenal`

Examples

```R
## Not run:
###--- Trivariate joint model for longitudinal data, ---###
###--- recurrent events and a terminal event ---###
data(colorectal)
data(colorectalLongi)

# Parameter initialisation for covariates - longitudinal and terminal part

# Survival data preparation - only terminal events
colorectalSurv <- subset(colorectal, new.lesions == 0)
initial.longi <- longiPenal(Surv(time1, state) ~ age + treatment + who.PS + prev.resection, tumor.size ~ year * treatment + age + who.PS, colorectalSurv, data.Longi = colorectalLongi, random = c("1", "year"), id = "id", link = "Random-effects", left.censoring = -3.33, n.knots = 6, kappa = 2, method.GH="Pseudo-adaptive", maxit=40, n.nodes=7)

# Parameter initialisation for covariates - recurrent part
initial.frailty <- frailtyPenal(Surv(time0, time1, new.lesions) ~ cluster(id) + age + treatment + who.PS, data = colorectal, recurrentAG = TRUE, RandDist = "LogN", n.knots = 6, kappa =2)
```
# Baseline hazard function approximated with splines
# Random effects as the link function, Calendar timescale
# (computation takes around 40 minutes)

model.spli.RE.cal <- trivPenal(Surv(time0, time1, new.lesions) ~ cluster(id) + age + treatment + who.PS + terminal(state),
formula.terminalEvent =~ age + treatment + who.PS + prev.resection,
tumor.size =~ year * treatment + age + who.PS, data = colorectal,
data.Longi = colorectalLongi, random = c("1", "year"), id = "id",
link = "Random-effects", left.censoring = -3.33, recurrentAG = TRUE,
n.knots = 6, kappa=c(0.01, 2), method.GH="Standard", n.nodes = 7,
init.B = c(-0.07, -0.13, -0.16, -0.17, 0.42, # recurrent events covariates
-0.16, -0.14, -0.14, 0.08, 0.86, -0.24, # terminal event covariates
2.93, -0.28, -0.13, 0.17, -0.41, 0.23, 0.97, -0.61)) # biomarker covariates

# Weibull baseline hazard function
# Random effects as the link function, Gap timescale
# (computation takes around 30 minutes)

model.weib.RE.gap <- trivPenal(Surv(gap.time, new.lesions) ~ cluster(id) + age + treatment + who.PS + prev.resection + terminal(state),
formula.terminalEvent =~ age + treatment + who.PS + prev.resection,
tumor.size =~ year * treatment + age + who.PS, data = colorectal,
data.Longi = colorectalLongi, random = c("1", "year"), id = "id",
link = "Random-effects", left.censoring = -3.33, recurrentAG = FALSE,
hazard = "Weibull", method.GH="Pseudo-adaptive", n.nodes=7)

## End(Not run)

**trivPenalNL**

*Fit a Non-Linear Trivariate Joint Model for Recurrent Events and a Terminal Event with a Biomarker Described with an ODE Population Model*

**Description**

Fit a non-linear trivariate joint model for a longitudinal biomarker, recurrent events and a terminal event using a semiparametric penalized likelihood estimation or a parametric estimation on the hazard functions.

The values $y_i(t)$ ($i = 1, \ldots, N$) for $N$ subjects represent the individual evolution of the biomarker e.g. tumor size expressed as the sum of the longest diameters (SLD) of target lesions. The dynamics of the biomarker are described by an ordinary differential equation (ODE) that includes the effect of the natural net growth and the treatment effect:
The model includes the following parameters (using the interpretation of tumor dynamics): \( \exp(K_{G,0}) \) the constant tumor growth rate, \( \exp(K_{D,0}) \) the drug-induced tumor decline rate, \( \lambda \) resistance effect to drug (exponential tumor decay change with time), \( \exp(y_0) \) the initial level of the biomarker and \( d_i \) is the treatment concentration (e.g. dose). The random effects \( b_i^T = (b_{y_0,i}, b_{G,i}, b_{D,i}, b_{\lambda,i})^T \) are gaussian variables with a diagonal covariance matrix \( B_i \). In the trivariate model we use the analytical solution of the equation with the population-based approach of the non-linear mixed effects model. We can also assume a transformation for the observations of the biomarker (one parametrical solution of the equation with the population-based approach of the non-linear mixed effects model).

The risks of the recurrent \( r_{ij}(\cdot) \) the risk of the \( j \)-th event of the individual \( i \) and terminal events \( \lambda_i \) the risk of the event of the individual \( i \) are represented by proportional hazard risk models. The joint model is constructed assuming that the processes are linked via a latent structure and includes the non-linear mixed effects model for the longitudinal data:

\[
\begin{align*}
g(t_{ik}) &= \exp[y_0 + b_{y_0,i} + t_{ik} \times \exp(K_{G,0} + b_{G,i} + X_{G,i}(t)^T \beta_G) \\
&+ d_i \times \exp(K_{D,0} + b_{D,i} - \lambda - b_{\lambda,i} + X_{D,i}(t)^T \beta_D) \\
&\times (\exp(-\exp(\lambda + b_{\lambda,i}t_{ik}) - 1)) + \epsilon_{ik} \\
r_{ij}(t|b_i) &= r_0(t) \exp(v_i + X_{R,j}(t)^T \beta_R + g(y_i(t))^T \eta_R) \\
\lambda_i(t|b_i) &= \lambda_0(t) \exp(\alpha v_i + X_{T,i}(t)^T \beta_T + h(y_i(t))^T \eta_T)
\end{align*}
\]

where \( X_{G,i}(t) \), \( X_{D,i}(t) \), \( X_{R,j}(t) \) and \( X_{T,i}(t) \) are vectors of possible time-varying fixed effects covariates and \( \beta_G, \beta_D, \beta_R \) and \( \beta_T \) are the associated coefficients. The random effects \( b_i \) are independent from the measurement error. The relationship between the biomarker and recurrent events is explained via \( g(y_i(t)) \) with coefficients \( \eta_R \) and between the biomarker and terminal event is explained via \( h(y_i(t)) \) with coefficients \( \eta_T \). Currently, only one form of the functions \( g(\cdot) \) and \( h(\cdot) \) is available: the random effects \( b_i \). The frailty term \( v_i \) is gaussian with mean 0 and variance \( \sigma_v \). Together with \( b_i \) constitutes the random effects of the model:

\[
u_i = \begin{pmatrix} b_i \\ v_i \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ B_1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \sigma_v^2 \end{pmatrix} \right)
\]

Any combination of the random effects \( b_i \), e.g. \( b_i = b_{y_0,i} \) or \( b_i = \{b_{G,i}, b_{D,i}, b_{\lambda,i}\} \) can be chosen for the model.

We consider that the longitudinal outcome can be a subject to a quantification limit, i.e. some observations, below a level of detection \( s \) cannot be quantified (left-censoring).
Arguments

- **formula**: 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 like in survival package. Interactions are possible using * or :.

- **formula.terminalEvent**: A formula object, only requires terms on the right to indicate which variables are modelling the terminal event. Interactions are possible using * or :.

- **biomarker**: Name of the variable representing the longitudinal biomarker.

- **formula.KG**: A formula object, only requires terms on the right to indicate which covariates related to the biomarker growth are included in the longitudinal sub-model. It must follow the standard form used for linear mixed-effects models. Interactions are possible using * or :.

- **formula.KD**: A formula object, only requires terms on the right to indicate which covariates related to the biomarker drug-induced decline are included in the longitudinal sub-model. It must follow the standard form used for linear mixed-effects models. Interactions are possible using * or :.

- **dose**: Name of the variable representing the drug concentration indicator.

- **time.biomarker**: Name of the variable of times of biomarker measurements.

- **data**: A 'data.frame' with the variables used in formula.

- **data.Longi**: A 'data.frame' with the variables used in formula.KG, formula.KD, biomarker, dose, time.biomarker and id.

- **random**: Names of parameters for which the random effects are included in the mixed model. The names must be chosen among "y0", "KG", "KD" and "lambda". Any combination of the mentioned names is allowed.

- **id**: Name of the variable representing the individuals.

- **link**: Type of link functions for the dependence between the biomarker and death and between the biomarker and the recurrent events: only "Random-effects" for the association directly via the random effects of the biomarker is allowed for the moment (option for a future extension).

- **BoxCox**: Should the Box-Cox transformation be used for the longitudinal biomarker? If there is no transformation, the argument must be equal to FALSE, otherwise the value of the transformation parameter must be given, then the transformed values are $y^\xi = (y^\xi - 1)/\xi$, where $\xi$ is the Box-Cox parameter.

- **left.censoring**: Is the biomarker left-censored below a threshold s? If there is no left-censoring, the argument must be equal to FALSE, otherwise the value of the threshold must be given.

- **recurrentAG**: Logical value. Is Andersen-Gill model fitted? If so indicates that recurrent event times with the counting process approach of Andersen and Gill is used. This formulation can be used for dealing with time-dependent covariates. The default is FALSE.

- **n.knots**: Integer giving the number of knots to use. Value required in the penalized likelihood estimation. It corresponds to the (n.knots+2) splines functions for the approximation of the hazard or the survival functions. We estimate I or M-splines.
of order 4. When the user set a number of knots equals to k (n.knots=k) then the number of interior knots is (k-2) and the number of splines is (k-2)+order. Number of knots must be between 4 and 20. (See Note in frailtyPenal function)

**kappa**
Positive smoothing parameters in the penalized likelihood estimation. The coefficient kappa of the integral of the squared second derivative of hazard function in the fit (penalized log likelihood). To obtain an initial value for kappa, a solution is to fit the corresponding Cox model using cross validation (See cross.validation in function frailtyPenal). We advise the user to identify several possible tuning parameters, note their defaults and look at the sensitivity of the results to varying them.

**maxit**
Maximum number of iterations for the Marquardt algorithm. Default is 300

**hazard**
Type of hazard functions: "Splines" for semiparametric hazard functions using equidistant intervals or "Splines-per" using percentile with the penalized likelihood estimation, "Weibull" for the parametric Weibull functions. The default is "Splines".

**init.B**
Vector of initial values for regression coefficients. This vector should be of the same size as the whole vector of covariates with the first elements for the covariates related to the recurrent events, then to the terminal event and then to the biomarker (interactions in the end of each component). Default is 0.5 for each.

**init.Random**
Initial value for variance of the elements of the matrix of the distribution of the random effects.

**init.Eta**
Initial values for regression coefficients for the link functions, first for the recurrent events ($\eta_R$) and for the terminal event ($\eta_T$).

**init.Alpha**
Initial value for parameter alpha

**init.Biomarker**
Initial values for biomarker parameters: $y_0, K_{G,0}, K_{D,0}$ and $\lambda$ (using this order).

**method.GH**
Method for the Gauss-Hermite quadrature: "Standard" for the standard non-adaptive Gaussian quadrature and "Pseudo-adaptive" for the pseudo-adaptive Gaussian quadrature (see Details). The default is "Standard". When the option "Pseudo-adaptive" is chosen, then a univariate model (non-linear mixed model for the biomarker) is fitted in order to obtain the estimations of the random effects $b_i$.

**init.GH**
Only when the option "Pseudo-adaptive" of the argument method.GH is chosen. If TRUE, the estimations of the biomarker parameters ($y_0, K_{G,0}, K_{D,0}$ and $\lambda$), $\sigma_s, \beta_G$ and $\beta_D$ from the univariate mixed model are used as the initial values of the parameters related to the biomarker.

**n.nodes**
Number of nodes for the Gauss-Hermite quadrature (from 5 to 32). The default is 9.

**LIMparam**
Convergence threshold of the Marquardt algorithm for the parameters (see Details), $10^{-3}$ by default.

**LIMlogl**
Convergence threshold of the Marquardt algorithm for the log-likelihood (see Details), $10^{-3}$ by default.

**LIMderiv**
Convergence threshold of the Marquardt algorithm for the gradient (see Details), $10^{-3}$ by default.

**print.times**
a logical parameter to print iteration process. Default is TRUE.
Details

Typical usage for the joint model

```
trivPenalNL(Surv(time,event)~cluster(id) + var1 + var2 +
terminal(death), formula.terminalEvent =~ var1 + var3, biomarker =
"biomarker.name", dose = "dose.name", time.biomarker = "time", formula.KG ~
var1, formula.KD ~ var2, data, data.Longi, ...)
```

The method of the Gauss-Hermite quadrature for approximations of the multidimensional integrals, i.e. length of random more than 2, can be chosen among the standard (non-adaptive) and pseudo-adaptive in which the quadrature points are transformed using the information from the fitted mixed-effects model for the biomarker (Rizopoulos 2012) or multivariate non-adaptive procedure proposed by Genz et al. 1996 and implemented in FORTRAN subroutine HRMSYM. The choice of the method is important for estimations. The standard non-adaptive Gauss-Hermite quadrature ("Standard") with a specific number of points gives accurate results but can be time consuming. The pseudo-adaptive quadrature uses transformed quadrature points to center and scale the integrand by utilizing estimates of the random effects from an appropriate non-linear mixed-effects model (this transformation does not include the frailty in the trivariate model, for which the standard method, with 20 quadrature points, is used). This method enables using less quadrature points while preserving the estimation accuracy and thus lead to a better computational time.

NOTE. Data frames data and data.Longi must be consistent. Names and types of corresponding covariates must be the same, as well as the number and identification of individuals.

Value

The following components are included in a 'trivPenalNL' object for each model:

- **b**: The sequence of the corresponding estimation of the coefficients for the hazard functions (parametric or semiparametric), the random effects variances and the regression coefficients.
- **call**: The code used for the model.
- **formula**: The formula part of the code used for the recurrent event part of the model.
- **formula.terminalEvent**: The formula part of the code used for the terminal event part of the model.
- **formula.KG**: The formula part of the code used for the longitudinal part of the model, for the biomarker growth dynamics.
- **formula.KD**: The formula part of the code used for the longitudinal part of the model, for the biomarker decline dynamics.
- **coef**: The regression coefficients (first for the recurrent events, then for the terminal event, then for the biomarker growth and then for the biomarker decline.
- **groups**: The number of groups used in the fit.
- **kappa**: The values of the smoothing parameters in the penalized likelihood estimation corresponding to the baseline hazard functions for the recurrent and terminal events.
- **logLikPenal**: The complete marginal penalized log-likelihood in the semiparametric case.
logLik
n.measurements
max_rep
n
n.events
n.deaths
n.iter
n.knots
n.strat
varH
varHIIH
xR
lamR
survR
xD
lamD
survD
medianR
medianD
typeof
npar
nvar
nvarRec
nvarEnd
nvarKG
nvarKD
noVarRec
noVarEnd
noVarKG

The marginal log-likelihood in the parametric case.
The number of biomarker observations used in the fit.
The maximal number of repeated measurements per individual.
The number of observations in 'data' (recurrent and terminal events) used in the fit.
The number of recurrent events observed in the fit.
The number of terminal events observed in the fit.
The number of iterations needed to converge.
The number of knots for estimating the baseline hazard function in the penalized likelihood estimation.
The number of stratum.
The variance matrix of all parameters (before positivity constraint transformation for the variance of the measurement error, for which the delta method is used).
The robust estimation of the variance matrix of all parameters.
The vector of times where both survival and hazard function of the recurrent events are estimated. By default seq(0,max(time),length=99), where time is the vector of survival times.
The array (dim=3) of baseline hazard estimates and confidence bands (recurrent events).
The array (dim=3) of baseline survival estimates and confidence bands (recurrent events).
The vector of times where both survival and hazard function of the terminal event are estimated. By default seq(0,max(time),length=99), where time is the vector of survival times.
The array (dim=3) of baseline hazard estimates and confidence bands.
The array (dim=3) of baseline survival estimates and confidence bands.
The value of the median survival and its confidence bands for the recurrent event.
The value of the median survival and its confidence bands for the terminal event.
The type of the baseline hazard function (0: "Splines", 2: Weibull").
The number of parameters.
The vector of number of explanatory variables for the recurrent events, terminal event, biomarker growth and biomarker decline.
The number of explanatory variables for the recurrent events.
The number of explanatory variables for the terminal event.
The number of explanatory variables for the biomarker growth.
The number of explanatory variables for the biomarker decline.
The indicator of absence of the explanatory variables for the recurrent events.
The indicator of absence of the explanatory variables for the terminal event.
The indicator of absence of the explanatory variables for the biomarker growth.
noVarKD  The indicator of absence of the explanatory variables for the biomarker decline.

LCV  The approximated likelihood cross-validation criterion in the semiparametric case (with $H$ minus the converged Hessian matrix, and $l(.)$ the full log-likelihood).

$$LCV = \frac{1}{n}(\text{trace}(H^{-1}_{pl}H) - l(.))$$

AIC  The Akaike information Criterion for the parametric case.

$$AIC = \frac{1}{n}(np - l(.))$$

n.knots.temp  The initial value for the number of knots.

shape.weib  The shape parameter for the Weibull hazard functions (the first element for the recurrences and the second one for the terminal event).

scale.weib  The scale parameter for the Weibull hazard functions (the first element for the recurrences and the second one for the terminal event).

random.effects.pred  The empirical Bayes predictions of the random effects (ie. using conditional posterior distributions).

global_chisq.testR  The binary variable equals to 0 when no multivariate Wald is given, 1 otherwise (for the recurrent part).

global_chisq.testT  The binary variable equals to 0 when no multivariate Wald is given, 1 otherwise (for the terminal part).

global_chisq.testKG  The binary variable equals to 0 when no multivariate Wald is given, 1 otherwise (for the biomarker growth).

global_chisq.testKD  The binary variable equals to 0 when no multivariate Wald is given, 1 otherwise (for the biomarker decline).

AG  The logical value. Is Andersen-Gill model fitted?

B1  The variance matrix of the random effects for the longitudinal outcome.

sigma2  The variance of the frailty term ($\sigma_v$).

alpha  The coefficient $\alpha$ associated with the frailty parameter in the terminal hazard function.

ResidualSE  The variance of the measurement error.

etaR  The regression coefficients for the link function $g(\cdot)$.

etaT  The regression coefficients for the link function $h(\cdot)$.

ne_re  The number of random effects $b$ used in the fit.

names.re  The names of variables for the random effects $b_i$.

link  The name of the type of the link functions.

leftCensoring  The logical value. Is the longitudinal outcome left-censored?
leftCensoring.threshold
For the left-censored biomarker, the value of the left-censoring threshold used for the fit.

prop.censored
The fraction of observations subjected to the left-censoring.

methodGH
The Gaussian quadrature method used in the fit.

n.nodes
The number of nodes used for the Gaussian quadrature in the fit.

K_G0
Value of the estimate of the biomarker growth parameter.

K_D0
Value of the estimate of the biomarker decay parameter.

lambda
Value of the estimate of the biomarker resistance to drug.

y_0
Value of the estimate of the biomarker initial level.

biomarker
Name of the variable associated with the biomarker in the data.

time.biomarker
Name of the variable associated with the time of measurements of the biomarker in the data.

dose
Name of the variable associated with the drug concentration in the data.

BoxCox
The logical value. Is the BoxCox transformation applied for the biomarker?

BoxCox_parameter
The value of the BoxCox transformation parameter.

alpha_p.value
p-value of the Wald test for the estimated coefficient $\alpha$.

sigma2_p.value
p-value of the Wald test for the estimated variance of the frailty term ($\sigma_v$).

etaR_p.value
p-values of the Wald test for the estimated regression coefficients for the link function $g(\cdot)$.

etaT_p.value
p-values of the Wald test for the estimated regression coefficients for the link function $h(\cdot)$.

y_0_p.value
p-value of the Wald test for the estimated biomarker initial level.

K_G0_p.value
p-value of the Wald test for the estimated biomarker growth parameter.

K_D0_p.value
p-value of the Wald test for the estimated biomarker decay parameter.

lambda_p.value
p-value of the Wald test for the estimated biomarker resistance to drug.

beta_p.value
p-values of the Wald test for the estimated regression coefficients.

Note
It is recommended to initialize the parameter values using the results from a corresponding reduced model (frailtyPenal for the recurrent and terminal part). See example.

Estimations of models with more than three random effects can be very long.

References


See Also

plot.trivPenalNL, print.trivPenalNL, summary.trivPenalNL

Examples

```r
## Not run:

###--- Non-linear trivariate joint model for longitudinal data, ---###
###--- recurrent events and a terminal event ---###

data(colorectal)
data(colorectalLongi)

# No information on dose - creation of a dummy variable
colorectalLongi$dose <- 1

# Parameters initialisation - estimation of a simplified model
# with two random effects (a frailty term and a random effect
# related to biomarker growth (KG))
initial.model <- trivPenalNL(Surv(time0, time1, new.lesions) ~ cluster(id) + age + treatment + terminal(state), formula.terminalEvent =~ age + treatment, biomarker = "tumor.size", formula.KG ~ 1, formula.KD ~ treatment, dose = "dose", time.biomarker = "year", data = colorectal, data.Longi = colorectalLongi, random = "KG", id = "id", recurrentAG = TRUE, n.knots = 5, kappa = c(0.01, 2), method.GH = "Pseudo-adaptive")

# Trivariate joint model with initial values for parameters
# (computation takes around 40 minutes)
model <- trivPenalNL(Surv(time0, time1, new.lesions) ~ cluster(id) + age + treatment + terminal(state), formula.terminalEvent =~ age + treatment, biomarker = "tumor.size", formula.KG ~ 1, formula.KD ~ treatment, dose = "dose", time.biomarker = "year", data = colorectal, data.Longi = colorectalLongi, random = c("y0", "KG"), id = "id", init.B = c(-0.22, -0.16, -0.35, -0.19, 0.04, -0.41, 0.23), init.Alpha = 1.86, init.Eta = c(0.5, 0.57, 0.5, 2.34), init.Biomarker = c(1.24, 0.81, 1.07, -1.53), recurrentAG = TRUE, n.knots = 5, kappa = c(0.01, 2), method.GH = "Pseudo-adaptive")
```
Identify weights

Description

This is a special function used in the context of the joint frailty models for data from nested case-control studies. It specifies weights defined by using `wts` function, and is used of `frailtyPenal` formula for fitting joint models.

Usage

wts(x)

Arguments

x A numeric variable which is supposed to indicate the weights

Value

x A variable identified as weights

See Also

frailtyPenal

Examples

## Not run:

data(dataNCC)
modJoint.ncc <- frailtyPenal(Surv(t.start,t.stop,event)~cluster(id)+cov1 +cov2+terminal(death)+wts(ncc.wts), formula.terminalEvent=~cov1+cov2,
data=dataNCC,n.knots=8,kappa=c(1.6e+10, 5.0e+03),recurrentAG=TRUE, RandDist="LogN")

print(modJoint.ncc)

## End(Not run)
Index

*Topic Surrogate
  ste, 129
*Topic concordance
  Cmeasures, 14
*Topic datasets
  bcos, 12
  colorectal, 17
  colorectallongi, 18
  dataAdditive, 19
  dataMultiv, 20
  dataNCC, 20
  dataNested, 21
  dataOvarian, 22
  gastadj, 44
  readmission, 126
*Topic effect
  ste, 129
*Topic file
  additivePenal, 8
  plot.Diffepoce, 84
  plot.epoce, 85
  plot.frailtyPenal, 86
  plot.longiPenal, 92
  plot.predFrailty, 97
  plot.predJoint, 97
  plot.predLongi, 98
*Topic loocv
  loocv, 73
*Topic methods
  multivPenal, 75
  plot.additivePenal, 83
  plot.jointNestedPenal, 87
  plot.jointPenal, 89
  plot.multivPenal, 94
  plot.nestedPenal, 95
  plot.trivPenal, 99
  plot.trivPenalNL, 100
  print.additivePenal, 118
  print.Cmeasures, 119
  print.frailtyPenal, 119
  print.jointNestedPenal, 120
  print.jointPenal, 121
  print.longiPenal, 121
  print.multivPenal, 122
  print.nestedPenal, 123
  print.prediction, 124
  print.trivPenal, 124
  print.trivPenalNL, 125
  summary.additivePenal, 132
  summary.frailtyPenal, 133
  summary.jointNestedPenal, 134
  summary.jointPenal, 135
  summary.jointSurroPenal, 137
  summary.jointSurroPenalSimul, 138
  summary.longiPenal, 140
  summary.multivPenal, 141
  summary.nestedPenal, 142
  summary.trivPenal, 143
  summary.trivPenalNL, 144
*Topic misc
  cluster, 13
  Diffeopece, 23
  epoce, 25
  event2, 29
  num.id, 81
  prediction, 104
  slope, 127
  subcluster, 131
  terminal, 148
  timedep, 149
  wts, 168
*Topic models
  frailtyPenal, 30
  longiPenal, 66
  multivPenal, 75
  trivPenal, 151
  trivPenalNL, 159
*Topic multiv
multivPenal, 75
print.multivPenal, 122
summary.multivPenal, 141

*Topic package
  frailtypack-package, 4

*Topic prediction
  looczv, 73
predict.jointSurroPenal, 102
ste, 129

*Topic ste
  ste, 129

*Topic surrogate
  looczv, 73
plot.jointSurroPenal, 91
predict.jointSurroPenal, 102
ste, 129

*Topic threshold
  ste, 129

additivePenal, 8, 83, 119, 128, 132

bcos, 12

CbootstrapFP (Cmeasures), 14
cindexes (Cmeasures), 14
cluster, 13, 41
Cmeasures, 14, 119
colorectal, 17
colorectallongi, 18
data.frame, 48, 64
dataAdditive, 19
dataMultiv, 20
dataNCC, 20
dataNested, 21
dataOvarian, 22
diffepoce, 23, 84
epoce, 25, 85
event2, 29, 81

factor.names (frailtyPenal), 30
for (multivPenal), 75
frailty (multivPenal), 75
frailtypack (frailtypack-package), 4
frailtypack-package, 4
frailtyPenal, 14, 16, 30, 49, 55, 56, 82, 86, 88, 90, 96, 120, 121, 123, 131, 133, 135, 136, 142, 149, 168
gastadj, 44
hazard, 45
jointSurroPenal, 46, 54, 60, 62, 64, 74, 92, 102, 103, 130, 138
jointSurroPenalSimul, 53, 54, 139
jointSurroTKendall, 60, 138
jointSurrSimul, 53, 60, 62, 62, 65
lines.additivePenal
  (plot.additivePenal), 83
lines.frailtyPenal (plot.frailtyPenal), 86
lines.jointNestedPenal
  (plot.jointNestedPenal), 87
lines.jointPenal (plot.jointPenal), 89
lines.jointSurroPenal
  (plot.jointSurroPenal), 91
lines.longiPenal (plot.longiPenal), 92
lines.nestedPenal (plot.nestedPenal), 95
lines.trivPenal (plot.trivPenal), 99
lines.trivPenalNL (plot.trivPenalNL), 100
longDat, 65
longiPenal, 66, 93, 122, 140
looczv, 73

model, 129
model (multivPenal), 75
multivariate (multivPenal), 75
multivPenal, 29, 75, 95, 122, 142

num.id, 41, 81

plot.additivePenal, 83
plot.Diffepoce, 84
plot.epoce, 85
plot.frailtyPenal, 83, 86, 96
plot.jointNestedPenal, 87
plot.jointPenal, 89
plot.jointSurroPenal, 91
plot.longiPenal, 72, 92
plot.multivPenal, 81, 94
plot.nestedPenal, 95
plot.predFrailty, 97
plot.predJoint, 97
plot.predLongi, 98
plot.trivPenal, 99, 158
plot.trivPenalNL, 100, 167
predict, 130
predict.jointSurroPenal, 102
prediction, 104, 124
print.additivePenal, 118
print.Cmeasures, 16, 119
print.frailtyPenal, 119
print.jointNestedPenal, 120
print.jointPenal, 121
print.longiPenal, 72, 121
print.multivPenal, 81, 122
print.nestedPenal, 123
print.predFrailty (print.prediction), 124
print.prediction, 124
print.predJoint (print.prediction), 124
print.predLongi (print.prediction), 124
print.summary.additivePenal
  (summary.additivePenal), 132
print.summary.frailtyPenal
  (summary.frailtyPenal), 133
print.summary.jointNestedPenal
  (summary.jointNestedPenal), 134
print.summary.jointPenal
  (summary.jointPenal), 135
print.summary.jointSurroPenal
  (summary.jointSurroPenal), 137
print.summary.jointSurroPenalSimul
  (summary.jointSurroPenalSimul), 138
print.summary.longiPenal
  (summary.longiPenal), 140
print.summary.nestedPenal
  (summary.nestedPenal), 142
print.summary.trivPenal
  (summary.trivPenal), 143
print.summary.trivPenalNL
  (summary.trivPenalNL), 144
print.trivPenal, 124, 158
print.trivPenalNL, 125, 167
readmission, 126
runShiny, 127
slope, 12, 127
statFP (Cmeasures), 14
STE, 137
ste, 129
subcluster, 41, 131
summary.additivePenal, 132
summary.frailtyPenal, 133
summary.jointNestedPenal, 134
summary.jointPenal, 135
summary.jointSurroPenal, 53, 62, 137
summary.jointSurroPenalSimul, 60, 138
summary.longiPenal, 72, 140
summary.multivPenal, 81, 141
summary.nestedPenal, 142
summary.trivPenal, 143, 158
summary.trivPenalNL, 144, 167
SurvIC, 41, 146
survival, 147
terminal, 41, 81, 148
timedep, 41, 149
timedep.names (frailtyPenal), 30
transfo.table (multivPenal), 75
trivPenal, 100, 125, 144, 151
trivPenalNL, 101, 125, 145, 159
waldtest (frailtyPenal), 30
wts, 168