Joint models for longitudinal and time-to-event data with shared random effects with jmsrepy library

1 Preamble

jmsrepy library is a partial adaptation of Dimitris Rizopoulos's JM R package. The aim of this library isn't today to be an exhaustive adaptation of JM R package, but to respond needs of our specific applied clinical project. In consequence, the library reproduce today only a part of what R package do. It will be improved in future with some functions which already exists in R package and also some features not handled by R package.

jmsrepy estimate parameters of a joint model for longitudinal and time-to-event data with shared random effects from a mixed linear model and a proportional risk model. jmsrepy also compute probability to survive after a future time.

2 Mathematical formulation

Joint models for longitudinal and time-to-event data with shared random effects here abbreviated **joint models** are used to simultaneously model instantaneous hazard and a longitudinal marker taking into account the correlation between longitudinal marker and instantaneous hazard. In practice we use joint models to estimate an instantaneous hazard using both baseline and longitudinal covariates. Then, joint model combine a proportional risk model which models instantaneous hazard and a mixed linear model which models the longitudinal marker. Instantaneous hazard and longitudinal process are linked by a function in instantaneous hazard part of joint model. This function can have many formulations, but it only depends on mixed linear model's random effects.

If we have N groups in our mixed linear model, the joint model is formulated as following for the i^{th} group at time t :

$$\begin{cases} h_i(t) = h_0(t)\exp(\gamma W_i + g(m_i(t-u), \alpha, b_i)) \\ Y_i(t) = m_i(t) + \epsilon_i(t) = X_i\beta + Z_i b_i + \epsilon_i(t) \\ b_i \sim N(0, D), \epsilon_i(t) \sim N(0, \sigma^2) \end{cases}$$

Where the different terms denotes :

- X_i : Design matrix of covariates associated with mixed linear model fixed effects to modeling longitudinal marker
- β : Vector of estimated linear model fixed effects to modeling longitudinal marker
- Z_i : Design matrix of covariates associated with mixed linear model random effects to modeling longitudinal marker (Z_i is a slice of X_i)
- b_i : Vector of estimated linear model random effects of the i^{th} group to modeling longitudinal marker
- ϵ_i : Error vector
- $h_0(t)$: Baseline risk at time t
- W_i : Baseline covariates of group i
- γ : Vector of regression coefficients associated with baseline covariates
- $g()$: Link function between instantaneous hazard and longitudinal process
- $m_i(t)$: Estimation of longitudinal marker at time t
- u : Delay
- α : Vector of coefficients quantifying effect of longitudinal marker on $h_i(t)$

As mentioned earlier, g() function can takes many formulations, see few examples bellow :

- $h_i(t)$ depends on current value of longitudinal marker:

$$g(m_i(t-u), \alpha, b_i) = \alpha m_i(t)$$

- $h_i(t)$ depends on evolution of longitudinal marker:

$$g(m_i(t-u), \alpha, b_i) = \alpha m_i'(t)$$

- $h_i(t)$ depends on current value and evolution of longitudinal marker:

$$g(m_i(t-u), \alpha, b_i) = \alpha_1 m_i(t) + \alpha_2 m_i'(t)$$

3 Input arguments

Following arguments are available for instantation of jm objet :

- *lme_object* : Instantiated and not fitted mixed linear model of longitudinal marker object. A *statsmodels.regression.mixed_linear_model.MixedLM* object is expected.
- *lme_object_fitted* : Fitted mixed linear model of longitudinal marker object. A *statsmodels.regression.mixed_linear_model.MixedLMResultsWrapper* object is expected. item *lme_data* : Data used in mixed linear model of longitudinal marker object. In other terms, it corresponds to a pandas DataFrame which contains longitudinal marker and covariates used to model it.

- *lme_formula* : R-type formula of longitudinal marker modeled by fixed effects. Takes this form : *'long_marker ∼ X1 + X2'*.
- *lme_re_formula* : R-type formula of random effects. Takes this form : *'∼ X1'*.
- *surv_object* : Fitted proportional risk model object. A *lifelines.CoxPHFitter* object is expected.
- *surv_data* : Data used to fit proportional risk model object. In other terms, it corresponds to an array-like which contains baseline covariates, time of event or censor and vector which indicates if event appeared (0 : censor, 1 : event).
- *time_var* : Column name in *lme_data* of variable which give time of each longitudinal measurement.
- *parametrization* : form of g() function (function which links longitudinal process and instantaneous hazard) :

parametrization	form of g()
'value'	$g(m_i(t-u), \alpha, b_i) = \alpha m_i(t)$
'slope'	$g(m_i(t-u), \alpha, b_i) = \alpha m_i'(t)$
'both'	$g(m_i(t-u), \alpha, b_i) = \alpha_1 m_i(t) + \alpha_2 m_i'(t)$

- *derivForm* : A dictionary which specifies the derivative form of mixed linear model formula. This argument is required only when *parametrization* is *'slope'* or *'both'*. The dictionary must contains following components :

dict key	associated value
fixed	formula representing the derivative of the fixed-effects part of the linear mixed model with respect to time
ind_fixed	list indicating position in fixed-effects vector of fixed-effects corresponding to the derivative
random	formula representing the derivative of the random-effects part of the linear mixed model with respect to time
ind_random	list indicating position in random-effects vector of random-effects corresponding to the derivative

  For example, if we have *lme_formula = 'long_marker ∼ 1 + X1'* and *lme_re_formula =∼ 1 + X1'*, then we will set *derivForm = dict(fixed='∼ 1', ind_fixed = [1], random='∼ 1', ind_random = [1])*
- *lag* : Delay in years
- *scale_wb* : A numeric scalar denoting a fixed value for the scale parameter of the Weibull hazard. If the value is *None*, parameter is estimated.
- *init* : A dictionary in which user can specify initial values of models parameters. The dictionary can contains following components (they all are optional, user can set all initial vales, none initial values or a part of initial values) :

dict key	associated value
betas	numpy array of linear mixed model fixed-effects parameters
d_vc	numpy array of linear mixed model random-effects variance-covariance matrix
sigma	numpy array of the measurement of error standard deviation for the linear mixed effects model
gammas	numpy array of coefficients associated to baseline covariates in proportional risk model
alpha_value	numpy array of association parameter between current value of longitudinal marker and instantaneous risk
alpha_slope	numpy array of association parameter between evolution of longitudinal marker and instantaneous risk
sigma_t	numpy array of scale parameter for the Weibull baseline risk function

4 Functions

Following functions are available after the object is instantiated

- *fit* : Function which fit the joint model on data
- *Summary* : Function which display the summary about fitted model informations (parameters value, log-likelihood, AIC etc)
- *Survfit* : Computes the probability at time s of surviving over time $s + t$

5 References

Rizopoulos, D. (2012) *Joint Models for Longitudinal and Time-to-Event Data: with Applications in R.* Boca Raton: Chapman and Hall/CRC.