Two Methods to Fit Longitudinal Sub-Model of Joint Model

Rong Li

Shanghai University School of Economics Shanghai China

Rongli_210@163.com

Keywords: Joint model; Longitudinal data; Survival sub-model; Linear mixed effect model; Cox proportional; Hazard model; Machine learning

Abstract. The joint model is a new shared parameter model consisting of longitudinal and survival data. In general, we often use a linear mixed-effects model to estimate the longitudinal sub-model and a Cox proportional hazard model to estimate the surviving sub-model. However, the linear model requires a hypothesis of normal distribution. Consequently, when the assumption is not valid, the estimation result will produce large deviations. In this paper, we use machine learning to estimate the longitudinal model, and the survival model still uses the Cox model. We compare the effects of the linear mixed-effects model and the machine learning method. These two estimation methods are compared by simulated data. The results of the comparison show that the residuals diagnosis outcomes of the machine learning survival sub-model are more in line with the theoretical results, while the residuals of the longitudinal sub-model are more dispersed than the linear model. Additionally, the running time of the machine learning is shorter than that of the linear mixed effect model.

Introduction

In recent years, more and more researchers are interested in joint models consisting of longitudinal data and time-to-event data, especially in the field of statistics and biological diseases. Longitudinal data refers to the repeated measurement of data by a group of individuals in time or space order. Each individual is observed multiple times at different times or under different test conditions, and the obtained data has the characteristics of time series and cross-section data. In many practical studies, participants were followed up repeatedly to obtain longitudinal data. The event time data refers to the time elapsed from a certain starting point to the occurrence of the event. The event time data we usually study is the time of death, also known as Survival data.

There are many approaches on the joint model with longitudinal data and survival data. The most typical one is two-stage method, unfortunately substantial research on this topic indicates that this approach results in biased estimates [12,3,11,15]. In addition, maximum likelihood estimation [2,5] is a common method to estimate parameters in joint model. In terms of the joint model, time cost is a difficult problem to solve. Therefore, in order to reduce the computational cost and improve the efficiency of the model, many researchers have introduced the R software packages. Rizopoulos (2010, 2014) introduced two software packages, JM and JMbayes, respectively based on the expected maximum and Bayesian methods, which plays an active role in promoting the R packages used to fit the joint model.

The machine learning methods have the unique advantage when fitting longitudinal sub-model, namely it does not need to consider any assumptions, compared with linear mixing effect. In this paper, we estimate the joint model by means of the machine learning method [21]. And the simulation data including the longitudinal and survival data is used. Through the simulation data, we compare the maximum likelihood with the aid of the JM packages and the machine learning methods, such as the classification tree, nearest neighbor, random forest and support vector regression. What' s more, the running time is also be used to compare.

Joint Model Framework

The joint model of longitudinal data and survival data is divided into two parts: longitudinal

  DOI: 10.25236/icess.2019.174

sub-model and survival sub-model. Linear mixed effect model is usually used to fit the longitudinal sub-model, while survival sub-model is almost the Cox proportional risk model [12].

For each subject $i=1,2,\cdots,n$, $y_i=\left(y_{i1}^T,y_{i2}^T,\cdots,y_{iK}^T\right)$ is the K-variate continuous outcome vector with possible right-censoring, where each y_{ik} denotes an $\left(n_{ik}\times 1\right)$-vector of observed longitudinal measurements for the k-th outcome type: $y_{ik}=\left(y_{i1k},y_{i2k},\cdots,y_{in_{ik}k}\right)^T$.Each outcome is measured at observed (possibly pre-specified) time t_{ijk} for $j=1,2,\cdots,n_{ik}$ $\left(k=1,2,\cdots,K\right)$, which can differ between subjects and outcomes. Let T_i denote the observed failure time for the i-th subject ($i=1,2,\cdots,n$), which is taken as the minimum of the true event time T_i^* and the censoring time C_i, i.e., $T_i=\min\left(T_i^*,C_i\right)$.Furthermore, we define the failure indicator $\delta_i=I\left(T_i^*\le C_i\right)$, where $I\left(\bullet\right)$ is the indicator function that takes the value 1 if the condition $T_i^*\le C_i$ is satisfied , and 0 otherwise. Thus, the observed data for the time-to-event outcome consist of the pairs $\left\{\left(T_i,\delta_i\right),i=1,2,\cdots,n\right\}$.
In terms of the longitudinal sub-model,

$$\begin{aligned} &y_{ik}\left(t\right)=m_{ik}\left(t\right)+\varepsilon_{ik}\left(t\right),\varepsilon_{ik}\sim N\left(0,\sigma_k^2\right)\\ &m_{ik}\left(t\right)=x_{ik}^T\left(t\right)\beta_k+z_{ik}^T\left(t\right)b_{ik}\end{aligned} \tag{1}$$

$y_{ik}\left(t\right)$ denotes the value of the k-th longitudinal outcome for the i-th subject at time point t, and $m_{ik}\left(t\right)$ denotes the true value of the subject-specific i underlying longitudinal covariate at time point t for the k-th outcome. ε_{ik} is the model error term, which we assume to be independent and identically distributed normal with mean 0 and variance σ_k^2. $x_{ik}\left(t\right)$ and $z_{ik}\left(t\right)$ denote the design vectors for the fixed-effects β_k and for the random effects b_{ik}, respectively. More specifically, the complete vector of random effects $\boldsymbol{b}_i=\left(b_{i1}^T,b_{i2}^T,\cdots,b_{1K}^T\right)^T$ is assumed to follow a multivariate normal distribution with mean zero and variance-covariance matrix $\boldsymbol{D}$.

As to the survival sub-model, we apply the extended Cox model (also known as the Andersen-Gill model)

$$h_i\left(t\left|M_i\left(t\right),\omega_i\right.\right)=h_0\left(t\right)\exp\left(\gamma^T\omega_i+\alpha m_i\left(t\right)\right) \tag{2}$$

Among them, $M_i\left(t\right)=\left\{m_i\left(u\right),0\le u<t\right\}$ denotes the history of the underlying longitudinal process up to t, $h_0\left(t\right)$ denotes the baseline hazard function, and $\omega_i\left(t\right)$ is a vector of exogenous, possibly time-varying (such as treatment method index, disease history, gender, etc.), covariates with corresponding regression coefficients γ. α describes the effect of the potential output m_i on the hazard h_i [11].

To sum up, the joint model with multivariate longitudinal data and survival data is as follows:

$$\begin{aligned} &y_{ik}\left(t\right)=m_{ik}\left(t\right)+\varepsilon_{ik}\left(t\right)\\ &\qquad=\boldsymbol{x}_{ik}^T\left(t\right)\beta_k+\boldsymbol{z}_{ik}^T\left(t\right)b_{ik}+\varepsilon_{ik}\left(t\right)\\ &h_i\left(t\right)=h_0\left(t\right)\exp\left(\gamma^T\boldsymbol{\omega}_i+\alpha m_i\left(t\right)\right)\\ &\varepsilon_{ik}\sim N\left(0,\sigma_k^2\right)\\ &\boldsymbol{b}_i\sim MVN\left(0,\boldsymbol{D}\right)\end{aligned} \tag{3}$$

Estimation Method

Linear Mixed-effects Model. Linear mixed-effects models are extensions of linear regression models for data that are collected and summarized in groups. These models describe the relationship between a response variable and independent variables, with coefficients that can vary with respect to one or more grouping variables. A mixed-effects model consists of two parts, fixed effects and random effects. Fixed-effects terms are usually the conventional linear regression part, and the random effects are associated with individual experimental units drawn at random from a population. The random effects have prior distributions whereas fixed effects do not. Mixed-effects models can represent the covariance structure related to the grouping of data by associating the common random effects to observations that have the same level of a grouping variable. The standard form of a linear mixed-effects model is

$$y = \underset{fixed}{X\beta} + \underset{random}{Zb} + \underset{error}{\varepsilon} \tag{4}$$

Mixed-effects models are also called multilevel models or hierarchical models depending on the context. Mixed-effects models are a more general term than the latter two. Mixed-effects models might include factors that are not necessarily multilevel or hierarchical, for example crossed factors. That is why mixed-effects is the terminology preferred here. Sometimes mixed-effects models are expressed as multilevel regression models (first level and grouping level models) that are fit simultaneously.

Machine Learning Methods. The Cox proportional hazard model is used to estimate the survival sub-model. With regard to the estimation of the longitudinal sub-model, this kind of Statistical models, such as the mixed effect models, requires a variety of assumptions about the data, which are mathematically convenient but not necessarily satisfied. Therefore, the results of simulation with these models will produce bias, to the extent. Recent years, the machine learning, namely the data mining technique or the algorithm model, has been widely used. The algorithm method has no assumptions about the data, and the simulated results can be judged by the means of the cross-validation, which deviates from the hypothetical distribution. Based on this method, the prediction effect of the model is comparably great, and the results of the cross-validation are relatively easy to be understood and accepted by many practical investigators. In the following, we will summarize the main four methods in machine learning [22].

The first is the decision tree. When used for regression, the decision tree is also called the regression tree. While used for classification, it is called classification tree. It is simple to use the decision tree. If there are new observations (only arguments known), we can go from the root node to an end node according to the size of each variable of the decision tree, where the mean of the dependent variable is the predicted value of the dependent variable of the observation's value.

The nearest neighbor method is probably the easiest way among the all algorithms modellings. It regresses or classifies test sets based on training sets. Each regression or classification problem has some independent variables that make up a multidimensional space. First assume a distance in space. For example, in the case of continuous independent variables, Euclidean distance, absolute distance, etc. can be used. In the classification problem, a point of the test set should be judged as the type of the majority of the k points closest to the point in the training set. The "majority voting method" generally has the weight. The closer the point, the greater the weight of voting. The predicted value of the dependent variable of a test set point should be equal to the average of the corresponding dependent variable values of the k closest training sets.

Random forest is another combination method, which is also generated by a decision tree, which is formed by randomly resampled samples. The characteristic is that the segmentation variables of each node of these decision trees are not generated by all independent variables, but generated by a small number of randomly selected variables. So not only the samples that each decision tree are generated, but the generation of each node of each tree are random. These decision trees randomly generated are large and are therefore called random forests. The voting (or average) of the results is equal.

The support vector machine(SVM) is a two-class model, whose purpose is to find a hyperplane to segment the sample. The principle of segmentation is to maximize the interval and finally transform it into a convex quadratic programming problem. The regression method developed by SVM is called support vector regression(SVR).

Five-fold cross-validation was used for the above methods to judge the reliability of the results. In the calculation, five models are established through five randomly established training sets, and five standardized mean square errors (NMSE) are obtained for the training set and the test set, respectively, and five average NMSEs are obtained [21]. Let $\bar{y}$ be the mean of the dependent variable, $\hat{y}$ is the predicted value of the model obtained from the training set against a data set (which may be the training set itself or the test set), where NMSE is defined

$$NMSE = \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \tag{5}$$

In general, the smaller the NMSE, the better.

Methods Comparison in Simulation Dataset

In this section, we will compare the maximum likelihood and the machine learning method, through the simulation dataset and the actual dataset respectively.

Firstly, we use the maximum likelihood estimation to fit the simulation data by means of the JM packages. The fitted results are showed in Table 1. It is obvious that all the tests are quite remarkably. And the NMSE is 0.000705.

Using the maximum likelihood estimation to fit the joint model requires a prerequisite, that is the normal distribution. But this assumption may not always be valid. When the assumption of the distribution of random effects is not established, the estimated results may have a large deviation. We now test the normality of the residuals of the joint model fitted by the traditional method. According to the Figure 1 and Table 2, the Q-Q plot and the Shapiro normality test, we can see that the residuals do not obey the normal distribution. That is to say, the hypothesis of the longitudinal sub-model is invalid.

Table 1 The main outcomes of the joint model by JM package

	Longitudinal submodel				Survival submodel	
Variables	Intercept	ltime	ctsxl	binxl	ctsx	binx
coeffecient	0.998	1.066	1.007	0.997	0.647	0.607
p-value	0	0	0	0	< 2e-16	3.65e-10

Table 2 The results of the Shapiro test

Shapiro-Wilk normality test		
Data	W statistics value	p-value
res.lmefit	0.9554	< 2.2e-16

Table 3 The NMSE comparison of the four machine learning methods

Method	Classification tree	Random forest	Nearest neighbour regression	Support vector regression
NMSE	0.6180	0.7388	0.3941	0.7154

Table 4 The running time of tne kknn and the lme [secs]

Model	Kknn	Lme
Time	3.3004	32.2907

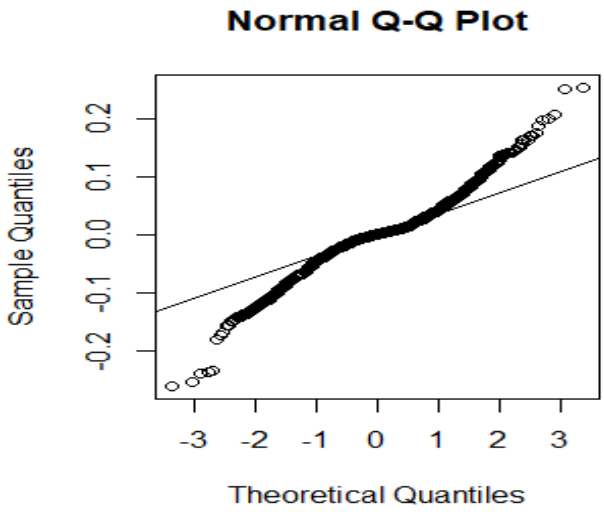

Figure 1. The Q-Q plot of the joint model by JM package

Next, we use the four machine learning methods to fit the longitudinal sub-models, while the survival sub-model is estimated by the Cox proportional hazard model. The NMSEs of the four methods are showed in Table 3. As we can see, the NMSE of kknn is the smallest among them. So, we will estimate the joint model only by the kknn. In the Figure 2, the left part is the residuals plot of the kknn, and the right is the JM. The slope of the trend line of the kknn method is larger than that of the trend line of the linear mixed-effects model.

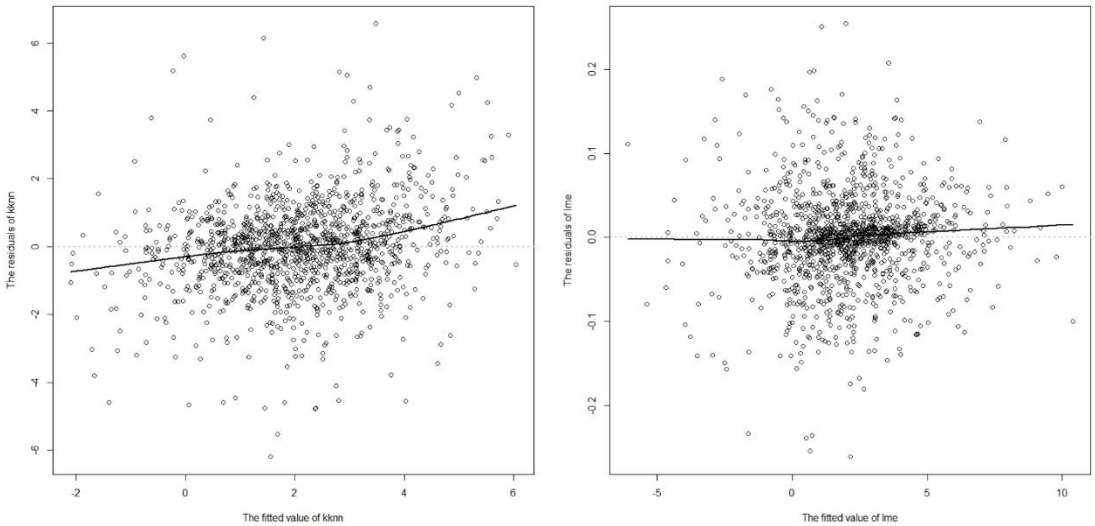

Figure 2. The plot of residuals and the fitted values of the linear mixed-effect model and kknn method

In the Figure 3, the left part is the residualss plot of the kknn, and the right is the JM. The residualss scattered points of the kknn regression are between the horizontal lines 7 and -7, mainly

between 1.5 and -1.5. Conversely, the residualss of the linear mixed effects range from -1 to 1, mainly from -0.15 to 0.15. It is clear that the residualss of the kknn is more dispersed than ones of the linear mixed effect regression.

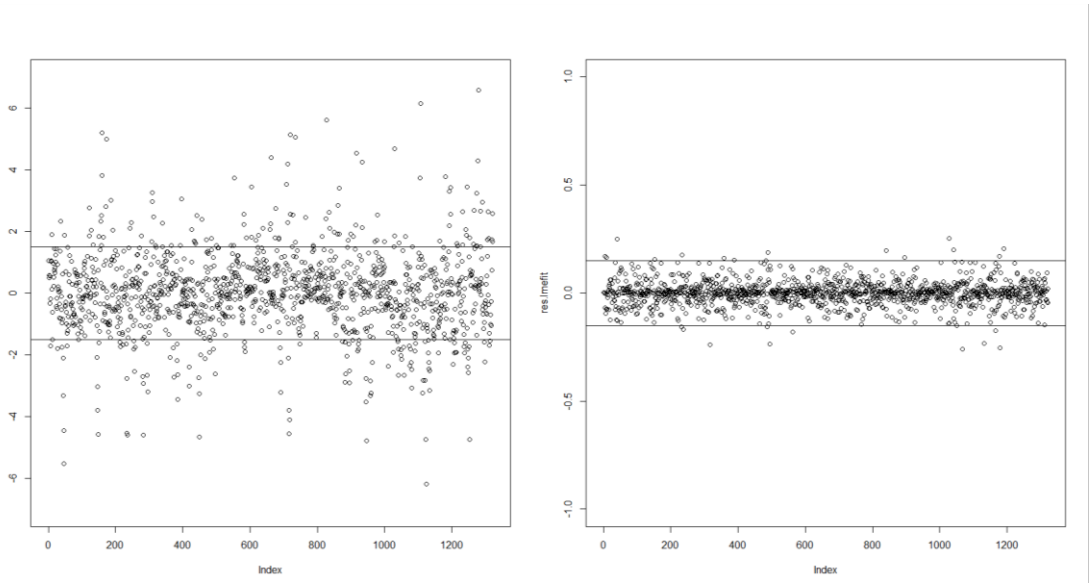

Figure 3. The residuals plot of the linear mixed-effect model and kknn method

As for the survival sub-model, we employ the Cox-Snell residuals [21]. The formula of the Cox-Snell residuals is as the following:

$$\begin{aligned} r_i^{tcs} &= \int_0^{T_i} R_i(s) h_0(s) \exp\left\{\gamma\omega_i + \alpha m_i(t)\right\} ds \\ &= N_i(T_i) - r_i^{tm}(T_i) \\ &= \delta_i - r_i^{tm} \end{aligned} \tag{6}$$

In Figure 4, the upper left part is the KM estimate of the Cox-Snell residuals of the nearest neighbour joint model, and the upper right is the KM estimate of the nearest neighbor Cox-Snell residuals; the bottom left is the KM estimate of the Cox-Snell residuals of the linear mixed effect joint model. On the upper right is the KM estimate of the Cox-Snell residuals of the linear mixed-effects joint model. The dashed line is its corresponding 95% point-by-point confidence interval, and the gray solid line is an exponential distribution function with a parameter of 1. Compare the Cox-Snell residualss of these two method, we can see that the KM estimation image of the Cox-snell residualss of the kknn joint model is closer to the ideal state than those of the linear mixed effect joint model. Therefore, the kknn method is better fitted to the survival sub-model.

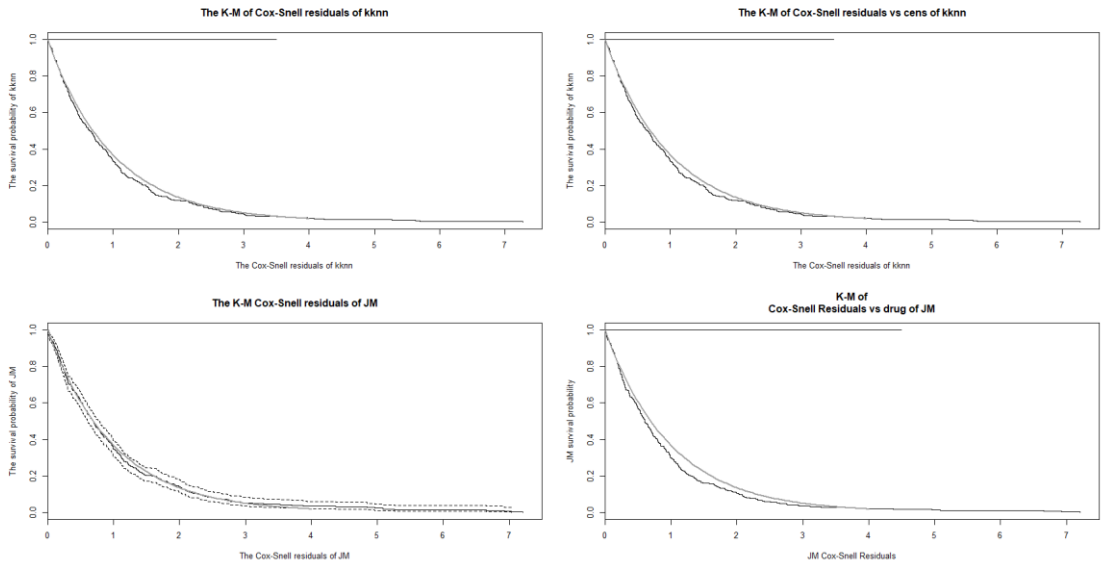

Figure 4. The comparison plot of two methods

In addition, we compare the running time of these two methods. The specific figures are showed in Table 4. Evidently, the kknn runs faster than the lme in fitting the joint model.

We can also compare these two methods of joint model in some actual dataset, such as the aids data [9]. The results are the same to that of the simulation data[21].

Summary

In this paper, the machine learning method is used to fit the joint model, which breaks the limitations of the traditional simulation method. The longitudinal sub-model is regressed by machine learning methods such as classification tree, nearest neighbor method, random forest and support vector regression. The survival sub-model is still estimated by the Cox proportional hazard model. According to the Dimitris diagnosis method, the fitted survival model is better than the survival model of the traditional joint model, and the residuals image is more in line with the theoretical results. In addition, the machine learning method is different from the traditional analog joint model method in that it does not consider various distribution hypotheses, and can also make full use of observed variables and observation information. From the experiments in this paper, we know that the residuals graph of machine learning is more scattered than the residuals graph of the linear mixed-effects model.

When using a conventional method such as a linear mixed-effects model to estimate a vertical sub-model, it is also necessary to verify whether the assumptions of the model are satisfied. When the assumptions are true, the NMSE of the linear mixed-effects model is naturally smaller than that of machine learning. Moreover, thanks to the years of accumulation, the traditional estimation method is more mature. In this case, the traditional method is better. However, when the assumption is not true, it is better to aid for the machine learning method to fit the joint model, because it is not necessary to consider the deviation of the result when the assumption is not established.

References

[1] U. G. Dafni, A. A. Tsiatis. Evaluating surrogate markers of clinical outcome when measured with error[J]. Biometrics, (1998): 1445-1462.

[2] X. Piulachs, R. Alemany and M. Guillen. Joint modelling of survival and emergency medical care usage in Spanish insureds aged 65+[J]. PloS one, (2016), 11(4).

[3] W. Ye, X. Lin and J. M. G. Taylor. Semiparametric modeling of longitudinal measurements and time-to-event data a two-stage regression calibration approach[J]. Biometrics, (2008), 64(4): 1238-1246.
[4] A. A. Tsiatis, M. Davidian. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error[J]. Biometrika, (2001), 88(2): 447-458.
[5] H. Lin, C. E. McCulloch and Mayne S T. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables[J]. Statistics in Medicine, (2002), 21(16): 2369-2382.
[6] Y. Y. Chi, J. G. Ibrahim. Joint models for multivariate longitudinal and multivariate survival data[J]. Biometrics, (2006), 62(2): 432-445.
[7] D. Rizopoulos, P. Ghosh. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event[J]. Statistics in medicine, (2011), 30(12): 1366-1380.
[8] J. G. Ibrahim, H. Chu and L. M. Chen. Basic concepts and methods for joint models of longitudinal and survival data[J]. Journal of Clinical Oncology, (2010), 28(16): 2796.
[9] D. Rizopoulos. JM: An R package for the joint modelling of longitudinal and time-to-event data[J]. Journal of Statistical Software (Online), (2010), 35(9): 1-33.
[10] M. J. Sweeting, S. G. Thompson. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture[J]. Biometrical Journal, (2011), 53(5): 750-763.
[11] D. Rizopoulos. *Joint models for longitudinal and time-to-event data: With applications in R*[M]. Chapman and Hall/CRC, 2012.
[12] A. A. Tsiatis, M. Davidian. Joint modeling of longitudinal and time-to-event data[J]. Statistica Sinica, (2004), 14(3):793-818.
[13] D. Rizopoulos. The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC[J]. ArXiv preprint arXiv:1404.7625, 2014.
[14] G. L. Hickey, P. Philipson and A. Jorgensen, et al.. joineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes[J]. BMC medical research methodology, (2018), 18(1): 50.
[15] K. Mauff, E. Steyerberg and I. Kardys, et al.. Joint models with multiple longitudinal outcomes and a time-to-event outcome[J]. ArXiv preprint arXiv:1808.07719, (2018).
[16] J. Barrett, P. Diggle and R. Henderson, et al.. Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), (2015), 77(1): 131-148.
[17] D. Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data[J]. Biometrics, (2011), 67(3): 819-829.
[18] F. Hsieh, Y. K. Tseng and J. L. Wang. Joint modeling of survival and longitudinal data: likelihood approach revisited[J]. Biometrics, (2006), 62(4): 1037-1043.
[19] P. Philipson, P. Diggle and I. Sousa, et al.. joineR: Joint modelling of repeated measurements and time-to-event data[J]. (2012).
[20] Q. Ye, L. Wu. Two-step and likelihood methods for joint models of longitudinal and survival data[J]. Communications in Statistics-Simulation and Computation, (2017), 46(8): 6019-6033.
[21]Z Wen. The Joint Model of Longitudinal and Survival Data—Based on Machine Learning Methods[J]. Statistical and Application, (2015), 4(04): 252. (In Chinese)
[22]X.Z. Wu. Complex Data Statistics Method: R-based Application [M]. Renmin University of China Press, 2013. (In Chinese)