Frailty Models in Survival Analysis

Faltus, Václav
2006

Frailty Models in Survival Analysis

Post-Graduate Student:
MGR. VÁCLAV FALTUS, M.SC.
Department of Medical Informatics
Instutite of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2
182 07 Praha 8
Czech Republic
faltus@euromise.cz

Supervisor:
DOC. ZDENĚK VALENTA, M.SC., PH.D.
Department of Medical Informatics
Instutite of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2
182 07 Praha 8
Czech Republic
valenta@euromise.cz

Field of Study:
Biomedical Informatics

The work was suported by the grant 1M06014 of the Ministry of Education of the Czech Republic

Abstract

The aim of this paper is to present an overview of the methods used in survival analysis and especially modeling survival data. Since the topic of my future Ph.D. thesis is *Statistical models for correlated survival data* we introduce the use of frailties as an equivalent of random effects in common statistical modeling together with its connection to correlation. Frailty model, how model with frailties is called, uses frailties as a parameter for individuals. Those who are most frail will experience an event earlier than others.

Keywords: survival analysis, frailty models

1. Introduction

Survival analysis is analysis of time to the occurrence of an event. Examples of such data arise in many diverse fields, such as medicine, biology, public health, epidemiology, engineering, economics, and demography. In epidemiology the event is not always a death, but it can be, for example, a first occurrence of relapse, duration of response to treatment, time to development of a disease, duration of stay in hospital, and duration of a seizure. Survival analysis attempts to answer questions such as: what is the fraction of a population which will survive past a certain time? Of those that survive, at what rate will they die or fail? Can multiple causes of death or failure be taken into account? How do particular circumstances or characteristics increase or decrease the probability of survival?

Analysis of survival data is very often complicated by the issue of censoring, where the event is known only to not have occurred during a certain period of time and by truncation, where the individuals enter the study only if they survive a sufficient length of time or individuals are included in the study only if the event has occurred by a given time. These two issues play a very important role in estimating model parameters and in modeling survival data. The theoretical basis for the analysis of survival data has been solidified by connecting it to the study of counting processes and martingale theory [6]. This theory, among many other benefits, substantially helped with accounting for censoring.

In this paper, we start with an introduction to the key concepts in survival analysis: the hazard, survival, and cumulative hazard functions. Then we turn to regressions models with emphasis on frailty model as an extension of the Cox proportional hazards model. The frailty model may be also called the shared frailty models because the frailty may be parameter shared among members of one group or family. We also discuss some models used in multivariate survival data and show the universality of the concept of frailties.

2. Basic concepts

Random variables analyzed in survival analysis are called the failure times T and they are always non-negative $T \geq 0$. T can either be discrete (taking a finite set of values, e.g. $a_1, a_2, \ldots, a_n$) or continuous

(defined on $[0, \infty)$).

The concept used in describing time-to-event phenomena is the survival function, the probability of an individual surviving beyond time t. It is defined as

$$S(t) = \mathrm{P}(T > t).$$

Survival function is non-increasing function with a value of 1 at the origin and 0 at infinity. When T is a continuous random variable then $S(t)$ is a non-increasing continuous function. When T is a continuous variable, the survival function is a complement of the cumulative distribution function, $S(t) = 1 - F(t)$, where $F(t) = 1 - P(T \leq t)$. The survival function is the integral of the probability density function $f(x)$. Then

$$S(t) = 1 - F(t) = 1 - P(T \leq t) = \int_t^\infty f(x)dx, \tag{1}$$

and

$$f(x) = -\frac{dS(x)}{dx}.$$

Fundamental in survival analysis is the hazard function. It is also known as the conditional failure rate in reliability, the force of mortality in demography, the intensity function in stochastic processes, the age-specific failure rate in epidemiology or simply as the hazard rate. The hazard rate is defined by

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{\mathrm{P}(T \in [t, t + \Delta t) | T \geq t)}{\Delta t} = \\ &= \lim_{\Delta t \to 0} \frac{\mathrm{P}(T \in [t, t + \Delta t))}{\Delta t \cdot \mathrm{P}(T \geq t)} = \\ &= \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t \cdot S(t)} = \frac{F'(t)}{S(t)}. \end{aligned} \tag{2}$$

Then

$$h(t) = \frac{F'(t)}{S(t)} = \frac{(1 - S(t))'}{S(t)} = \frac{-S'(t)}{S(t)} = -\ln'(S(t)). \tag{3}$$

A related quantity is the cumulative hazard function, defined by

$$H(t) = \int_{y=0}^{t} h(y)dy = -\ln S(t) + \ln S(0) = -\ln(S(t)). \tag{4}$$

Then

$$S(t) = e^{-H(t)}.$$

3. Modeling survival data

A problem frequently encountered in analyzing survival data is that of adjusting the survival function to account for concomitant information (sometimes referred to as covariates, explanatory variables or independent variables).

A popular concept for modeling the relationship of covariates to a survival outcomes is represented by the Cox proportional hazards model. Let X_{ij} be the covariate of the ith person, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$. The set of covariates then, like in linear regression, forms an $n \times p$ matrix and X_i is used to denote the covariate vector for subject i (the ith row of the matrix).

The Cox model specifies the hazard for individual i as

$$h_i(t) = h_0(t) \cdot e^{X_i\beta}, \tag{5}$$

where h_0 is unspecified nonnegative function of time called the baseline hazard, and β is $p \times 1$ column vector of coefficients. Event rates cannot be negative (observed deaths cannot unhappen), and the exponential thus plays the an important role in ensuring that the final estimates are a physical possibility.

Because the hazard ratio for two subjects with fixed covariate vectors X_i and X_j,

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t) \cdot e^{X_i\beta}}{h_0(t) \cdot e^{X_j\beta}} = \frac{e^{X_i\beta}}{e^{X_j\beta}} = e^{[X_i - X_j]\beta} \tag{6}$$

is constant over time, the model is also known as the proportional hazards model. Estimations of β is based on the partial likelihood function introduced by Cox. Commonly used algorithm for finding approximations to the roots of real-valued function is the Newton-Raphson algorithm. The idea of the method is that one starts with a value which is reasonably close to the true root, then replaces the function by its tangent and computes the root of this tangent. This root of the tangent will typically be a better approximation to the function's root. The procedure is iterative and compares the nth guess with $n + 1$th, with the initial value usually set to 0. This algorithm is robust for the Cox partial likelihood. Convergence problem are very rare and easily addressed by simple methods such as step-halving.

3.1. The concept of frailty

In the last several years there has been active research in the area of survival models involving random effects. In this setting, a random effect is a continuous variable, that describes excess of risk or frailty for distinct categories, such as individuals or families. The idea is that individuals are characterized by frailty parameter that reflects their state of health. Those who are most frail will die earlier than the others.

Frailty models are also used in making adjustments for overdispersion in survival studies. Here, the frailty represents the total effect on survival of the covariates not measured when collecting information on individual subjects. If these effects are ignored, the resulting survival estimates may be misleading. Corrections for overdispersion allow for adjustments for other important effects that were unaccounted for in the study.

Computationally, frailty parameters are usually viewed as unobserved covariates. This leads to the use of EM (expectation-maximization) algorithm as an estimation tool. EM algorithm is an algorithm for obtaining maximum likelihood estimates of parameters in probabilistic models, where the model depends on unobserved latent variables. EM alternates between performing an expectation (E) step, which computes an expectation of the likelihood by including the latent variables as if they were observed, and a maximization (M) step, which computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood obtained at the E step. The parameters obtained at the M step are then used to begin another E step, and the process is repeated. However, the algorithm is slow, proper variance estimates require further computation, and no implementation appears in any of the more widely available packages. The computation can be approached instead as a penalized Cox model.

The most common model used is a shared frailty model. As we will see, it is an extension of the proportional hazards. Assume that each subject $i, i = 1, \ldots, n$, is a member of a single group $j, j = 1, \ldots, q$. Then we write the proportional hazards model

$$h_i(t) = h_0(t) \cdot e^{(X_i\beta + Z_i\omega)}, \tag{7}$$

where X_i and Z_i are the ith rows of covariate matrices $X_{n\times p}$ and $Z_{n\times p}$ respectively. X and β correspond to p fixed effects in the model, ω is a vector containing the q unknown random effects or frailties, and Z is a design matrix – Z_{ij} equals 1 if subject i is a member of family j, 0 otherwise.

For subject i, which is a member of the jth family the proportional hazards model can be also written as

$$h_{i(j)}(t) = h_0(t) \cdot w_j \cdot e^{X_i\beta}, \tag{8}$$

where the w_j is the frailty parameter for family j. This can be easily rewritten in the form of equation 7, with i ranging over all subjects,

$$w_j = e^{\omega_j},$$

and Z a matrix of indicator variables such that $Z_{ij} = 1$ if subject i is a member of family j and 0 otherwise. In this model, each individual can belong to only one family. The frailty parameter w_j has multiplicative effect on the hazard rates.

According to a penalty function and the choice of a design matrix we distinguish between gamma frailty model and Gaussian random effects model. The advantage of frailty having either of these two distributions is that the shared frailty model can be written exactly as a penalized likelihood and further estimation can be done more easily. Additional assumptions are discussed below.

Penalty function

$$p(\omega) = (1/\omega) \cdot \sum (\omega_i - e^{\omega_i})$$

and the design matrix Z as defined for shared frailty model gives us the equivalence with gamma frailty model. The ω_is are distributed as the logs of iid (independent identically distributed) random variables and the tuning parameter θ is their variance. For this frailty distribution, the correlation of subject within groups is $\theta/(2+\theta)$.

Penalty function

$$p(\omega) = (1/2\omega) \cdot \sum \omega_i^2$$

and general design matrix Z gives the Gaussian random effects model. The tuning parameter θ of the penalty function is the variance of the ω_is.

The fact that both penalties are the log-likelihoods for a random sample of $\omega_1, \ldots, \omega_n$ from the appropriate distribution raises the question whether other frailty distributions can be accommodated within penalized framework. From the viewpoint of the penalized fitting procedure θ is as nuisance or "tuning" parameter of the computation.

The variance can be a fixed parameter set by the user, either set directly or chosen indirectly by specifying the degrees of freedom for the random effects term. Next option is to seek an overall best variance by minimizing the AIC (Akaike's Information Criterion) or corrected AIC. The idea behind the AIC is to examine the complexity of the model together with goodness of its fit to the sample data, and to produce a measure which balances between the two. The formula for computing AIC involves the likelihood and number of parameters of examined model. The corrected AIC uses much larger penalty as the number of parameters approaches the sample size n. This leads to models which are a bit more conservative in the amount of model parameters.

For the gamma frailty, the profile likelihood for θ can be used. This gives a global solution that is identical [6] to the EM algorithm for a shared gamma frailty model. When a Gaussian distribution of the frailty is chosen, the variance θ of the random effect can be chosen based on an approximate REML (restricted maximum likelihood) equation. Restricted maximum likelihood serves to estimate the variance of any distribution and is also based on the likelihood principle.

3.2. Shared Gamma frailty models

As it was said, the most commonly used estimation procedure in frailty models is the EM algorithm. It gives discrete estimator of the distribution and does not allow direct estimation of the hazard function. Rondeaou et al. [2] present how to use maximum penalized likelihood estimation in estimating continuous hazard function in a shared gamma-frailty model with right-censored and left-truncated data. Instead of penalizing the frailties [6] the hazard function is penalized.

Let us consider the model in which the hazard function will partly depend on an unobservable random variable thought to act multiplicatively on the hazard and allow for stratum-specific baseline hazards. For

the jth individual ($j = 1, \ldots, n_{ih}$) of the hth stratum ($h = 1, \ldots, K$) and the ith group ($i = 1, \ldots, G$), let T_{ihj} denote the survival times under study. We also assume the independence of survival times and left-truncation times. The censoring times are assumed independent of the failure times and of the frailties Z_i.

Then the hazard function conditional on the frailty is

$$h_{ihj}(t|Z_i) = Z_i \cdot h_{0h}(t) \cdot e^{X_{ihj}\beta}, \tag{9}$$

where $h_{0h}(t)$ is the baseline hazard function for stratum h; $X_{ihj} = (X_{1ihj}, X_{2ihj}, \ldots, X_{pihj})$ denotes the covariate row vector for the j^{th} individual of stratum h and group i, and β is the corresponding vector of regression parameters.

Conditionally on the frailty Z_i, the failure times $T_{ih1}, T_{ih2}, \ldots, T_{ihn_{ih}}$ are assumed to be independent. It is also assumed that the Z_i's are independently and identically distributed from a gamma distribution with mean 1 and unknown variance θ. The probability density function is then

$$g(z) = \frac{z^{(1/\theta-1)} \cdot e^{-z/\theta}}{\Gamma(1/\theta) \cdot \theta^{1/\theta}}. \tag{10}$$

Large values of θ signify a closer positive relationship between the subjects of the same group and greater heterogeneity among the groups.

Instead of the use of the Cox partial likelihood the full likelihood obtained by integrating out the frailty Z_i from the joint likelihood is presented. Because of allowance for left-truncated data the likelihood is expressed conditional on involved failure times. Rondeau et al. [2] give the exact formula of the full log-likelihood which can be directly maximized to obtain the estimates of β, θ and the baseline hazard function $h_0(t)$, which is assumed to be smooth. Maximum penalized estimators (MPnLE) of $h_0(t)$, β and θ are then obtained by maximizing function consisting of the full log-likelihood and the penalty. The penalty has large values for rough functions and contains also smoothing parameter specific for particular stratum. The idea is that the penalty is for large values of the smoothing parameter forced toward zero and if the smoothing parameter is small, then the main contribution to penalized likelihood will be the full log-likelihood mentioned above.

Rondeau et al. also show the way how to obtain estimators of the baseline hazard function. It is an approximation obtained by using splines. Splines are polynomial functions which are combined linearly to give the interpolation of a function. In addition, confidence bands for the baseline hazard estimators can be given using the variance estimates. The choice of smoothing parameter can be done heuristically by plotting several curves and by choosing that which seems most realistic.

3.3. Nested Gamma frailty models

The frailty model is a random effect survival model, which allows for unobserved heterogeneity or for statistical dependence between observed survival data. The nested frailty model accounts for the hierarchical clustering of the data by including two nested random effects. They substitute the frailty Z_i from equation 9, only the structure is a bit more complicated by the fact that one frailty is nested into another one which refers to higher hierarchical level. Thus nested frailty models are particularly appropriate when data are clustered at several hierarchical levels and it does not matter if the situation arised naturally or by design of the study. In such cases it is important to estimate the parameters of interest as accurately as possible by taking into account the hierarchical structure of the data. Rondeau et. al [3]. present a maximum penalized likelihood estimation (MPnLE) to estimate non-parametrically a continuous hazard function in a nested gamma-frailty model with right-censored and left-truncated data. The estimators for the regression coefficients and the variance components of the random effects are obtained simultaneously.

To demonstrate the MPnLE method and the nested frailty model two examples were given [3]. One was for modeling the effect of particulate air pollution on mortality in different areas with two levels of geographical regrouping. The other application concerned recurrent infection times of patients from different hospitals.

Simulation study was performed and it proved that the semi-parametric approach yields satisfactory results. Using a shared frailty model instead of nested frailty model with two levels of regrouping leads to inaccurate estimates, with an overestimation of the variance of the random effects. Next, even when the frailty effects are fairly small in magnitude, they are important since they alter the results in a systematic pattern.

Let us consider a multilevel proportional hazards model with two sets of nested random effects that act multiplicatively on the hazard, so that a large value of these variables increases the hazard. There are several types of models for expressing how the hazard function depends on the explanatory variables. The most popular model when analyzing epidemiological survival data is the proportional hazards model even if some other models as additive or accelerated failure time can be used. Proportional hazards models are semi-parametric and fairly flexible and their covariates can be time-dependent. We treat the case of a cohort with G independent clusters $(i = 1, \ldots, G)$. Within the ith cluster, there are J_i correlated sub-clusters $(j = 1, \ldots, J_i)$. We treat the case of right-censored and left-truncated data. T_{ijk} denotes the survival times under study for subject k $(k = 1, \ldots, K_{ij})$ from subgroup j, and group i. We also assume the independence of survival times T_{ijk} and left-truncation times.

Then we define two random effects v_i and w_{ij} and assume that the cluster-level random effects v_i and the sub-cluster random effects w_{ij} are independent and gamma-distributed random effects $(\Gamma(1/\alpha; 1/\alpha))$ and $(\Gamma(1/\eta; 1/\eta))$ with $E(v_i) = 1, var(v_i) = \alpha$ and $E(w_{ij}) = 1, var(w_{ij}) = \eta$.

For identifiability, they have a mean equal to 1 at birth (i.e. at duration $t_{ijk} = 0$). If the variance is null, then observations from the same group are independent. A larger variance implies greater heterogeneity in frailty across groups and a greater correlation of the survival times for individuals belonging to the same group. Mainly for reasons of mathematical convenience, the frailty terms are often assumed to follow a gamma distribution. Gamma distribution is chosen because of its correspondence to shared frailty model and the possibility to express the shared frailty model as penalized likelihood. To construct the likelihood function, apart from the usual assumption of independent censoring, the censoring must be non-informative for v_i and w_{ij}.

The hazard function conditional on the two frailties v_i and w_{ij}, for individual (i, j, k) is

$$h_{ijk}(t|v_i, w_{ij}) = v_i \cdot w_{ij} \cdot h_0(t) \cdot e^{X_{ijk}\beta}, \tag{11}$$

where $h_0(t)$ is the baseline hazard function; $X_{ijk} = (X_{1ijk}, \ldots, X_{pijk})$ denotes the covariate row vector for the kth individual, with p the number of covariates, and β is the corresponding vector of regression parameters.

Rondeau et al. estimated the baseline hazard function in a nested frailty model non-parametrically using a penalized full likelihood assuming both left-truncated and right-censored survival data. This makes it possible to estimate simultaneously the regression coefficients, the variance component parameters and especially a smooth hazard function, which cannot be correctly estimated using conventional non-parametric methods.

A major advantage of nested frailty models is their ability to analyze data that are also correlated at several different hierarchical levels. The eventual hierarchical structure of the data needs to be taken into account in survival analysis to obtain accurate inferences. Ignoring random cluster effects may result in overlooking the importance of certain cluster effects and calls into question the validity of traditional statistical techniques such as the shared frailty model. The nested frailty model proved helpful in diagnosing the source of correlation in data. This method can be easily extended to bivariate frailty models, making it possible to treat two events simultaneously per subject.

3.4. Multivariate survival data

The shared frailty model can also be viewed as a specific kind of the common risk model. The frailty is the term that describes the common risk, acting as a factor on the hazard function. In most settings the common risks are assumed random and therefore we can speak about mixture model. Given the frailties the observations are assumed independent. This is then called conditional independence.

The bivariate survival function can be defined as $S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$ and in the multivariate case obviously $S(t_1, t_2, \ldots, t_n) = P(T_1 > t_1, T_2 > t_2, \ldots, T_n > t_n)$. P. Hougaard [7] shows also the forms of the bivariate survival function by assuming certain distributions of the frailty. It is also shown that the dependence can be assessed by by the rank-based correlation-type measures like Kendall's τ or Spearman's ρ and these measures depend only on the frailty distribution. That is, they are independent of the hazard function.

In many cases, we may need extensions of the common risk models discussed above and, similarly, more models with varying degrees of dependence. The general frailty approach can be used to create a random treatment by group interaction or other models with several sources of variation. Nested model discussed by Rondeau et al. may be seen as special univariate case of multivariate models discussed in [7]. It was already mentioned above that the two levels in nested model can be extended. This situation can arise in the field of genetics where there are several sources of variation present. For twins the question may be whether the dependence is the same for monozygotic and dizygotic twins. P. Hougaard shows also an example of the nested trivariate parallel data model, which is a model that would be applicable for a sibling group where individuals 1 and 2 are twins and individual 3 is a single birth. Survival times of the first two siblings T_1, T_2 may be strongly dependent. T_3 may show more modest dependence. This scenario can be accomplished by three frailties. In general, each frailty may have different distribution. P. Hougaard proposes the use of semi-parametric estimate. That requires extension for a program for the shared frailty models and seems to be complicated.

So far we discussed multiplicative frailty models but additive models are plausible too, mainly because the multiplicative models are not able to handle all dependence structures. Therefore, the additive models seem to be more operational. The disadvantage of the additive model is that more parameters are needed to define the models where the parameters cannot be identified from subsets of the data.

Frailty can be also modeled as stochastic process. P. Hougaard discusses independent increments frailty models, piecewise gamma model, moving average model, hidden cause of death model and the Woodbury-Manton model. All these models apply when there is a instantaneous and/or short-term frailty. The ideas are given, however the mathematical complexity of these models is a major problem. Instantaneous dependence has a mathematical appeal allowing for many theoretical evaluations. It is obtained by using a frailty varying randomly over time, where the hazard is described by the independent increments of the process. Short-term dependence models seem to be a more relevant extension of the frailty model and from mathematical point of view, there seems to be a lot of potential. The key problem is the computational burden. Models with with time-varying frailties seem particularly important for recurrent events data, but most such model treat only additive piecewise constant frailties and with a limited number of intervals. [7]

4. Conclusion

Short overview of frailty models used in survival analysis was given together with discussion and references to available literature. Discussion concerning censoring, truncation, estimation of parameters using the likelihood, partial likelihood and other methods together with numerical procedures exceeds the scope of this paper and thus were only slightly mentioned. The use and rising popularity of frailty models used in analysis of correlated censored and possibly truncated data is relatively new and active research in this field is taking place. In section about multivariate survival data there is a bit of a discussion about correlation and conditional independence. In my prospective Ph.D. thesis I would like to focus on statistical aspects of models involving frailty parameters and this paper was in my opinion a good start to accomplish that goal.

References

[1] C. A. McGilchrist and C. W. Aisbett, "Regression with Frailty in Survival Analysis", *Biometrics*, vol. 47, pp. 461–466, 1991.

[2] Virginie Rondeau, Daniel Commenges, and Pierre Joly, "Maximum Penalized Likelihood Estimation in a Gamma-Frailty Model" *Lifetime Data Analysis*, vol. 9, pp. 139–153, 2003.

[3] Virginie Rondeau, Daniel Commenges, and Pierre Joly, "Nested frailty models using maximum penalized likelihood estimation" *Statistics in Medicine*, 2006.

[4] Erik Parner, "Asymptotic Theory for the Correlated Gamma-Frailty Model", *The Annals of Statistics*, vol. 26(1), pp. 183–214, 1998.

[5] John P. Klein and Melvin L. Moeschberger, "Survival Analysis: Techniques for Censored and Truncated Data", *Springer*, 1997.

[6] Terry M. Therneau and Patricia M. Grambsch, "Modelling Survival Data: Extending the Cox Model", *Springer*, 2000.

[7] Philip Hougaard "Analysis of Multivariate Survival Data", *Springer*, 2000.

[8] Vladimír Bencko, Karel Hrach, Marek Malý, Hynek Pikhart, Jindra Reissigová, Štěpán Svačina, Marie Tomečková, and Jana Zvárová, "Statistické metody v epidemiologii: svazek 2", *Nakladatelství Karolinum*, 2003.