Methods of the Survival Analysis

J. Fürstová[1]

[1]Faculty of Medicine and Dentristy, Palacky University, Olomouc, Czech Republic

Supervisor: Doc. Zdeněk Valenta, M.Sc., M.S., Ph.D.

Summary

The survival analysis is a set of statistical methods dealing with time-to-event data. In biomedical applications the event of interest is usually relapse of the disease or death. A special feature of the survival analysis is censoring and truncation of data. When censoring or truncation occurs some information about the patients' survival is lost, e.g. some patients are lost to follow-up or the study ends before all the patients die. The survival analysis methods are used for estimation of the survival time distribution, for identification of risk factors that affect the survival time, and also for predicting the survival time when risk factors are present. Survival analysis methods have been further developed by the means of counting processes and martingale theory. Univariate survival analysis methods have been extended to multivariate setting. The multivariate survival analysis covers the field where independence between survival times cannot be assumed. Multi-state models and frailty models represent the two main approaches of multivariate methods.

Keywords: survival analysis, survival function, hazard function, cumulative hazard function, censoring, truncation, Kaplan-Meier estimator, Nelson-Aalen estimator, Cox PH model, partial likelihood, counting process, history, filtration, martingale, competing risk, multi-state models, frailty models

1. Introduction

The survival analysis is a collection of statistical methods for analyzing time-to-event data. The commencement of the survival analysis dates back to the 18th century when analyses of mortality experience of human populations started. During the World War II, the survival analysis focused on engineering - reliability of military equipment was being analyzed. After the World War II the interest turned towards economics and medicine. In 1960s, after the fundamental article of E. L. Kaplan and P. Meier [9] had been published, medical applications of the survival analysis shifted to the center of statistical focus.

2. Basic concepts in the survival analysis

2.1 The survival and hazard function

Let X be the time until some specified event occurs, i.e. X is a non-negative real valued random variable having continuous distribution with finite expectation. There are several functions characterizing the distribution of X:

- The probability density of X: $f(x)$, $x \geq 0$.
- The survival function:

$$S(x) = P(X > x) = \int_x^\infty f(u)du$$
$$= 1 - F(x),$$

where $F(x)$ is the cumulative distribution function. The survival function describes the probability of an individual surviving beyond time x (experiencing the event after time x).

- The hazard function:

$$\lambda(x) = \lim_{\Delta x \to 0^+} \frac{P(x \leq X < x + \Delta x \mid X \geq x)}{\Delta x},$$

for all $x > 0$. The hazard function represents a conditional probability rate at which an individual alive at time x will experience an event in the next instant. There is a close relationship between the hazard and the survival functions:

$$\lambda(x) = \lim_{\Delta x \to 0^+} \frac{1}{\Delta x} \frac{P(x \leq X < x + \Delta x)}{P(X \geq x)}$$

$$= \lim_{\Delta x \to 0^+} \frac{\frac{1}{\Delta x}\int_x^{x+\Delta x} f(u)du}{\int_x^\infty f(u)du} = \frac{f(x)}{S(x)}$$

$$= -\frac{\frac{dS(x)}{dx}}{S(x)} = -\frac{d}{dx}\ln S(x).$$

- The cumulative hazard function:

$$\Lambda(x) = \int_0^x \lambda(u)du = -\ln S(x).$$

Thus

$$S(x) = \exp(-\Lambda(x)) = \exp\left(-\int_0^x \lambda(u)du\right).$$

If X is a discrete random variable taking values $x_1 < x_2 <\ldots$ with associated probability mass function $f(x_i) = P(X = x_i)$, $i = 1, 2, \ldots$, the survival function is

$$S(x) = \sum_{j:x_j > t} f(x_j)$$

the hazard at x_i is

$$\lambda_i = P(X = x_i \mid X \geq x_i) = \frac{f(x_i)}{S(x_i^-)}, i = 1,2,\ldots,$$

where $S(x^-) = \lim_{t \to x^-} S(t)$. The survival function and probability mass function can be also written as (see [8])

$$S(x) = \prod_{j:x_j \leq x} (1 - \lambda_j), \quad f(x_i) = \lambda_i \prod_{j=1}^{i-1} (1 - \lambda_j).$$

More generally, the distribution of X may have both discrete and continuous components. The approach to the discrete and continuous parts can be unified through the notion of a product integral: Let $\lambda_c(x)$ be the continuous component of the hazard function, and let λ_1, λ_2,... be the discrete components at times $x_1 < x_2 <\ldots$ The overall survival function is then

$$S(x) = \exp(-\int_0^x \lambda_c(u)du) \prod_{j:x_j \leq x} (1 - \lambda_j)$$

and the cumulative hazard function is

$$\Lambda(x) = \int_0^x \lambda_c(u)du + \sum_{j:x_j \leq x} \ln(1 - \lambda_j).$$

Let $d\Lambda(x)$ be a differential increment of the cumulative hazard over the interval $[x, x+dx)$:

$$
\begin{aligned}
d\Lambda(x) &= \Lambda(x^- + dx) - \Lambda(x^-) \\
&= P(X \in [x, x+dx) \mid X \geq x) \\
&= \begin{cases} -\ln(1-\lambda_i) & \textit{for } x = x_i,\ i = 1,2,\ldots \\ \lambda_c(x)dx & \textit{otherwise.} \end{cases}
\end{aligned}
$$

The survival function in the discrete, continuous, or mixed cases can then be written as

$$S(x) = \mathsf{P}_0^x(1 - d\Lambda(u)),$$

where

$$\mathsf{P}_0^x(1 - d\Lambda(u)) = \lim_{r\to\infty} \prod_{k=1}^{r} (1 - (\Lambda(u_k) - \Lambda(u_{k-1})))$$

is the product integral [8].

2.2 Censoring and truncation

Survival data possess a special feature of censoring, compared to other statistical data. Censoring is used when the survival time is not known exactly, the event is only known to have occurred within some time interval. There are several types of censoring: right, left and interval. In biomedical applications, right censoring is the most common type of censoring. It occurs when the survival time is incomplete on the right-hand side of the follow-up period, i.e. the study ends before all patients experience the event or a patient is lost to follow-up (dies due to reasons other than the event of interest, withdraws from the study, moves to another city, etc.).

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) survival times and $C_1, C_2, \ldots C_n$ be i.i.d. censoring times. The lifetime X_i of the i-th individual will be known if, and only if, $X_i < C_i$. If $C_i < X_i$ the event time will be censored at C_i. Thus it is convenient to represent the survival experience of a group of patients by the pairs of random variables (T_i, δ_i) where $T_i = min(X_i, C_i)$, $\delta_i = \mathrm{I}(X_i < C_i)$ and I is an indicator of the event's occurring, having value one if the event occurs, and zero otherwise.

Another feature, common in survival data, is a truncation. Truncation occurs when only those individuals whose event time lies within a certain time interval (T_L, T_R) are observed. For left truncation, $T_R = \infty$ in case of right truncation, $T_L = 0$. Individuals, whose event time is not in this interval, are not observed and no information on these subjects is available. This is in contrast to censoring where there is at least partial information available on each patient. When data are truncated, a conditional distribution has to be used in constructing the likelihood (see [10]).

A critical assumption for the likelihood construction is the independence of lifetimes and censoring times. Censoring is said to be independent if the failure rates that apply to individuals on trial at each time $t > 0$ are the same as those that would have applied had there been no censoring [8]. Thus the requirement is that at each time t

$$\lim_{\Delta t\to 0} \frac{P(T \in [t, t+\Delta t) \mid T \geq t)}{\Delta t} = \lim_{\Delta t\to 0} \frac{P(T \in [t, t+\Delta t) \mid T \geq t, Y(t) = 1)}{\Delta t},$$

where $Y(t) = 1$ indicates that the individual is at risk of failure at time t (has neither failed nor been censored prior to t).

2.3 Counting processes and martingales

An alternative approach to develop inference procedures for censored data involves counting processes. A counting process $N = \{N(t), t \geq 0\}$ is a stochastic process with $N(0) = 0$ whose value at time t counts the number of events that have occurred in the interval $(0,t]$. The sample paths (realizations) of N are nondecreasing, right-continuous step functions that jump whenever an event (or events) occur. In the counting process formulation, the pair of variables (T_i, δ_i) introduced in Section 2.2 is replaced with the pair of functions $N_i(t), Y_i(t), i = 1,\ldots,n$ where

$$N_i(t) = no.\ of\ \ events\ observed\ in\ [0,t]\ \ for\ unit\ i$$

$$Y_i(t) = \begin{cases} 1 & unit\ \ i\ is\ at\ risk\ at\ time\ \ t, \\ 0 & otherwise. \end{cases}$$

$N_i(t)$ is a counting process, while $Y_i(t)$ is a predictable process, i.e. a process whose value at time t is known infinitesimally before t at time t. This process has left-continuous sample paths. Right-censored survival data are included in this formulation as a special case:

$N_i(t) = \mathrm{I}(T_i \leq t, \delta_i = 1)$ and $Y_i(t) = \mathrm{I}(T_i \geq t)$.

To deal with all on-study information of each patient, a term history (or filtration) is used. A history, denoted $\{\mathsf{F}_t, t \geq 0\}$ is a σ-algebra generated by N_i and Y_i:

$$\mathsf{F}_t = \sigma(N_i(s), Y_i(s^+), i = 1,\ldots,n; 0 \leq s \leq t),$$

where $Y_i(s^+) = \lim_{u\to s^+} Y_i(u)$.

Thus F_t contains the information up to and including time t. The information in F_t increases with increasing time on study, i.e. $\mathsf{F}_s \subseteq \mathsf{F}_t$ for $s \leq t$ [4]. Let $dN_i(t)$ denote the increment of over the time interval $[t, t+dt)$:

$$dN_i(t) = N_i((t+dt)^-) - N_i(t^-).$$

For each $t > 0$ let

$$\mathsf{F}_{t^-} = \sigma(N_i(s), Y_i(s), i = 1,\ldots,n; 0 \leq s < t)$$

denote the full history of the processes $N_i(s)$, $Y_i(s)$, $i = 1,\ldots,n$ up to but not including t. Then (see [4]):

$$\mathrm{E}(dN_i(t) \mid \mathsf{F}_{t^-}) = Y_i(t)\lambda_i(t)dt,$$

where $\lambda_i(t)$ is the hazard function. The process

$$\Lambda_i(t) = \int_0^t Y_i(s)\lambda_i(s)ds, \quad t \geq 0,$$

is called the intensity process. At each fixed t, this process is a random variable which approximates the number of jumps by N_i over $(0,t]$. In fact, $\mathrm{E}N_i(t) = \mathrm{E}\Lambda_i(t)$ and thus [4].

$$\mathrm{E}(N_i(t) \mid \mathsf{F}_{t^-}) = \mathrm{E}(\Lambda_i(t) \mid \mathsf{F}_{t^-}) = \Lambda_i(t)$$

For any given i define the process

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda_i(s)ds, \quad t \geq 0, i = 1,\ldots,n. \tag{1}$$

Equivalently, the process can be defined

$$M_i(t) = \int_0^t dM_i(s),$$

where

$$dM_i(t) = dN_i(t) - Y_i(t)\lambda_i(t)dt.$$

It can be seen that $E(dM_i(t)|F_{t-}) = 0$ for all t and $E(M_i(t)|F_s) = M_i(s)$ for all $s \leq t$ [4]. A process that satisfies these (equivalent) conditions is a martingale. According to the Doob-Meier decomposition theorem (see [4]), any counting process may be uniquely decomposed as a sum of a martingale and a compensator C which is a predictable, right-continuous process with $C(0) = 0$. As an example, according to (1)

$$\begin{aligned} N_i(t) &= M_i(t) + \int_0^t Y_i(s)\lambda_i(s)ds \\ &= M_i(t) + \Lambda_i(t), \end{aligned} \qquad (2)$$

where $M_i(t)$ is the counting process martingale corresponding to $N_i(t)$ and $\Lambda_i(t)$ is the compensator of the counting process N_i with respect to the filtration F_{t-}. In terms of differential increments, the process (2) can be equivalently written as

$$dN_i(t) = dM_i(t) + Y_i(t)\lambda_i(t)dt.$$

The approach using martingale methods is very useful in yielding results for censored and truncated data, especially for calculating and verifying asymptotic properties of test statistics and estimators.

3. Non-parametric and semi-parametric models

A principle objective of the survival analysis focuses on estimation of basic quantities (the survival and hazard function) based on censored data. To analyze survival data parametrically, assumptions about the distribution of the failure times would have to be made. To avoid such assumptions, it is common to use non-parametric models. The simplest non-parametric estimate of a distribution function is the empirical distribution function

$$F_n(x) = \frac{no.\ of\ \ sample\ values \leq x}{n},$$

when a continuous distribution is estimated by a discrete one. For an uncensored sample of n distinct failure times, the empirical survival function is then estimated by $S_n(t) = 1 - F_n(t)$. The only problem with this approach is the censoring - it is not taken into account in standard statistical methods. Important steps in the development of appropriate methods were done by Kaplan and Meier [9] and Cox [3].

3.1 Kaplan-Meier and Nelson-Aalen estimators

The Kaplan-Meier estimator (called also the product-limit estimator) estimates the survival function by

$$\hat{S}(t) = \prod_{i:t_i \leq t}\left(1 - \frac{d_i}{R_i}\right),$$

where d_i there are events observed at time t_i and R_i is the number of individuals still at risk at time t_i (uncensored survivors just before t_i). The variance of the estimator can be estimated using Greenwood's formula (see [11]):

$$\hat{V}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{i:t_i \leq t} \frac{d_i}{R_i(R_i - d_i)}.$$

The product $\hat{\Lambda}(t) = -\ln(\hat{S}(t))$ can also be used to estimate the cumulative hazard function:

An alternative estimator of the cumulative hazard function was proposed by Nelson in 1972 [12] and rediscovered by Aalen in 1978 [1]:

$$\tilde{\Lambda}(t) = \sum_{i:t_i \leq t} \frac{d_i}{R_i}.$$

The variance of the Nelson-Aalen estimator was estimated by Aalen using counting process techniques and is given by

$$\tilde{V}(\tilde{\Lambda}(t)) = \sum_{i:t_i \leq t} \frac{d_i}{R_i^2}.$$

Based on the Nelson-Aalen estimator of the cumulative hazard function, an alternative estimator of the survival function becomes

$$\tilde{S}(t) = \exp(-\tilde{\Lambda}(t)).$$

Suppose now that n individuals from a homogeneous population are put on a study at time 0. Let N_i be the counting process and Y_i be the at-risk process of the i-th individual, as described in Section 2.3. Let

$N.(t) = \sum_{i=1}^n N_i(t)$ and $Y.(t) = \sum_{i=1}^n Y_i(t), \ 0 < t < \infty.$

$N.(t)$ denotes the total number of observed failures in the interval, while $Y.(t)$ is the number of individuals in the entire study group that are at risk at time t. The Nelson-Aalen estimator of the cumulative hazard can be written in the counting process notation as

$$\tilde{\Lambda}(t) = \int_0^t \frac{J(u)}{Y.(u)} dN.(u),$$

where $J(u) = I(Y.(u) > 0$ with the convention that 0/0 is interpreted as 0 [8]. The Kaplan-Meier estimator of the survival function is then

$$\hat{S}(t) = \prod_{u \leq t}(1 - d\tilde{\Lambda}(u)).$$

A common interest is to compare two or more samples, i.e. to test whether there is a significant difference in survival experience of distinct groups of patients. Several generalizations of standard non-parametric tests have been developed to deal with censored and truncated data. The most common tests are the log-rank test, Gehan-Wilcoxon test and Peto-Peto test. For more information, see e.g. [10].

3.2 The Cox model

In clinical studies we typically examine the association of several risk factors with the occurrence of the event of interest. Various patients' characteristics may be associated with patients' survival experience (e.g. age, sex, blood pressure, ...). The Cox proportional hazards model has become a popular approach to modeling covariate effects on survival. In this model the intensity process (hazard) for the i-th subject is

$$\lambda_i(t) = Y_i(t)\lambda_0(t)\exp(\beta^T X_i),$$

where $Y_i(t)$ is the at-risk process, λ_0 is the baseline hazard (common to all individuals in the study population), X_i is the vector of covariates of individual i and β is a vector of unknown regression parameters. In this model, the ratio of hazard functions of two individuals is constant (the baseline hazard $\lambda_0(t)$ is canceled out), thus the temporal effect is separated from the effect of the covariates. Estimation of the regression coefficients is based on maximizing of the partial likelihood function, which was introduced by Cox in 1972 [3].

The partial likelihood function for β reads

$$L(\beta)=\prod_{j=1}^{k}\frac{\exp(\beta^T X_j)}{\sum\limits_{l\in R(t_j)}\exp(\beta^T X_l)},$$

where $t_1 < ... < t_k$ are the uncensored failure times of the study group, $R(t_j)$ is the set of subjects at risk of failure at time t_j (just prior to time t_j), and X_j denotes the covariate vector for an individual failing at t_j. The partial likelihood function is treated as a standard likelihood, and inference is carried out by usual means.

4. Multivariate survival analysis

In most clinical applications the univariate survival analysis assumes that the observed survival times are mutually independent (i.i.d. failure times). In practice, however, dependence can occur for very different kinds of data, e.g. survival of twins or other several individuals, similar organs, recurrent events or multi-state events. The multivariate survival analysis covers the field where independence between survival times cannot be assumed. According to [7], the various approaches to analyzing multivariate survival data fall into four main categories: multi-state models, frailty models, marginal modeling and non-parametric methods. The data structure should be considered as well. The data can be parallel (where the number of failures is fixed by the design of the study) or longitudinal (where the number of failures is random for each object under study). The data sets are classified into six types: several individuals, similar organs, recurrent events, repeated measurements, different events and competing risks. Relation of the data types to the two main approaches of analysis (multi-state and frailty models) is described in Table 1. Only these two approaches to analyzing multivariate survival data are presented in this paper. For more information on marginal and non-parametric methods, see e.g. [7], [16], or [8].

Tab. 1. Overview of data types and approaches; x means relevant, blank not relevant. Adopted from [7].

Type of data	Multi-state	Frailty
Several individuals	x	x
Similar organs	x	x
Recurrent events	x	x
Repeated measurements		x
Different events	x	
Competing risks	x	

4.1 Competing risk and multi-state models

Multi-state models are commonly used for describing the development of longitudinal data. They model stochastic processes, which at any time point occupy one of a set of discrete states. In medicine, the states can be e.g. healthy, diseased, and dead. A change of state is called a transition. The competing risk model is an example of multi-state modeling. In competing risks, various causes of death “compete” in the life of patient, and occurrence of one event precludes occurrence of the other events. There are generally three areas of interest in the analysis of competing risks [8]:

1. Studying the relationship between a vector of covariates and the rate of occurrence of specific types of failure.
2. Analyzing whether patients at high risk of one type of failure are also at high risk for others.
3. Estimating the risk of one type of failure after removing others.

Suppose that individuals under study can experience any one of m distinct failure types. For each individual, the underlying failure time T and a covariate vector X are known. The overall hazard function at time t is

$$\lambda(t,X)=\lim_{\Delta t\to 0^+}\frac{P(t\le T<t+\Delta t\,|\,T\ge t,X)}{\Delta t}.$$

To model competing risks, a cause-specific hazard function is considered:

$$\lambda_j(t,X)=\lim_{\Delta t\to 0^+}\frac{P(t\le T<t+\Delta t, J=j\,|\,T\ge t,X)}{\Delta t}$$

for $j = 1,...,m$ is a random variable representing the type of failure, and $t > 0$. In words, $\lambda_j(t,X)$ specifies the rate of type j failures, given and in the presence of all other failure types [8]. If only one of the failure types can occur, then

$$\lambda(t,X)=\sum_{j=1}^{m}\lambda_j(t,X)$$

due to the law of total probability.

It is possible to calculate the Kaplan-Meier estimator for each type of failure separately, but it is difficult to give this a survival function interpretation and therefore this is not recommended [7]. Instead, generalizations of the Kaplan-Meier and Nelson-Aalen estimators can be made (see e.g. [8]). The generalized estimator includes all causes of failure and is usually denoted the Aalen-Johanson estimator.

The Cox model for the cause-specific hazard functions can be considered:

$$\lambda_j(t,X)=\lambda_{0j}(t)\exp(\beta_j^T X),\quad j=1,\ldots,m.$$

Both the baseline hazards λ_{0j} and the regression coefficients β_j vary arbitrarily over the failure types. Estimation and comparison of the coefficients β_j can be conducted by applying asymptotic likelihood techniques individually to the m factors.

A traditional approach to multi-state models is based on the Markov models. Consider first a homogeneous population with no covariates. Let $A(t)$ be the state occupied at time $t \ge 0$ with probability model of $A(t)$ being the Markov process. The individuals under study move among $m > 1$ discrete states.

If a randomly chosen individual is in state *i* at time t^-, the transition rate (or intensity) from *i* to *j* at time *t* is given by

$$d\Lambda_{ij}(t) = P[A(t^- + dt) = j \mid A(u), 0 \le u < t, A(t^-) = i]$$

$$= P[A(t^- + dt) = j \mid A(t^-) = i], \quad t > 0,$$

which holds for all $A(u)$, $0 \le u < t$ with $A(t^-) = i$ and $i, j \in \{1,\dots,m\}$, $j \neq i$. The process is memoryless in that only the current state occupied is relevant in specifying the transition rates [8]. In the continuous case, $d\lambda_{ij}(t) = \lambda_{ij}(t)dt$ for all $i,j = 1,\dots,m$ so that $\lambda_{ij}(t)$, $i \neq j$ is the continuous-time intensity function for *i*-to-*j* transitions. Estimation of the cumulative intensity functions $\lambda_{ij}(t)$ proceeds as follows [8]: consider a possibly right-censored sample of *n* individuals. For $k = 1,\dots,n$, let $N_{ijk}(t)$ be the right continuous process that counts the number of observed direct *i*-to-*j* transitions for *k*-th individual, $i,j = 1,\dots,m, i \neq j$. Let $Y_{ik}(t)$ be the corresponding at-risk process. Define the filtration process as

$$\mathrm{F}_t = (N_{ijk}(t), Y_{ik}(u^+), 0 \le u \le t),$$

For $k = 1,\dots,n$; $i,j = 1,\dots,m$ and suppose that censoring is independent, so that

$$P(dN_{ijk}(t) = 1 \mid \mathrm{F}_{t^-}) = Y_{ik}(t)d\Lambda_{ij}(t),$$

which must hold for all *i,j,k* and $t > 0$. The Nelson-Aalen estimator of $\lambda_{ij}(t)$ is then given by

$$d\hat{\Lambda}_{ij}(t) = \frac{dN_{ij}.(t)}{Y_i.(t)}$$

for all $i \neq j$.

When the vector of covariates *X* is present, the continuous-time modulated Markov model can be specified for the underlying intensity function

$$\lambda_{ijk}(t) = \lim_{dt \to 0^+} \frac{P(A_k(t^- + dt) = j \mid A_k(t^-) = i, X)}{dt}.$$

Parametric and semi-parametric models for λ_{ijk} are obtained analogously as earlier and may be found in [8].

4.2 Frailty models

Frailty models represent an extension of the Cox proportional hazards model. The concept of frailty provides a way to introduce random effects into the model to account for association (correlation) and unobserved heterogeneity. This heterogeneity may be difficult to assess but is nevertheless of a great importance. The frailty is an unobserved random factor that modifies multiplicatively the hazard function of an individual or a group of individuals. The key idea of these models is that individuals most "frail" die earlier than the others [16]. The frailty models are relevant to lifetimes of several individuals, similar organs and repeated measurements. They are not generally relevant for the case of different events [7].

First, bivariate models will be considered. Let

$$S_{12}(t_1, t_2) = P(T_1 \ge t_1, T_2 \ge t_2)$$

be the joint survival function for the two survival times T_1 and T_2 where $S_{12}(t,t)$ is the probability that both subjects under study will be alive at time *t*.

The marginal survival functions are then

$$S_1(t_1) = P(T_1 \ge t_1) = S_{12}(t_1, 0)$$

$$S_2(t_2) = P(T_2 \ge t_2) = S_{12}(0, t_2).$$

If T_1 and T_2 are independent, $S_{12}(t_1,t_2) = S_1(t_1)S_2(t_2)$. The joint hazard function is

$$\lambda_{12}(t_1,t_2) = \lim_{\Delta t \to 0^+} \frac{P(T_1 \in [t_1, t_1 + \Delta t), T_2 \in [t_2, t_2 + \Delta t) \mid T_i \ge t_i)}{\Delta t^2},$$

and the marginal hazards are

$$\lambda_i(t_i) = \lim_{\Delta t \to 0^+} \frac{P(T_i \in [t_i, t_i + \Delta t) \mid T_i \ge t_i)}{\Delta t},$$

For $i = 1,2$. To address heterogeneity in the survival times it is assumed that the lifetimes are conditionally independent, i.e. T_1 and T_2 are independent given the random effect *Z* called frailty:

$$S_{12}(t_1, t_2 \mid Z) = S_1(t_1 \mid Z) S_2(t_2 \mid Z).$$

Usually, the frailty is assumed to act multiplicatively on the hazard, so that

$$\lambda_i(t_i) = Z\lambda_{0i}(t_i) \quad and \quad S_i(t_i \mid Z) = S_{0i}(t_i)^Z$$

for some baseline hazard $\lambda_{0i}(t)$ and baseline survival function $S_{0i}(t)$ (when known covariates X_i are present, the hazard may be expressed as

$$\lambda_{0i}(t_i) = \lambda_0(t_i)\exp(\beta^T X_i)$$

through the Cox regression model). Under the assumption of multiplicative frailty, the cumulative hazards are

$$\Lambda_i(t_i) = Z\Lambda_{0i}(t_i).$$

The conditional joint survival function is then

$$S_{12}(t_1, t_2 \mid Z) = S_{01}(t_1)^Z S_{02}(t_2)^Z$$

$$= \exp(-Z\Lambda_{01}(t_1))\exp(-Z\Lambda_{02}(t_2))$$

$$= \exp(-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_2))).$$

As the frailty *Z* is an unobserved effect, it needs to be 'integrated out' of the survival function. This is done by the Laplace transform, which is defined for a random variable *Z* as

$$L_g(s) = \int \exp(-sz)g(z)dz = \mathrm{E}(\exp(-sZ)).$$

Where $g(z)$ is the probability density of *Z*. For the bivariate survival function thus

$$S_{12}(t_1,t_2) = \int_0^\infty S_{12}(t_1, t_2 \mid Z)g(z)dz$$

$$= \int_0^\infty \exp(-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_2)))g(z)dz$$

$$= L_g(\Lambda_{01}(t_1) + \Lambda_{02}(t_2)),$$

where the Laplace transform of $g(z)$ is evaluated at

$$s = \Lambda_{01}(t_1) + \Lambda_{02}(t_2).$$

In many applications, the frailty *Z* is assumed to follow some distribution with the explicit Laplace transform. A standard (and most widely used) distribution for frailty is the gamma distribution.

The random variable Z is gamma distributed with parameters k and Θ ($Z \sim \Gamma(k, \Theta)$), if its probability density function is

$$g(z) = \frac{\theta^k z^{k-1} \exp(-\theta z)}{\Gamma(k)}, \quad k,\theta > 0 \ and \ z > 0,$$

With $\mathrm{E}Z = \dfrac{k}{\theta}, \quad \mathrm{var}Z = \dfrac{k}{\theta^2}.$

The gamma function in the denominator of the probability density function is defined as

$$\Gamma(k) = \int_0^\infty u^{k-1} \exp(-u) du, \qquad for \ k > 0.$$

It satisfies $\Gamma(k+1) = k\Gamma(k).$

The gamma distribution fits very well to failure data and is also convenient from computational and analytical point of views [19].

Suppose the common frailty component Z has a gamma distribution with parameters $k = \Theta = 1/\sigma^2$. The Laplace transform of the gamma density is then

$$L(s) = \left(\frac{\theta}{\theta + s}\right)^k,$$

Which leads to (see [7])

$$S_{12}(t_1, t_2) = \left(\frac{1}{1 + \sigma^2 \Lambda_{01}(t_1) + \sigma^2 \Lambda_{02}(t_2)}\right)^{1/\sigma^2}.$$

To extend the bivariate model to a multivariate one, consider a set of clustered data where for the j-th individual in the i-th group (or cluster) there are the observation times t_{ij} and the vector of covariates X_{ij}. The assumption is, again, that given X_{ij} and a random effect Z_i the m_i lifetimes in group i are independent. Thus the joint distribution of these lifetimes given Z_i is the product of the marginal distributions given Z_i. The marginal hazards then satisfy

$$\lambda_{ij}(t_{ij} \mid X_{ij}, Z_i) = Z_i \lambda_{0ij}(t_{ij} \mid X_{ij}).$$

When the hazards are modeled using the Cox proportional hazards,

$$\lambda_{0ij}(t_{ij} \mid X_{ij}) = \lambda_0(t_{ij}) \exp(\beta^T X_{ij}).$$

If the cluster-specific random effects Z_i have independent gamma distributions, then the unconditional survival for the m_i lifetimes in cluster is

$$S_i(t_i, X_i) = \int_0^\infty \prod_j S_{ij}(t_{ij} \mid X_{ij}, Z_i) g(z_i) dz_i,$$

where $t_i = (t_{i1}, t_{i2}, \ldots, t_{im_i})^T, \ X_i = (X_{ij})_{m_i \times n}$

This can be solved using the Laplace transform (see [14])

$$S(t_i, X_i) = \left(\frac{1}{\psi}\right)^{1/\sigma^2},$$

Where

$$\psi = 1 + \sigma^2 \Lambda_0(t_{i1}) \exp(\beta^T X_{i1}) + \ldots + \sigma^2 \Lambda_0(t_{im_i}) \exp(\beta^T X_{im_i}).$$

Different choices of distribution for the frailty Z are possible, e.g. the family of positive stable distributions or the PVF (power variance function) family. For more information about these, see [7]. The frailty Z may also be treated non-parametrically. Although it is desirable to have completely non-parametric estimate of the survival function, the estimates are mathematically complicated and are not of major importance [7].

Statistical models that use counting process notation and are convenient for these types of analyses are slightly different from those used until now. In the previously used models, the intensity process $\lambda(t)$ at the follow-up time t given the covariates X was

$$\lambda(t)dt = P(dN(t) = 1 \mid N(s), 0 \le s < t, X).$$

In this expression it is assumed that jumps in N are of a unit size only. However, recurrent and correlated failure time data include jumps of a size greater than one (more than one event can be recorded for an individual at a specific follow-up time). Thus it is natural to model the mean jump in N across time:

$$d\Lambda(t) = \mathrm{E}(dN(t) \mid N(s), 0 \le s < t, X).$$

in the cumulative intensity process. The Cox-type model for the intensity process is then

$$d\Lambda(t) = d\Lambda_0(t) \exp(\beta^T X).$$

For more details, see [8].

5. Conclusion

The survival analysis is a collection of specific statistical methods. In this paper, a short overview of these methods was presented. The standard univariate models were extended to multivariate models dealing with parallel and longitudinal data. The two major multivariate concepts were introduced: multi-state and frailty models.

Acknowledgments

The paper has been supported by the project SVV-2010-265513.

References

[1] Aalen O. O.: Nonparametric Inference for a Family of Counting Processes. Annals of Statistics 6 (1978), 701726.

[2] Andersen P. K., Gill R. D.: Cox's regression model for counting processes: a large sample study. The Annals of Statistics 10 (1982), 11001120.

[3] Cox D. R.: Regression Models and Life-Tables. Journal of the Royal Statistical Society B 34 (1972), 187220.

[4] Fleming T. R., Harrington D. P.: Counting Processes and Survival Analysis. John Wiley & Sons, New York, 1991.

[5] Fürstová J.: Multivariate Methods of Survival Analysis. Doktorandský den 2010, Matfyzpress, Praha, 2010.

[6] Gill R. D.: Understanding Cox's regression model: a martingale approach. Journal of the American Statistical Association 79 (1984), 441447.

[7] Hougaard P.: Analysis of Multivariate Survival Data. Springer, New York, 2000.

[8] Kalbfleisch J. D., Prentice R. L.: The Statistical Analysis of Failure Time Data. John Wiley & Sons, New York, 2002.

[9] Kaplan E. L., Meier P.: Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53 (1958), 457481.

[10] Klein J. P., Moeschberger M. L.: Survival Analysis. Techniques for Censored and Truncated Data. Springer, New York, 2003.

[11] Miller R. G., Gong G., Muñoz A.: Survival Analysis. John Wiley & Sons, New York, 1998.

[12] Nelson W.: Theory and Applications of Hazard Plotting for Censored Failure Data. Technometrics 14 (1972), 945965.

[13] Prentice R. L., Williams B. J., Peterson A. V.: On the regression analysis of multivariate failure time data. Biometrika 68 (1981), 373379.

[14] Rodríguez G.: Multivariate Survival Models. available at http://data.princeton.edu/, cited on April 10, 2010.

[15] Self S. G., Prentice R. L.: Commentary on Andersen and Gill's “Cox's regression model for counting processes: a large sample study”. The Annals of Statistics 10 (1982), 11211124.

[16] Therneau T. M., Grambsch P. M.: Modeling Survival Data. Extending the Cox Model. Springer, New York, 2000.

[17] Vaupel J. W., Manton K. G., Stallard E.: The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16 (1979), 439454.

[18] Wei L. J., Lin D. Y., Weissfeld L.: Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association 84 (1989), 10651073.

[19] Wienke A.: Frailty Models in Survival Analysis. Habilitation. Martin-Luther-Universität Halle-Wittenberg, 2007. available at http://sundoc.bibliothek.uni-halle.de/habil-online/

Contact
Mgr. Jana Fürstová
Faculty of Medicine and Dentristy
Palacky University
Tř. Svobody 8
771 26 Olomouc
Czech Republic
e-mail:jana.furstova@email.cz