Journal of Statistical Planning and Inference 151-152 (2014) 1–16

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Review

An overview of semiparametric models in survival analysis

Shaojun Guo [a], Donglin Zeng [b,*]

[a] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, People's Republic of China
[b] Department of Biostatistics, University of North Carolina, United States

ARTICLE INFO

Article history:
Received 12 April 2013
Received in revised form
15 October 2013
Accepted 17 October 2013
Available online 28 October 2013

Keywords:
Proportional hazards model
Partial likelihood
Proportional odds model
Transformation model
Frailty model
Right censoring
Interval censoring
Nonparametric maximum likelihood
Counting process
Empirical process

ABSTRACT

We provide an overview of semiparametric models commonly used in survival analysis, including proportional hazards model, proportional odds models and linear transformation models. The applications of these models to different types of censored data, either univariate or multivariate survival analysis, are given. For each case, inference procedures using censored observations are discussed.

Contents

* Corresponding author. Tel.: +1 919 966 7273.
E-mail addresses: guoshaoj@amss.ac.cn (S. Guo), dzeng@email.unc.edu (D. Zeng).

0378-3758/$ - see front matter
http://dx.doi.org/10.1016/j.jspi.2013.10.008

1. Introduction

In survival analysis, one of the most important goals is to estimate the association between risk factors and time-to-event and make future prediction of subject's survival probabilities. One major challenge in achieving this goal is that censored data are usually observed in studying time-to-event, where only partial information instead of precise measurement of time-to-event is available for some subjects. Since the seminal paper of Cox (1972) on proportional hazards model, numerous semiparametric models have been developed to analyze censored data in survival analysis, including proportional odds model and linear transformation models as well as their generalization from univariate survival analysis to multivariate survival analysis. At the same time, inefficient and efficient inference procedures have been proposed to derive the estimation of model parameters which are used to predict future survival probabilities. These procedures have been largely helped by the substantial development of mathematical theories including counting process theory (Fleming and Harrington, 1991) and empirical process theory (van der Vaart and Wellner, 1996).

In this paper, we aim to provide an overview of semiparametric models in survival analysis as well as theoretical development for these models. Since the work in this area is very extensive (maybe the most extensive area in statistics), our review is never meant to be complete and comprehensive. The paper is organized as follows: in Section 2, we describe the challenge of data structure specific to survival analysis. We then start to review semiparametric models for each type of survival data, including univariate survival data with right censored observations (Section 3), multivariate survival data with right censored observations (Section 4), and interval censored data and joint analysis of multiple outcomes (Section 5). Two real examples are analyzed in Section 5 to illustrate the power of semiparametric models. In Section 7, we conclude the paper by further listing some other semiparametric models and work in survival analysis.

2. Challenges in analysis of survival data

One of the major features that distinguishes the statistical analysis of survival data from studying other outcomes such as continuous or categorical outcomes is the presence of censoring. For censored subjects, only partial information instead of accurate measurement of survival time is observed. Usually, the censoring mechanism is classified into either right censoring or interval censoring (Kalbfleisch and Prentice, 2002) as follows: let T be the survival time of interest. For right-censored data, we either observe T or observe T greater than a time C. Mathematically, if we let C be potential censoring time, then the observed data can be expressed as $(Y=\min(T,C),\ \Delta=I(T\le C))$, where $I(\cdot)$ is the indicator function and Δ is called censoring indicator. For interval censored data, one only observes that T belongs to an interval, denoted by $[L,U]$, and the observed data can be summarized as $(L, U, I(T\le L), I(T\le U))$. When $L=U$, i.e., T is observed to be less or larger than L, this special interval censored data are called current status data. One key difference between right-censored data and interval censored data is that for the former, there is still some chance that we can accurately measure the survival time T. Therefore, right-censored data should provide more information regarding T as compared to interval censored data.

In survival analysis, another complication is different nature of outcome of interest. Typical time-to-event is a single outcome, such as patient's cancer survival or time to disease. However, multivariate survival outcomes are often investigated in many studies. They can be the survival events from a clustered group of subjects (clustered survival time), different types of survival outcomes from the same subject (multiple types of survival events), difference cause-specific survival time (competing risks), the repeated occurrence of the same type of event in the same subject (recurrent event), and the presence of multiple outcomes of which survival event is one (joint outcomes). In addition to event outcomes, another level of the complexity in survival analysis involves time-dependent covariates and time-varying effects. Furthermore, in practice, different study designs may be adopted in the study of time-to-event. All these pose very different challenges in survival analysis from analysis of non-censored observations.

Before introducing all kinds of semiparametric models, we list a few real examples below to illustrate different data structure in survival analysis.

Example 1 (*Lung cancer study*). This study was from the Veterans Administration lung cancer study group (Prentice, 1973; Kalbfleisch and Prentice, 2002). In a subgroup of 97 patients without prior therapy, survival times ranges from 1 to 587 days and 6 of them are censored. This is a typical survival analysis with right-censored data.

Example 2 (*rhDNase study*). A randomized clinical trial was conducted to assess the efficacy of rhDNae, a highly purified recombinant enzyme, in reducing exacerbations of respiratory symptoms for patients with cystic fibrosis. See Therneau and Hamilton (1997). In this clinical trial, a total of 321 patients were assigned to rhDNase and 324 were assigned to placebo. By the end of follow-up (approximately 170 days), 65 patient in rhDNase group experienced 1 exacerbation and 39 experienced at least 2 exacerbations. In the placebo group, 97 patients experienced 1 exacerbation and 42 experienced at least 2 exacerbations. The outcome of interest in this study was exacerbation so it belongs to recurrent event category as discussed earlier.

Example 3 (*Lung tumor study*). Hoel and Walberg (1972) provided a set of data for 144 RFM mice in tumorigenicity experiment that involved lung tumors. The goal of this study was to identify whether a suspected environment accelerated the time until tumor onset in experimental animals. The data consisted of the death time or sacrifice time of each animal measured in days and an indicator whether lung tumor was present or absent at time of death. Although the time to tumor onset was of interest, it was only known to be less than or greater than the observed time of death or sacrifice. These data belong to the category of the interval censored data, specifically, the current status data.

Example 4 (*Colon cancer study*). In one colon cancer study (Lin, 1994), the investigators wished to study the efficacy of adjuvant therapy on recurrent of cancer and death for patients with resected colon cancer. One goal of the study was to characterize the dependence between cancer recurrence and death. In the study, 315, 310 and 304 with stage C disease received observation, levamisole alone and levamisole combined with 5-fluorouracial, respectively. For each group, there were 155 patients in the observation group, 144 in the levamisole alone group and 103 in levamisole combined with 5-fluorouracil group who had cancer recurrence and the number of deaths from each group were in turn 114, 109 and 78. Since both survival outcomes of cancer recurrence and death were of interest, the analysis needs to account for multiple outcomes and the semicompeting risk nature due to death.

Example 5 (*Diabetic retinopathy study*). The well-known Diabetic Retinopathy study (Huster et al., 1989) was conducted to assess the effectiveness of laser photocoagulation in delaying visual loss among patients with diabetic retinopathy. Either eye of each patient was randomly selected to receive the laser treatment while the other eye was used as a control. The failure time of interest is the time to visual loss as measured by visual acuity less than 5/200. Since the event times from both eyes were correlated, the data presented from this study are clustered survival time.

For data with specific censoring mechanisms and univariate or multivariate nature, different models and inference procedures are needed to analyze data. From the next section, we will review some commonly used models and methods for each category of survival data. Specifically, we will first focus on univariate and right censored data; we will then discuss multivariate survival analysis; finally, we will review the work on interval censored data and joint analysis.

3. Univariate analysis with right censored data

3.1. Notation

Throughout this subsection, for each subject i, T_i and C_i are the survival time and the censoring time, respectively. Furthermore, we denote $Z_i(t)$ as the covariates which may depend on the time t. In the presence of censoring, the data $\{T_i, Z_i\}_{i=1}^n$ are not completely observable. Instead, we observe the data $\{Y_i, \Delta_i, Z_i(\cdot),\ i=1,\ldots,n\}$, where $Y_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \le C_i)$, $i=1,\ldots,n$.

Counting process notation is also commonly used in survival analysis. Specifically, for each i, we denote $N_i(t) = \Delta_i I(Y_i \le t)$ as the counting process associated with the observed event time, and $R_i(t) = I(Y_i \ge t)$ as the at-risk process. We let τ be the duration of the study. In the following analysis, we assume that the censoring C is noninformative, that is, C is independent of the survival time T given the covariate $Z(\cdot)$. However, generalization to informative censoring C is also possible by using additional auxiliary covariates.

3.2. Cox's proportional hazards model

The most popular semiparametric model for data fitting is the Cox (1972) proportional hazards model, which is the cornerstone of modern survival analysis. Given the vector of covariates Z, this model is specified by a hazard function

$$\lambda(t|Z) = \lambda(t)\exp(\beta^T Z), \tag{1}$$

where β is a vector of unknown regression coefficients and $\lambda(t)$ is an unknown baseline hazard function. The covariate effects act multiplicatively on the hazard function, and the exponential of the coefficient β gives the constant hazard rate ratio for an increase of one unit for the covariate in question. Cox (1972) also generalized model (1) to the discrete time by assuming

$$\frac{\lambda(t|Z)\,dt}{1-\lambda(t|Z)\,dt} = \exp\left(\beta^T Z\right)\frac{\lambda(t)\,dt}{1-\lambda(t)\,dt}. \tag{2}$$

Model (2) is a logistic model and in the continuous case model (2) reduces to model (1). To estimate the regression coefficients, Cox (1972, 1975) introduced the partial likelihood principle to eliminate the infinite-dimensional baseline hazard function. Specifically, the partial likelihood for β is

$$L_n(\beta)=\prod_{i=1}^{n}\left[\frac{\exp(\beta^T Z_i)}{\sum_{j=1}^{n}R_j(Y_i)\exp(\beta^T Z_j)}\right]^{\Delta_i},$$

whose logarithm yields

$$l_n(\beta)=\sum_{i=1}^{n}\int_0^{\tau}\left\{\beta^T Z_i-\log\left[\sum_{j=1}^{n}R_j(t)\exp(\beta^T Z_j)\right]\right\}dN_i(t). \tag{3}$$

The objective function (3) does not involve the baseline function $\lambda_0(t)$ and is almost always strictly concave, yielding a stable estimation. Let its solution be $\widehat{\beta}_n$ and the Fisher information matrix

$$\Sigma(\beta)=\int_0^{\tau}\left\{s^{(2)}(\beta,t)-\left[\frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}\right]^{\otimes 2}\right\}s^{(0)}(\beta,t)\lambda(t)\,dt,$$

where $s^{(k)}(\beta,t)=E[Z^{\otimes k}R(t)\exp(\beta^T Z)]$, $k=0,1,2$. Then under some regularity conditions, $\sqrt{n}(\widehat{\beta}_n-\beta)$ is asymptotically normal with mean 0 and covariance matrix $[\Sigma(\beta)]^{-1}$. Furthermore, $\widehat{\beta}$ is semiparametrically efficient in terms that its asymptotic variance is minimal among all locally regular estimators. The asymptotic properties were established in a seminal paper of Anderson and Gill (1982), where they extended the Cox proportional hazard model to general counting processes and also established the asymptotic properties of the associated Breslow (1972) estimator of the cumulative baseline hazard function via an elegant counting-process martingale theory (Fleming and Harrington, 1991).

We remark that the partial likelihood is actually a profile likelihood or a special case of nonparametric maximum likelihood. Specifically, we maximize the log-likelihood over (β,Λ). Unfortunately, such a maximum does not exist if $\Lambda(\cdot)$ is only restricted to be absolutely continuous because we can always choose some function $\Lambda(\cdot)$ with fixed values at the Y_i while letting $\lambda(Y_i)$ go to infinity for some Y_i with $\Delta_i=1$. Instead, we restrict our maximization to the parameter space

$$\{(\beta,\Lambda):\beta\in\Theta,\Lambda(t)\text{ is an increasing step function with jumps only at }Y_i's\}$$

and let $\lambda(t)$ to be the jump size of $\Lambda(t)$ at time t, denoted by $\Lambda\{t\}$. The modified log-likelihood function is of the form

$$l_n(\beta,\Lambda)=\sum_{i=1}^{n}\left\{\Delta_i[\log\lambda\{Y_i\}+\beta^T Z_i]-\int_0^{\infty}R_i(s)\exp(\beta^T Z_i)\,d\Lambda(s)\right\}.$$

The maximum of l_n with respect to $\lambda\{Y_i\}$ at the failure time Y_i is obtained at

$$\widehat{\lambda}\{Y_i\}=\Delta_i\bigg/\left[\sum_{j=1}^{n}R_j(Y_i)\exp(\beta^T Z_j)\right].,\quad i=1,\ldots,n.$$

Substituting the above formula into the previous expression, we obtain the profile log-likelihood function for β is

$$\max_{\Lambda} l_n(\beta,\Lambda)=\sum_{i=1}^{n}\int_0^{\tau}\left\{\beta^T Z_i-\log\left[\sum_{j=1}^{n}R_j(t)\exp(\beta^T Z_j)\right]\right\}dN_i(t),$$

which is exactly the log-partial likelihood function.

The key assumption in this model is the proportional hazards model. Therefore, many methods have also been developed to check this assumption. The commonly used graphical and numerical ways include plot of logarithm of cumulative hazard functions, plot of Schoenfeld residuals, or incorporating interaction between time and covariates. If it is observed that violation of the proportional hazards assumption is likely, one remedy is to stratify the data into some subgroups and apply the model for each stratum. Another strategy to fix proportional hazards assumption is to include the time-varying covariates in the model. In this case, the covariates in model (1) are indexed by time and the setup of the partial likelihood functions is still applicable. However, time-varying covariates must be used with caution since they involve constructing a function of time that is usually not self-evident and may be suggested by biological hypothesis. In addition, Zeng and Lin (2007c) pointed out that adjustment of non-proportionality through the use of manufactured time-varying covariates in the form of $Zf(t)$ was restrictive and data driven, leading to untransparent effect of covariates on the survival time.

3.3. *Proportional odds model*

An alternative model to capture the non-proportionality is the proportional odds model (Bennett, 1983). Under this model, the hazard ratio between two sets of covariate values converges to unity rather than staying constant as time increases. This property is desirable when the initial effects of treatment, or the differences between stages of the disease at diagnosis, tend to diminish with time, and the survival probabilities of different groups of patients become more similar.

The survival function S_Z, given the vector of covariates Z, is parameterized by

$$-\log\left\{\frac{S_Z(t)}{1-S_Z(t)}\right\}=G(t)+\beta^T Z,$$

where G is an arbitrary baseline log-odds and β is a vector of regression coefficients.

Unlike the partial likelihood in Cox's model, the estimation and inference for the parameter β is much more challenging under the proportional odds model. Several contributions have been made by Dabrowska and Doksum (1988), Cheng et al. (1995), Murphy et al. (1997), Shen (1998) and so on. In particular, Murphy et al. (1997) demonstrated that under some regularity conditions, the nonparametric maximum likelihood approach still works and the nonparametric maximum likelihood estimator (hereafter NPMLE) was asymptotically semiparametric efficient in the Fisher sense. To be specific, Let $H(t)=\exp\{G(t)\}$ and $h(t)=dH(t)/dt$. Then, under right censoring, the (observed data) log-likelihood is

$$l_n(\beta,H)=\sum_{i=1}^{n}\{\Delta_i[\log h(Y_i)-Z_i^T\beta]-(1+\Delta_i)\log[H(Y_i)+\exp(-Z_i^T\beta)]\}.$$

As discussed in Cox's model, let $H(\cdot)$ to be an increasing step function with jumps only at the observed failure times and $h(t)$ to be the jump size of $H(t)$ at t, denoted by $h\{t\}$. Then the modified log-likelihood function is

$$l_n(\beta,H)=\sum_{i=1}^{n}\{\Delta_i[\log h\{Y_i\}-Z_i^T\beta]-(1+\Delta_i)\log[H(Y_i)+\exp(-Z_i^T\beta)]\}. \tag{4}$$

The maximum of (4) exists and can be solved by an iterative algorithm. Theoretically, Murphy et al. (1997) showed that this maximum profile likelihood estimator was consistent, asymptotically normal, and semiparametrically efficient. Furthermore, they demonstrated that the profile likelihood could be treated as a parametric likelihood and provided the asymptotic distribution for the profile likelihood ratio statistic. For more details, see Murphy et al. (1997).

Instead of the nonparametric maximum likelihood estimation, another method to estimate β is maximizing the marginal likelihood of the ranks (Pettitt, 1984). Since this marginal likelihood cannot be calculated explicitly for all β, Pettitt (1984) used an approximation based on a Taylor expansion on the logarithm of the marginal likelihood at $\beta=0$, however, the resulting estimator is biased and inconsistent. Lam and Leung (2001) employed the technique of importance sampling to express the marginal likelihood as an expectation with respect to some distribution. Numerically, the bias of their estimator was negligible and the corresponding standard errors were comparable with the asymptotically efficient profile likelihood estimator. However, their method is computationally intensive and the theoretical properties of the maximum marginal likelihood estimator have not been investigated.

3.4. Transformation models

Both the proportional hazards and proportional odds models belong to the class of linear transformation models, which relate an unknown transformation of the survival time linearly to covariates. Let T be the survival time and Z a corresponding vector of covariates. This model can be written as

$$H(T)=-\beta^T Z+\varepsilon, \tag{5}$$

where $H(\cdot)$ is an unknown monotone transformation function, ε is a random variable with a known distribution and is independent of Z, and β is a vector of unknown regression coefficients. If ε follows the extreme value distribution, model (5) becomes to the proportional hazard model, while if ε follows the standard logistic distribution, model (5) reduces to the proportional odds model.

Several papers proposed general estimators for the regression coefficients in model (5). Dabrowska and Doksum (1988) provided estimators in the two sample problem based on marginal likelihood of ranks and Markov chain Monte Carlo method. Their estimators suffer from severe bias under heavy censoring and the bias cannot be reduced by increasing the size of Monte Carlo simulation (Lam and Kuk, 1997). Cheng et al. (1995) derived inference procedures from a class of generalized estimating equations based on dichotomous variables of pairwise ranks. They adjusted censoring by the inverse weight of the Kaplan–Meier estimator for the survival function of the censoring variable, so the validity of their procedures relies on the assumption that censoring variable is independent of the covariates. Chen et al. (2002) mentioned that such an assumption was often too restrictive, even for randomized clinical trials. Instead, Chen et al. (2002) proposed an estimator using the martingale-based estimating equations and the estimating equations precisely become the Cox partial likelihood score equation for the proportional hazard model. Although Dabrowska and Doksum (1988), Cheng et al. (1995) and Chen et al. (2002) established the consistency and asymptotic normality for their estimators, none of the estimators is semiparametrically efficient. Recently, Kosorok et al. (2004) considered a class of frailty models for independent observations which results in a class of transformation models as one parameter extension of the proportional hazards model, and they further studied semiparametrically efficient estimation in their class.

Model (5) can be rewritten as

$$g(\Lambda_Z(t))=G(t)+Z^T\beta,$$

where $\Lambda_Z(t)$ denotes the cumulative hazards function of T given Z, and $g(\cdot)$ is a known function depending on the distribution of ε in (5). This formulation in terms of the cumulative hazards function allows one to incorporate time-dependent covariates in the transformation models. Specifically, to accommodate time-varying covariates, Zeng and Lin (2006) suggested to model the cumulative hazards function $\Lambda_Z(t)$ for T given $\{Z(s), s \le t\}$ directly. They proposed a broad class of transformation models, which assumes

$$\Lambda_Z(t) = G\left(\int_0^t \exp\{\beta^T Z(s)\}\, d\Lambda(s)\right), \tag{6}$$

where $G(\cdot)$ is a continuously differentiable and strictly increasing function, β is a vector of unknown regression parameters, and $\Lambda(\cdot)$ is an unspecified increasing function with $\Lambda(0)=0$. Two classes of transformation functions are very useful in practice: One is the class of Box–Cox transformations

$$G(x) = \{(1+x)^\rho - 1\}/\rho, \quad \rho \ge 0.$$

If $\rho = 0$, it corresponds to $G(x) = \log(1+x)$. The other is the class of logarithmic transformations:

$$G(x) = \log(1+rx)/r, \quad r \ge 0.$$

Here if $r=0$, we assume that $G(x) = x$. Interestingly, this class of semiparametric transformation models include two useful models we have discussed: proportional hazards and odds models. Moreover, if Z is time invariant, the transformation model (6) reduces to the linear transformation model (5).

Under model (6), Zeng and Lin (2006) proposed the nonparametric maximum likelihood approach to estimate Λ and β simultaneously. To be precise, the hazards rate for T_i given $\{Z_i(s), 0 \le s \le t\}$ is $R_i(t) \exp\{\beta^T Z_i(t)\}\lambda(t) G'(\int_0^t R_i(s) \exp\{\beta^T Z_i(s)\}\, d\Lambda(s))$, where $\lambda(t) = \Lambda'(t)$. Thus, the log-likelihood function concerning the parameters (β, Λ) is of the form

$$\sum_{i=1}^n \left\{ \int_0^\tau \log \lambda(t)\, dN_i(t) + \int_0^\tau \log G'\left(\int_0^t R_i(s) \exp\{\beta^T Z_i(s)\}\, d\Lambda(s)\right) dN_i(t) + \int_0^\tau \beta^T Z_i(t)\, dN_i(t) - G\left(\int_0^\tau R_i(s) \exp\{\beta^T Z_i(s)\}\, d\Lambda(s)\right) \right\}.$$

In the NPMLE approach, we let $\Lambda(\cdot)$ be an increasing step function with jumps only at the observed survival times and $\lambda(t)$ to be the jump size of $\Lambda(t)$ at t, denoted by $\Lambda\{t\}$. The modified log-likelihood becomes

$$\sum_{i=1}^n \left\{ \int_0^\tau \log \Lambda\{t\}\, dN_i(t) + \int_0^\tau \log G'\left(\int_0^t R_i(s) \exp\{\beta^T Z_i(s)\}\, d\Lambda(s)\right) dN_i(t) + \int_0^\tau \beta^T Z_i(t)\, dN_i(t) - G\left(\int_0^\tau R_i(s) \exp\{\beta^T Z_i(s)\}\, d\Lambda(s)\right) \right\}.$$

It is shown that the maximum of this function exists. Under certain regularity conditions, Zeng and Lin (2006) proved that the resulting NPMLEs for β are asymptotically normal and also efficient using the empirical process theory.

3.5. *Additive hazards model*

An additive hazards model takes the form

$$\lambda_Z(t) = \lambda(t) + \beta^T Z(t).$$

This model relates the effect of the covariate $Z(t)$ linearly to the hazards function $\lambda_Z(t)$. The parameter β denotes the risk difference while β represents the logarithm of the risk ratio in Cox's proportional hazards model. The model has been studied by Cox and Oakes (1984), and Lin and Ying (1994, 1995, 1997) among others.

Under right censoring, the log-likelihood function concerning the parameter (β, Λ) is

$$l_n(\beta, \lambda) = \sum_{i=1}^n \left[\int_0^\tau \log\{\lambda(t) + \beta^T Z_i(t)\}\, dN_i(t) - \int_0^\tau R_i(t)\{\lambda(t) + \beta^T Z_i(t)\}\, dt \right].$$

It is easily shown that the maximum of the above function over (β, λ) does not exist and the nonparametric maximum likelihood approach does not work either. Instead, Lin and Ying (1994) proposed the following estimating equation:

$$U_n(\beta) = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \overline{Z}(t)\}\{dN_i(t) - \beta^T Z_i(t) R_i(t)\, dt\},$$

where $\overline{Z}(t) = \sum_{i=1}^n Z_i R_i(t) / \sum_{i=1}^n R_i(t)$. Since $U_n(\beta)$ is linear in β, the resulting estimator has an explicit expression

$$\hat{\beta}_n = \left[\sum_{i=1}^n \int_0^\tau \{Z_i(t) - \overline{Z}(t)\}^{\otimes 2} R_i(t)\, dt \right]^{-1} \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \overline{Z}(t)\}\, dN_i(t).$$

Furthermore, Lin and Ying (1994) applied the counting process martingale theory to show that β_n is consistent and asymptotically normal. However, it is not asymptotically efficient.

Zeng et al. (2005) provided a unified framework for deriving an efficient estimator for the regression parameter under a broad class of transformed hazards models, including the additive hazards model. In this class, the hazard function for the survival time given covariate $\{Z(s), s \le t\}$ takes the form

$$G\{\lambda_Z(t)\} = \mu(t) + \beta^T Z(t),$$

where $\mu(t)$ is an unknown baseline hazard function, and $G(\cdot)$ is a known and increasing transformation function. The commonly used transformation functions $G(\cdot)$ have been mentioned before and $G(x) = x$ yields the additive hazards model. Take $H(t) = G^{-1}(t)$. Under right censoring, the log-likelihood function is proportional to

$$l_n(\beta, \mu) = \sum_{i=1}^{n} \left[\int_0^\tau \log H\{\mu(t) + \beta^T Z_i(t)\}\, dN_i(t) - \int_0^\tau R_i(t) H\{\mu(t) + \beta^T Z_i(t)\}\, dt \right].$$

For inference, Zeng et al. (2005) proposed a sieve maximum likelihood approach. In their sieve method, the functional parameter $\mu(t)$ is restricted to a finite dimensional functional space $\mathcal{S}_n$ with the dimension d_n:

$$\mathcal{S}_n = \left\{ \mu(t) : \mu(t) = \sum_{j=1}^{d_n} \alpha_k B_k(t), \sum_{j=1}^{d_n} |\alpha_j| \le D_n \right\},$$

where $\{B_k(t), k = 1, \ldots, d_n\}$ are basis functions such as B-splines and wavelet basis functions and the dimension d_n depends on the sample size n. They maximized the above log-likelihood function over $\mathcal{B}_0 \times \mathcal{S}_n$ and obtained the final estimators $\widehat{\beta}_n$ and $\widehat{\mu}(t)$, where $\mathcal{B}_0$ is a prespecified bounded open set containing the true β. Zeng et al. (2005) showed that $\widehat{\beta}_n$ is asymptotically normal and semiparametrically efficient.

3.6. Accelerated failure time model

Another important alternative to the Cox proportional hazard model is the accelerated failure time model. This model provides a natural formulation of the effects of covariates on potentially censored response variable and is in many ways more appealing because of its quite direct physical interpretation, especially when the response variable does not pertain to survival time (Zeng and Lin, 2007b). The accelerated failure time model specifies that

$$\log(T) = -\beta^T Z + \varepsilon,$$

where ε is an error term with a completely unspecified distribution function. Its hazard function given covariate Z is then

$$\lambda_Z(t) = \lambda\{t \exp(\beta^T Z)\} \exp(\beta^T Z),$$

where $\lambda(\cdot)$ is the hazard function of the error $\exp(\varepsilon)$, which is completely unknown. The accelerated failure time model characterizes a linear relationship directly between the logarithm of the survival time T and the covariate Z. Over the last three decades, there has been considerable progress in the estimation and inference for the parameter β in the presence of right-censoring, to name a few, Ying (1993), Fygenson and Ritov (1994), Jin et al. (2003), Lin and Chen (2013) and so on.

The rank approach has been used for inference in the accelerated failure time model. Define $e_i(\beta) = \log(Y_i) - Z_i^T \beta$, $N_i(\beta, t) = \Delta_i I(e_i(\beta) \le t)$ and $R_i(\beta, t) = I(e_i(\beta) \ge t)$ for $i = 1, \ldots, n$. Write $S_n^{(k)}(\beta, t) = n^{-1} \sum_{i=1}^n Z_i^{\otimes k} R_i(\beta, t)$ $(k = 0, 1)$ and $\overline{Z}(\beta, t) = S_n^{(1)}(\beta, t) / S_n^{(0)}(\beta, t)$. Then a general rank estimating function for β takes the form

$$l_{\phi,n}(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\beta, t)[Z_i - \overline{Z}(\beta, t)]\, dN_i(\beta, t),$$

where $\phi(\cdot)$ is a possibly data dependent weight function. The rank estimator of β is obtained by solving the equation $l_{\phi,n}(\beta) = 0$ if solvable. The choice of the weight function $\phi(\cdot)$ is very important for the estimation efficiency and computational feasibility. In practice, the two weights, $\phi = 1$ and $\phi(\beta, t) = S_n^{(0)}(\beta, t)$, are often used, which correspond to the log-rank and Gehan statistics, respectively. In particular, when $\phi(\beta, t) = S_n^{(0)}(\beta, t)$, $l_{\phi,n}(\beta)$ can be re-written as

$$l_{\phi,n}(\beta) = \frac{1}{n} \sum_{i,j=1}^{n} \Delta_i (Z_i - Z_j) I\big(e_i(\beta) \le e_j(\beta)\big).$$

It is a monotone estimating function as discussed in Fygenson and Ritov (1994) and the estimator of β can be obtained by minimizing the following objective function via linear programming:

$$L_{\phi,n}(\beta) = \frac{1}{n} \sum_{i,j=1}^{n} \Delta_i [e_i(\beta) - e_j(\beta)]^-,$$

observing that $(Z_i - Z_j) I(e_i(\beta) \le e_j(\beta))$ is the gradient in β of $[e_i(\beta) - e_j(\beta)]^-$, where $a^- = I_{(a<0)}|a|$. For other general weight functions, Jin et al. (2003) proposed an iterative linear programming. Particularly, for $\phi(t) = \lambda'(t)/\lambda(t)$ or $\phi_n(t) \to \lambda'(t)/\lambda(t)$ in probability, the resulting weighted rank estimator is asymptotically semiparametric efficient. On the basis of this fact, Lin and Chen (2013) proposed a one-step method to derive an efficient estimator for β.

To accommodate time-dependent covariates, Zeng and Lin (2007b) considered a general accelerate failure time model:

$$\lambda_Z(t) = \lambda\left(\int_0^t \exp\{\beta^T Z(s)\}\, ds\right) \exp\{\beta^T Z(t)\},$$

where $\lambda(\cdot)$ is an unknown baseline function of the error term $\exp(\varepsilon)$. They showed that the nonparametric maximum likelihood method does not work this. To estimate the parameter β efficiently, they used kernel smoothing to construct a smooth approximation to the profile likelihood function for the regression parameter. They showed that the kernel-smoothed profile likelihood estimator is consistent and asymptotically normal with the asymptotic covariance matrix that attains the semiparametric efficiency bound.

4. Multivariate analysis with right censored data

In this section, we give an overview of semiparametric models for analyzing multivariate survival data. Specifically, these multivariate survival data include clustered failure time data, recurrent events and multiple types of survival events.

4.1. Clustered failure time data

Clustered failure time data arise when the study subjects are sampled in clusters. Because of clustering, the failure times within the same cluster tend to be correlated. Therefore, it is necessary to account for this within-cluster dependence in modeling and inference. One approach to model the intro-cluster dependence is the proportional hazards frailty model, under which the hazard function for the jth subject of the ith cluster associated with covariates $X_{ij}(\cdot)$ takes the form

$$\lambda(t|X_{ij}, \xi_i) = \xi_i \lambda(t) \exp\{\beta^T X_{ij}(t)\}, \quad i = 1, \ldots, n; \;\; j = 1, \ldots, n_i, \tag{7}$$

where $\lambda(\cdot)$ is an unspecified baseline hazard function, β is a vector of unknown regression parameters, and ξ_i is an unobserved frailty for the ith cluster assumed to follow a gamma-distribution. The consistency and asymptotic distribution of the nonparametric maximum likelihood estimator for this model have been rigorously studied by Murphy (1994, 1995) for the case of no covariates and by Parner (1998) for the case with covariates.

More general transformation models have also been considered to fit clustered failure time data. Cai et al. (2002) considered the class of models given in Section 3.4 with a scalar random effect (i.e., $Z_{ij} \equiv 1$). Their estimation was based on minimizing the empirical sum of squares of the differences between certain observed quantities and their expected values. The estimators are not asymptotically efficient, and the variance estimation is computationally demanding. Recently, Zeng et al. (2008) studied a broad class of transformation models with random effects. In their model, the cumulative hazard function of T_{ij}, the jth failure time of the ith cluster, is related to $X_{ij}(\cdot)$ and $Z_{ij}(\cdot)$ as follows:

$$\Lambda(t|\overline{X}_{ij}(t), \overline{Z}_{ij}(t), b_i) = G\left(\int_0^t \exp\{X_{ij}(s)^T \beta + Z_{ij}(s)^T b_i\}\, d\Lambda(s)\right), \quad i = 1, \ldots, n; \;\; j = 1, \ldots, n_i, \tag{8}$$

where $\overline{X}_{ij}(t)$ and $\overline{Z}_{ij}(t)$ denote the histories of $X_{ij}(\cdot)$ and $Z_{ij}(\cdot)$ over $[0, t]$, $G(\cdot)$ is a known increasing function with $G(0) = 0$ and $G(\infty) = \infty$, $\Lambda(\cdot)$ is an unspecified increasing function, β is a set of unknown regression parameters, and b_i is a set of unobserved mean-zero random effects for the ith cluster with a density function $\psi(b_i; \gamma)$ (with respect to a σ-finite measure $\mu(b_i)$) indexed by a d_2-dimensional parameter γ. If both X_{ij} and Z_{ij} are time-independent, then model (8) reduces to linear transformation models

$$H(T_{ij}) = -X_{ij}^T \beta - Z_{ij}^T b_i + \log \varepsilon_{ij}, \quad i = 1, \ldots, n; \;\; j = 1, \ldots, n_i,$$

where $H(x) = \log \Lambda(x)$. Thus, the proposed model is reminiscent of the linear mixed-effects model (Laird and Ware, 1982) for longitudinal data. Nonparametric maximum likelihood estimation has been used for inference in Zeng et al. (2008).

4.2. Recurrent events data

Recurrent events data are commonly encountered in many scientific studies where the event of interest may be experienced repeatedly for each study subject. Examples of recurrent events include multiple infection episodes and tumor recurrences often arise. To handle the recurrent events data, it is very natural and convenient to represent the recurrent event times as a non-decreasing counting process. Let $N^*(t)$ denote the number of events that the subject has experienced by time t, and let $Z(t)$ be a possibly time-dependent vector. Models for analyzing recurrent events can often be classified into intensity models or rate models.

For intensity models, one tends to model the functional form of the following intensity function:

$$\lambda_Z(t) = \lim_{\Delta t \to 0+} \frac{P\{N^*(t+\Delta t) - N^*(t) = 1 | N^*(s), Z(s), 0 \le s < t\}}{\Delta t}.$$

The most popular proportional hazards intensity model studied by Anderson and Gill (1982) specifies that the intensity function given $\{Z(s), s \le t\}$ takes the form

$$\lambda_Z(t) = \lambda_0(t) \exp\{\beta^T Z(t)\},$$

where $\lambda_0(\cdot)$ is an unspecified baseline intensity function. Anderson and Gill (1982) developed an elegant asymptotic theory using the powerful martingale theory for the partial likelihood estimation of β. The above model assumes that the occurrence of a future event is independent of the prior event history unless such dependence is represented by suitable time-varying covariates. This assumption is slightly strong and often violated. Instead, a more appropriate model is to incorporate unobserved random effects or frailties into the intensity model to accommodate the dependence between events. One example is the proportional intensity model with gamma-frailty,

$$\lambda(t|Z,\xi) = \xi\lambda(t) \exp\{\beta^T Z(t)\}.$$

To accommodate non-proportional intensity and allow various frailty distributions, Zeng and Lin (2007a) proposed a broad class of semiparametric transformation intensity models with random effects and provided a unified framework of estimation and inference for the parameters of interest. Specifically, they used a transformation to model the cumulative intensity function of $N^*(t)$, which is defined by

$$\Lambda_Z(t|X,Z;b) = G\left(\int_0^t \exp\{\beta^T Z(s) + b^T X(s)\}\, d\Lambda(s)\right),$$

where β is a set of unknown regression parameters, b is a set of random effects with density function $\psi(b;\gamma)$ indexed by parameter γ, Z and X are the covariate processes associated with the fixed and random effects, and $G(\cdot)$ is a transformation function with $G(0)=0$ and $G(\infty)=\infty$. For the distribution of random effects, one may specify the multivariate normal distribution for ψ. Under this model, the log-likelihood function concerning the parameters (Λ, β, γ) is proportional to

$$\sum_{i=1}^n \log \int_b \prod_{t \le \tau} \left\{ Y_i(t) \exp\{\beta^T Z_i(s) + b^T X_i(s)\}\lambda(t) \times G'\left(\int_0^t Y_i(s) \exp\{\beta^T Z_i(s) + b^T X_i(s)\}\, d\Lambda(s)\right)\right\}^{\Delta N_i(t)}$$
$$\times \exp\left\{ -G\left(\int_0^\tau Y_i(s) \exp\{\beta^T Z_i(s) + b^T X_i(s)\}\, d\Lambda(s)\right)\right\}\psi(b;\gamma)\, db,$$

where $G'(t) = dG(t)/dt$ and $\Delta N_i(t)$ denotes the jump of N_i at time t. Zeng and Lin (2007a) proposed the nonparametric maximum likelihood estimation and also provided a simple and reliable EM algorithm for the maximization of the nonparametric likelihood function. They showed that under some regularity conditions, the NPMLEs for (β,γ) are consistent, asymptotically normal and semiparametrically efficient.

Another group of models for recurrent events are called rate models, where one tends to model marginal mean of $dN^*(t)$ given covariates. Particularly, Pepe and Cai (1993) and Lin et al. (2000) studied the following rate model:

$$d\mu_Z(t) = E\{dN^*(t)|Z(t)\} = \exp\{\beta^T Z(t)\}\, d\Lambda(t).$$

See also Lawless and Nadeau (1995) and Lawless et al. (1997). To estimate β, Lin et al. (2000) proposed a similar partial likelihood function for maximization. However, since $N^*(t) - \int_0^t \exp(\beta^T Z(s))\lambda(s)\, ds$ is no longer a martingale, the counting process martingale theory is not applicable. Lin et al. (2000) gave a rigorous theoretical justification for this partial likelihood-like estimation for β using model empirical process theory. When $Z(\cdot)$ is time-invariant, this model implies that the conditional mean of $N^*(t)$ given Z satisfies

$$m_Z(t) = E\{N^*(t)|Z\} = \exp(\beta^T Z) m_0(t),$$

which shows that the mean functions associated with any two sets of covariates are proportional, independent of time. Lin et al. (2001) argued that this was strong in practice through an examination of one real data. To allow non-proportional structure, they also proposed a transformation model:

$$m_Z(t) = E\{N^*(t)|Z\} = G\{\exp(\beta^T Z) m_0(t)\},$$

where the link function $G(\cdot)$ is a two continuously differentiable and strictly increasing function.

4.3. Multiple types of events

Prentice et al. (1981) was among the first to extend Cox's model to different types of survival events. Let T_{ik} denote the failure time of event type k $(k=1,\ldots,K)$. For the ith subject, conditional on $Z(t)$ and the event history $\mathcal{N}_i(t) = \{N_i(s), s \le t\}$ up to time t, the hazard or intensity function for the kth event is defined by

$$\lambda(t|\mathcal{N}_i(t), Z(t)) = R_{ik}(t)\lambda_{0k}(t) \exp\{\beta_k^T Z_{ik}(t)\},$$

or

$$\lambda(t|\mathcal{N}_i(t), Z(t)) = R_{ik}(t)\lambda_{0k}(t - T_{i,k-1}) \exp\{\beta_k^T Z_{ik}(t)\},$$

where $\lambda_{0k}(\cdot)$ is the unknown baseline hazard function and β_k is the unknown parameter depending on k. For the risk set, $R_{ik}(t) = I(Y_{i,k-1} < t \le Y_{ik})$ for the first model and $R_{ik}(t) = I(Y_{i,k-1} + t \le Y_{ik})$ for the second model. Prentice et al. (1981) showed that, by modifying the risk sets for each k, the partial likelihood approach works for estimating the parameter β_k.

Instead of modeling the conditional distribution of the kth event, Wei et al. (1989) proposed to model the marginal hazard function $\lambda_{ik}(t)$ of the kth type of failure of the ith subject with Cox's proportional hazard model as follows:

$$\lambda_{ik}(t|Z(t)) = \lambda_{0k}(t)\exp\{\beta_k^T Z_{ik}(t))\}.$$

Let $\mathcal{R}_k(t) = \{j : Y_{jk} \ge t\}$. Then, the partial likelihood for the kth type of failure is

$$\sum_{i=1}^{n}\left[\frac{\exp\{\beta^T Z_{ik}(Y_{ik})\}}{\sum_{j\in\mathcal{R}_k(Y_{ik})}\exp\{\beta^T Z_{jk}(Y_{jk})\}}\right]^{\Delta_{ik}}.$$

Thus, for each parameter β_k, the partial likelihood approach is applied directly.

Although the marginalization in these models is robust, the marginal assumption ignores the information among failure times, leading to inefficient estimation procedures. To explicitly account for between-event dependence, Zeng et al. (2009) studied a gamma-frailty transformation model, which is defined as, given a frailty ξ_i and covariate Z_i,

$$\Lambda_{ik}(t|Z_i,\xi_i) = G_k\{\Lambda_k(t)\exp(\beta_k^T X_i)\}\xi_i,$$

where ξ_i follows the gamma distribution with mean 1 and variance θ. Here for easy presentation, assume that the covariate Z is time independent. This model has at least two merits. The model is a natural generalization of the usual frailty model and yields a marginally linear transformation model for each failure time. Second, β_k's has the same interpretation as the coefficients in the usual linear transformation models. Nonparametric maximum likelihood estimation was used for inference yielding the efficient estimators.

Another special group of multiple type events is called competing risk, where one survival event (for example, death), if occurred first, censors the other events. For competing risk data, only cause-specific hazards or sub-distribution are well defined. Specifically, the cause-specific hazard (CSH) function is defined as

$$\lambda_{k,C}(t|Z) = \lim_{h\to 0}\Pr(T\in[t,t+h), \quad \varepsilon = k|Z, T\ge t)/h, \quad k=1,\ldots,K,$$

where ε is the cause-specific indicator and $\lambda_{k,C}(t|Z)$ is the rate of failure at time t from cause k. Naturally, the cause-specific cumulative hazard function

$$\Lambda_{k,C}(t|Z) = \int_0^t \lambda_{k,C}(u|Z)\,du.$$

The so-called sub-distribution or cumulative incidence function (CIF) is defined as

$$F_k(t|Z) = \Pr(T\le t, \varepsilon = k|Z), \quad k=1,\ldots,K.$$

We note that $F_k(t|Z)$ is not a proper distribution in the presence of all causes of failure as the probability that the subject will fail from cause k, $F_k(\infty|Z) < 1$.

When the focus is on the cause-specific hazards for a particular cause, which can apply the standard clustered survival data, which can be applied by treating failures from other causes as censored observations. For modeling cumulative incidence function, the most commonly used model for relating covariates and cumulative incidence function pertains to the Fine and Gray (1999) model. In this model, one assumes the following proportional hazards expression for the sub-distribution of a competing risk:

$$\lambda_1(t|Z) = \lambda_1(t)\exp(\beta^T Z),$$

where $\lambda_1(t|Z)$ is the conditional sub-distribution hazard for cause 1 given covariates Z, $\lambda_1(t)$ is the completely unspecified baseline sub-distribution hazard function and β is unknown regression parameters. Given n i.i.d. observations $\{X_i \equiv \min(T_i, C_i), \varepsilon_i\Delta_i, Z_i\}$, Fine and Gray (1999) proposed to maximize a variation of the partial log-likelihood for standard survival data by adapting inverse probability of censoring weighting techniques

$$\sum_{i=1}^{n} I(\varepsilon_i\Delta_i = 1)\left\{\beta^T Z_i - \log\left(\sum_{j=1}^{n} w_j(X_i)\exp(\beta^T Z_j)\right)\right\},$$

where $w_i(t) = I(X_i \ge t \cup \varepsilon_i\Delta_i > 1)\widehat{G}(t)/\widehat{G}(X_i\wedge t)$ is the weight for subject i at time t, and $\widehat{G}$ is the Kaplan–Meier estimate of the survival function of the censoring time. One key feature of the above partial log-likelihood is that subjects with failure from other causes remain in the risk at time t as long as $C_i > t$. If there is only a single cause of failure, the above partial likelihood reduces to the typical partial likelihood for the Cox model.

When competing risks data are clustered, Katsahian et al. (2006) extended Fine and Gray's model by including random effects or frailties in the sub-distribution hazard. Specifically, the conditional sub-distribution hazard of subject j in the ith cluster given the cluster-specific random effect u_i takes the form

$$\lambda_1(t|Z_{ij}, u_i) = \lambda_1(t)\exp(\beta^T Z_{ij} + u_i).$$

The random effects u_i's are assumed to be Gaussian with mean 0 and variance θ. The mean of the frailties is fixed at 0 to ensure the identifiability of the model. The conditional partial log-likelihood given the frailties is then expressed as

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} I(\varepsilon_{ij}\Delta_{ij}=1)\left\{\beta^T Z_{ij}+u_i-\log\left(\sum_{i'=1}^{n}\sum_{j'=1}^{n_i} w_{i'j'}(X_{ij})\exp(\beta^T Z_{i'j'}+u_{i'})\right)\right\},$$

where $w_{ij}(t)=I(X_{ij}\ge t\cup \varepsilon_{ij}\Delta_{ij}>1)\widehat{G}(t)/\widehat{G}(Y_{ij}\wedge t)$. To obtain the estimators of the regression parameters, one can maximize the marginal partial likelihood by integrating out the frailties from the conditional partial likelihood.

5. Interval censored data and joint analysis

5.1. Semiparametric models for interval censored data

The work on interval censored data is relatively less as compared to right censored data. One big challenge for interval censored data is the development of efficient estimation procedure and the corresponding computation. Here, we briefly review some key reference while refer interested reader to books by Sun (2006) and Chen et al. (2013).

Current status data for independent subjects have been studied widely in the literature. Especially, Huang (1996) proposed the proportional hazards model (1) and studied the efficient estimation based on the nonparametric maximum likelihood estimation. Let the observed data are $\{(C_i,\Delta_i,Z_i), i=1,\dots,n\}$, where for each i, C_i is the censoring time and $\Delta_i=I(T_i\le C_i)$ indicating T_i has occurred or not. The observed-data log-likelihood function for (β,Λ) is

$$l_n(\beta,\Lambda)=\sum_{i=1}^{n}[\Delta_i\log\{1-\exp(-\Lambda(C_i)\exp(\beta^T Z_i))\}-(1-\Delta_i)\exp(\beta^T Z_i)\Lambda(C_i)].$$

Note that this log-likelihood function is concave with respect to the cumulative hazards function Λ. The NPMLE of $(\beta,\ \Lambda)$ can be obtained by maximizing $l_n(\beta,\Lambda)$ by restricting $\Lambda(\cdot)$ to be an increasing step function with jumps only at the censoring times Y_i's. In addition to the semiparametric efficiency for β, Huang (1996) showed that convergence rate for the cumulative hazards function was no longer $n^{1/2}$ but $n^{1/3}$. Inference based on imputation for this model was later given by Satten et al. (1998). Rossini and Tsiatis (1996) examined the proportional odds model for the current status data and proposed the sieve estimation method. A general partial linear transformation model was considered by Ma and Kosorok (2005), which takes the form

$$H(T)=\beta^T Z+h(W)+\varepsilon,$$

where $h(\cdot)$ is an unknown smooth function and ε has a known distribution F. Based on the observed data $\{(C_i,\Delta_i,Z_i,W_i), i=1,\dots,n\}$, the log-likelihood for (β,h,H) can be expressed as

$$l_n(\beta,h,H)=\sum_{i=1}^{n}\{\Delta_i\log F[\beta^T Z_i+h(W_i)+H(C_i)]+(1-\Delta_i)\log\{1-F[\beta^T Z_i+h(W_i)+H(C_i)]\}\}.$$

To control the degree of smoothness of h, penalization or sieves may be used. Ma and Kosorok (2005) proposed the penalized maximum likelihood approach for (β,h,H), i.e., maximizing

$$l_n(\beta,h,H)-\lambda_n^2 J^2(h),$$

under the constraints that $\beta\in\Theta_0$, h belongs to a function space with $J(h)<\infty$ and H is non-decreasing, where λ_n is a tuning parameter and $J(h)$ is a penalty function. The penalized estimator of β is shown to be semiparametrically efficient while the estimators of H and h are $n^{1/3}-$consistent. Contrast to the multiplicative hazards models, the additive hazards model with current status data was also studied in Lin et al. (1998), with an additional assumption that the monitoring time C follows a proportional hazards model

$$\lambda_c(t|Z)=\lambda_{c,0}(t)\exp(\beta^T Z(t)).$$

This assumption is very critical and appealing since it converts the analysis of the current status data under the additive hazards model into the standard Cox regression. Their estimation was based on the martingale equations and later improved by Martinussen and Scheike (2002). Further efficiency issue on the additive hazards model via nonparametric maximum likelihood was considered in Ghosh (2001).

Although various statistical methods have been proposed to study the effects of covariates for current status data, studies of general interval-censored data in the literature have been relatively limited. Among those available, Finkelstein (1986) studied the proportional hazards model and Rabinowitz et al. (1995) considered the accelerated failure time models. Later, Zeng et al. (2006a) studied the additive risk model

$$\lambda(t|Z(s),0\le s\le t)=\lambda(t)+\beta^T Z(t)$$

for case II of interval censored data and provided the efficient estimation for β via nonparametric maximum likelihood. Specifically, the observed data are summarized as

$$\{\Delta_{1i}, \Delta_{2i}, (\Delta_{1i}+\Delta_{2i})U_i, (1-\Delta_{1i})V_i, Z(t), i=1,\ldots,n\},$$

where $\Delta_{1i} = I(T_i \le U_i)$, $\Delta_{2i} = I(U_i < T_i \le V_i)$ with left and right examination times U_i's and V_i's. Let $G(t) = \exp\{-\int_0^t \lambda(s)\,ds\}$ and $\overline{Z}(t) = \int_0^t Z(s)\,ds$. Under regularity conditions, the observed log-likelihood can be expressed as

$$l_n(\beta, G) = \sum_{i=1}^n \{\Delta_{1i}\log[1-G(U_i)\exp(-\beta^T\overline{Z}(U_i))] + \Delta_{2i}\log[G(U_i)\exp(-\beta^T\overline{Z}(U_i)) - G(V_i)\exp(-\beta^T\overline{Z}(V_i))] + (1-\Delta_{1i}-\Delta_{2i})\log[G(V_i)\exp(-\beta^T\overline{Z}(V_i))]\}.$$

The maximum likelihood estimator for (β, G) was obtained by maximizing $l_n(\beta, G)$ over the parameter space,

$$\Theta = \{(\beta, G) : \|\beta\| \le B, G(t) \text{ is a non-decreasing step function with jumps only at observed examination times and } G(0) = 1\}.$$

5.2. Joint models for repeated measures and failure times

Analysis of repeated measurements and clinical event time has received much attention in the last decade. Joint modeling has been proposed for such analysis. With joint modeling, one can obtain the most information from the data and provide unbiased and efficient estimation. Joint modeling includes both the model for repeated measurements and the model for event time. In general, a mixed-effect model (Diggle et al., 1994, Chapter 9) with normal random effects is used to model repeated measurements, while a proportional hazards model (Cox, 1972) is used to model the hazard function of failure time. The covariates in the models can be the subject's baseline variables, time for entry, or error-free confounders, etc. Random effects are used in both the mixed model and the proportional hazards model to account for the dependence between repeated measurements and survival time due to unobserved heterogeneity.

In some literature, such a joint model is described as either a selection model or a pattern-mixture model, depending upon how they are derived. When the conditional distribution of the survival time given repeated measurements is modeled, the derived joint model is called a selection-model; when the conditional distribution of repeated measurements given the survival time is modeled, the derived joint model is called a pattern-mixture model. Selection models have been studied by many authors in different contexts (Tsiatis et al., 1995; Wulfsohn and Tsiatis, 1997; Xu and Zeger, 2001a, 2001b). On the other hand, Wu and Carroll (1988), Wu and Bailey (1989), and Hogan and Laird (1997) proposed pattern-mixture models. In some studies where repeated measurements are considered as an internal covariate predicting risk of survival time, joint analysis is regarded as a missing covariate problem or measurement error problem in a proportional hazards model. For instance, Chen and Little (1999) considered the missing covariate problem in the proportional hazards model, although the covariates that were assumed to be time-independent. In most joint analysis literature, nonparametric maximum likelihood estimation is used (e.g., Tsiatis et al., 1995; Wulfsohn and Tsiatis, 1997). Computationally, the EM algorithm (Dempster et al., 1977) is often used to calculate the maximum likelihood estimates where random effects are treated as missing. The asymptotic properties of the proposed NPMLE were established for the first time by Zeng and Cai (2005).

To relax the parametric assumption of the random effects, Song et al. (2002) proposed another semiparametric model in which parametric assumptions on the distribution of random effects may be relaxed to those following a smooth density. They took a likelihood-based approach with the EM algorithm for the estimation procedure. Later, Song and Wang (2008) proposed an even more flexible semiparametric model by adapting time-varying coefficients to the proportional hazards model of the failure time T, which is

$$\lambda(t|Z) = \lambda_0(t)\exp\{\beta(t)Z(t)\}$$

and allows the effect of coefficients to vary over time, in addition to no distributional assumptions on the underlying longitudinal covariate processes $Z(t)$. An estimation procedure was constructed based on the conditional score estimators, and asymptotic properties of the estimators were derived based on martingale and empirical process theories.

5.3. Joint models for recurrent and terminal events

In many applications, especially in medical studies, the censoring time (i.e., time to the terminal event) is likely to be correlated with the recurrent event times. Marginal models have been proposed to analyze recurrent event data in the presence of a single terminal event (a univariate informative censoring time); see Cook and Lawless (1997) and Ghosh and Lin (2000, 2002). To capture the recurrence given an individual's past event history, a more attractive approach is to adopt a joint modeling approach for both recurrent events and informative censoring times. Approaches along these lines include the shared frailty model by Wang et al. (2001), Huang and Wang (2004), and Liu et al. (2004), which model the recurrent event and informative censoring time separately but allow a common frailty shared by these two models. Let $N(t)$ denote the number of events occurring before or at time t, and let T be the time to the terminal event. To be specific, Huang and Wang

(2004) assumed that the intensity function of recurrent event process $N(t)$ is

$$\lambda_R(t|Z,\xi)=\xi\lambda_{R,0}(t)\exp(\alpha^T Z) \tag{9}$$

and the hazards function of T takes the form

$$\lambda_T(t|Z,\xi)=\xi\lambda_{T,0}(t)\exp(\beta^T Z),$$

where ξ is nonnegative-valued latent variable. To estimate α and β, they proposed an estimation procedure by first estimating the value of the latent variable ξ from recurrent event data and then using the estimated values to estimate β through the partial likelihood. Liu et al. (2004) modeled ξ as the gamma frailty with the density $f_\theta(\cdot)$ and the hazards function of T as

$$\lambda_T(t|Z,\xi)=\xi^\gamma\lambda_{T,0}(t)\exp(\beta^T Z). \tag{10}$$

The parameter γ characterizes how different effects the frailty can have on the two hazards. Based on the full log-likelihood under models (9) and (10), they proposed an EM algorithm for parameter estimation. More general classes of transformation models and their theoretical properties have been more recently considered by Zeng and Lin (2009) and Zeng and Cai (2010a, 2010b). Zeng and Lin (2009) specified that conditional on the covariate Z and random effects b, the cumulative intensity function $\Lambda_R(t|Z,b)$ of $N(t)$ and the cumulative hazards function $\Lambda_T(t|Z,b)$ of T take the forms

$$\Lambda_R(t|Z,b)=H\left(\int_0^t \exp\{\alpha^T Z(s)+b^T\tilde{Z}(s)\}\,d\Lambda_{R,0}(s)\right)$$

and

$$\Lambda_T(t|Z,b)=G\left(\int_0^t \exp\{\alpha^T Z(s)+b^T(\phi\bigcirc\tilde{Z}(s))\}\,d\Lambda_{T,0}(s)\right),$$

respectively. Combining the techniques in Section 3.4 with those in Section 4.2, Zeng and Lin (2009) constructed the observed-data log-likelihood function for $(\alpha,\beta,\Lambda_{R,0},\Lambda_{T,0})$ and recommended the nonparametric maximum likelihood approach via a simple EM algorithm. Alternatively, Zeng and Cai (2010a, 2010b) considered a linear transformation model for T,

$$H(T)=\beta^T Z+\varepsilon$$

while modeled the rate function of $N(t)$ using an additive model given a subject-specific latent variable ξ,

$$E\{dN(t)|Z,T>t,\xi\}=I(T>t)\{dR(t,\xi)+\gamma^T Z dt\},$$

where $R(t,\xi)$ is an unknown subject-specific baseline cumulative rate function. The latent effect ν may be associated with the terminal event residual ε and explains the dependence between the recurrent event process and the terminal event. In contrast to parametric latent effects in Zeng and Lin (2009), such an association is fully nonparametric. The estimators of β and H can be found in Zeng and Lin (2006). To estimate γ, they found two surrogates $\overline{N}_i(t)$ and $\overline{Z}_i(t)$ based on estimates of β and H and proposed the following estimating equation for inference:

$$\sum_{i=1}^n\int \omega(t)I(Y_i>t)\{Z_i-\overline{Z}_i(t)dN_i(t)-d\overline{N}_i(t)-\gamma^T\{Z_i-\overline{Z}_i(t)\}\,dt\}=0,$$

where $\omega(t)$ is a deterministic weight function and $Y_i=\min(T_i,C_i)$ with the censoring times C_i for each $i=1,\dots,n$.

6. Real data analysis

In this section, we use two real examples described in Section 1 to show the power of semiparametric models in survival analysis.

Example 1 (*Lung cancer study (continued)*). As in the existing literature, we confined our attention to the subset of 97 patients without prior therapy and consider the two kinds of covariates: performance status (Karnofsky rating), a measure of general fitness and tumor type, namely, large, adeno, small or squamous. This dataset has been extensively analyzed by many authors. In particular, Cheng et al. (1995) and Murphy et al. (1997) fitted a proportional odds model. Recently, Chen et al. (2002) and Zeng and Lin (2006) used a class of transformation models (6) to fit the data, along with the transformation $G(x)=\log(1+rx)/r$ with $r\in[0,1]$. Table 1 summarized the estimated results, which is from Zeng and Lin (2006). They all showed that except for the covariate 'squamous versus large', all other covariates' effects are statistically significant. Since transformation models (6) include the proportional hazards and odds model, we could look at how well both of them perform from the likelihood principle. Zeng and Lin (2006) depicted that the likelihood at $r=1$ was slightly smaller than the maximum of the likelihood function over $r\in[0,1]$ but larger than the likelihood at $r=0$. In this sense, the proportional odds model is a good choice.

Example 5 (*(Diabetic retinopathy study (continued)*). As mentioned in the previous sections, the failure time of interest is the time to visual loss as measured by visual acuity less than 5/200. A subset of 197 high-risk patients has been extensively

Table 1
Summarized results for the lung cancer data in Example 1.

Covariate	$r=0$	$r=1$	$r=1.5$	$r=2$
Performance	−0.024 (0.006)	−0.053 (0.010)	−0.063 (0.012)	−0.072 (0.014)
Adeno vs large tumor	0.851 (0.348)	1.314 (0.554)	1.497 (0.012)	1.679 (0.712)
Small vs large tumor	0.547 (0.321)	1.383 (0.524)	1.605 (0.596)	1.814 (0.661)
Squam vs large tumor	−0.215 (0.347)	−0.181 (0.588)	−0.075 (0.675)	0.045 (0.749)

Note: Standard error estimates are shown in parentheses.

Table 2
Parameter estimates for the diabetic retinopathy study in Example 5.

Parameter	$\xi=0$	$\xi=0.3$	$\xi=1$
β_1	−0.523 (0.231)	−0.564 (0.250)	−0.659 (0.295)
β_2	0.421 (0.264)	0.447 (0.288)	0.496 (0.345)
β_3	−0.999 (0.369)	−1.073 (0.398)	−1.234 (0.466)
σ	1.038 (0.191)	1.114 (0.207)	1.296 (0.251)

Note: Standard error estimates are shown in parentheses.

analyzed, for example, by Cai et al. (2002) and Zeng et al. (2005, 2008). Three covariates are often considered: X_{1ij} indicates, by the value versus 0, whether or not the jth eye ($j=1$ for the left eye and $j=2$ for the right eye) of the ith patient was treated with laser photocoagulation, $X_{2i1}=X_{2i2}$ indicates, by the value 1 versus 0, whether the ith patient had adult-onset or juvenile-onset diabetics and $X_{3ij}=X_{1ij}\cdot X_{2ij}$ is the interaction between X_{1ij} and X_{2ij}. Zeng et al. (2008) fitted a semiparametric transformation model (8) with a Gaussian random effect b_i to account for the dependence between both eyes of the same patient. The transformation $G(x)$ was taken from the class $\{1-(1+\xi x)^{-1/\xi}, \xi\in[0,1]\}$. In particular, $\xi=0$ corresponds to the proportional hazards model and $\xi=1$ to the proportional odds model. Zeng et al. (2008) showed that the observed-data likelihood function is maximized at $\xi=0.3$. The results of the analysis were summarized in Table 2, as well as the estimated results for $\xi=0$ and $\xi=1$ for comparison. They show that there is a high degree of dependence between the failure times of both eyes from the same patient. Both the treatment indicator and the interaction term are significant whereas the diabetic type is not. For more details, see Zeng et al. (2008).

These data have also been analyzed using different models, such as additive transformation models (Zeng and Cai, 2010a, 2010b) and additive mixed effect models (Cai and Zeng, 2011). Although the estimated results are somewhat different, those findings of both the direction and significance of the covariate's effects coincide with the above.

7. Conclusions

We have provided an overview of semiparametric models in survival analysis. However, since this area is tremendously extensive, our review has inevitably missed a lot of models and methods in the literature. Particularly, they include mixture models and promotion time models for survival data with a cure fraction (Yakovlev and Tsodikov, 1996; Tsodikov, 1998; Chen et al., 1999; Zeng et al., 2006a, 2006b), missing data and measurement error problems in survival models (Li and Ryan, 2004), inference method for two-phase sampling including case-cohort and nested case-control studies (Self and Prentice, 1988; Breslow et al., 2009), and double robust estimation using auxiliary covariates (van der Laan and Robins, 2003). There are also many current works in survival analysis which are not mentioned in the review. These areas are but not limited to single index models with censored data, variable selection in semiparametric models with censored data, and statistical learning using censored data.

Acknowledgments

The authors are grateful to the Editor and the referee for comments and suggestions that lead to substantial improvements in the paper. The research of S. Guo was partially supported by the Chinese NSF Grants (11101408) and Key Lab of Random Complex Structures and Data Science of Chinese Academy of Sciences.

References

Anderson, P.K., Gill, R.D., 1982. Cox regression model for counting process: a large sample study. Ann. Statist. 10, 1100–1120.
Bennett, S., 1983. Analysis of survival data by the proportional odds model. Statist. Med. 2, 273–277.
Breslow, N.E., 1972. Discussion of the paper by D.R. Cox. J. Roy. Statist. Soc. Ser. B 34, 216–217.

Breslow, N.E., Lumbley, T., Ballantyne, C.M., Chambless, L.E., Kulich, M., 2009. Using the whole cohort in the analysis of case-cohort data. Amer. J. Epidem. 169, 1398–1405.
Cai, J., Zeng, D., 2011. Additive mixed effect model for clustered failure time data. Biometrics 67 (4), 1340–1351.
Cai, T., Cheng, S.C., Wei, L.J., 2002. Semiparametric mixed-effects models for clustered failure time data. J. Amer. Statist. Assoc. 97, 514–522.
Chen, D.G., Sun, J., Peace, K.E. (Eds.), 2013. Interval-Censored Time-to-Event Data: Methods and Applications, vol. 52. CRC Press.
Chen, K., Jin, Z., Ying, Z., 2002. Semiparametric analysis of transformation models with censored data. Biometrika 89, 659–668.
Chen, M.H., Ibrahim, J.G., Sinha, D., 1999. A new Bayesian model for survival data with a surviving fraction. J. Amer. Statist. Assoc. 94, 909–919.
Cheng, S.C., Wei, L.J., Ying, Z., 1995. Analysis of transformation models for censored data. Biometrika 82, 835–845.
Chen, H.Y., Little, R.J.A., 1999. Proportional hazards regression with missing covariates. J. Amer. Statist. Assoc. 94, 896–908.
Cook, R.J., Lawless, J.F., 1997. Marginal analysis of recurrent events and a terminal event. Statist. Med. 16, 911–924.
Cox, D.R., 1972. Regression models and life-tables (with discussion). J. Roy. Statist. Soc. Ser. B 34, 187–220.
Cox, D.R., 1975. Partial likelihood. Biometrika 62, 269–276.
Cox, D.R., Oakes, D., 1984. Analysis of Survival Data. Chapman and Hall, London.
Dabrowska, D.M., Doksum, K.A., 1988. Partial likelihood in transformation models with censored data. Scand. J. Statist. 15, 1–23.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. Ser. B 39, 1–38.
Diggle, P.J., Liang, K.Y., Zeger, S.L., 1994. Analysis of Longitudinal Data. Oxford University Press, New York.
Fine, J., Gray, R.J., 1999. A proportional hazards model for the subdistribution of a competing risk. J. Amer. Statist. Assoc. 94, 496–509.
Finkelstein, D.M., 1986. A proportional hazards model for interval-censored failure time data. Biometrics 42, 845–854.
Fleming, T.R., Harrington, D.P., 1991. Counting Processes and Survival Analysis. Wiley, New York.
Fygenson, M., Ritov, Y., 1994. Monotone estimating equations for censored data. Ann. Statist. 22, 732–746.
Ghosh, D., 2001. Efficiency considerations in the additive hazards model with current status data. Statist. Neerlandica 55, 367–376.
Ghosh, D., Lin, D.Y., 2000. Nonparametric analysis of recurrent events and death. Biometrics 56, 554–562.
Ghosh, D., Lin, D.Y., 2002. Marginal regression models for recurrent and terminal events. Statist. Sinica 12, 663–688.
Hogan, J.W., Laird, N.M., 1997. Mixture models for the joint distribution of repeated measures and event times. Statist. Med. 16, 239–257.
Hoel, D.G., Walberg, H.E., 1972. Statistical analysis of survival experiments. J. Nat. Canad. Inst. 49, 361–372.
Huang, J., 1996. Efficient estimation for the proportional hazard model with interval censoring. Ann. Statist. 24, 540–568.
Huang, C., Wang, M., 2004. Joint modeling and estimation for recurrent event processes and failure time data. J. Amer. Statist. Assoc. 99, 1153–1165.
Huster, W.J., Brookmeyer, R., Self, S.G., 1989. Modelling paired survival data with covariates. Biometrics 45, 145–156.
Jin, Z., Lin, D.Y., Wei, L.J., Ying, Z., 2003. Rank-based inference for the accelerated failure time model. Biometrika 90, 341–353.
Kalbfleisch, J.D., Prentice, R.L., 2002. The Statistical Analysis for Failure Time Data, 2nd ed. John Wiley and Sons, New York.
Katsahian, S., Resche-Rigon, M., Chevret, S., Porcher, R., 2006. Analysing multicenter competing risks data with a mixed proportional hazards model for the subdistribution. Statist. Med. 25, 4267–4278.
Kosorok, M.R., Lee, B.L., Fine, J.P., 2004. Robust inference for proportional hazards univariate frailty regression models. Ann. Statist. 32, 1448–1491.
Lam, K.F., Kuk, Y.C., 1997. A marginal likelihood approach to estimation in frailty model. J. Amer. Statist. Assoc. 92, 985–990.
Lam, K.F., Leung, T.L., 2001. Marginal likelihood estimation for proportional odds models with right censored data. Lifetime Data Anal. 7, 39–54.
Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–974.
Lawless, J.F., Nadeau, C., 1995. Some simple robust methods for the analysis of recurrent events. Technometrics 37, 158–168.
Lawless, J.F., Nadeau, C., Cook, R.J., 1997. Analysis of mean and rate functions for recurrent events. In: Lin, D.Y., Fleming, T.R. (Eds.), Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis. Springer-Verlag, New York, pp. 37–49.
Li, Y., Ryan, L., 2004. Survival analysis with heterogeneous covariate measurement error. J. Amer. Statist. Assoc. 99, 724–735.
Lin, D.Y., 1994. Cox regression analysis of multivariate failure time data: the marginal approach. Statist. Med. 13 (21), 2233–2247.
Lin, Y., Chen, K., 2013. Efficient estimation of the censored linear regression model. Biometrika 100 (2), 525–530, http://dx.doi.org/10.1093/biomet/ass073.
Lin, D.Y., Oakes, D., Ying, Z.L., 1998. Additive hazards regression with current status data. Biometrika 85, 289–298.
Lin, D.Y., Ying, Z., 1994. Semiparametric analysis of the additive risk model. Biometrika 81, 61–71.
Lin, D.Y., Ying, Z., 1995. Semiparametric analysis of general additive multiplicative intensity models for counting processes. Ann. Statist. 23, 1712–1734.
Lin, D.Y., Ying, Z., 1997. Additive hazards regression models for survival data. In: Proceedings of the First Seattle Symposium in Biostatistics. Springer-Verlag, New York.
Lin, D.Y., Wei, L.J., Yang, I., Ying, Z., 2000. Semiparametric regression for the mean and rate functions of recurrent events. J. Roy. Statist. Soc. Ser. B 62, 711–730.
Lin, D.Y., Wei, L.J., Ying, Z., 2001. Semiparametric transformation models for point processes. J. Amer. Statist. Assoc. 96, 620–628.
Liu, L., Wolfe, R.A., Huang, X., 2004. Shared frailty models for recurrent events and a terminal event. Biometrics 60, 747–756.
Ma, S., Kosorok, M., 2005. Penalized log-likelihood estimation for partly linear transformation models with current status data. Ann. Statist. 33, 2256–2290.
Martinussen, T., Scheike, T.H., 2002. Efficient estimation in additive hazards regression with current status data. Biometrika 89 (3), 649–658.
Murphy, S.A., 1994. Consistency in a proportional hazards model incorporating a random effect. Ann. Statist. 22, 712–731.
Murphy, S.A., 1995. Asymptotic theory for the frailty model. Ann. Statist. 23, 182–198.
Murphy, S.A., Rossini, A.J., van der Vaart, A.W., 1997. Maximum likelihood estimation in the proportional odds model. J. Amer. Statist. Assoc. 92, 968–976.
Parner, E., 1998. Asymptotic theory for the correlated gamma-frailty model. Ann. Statist. 26, 183–214.
Pepe, M.S., Cai, J., 1993. Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. J. Amer. Statist. Assoc. 88, 811–820.
Pettitt, A.N., 1984. Proportional odds models for survival data and estimates using ranks. Appl. Statist. 33, 169–175.
Prentice, R.L., 1973. Exponential survivals with censoring and explanatory variables. Biometrika 60, 278–288.
Prentice, R.L., Williams, B.J., Peterson, A.V., 1981. On the regression analysis of multivariate failure time data. Biometrika 68, 373–379.
Rabinowitz, D., Tsiatis, A., Aragon, J., 1995. Regression with interval-censored data. Biometrika 82, 501–513.
Rossini, A.J., Tsiatis, A., 1996. A semiparametric proportional odds regression model for the analysis of current status data. J. Amer. Statist. Assoc. 91, 713–721.
Satten, G.A., Datta, S., Williamson, J.M., 1998. Inference based on imputed failure times for the proportional hazards model with interval-censored data. J. Amer. Statist. Assoc. 93, 318–327.
Self, S.G., Prentice, R.L., 1988. Asymptotic distribution theory and efficiency results for case-cohort studies. Ann. Statist. 16, 64–81.
Shen, X., 1998. Proportional odds regression and sieve maximum likelihood estimation. Biometrika 85, 165–177.
Song, X., Davidian, M., Tsiatis, A.A., 2002. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. Biometrics 58, 742–753.
Song, X., Wang, C., 2008. Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. Biometrics 64, 557–566.
Sun, J., 2006. The Statistical Analysis of Interval-censored Failure Time Data. Springer-Verlag, New York.
Therneau, T.M., Hamilton, S.A., 1997. rhDNase as an example of recurrent event analysis. Statist. Med. 16, 2029–2047.
Tsiatis, A.A., DeGruttola, V., Wulfsohn, M.S., 1995. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. J. Amer. Statist. Assoc. 90, 27–37.
Tsodikov, A., 1998. A proportional hazards model taking account of long-term survivors. Biometrics 54, 1508–1516.
van der Laan, M.J., Robins, J., 2003. Unified Methods for Censored and Longitudinal Data and Causality. Springer-Verlag, New York.

van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer-Verlag, New York.
Wang, M., Qin, J., Chiang, C., 2001. Analyzing recurrent event data with informative censoring. J. Amer. Statist. Assoc. 96, 1057–1065.
Wei, L.J., Lin, D.Y., Weissfeld, L., 1989. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J. Amer. Statist. Assoc. 84, 1065–1073.
Wu, M.C., Bailey, K.R., 1989. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. Biometrics 44, 175–188.
Wu, M., Carroll, R., 1988. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. Biometrics 53, 330–339.
Wulfsohn, M.S., Tsiatis, A.A., 1997. A joint model for survival and longitudinal data measured with error. Biometrics 53, 330–339.
Xu, J., Zeger, S.L., 2001a. The evaluation of multiple surrogate endpoints. Biometrics 57, 81–87.
Xu, J., Zeger, S.L., 2001b. Joint analysis of longitudinal data comprising repeated measures and times to events. Appl. Statist. 50, 375–387.
Yakovlev, A.Y., Tsodikov, A.D., 1996. Stochastic Models of Tumor Latency and Their Biostatistical Applications. World Scientific, New Jersey.
Ying, Z., 1993. A large-sample theory of rank estimation for censored regression data. Ann. Statist. 21, 76–99.
Zeng, D., Cai, J., 2005. Asymptotic results for maximum likelihood estimates in joint analysis of repeated measurements and survival time. Ann. Statist. 33, 2132–2163.
Zeng, D., Cai, J., 2010a. A semiparametric additive rate model for recurrent events with an informative terminal event. Biometrika 97, 699–712.
Zeng, D., Cai, J., 2010b. Additive transformation models for clustered failure time data. Lifetime Data Anal. 16 (3), 333–352.
Zeng, D., Cai, J., Shen, Y., 2006a. Semiparametric additive risks model for interval-censored data. Statist. Sinica 16, 287–302.
Zeng, D., Yin, G., Ibrahim, J., 2006b. Semiparametric transformation models for survival data with a cure fraction. J. Amer. Statist. Assoc. 101, 670–684.
Zeng, D., Chen, Q., Ibrahim, J., 2009. Gamma frailty transformation models for multivariate survival data. Biometrika 96, 277–292.
Zeng, D., Lin, D.Y., 2006. Efficient estimation of semiparametric transformation models for counting processes. Biometrika 93, 627–640.
Zeng, D., Lin, D.Y., 2007a. Semiparametric transformation models with random effects for recurrent events. J. Amer. Statist. Assoc. 102, 167–180.
Zeng, D., Lin, D.Y., 2007b. Efficient estimation in the accelerated failure time model. J. Amer. Statist. Assoc. 102, 1387–1396.
Zeng, D., Lin, D.Y., 2007c. Maximum likelihood estimation in semiparametric models with censored data (with discussion). J. Roy. Statist. Soc. Ser. B 69, 507–564.
Zeng, D., Lin, D.Y., 2009. Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. Biometrics 65, 746–752.
Zeng, D., Lin, D.Y., Lin, X., 2008. Semiparametric transformation models with random effects for clustered failure time data. Statist. Sinica 18, 355–377.
Zeng, D., Yin, G., Ibrahim, J., 2005. Inference for a class of transformed hazards models. J. Amer. Statist. Assoc. 100, 1000–1008.