Discussion Paper, Series A, No. 2010-231

An Approximate Likelihood Procedure for Competing Risks Data

Akio Suzukawa

November 22, 2010

Graduate School of Economics & Business Administration
Hokkaido University
Kita 9 Nishi 7, Kita-Ku, Sapporo 060-0809, JAPAN

An approximate likelihood procedure for competing risks data

Akio Suzukawa

Graduate School of Economics and Business Administration
Hokkaido University

Abstract

Parametric estimation of cause-specific hazard functions in a competing risks model is considered. An approximate likelihood procedure for estimating parameters of cause-specific hazard functions based on competing risks data subject to right censoring is proposed. In an assumed parametric model that may have been misspecified, an estimator of a parameter is said to be consistent if it converges in probability to the pseudo-true value of the parameter as the sample size becomes large. Under censorship, the ordinary maximum likelihood method does not necessarily give consistent estimators. The proposed approximate likelihood procedure is consistent even if the parametric model is misspecified. An asymptotic distribution of the approximate maximum likelihood estimator is obtained, and the efficiency of the estimator is discussed. Datasets from a simulation experiment, an electrical appliance test and a pneumatic tire test are used to illustrate the procedure.

Keywords and Phrases: Aalen-Johansen estimator, cause-specific cumulative incidence function, censored data, Kaplan-Meier estimator.

1 Introduction

In reliability analysis or analyses in medical studies, more than one risk factor is often present. The failure of an individual is attributed to one of the risk factors, or causes, and is called the cause of failure. The competing risks data comprise the time to failure and the cause of failure. There have been many analyses of such data; results and references are given by Chiang (1968), David and Moeschberger (1978), Crowder (2001) and Kalbfleisch and Prentice (2002).

Suppose that an individual is exposed to k mutually exclusive causes of failure and that observations of a random sample of n individuals are made. Denote the failure time of the i-th individual by T_i. Its distribution function is denoted by $F_0(t) = P(T_i \le t)$. The corresponding survival function is denoted by $\bar{F}_0 = 1 - F_0$. Let V_i be the cause of failure of the i-th individual, which takes on values in a set of all causes $C = \{1, 2, \ldots, k\}$. We consider the independent right censoring of the failure times. Let Y_i be the censoring time with distribution function G and survival function $\bar{G} = 1 - G$. For each i, define

$$X_i = \min(T_i, Y_i) \quad \text{and} \quad \delta_i = I(T_i \le Y_i),$$

where $I(A)$ denotes the indicator of a set A. Observation of the i-th individual is either of the form $(X_i = x_i, V_i = v_i, \delta_i = 1)$ or $(X_i = x_i, \delta_i = 0)$.

The identifiable probabilistic aspect in the competing risks set-up is the joint distribution of T_i and V_i. It can be specified in terms of the cumulative incidence functions $F_0^{(j)}(t) = P(T_i \le t,\ V_i = j)$, $j \in C$. When F_0 is continuous, the sub-density of a cause $j \in C$ is defined by $f_0^{(j)}(t) = dF_0^{(j)}(t)/dt$, and the cause-specific hazard function is defined by $\lambda_0^{(j)}(t) = f_0^{(j)}(t)/\bar{F}_0(t)$. The overall hazard function is given by $\lambda_0(t) = \sum_{j \in C} \lambda_0^{(j)}(t)$.

In a nonparametric set-up, the overall survival function $\bar{F}_0$ is estimated by Kaplan and Meier's (1958) estimator:

$$\bar{F}_n(t) = \prod_{i: t_{(i)} \le t} (1 - d_i/n_i),$$

where $t_{(1)} < \cdots < t_{(m)}$ are the distinct times at which failures occur, and d_i and n_i are the number of failures and individuals at risk at $t_{(i)}$, respectively. A nonparametric maximum likelihood estimator of the cumulative incidence function $F_0^{(j)}$ was derived by Aalen (1976), and it can be thought of as a special case of the Aalen-Johansen theory of estimation for time-inhomogeneous Markov processes (Aalen and Johansen 1978). The estimator, which is termed Aalen-Johansen estimator, is defined by

$$F_n^{(j)}(t) = \sum_{i:t_{(i)} \leq t} \bar{F}_n(t_{(i)}-)\, d_i^{(j)}/n_i,$$

where $d_i^{(j)}$ is the number of individuals failing by cause j at $t_{(i)}$.

In a parametric set-up, the overall density $f_0(t) = dF_0(t)/dt$, or equivalently the overall hazard $\lambda_0(t)$, can be estimated by the ordinary likelihood procedure. Oakes (1986) proposed an approximate likelihood procedure for estimating unknown parameters of the overall density or the overall hazard. Suzukawa *et al.* (2001) investigated asymptotic properties of the approximate maximum likelihood estimator in comparison with those of the ordinary maximum likelihood estimator. In this paper, Oakes' (1986) approximate likelihood procedure is generalized to the competing risks framework.

In many practical situations, we are interested not in all causes but in only some of the causes. Decompose all causes C as $C = C_1 \cup C_2$, where C_1 and C_2 are disjoint. Our main concern is in the cause-specific hazards of sub-causes C_1. Denote the cumulative cause-specific hazard of cause j by $\Lambda^{(j)}$. In the nonparametric set-up, $\Lambda^{(j)}$ is estimated by the Nelson-Aalen estimator:

$$\Lambda_n^{(j)}(t) = \sum_{i:t_{(i)} \leq t} d_i^{(j)}/n_i.$$

Suzukawa and Taneichi (2003) proposed semiparametric estimators of $\Lambda^{(j)}$, $j \in C_1$, which are asymptotically more efficient than the above nonparametric estimators.

In a parametric set-up, a parametric model,

$$\mathcal{M}^{C_1} = \left\{ \{\lambda^{(j)}(\cdot;\boldsymbol{\theta})\}_{j\in C_1};\, \boldsymbol{\theta} \in \Theta \right\} \tag{1}$$

is assumed for the cause-specific hazards of sub-causes C_1, where $\boldsymbol{\theta}$ is a p-dimensional vector of unknown parameters and Θ is its parameter space. However, the assumed parametric model $\mathcal{M}^{C_1}$ does not necessarily contain the true cause-specific hazards. If it does not, it is a misspecified model. One of the simplest parametric model is a constant hazard model:

$$\mathcal{M}^{C_1}_{const} = \left\{ \{\lambda^{(j)}(\cdot;\theta^{(j)}) = \theta^{(j)}\}_{j\in C_1};\, \theta^{(j)} > 0, j \in C_1 \right\},$$

in which the cause-specific hazards for sub-causes C_1 are assumed to be independent of time. In almost cases, this seems to be a misspecified model. However, if we want to approximate each of the cause-specific hazards in C_1 with a constant value, the model $\mathcal{M}^{C_1}_{const}$ will be assumed, and estimates of the parameters $\theta^{(j)}$, $j \in C_1$ will be used as approximated values. In such a situation, it is important to estimate unknown parameters in a parametric model that may have been misspecified. In this paper, an approximate likelihood procedure for estimating parameters in model (1), which is not necessarily correct, is proposed.

The structure of this paper is as follows. In Section 2, it is pointed out that the ordinary likelihood procedure is unsuitable under the condition of misspecification of parametric models, and a formulation of the approximate likelihood procedure is presented. These procedures are asymptotically compared in Section 3. In Section 4, these procedures are illustrated using datasets from a simulation experiment, an electrical appliance test (Nelson 1970) and a pneumatic tire test (Davis and Lawrance 1989).

2 Parametric estimation of cause-specific hazard functions

2.1 Kullback-Leibler information in terms of cause-specific hazards

For any set of sub-densities $\{f^{(j)}\}_{j\in C}$, Kullback-Leibler information of the set of true sub-densities $\{f_0^{(j)}\}_{j\in C}$ relative to $\{f^{(j)}\}_{j\in C}$ is defined by

$$\mathrm{KL}\left[\{f_0^{(j)}\}_{j\in C},\ \{f^{(j)}\}_{j\in C}\right] = \sum_{j\in C}\int_0^\infty f_0^{(j)}(t)\log\frac{f_0^{(j)}(t)}{f^{(j)}(t)}\,dt. \tag{2}$$

This is a measure of the divergence of sub-densities. However, we are interested in the divergence of cause-specific hazards rather than the divergence of sub-densities.

When a set of sub-densities $\{f^{(j)}\}_{j\in C}$ is given, the corresponding set of cause-specific hazards $\{\lambda^{(j)}\}_{j\in C}$ is given by

$$\lambda^{(j)}(t) = f^{(j)}(t)/\bar{F}(t),$$

where $\bar{F}(t) = \sum_{j\in C}\int_t^\infty f^{(j)}(u)\,du$. Conversely, for a given set of cause-specific hazards $\{\lambda^{(j)}\}_{j\in C}$, the corresponding set of sub-densities $\{f^{(j)}\}_{j\in C}$ is given by $f^{(j)}(t) = \lambda^{(j)}(t)\exp\{-\Lambda(t)\}$, where $\Lambda(t) = \sum_{j\in C}\Lambda^{(j)}(t)$ and $\Lambda^{(j)}(t) = \int_0^t \lambda^{(j)}(u)\,du$. Thus, the Kullback-Leibler information defined by (2) can be thought as a measure of the divergence of the true cause-specific hazards $\{\lambda_0^{(j)}\}_{j\in C}$ relative to the cause-specific hazards $\{\lambda^{(j)}\}_{j\in C}$, and it can be expressed as

$$\begin{aligned}
&\mathrm{KL}\left[\{\lambda_0^{(j)}\}_{j\in C},\ \{\lambda^{(j)}\}_{j\in C}\right]\\
&= \sum_{j\in C}\int_0^\infty f_0^{(j)}(t)\log\left\{\frac{\lambda_0^{(j)}(t)}{\lambda^{(j)}(t)}\frac{\bar{F}_0(t)}{\bar{F}(t)}\right\}dt\\
&= \sum_{j\in C}\int_0^\infty f_0^{(j)}(t)\log\frac{\lambda_0^{(j)}(t)}{\lambda^{(j)}(t)}\,dt + \sum_{j\in C}\int_0^\infty f_0^{(j)}(t)\log\frac{\bar{F}_0(t)}{\bar{F}(t)}\,dt\\
&= \sum_{j\in C}\int_0^\infty f_0^{(j)}(t)\log\frac{\lambda_0^{(j)}(t)}{\lambda^{(j)}(t)}\,dt + \int_0^\infty f_0(t)\log\frac{\bar{F}_0(t)}{\bar{F}(t)}\,dt\\
&= \sum_{j\in C}\int_0^\infty f_0^{(j)}(t)\log\frac{\lambda_0^{(j)}(t)}{\lambda^{(j)}(t)}\,dt - \int_0^\infty f_0(t)\{\Lambda_0(t)-\Lambda(t)\}\,dt\\
&= \sum_{j\in C}\left[\int_0^\infty f_0^{(j)}(t)\log\frac{\lambda_0^{(j)}(t)}{\lambda^{(j)}(t)}\,dt - \int_0^\infty f_0(t)\{\Lambda_0^{(j)}(t)-\Lambda^{(j)}(t)\}\,dt\right].
\end{aligned}$$

From this, for each cause j, the divergence of the true cause-specific hazard $\lambda_0^{(j)}$ relative to $\lambda^{(j)}$ can be measured by

$$\int_0^\infty f_0^{(j)}(t)\log\frac{\lambda_0^{(j)}(t)}{\lambda^{(j)}(t)}\,dt - \int_0^\infty f_0(t)\{\Lambda_0^{(j)}(t)-\Lambda^{(j)}(t)\}\,dt.$$

When the parametric model $\mathcal{M}^{C_1}$ of (2) is assumed for sub-causes C_1, the divergence of the true cause-specific hazards $\{\lambda_0^{(j)}\}_{j\in C_1}$ relative to the parametric cause-specific hazards $\{\lambda^{(j)}(\cdot;\boldsymbol{\theta})\}_{j\in C_1}$ can be measured by

$$\begin{aligned}
&\sum_{j\in C_1}\left[\int_0^\infty f_0^{(j)}(t)\log\frac{\lambda_0^{(j)}(t)}{\lambda^{(j)}(t;\boldsymbol{\theta})}\,dt - \int_0^\infty f_0(t)\{\Lambda_0^{(j)}(t)-\Lambda^{(j)}(t;\boldsymbol{\theta})\}\,dt\right]\\
&\quad= \sum_{j\in C_1}\left[\int_0^\infty f_0^{(j)}(t)\log\lambda_0^{(j)}(t)\,dt - \int_0^\infty f_0(t)\Lambda_0^{(j)}(t)\,dt\right]\\
&\qquad- \sum_{j\in C_1}\left[\int_0^\infty f_0^{(j)}(t)\log\lambda^{(j)}(t;\boldsymbol{\theta})\,dt - \int_0^\infty f_0(t)\Lambda^{(j)}(t;\boldsymbol{\theta})\,dt\right],
\end{aligned}$$

where $\Lambda^{(j)}(t;\boldsymbol{\theta}) = \int_0^t \lambda^{(j)}(u;\boldsymbol{\theta})\,du$. The first term on the right-hand side is independent of $\boldsymbol{\theta}$. Put the second term as

$$\eta(\boldsymbol{\theta}) = \sum_{j\in C_1}\left[\int_0^\infty f_0^{(j)}(t)\log\lambda^{(j)}(t;\boldsymbol{\theta})\,dt - \int_0^\infty f_0(t)\Lambda^{(j)}(t;\boldsymbol{\theta})\,dt\right]. \tag{3}$$

Since it is desired to minimize the divergence, the best $\boldsymbol{\theta}$ in the parameter space Θ is a maximizer of $\eta(\boldsymbol{\theta})$. Supposing that $\eta(\boldsymbol{\theta})$ has a maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0^* \in \Theta$, it is said that $\boldsymbol{\theta}_0^*$ is the pseudo-true vector of $\boldsymbol{\theta}$. The pseudo-true $\boldsymbol{\theta}_0^*$ gives the best approximation to the true cause-specific hazards $\{\lambda_0^{(j)}\}_{j\in C_1}$ in the model $\mathcal{M}^{C_1}$.

When $\mathcal{M}^{C_1}$ includes the true cause-specific hazards $\{\lambda_0^{(j)}\}_{j\in C_1}$, there exists the true vector $\boldsymbol{\theta}_0$ such that $\lambda_0^{(j)}(t) = \lambda^{(j)}(t;\boldsymbol{\theta}_0)$ holds for all $j \in C_1$ and all $t > 0$. Noting

$$\log\frac{\lambda^{(j)}(t;\boldsymbol{\theta}_0)}{\lambda^{(j)}(t;\boldsymbol{\theta})} = -\log\frac{\lambda^{(j)}(t;\boldsymbol{\theta})}{\lambda^{(j)}(t;\boldsymbol{\theta}_0)} \geq 1 - \frac{\lambda^{(j)}(t;\boldsymbol{\theta})}{\lambda^{(j)}(t;\boldsymbol{\theta}_0)},$$

we have, for any $\boldsymbol{\theta} \in \Theta$,

$$\begin{aligned}
&\eta(\boldsymbol{\theta}_0) - \eta(\boldsymbol{\theta})\\
&= \sum_{j\in C_1}\left[\int_0^\infty f_0^{(j)}(t)\log\frac{\lambda^{(j)}(t;\boldsymbol{\theta}_0)}{\lambda^{(j)}(t;\boldsymbol{\theta})}\,dt - \int_0^\infty f_0(u)\{\Lambda_0^{(j)}(u;\boldsymbol{\theta}_0) - \Lambda^{(j)}(u;\boldsymbol{\theta})\}\,du\right]\\
&\geq \sum_{j\in C_1}\left[\int_0^\infty f_0^{(j)}(t)\left\{1 - \frac{\lambda^{(j)}(t;\boldsymbol{\theta})}{\lambda^{(j)}(t;\boldsymbol{\theta}_0)}\right\}dt - \int_0^\infty f_0(u)\left\{\int_0^u \lambda_0^{(j)}(t;\boldsymbol{\theta}_0) - \lambda^{(j)}(t;\boldsymbol{\theta})\,dt\right\}du\right]\\
&= \sum_{j\in C_1}\left[\int_0^\infty \bar{F}_0(t)\{\lambda^{(j)}(t;\boldsymbol{\theta}_0) - \lambda^{(j)}(t;\boldsymbol{\theta})\}\,dt\right.\\
&\qquad\left. - \int_0^\infty \{\lambda_0^{(j)}(t;\boldsymbol{\theta}_0) - \lambda^{(j)}(t;\boldsymbol{\theta})\}\left\{\int_t^\infty f_0(u)\,du\right\}dt\right] = 0.
\end{aligned}$$

Thus, if $\mathcal{M}^{C_1}$ includes the true hazards, then $\eta(\boldsymbol{\theta})$ has a maximum at $\boldsymbol{\theta}_0$. In this case, the pseudo-true $\boldsymbol{\theta}_0^*$ is the true $\boldsymbol{\theta}_0$.

Example 1. Let $\mu_0 = \int_0^\infty \bar{F}_0(t)\,dt < \infty$. Consider the case in which, only for cause 1, the constant hazard model

$$\mathcal{M}_{const}^{(1)} = \{\lambda^{(1)}(\cdot;\theta) = \theta;\ \theta > 0\}. \tag{4}$$

is assumed. Then, η of (3) is given by

$$\eta(\theta) = \int_0^\infty f_0^{(1)}(t)\log\theta\,dt - \theta\int_0^\infty tf_0(t)\,dt = \pi_0^{(1)}\log\theta - \theta\mu_0,$$

where $\pi_0^{(1)} = \int_0^\infty f_0^{(1)}(t)\,dt$. Obviously, $\eta(\theta)$ has a maximum at $\theta = \theta_0^* = \pi_0^{(1)}/\mu_0$, which is the pseudo-true value. When the model $\mathcal{M}_{const}^{(1)}$ is correct, there exists a true parameter $\theta_0 > 0$ such that $\lambda_0^{(1)}(t) = \lambda^{(1)}(t;\theta_0) = \theta_0$ for all t. In this case, the pseudo-true value θ_0^* is the true value θ_0 since

$$\theta_0^* = \pi_0^{(1)}/\mu_0 = \int_0^\infty f_0^{(1)}(t)\,dt/\mu_0 = \int_0^\infty \theta_0\bar{F}_0(t)\,dt/\mu_0 = \theta_0. \quad \square$$

2.2 Likelihood procedure

The purpose here is to explore the likelihood procedure in the parametric model $\mathcal{M}^{C_1}$ of (1), which may have been misspecified. It is well known that the log-likelihood is given by

$$L_n(\boldsymbol{\theta}) = \sum_{i=1}^n\sum_{j\in C_1}\left\{\delta_i I(V_i = j)\log\lambda^{(j)}(X_i;\boldsymbol{\theta}) - \Lambda^{(j)}(X_i;\boldsymbol{\theta})\right\} \tag{5}$$

(Kalbfleisch and Prentice 1980, Crowder 2001). The maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}}_n$ is a maximizer of $L_n(\boldsymbol{\theta})$. By the law of large numbers, as $n \to \infty$, $L_n(\boldsymbol{\theta})/n$ converges in probability to a function

$$\begin{aligned}\xi(\boldsymbol{\theta}) &= \sum_{j\in C_1} E\left\{\delta_i I(V_i = j)\log\lambda^{(j)}(X_i;\boldsymbol{\theta}) - \Lambda^{(j)}(X_i;\boldsymbol{\theta})\right\} \\ &= \sum_{j\in C_1}\left\{\int_0^\infty f_0^{(j)}(x)\bar{G}(x)\log\lambda^{(j)}(x;\boldsymbol{\theta})\,dx - \int_0^\infty h(x)\Lambda^{(j)}(x;\boldsymbol{\theta})\,dx\right\},\end{aligned}$$

where $h(x) = f_0(x)\bar{G}(x) + g(x)\bar{F}_0(x)$ and g is the probability density function of the censoring distribution G. Assuming that $\xi(\boldsymbol{\theta})$ has a maximum at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_0$, under suitable regularity conditions, the MLE $\hat{\boldsymbol{\theta}}_n$ converges in probability to $\tilde{\boldsymbol{\theta}}_0$ as $n \to \infty$. The important point to note is that the limit $\tilde{\boldsymbol{\theta}}_0$ generally depends on the censoring distribution G, which is a nuisance. In general, $\tilde{\boldsymbol{\theta}}_0$ differs from the pseudo-true $\boldsymbol{\theta}_0^*$.

When $\mathcal{M}^{C_1}$ includes the true cause-specific hazards $\{\lambda_0^{(j)}\}_{j\in C_1}$, there exists the true $\boldsymbol{\theta}_0$ such that $\lambda_0^{(j)}(t) = \lambda^{(j)}(t;\boldsymbol{\theta}_0)$ holds for all $j \in C_1$ and all $t > 0$. Thus, we have, for any $\boldsymbol{\theta} \in \Theta$,

$$\begin{aligned}&\xi(\boldsymbol{\theta}_0) - \xi(\boldsymbol{\theta}) \\ &= \sum_{j\in C_1}\left[\int_0^\infty f_0^{(j)}(x)\bar{G}(x)\log\frac{\lambda^{(j)}(x;\boldsymbol{\theta}_0)}{\lambda^{(j)}(x;\boldsymbol{\theta})}\,dx - \int_0^\infty h(x)\{\Lambda^{(j)}(x;\boldsymbol{\theta}_0) - \Lambda^{(j)}(x;\boldsymbol{\theta})\}\,dx\right] \\ &\geq \sum_{j\in C_1}\left[\int_0^\infty f_0^{(j)}(x)\bar{G}(x)\left\{1 - \frac{\lambda^{(j)}(x;\boldsymbol{\theta})}{\lambda^{(j)}(x;\boldsymbol{\theta}_0)}\right\}dx - \int_0^\infty h(x)\{\Lambda^{(j)}(x;\boldsymbol{\theta}_0) - \Lambda^{(j)}(x;\boldsymbol{\theta})\}\,dx\right] \\ &= \sum_{j\in C_1}\left[\int_0^\infty\left\{\bar{F}_0(x)\bar{G}(x) - \int_x^\infty h(u)\,du\right\}\{\lambda^{(j)}(x;\boldsymbol{\theta}_0) - \lambda^{(j)}(x;\boldsymbol{\theta})\}\,dx\right] = 0.\end{aligned}$$

Thus, if $\mathcal{M}^{C_1}$ is correct, then $\xi(\boldsymbol{\theta})$ has a maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and, as $n \to \infty$, the MLE $\hat{\boldsymbol{\theta}}_n$ converges in probability to the true $\boldsymbol{\theta}_0$ under some regularity conditions.

However, in misspecified parametric models, $\tilde{\boldsymbol{\theta}}_0$, which is a limit of MLE, generally differs from the pseudo-true $\boldsymbol{\theta}_0^*$. Thus, MLE is not always consistent. In general, MLE does not give the best approximation in the assumed parametric model to the true hazards even if the sample size is sufficiently large.

Example 2. As a simple example, we consider the case of Example 1 again. In the model $\mathcal{M}_{const}^{(1)}$ of (4), the log-likelihood function is

$$L_n(\theta) = \sum_{i=1}^n\{\delta_i I(V_i = 1)\log\theta - \theta X_i\},$$

and MLE is given by

$$\hat{\theta}_n = \{\sum_{i=1}^n \delta_i I(V_i = 1)\}/(\sum_{i=1}^n X_i). \tag{6}$$

According to the law of large numbers, as $n \to \infty$, $\hat{\theta}_n$ converges in probability to

$$\tilde{\theta}_0 = E\{\delta_i I(V_i = 1)\}/E(X_i) = \left\{\int_0^\infty f_0^{(1)}(x)\bar{G}(x)\,dx\right\}/\left\{\int_0^\infty \bar{F}_0(x)\bar{G}(x)\,dx\right\},$$

which maximizes

$$\begin{aligned}\xi(\theta) &= \int_0^\infty f_0^{(1)}(x)\bar{G}(x)\log\lambda^{(1)}(x;\theta)\,dx - \int_0^\infty h_0(x)\Lambda^{(1)}(x;\theta)\,dx \\ &= (\log\theta)\int_0^\infty f_0^{(1)}(x)\bar{G}(x)\,dx - \theta\int_0^\infty \bar{F}_0(x)\bar{G}(x)\,dx.\end{aligned}$$

If $\mathcal{M}_{const}^{(1)}$ is correct, there exists the true parameter $\theta_0 > 0$ such that $\lambda_0^{(1)}(t) = \lambda^{(1)}(t;\theta_0) = \theta_0$ holds for all t. In this case,

$$\tilde{\theta}_0 = \left\{ \int_0^\infty \theta_0 \bar{F}_0(x)\bar{G}(x)\, dx \right\} / \left\{ \int_0^\infty \bar{F}_0(x)\bar{G}(x)\, dx \right\} = \theta_0,$$

and hence $\hat{\theta}_n$ is consistent.

If $\bar{G} \equiv 1$ (no censorship), then $\tilde{\theta}_0 = \theta_0^* = \pi_0^{(1)}/\mu_0$ (pseudo-true value). However, $\tilde{\theta}_0$ is generally dependent on $\bar{G}$, and it is generally different from the pseudo-true value. We will make stronger assumptions to illustrate how $\tilde{\theta}_0$ differs from the pseudo-true value θ_0^* when $\mathcal{M}_{const}^{(1)}$ is a misspecified model.

Assume that the true cause-specific hazard of cause 1 is proportional to the overall hazard: $\lambda_0^{(1)}(t) = \pi_0^{(1)}\lambda_0(t)$ and the overall hazard has a Weibull form: $\lambda_0(t) = \alpha t^{\alpha-1}$, where $\alpha > 0$. In this case, $\mathcal{M}_{const}^{(1)}$ is correct if and only if $\alpha = 1$. The pseudo-true value is given by $\theta_0^* = \pi_0^{(1)}/\Gamma(1+1/\alpha)$, where Γ is the gamma function. We also assume proportional censoring: $\bar{G}(t) = \{\bar{F}_0(t)\}^\beta$. Then the censoring proportion q is given by $q = P(\delta_i = 0) = \beta/(1+\beta)$. Under these assumptions, the MLE of (6) converges in probability to $\tilde{\theta}_0 = \theta_0^*(1-q)^{1-1/\alpha}$ as $n \to \infty$. A necessary and sufficient condition for $\tilde{\theta}_0 = \theta_0^*$ is $\alpha = 1$ (correct specification) or $q = 0$ (no censorship). In other words, under the condition of misspecification ($\alpha \neq 1$) and censorship ($q > 0$), the MLE defined by (6) does not converge in probability to the pseudo-true value as $n \to \infty$.

The inconsistency can be measured by

$$\eta(\theta_0^*) - \eta(\tilde{\theta}_0) = \pi_0^{(1)} \left\{ (1-q)^{1-1/\alpha} - 1 - (1-1/\alpha)\log(1-q) \right\}.$$

When $\alpha \neq 1$ (misspecification), this is strictly increasing with the censoring proportion q. Thus, the divergence of $\lambda^{(1)}(t;\tilde{\theta}_0) = \tilde{\theta}_0$, which is a limit of the MLE, with respect to the true hazard increases with increase in the censoring proportion.

Figure 1 shows how the hazards $\lambda^{(1)}(t;\theta_0^*) = \theta_0^*$ (dotted line) and $\lambda^{(1)}(t;\tilde{\theta}_0) = \tilde{\theta}_0$ (dashed line) differ from the true hazard $\lambda_0^{(1)}(t) = \pi_0^{(1)}\alpha t^{\alpha-1}$ when $\alpha = 1.5$ (misspecification), $\pi_0^{(1)} = 0.2$ and $q = 1/3, 2/3$. The hazard $\lambda^{(1)}(t;\theta_0^*) = \theta_0^*$ (dotted line) is the pseudo-true value, which is a limit of the MLE under no censorship. It gives the best constant approximation to the true Weibull hazard. However, it can be seen that the hazard $\lambda^{(1)}(t;\tilde{\theta}_0) = \tilde{\theta}_0$, which is the limit of the MLE under censorship, is so different from the true hazard that censoring is heavy. Under the condition of the misspecification and heavy censorship, the MLE is not suitable. □

Example 3. Assume that the true cause-specific hazard of cause 1 is a piecewise-constant hazard:

$$\lambda_0^{(1)}(t) = \theta_{01}^{(1)} I(0 < t \leq a) + \theta_{02}^{(1)} I(t > a),$$

where $a > 0$ is a fixed time point. An assumed parametric model for cause 1 is $\mathcal{M}_{const}^{(1)}$ of (4). Unless $\theta_{01}^{(1)} = \theta_{02}^{(1)}$, this is a misspecified model.

Suppose that the true overall hazard is $\lambda_0(t) \equiv 1$. Then, $\eta(\theta)$ defined by (3) is

$$\eta(\theta) = \{\theta_{01}^{(1)} + e^{-a}(\theta_{02}^{(1)} - \theta_{01}^{(1)})\}\log\theta - \theta,$$

and it has a maximum at $\theta = \theta_0^* = \theta_{01}^{(1)} + e^{-a}(\theta_{02}^{(1)} - \theta_{01}^{(1)})$. This is the pseudo-true value, and it gives the best constant approximation to the true piecewise-constant hazard. In the left figure of Figure 2, the solid line shows the true piecewise-constant hazard with $a = 1$, $\theta_{01}^{(1)} = 0.4$ and $\theta_{02}^{(1)} = 0.1$, and the dotted line shows the pseudo-true value, $\theta_0^* = 0.290$. The figure on the right shows corresponding cumulative cause-specific hazards.

Suppose that the hazard of censoring time is constant ($= \beta$). Then censoring proportion is given by $q = \beta/(1+\beta)$. MLE of θ in the model $\mathcal{M}_{const}^{(1)}$ is given by (6), which converges in probability to

$$\tilde{\theta}_0 = \theta_{01}^{(1)} + e^{-a/(1-q)}(\theta_{02}^{(1)} - \theta_{01}^{(1)}) \text{ as } n \to \infty.$$

It holds that $\tilde{\theta}_0 = \theta_0^*$ if and only if $q = 0$ (no censorship) or $\theta_{01}^{(1)} = \theta_{02}^{(1)}$ (correct specification). The dashed lines in the left figure of Figure 2 show $\tilde{\theta}_0$ for $q = 1/3$ and $q = 2/3$. Inconsistency of the MLE is remarkable for $q = 2/3$ (heavy censorship). □

Example 4. We will take an example of MLE being consistent even if an assumed parametric model is misspecified. Assume that the true cause-specific hazards of causes 1 and 2 are constant; $\lambda_0^{(j)}(t) = \theta_0^{(j)}$, $j = 1, 2$. An assumed parametric model is

$$\mathcal{M}_{const}^{(1)=(2)} = \{\lambda^{(1)}(\cdot;\theta) = \lambda^{(2)}(\cdot;\theta) = \theta;\ \theta > 0\}. \tag{7}$$

In this model, it is assumed that the hazards of causes 1 and 2 are not only constant but also the same. This model is correct if $\theta_0^{(1)} = \theta_0^{(2)}$. The pseudo-true value is given by $\theta_0^* = (\theta_0^{(1)} + \theta_0^{(2)})/2$, and MLE is given by

$$\hat{\theta}_n = \left\{ \sum_{i=1}^n \delta_i I(V_i = 1 \text{ or } 2) \right\} \Big/ \left(2 \sum_{i=1}^n X_i \right). \tag{8}$$

It converges in probability to

$$\frac{E\{\delta_i I(V_i = 1 \text{ or } 2)\}}{2E(X_i)} = \frac{(\theta_0^{(1)} + \theta_0^{(2)}) \int_0^\infty \bar{F}_0(t)\bar{G}(t)\,dt}{2\int_0^\infty \bar{F}_0(t)\bar{G}(t)\,dt} = \theta_0^* \quad \text{as } n \to \infty.$$

Thus, in this case, the MLE is consistent though the assumed model is misspecified. □

2.3 Approximate likelihood procedure

As we have already seen in Section 2.1, the purpose of parametric estimation in model (1), which may have been misspecified, is to know the pseudo-true $\boldsymbol{\theta}_0^*$ which maximizes $\eta(\boldsymbol{\theta})$ of (3). The objective function $\eta(\boldsymbol{\theta})$ can be expressed as

$$\eta(\boldsymbol{\theta}) = \sum_{j \in C_1} \left[\int_0^\infty \log \lambda^{(j)}(t;\boldsymbol{\theta})\, dF_0^{(j)}(t) - \int_0^\infty \Lambda^{(j)}(t;\boldsymbol{\theta})\, dF_0(t) \right].$$

Replacing the cumulative incidence function $F_0^{(j)}$ and the overall lifetime distribution F_0 by the Aalen-Johansen estimator $F_n^{(j)}$ and the Kaplan-Meier estimator F_n, respectively, the approximated log-likelihood is defined by

$$L_n^*(\boldsymbol{\theta}) = n \sum_{j \in C_1} \left\{ \int_0^\infty \log \lambda^{(j)}(t;\boldsymbol{\theta})\, dF_n^{(j)}(t) - \int_0^\infty \Lambda^{(j)}(t;\boldsymbol{\theta})\, dF_n(t) \right\}. \tag{9}$$

The approximated maximum likelihood estimator (AMLE) $\hat{\boldsymbol{\theta}}_n^*$ is a maximizer of $L_n^*(\boldsymbol{\theta})$.

When there is no censoring at all, the Aalen-Johansen estimator and the Kaplan-Meier estimator reduce to empirical (sub-)distribution functions. Thus, under no censorship, the approximated log-likelihood $L_n^*(\boldsymbol{\theta})$ coincides with the log-likelihood $L_n(\boldsymbol{\theta})$ of (5), and hence AMLE and MLE are the same.

Define $\tau_H = \inf\{x : H(x) = 1\}$, i.e., τ_H is the right extreme of H. Then, from Stute and Wang's (1993) results for Kaplan-Meier integrals, it can be seen that, as $n \to \infty$,

$$\int_0^\infty \Lambda^{(j)}(t;\boldsymbol{\theta})\, dF_n(t) \overset{\text{a.s.}}{\to} \int_0^{\tau_H} f_0(t)\Lambda^{(j)}(t;\boldsymbol{\theta})\, dt, \quad j \in C_1.$$

Suzukawa (2002) has shown that for any F-integrable function φ, the Aalen-Johansen integral $\int_0^\infty \varphi\, dF_n^{(j)}$ almost surely converges to $\int_0^{\tau_H} \varphi\, dF^{(j)}$ as $n \to \infty$. From this result, it is seen that as $n \to \infty$,

$$\int_0^\infty \log \lambda^{(j)}(t;\boldsymbol{\theta})\, dF_n^{(j)}(t) \overset{\text{a.s.}}{\to} \int_0^{\tau_H} f_0^{(j)}(t)\Lambda^{(j)}(t;\boldsymbol{\theta})\, dt, \quad j \in C_1.$$

Thus, $L_n^*(\boldsymbol{\theta})/n$ almost surely converges to

$$\sum_{j \in C_1} \left\{ \int_0^{\tau_H} f_0^{(j)}(t)\Lambda^{(j)}(t;\boldsymbol{\theta})\, dt - \int_0^{\tau_H} f_0(t)\Lambda^{(j)}(t;\boldsymbol{\theta})\, dt \right\}. \tag{10}$$

Let $\tau_{F_0} = \inf\{x : F_0(x) = 1\}$ and $\tau_G = \inf\{x : G(x) = 1\}$. Since $\tau_H = \min(\tau_{F_0}, \tau_G)$, (10) is equal to $\eta(\boldsymbol{\theta})$ of (3) if $\tau_{F_0} \le \tau_G$. Thus, when $\tau_{F_0} \le \tau_G$, AMLE $\hat{\boldsymbol{\theta}}_n^*$ converges in probability to the pseudo-true $\boldsymbol{\theta}_0^*$ under suitable regularity conditions. In a large number of practical situations, $\tau_{F_0} = \tau_G = \infty$, and hence the assumption $\tau_{F_0} \le \tau_G$ is satisfied.

We next consider the asymptotic distribution of the AMLE under the condition $\tau_{F_0} \le \tau_G$. Define

$$Q_1(t;\boldsymbol{\theta}) = \exp\{-\sum_{j \in C_1} \Lambda^{(j)}(t;\boldsymbol{\theta})\}, \quad q^{(j)}(t;\boldsymbol{\theta}) = \lambda^{(j)}(t;\boldsymbol{\theta})Q_1(t;\boldsymbol{\theta}), \quad j \in C_1,$$

and let

$$\varphi^{(j)}(t;\boldsymbol{\theta}) = \begin{cases} \log q^{(j)}(t;\boldsymbol{\theta}) & j \in C_1 \\ \log Q_1(t;\boldsymbol{\theta}) & j \in C_2. \end{cases}$$

Then the approximated log-likelihood $L_n^*(\boldsymbol{\theta})$ can be expressed as

$$L_n^*(\boldsymbol{\theta}) = n \sum_{j \in C} \int_0^\infty \varphi^{(j)}(t;\boldsymbol{\theta})\, dF_n^{(j)}(t).$$

The AMLE $\hat{\boldsymbol{\theta}}_n^*$ is a solution of an equation

$$\mathbf{0} = \frac{\partial}{\partial \boldsymbol{\theta}} L_n^*(\hat{\boldsymbol{\theta}}_n^*) = \frac{\partial}{\partial \boldsymbol{\theta}} L_n^*(\boldsymbol{\theta}_0^*) + \left\{ \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} L_n^*(\bar{\boldsymbol{\theta}}_n^*) \right\} (\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}_0^*),$$

where $\bar{\boldsymbol{\theta}}_n^*$ lies between $\hat{\boldsymbol{\theta}}_n^*$ and $\boldsymbol{\theta}_0^*$. Thus, we have

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}_0^*) = \left\{ -n^{-1} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} L_n^*(\bar{\boldsymbol{\theta}}_n^*) \right\}^{-1} \left\{ n^{-1/2} \frac{\partial}{\partial \boldsymbol{\theta}} L_n^*(\boldsymbol{\theta}_0^*) \right\}. \tag{11}$$

The first term on the right-hand side is the sum of Aalen-Johansen integrals:

$$\sum_{j \in C} \int_0^\infty -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \varphi^{(j)}(t;\bar{\boldsymbol{\theta}}_n^*)\, dF_n^{(j)}(t),$$

and it converges in probability to a $p \times p$ matrix

$$\boldsymbol{J}^*(\boldsymbol{\theta}_0^*) = \sum_{j \in C} \int_0^\infty -f_0^{(j)}(t) \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \varphi^{(j)}(t;\boldsymbol{\theta}_0^*)\, dt. \tag{12}$$

On the other hand, the second term on the right-hand side of (11) can be written as

$$\begin{aligned} n^{-1/2} \frac{\partial}{\partial \boldsymbol{\theta}} L_n^*(\boldsymbol{\theta}_0^*) &= \sum_{j \in C} n^{1/2} \int_0^\infty \frac{\partial}{\partial \boldsymbol{\theta}} \varphi^{(j)}(t;\boldsymbol{\theta}_0^*)\, dF_n^{(j)}(t) \\ &= [\boldsymbol{I}_p, \ldots, \boldsymbol{I}_p] \times \boldsymbol{s}_n, \end{aligned}$$

where $\boldsymbol{I}_p$ is the p-dimensional identity matrix, and $\boldsymbol{s}_n$ is a pk-dimensional vector defined by

$$\boldsymbol{s}_n = \begin{bmatrix} n^{1/2} \int_0^\infty \frac{\partial}{\partial \boldsymbol{\theta}} \varphi^{(1)}(t;\boldsymbol{\theta}_0^*)\, dF_n^{(1)}(t) \\ \vdots \\ n^{1/2} \int_0^\infty \frac{\partial}{\partial \boldsymbol{\theta}} \varphi^{(k)}(t;\boldsymbol{\theta}_0^*)\, dF_n^{(k)}(t) \end{bmatrix}.$$

Suppose the following assumptions

(A1): for each $j \in C$, there exists a $p \times p$ matrix

$$\int_0^\infty \frac{f_0^{(j)}(t)}{\bar{G}(t)} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \varphi^{(j)}(t; \boldsymbol{\theta}_0^*) \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}'} \varphi^{(j)}(t; \boldsymbol{\theta}_0^*) \right\} dt.$$

and

(A2): for all $j \in C$,

$$\int_0^\infty f_0^{(j)}(t) \sqrt{C(t)} \left| \frac{\partial}{\partial \boldsymbol{\theta}} \varphi^{(j)}(t; \boldsymbol{\theta}_0^*) \right| dt < \infty,$$

where $C(t) = \int_0^t dG(x)/\{\bar{H}(x)\bar{G}(x)\}$.

Then, similarly to Theorem 2 of Suzukawa (2002), it can be shown that the limiting distribution of $\boldsymbol{s}_n$ is pk-variate normal with a mean vector

$$\left[\boldsymbol{\psi}^{(1)}(0; \boldsymbol{\theta}_0^*), \ldots, \boldsymbol{\psi}^{(k)}(0; \boldsymbol{\theta}_0^*)\right]'$$

and a covariance matrix

$$\boldsymbol{\Omega}(\boldsymbol{\theta}_0^*) = \begin{bmatrix} \boldsymbol{\Omega}_{11}(\boldsymbol{\theta}_0^*) & \cdots & \boldsymbol{\Omega}_{1k}(\boldsymbol{\theta}_0^*) \\ \vdots & & \vdots \\ \boldsymbol{\Omega}_{k1}(\boldsymbol{\theta}_0^*) & \cdots & \boldsymbol{\Omega}_{kk}(\boldsymbol{\theta}_0^*) \end{bmatrix},$$

where, for $0 \le t < \infty$,

$$\boldsymbol{\psi}^{(j)}(t; \boldsymbol{\theta}) = \int_t^\infty f_0^{(j)}(x) \frac{\partial \varphi^{(j)}(x; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} dx, \quad j \in C,$$

and, for $0 \le j, l \le k$, $\boldsymbol{\Omega}_{jl}(\boldsymbol{\theta})$ is a $p \times p$ matrix defined by

$$\boldsymbol{\Omega}_{jl}(\boldsymbol{\theta}) = \begin{cases} \displaystyle\int_0^\infty \frac{f_0^{(j)}(t)}{\bar{G}(t)} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \varphi^{(j)}(t; \boldsymbol{\theta}_0^*) \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}'} \varphi^{(j)}(t; \boldsymbol{\theta}_0^*) \right\} dt & \\ \displaystyle -\int_0^\infty \frac{\bar{F}_0(t)}{\{\bar{H}(t)\}^2} \boldsymbol{\psi}^{(j)}(t; \boldsymbol{\theta}) \boldsymbol{\psi}^{(j)'}(t; \boldsymbol{\theta})\, dG(t) - \boldsymbol{\psi}^{(j)}(0; \boldsymbol{\theta}) \boldsymbol{\psi}^{(j)'}(0; \boldsymbol{\theta}), & j = l, \\ \displaystyle -\int_0^\infty \frac{\bar{F}_0(t)}{\{\bar{H}(t)\}^2} \boldsymbol{\psi}^{(j)}(t; \boldsymbol{\theta}) \boldsymbol{\psi}^{(l)'}(t; \boldsymbol{\theta})\, dG(t) - \boldsymbol{\psi}^{(j)}(0; \boldsymbol{\theta}) \boldsymbol{\psi}^{(l)'}(0; \boldsymbol{\theta}), & j \ne l. \end{cases}$$

Thus, the limiting distribution of $n^{-1/2} \partial L_n^*(\boldsymbol{\theta}_0^*)/\partial \boldsymbol{\theta}$ is p-variate normal with a mean vector

$$\sum_{j \in C} \psi^{(j)}(0; \boldsymbol{\theta}_0^*) = \frac{\partial}{\partial \boldsymbol{\theta}} \eta(\boldsymbol{\theta}_0^*) = 0.$$

We redefine the function η of (3) as

$$\eta(t; \boldsymbol{\theta}) = \sum_{j \in C_1} \left[\int_t^\infty f_0^{(j)}(x) \log \lambda^{(j)}(x; \boldsymbol{\theta})\, dx - \int_t^\infty f_0(x) \Lambda^{(j)}(x; \boldsymbol{\theta})\, dx \right],$$

($\eta(\boldsymbol{\theta})$ of (3) is $\eta(0; \boldsymbol{\theta})$). Then we have

$$\sum_{j \in C} \psi^{(j)}(t; \boldsymbol{\theta}_0^*) = \frac{\partial}{\partial \boldsymbol{\theta}} \eta(t; \boldsymbol{\theta}_0^*),$$

and we can express the limiting covariance matrix of $n^{-1/2} \partial L_n^*(\boldsymbol{\theta}_0^*)/\partial \boldsymbol{\theta}$ as

$$\begin{aligned} \boldsymbol{I}^*(\boldsymbol{\theta}_0^*) &= \sum_{j \in C} \sum_{l \in C} \boldsymbol{\Omega}_{jl}(\boldsymbol{\theta}_0^*) \qquad (13) \\ &= \sum_{j \in C} \left[\int_0^\infty \frac{f_0^{(j)}(t)}{\bar{G}(t)} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \varphi^{(j)}(t; \boldsymbol{\theta}_0^*) \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}'} \varphi^{(j)}(t; \boldsymbol{\theta}_0^*) \right\} dt \right] \\ &\quad - \int_0^\infty \frac{\bar{F}_0(t)}{\{\bar{H}(t)\}^2} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \eta(t; \boldsymbol{\theta}_0^*) \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}'} \eta(t; \boldsymbol{\theta}_0^*) \right\} dG(t). \end{aligned}$$

Thus, we obtain the following results.
Theorem 1. *Under the assumptions* $\tau_{F_0} \le \tau_G$, (A1) *and* (A2)*, it holds that*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^*_n - \boldsymbol{\theta}^*_0) \xrightarrow{\mathrm{d}} N_p\left(0,\ \boldsymbol{J}^{*-1}(\boldsymbol{\theta}^*_0) I^*(\boldsymbol{\theta}^*_0) \boldsymbol{J}^{*-1}(\boldsymbol{\theta}^*_0)\right) \quad \text{as } n \to \infty,$$

where $p \times p$ *matrices* $\boldsymbol{J}^*(\boldsymbol{\theta}^*_0)$ *and* $\boldsymbol{I}^*(\boldsymbol{\theta}^*_0)$ *are defined by* (12) *and* (13)*, respectively.*
Corollary 1. *When* $\mathcal{M}^{C_1}$ *of* (1) *contains the true hazards, under the same assumptions as Theorem 1, it holds that*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^*_n - \boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} N_p\left(0,\ \boldsymbol{J}^{*-1}_0(\boldsymbol{\theta}_0) \boldsymbol{I}^*_0(\boldsymbol{\theta}_0) \boldsymbol{J}^{*-1}_0(\boldsymbol{\theta}_0)\right) \tag{14}$$

as $n \to \infty$*, where*

$$\begin{aligned} \boldsymbol{J}^*_0(\boldsymbol{\theta}_0) &= \sum_{j \in C_1} \int_0^\infty \lambda^{(j)}(t;\boldsymbol{\theta}_0) \bar{F}_0(t) \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda^{(j)}(t;\boldsymbol{\theta}_0) \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}'} \log \lambda^{(j)}(t;\boldsymbol{\theta}_0) \right\} dt \quad \text{and} \\ \boldsymbol{I}^*_0(\boldsymbol{\theta}_0) &= \sum_{j \in C_1} \int_0^\infty \frac{\lambda^{(j)}(t;\boldsymbol{\theta}_0) \bar{F}_0(t)}{\bar{G}(t)} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda^{(j)}(t;\boldsymbol{\theta}_0) \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}'} \log \lambda^{(j)}(t;\boldsymbol{\theta}_0) \right\} dt. \end{aligned}$$

3 Asymptotic comparison of estimators

If both MLE and AMLE are consistent, we are interested in their efficiencies. In this section, we consider asymptotic comparison of these estimators. When $\mathcal{M}^{C_1}$ is correct, under the ordinary regularity conditions for the likelihood (5), the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ is $N_p\left(0,\ \boldsymbol{I}_0^{-1}(\boldsymbol{\theta}_0)\right)$, where

$$\boldsymbol{I}_0(\boldsymbol{\theta}_0) = \sum_{j \in C_1} \int_0^\infty \lambda^{(j)}(t;\boldsymbol{\theta}_0) \bar{F}_0(t) \bar{G}(t) \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda^{(j)}(t;\boldsymbol{\theta}_0) \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}'} \log \lambda^{(j)}(t;\boldsymbol{\theta}_0) \right\} dt.$$

Theorem 2. *When the assumed parametric model* $\mathcal{M}^{C_1}$ *of* (1) *is correct, the MLE* $\hat{\boldsymbol{\theta}}_n$ *is asymptotically more efficient than the AMLE* $\hat{\boldsymbol{\theta}}^*_n$.

Proof.
The asymptotic covariance matrices of the AMLE $\hat{\boldsymbol{\theta}}^*_n$ and MLE $\hat{\boldsymbol{\theta}}_n$ are given by

$$\boldsymbol{J}^{*-1}_0(\boldsymbol{\theta}_0) I^*_0(\boldsymbol{\theta}_0) \boldsymbol{J}^{*-1}_0(\boldsymbol{\theta}_0) \quad \text{and} \quad \boldsymbol{I}_0^{-1}(\boldsymbol{\theta}_0),$$

respectively. Define, for $j \in C_1$,

$$\boldsymbol{l}_i^{(j)} = \delta_i\, I(V_i = j)\, \frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda^{(j)}(X_i;\boldsymbol{\theta}_0),$$

then we can express as

$$\boldsymbol{I}_0(\boldsymbol{\theta}_0) = \sum_{j \in C_1} E\left\{ \boldsymbol{l}_i^{(j)} \boldsymbol{l}_i^{(j)'} \right\}, \quad \boldsymbol{I}^*_0(\boldsymbol{\theta}_0) = \sum_{j \in C_1} E\left[\left\{ \boldsymbol{l}_i^{(j)}/\bar{G}(X_i) \right\} \left\{ \boldsymbol{l}_i^{(j)}/\bar{G}(X_i) \right\}' \right]$$

and

$$\boldsymbol{J}^*_0(\boldsymbol{\theta}_0) = \sum_{j \in C_1} E\left[\boldsymbol{l}_i^{(j)} \left\{ \boldsymbol{l}_i^{(j)}/\bar{G}(X_i) \right\}' \right].$$

For any non-zero p-dimensional vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, we have expressions

$$\boldsymbol{a}' \boldsymbol{I}^*_0(\boldsymbol{\theta}_0) \boldsymbol{a} = \sum_{j \in C_1} E\left[\left\{ \boldsymbol{a}' \boldsymbol{l}_i^{(j)}/\bar{G}(X_i) \right\}^2 \right], \quad \boldsymbol{b}' \boldsymbol{I}_0(\boldsymbol{\theta}_0) \boldsymbol{b} = \sum_{j \in C_1} E\left\{ \left(\boldsymbol{b}' \boldsymbol{l}_i^{(j)} \right)^2 \right\}$$

and

$$\boldsymbol{b}'\boldsymbol{J}_0^*(\boldsymbol{\theta}_0)\boldsymbol{a} = \sum_{j \in C_1} E\left[\left(\boldsymbol{b}'\boldsymbol{l}_i^{(j)}\right)\left\{\boldsymbol{a}'\boldsymbol{l}_i^{(j)}/\bar{G}(X_i)\right\}\right].$$

From Cauchy-Schwarz inequality, it holds that

$$\{\boldsymbol{a}'\boldsymbol{I}_0^*(\boldsymbol{\theta}_0)\boldsymbol{a}\} \times \{\boldsymbol{b}'\boldsymbol{I}_0(\boldsymbol{\theta}_0)\boldsymbol{b}\} \geq \{\boldsymbol{b}'\boldsymbol{J}_0^*(\boldsymbol{\theta}_0)\boldsymbol{b}\}^2.$$

Since $\boldsymbol{b}$ is arbitrary, we can put $\boldsymbol{b} = \boldsymbol{I}_0^{-1}(\boldsymbol{\theta}_0)\boldsymbol{J}_0^*(\boldsymbol{\theta}_0)\boldsymbol{a}$. Then we have

$$\boldsymbol{a}'\boldsymbol{I}_0^*(\boldsymbol{\theta}_0)\boldsymbol{a} \geq \boldsymbol{a}'\boldsymbol{J}_0^*(\boldsymbol{\theta}_0)\boldsymbol{I}_0^{-1}(\boldsymbol{\theta}_0)\boldsymbol{J}_0^*(\boldsymbol{\theta}_0)\boldsymbol{a}.$$

This implies that $\boldsymbol{J}_0^{*-1}(\boldsymbol{\theta}_0)\boldsymbol{I}_0^*(\boldsymbol{\theta}_0)\boldsymbol{J}_0^{*-1}(\boldsymbol{\theta}_0) - \boldsymbol{I}_0^{-1}(\boldsymbol{\theta}_0)$ is positive semi-definite. □

The above result shows that information is lost by the approximated likelihood method when an assumed parametric model is correct. Similar results have been shown by Oakes (1986) and by Suzukawa *et al.* (2001) in a non-competing risks framework.

Example 5. Assume that the overall lifetime distribution and the censoring distribution are exponential distributions with hazards 1 and β, respectively. Suppose that the true cause-specific hazard of cause 1 is constant; $\lambda_0^{(1)}(t) = \pi_0^{(1)}$, where $0 < \pi_0^{(1)} < 1$ is the probability of failure by cause 1. Then the censoring proportion is $q = P(\delta_i = 0) = \beta/(1+\beta)$. An assumed parametric model is the constant model $\mathcal{M}_{const}^{(1)}$ of (4). This model is correct, and the true parameter is $\theta_0 = \pi_0^{(1)}$. MLE of θ is given by (6) and it is a consistent estimator of $\theta_0 = \pi_0^{(1)}$. Moreover, it can be shown that, as $n \to \infty$,

$$\sqrt{n}(\hat{\theta}_n - \pi_0^{(1)}) \xrightarrow{\text{d}} N(0,\, \pi_0^{(1)}/(1-q)).$$

On the other hand, AMLE is given by

$$\hat{\theta}_n^* = F_n^{(1)}(\infty)/\int_0^\infty t\, dF_n(t) \tag{15}$$

and it is also consistent to $\theta_0 = \pi_0^{(1)}$. If $q < 1/2$, the assumptions (A1) and (A2) in Section 2.3 are satisfied. Then it is seen that

$$\sqrt{n}(\hat{\theta}_n^* - \pi_0^{(1)}) \xrightarrow{\text{d}} N(0,\, \pi_0^{(1)}(1-q)/(1-2q)) \quad \text{as } n \to \infty.$$

The asymptotic relative efficiency (ARE) of the AMLE $\hat{\theta}_n^*$ with respect to the MLE $\hat{\theta}_n$ is given by

$$\text{ARE}(\hat{\theta}_n^*; \hat{\theta}_n) = (1-2q)/(1-q)^2 \quad \text{for } 0 \leq q < 1/2.$$

It is a strictly decreasing function of the censoring proportion q and is less than one unless $q = 0$ (no censoring). This indicates that information loss by the AMLE increases with increase in the censoring proportion. □

Example 6. Let us consider the case of Example 4 again. This is a case in which MLE is consistent although an assumed parametric model is not always correct. Assume that the true cause-specific hazards of causes 1 and 2 are constant; $\lambda_0^{(j)}(t) = \theta_0^{(j)}$, $j = 1, 2$. An assumed parametric model is (7), in which the two hazards are assumed to be not only constant but also the same. In this model, the pseudo-true value is $\theta_0^* = (\theta_0^{(1)} + \theta_0^{(2)})/2$. MLE $\hat{\theta}_n$ is given by (8). If $\mu_X = E(X_i) < \infty$, it converges in probability to the pseudo-true value as $n \to \infty$. We can write as

$$\sqrt{n}(\hat{\theta}_n - \theta_0^*) = \left[n^{-1/2}\sum_{i=1}^n \{\delta_i I(V_i = 1 \text{ or } 2) - (\theta_0^{(1)} + \theta_0^{(2)})X_i\}\right]/(2\bar{X}),$$

where $\bar{X} = \sum_{i=1}^{n} X_i/n$. Under the condition $E(X_i^2) < \infty$, the asymptotic distribution of the numerator is $N(0, 2\theta_0^* \mu_X)$ and the denominator converges in probability to $2\mu_X$ as $n \to \infty$. Thus, it is seen that

$$\sqrt{n}(\hat{\theta}_n - \theta_0^*) \xrightarrow{\text{d}} N(0,\ \theta_0^*/(2\mu_X)) \quad \text{as } n \to \infty.$$

On the other hand, AMLE is given by

$$\hat{\theta}_n^* = \left\{ F_n^{(1)}(\infty) + F_n^{(1)}(\infty) \right\} / \left\{ 2 \int_0^{\infty} t\, dF_n(t) \right\}.$$

If $\tau_{F_0} \leq \tau_G$, it converges in probability to the pseudo-true value $\theta_0^* = (\theta_0^{(1)} + \theta_0^{(2)})/2$ as $n \to \infty$. From Theorem 1, it is seen that, under assumptions $\tau_{F_0} \leq \tau_G$ and $\int_0^\infty \bar{F}_0(t)\{\bar{G}(t)\}^{-1}\, dt < \infty$,

$$\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) \xrightarrow{\text{d}} N\left(0,\ \frac{\theta_0^*}{2\mu_0^2} \int_0^{\infty} \bar{F}_0(t)\{\bar{G}(t)\}^{-1}\, dt\right) \quad \text{as } n \to \infty,$$

where $\mu_0 = E(T_i) = \int_0^\infty \bar{F}_0(t)\, dt$. The ARE of $\hat{\theta}_n^*$ with respect to $\hat{\theta}_n$ is

$$\frac{\mu_0^2}{\mu_X \times \int_0^\infty \bar{F}_0(t)\{\bar{G}(t)\}^{-1}\, dt} = \frac{\left\{\int_0^\infty \bar{F}_0(t)\, dt\right\}^2}{\left\{\int_0^\infty \bar{F}_0(t)\bar{G}(t)\, dt\right\} \times \left\{\int_0^\infty \bar{F}_0(t)\{\bar{G}(t)\}^{-1}\, dt\right\}} \leq 1.$$

The last inequality follows from the Cauchy-Schwarz inequality. Thus, the MLE is asymptotically more efficient than the AMLE. □

4 Applications

4.1 Application to simulated data

We assume that there are two competing risks and let $U^{(j)}$, $j = 1, 2$ be the potential unobservable time to occurrence of the jth risk. Suppose that the joint survival function of the potential times is

$$P(U^{(1)} > u_1, U^{(2)} > u_2) = \{1 + \rho(u_1 + u_2)\}^{-1/\rho} \tag{16}$$

with a parameter $\rho \geq 0$. This has been discussed as an example of potential times in competing risks by Klein and Moeschberger (2003). The parameter ρ measures dependence between $U^{(1)}$ and $U^{(2)}$. Here, $\rho = 0$ implies independence and $\rho > 0$ implies dependence. Kendall's tau is $\rho/(\rho + 2)$. Marginal survival functions are

$$P(U^{(1)} > u) = P(U^{(2)} > u) = (1 + \rho u)^{-1/\rho}.$$

Cause-specific hazard functions are

$$\lambda_0^{(1)}(t) = \lambda_0^{(2)}(t) = (1 + 2\rho t)^{-1}. \tag{17}$$

Cause-specific cumulative incidence functions are

$$F_0^{(1)}(t) = F_0^{(2)}(t) = \frac{1}{2}\left\{1 - (1 + 2\rho t)^{-1/\rho}\right\}. \tag{18}$$

The potential times $U^{(1)}$ and $U^{(2)}$ are unobservable. When there is no censorship, we can observe the failure time $T = \min(U^{(1)}, U^{(2)})$ and the cause of failure. The dependence parameter ρ is not testable with only the observable variates.

The joint survival function (16) is constructed by Clayton's (1978) bivariate copulas. Thus, random variates $(U^{(1)}, U^{(2)})$ can be generated by Genest and MacKay (1986)'s algorithm. For $\rho = 1/2$, $n = 300$ samples were generated. Scatter plots of them are presented in Figure 3. Kendall's tau of the population is $\rho/(\rho + 2) = 0.2$, and Kendall's tau of the simulated samples is 0.189. The sample correlation coefficient is 0.396.

We also assume that the censoring time Y is distributed as the standard exponential distribution. We can only observe $\min(U^{(1)}, U^{(2)}, Y)$ and cause of the failure. In the simulated competing risks data, 90 (30.0%) and 101 (33.7%) samples failed by cause 1 and 2, respectively, and 109 (36.3%) samples were rightly censored.

Since the two causes are symmetric, we focus only on cause 1. In Figure 4, the curve shows the true cumulative incidence function (18) with $\rho = 1/2$. The step function is its Aalen-Johansen estimate obtained from the simulated competing risks data. From this, we can see that the true cumulative incidence curve is well-estimated by the nonparametric Aalen-Johansen estimates.

We consider parametric estimation of the cause-specific hazard of cause 1. If the potential times are independent ($\rho = 0$), then the cause-specific hazard given by (17) is constant. Here the true value of ρ is 1/2 and the potential times are dependent. However, the dependency cannot be seen from the competing risks data. If we have mistakenly considered that the potential times are independent ($\rho = 0$), then the constant hazard model (4) would be assumed. This is a misspecified parametric model.

Under the misspecified model (4), the unknown constant parameter θ can be estimated by MLE and AMLE. The MLE $\hat{\theta}_n$ is given by (6) and the AMLE $\hat{\theta}_n^*$ is given by (15). From the simulated samples, we obtained two estimates of θ as $\hat{\theta}_n = 0.770$ and $\hat{\theta}_n^* = 0.670$. The left figure of Figure 5 shows these constant estimates and the true cause-specific hazard function. The right figure shows the corresponding cumulative hazard functions. The cause-specific hazard of cause 1 is overly estimated by the MLE (dashed line). On the other hand, the estimated line by the AMLE (dotted line) is nearer the true curve (solid curve).

4.2 Application to failure data for electrical appliance test

Let us consider the data shown in Table 1 (Nelson 1970; Lawless 1982), which shows failure times of 36 electrical appliances. The appliances were operated repeatedly by an automatic testing machine. Each failure time in the table is the number of cycles of use completed until the appliance failed. The appliances were exposed to 18 causes of failure. Each cause is denoted by a failure code number. There were some censored observations, since it was not always possible to operate the testing machine long enough for an appliance to fail. Failure codes of the censored observations are indicated by 0.

Only failure codes 6 and 9 appear more than twice. Figure 6 shows Nelson-Aalen estimates of cumulative cause-specific hazard functions and Aalen-Johansen estimates of cumulative incidence functions. It can be seen that causes 6 and 9 are main causes of failure. Estimates of probabilities that the appliance fails by causes 6 and 9 are 0.198 and 0.509, respectively. We shall focus on these two causes.

Let $C_1 = \{6, 9\}$, which is a set of sub-causes in which we are interested. Denote the cause-specific hazard functions of causes 6 and 9 by $\lambda^{(6)}$ and $\lambda^{(9)}$, respectively. The following three parametric models are considered.
Constant hazard model:

$$\mathcal{M}_{\text{const}}^{C_1} = \left\{\lambda^{(j)}(t; \alpha^{(j)}) = \alpha^{(j)};\ \alpha^{(j)} > 0, j \in C_1\right\}, \tag{19}$$

Weibull hazard model:

$$\mathcal{M}_{\text{Weibull}}^{C_1} = \left\{\lambda^{(j)}(t; \alpha^{(j)}, \beta^{(j)}) = \alpha^{(j)}\beta^{(j)}t^{\beta^{(j)}-1};\ \alpha^{(j)} > 0, \beta^{(j)} > 0, j \in C_1\right\}, \tag{20}$$

Log-Logistic (LL) hazard model:

$$\mathcal{M}_{\text{LL}}^{C_1} = \left\{\lambda^{(j)}(t; \alpha^{(j)}, \beta^{(j)}) = \frac{\alpha^{(j)}\beta^{(j)}t^{\beta^{(j)}-1}}{1 + \alpha^{(j)}t^{\beta^{(j)}}};\ \alpha^{(j)} > 0, \beta^{(j)} > 0, j \in C_1\right\}. \tag{21}$$

The model $\mathcal{M}_{\text{const}}^{C_1}$ is the simplest model, and it is important in order to approximate a hazard function by a constant value. The model $\mathcal{M}_{\text{Weibull}}^{C_1}$ is a generalization of the constant model. In $\mathcal{M}_{\text{Weibull}}^{C_1}$, if $\beta^{(j)} < 1$ ($\beta^{(j)} > 1$), then $\lambda^{(j)}$ is decreasing (increasing). If $\beta^{(j)} = 1$, then $\lambda^{(j)}$ is constant. From the Nelson-Aalen estimate of the cumulative cause-specific hazard of cause 9 (the solid step function in the left figure of Figure 6), we can see that the cumulative hazard of cause 9 is convex until 4000 cycles and it is concave after 4000 cycles. Thus, the cause-specific hazard of cause 9 is increasing until 4000 cycles, and it is decreasing after 4000 cycles. The model $\mathcal{M}_{\rm LL}^{C_1}$ includes such type of hazard function. In $\mathcal{M}_{\rm LL}^{C_1}$, if $\beta^{(j)} > 1$, then $\lambda^{(j)}(t; \alpha^{(j)}, \beta^{(j)})$ is increasing for $t < \{(\beta^{(j)} - 1)/\alpha^{(j)}\}^{1/\beta^{(j)}}$ and is decreasing for $t > \{(\beta^{(j)} - 1)/\alpha^{(j)}\}^{1/\beta^{(j)}}$.

In each model, estimates of parameters are obtained by MLE and AMLE. The estimates are given in Table 2. Figure 7 shows estimates of the cumulative cause-specific hazards of causes 6 and 9. The figures on the left and right show estimates for causes 6 and 9, respectively. In these figures, the solid step functions are the Nelson-Aalen estimates of the cumulative cause-specific hazards, and the dotted step functions are approximate pointwise 90% confidence limits of them. The limits are obtained by log-transformation. The solid and dashed curves show parametric estimates by the MLE and AMLE, respectively.

For each cause, the Nelson-Aalen estimate does not look like a straight line. It is quite likely that the constant model $\mathcal{M}_{\rm const}^{C_1}$ is not correct. Obviously, in this model, the hazards are under-estimated by the MLE. The estimates by the MLE do not give good approximations to the Nelson-Aalen estimates. Using the simplest model $\mathcal{M}_{\rm const}^{C_1}$, the AMLE is superior to the MLE.

In cause 6, the Weibull and LL models give similar estimates, and there is no notable difference between the MLE and AMLE. On the other hand, for cause 9, the parametric estimates show different features between the MLE and AMLE. Generally, in the Weibull and LL models, estimates obtained by the MLE is well-fitted. However, they are strongly influenced by two observations failed by cause 9 after 5000 cycles, and hence their fitness is poor before 5000 cycles. On the other hand, until 5000 cycles, the estimates by the AMLE are similar to the nonparametric Nelson-Aalen estimates.

4.3 Application to failure data for pneumatic tire test

Let us consider the data shown in Table 3 from a laboratory test on 171 pneumatic tires (Davis and Lawrance 1989). The test involved rotating the tires against a steel drum until some type of failure occurred. Failures were classified into six modes: open joint on the inner liner (failure mode 1), rubber chunking on the shoulder (2), loose casing low on the sidewall (3), cracking of the tread rubber (4), cracking on the sidewall (5) and all other causes (6). There were 21 censored observations (failure code 0).

Figure 8 shows Nelson-Aalen estimates of cumulative cause-specific hazards and Aalen-Johansen estimates of cumulative incidences. It can be seen that cause-specific hazards of some failure codes change at 200 hours. In this test, the inflation pressure in unfailed tires was reduced at 200 hours . Naturally this changes the hazards. Thus, as a simple parametric model, a piecewise-constant hazard model,

$$\lambda^{(j)}(t;\, \theta_1^{(j)}, \theta_2^{(j)}) = \theta_1^{(j)} I(t \leq 200) + \theta_2^{(j)} I(t > 200), \quad j = 1, 2, \ldots, 6, \tag{22}$$

is considered.

Table 4 shows estimates of the parameters obtained by the MLE and AMLE. In estimates of $\theta_1^{(j)}$, $j = 1, \ldots, 6$, which are hazards before 200 hours, differences between the MLE and AMLE are slight. Notable differences appear in $\theta_2^{(j)}$, $j = 1, \ldots, 6$, which are hazards after 200 hours. Figure 9 shows cumulative cause-specific hazards. The dotted step-functions are the Nelson-Aalen estimates, and the solid and dashed lines are the parametric estimates obtained by the MLE and AMLE, respectively. In each cause of failure, the difference between the MLE and AMLE is not so large. The Nelson-Aalen estimates are approximately piecewise-linear. The piecewise-constant model (22) may be well-fitted. In this case, it is better to use the estimates obtained by the MLE.

The effect on each cause of the reduction of inflation pressure can be evaluated by the ratio $\theta_2^{(j)}/\theta_1^{(j)}$. Based on the estimates obtained by the MLE, the ratios are 5.66 (cause 1), 15.3 (cause 2), 1.91 (cause 3), 33.9 (cause 4), 3.64 (cause 5) and 7.63(cause 6). The reduction of inflation pressure accelerates failure due to each cause. The acceleration is remarkable for causes 2 and 4.

5 Concluding remarks

This paper presents a procedure for estimating parameters of cause-specific hazard functions in competing risks. When there are censored observations, the ordinary likelihood procedure is not always appropriate. If an assumed parametric model is misspecified, it is not always consistent to the pseudo-true parameters. On the other hand, the proposed approximate maximum likelihood estimators (AMLE) are consistent to the pseudo-true parameters under misspecification and censorship. In other words, when sample size is sufficiently large, the best approximations to the true hazards are given by AMLE even if the assumed parametric model is misspecified.

The asymptotic normality of AMLE was shown in this paper. Under correct parametric models, AMLE is asymptotically inefficient compared with the ordinary maximum likelihood estimator (MLE). In conclusion, AMLE is superior to MLE when fitness of an assumed parametric model is not so good. On the other hand, when the model is well-fitted, AMLE is inferior to MLE. Selection of an appropriate parametric model is very important in analysis of censored competing risks data. It is debatable which estimating procedure should be used in practical situations. There is room for further investigation on this point.

References

[1] Aalen, O. O. (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.* **3**, 15–27.

[2] Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations, *Scandinavian Journal of Statistics*, **5**, 141–150.

[3] Chiang, C. L. (1968). *Introduction to Stochastic Process in Biostatistics.* New York: John Wiley and Sons.

[4] Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141–151.

[5] Crowder, M. J. (2001). *Classical Competing Risks.* London: Chapmann and Hall.

[6] David, H. A. and Moeschberger, M. L. (1978). *Theory of Competing Risks.* London: Griffin.

[7] Davis, T. P. and Lawrance, A. J. (1989). The likelihood for competing risk survival analysis. *Scand. J. Statist.* **16**, 23–28.

[8] Genest, C. and MacKay, J. (1986). Copules archmédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canad. J. Statist.* **14**, 145–159.

[9] Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data, 2nd ed.* New York: John Wiley and Sons.

[10] Kaplan, E. L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457–481.

[11] Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis, Techniques for Censored and Truncated Data, 2nd ed.* New York: Springer.

[12] Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* New York: John Wiley and Sons.

[13] Nelson, W. B. (1970). Hazard plotting methods for analysis of life data with different failure codes. *J. Qual. Technol.* **2**, 126–149.

[14] Oakes, D. (1986). An approximate likelihood procedure for censored data. *Biometrics*, **42**, 177–182.

[15] Stute, W. and Wang, J. L. (1993). The strong law under random censorship. *Annals of Statistics*, **21**, 1591–1607.

[16] Suzukawa, A. (2002). Asymptotic properties of Aalen-Johansen integrals for competing risks data. *J. Jpn. Statist. Soc.* **32**, 77–93.

[17] Suzukawa, A., Imai, H. and Sato, Y. (2001). Kullback-Leibler information consistent estimation for censored data. *Ann. Inst. Statist. Math.* **53**, 262–276.

[18] Suzukawa, A. and Taneichi, N. (2003). Semiparametric estimation based on parametric modeling of the cause-specific hazard ratios in competing risks. *J. Multi. Anal.* **87**, 80–100.

Table 1: *Failure data for electrical appliance test*

Failure time	Code	Failure time	Code	Failure time	Code	Failure time	Code
11	1	35	15	49	15	170	6
329	6	381	6	708	6	958	10
1062	5	1167	9	1594	2	1925	9
1990	9	2223	9	2327	6	2400	9
2451	5	2471	9	2551	9	2565	0
2568	9	2694	9	2702	10	2761	6
2831	2	3034	9	3059	6	3112	9
3214	9	3478	9	3504	9	4329	9
6367	0	6976	9	7846	9	13403	0

Table 2: Estimates of parameters in models (19), (20) and (21) for electrical appliance tests

Model	Constant (19)	Weibull (20)		Log-Logistic (21)	
Parameter	$\alpha^{(6)}$	$\alpha^{(6)}$	$\beta^{(6)}$	$\alpha^{(6)}$	$\beta^{(6)}$
MLE	7.05×10^{-5}	6.97×10^{-4}	0.721	2.27×10^{-3}	0.589
AMLE	8.34×10^{-5}	4.51×10^{-4}	0.791	3.05×10^{-4}	0.859
Parameter	$\alpha^{(9)}$	$\alpha^{(9)}$	$\beta^{(9)}$	$\alpha^{(9)}$	$\beta^{(9)}$
MLE	1.71×10^{-4}	8.57×10^{-7}	1.628	1.42×10^{-8}	2.192
AMLE	2.15×10^{-4}	6.78×10^{-9}	2.249	7.09×10^{-15}	4.032

Table 3: *Failure data for pneumatic tire test*

Time	Code	Time	Code	Time	Code	Time	Code	Time	Code	Time	Code	Time	Code
6	0	30	0	47	5	72	1	74	3	81	1	84	3
84	3	84	3	90	3	96	1	101	6	105	5	105	3
106	4	107	1	111	6	111	4	111	4	118	3	118	4
119	3	120	4	126	5	131	1	132	1	133	6	133	3
135	4	135	3	136	6	137	3	142	5	144	3	148	3
153	3	155	1	157	6	158	5	159	0	162	1	165	4
172	5	177	2	179	3	181	4	188	1	188	6	191	6
193	3	195	4	197	5	198	6	200	0	200	2	201	4
203	4	204	3	204	6	205	4	205	6	206	4	207	4
207	4	207	1	207	4	208	1	208	6	208	4	208	3
209	1	209	4	210	6	210	6	210	4	210	6	211	4
212	4	213	4	214	4	215	4	215	4	215	4	215	3
216	4	217	4	220	5	222	4	222	4	224	4	224	6
225	4	225	5	226	4	227	4	227	2	228	4	229	4
229	6	230	5	230	1	230	2	231	4	232	1	232	2
233	1	233	4	233	4	234	4	234	4	236	4	237	4
239	6	241	4	241	4	243	4	244	1	244	3	246	6
246	4	249	4	250	4	252	4	253	4	255	4	258	4
259	0	262	5	265	4	266	1	268	2	269	4	270	5
270	2	271	4	271	4	281	4	281	3	285	1	285	4
286	4	286	4	295	1	297	2	299	3	300	0	300	0
300	4	300	4	300	0	300	0	300	0	300	0	300	0
300	0	300	0	300	0	300	0	300	0	300	0	300	0
300	0	306	4	306	4	314	4	318	6	320	4	332	4
335	0	342	6	347	4								

Table 4: *Estimates of parameters in the model* (22) *for failure data of pneumatic tires*

Failure Code	1	2	3
Parameters	$(\theta_1^{(1)}, \theta_2^{(1)})$	$(\theta_1^{(2)}, \theta_2^{(2)})$	$(\theta_1^{(3)}, \theta_2^{(3)})$
MLE $\times 10^{-3}$	$(0.295, 1.670)$	$(0.066, 1.002)$	$(0.525, 1.002)$
AMLE $\times 10^{-3}$	$(0.296, 1.567)$	$(0.066, 0.944)$	$(0.525, 0.941)$
Failure Code	4	5	6
Parameters	$(\theta_1^{(4)}, \theta_2^{(4)})$	$(\theta_1^{(5)}, \theta_2^{(5)})$	$(\theta_1^{(6)}, \theta_2^{(6)})$
MLE $\times 10^{-3}$	$(0.295, 10.02)$	$(0.230, 0.835)$	$(0.262, 2.004)$
AMLE $\times 10^{-3}$	$(0.296, 11.20)$	$(0.230, 0.785)$	$(0.263, 2.617)$

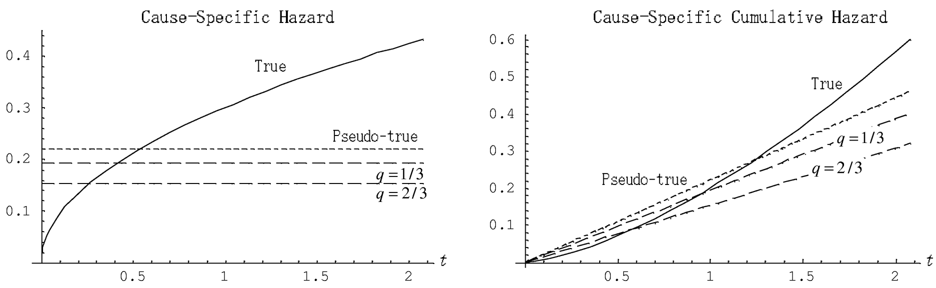

Figure 1: Limits of MLE under the condition of misspecification and censorship (Example 2). The solid curve in the left figure is the true cause-specific hazard of cause 1, $\lambda_0^{(1)}(t) = 0.2 \times 1.5\, t^{1.5-1}$, and the solid curve in the right figure is a corresponding cumulative hazard, $\Lambda_0^{(1)}(t) = 0.2 \times t^{1.5}$. The assumed parametric model is $\mathcal{M}_{const}^{(1)}$ of (4). In each figure, the dotted line is the pseudo-true hazard in $\mathcal{M}_{const}^{(1)}$. The dashed lines are limits of the MLE under the conditions of censoring proportion $q = 1/3$ and $2/3$.

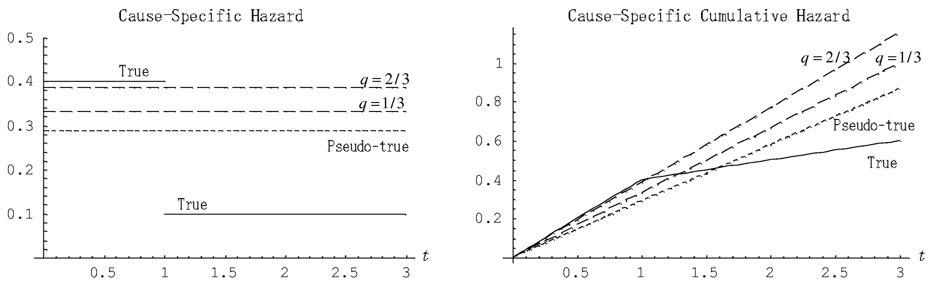

Figure 2: Limits of MLE under the condition of misspecification and censorship (Example 3). The solid line in the left figure is the true cause-specific hazard of cause 1, $\lambda_0^{(1)}(t) = 0.4I(t \leq 1) + 0.1I(t > 1)$, and the solid line in the right figure is a corresponding cumulative hazard. The assumed parametric model is $\mathcal{M}_{const}^{(1)}$ of (4). In each figure, the dotted line is the pseudo-true hazard in $\mathcal{M}_{const}^{(1)}$. The dashed lines are limits of the MLE as $n \to \infty$ under the conditions of censoring proportion $q = 1/3$ and $2/3$.

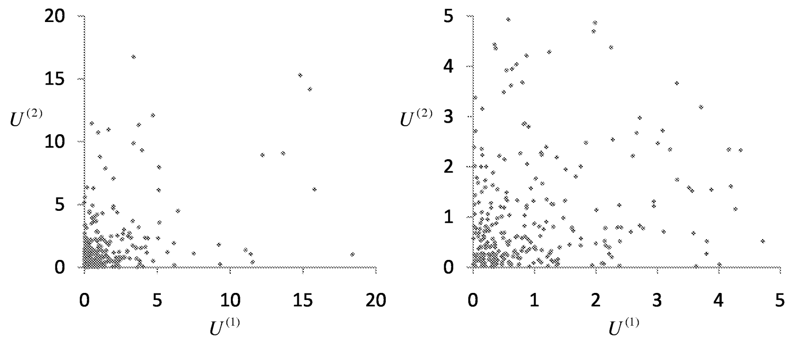

Figure 3: Scatter plots of simulated samples ($n = 300$) from the bivariate survival function (16) with $\rho = 1/2$. The left figure shows plots of all samples. The right figure is an expansion of the origin.

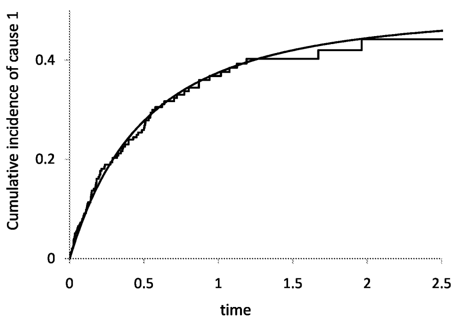

Figure 4: The cumulative incidence functions of cause 1 for simulated data. The curve is the true cumulative incidence function (18) with $\rho = 1/2$. The step-function is its Aalen-Johansen estimate based on the simulated samples.

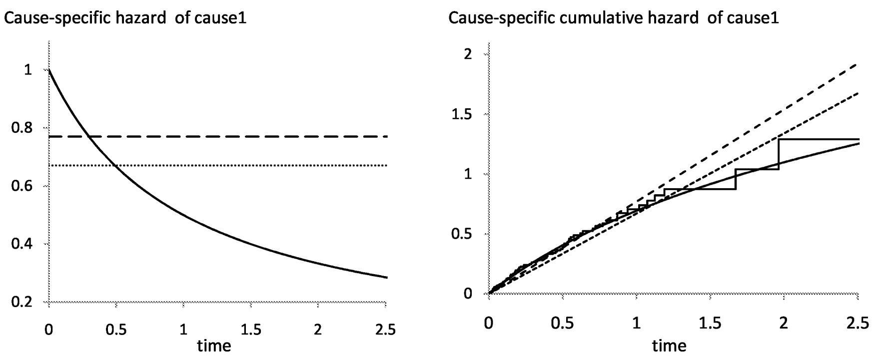

Figure 5: The cause-specific hazard functions of cause 1 (left) and the corresponding cumulative hazard functions (right) for the simulated data. In both figures, the solid curves are true. The dashed and dotted lines are estimates by MLE and AMLE, respectively. In the right figure, the solid step-function is the nonparametric Nelson-Aalen estimate.

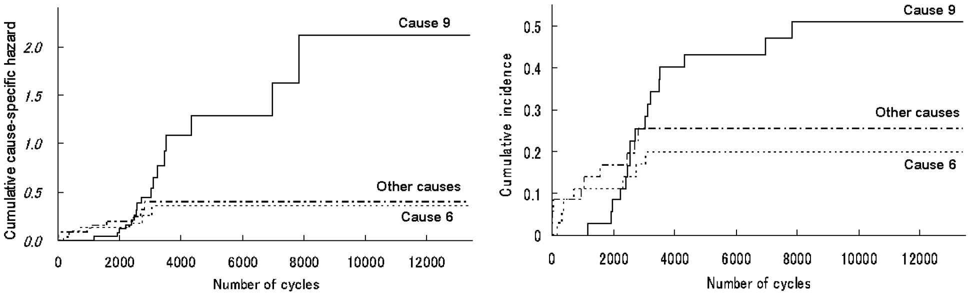

Figure 6: Nelson-Aalen estimates of cumulative cause-specific hazards and Aalen-Johansen estimates of cumulative incidences for electrical appliance tests.

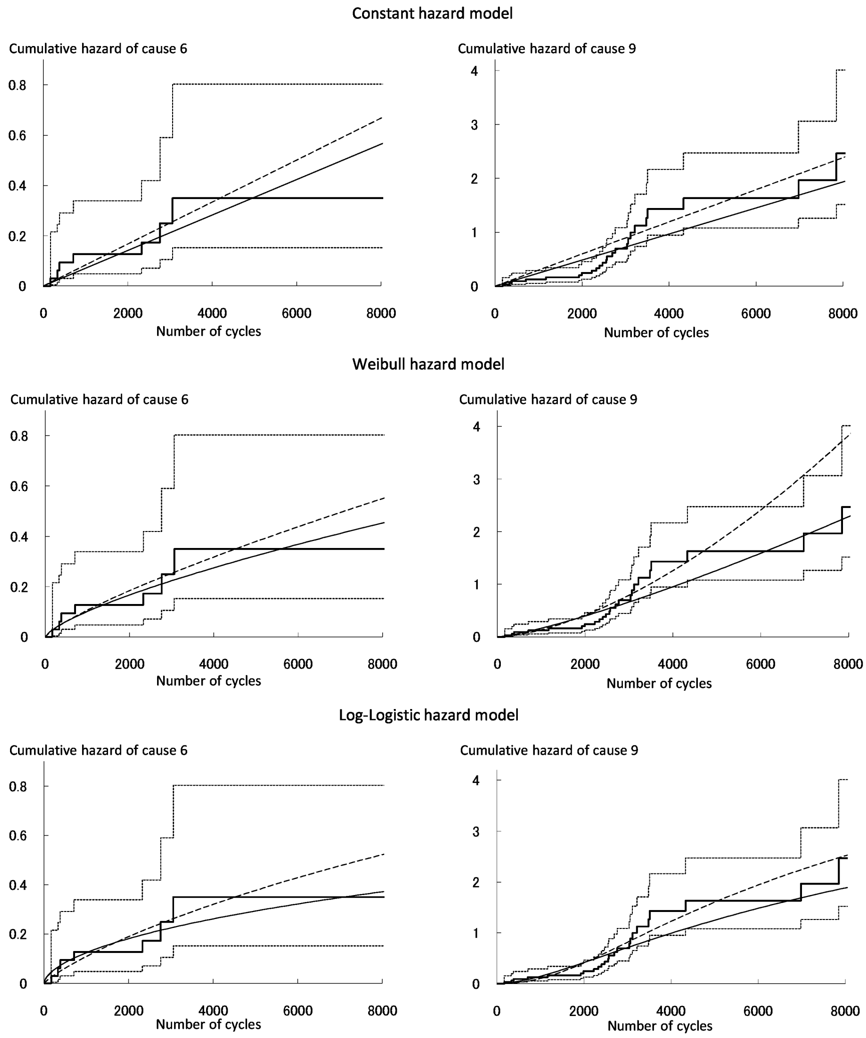

Figure 7: Estimates of cumulative cause-specific hazards of causes 6 and 9 (appliance test data). The figures on the left and right show estimates for causes 6 and 9, respectively. In each figure, the thick step-function is the Nelson-Aalen estimate and the dotted step-functions are approximate pointwise 90% confidence limits based on log-transformation. Assumed parametric models are (19), (20) and (21). In each figure, the solid and dashed curves show estimates obtained by the MLE and AMLE, respectively.

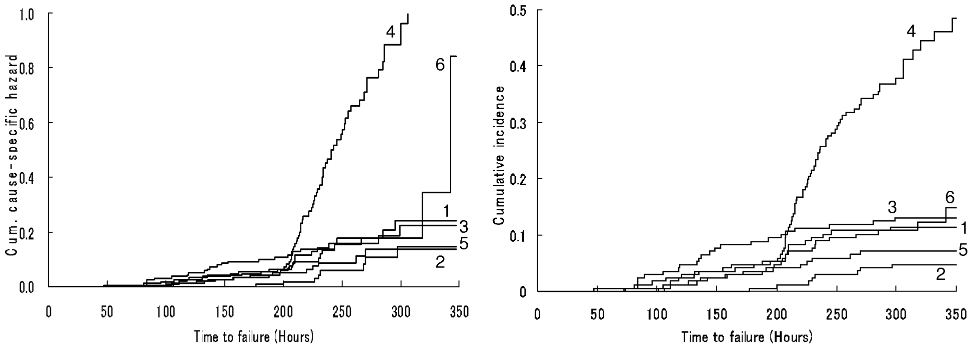

Figure 8: Nelson-Aalen estimates of cumulative cause-specific hazards and Aalen-Johansen estimates of cumulative incidences for failure data of pneumatic tires.

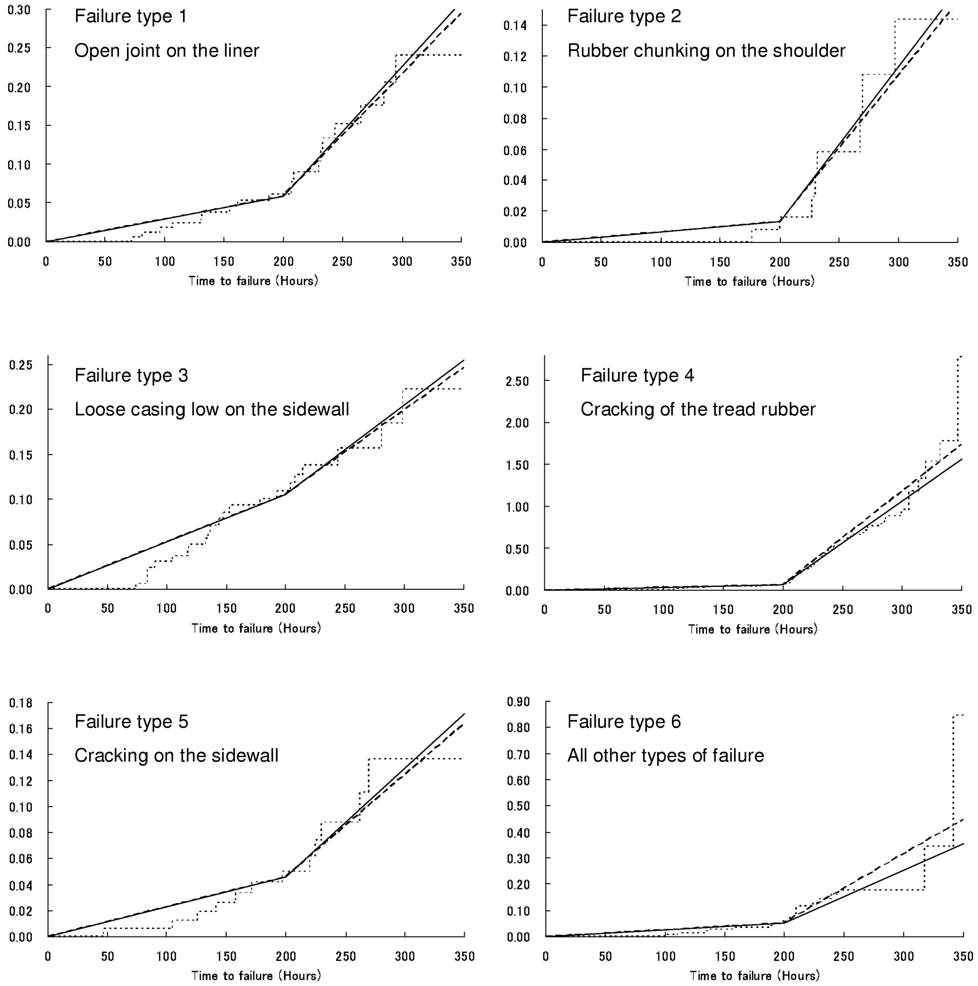

Figure 9: Estimates of cumulative cause-specific hazards for failure of pneumatic tires. The dotted step-functions are Nelson-Aalen estimates. The solid and dashed lines are estimates obtained by the MLE and AMLE in the piecewise-constant model (22), respectively.