ASYMPTOTIC DISTRIBUTIONS OF SEMIPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATORS WITH ESTIMATING EQUATIONS FOR GROUP-CENSORED DATA

DI CHEN[1], JYE-CHYI LU[2*] AND SHU CHUAN LIN[2]

1: Department of Statistics, North Carolina State University,
Raleigh, NC 27695-8203, U.S.A.

2: School of Industrial and Systems Engineering, Georgia Institute of Technology,
Atlanta, GA 30332-0205, U.S.A.

*: Author to whom correspondence should be addressed.

Summary

Semiparametric maximum likelihood estimation with estimating equations (SMLE) introduced in Qin & Lawless (1994) is more flexible than the traditional methods such as the parametric maximum likelihood estimation (MLE), Cox's (1972) proportional hazards model, accelerated failure time model, quasi-likelihood (Wedderburn, 1974) and generalized estimating equations (Liang & Zeger, 1986) with much less restrictions on distributions and regression-models. The needed information about distribution and regression structures is incorporated in estimating equations of the SMLE to improve the estimation quality of nonparametric methods. The likelihood of the SMLE in censored data cases involve several complicated implicit functions without closed-form expressions, and the first derivatives of the log-profile-likelihood cannot be expressed as summations of independent and identically distributed random variables. It is quite challenging to derive asymptotic properties of the SMLE in censored data cases. For group-censored data, the authors verify that all the implicit functions are well defined and obtain the asymptotic distributions of the SMLE for model parameters and lifetime distributions. This article uses several examples to compare the SMLE, the regular nonparametric likelihood estimation method and the parametric MLEs in terms of their asymptotic efficiencies. The article also present a real life example to illstrate the application of SMLE method. Various asymptotic distributions of the likelihood ratio statistics are derived for testing the adequacy of estimating equations and a partial set of parameters being equal to some known values.

Key words: Asymptotics; Generalized Estimating Equations; Likelihood Ratio Test; Maximum Likelihood Estimation.

AMS Subject Classification: Primary 62F10; secondary 62J02.

1. Introduction

Qin & Lawless (1994) and Lu, Chen & Gan (2002) used many examples to show that semiparametric maximum likelihood estimation with estimating equations (SMLE) method is very flexible in many studies including estimating failure time distributions and regression parameters. For example, the distributions of the failure times in testing printed circuit boards at different humidity levels studied in Meeker & LuValle (1995) are not the same after scale changes as suggested in the regular accelerated failure time models (AFT; Nelson, 1990, Chapter 4; Meeker & Escobar, 1998, Chapter 18). Because the SMLE method is not restricted to the distribution structures assumed in other semiparametric methods, e.g., the AFT and proportional hazards (PH; Cox, 1972) models, it should perform much better compared to the parametric MLE when the model is mis-specified.

Similar to MLE procedure, the SMLE has many nice statistical properties. To discuss the properties of SMLE, asymptotic theory plays an important role in deriving the large sample approximation of the distribution of the MLE and likelihood ratio test (LRT) statistics. In the complete sample case, the asymptotic theory of the SMLE is similar to the theory for the maximum empirical likelihood estimate (MELE). Suppose $G(t, \boldsymbol{\theta})$ is a vector of r estimating functions, and $\boldsymbol{\theta}$ is a p-dimensional parameter vector. Owen (1988, 1990) studied asymptotic properties of the MELE in $r = p$ situation. Qin & Lawless (1994, 1995) discussed more general cases with $r \geq p$. Because censored data usually occurred in real-life applications, extension of the results given in the complete sample to censored data is necessary.

The main purpose of this article is to provide technical details of SMLE's asymptotic distributional results focusing on group-censored data. Section 5 uses several numerical examples to show the efficiency and the application of the proposed SMLE methods. In general, when compared to the parametric MLE with a "correct" distribution used, the SMLE does not lose much efficiency asymptotically. On the other hand, when the parametric MLE mis-specifies the underlying distribution or does not use the information of distribution structure, the SMLE can be much more efficient than the mis-specified parametric MLEs or regular nonparametric MLE method, i.e., unconstrained SMLE. See Tables 1 and 2 for numerical values of the comparison.

Because the likelihood for the censored data in the SMLE approach has no closed expression and is determined by a set of complicate implicit functions, derivation of its asymptotic distribution is much more difficult than complete sample cases. Moreover, the log-profile-likelihood cannot be expressed as a sum of independent and identically distributed (i.i.d.) random variables. Therefore, it is quite challenging to derive the asymptotic properties of the SMLE for the censored data case.

This article obtains nice results for the grouped data with random censorship. The rest of the article is arranged as bellow. Section 2 briefly introduces the SMLE procedure. Section 3 presents the consistency of the SMLE estimators. Section 4 derives asymptotic distributions of the SMLE for various quantities. Section 5 shows some numerical results of asymptotic efficiencies of the SMLE and a real life example. Section 6 gives asymptotic distributions of the LRT statistics for simple hypothesis $H_{10} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, testing partial components of the parameters and model hypothesis $H_{20} : E(G(T, \boldsymbol{\theta})) = 0$. A skeleton of the proofs is presented in the Appendix. Appendix A is a proof of Lemma 3.1, and Appendix B shows the differentiability of some implicit functions used in SMLE procedures. The proofs of a few lemmas and the asymptotic normality of the likelihood are presented in Appendix C and D.

2. Semiparametric maximum likelihood estimation with estimating equations (SMLE)

2.1 Nonparametric likelihood for group-censored data

Let T_i and W_i be the lifetime and censoring time for the ith subject, $i = 1, 2, \ldots, n$, respectively. Assume that lifetime T and censoring time W are independent. With the random censorship, for the ith subject, one observes only $Y_i = \min(T_i, W_i)$ and $\Delta_i = I(T_i \le W_i)$, where $I(.)$ is the indicator function.

Let d_{ni} and c_{ni} be the numbers of deaths (lifetime data) and withdraws (censored data) observed at pre-fixed k time-points $0 < \xi_1 < \xi_2 < \ldots < \xi_k$, respectively, i.e., $d_{ni} = \sum_{j=1}^{n} I(Y_j = \xi_i,\ \Delta_j = 1)$, $c_{ni} = \sum_{j=1}^{n} I(Y_j = \xi_i,\ \Delta_j = 0)$, $i = 1, 2, \ldots, k$, where $0 < \xi_1 < \xi_2 < \ldots < \xi_k$ are known. In this article, denote $d_{nk+1} = n - \sum_{j=1}^{k}(d_{nj} + c_{nj})$ the number of the subjects whose lifetimes are beyond ξ_k.

This article considers T and W as discrete random variables with the following probability mass functions, respectively, $\Pr(T = \xi_i) = P_i > 0$, and $\Pr(W = \xi_i) = Q_i > 0$, $i = 1, 2, \ldots, k+1$, where $\xi_{k+1} \equiv L$, the experiment ending time ($L > \xi_k$), and $\sum_{i=1}^{k+1} P_i = 1$ and $\sum_{i=1}^{k+1} Q_i = 1$. With these definitions and the independence between T and W, we have

$$\begin{aligned} \Pr(Y = \xi_i, \Delta = 1) &= \Pr(T = \xi_i, W \ge \xi_i) = P_i \sum_{m=i}^{k+1} Q_m, i = 1, 2, \ldots, k+1, \\ \Pr(Y = \xi_i, \Delta = 0) &= \Pr(W = \xi_i, T > \xi_i) = Q_i \sum_{m=i+1}^{k+1} P_m, i = 1, 2, \ldots, k. \end{aligned} \tag{1}$$

Thus, the nonparametric likelihood of the group-censored data d_{ni}'s and c_{ni}'s is

$$L_n = \prod_{i=1}^{k+1} \left(P_i \sum_{m=i+1}^{k+1} Q_m \right)^{d_{ni}} \cdot \prod_{i=1}^{k} \left(Q_i \sum_{m=i+1}^{k+1} P_m \right)^{c_{ni}}.$$

Treating the censoring time probabilities Q_i, $i = 1, 2, \ldots, k+1$, as nuisance parameters. Then the nonparametric likelihood is simplified to

$$L_n = \prod_{i=1}^{k+1} P_i^{d_{ni}} \cdot \prod_{i=1}^{k} \left(\sum_{m=i+1}^{k+1} P_m \right)^{c_{ni}}. \tag{2}$$

Although the data is in groups, random censoring is involved through c_{ni} counts.

The likelihood (2) is derived for one-sample group-censored data. Its extension to independent multiple samples with covariates is straight forward and thus skipped here. See Lu, Chen & Gan (2002) for details.

2.2 The semiparametric model given by estimating equations

Let $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \boldsymbol{R}^p$ be the p-dimensional parameter vector associated with the discrete distribution $\boldsymbol{P} = (P_1, P_2, \ldots, P_{k+1})$, where $\boldsymbol{\Theta}$ is the parameter space which contains a neighborhood of true parameter $\boldsymbol{\theta}_0$. Suppose that the information about $\boldsymbol{\theta}$ and $\boldsymbol{P}$ is available in terms of $r \geq p$ functionally independent unbiased estimating equations $\mathrm{E}_P(G(T, \boldsymbol{\theta})) = 0$, where $G(T, \boldsymbol{\theta})$ is from $(0, \infty) \times \boldsymbol{\Theta}$ into $\boldsymbol{R}^r$. The parameter of interest, $\boldsymbol{\theta}$, and the distribution $\boldsymbol{P}$ are associated with each other by the estimating equations $\mathrm{E}_P(G(T, \boldsymbol{\theta})) = 0$. In other words, distribution $\boldsymbol{P} = (P_1, P_2, \ldots, P_{k+1})$, can be considered as the implicit functions of $\boldsymbol{\theta}$ that are described by equation $\sum_{i=1}^{k+1} P_i G(T, \xi_i, \boldsymbol{\theta}) = 0$.

For given $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, denoted by $\mathcal{B}_{\boldsymbol{\theta}}$ the set of all vectors of $\boldsymbol{b} = (p_1, p_2, \ldots, p_{k+1})$ with non-negative entries such that $\sum_{i=1}^{k+1} p_i = 1$ and $\sum_{i=1}^{k+1} p_i G(T, \xi_i, \boldsymbol{\theta}) = 0$. SMLE $\hat{\boldsymbol{\theta}}_n \in \boldsymbol{\Theta}$ of $\boldsymbol{\theta}$ searches $\boldsymbol{b}$ in $\mathcal{B}_{\boldsymbol{\theta}}$ to maximize the following function. Let

$$\begin{aligned} H_n &= \frac{1}{n} \log L_n - \lambda_0 (\sum_{i=1}^{k+1} P_i - 1) - \boldsymbol{\lambda}^\top \sum_{i=1}^{k+1} P_i G(T, \xi_i, \boldsymbol{\theta}) \\ &= \frac{1}{n} \sum_{i=1}^{k+1} d_{ni} \log P_i + \frac{1}{n} \sum_{i=1}^{k} c_{ni} \log (\sum_{j=i+1}^{k+1} P_j) - \lambda_0 (\sum_{i=1}^{k+1} P_i - 1) - \boldsymbol{\lambda}^\top \sum_{i=1}^{k+1} P_i G(T, \xi_i, \boldsymbol{\theta}), \end{aligned} \tag{3}$$

where λ_0 and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ are Lagrange multiplier quantities.

Taking derivatives with respect to P_i, λ_0, $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$, we have

$$\frac{\partial H_n}{\partial P_i} = \frac{d_{ni}}{n P_i} + \frac{1}{n} \sum_{m=1}^{i-1} \frac{c_{nm}}{(\sum_{j=m+1}^{k+1} P_j)} - \lambda_0 - \boldsymbol{\lambda}^\top G(T, \xi_i, \boldsymbol{\theta}), \; i = 1, 2, \ldots, k+1,$$

$$\frac{\partial H_n}{\partial \lambda_0} = 1 - \sum_{i=1}^{k+1} P_i, \quad \frac{\partial H_n}{\partial \boldsymbol{\lambda}} = - \sum_{i=1}^{k+1} P_i G(T, \xi_i, \boldsymbol{\theta})$$

and

$$\frac{\partial H_n}{\partial \boldsymbol{\theta}} = -\boldsymbol{\lambda}^\top \sum_{i=1}^{k+1} P_i \frac{\partial G(T, \xi_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

Equating the above derivatives to zeros, one can obtain $\hat{\boldsymbol{\theta}}_n$, the SMLE of $\boldsymbol{\theta}$, by solving the following equations:

$$\boldsymbol{\lambda}_n^\top(\boldsymbol{\theta}) \sum_{i=1}^{k+1} P_{ni}(\boldsymbol{\theta}) \frac{\partial G(T, \xi_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0, \tag{4}$$

where

$$P_{ni}(\boldsymbol{\theta}) = \frac{d_{ni}}{n(1 - a_{ni}(\boldsymbol{\theta}) + \boldsymbol{\lambda}_n^\top(\boldsymbol{\theta}) G(T, \xi_i, \boldsymbol{\theta}))}, \quad i = 1, 2, \cdots, k, \quad P_{n(k+1)} = 1 - \sum_{i=1}^{k} P_{ni}, \tag{5}$$

in which

$$a_{ni}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{m=1}^{i-1} \frac{c_{nm}}{\sum_{j=m+1}^{k+1} P_{nj}(\boldsymbol{\theta})}, \; i = 2, \cdots, k, \quad a_{n1} = 0, \tag{6}$$

and $\boldsymbol{\lambda}_n(\boldsymbol{\theta})$ is given by

$$\sum_{i=1}^{k+1} \frac{d_{ni} G(T, \xi_i, \boldsymbol{\theta})}{1 - a_{ni}(\boldsymbol{\theta}) + \boldsymbol{\lambda}_n^\top(\boldsymbol{\theta}) G(T, \xi_i, \boldsymbol{\theta})} = 0. \tag{7}$$

Note that $\lambda_0 = 1$ by using the fact that $\sum_{i=1}^{k+1} P_i \partial H_n / \partial P_i = 0$. In the case that no censoring is observed, $a_{ni} = 0$ and P_{ni} reduces to the results as obtained by Qin & Lawless (1994).

Because equation (4) is defined by a series of implicit functions without closed-form expressions, it is difficult to study the properties of SMLE $\hat{\boldsymbol{\theta}}_n$. In particular, for this group-censored data case, because $\boldsymbol{\lambda}_n(\boldsymbol{\theta})$ determined by (7) depends on $\boldsymbol{\theta}$, $P_{ni}(\boldsymbol{\theta})$ and $a_{ni}(\boldsymbol{\theta})$, the left side of (7) is no longer a sum of i.i.d. random variables like that in completed sample case. Thus, the procedures used in Owen (1988, 1990) and Qin & Lawless (1994, 1995) are not applicable. Furthermore, in the survival analysis literature, the asymptotic distribution of $\hat{F}_T(t)$ has been derived only in some special cases (e.g., Breslow & Crowley (1974) and Tsiatis (1981)). The SMLE of survival functions,

$$\hat{F}_T(t) = \sum_{\xi_i \le t} \frac{d_{ni}}{n(1 - a_{ni}(\boldsymbol{\theta}_n) + \boldsymbol{\lambda}_n^\top(\hat{\boldsymbol{\theta}}_n) G(T, \xi_i, \hat{\boldsymbol{\theta}}_n))},$$

is very complicated. To derive the asymptotic distribution of $\hat{F}_T(t)$, the convergence of $a_{ni}(\hat{\boldsymbol{\theta}}_n)$ and $\boldsymbol{\lambda}_n(\hat{\boldsymbol{\theta}}_n)$ as well as their convergence rates are needed. Because these quantities involve many implicit functions, investigating their asymptotic properties is quite challenging.

Before discussing the asymptotic properties of the SMLE, we need to show that the equations (5) - (7) are well defined, i.e., their solutions exist uniquely. Under some mild regularity conditions, the next section presents the existence of these solutions and the consistency of the SMLE $\hat{\boldsymbol{\theta}}_n$.

3. Consistency of the SMLE $\hat{\boldsymbol{\theta}}$

The following regularity conditions are necessary in discussing the properties of the SMLE $\hat{\boldsymbol{\theta}}_n$.

(**R.1**) Parameter space $\boldsymbol{\Theta} \subset R^p$ is compact without isolated points and contains a neighborhood of true parameter $\boldsymbol{\theta}_0$, and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\{0 < |\max_{1\leq i\leq k+1} G(\xi_i, \boldsymbol{\theta})|\} < \infty$,.

(**R.2**) Denoted by $G(t, \boldsymbol{\theta}) = (g_1(t, \boldsymbol{\theta}), g_2(t, \boldsymbol{\theta}), \ldots, g_r(t, \boldsymbol{\theta}))$. For given $0 < \xi_1 < \xi_2 \ldots < \xi_k < \xi_{k+1} = L$, let $\boldsymbol{g}_j = (g_j(\xi_1, \boldsymbol{\theta}), g_j(\xi_2, \boldsymbol{\theta}), \ldots, g_j(\xi_{k+1}, \boldsymbol{\theta}))$, $j = 1, 2, \ldots, r$, and $\boldsymbol{G} = (\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_r)$. For every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, assume that $r \times r$ matrix $\boldsymbol{G}^\top \boldsymbol{G}$ is nonsingular,

(**R.3**) $\mathrm{E}(||G(T, \boldsymbol{\theta})||^3) < \infty$ and $G(T, \boldsymbol{\theta})$ is second-order differentiable with respect to $\boldsymbol{\theta}$, i.e., $\partial^2 G(T, \boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ exists for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

The regularity condition **R1** is to ensure that the maximum of the likelihood function $|H_n(\boldsymbol{\theta})|$ defined in (3) exists in the interner of $\boldsymbol{\Theta}$. **R2** and **R3** state the non-singularity, continuity and differentiability of estimating function $G(t, \boldsymbol{\theta})$ to ensure that equation (4) is well defined and the SMLE $\hat{\boldsymbol{\theta}}_n$ will be found in the ball $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < n^{-1/3}$ with probability 1, when n is sufficiently large. The following result shows that equation (4) is well defined. See Appendix A for its proof.

Lemma 3.1 *Under regularity condition* **R2**, *for given* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, *solutions of the implicit functions* $\boldsymbol{\lambda}(\boldsymbol{\theta})$ *(5) - (7) exist uniquely.*

It follows from Lemma 3.1 that $P_i(\boldsymbol{\theta})$, $i = 1, 2, \ldots, k+1$ given by (5) are well defined. Therefore, the log-profile-likelihood of $\boldsymbol{\theta}$ is

$$H_n(\boldsymbol{\theta}) = \sum_{i=1}^{k+1} d_{ni} \log P_{ni}(\boldsymbol{\theta}) + \sum_{i=1}^{k} c_{ni} \log\left(\sum_{m=i+1}^{k+1} P_{nm}(\boldsymbol{\theta})\right). \tag{8}$$

Under the regularity conditions **R2**, **R3**, Appendix B shows that $P_{ni}(\boldsymbol{\theta})$, $i = 1, 2, \ldots, k+1$, and $\boldsymbol{\lambda}_n(\boldsymbol{\theta})$ are differentiable with respect to $\boldsymbol{\theta}$. Thus, SMLE $\hat{\boldsymbol{\theta}}_n$ can be obtained by solving equations $\partial H_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0$. From (8), we know that it is equivalent to solving the following equation

$$\sum_{i=1}^{k+1} \frac{d_{ni}}{P_{ni}(\boldsymbol{\theta})} \frac{\partial P_{ni}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \sum_{i=1}^{k}\left(\frac{c_{ni}}{\sum_{m=i+1}^{k+1} P_{nm}(\boldsymbol{\theta})} \sum_{j=i+1}^{k+1} \frac{\partial P_{nj}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right) = 0.$$

For the notation simplicity, the "dot" notation will be used to define the derivative of quantities with respect to $\boldsymbol{\theta}$, e.g., $\dot{P}_{ni}(\boldsymbol{\theta}) = \partial P_{ni}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. Because

$$\begin{aligned}
&\sum_{i=1}^{k+1} \frac{d_{ni}}{P_{ni}(\boldsymbol{\theta})} \dot{P}_{ni}(\boldsymbol{\theta}) + \sum_{m=1}^{k} \frac{c_{nm}}{\sum_{j=m+1}^{k+1} P_{nj}(\boldsymbol{\theta})} \sum_{i=m+1}^{k+1} \dot{P}_{ni}(\boldsymbol{\theta}) \\
=\ & \sum_{i=1}^{k+1}\left(n(1 - a_{ni}(\boldsymbol{\theta}) + \boldsymbol{\lambda}_n^\top G(T, \xi_i, \boldsymbol{\theta})) + \sum_{m=1}^{i-1} \frac{c_{nm}}{\sum_{j=m+1}^{k+1} P_{nj}(\boldsymbol{\theta})}\right) \dot{P}_{ni}(\boldsymbol{\theta}),
\end{aligned} \tag{9}$$

and recall that

$$a_{ni} = \sum_{m=1}^{i-1} \frac{c_{nm}}{\sum_{j=m+1}^{k+1} P_{nj}(\boldsymbol{\theta})}, \quad \sum_{i=1}^{k+1} \dot{P}_{nj}(\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}}\left(\sum_{i=1}^{k+1} P_{nj}(\boldsymbol{\theta})\right) = 0,$$

and $\sum_{i=1}^{k+1} G(T,\xi_i,\boldsymbol{\theta})\dot{P}_{ni}(\boldsymbol{\theta}) = -\sum_{i=1}^{k+1} \dot{G}(T,\xi_i,\boldsymbol{\theta})P_{ni}(\boldsymbol{\theta})$, it follows that the right hand side of (9) is $-n\boldsymbol{\lambda}_n^\top(\boldsymbol{\theta})\sum_{i=1}^{k} \dot{G}(T,\xi_i,\boldsymbol{\theta})P_{ni}(\boldsymbol{\theta})$. That is,

$$\frac{\partial H_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -n\boldsymbol{\lambda}_n^\top(\boldsymbol{\theta})\sum_{i=1}^{k} \dot{G}(T,\xi_i,\boldsymbol{\theta})P_{ni}(\boldsymbol{\theta}). \tag{10}$$

Therefore, equation $\partial H_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0$ is the same as equation (4). Similar to the results of Qin & Lawless (1994, Lemma 1) in the completed sample case, we have the following conclusions for the group-censored data. The proof of Theorem 3.1 is tedious but straightforward by following the similar procedures used by Qin & Lawless (1994) and Owen (1990), and thus, it is omitted.

Theorem 3.1 *Let $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ be the true value of the parameter. Under the regularity conditions* **R1-R3**, *we have*

(1) *$\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\lambda}}_n = \boldsymbol{\lambda}_n(\hat{\boldsymbol{\theta}})$ satisfy equations (4) and (7).*

(2) *$\boldsymbol{\lambda}_n(\boldsymbol{\theta}) \xrightarrow{w.p.1} 0$ and $\boldsymbol{\lambda}_n(\boldsymbol{\theta}) = O_p(n^{-1/2})$, for $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le n^{-1/3}\}$.*

(3) *When sample size n is large, with probability 1, $H_n(\boldsymbol{\theta})$ attains its maximum value at some point $\hat{\boldsymbol{\theta}}_n$ in the interior of the ball $||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le n^{-1/3}$. Thus, the SMLE $\hat{\boldsymbol{\theta}}_n$ is a strongly consistent estimate of $\boldsymbol{\theta}$.*

4. Asymptotic distribution of the SMLE

4.1 Asymptotic normality of $\partial H_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$

Before deriving of the asymptotic normality of $\partial H_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, we need to study the convergence of $\hat{a}_{ni}$ and the consistency of $\hat{P}_{ni}$ first. Replacing the parameter $\boldsymbol{\theta}$ in equations (5) and (6) by its SMLE $\hat{\boldsymbol{\theta}}_n$ leads to $\hat{P}_{ni} = n^{-1}d_{ni}/(1-a_{ni}+\boldsymbol{\lambda}_n^\top G(T,\xi_i,\hat{\boldsymbol{\theta}}_n))$, $i=1,2,\ldots,k$, and $\hat{P}_{n(k+1)} = 1-\sum_{i=1}^{k}\hat{P}_i$, where $a_{n1}=0$, $a_{ni} = n^{-1}\sum_{m=1}^{i-1} c_{nm}/(\sum_{j=m+1}^{k+1}\hat{P}_{nj})$, $i=2,\ldots,k$.

The following lemmas are needed to derive the asymptotic results given in Theorem 4.1. See Appendix C for their proofs.

Lemma 4.1 *With the above notations, we have that $a_{ni} \xrightarrow{w.p.1} \sum_{j=1}^{i-1} Q_j$, $i=2,3,\ldots,k$, and $\hat{P}_{ni} \xrightarrow{w.p.1} P_i$, $i=1,2,\ldots,k$.*

Lemma 4.2 *Under the conditions of Theorem 3.1, for given $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le n^{1/3}\}$,*

$$\boldsymbol{\lambda}_n(\boldsymbol{\theta}) = -\left[\sum_{i=1}^{k+1}\frac{d_{ni}G(T,\xi_i,\boldsymbol{\theta})G^\top(T,\xi_i,\boldsymbol{\theta})}{n(1-a_{ni})^2}\right]^{-1}\left[\sum_{i=1}^{k+1}\frac{d_{ni}G(T,\xi_i,\boldsymbol{\theta})}{n(1-a_{ni})}\right] + o_p(n^{-1/2}), \tag{11}$$

where $P_{ni}(\boldsymbol{\theta})$, a_{ni} and $\boldsymbol{\lambda}_n(\boldsymbol{\theta})$ are the solution of equations (5) - (7).

Let

$$\Sigma_{n1} = \sum_{i=1}^{k+1} \frac{d_{ni} G(T, \xi_i, \boldsymbol{\theta}_n) G^\top(T, \xi_i, \hat{\boldsymbol{\theta}}_n)}{n(1-a_{ni})^2}, \quad \text{and} \quad \Sigma_1 = \sum_{i=1}^{k+1} \frac{P_i G(T, \xi_i, \boldsymbol{\theta}) G^\top(T, \xi_i, \boldsymbol{\theta})}{\sum_{m=i}^{k+1} Q_m}. \tag{12}$$

According to Lemma 4.1, we know that

$$\Sigma_{n1} \quad \stackrel{w.p.1}{\longrightarrow} \quad \sum_{i=1}^{k+1} \frac{P_i \sum_{m=i}^{k+1} Q_m G(T, \xi_i, \boldsymbol{\theta}) G^\top(T, \xi_i, \boldsymbol{\theta})}{(\sum_{m=i}^{k+1} Q_m)^2} = \Sigma_1.$$

Let $\boldsymbol{Z}_n^* = n^{1/2} \sum_{i=1}^{k+1} d_{ni} G(T, \xi_i, \boldsymbol{\theta})/n(1-a_{ni})$. Appendix D shows that $\boldsymbol{Z}_n^* \stackrel{d}{\longrightarrow} N_r(0, \Sigma_1)$, where Σ_1 is given in (12). Then, Applying the results of $\boldsymbol{\lambda}_n(\boldsymbol{\theta})$ given in (11) to equation (10), we have

$$\frac{1}{\sqrt{n}} \frac{\partial H_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\boldsymbol{A}_n^\top \boldsymbol{B}_n^{-1} \boldsymbol{Z}_n^* + o_p(n^{-1/2}), \tag{13}$$

where $\boldsymbol{A}_n = \sum_{i=1}^{k+1} \dot{G}(T, \xi_i, \boldsymbol{\theta}) P_{ni}(\boldsymbol{\theta})$, $\boldsymbol{B}_n = n^{-1} \sum_{i=1}^{k+1} d_{ni} G(T, \xi_i, \boldsymbol{\theta}) G^\top(T, \xi_i, \boldsymbol{\theta})/(1-a_{ni})^2$. Under the regularity conditions **R.2** and **R.3**, we know that

$$\lim_{n\to\infty} \boldsymbol{A}_n = \sum_{i=1}^{k+1} \dot{G}(T, \xi_i, \boldsymbol{\theta}) P_i \equiv \Sigma_2 \tag{14}$$

exists with probability 1, and $\boldsymbol{B}_n = n^{-1} \sum_{i=1}^{k+1} d_{ni} G(T, \xi_i, \boldsymbol{\theta}) G^\top(\xi_i, \boldsymbol{\theta})/(1-a_{ni})^2 \stackrel{w.p.1}{\longrightarrow} \Sigma_1$. Then, it follows from (13) that $n^{-1/2} \partial H_n(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \stackrel{d}{\longrightarrow} N_r(0, \Sigma_q)$, where $\Sigma_q = \Sigma_2^\top \Sigma_1^{-1} \Sigma_1 \Sigma_1^{-1} \Sigma_2 = \Sigma_2^\top \Sigma_1^{-1} \Sigma_2$. We state the this conclusion formly as the next theorem.

Theorem 4.1 *Under the conditions of Theorem 3.1, for the derivatives of the log-profile-likelihood given in (10),* $n^{-1/2} \partial H_n(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \stackrel{d}{\longrightarrow} N_r(0, \Sigma_q)$, as $n \to \infty$, *where* $\Sigma_q = \Sigma_2^\top \Sigma_1^{-1} \Sigma_2$, *and* Σ_1 *and* Σ_2 *are as defined in (12) and (14), respectively.*

4.2 Asymptotic distribution of the SMLE $\hat{\boldsymbol{\theta}}_n$

Applying Taylor's expansion to the left hand side of (4) at $\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_0$, we obtain that

$$n^{-1/2} \boldsymbol{\lambda}_n^\top \sum_{i=1}^{k+1} \dot{G}(T, \xi_i, \boldsymbol{\theta}) P_{ni} = -n^{-1} \partial [\boldsymbol{\lambda}_n^\top \sum_{i=1}^{k+1} \dot{G}(T, \xi_i, \boldsymbol{\theta}) P_{ni}] / \partial \boldsymbol{\theta} \sqrt{n} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) + o_p(\sqrt{n} \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|),$$

where $\boldsymbol{\lambda}_n = \boldsymbol{\lambda}_n(\boldsymbol{\theta})$ and $P_{ni} = P_{ni}(\boldsymbol{\theta})$ are determined by (5) and (6) implicitly. Before deriving the asymptotic distribution of the SMLE $\hat{\boldsymbol{\theta}}_n$, we need to show that the following limit exists with probability 1, and the limiting matrix is nonsingular (rank $= r$),

$$\lim_{n\to\infty} \frac{1}{n} \frac{\partial}{\partial \boldsymbol{\theta}} \left(\boldsymbol{\lambda}_n^\top(\boldsymbol{\theta}) \sum_{i=1}^{k+1} \dot{G}(T, \xi_i, \boldsymbol{\theta}) P_{ni}(\boldsymbol{\theta}) \right) \equiv \Sigma_*. \tag{15}$$

Assume that the derivative and the limit are exchangeable. The above limit can be written as

$$\lim_{n\to\infty}\frac{1}{n}\left[\frac{\partial\boldsymbol{\lambda}_n^\top(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\sum_{i=1}^{k+1}\dot{G}(T,\xi_i,\boldsymbol{\theta})P_{ni}(\boldsymbol{\theta})+\boldsymbol{\lambda}_n^\top(\boldsymbol{\theta})\frac{\partial}{\partial\boldsymbol{\theta}}\left(\sum_{i=1}^{k+1}\dot{G}(T,\xi_i,\boldsymbol{\theta})P_{ni}(\boldsymbol{\theta})\right)\right]. \tag{16}$$

Because the derivative of $\partial\left(\sum_{i=1}^{k+1}\dot{G}(T,\xi_i,\boldsymbol{\theta})P_{ni}(\boldsymbol{\theta})\right)/\partial\boldsymbol{\theta}$ exists, and $\boldsymbol{\lambda}_n(\boldsymbol{\theta})\xrightarrow{w.p.1}0$, under the conditions **R.2, R.3**, it can be shown that the limit of the second term of equation (16) is zero with probability 1. Then, the limit in (15) is equal to $\lim_{n\to\infty}\partial\boldsymbol{\lambda}_n^\top(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\sum_{i=1}^{k+1}\dot{G}(T,\xi_i,\boldsymbol{\theta})P_{ni}(\boldsymbol{\theta})$.

Under the conditions of Theorem 3.1, it follows from the expression of $\partial\boldsymbol{\lambda}_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ (see Appendix B) that we have $\lim_{n\to\infty}\partial\boldsymbol{\lambda}_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}=\Sigma_1^{-1}\Sigma_2$ w.p.1. Thus, from (14), we know that $\Sigma_*=\lim_{n\to\infty}=\Sigma_2^\top\Sigma_1^{-1}\Sigma_2$. With this result, the limiting covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})$ becomes $\Sigma=[\Sigma_2^\top\Sigma_1^{-1}\Sigma_2]^{-1}$.

Therefore, it follows from the Taylor expansion and the results given in Lemma 3.3 and Theorem 3.1 that the asymptotic distribution of the SMLE $\hat{\boldsymbol{\theta}}_n$ can be derived. The result is stated in the next theorem.

Theorem 4.2 *Let $\boldsymbol{\theta}$ be the true parameter and $\hat{\boldsymbol{\theta}}_n$, the solution of the equation in Theorem 3.1, be the SMLE of $\boldsymbol{\theta}$. Under the regularity conditions* **R.2, R.3**, *we have $\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})\xrightarrow{d}N_p(0,\Sigma)$, where $\Sigma=[\Sigma_2^\top\Sigma_1^{-1}\Sigma_2]^{-1}$, Σ_1 and Σ_2 are given in (12) and (14), respectively .*

The following offer some interpretation of the covariance matrix Σ. Recall from (12) and (14) that Σ_1 and Σ_2 are the following expectations, $\Sigma_1=\mathrm{E}[G(T,\boldsymbol{\theta})G^\top(T,\boldsymbol{\theta})/S_W(T)]$, $\Sigma_2=\mathrm{E}[\dot{G}(T,\boldsymbol{\theta})]$, where $S_W(t)=\sum_{i:\xi_i\geq t}Q_i$. The denominator of Σ_1 represents the proportion of uncensored data in the observations. The more uncensored data are observed, the smaller asymptotic variance of the SMLE $\hat{\boldsymbol{\theta}}_n$ is. For example, in the completed sample case, this denominator becomes one, and the result of Theorem 4.2 is the same result as given in Qin & Lawless (1994). In the Type-I censoring case that the censoring time $w_1,w_2,\ldots,w_n$ are as fixed (not random variables). The denominator $S_W(T)=Pr(W>T)=F_T(W)$ can be replaced by $Q=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n F_T(w_i)$ (assume it exists). Then, Σ_1 becomes $\mathrm{E}(G(T,\boldsymbol{\theta})G^\top(T,\boldsymbol{\theta}))/Q$.

As a simple illustration, assume that T has an exponential distribution with the mean parameter θ. Based on the classical parametric MLE result for the exponential distribution (Lawless (1982), page 105-107) in the type-I censored data case, its asymptotic variance is $\boldsymbol{\theta}^2/Q^*$, where $Q^*=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n(1-e^{-w_i/\boldsymbol{\theta}})$. In this case, from Theorem 4.2 the SMLE with a single estimating equation, $G(T,\theta)=T-\theta$, has the limiting variance as

$$\sigma^2=\frac{\mathrm{E}(T-\theta)^2}{Q}=\frac{\mathrm{Var}(T)}{Q}=\theta^2/Q.$$

The quantity Q^* matches the quantity Q defined above, and the asymptotic variances are the same in the parametric MLE and SMLE approaches. See Lu, Chen & Gan (2002) for additional comparisons.

4.3 Asymptotic normal distributions of the SMLEs $\hat{P}_{ni}$ and $\hat{F}_n(t)$

Let us consider the asymptotic distribution of $\hat{P}_{ni} = P_{ni}(\hat{\boldsymbol{\theta}}_n)$ first. Because

$$\hat{P}_{ni} = \frac{d_{ni}}{(1 - a_{ni} + \boldsymbol{\lambda}_n^\top(\hat{\boldsymbol{\theta}}_n)G(T, \xi_i, \hat{\boldsymbol{\theta}}_n))} \text{ and } \|\boldsymbol{\lambda}_n\| = O_p(n^{-1/2}),$$

Taylor expansion of $\hat{P}_{ni}$ at $\boldsymbol{\lambda}_n = 0$ results in

$$\hat{P}_{ni} = \frac{d_{ni}}{n(1-a_{ni})} + \frac{d_{ni}G^\top(\xi_i, \hat{\boldsymbol{\theta}}_n)\boldsymbol{\lambda}_n(\hat{\boldsymbol{\theta}}_n)}{n(1-a_{ni})^2} + o_p(n^{-1/2}).$$

From (11) and (12),

$$\boldsymbol{\lambda}_n = \Sigma_{n1}^{-1}\left(\sum_{i=1}^{k+1} \frac{d_{ni}G(T, \xi_i, \hat{\boldsymbol{\theta}}_n)}{n(1-a_{ni})}\right) + o_p(n^{-1/2}).$$

Therefore, $\sqrt{n}(\hat{P}_{ni} - P_i) =$

$$\begin{aligned} & \sqrt{n}\left(\frac{d_{ni}}{n(1-a_{ni})} - P_i\right) + \sqrt{n}\left(\frac{d_{ni}G^\top(\xi_i, \hat{\boldsymbol{\theta}}_n)\boldsymbol{\lambda}_n(\hat{\boldsymbol{\theta}}_n)}{n(1-a_{ni})^2}\right) + o_p(1) \\ = \ & \sqrt{n}\left(\frac{d_{ni}}{n(1-a_{ni})} - P_i\right) + \frac{d_{ni}G^\top(\xi_i, \hat{\boldsymbol{\theta}}_n)}{n(1-a_{ni})^2}\Sigma_{n1}^{-1}\left(\sum_{i=1}^{k+1} \frac{d_{ni}G(T, \xi_i, \hat{\boldsymbol{\theta}}_n)}{n(1-a_{ni})}\right) + o_p(1). \end{aligned}$$

Denoted by $Z_{ni}^{(1)}$ and $Z_{ni}^{(2)}$ the first and the second terms, respectively, in the above expression. It is obvious that $Z_{ni}^{(1)}$ is asymptotic normal. Appendix D shows that $Z_{ni}^{(2)}$ is also asymptotic normal. Therefore, the asymptotic distribution of $\hat{P}_{ni}$ is obtained as stated in the next theorem with the asymptotic variance equal to the limit of the following quantity as $n \to \infty$:
$\sigma_{ni}^2 = \mathrm{Var}(Z_{ni}^{(1)}) + \mathrm{Var}(Z_{ni}^{(2)}) - 2\mathrm{Cov}(Z_{ni}^{(1)}, Z_{ni}^{(2)})$.

Theorem 4.3 *Under the conditions of Theorem 3.1,* $\sqrt{n}(\hat{P}_{ni} - P_i) \xrightarrow{d} N(0, \sigma_i^2)$, *where*

$$\sigma_i^2 = \frac{P_i}{\sum_{j=i}^{k+1} Q_j} - P_i{}^2 - \left(\frac{P_i}{\sum_{j=i}^{k+1} Q_j}\right)^2 G^\top(\xi_i, \boldsymbol{\theta})\Sigma_1^{-1}G(T, \xi_i, \boldsymbol{\theta}).$$

Next, let us discuss the asymptotic distribution of $\hat{F}_n(t) = \sum_{i:\xi_i \le t} P_{ni}(\hat{\boldsymbol{\theta}}_n) = \sum_{i=1}^{k+1} P_{ni}(\hat{\boldsymbol{\theta}}_n)I(\xi_i \le t)$. First, we show that $\sqrt{n}(\hat{F}_n(t) - F_T(t))$ can be written in terms of $\sum_{i=1}^{k+1} b_{ni}(d_{ni}/n - \phi_i) + o_p(1)$ for some quantities b_{ni}, where $F_T(t) = \sum_{i=1}^{k+1} P_i I(\xi_i \le t)$, $\phi_i = \mathrm{E}(d_{ni}/n) = P_i \sum_{j=i}^{k+1} Q_j$. Note that Taylor expansion of $\hat{F}_n(t) = \sum_{\xi_i \le t} P_{ni}(\hat{\boldsymbol{\theta}}_n)$ at the true parameters $\boldsymbol{\theta}$ yields

$$\begin{aligned} \hat{F}_n(t) \ = \ & \sum_{i=1}^{k+1} P_{ni}(\boldsymbol{\theta})I(\xi_i \le t) + \\ & \sum_{i=1}^{k+1} \frac{d_{ni}I(\xi_i \le t)}{n(1 - a_{ni} + \boldsymbol{\lambda}_n^\top(\boldsymbol{\theta})G(T, \xi_i, \boldsymbol{\theta}))}\left(G^\top(\xi_i, \boldsymbol{\theta})\frac{\partial \boldsymbol{\lambda}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) + o_p(n^{-1/2}). \end{aligned}$$

Moreover, recall that $\partial \boldsymbol{\lambda}_n/\partial \boldsymbol{\theta} \to \Sigma_1^{-1}\Sigma_2$, and, based on (13), we know that $(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = \Sigma_{n*}^{-1}\boldsymbol{A}_n^\top \boldsymbol{B}_n^{-1}\sum_{i=1}^{k+1} d_{ni}G(T,\xi_i,\boldsymbol{\theta}$ $a_{ni}) + o(n^{-1})$, where Σ_{n*} is the left hand side of (15). Because $\boldsymbol{A}_n = \Sigma_2 + o_p(1)$, $\boldsymbol{B}_n = \Sigma_1 + o_p(1)$, and

$\Sigma_{n*} = \Sigma_* + o_p(1) = \Sigma_2^\top\Sigma_1^{-1}\Sigma_2 + o_p(1)$, $\sqrt{n}(\hat{F}_n(t) - F_T(t)) =$

$$\sum_{i=1}^{k+1}\frac{I(\xi_i \le t)}{(1-a_{ni})}\left(\frac{d_{ni}}{n} - \phi_i\right) - \rho^\top(t)\sum_{i=1}^{k+1}\frac{G(T,\xi_i,\boldsymbol{\theta})}{(1-a_{ni})}\left(\frac{d_{ni}}{n} - \phi_i\right) + o_p(1)$$

$$= \sum_{i=1}^{k+1}\frac{I(\xi_i \le t) - \rho^\top(t)G(T,\xi_i,\boldsymbol{\theta})}{(1-a_{ni})}\left(\frac{d_{ni}}{n} - \phi_i\right) + o_p(1),$$

where

$$\rho^\top(t) = \beta^\top(t)\Sigma_1^{-1}\Sigma_2[\Sigma_2^\top\Sigma_1^{-1}\Sigma_2]^{-1}\Sigma_2^\top\Sigma_1^{-1}, \quad \text{and} \quad \beta(t) = \sum_{\xi_i \le t} G(T,\xi_i,\boldsymbol{\theta})P_i. \tag{17}$$

Because d_{ni}, $i = 1,2,\ldots,k+1$, have a multinomial distribution with cell probabilities ϕ_i, $i = 1,2,\ldots,k+1$, respectively, $\sqrt{n}(\hat{F}_n(t) - F_T(t)) \xrightarrow{d} N(0,\sigma_{\hat{F}}^2(t))$, where $\sigma_{\hat{F}}^2(t) = \Sigma_{\hat{F}}(t) - u_{\hat{F}}^2(t)$,

$$\begin{aligned}
\Sigma_{\hat{F}}(t) &= \sum_{i=1}^{k+1}\left(\frac{I(\xi_i \le t) - \rho^\top(t)G(T,\xi_i,\boldsymbol{\theta})}{\sum_{j=i}^{k+1}Q_j}\right)^2 P_i\sum_{j=i}^{k+1}Q_j, \quad \text{and}\\
u_{\hat{F}}(t) &= \sum_{i=1}^{k+1}\frac{I(\xi_i \le t) - \rho^\top(t)G(T,\xi_i,\boldsymbol{\theta})}{\sum_{j=i}^{k+1}Q_j}P_i\sum_{j=i}^{k+1}Q_j = \sum_{i=1}^{k+1}I(\xi_i \le t)P_i = F_T(t).
\end{aligned}$$

After some algebraic manipulations, we have $\Sigma_{\hat{F}}^2(t) = \sum_{i:\xi_i\le t}P_i/(\sum_{j=i}^{k+1}Q_j) - \rho^\top(t)\Sigma_1\rho(t)$. Therefore,

$$\begin{aligned}
\sigma_{\hat{F}}(t) &= \Sigma_{\hat{F}}(t) - u_{\hat{F}}^2(t) = \sum_{i:\xi_i\le t}P_i/(\sum_{j=i}^{k+1}Q_j) - F_T^2(t) - \rho^\top(t)\Sigma_1\rho(t)\\
&= \sum_{i:\xi_i\le t}P_i/(\sum_{j=i}^{k+1}Q_j) - F_T^2(t) - \beta^\top(t)\Sigma_1^{-1}\Sigma_2[\Sigma_2^\top\Sigma_1^{-1}\Sigma_2]^{-1}\Sigma_2^\top\Sigma_1^{-1}\beta(t),
\end{aligned} \tag{18}$$

where $\beta(t)$ is given in (17) and Σ_1 and Σ_2 are given by (12) and (14), respectively. We state this result formally in the next theorem.

Theorem 4.4 *Let $\boldsymbol{P} = \{P_1,\ldots,P_{k+1}\}$ and $\boldsymbol{Q} = \{Q_1,\ldots,Q_{k+1}\}$ be the true discretized distributions of the lifetime T and censoring time W, at the pre-fixed time-points $0 < \xi_1 < \xi_2 < \ldots < \xi_k < L \equiv \xi_{k+1}$ as defined in (4), and $F_T(t) = \sum_{i:\xi_i\le t}P_i$ and $\hat{F}_n(t) = \sum_{\xi_i\le t}P_{ni}(\hat{\boldsymbol{\theta}}_n)$. Under the conditions of Theorem 3.1, we have that $\sqrt{n}(\hat{F}_n(t) - F_T(t,\boldsymbol{\theta})) \xrightarrow{d} N(0,\sigma_{\hat{F}}^2(t))$, where $\sigma_{\hat{F}}^2(t)$ is given in (18).*

In the complete sample case, $S_W(T) = 1$. Thus, $\sigma_{\hat{F}_n}^2(t) < F_T(t)(1 - F_T(t))$ as long as $\beta(t) = \sum_{i:\xi_i\le t}G(T,\xi_i,\boldsymbol{\theta})P_i = \mathrm{E}(G(T,\boldsymbol{\theta})I(T \le t)) \neq 0$.

5. Numerical examples

5.1 Asymptotic efficiency studies

This section provides some numerical comparisons (in the sense of asymptotic relative efficiency) of the parametric (MLE), constrained SMLE, unconstrained SMLE (uSMLE), and mis-specified parametric MLE (misMLE) in both the completed and censored data cases.

Consider type II censored data with the censoring rates, 10%, 20% and 30%. Let $0 = \xi_0 < \xi_1 = h < \xi_2 = 2h < \ldots < kh = \xi_k < (k+1)h = L$ be k pre-fixed points, and L is the ending time of the experiment. This setup allows us the study the effect of changing interval sizes by varying a single constant h. See Tables 1 and 2 for details. Thus, the profile likelihood of θ becomes

$$L(T,\theta) = \prod_{i=1}^{k} (S_T(ih,\theta) - S_T((i+1)h,\theta))^{d_i} S_T(kh,\theta)^{c_{k+1}}, \tag{19}$$

where d_i is the number of deaths observed in the interval $[ih,(i+1)h)$, $i = 1, 2, \ldots, k$, and c_{k+1} is the number of withdrawals after $\xi_k = hk$.

Consider the case that the lifetime T is from an exponential distribution $Exp(\theta)$ with mean $\theta = 1.5$ or a log-normal distribution that $Y = log(T)$ has a normal distribution with mean $\mu = 0.0589$ and variance $\sigma^2 = 0.693$ such that $\mathrm{Var}(T) = \mathrm{E}(T)^2$. Note that the lognormal mean parameter $\theta = \mathrm{E}(T) = e^{\mu+\sigma^2/2}$, and the variance parameter $\mathrm{Var}(T) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$.

Denoted by $\hat{\theta}_{MLE} = \hat{\theta}_1$ the MLE from the parametric models. Then, the asymptotic variance of $\sqrt{n}(\hat{\theta}_1 - \theta)$ can be obtained from the expectation of the Fisher information matrix $[\mathrm{E}(\partial^2 \log L(T,\theta)/\partial\theta^2)]^{-1}$, where $L(T,\theta)$ is defined in (19) with $S_T(t,\theta) = e^{-t/\theta}$ in the exponential case.

For the MLE $\hat{\theta}_1$ of the lognormal mean, the asymptotic of variance of $\sqrt{n}(\hat{\theta}_1 - \theta)$ can be obtained from applying the δmethod to the Fisher information matrix of the MLEs $\hat{\mu}$ and $\hat{\sigma}$, $\boldsymbol{I}(\mu,\sigma) = \mathrm{E}(\partial^2 \log L(Y,\mu,\sigma)/\partial\mu\partial\sigma)$, where $L(Y,\mu,\sigma)$ is defined in (19) with

$$S_Y(y,\mu,\sigma) = \int_y^\infty \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 1 - \Phi\left(\frac{y-\mu}{\sigma}\right).$$

Denoted by $\hat{\theta}_2$ the constrained SMLE of θ with model assumption $\mathrm{Var}(T) = \mathrm{E}(T)^2$. The estimating equations for the SMLE are $G_2(T,\theta) = (T - \theta, T^2 - 2\theta^2) = 0$. Denoted by $\hat{\theta}_3$ the unconstrained SMLE (uSMLE) of mean parameter θ. The estimating equation of the uSMLE is $G_1(T,\theta) = T - \theta$. Note that this estimating function does not impose any constraint on θ, but define θ as the mean parameter. In the completed sample case, $\hat{\theta}_3$ is nothing but the sample mean, and is the same as the MLE of θ in the exponential case. The asymptotic variance of both uSMLE $\sqrt{n}(\hat{\theta}_2 - \theta)$ and SMLE $\sqrt{n}(\hat{\theta}_3 - \theta)$ are as given in Theorem 4.2.

The parametric MLE is sensitive to assumption of the true models, and always loses efficiency if the model is mis-specified. We use the exponential MLE in the log-normal situation to assess the efficiency loss in the case of mis-specified MLE. Denoted by $\hat{\theta}_4$ the mis-secified MLE (misMLE) from using the exponential likelihood with mean $\theta = 1.5$, i.e., $\hat{\theta}_4$ is obtained by solving likelihood equation $\partial \log L(T, \theta)/\partial\theta = 0$, where $L(T, \theta)$ is defined in (19) with $S_T(t, \theta) = e^{-t/\theta}$. However, the asymptotic mean square error of $\hat{\theta}_4$, $\text{asvar}(\hat{\theta}_4) = \lim_{n\to\infty} \mathrm{E}((\hat{\theta}_4 - \theta)^2)$ is calculated with respect to the true model, the log-normal distribution.

Define the asymptotic relative efficiency of $\hat{\theta}_i$, $i = 2, 3$ and 4, as follows:

$$\text{asyeff}(\hat{\theta}_i) = \left(\frac{\text{asy. MSE of MLE}}{\text{asy. MSE of } \hat{\theta}_i}\right) \times 100\%.$$

For various interval size h, Table 1 presents the asymptotic relative efficiency of SMLE $\hat{\theta}_2$ and uSMLE $\hat{\theta}_3$ compared with the parametric MLE in the exponential case.

(Please place Table 1 here.)

Note that, the exponential distribution has the property of $\mathrm{Var}(T) = \mathrm{E}(T)^2$. In the completed sample case, with smaller interval sizes, uSMLE and SMLE are almost fully asymptotically efficient. However, the loss of information in uSMLE is faster than in SMLE as the interval size becomes larger. For example, when interval size $h = 1.25$, SMLE remains more than 90% efficiency, while uSMLE is about 70%.

In censoring cases, the advantage of SMLE is clearer. For the 10%, 20% and 30% censoring cases, and all interval sizes used, SMLE has about 90%, 80% and 70% efficiency, respectively, while uSMLE only has about 40%, 35% and 33% efficiency, respectively.

It should be noted that varying the interval size does not change the conclusion of the asymptotic efficiency much in both completed and censored sample cases. For example, even in the exponential distribution case, in which variance is equal to the square of mean, the constrained SMLE has an average of 100% better asymptotic efficiency than non-constrained SMLE if the lifetime data are censored.

For various interval size h, Table 2 presents the asymptotic relative efficiency of SMLE $\hat{\theta}_2$, uSMLE $\hat{\theta}_3$ and misMLE $\hat{\theta}_4$ compared with parametric MLE in the log-normal case.

(Please place Table 2 here.)

Table 2 shows that, in the completed sample case, SMLE has an average of 80% efficiency, uSMLE has an average of 70% efficiency, and misMLE has about 65% efficiency. In this case, because exponential MLE from grouped data is very close to the sample mean (the same as uSMLE in this case), the misMLE does not loss efficiency much. However, in censoring cases, the performance of misMLE is terribly bad.

In the light censoring case (10% censoring), the performances of SMLE and uSMLE are almost same, which is about 60% efficiency, but SMLE is slightly better than uSMLE uniformly over all interval sizes. The asymptotic efficiency of the misMLE is slightly over 10% for interval size $h \leq 0.2$, and about 6.5% for $0.3 \leq h \leq 0.5$. The efficiency of misMLE is sensitive to the interval size. The larger interval size h is, the less efficiency it is.

For 20% and 30% censoring cases, and all interval sizes used, SMLE has about 48% and 43% efficiency, respectively, while uSMLE only has about 38% and 30% efficiency, respectively. The constrained SMLE has an average of 25% better asymptotic efficiency than the non-constrained SMLE if the lifetime data are censored more than 20%. The performance of misMLE is unacceptable in this case. When censoring is more than 20%, misMLE only has about 10% efficiency for small interval sizes ($h \leq 0.3$), and almost has no efficiency at all (less than 3%) if the interval size $h \geq 0.4$. Moreover, misMLE has problems other than the efficiency. The misMLE obtained from exponential likelihood is not consistent. It is easy to know that $\Pr(\lim_{n\to\infty} \hat{\theta}_4 \neq \theta) = 1$.

From Tables 1 and 2, we can conclude that the SMLE has an average of 15% better asymptotic efficiency than uSMLE in completed sample cases. If the distribution assumption is correct for the parametric MLE, the performance of the SMLE is rather close to the MLE with more than 80% asymptotic efficiency. When the distribution assumption is incorrect, we know that misMLE performs terribly. In the censored sample cases, SMLE has an average of 100% better asymptotic efficiency than uSMLE. The misMLE has almost no efficiency if the censoring is more than 20%. The interval size for the data grouping does not impact SMLE efficiency much in both completed and censored samples.

In conclusion, when compared to the parametric MLE with a "correct" distribution used, the SMLE does not lose much efficiency asymptotically. On the other hand, when the parametric MLE mis-specifies the underlying distribution or does not use the information of distribution structure such as variance is a function of mean, the SMLE can be much more efficient than the mis-specified parametric MLEs or regular nonparametric MLE method, i.e., unconstrained SMLE.

5.2 Real life example: HIV Data

This section presents an example of using SMLE method to evaluate the survival time of HIV+ infected subjects. It was a follow-up study conducted by Health Maintenance Organization (Hosmer & Lemeshow, 1999, p.1-17 and 76-77). 100 HIV+ subjects were enrolled in the study from January 1, 1989 to December 31, 1991. The study ended on December 31, 1995. After a confirmed diagnosis of HIV, subjects were followed until death due to AIDS or AIDS-related events. The survival time is considered as censored if the subject was lost to follow-up. Data presented in Figure 1 for 100 subjects are the number of deaths (closed symbol) and censoring (open symbol) by time. The survival time is in month. The data have the group-censored structure as studied in this paper.

(Please place Figure 1 here.)

Note that the data collected in the follow-up study was ended at the 60th month, and two subjects were censored at the end. Conservatively, we assume that the distribution of the survival probability has a very lighter tail beyond $t = 65$, and assign the survival probability beyond survival time $t = 60$ at time point $L = \xi_{k+1} = 65$.

The commonly used Kaplan-Meier (KM) method provides an estimate the survival probability as given in Figure 2. The mean survival time and the standard deviation (SD) are 14.86 and 18.34, respectively. Note that the mean and the SD are rather close, and the survival function resembles to the exponential. Thus, the exponential distribution is entertained, and the estimate of survival function is plotted in Figure 2. Its estimates of survival probabilities for time smaller than 24 months are higher, but for time greater than 25 months are lower.

In the next experiment, we try the semi-parametric approach with the SMLE. The constraint "$Variance = Mean^2$" from the exponential property is used in the estimating equations $G(T, \theta)$. However, we note that this constraint does not impose the full distribution assumption like the exponential model.

In this case, $G(T, \theta) = (T - \theta, T^2 - 2\theta^2)$, where $\theta = \mathrm{E}(T)$ is the mean parameter, and $\partial G/\partial\theta = (-1, -4\theta)$. It follows from equation (4) that we have $\lambda_1 = -4\lambda_2\theta$, and then, equation (7) can be written as following two equations:

$$\sum_{i=1}^{k+1} \frac{d_i(\xi_i - \theta)}{1 - a_i + \lambda_2(\xi^2 - 4\xi_i\theta + 2\theta^2)} = 0 \quad \text{and} \quad \sum_{i=1}^{k+1} \frac{d_i(\xi_i^2 - 2\theta^2)}{1 - a_i + \lambda_2(\xi^2 - 4\xi_i\theta + 2\theta^2)} = 0. \tag{20}$$

Solving the equations in (20) simultaneously with careful grid search technique and the initial value $\lambda_2 = 0$ and $\theta = 14.86$, we have the solution of $\lambda_2 = 0.000545$ and $\theta = 19.414$. Then the $\hat{P}_i$ can be calculated with (5), i.e., $\hat{P}_i = d_i/n(1 - a_i(\boldsymbol{\theta}) + \boldsymbol{\lambda}_2(\xi^2 - 4\xi_i\theta + 2\theta^2)$, $i = 1, 2, \cdots, k$, where $a_1 = 0$, and $a_i = a_{i-1} + c_{nm}/n\hat{S}(\xi_{i-1})$, $i = 2, \cdots, k$, and $\hat{P}_{k+1} = 1 - \hat{S}(\xi_k) = 1 - \sum_{i=1}^{k} \hat{P}_i$.

SMLE's estimate of survival function is plotted in Figure 2. For time smaller than 15 months, SMLE estimate is between the exponential and the KM estimates. For time greater than 15 months, the SMLE estimate is higher than the other two estimates. Because that there are lots of censored data occurred in the early stage, e.g., 17% of subjects are censored before 25 months, the use of certain "knowledge" about the modeling distribution in the constraint can make a difference in the estimate. Investigations of the standard deviation (SD) of the survival function show that the SMLE has smaller SD than the KM and exponential estimates in the early stages. See Figure 3 for details.

According to Lawless (1982, p.74), the SD for Kaplan-Meier estimate is calculated by $\sigma(t)_{KM} = \hat{S}(t)_{KM}\sqrt{\sum_{j:\xi_j\le t} d_j/n_j(n_j-d_j)}$, where $\hat{S}(t)_{KM}$ is the Kaplan-Meier estimate of survival probability.

The SD of the SMLE is calculated according to (18). Based on Lemma 4.1 and equation (12) and (14), the following formula is used in calculation.

$$\sigma^2(t)_{SMLE} = \sum_{i:\xi_i\le t} \hat{P}_i/(1-a_i) - \hat{F}_T^2(t) - \beta^\top(t)\boldsymbol{B}^{-1}\boldsymbol{A}[\boldsymbol{A}^\top\boldsymbol{B}^{-1}\boldsymbol{A}]^{-1}\boldsymbol{A}^\top\boldsymbol{B}^{-1}\beta(t),$$

where $\boldsymbol{A} = \sum_{i=1}^{k+1} \dot{G}(T,\xi_i,\theta)\hat{P}_i = (-1,-4\theta)$, $\boldsymbol{B} = \sum_{i=1}^{k+1} \hat{P}_i G(T,\xi_i,\theta)G^\top(T,\xi_i,\theta)/n(1-a_i)$, and $\beta(t) = \sum_{i:\xi_i\le t} G(T,\xi_i,\theta)\hat{P}_i$.

The SD of SMLE for survival probability is smaller then that of Kaplan-Meier estimate for the survival time t ¡ 36 months. But, for longer survival time, the SD of SMLE remains about 0.049. It is reasonable in the sense that due to the heavy censoring at early stages, the estimating survival probability for longer survival time is harder. From the first term in (18), we can see that the censoring has an important impact on the asymptotic SD. When the censoring is heave in early stage, the denominator $\sum_{i:\xi_i\ge t}\hat{Q}_i = 1-a_i$ in the first term would be smaller for longer survival time t, which could result in a larger SD for SMLE $\hat{S}(t)$ there.

(Please place Figures 2 and 3 here.)

6. Semiparametric likelihood ratio tests (SLRT)

The large sample confidence intervals of P_i and $F_T(t)$ for a given t can be obtained from the normal theory with the asymptotic distributions given in Theorems 4.3 and 4.4, respectively, or from inverting likelihood ratio tests. This section presents the SLRTs for a simple hypothesis $H_{10}:\boldsymbol{\theta}=\boldsymbol{\theta}_0$, testing of partial components of parameters, and a hypothesis $H_{20}:\mathrm{E}(G(T,\boldsymbol{\theta}))=0$ for checking the adequacy of the estimating equation model.

The SLRT statistic for the hypothesis $H_{10}:\boldsymbol{\theta}=\boldsymbol{\theta}_0$ is $\Lambda_n(\boldsymbol{\theta}_0) = 2n(H_n(\hat{\boldsymbol{\theta}}_n)-H_n(\boldsymbol{\theta}_0))$, where $H_n(\boldsymbol{\theta})$ is given in (8). Taylor expansion of $nH_n(\boldsymbol{\theta}_0)$ at $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n$ yields

$$nH_n(\boldsymbol{\theta}_0) = nH_n(\hat{\boldsymbol{\theta}}_n) + \frac{n}{2}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)^\top\frac{\partial^2 H_n(\hat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)+o(1).$$

It follows from Theorem 4.1 that $\partial^2 H_n(\hat{\boldsymbol{\theta}}_n)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top = \hat{\Sigma}_q \xrightarrow{w.p.1} \Sigma_2^\top\Sigma_1^{-1}\Sigma_2$.

From the above two equations, we note that $\Lambda_n(\boldsymbol{\theta}_0)$ has the same asymptotic distribution as $n(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)^\top[\Sigma_2^\top\Sigma_1^{-1}\Sigma_2](\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)$ does. Then, applying Theorem 4.2, we obtain the needed result as stated in the next theorem.

Theorem 6.1 *Under the regularity conditions* **R.1 - R.3***, when* $H_{10} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ *is true,* $\Lambda_n(\boldsymbol{\theta}_0) = 2n(H_n(\hat{\boldsymbol{\theta}}_n) - H_n(\boldsymbol{\theta}_0)) \xrightarrow{d} \chi^2(p)$*, where* p *is the number of the parameters in* $\boldsymbol{\theta}$*.*

Consider testing hypothesis $H_{20} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$ of partial components of parameters in $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are m and $p-m$ dimensional parameter vectors, respectively. The SLRT statistic for testing H_{20} is $\Lambda_{2n}(\boldsymbol{\theta}_{10}) = 2n(H_n(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) - H_n(\boldsymbol{\theta}_{10}, \tilde{\boldsymbol{\theta}}_2))$, where $\tilde{\boldsymbol{\theta}}_2$ maximizes $H_n(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_2)$ with respect to $\boldsymbol{\theta}_2$. One can derive the asymptotic distribution of Λ_{2n} based on the fact that, under the regularity conditions **R.1 - R.3**, there is $\hat{\boldsymbol{\theta}}_2 - \tilde{\boldsymbol{\theta}}_2 = o_p(n^{-1/2})$ when $H_{20} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$ is true. The fact can be shown by applying Taylor expansions to $nH_n(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})$ at $(\boldsymbol{\theta}_{10}, \hat{\boldsymbol{\theta}}_2)$ and $(\boldsymbol{\theta}_{10}, \tilde{\boldsymbol{\theta}}_2)$, respectively. Then, concentrating on $\boldsymbol{\theta}_1$ parameters and using the results in Theorems 4.1 and 4.2, we obtain the next theorem similar to Theorem 6.1.

Theorem 6.2 *When* $H_{20} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$ *is true, under the regularity conditions* **R.1 - R.3***,* $\Lambda_{2n}(\boldsymbol{\theta}_{10}) = 2n(H_n(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) - H_n(\boldsymbol{\theta}_{10}, \tilde{\boldsymbol{\theta}}_2)) \xrightarrow{d} \chi^2(m)$*, where* m *is the number of parameters in* $\boldsymbol{\theta}_1$*, and* $\tilde{\boldsymbol{\theta}}_2$ *maximizes* $H_n(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_2)$ *with respect to* $\boldsymbol{\theta}_2$*.*

It is important to test if the model $\mathrm{E}(G(T, \boldsymbol{\theta})) = 0$ assumed in the estimating equations is true. For testing this hypothesis, the SLRT statistic is $\Lambda_{2n}(\boldsymbol{\theta}) = 2n(H_n^* - H_n(\hat{\boldsymbol{\theta}}_n))$, where H_n^* is the maximum of the empirical likelihood in the NPMLE approach without any constraint (given by the Kaplan-Meier estimate), and $H_n(\hat{\boldsymbol{\theta}}_n)$ is the maximum of the same likelihood in the SMLE approach.

To derive the asymptotic distribution of $\Lambda_{2n}(\boldsymbol{\theta})$, we first recall that $G(t, \boldsymbol{\theta}) = (g_1(t, \boldsymbol{\theta}), g_2(t, \boldsymbol{\theta}), \cdots, g_r(t, \boldsymbol{\theta}))$, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_p)$. Note that any p of r ($r > p$) equations $\mathrm{E}(g_i(T, \boldsymbol{\theta})) = 0$, $i = 1, 2, \ldots, r$, can define parameter $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_p)$. Conveniently, let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_p)$ be defined by the first p equations, i.e., $\mathrm{E}(g_i(T, \boldsymbol{\theta})) = 0$, $i = 1, 2, \ldots, p$. Denoted by $\beta_j = \mathrm{E}(g_j(T, \boldsymbol{\theta}))$, $j = p+1, p+2, \ldots, r$, and $\tau = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_p, \beta_1, \beta_2, \ldots, \beta_{r-p})$. Then, the r-dimensional parameter vector τ can be determined by the following estimating equations, $\mathrm{E}(G^*(T, \delta)) = 0$, where $G^*(t, \delta) = (g_1^*(t, \delta), g_2^*(t, \delta), \ldots, g_r^*(t, \delta))$, and $g_i^*(t, \delta) = g_i(t, \boldsymbol{\theta})$, $i = 1, 2, \ldots, p$, $g_i^*(t, \delta) = g_i(t, \boldsymbol{\theta}) - \beta_{i-p}$, $i = p+1, p+2, \ldots, r$. With this re-parameterization, testing the model $H_{20} : \mathrm{E}(G(T, \boldsymbol{\theta})) = 0$ is equivalent to testing $H_{20}^* : \beta = 0$, where $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_{r-p})$.

Let us use a simple example to illustrate this re-parameterization process. Consider testing a simple hypothesis that $H_{20}: \ \mathrm{E}(T)^2 = \mathrm{Var}(T)$. The G function in this case is $G(T, \boldsymbol{\theta}) = (T - \boldsymbol{\theta}, T^2 - 2\boldsymbol{\theta}^2)$ with $r = 2$ and $p = 1$. One can choose either function in G to define $\boldsymbol{\theta}$. Let θ be defined by $\mathrm{E}(T - \theta) = 0$, i.e., $\theta = \mathrm{E}(T)$. Let $\beta = \mathrm{E}(T^2 - 2\theta^2)$, and $\tau = (\theta, \beta)$. Then, $G^*(T, \tau) = (T - \theta, T^2 - 2\theta^2 - \beta)$. Because $\beta = \mathrm{E}(T^2 - 2\theta^2) = \mathrm{E}(T^2) - \mathrm{E}(T)^2 - \theta^2 = \mathrm{Var}(T) - \theta = \mathrm{Var}(T) - \mathrm{E}(T)^2$, testing model $H_{20}: \ \mathrm{E}(T)^2 = \mathrm{Var}(T)$ is equivalent to testing $H_{20}^* : \beta = 0$. By augmenting the parameters $\boldsymbol{\theta}$ with β in the new estimating equations, we change the problem of testing the model defined by the estimating equations into testing partial set of parameters $\tau = (\theta, \beta)$ equal to a set of specified values.

In the above example, the SLRT statistic in this case becomes $\tilde{\Lambda}_{2n} = 2n(H_n(\hat{\tau}) - H_n(\hat{\theta}, 0))$. Because in the case of $p = r$ the estimating equations $\mathrm{E}(G^*(T, \tau)) = 0$ makes no constraint on parameter τ, $H_n(\hat{\tau})$ is then equal to H_n^*, the maximum of the empirical likelihood in the NPMLE approach. With $H_n(\hat{\theta}_n, 0) = H_n(\hat{\theta}_n)$, the maximum likelihood in the SMLE approach, we have $\tilde{\Lambda}_{2n} = 2n(H_n(\hat{\tau}) - H_n(\hat{\theta}_n, 0)) = 2n(H_n^* - H_n(\hat{\theta}_n)) = \Lambda_{2n}$. Then, it follows from Theorem 6.2 that we have the following theorem.

Theorem 6.3. *When the model $\mathrm{E}(G(T, \boldsymbol{\theta})) = 0$ is true, under the regularity conditions* **R.1 - R.3***, we have $\Lambda_{2n}(\boldsymbol{\theta}) = 2n(H_n^* - H_n(\hat{\boldsymbol{\theta}}_n)) \xrightarrow{d} \chi^2(r-p)$, where r is the number of the functions in $G(T, \boldsymbol{\theta})$ and p is the number of the parameters in $\boldsymbol{\theta}$.*

Appendix

Here we give proofs for the results in Sections 3 and 4.

A. The proof of Lemma 3.1

Recall that $G(\xi, \theta) = (g_1(\xi, \theta), g_2(\xi, \theta), \ldots, g_r(\xi, \theta))$. Denoted $\mathbf{G} = (\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_r)$, where $\boldsymbol{g}_j = (g_j(\xi_0, \theta), g_j(\xi_1, \theta), \ldots, g_j(\xi_k, \theta))$, $j = 1, 2, \ldots, r$, and $\boldsymbol{g}_j \in R^{k+1}$, and $\mathbf{G}$ is a $(k+1) \times r$ matrix. Let $\mathcal{C}(\mathbf{G})$ denote the linear space spanned by the column vectors of matrix $\mathbf{G}$, $\mathcal{C}^{\perp}(\mathbf{G})$ is the orthogonal complement of $\mathcal{C}(\mathbf{G})$. Then, $\sum_{j=1}^{k+1} P_j G(\xi_{j-1}, \theta) = 0$ implies $\mathbf{G}^{\top}\boldsymbol{P} = 0$, i.e., $\boldsymbol{P} \in \mathcal{C}^{\perp}(\mathbf{G})$. I t follows from the linear space theory that $\boldsymbol{P}$ has the general form $\boldsymbol{P} = [I_{k+1} - \mathbf{G}(\mathbf{G}^{\top}\mathbf{G})^{-1}\mathbf{G}^{\top}]\alpha$, for some $\alpha \in R^{k+1}$. For $\boldsymbol{P} = (P_1, P_2, \ldots, P_{k+1})^{\top}$ given in (5), let $a_1 = 0$, $a_j = n^{-1}\sum_{m=1}^{j-1} c_m / \sum_{l=m+1}^{k+1} P_l$, $j = 2, 3, \ldots, k+1$, as given in (6), and $\beta_j{}^* = d_j/nP_j - (1 - a_j)$, $j = 1, 2, \ldots, k+1$. Then the equation system (5) can be expressed equivalently as $\mathbf{G}\boldsymbol{\lambda} = \boldsymbol{\beta}^*$.

The linear equation system with respect to $\boldsymbol{\lambda}$ has a solution, if and only if $\boldsymbol{\beta}^* \in C(\mathbf{G})$, i.e.,$\beta^{\top}\boldsymbol{P} = 0$. Note that we have $\boldsymbol{\beta}^{*\top}\boldsymbol{P} = \sum_{j=1}^{k+1} P_j\beta_j = n^{-1}(\sum_{j=1}^{k+1} d_j - n + \sum_{m=1}^{k} c_m) = 0$. Then, we know that the equation system of (5) - (7) has (at least) one solution, and it follows from linear space theory that the solution is unique if the rank of matrix $\mathbf{G}$ is r.

B. The derivatives of $P_i(\boldsymbol{\theta})$ and $\boldsymbol{\lambda}(\boldsymbol{\theta})$

Denoted Ψ be the lefe hand side of (7) and $\boldsymbol{P} = (P_1, P_2, \ldots, P_{k+1})$. Then, $\Psi = \boldsymbol{G}^{\top}\boldsymbol{P}$, where $\boldsymbol{G}$ is the $(k+1) \times r$ matrix with the ith row defined as $G^{\top}(\xi_i, \boldsymbol{\theta})$. Let $\boldsymbol{B}$ be the $(k+1) \times (k+1)$ diagonal matrix with the ith diagonal element $b_i = d_i/(1 - a_i + \boldsymbol{\lambda}^\top G(T, \xi_i, \boldsymbol{\theta}))^2$, $i = 1, 2, \ldots, k+1$. Then, we have $-\partial\Psi/\partial\boldsymbol{\lambda} = \boldsymbol{G}^\top \boldsymbol{B}\boldsymbol{G} > 0$, provided by regularity condition **R.2** that $\boldsymbol{G}^\top\boldsymbol{G} > 0$.

Note that $P_i(\boldsymbol{\theta})$ defined in equation (5) involves the vector $\boldsymbol{\lambda}$ obtained from solving equation (7), and $\partial\boldsymbol{\lambda}/\partial\boldsymbol{P} = -[\partial\Psi/\partial\boldsymbol{\lambda}]^{-1}\partial\Psi/\partial\boldsymbol{P} = -[\boldsymbol{G}^\top\boldsymbol{B}\boldsymbol{G}]^{-1}\partial\Psi/\partial\boldsymbol{P}$, where $\partial\Psi/\partial\boldsymbol{P} = \boldsymbol{G}^\top\boldsymbol{B}\dot{\boldsymbol{A}}$, $\dot{\boldsymbol{A}} = \partial\boldsymbol{A}/\partial\boldsymbol{P} = [\partial a_i/\partial P_j]_{(k+1)\times(k+1)}$,

$$\frac{\partial a_i}{\partial P_j} = \begin{cases} 0, & j = 1, \\ -\dfrac{1}{n}\displaystyle\sum_{m=1}^{i-1} \frac{c_{m+1}}{(\sum_{l=m}^{k+1} P_l)^2}, & j \geq i, \\ -\dfrac{1}{n}\displaystyle\sum_{m=1}^{j-1} \frac{c_{m+1}}{(\sum_{l=m}^{k+1} P_l)^2}, & 1 < j < i, i = 2, 3, \ldots, k+1, \end{cases}$$

and $\partial a_1/\partial P_j = 0$, $j = 1, 2, \ldots, k+1$. Therefore, $\partial\boldsymbol{\lambda}/\partial\boldsymbol{P} = -[\boldsymbol{G}^\top\boldsymbol{B}\boldsymbol{G}]^{-1}\boldsymbol{G}^\top\boldsymbol{B}\dot{\boldsymbol{A}}$.

With similar algebraic manipulations, we have $\partial\boldsymbol{\lambda}/\partial\boldsymbol{\theta} = -[\boldsymbol{G}^\top\boldsymbol{B}\boldsymbol{G}]^{-1}[\partial\Psi/\partial\boldsymbol{\theta}]$. Note that $\partial\Psi/\partial\boldsymbol{\theta} = \sum_{i=1}^{k+1} P_i\dot{\boldsymbol{G}}(T, \xi_i, \boldsymbol{\theta}) - \boldsymbol{G}^\top\boldsymbol{B}\boldsymbol{G}^*$ is well defined according to regularity condition **R.3**, where $\dot{\boldsymbol{G}}(\xi_i, \boldsymbol{\theta}) = \partial G(T, \xi_i, \boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is a $r \times p$ matrix, and $\boldsymbol{G}^*$ is a $(k+1) \times p$ matrix with ith row given as $\boldsymbol{\lambda}^\top\dot{\boldsymbol{G}}(\xi_i, \boldsymbol{\theta}), i = 1, 2, \ldots, k+1$. Let $\boldsymbol{\Gamma} = \sum_{i=1}^{k+1} P_i\dot{\boldsymbol{G}}(\xi_i, \boldsymbol{\theta})$, a $r \times p$ matrix, then we get $\partial\boldsymbol{\lambda}/\partial\boldsymbol{\theta} = -[\boldsymbol{G}^\top\boldsymbol{B}\boldsymbol{G}]^{-1}[\boldsymbol{\Gamma} - \boldsymbol{G}^\top\boldsymbol{B}\boldsymbol{G}^*]$.

Based on equations (5) and (7) that, for given $\boldsymbol{\theta}$, $P_i, i = 1, 2, \ldots, k+1$ can be determined by the following equation $L_i = d_i/P_i - 1 + a_i - \boldsymbol{\lambda}^\top G(T, \xi_i, \boldsymbol{\theta}) = 0$, $i = 1, 2, \ldots, k+1$.

Denote $\boldsymbol{L} = (L_1, L_2, \ldots, L_{k+1})$, and $\boldsymbol{D} = diag(d_1/P_1^2, d_2/P_2^2, \ldots, d_{k+1}/P_{k+1}^2)$. Then, it follows that $\partial\boldsymbol{L}/\partial\boldsymbol{P} = -\boldsymbol{D} - \dot{\boldsymbol{A}} - \boldsymbol{G}[\boldsymbol{G}^\top\boldsymbol{B}\boldsymbol{G}]^{-1}\boldsymbol{G}^\top\boldsymbol{B}\dot{\boldsymbol{A}}$ and $\partial\boldsymbol{L}/\partial\boldsymbol{\theta} = \boldsymbol{G}^* - \boldsymbol{G}[\boldsymbol{G}^\top\boldsymbol{B}\boldsymbol{G}]^{-1}[\boldsymbol{\Gamma} - \boldsymbol{G}^\top\boldsymbol{B}\boldsymbol{G}^*]$.

The matrix $\boldsymbol{D}$ is full ranked by definition, and the rank of $\dot{\boldsymbol{A}}$ is at most k because of the first column and the first row of $\dot{\boldsymbol{A}}$ are all 0's. Therefore, the inverse of $\partial\boldsymbol{L}/\partial\boldsymbol{P}$ exists. Thus, we know that $\partial\boldsymbol{P}/\partial\boldsymbol{\theta} == -[\partial\boldsymbol{L}/\partial\boldsymbol{P}]^{-1}[\partial\boldsymbol{L}/\partial\boldsymbol{\theta}]$ is well defined. Therefore, $P_i(\boldsymbol{\theta})$, $i = 1, 2, \ldots, k+1$, and $\boldsymbol{\lambda}(\boldsymbol{\theta})$ are differentiable with respect to $\boldsymbol{\theta}$.

C1. Proof of Lemma 4.1

We will use the inductive method in this proof. When $i = 1$, $\hat{P}_{n1} = d_{n1}/n(1 + \boldsymbol{\lambda}_n^\top G(\xi_1, \hat{\boldsymbol{\theta}}_n))$. Under the given condition, $\lim_{n\to\infty} \hat{P}_{n1} = \lim_{n\to\infty} d_{n1}/n = P_1 \sum_{i=1}^{k+1} Q_i = P_1$ w.p.1.

Now, suppose $\hat{P}_{ni} \stackrel{w.p.1}{\longrightarrow} P_i$, for all $i \leq m$. It follows from (1) that $c_{nj}/n \stackrel{w.p.1}{\longrightarrow} Q_j \sum_{i=j+1}^{k+1} P_i$. Then,

$$\hat{a}_{n,m+1} = \sum_{j=1}^{m} \frac{c_{nj}}{n} \frac{1}{1 - \sum_{i=1}^{j} \hat{P}_{ni}} \stackrel{w.p.1}{\longrightarrow} \sum_{j=1}^{m} Q_j \sum_{i=j+1}^{k+1} P_j \frac{1}{1 - \sum_{i=1}^{j} P_i} = \sum_{j=1}^{m} Q_j.$$

Similarly, from (1), we know that $d_{n,m+1}/n \stackrel{w.p.1}{\longrightarrow} P_{m+1} \sum_{i=m+1}^{k+1} Q_i = P_{m+1}(1 - \sum_{i=1}^{m} Q_i)$. Therefore,

$$\lim_{n\to\infty} \hat{P}_{n,m+1} = \lim_{n\to\infty} \frac{d_{n,m+1}}{n(1 - \hat{a}_{n,m+1})} = \frac{P_{m+1}(1 - \sum_{i=1}^{m} Q_i)}{(1 - \sum_{i=1}^{m} Q_i)} = P_{m+1} \quad \text{(w.p.1)}.$$

C2. Proof of Lemma 4.2

Denote $h(\boldsymbol{\lambda}_n) = \sum_{i=1}^{k+1} d_{ni}G(T,\xi_i,\boldsymbol{\theta})/n(1 - a_{ni} + \boldsymbol{\lambda}_n^\top G(T,\xi_i,\boldsymbol{\theta}))$. Because $\boldsymbol{\lambda}_n = O_p(n^{-1/2})$, Taylor expansion of $h(\boldsymbol{\lambda}_n)$ at $\boldsymbol{\lambda}_n = 0$ yields $0 = h(\boldsymbol{\lambda}_n) =$

$$h(0) + \left[\frac{\partial h(0)}{\partial \boldsymbol{\lambda}_n}\right]^\top \boldsymbol{\lambda}_n + o_p(n^{-1/2}) = \sum_{i=1}^{k+1} \frac{d_{ni}G(T,\xi_i,\boldsymbol{\theta})}{n(1-a_{ni})} + \sum_{i=1}^{k+1} \frac{d_{ni}G(T,\xi_i,\boldsymbol{\theta})G^\top(\xi_i,\boldsymbol{\theta})}{n(1-a_{ni})^2}\boldsymbol{\lambda}_n + o_p(n^{-1/2}),$$

and then we have (11).

D. Asymptotic normality of $\boldsymbol{Z}_n^*$

It follows from Lemma 4.1 that $\boldsymbol{Z}_n^*$ and

$$\boldsymbol{Z}_n^{**} = \sqrt{n}\sum_{i=1}^{k+1} \frac{G(T,\xi_i,\boldsymbol{\theta})d_{ni}}{n\sum_{m=i}^{k+1} Q_m} = \sum_{i=1}^{k+1} \frac{G(T,\xi_i,\boldsymbol{\theta})}{\sum_{m=i}^{k+1} Q_m}\sqrt{n}\left(\frac{d_{ni}}{n} - P_i\sum_{m=i}^{k+1} Q_m\right)$$

have the same asymptotic distribution. In the group-censored data case, these d_{ni} have a multinomial distribution with cell probabilities $\phi_i \equiv P_i \sum_{m=i}^{k+1} Q_m$, $i = 1,\ldots,k+1$, it follows from the asymptotic normality of multinomial distribution that $\boldsymbol{Z}_n^{**}$ is asymptotically normal with mean 0 and asymptotic covariance matrix

$$\Sigma = \sum_{i=1}^{k+1} \frac{G(T,\xi_i,\boldsymbol{\theta})\phi_i(1-\phi_i)G(T,\xi_i,\boldsymbol{\theta})^\top}{(\sum_{m=i}^{k+1} Q_m)^2} - \left[\sum_{i=1}^{k+1} \frac{G(T,\xi_i,\boldsymbol{\theta})}{\sum_{m=i}^{k+1} Q_m}\phi_i\right]\left[\sum_{i=1}^{k+1} \frac{G(T,\xi_i,\boldsymbol{\theta})}{\sum_{m=i}^{k+1} Q_m}\phi_i\right]^\top.$$

Finally, after some algebraic manipulations, it is easy to show that $\Sigma = \Sigma_1$.

References

Breslow, N. & Crowley, J. (1974), A large sample study of the life table and product limit estimates under random censorship, *The Annals of Statistics*, **2**(3), 437-453.

Cox, D. R. (1972), Regression models and life-tables, *Journal of the Royal Statistical Society*, **B. 34**, 187-202.

Hosmer Jr., D. W. & Lemeshow, S. (1999), *Applied Survival Analysis*, John Wiley & Sons, Inc. New York.

Jones, M. P. (1997), A class of semiparametric regression for the accelerated failure time model, *Biometrika*, **84**, 73-84.

Kaplan, E. L. & Meier, P. (1958), Nonparametric estimation from incomplete observation. *J. Amer. Statist. Assoc.*, **58**, 457-481.

Krall, J., Uthoff, V. & Harley, J. (1975), A step-up procedure for selecting variables associated with survival, *Biometrics*, **31**, 49-57.

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

Liang, K. Y. and Zeger, S. L. (1986), Longitudinal data analysis using generalized liners models, *Biometrika*, **73**, 12-22.

Lu, J. C., Chen, D. & Gan, N. (2002), Semi-Parametric Modeling And Likelihood Estimation with Estimating Equations, *Aust. N.Z. J. Stat.*, **44**(2), 193-212.

Meeker, W. Q. Jr. & Escobar, L. A. (1998), *Statistical Methods for Reliability Data.* John Wiley & Sons, New York.

Meeker, W. Q. & LuValle, M. J. (1995), An accelerated life test model based on reliability kinetics, Technometrics, **37**, 133-146.

Nelson, W. (1990), *Accelerated Testing*, John Wiley & Sons, New York.

Owen, A. B. (1988), Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, **75**, 237-249.

Owen, A. B. (1990), Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18**, 90-120.

Qin, J. & Lawless, J. F. (1994) Empirical likelihood and general estimating equations, *Ann. Statist.*, **22**, 300-325.

Qin, J. & Lawless, J. F. (1995), Estimating equations, empirical likelihood and constraints on parameters, *Canadian J. Statist.*, **23**, 145-159.

Tsiatis, A. (1981), A large sample study of Cox's regression model, *The Annals of Statistics*, **9**(1), 93-108.

Wedderburn, R. W. M. (1974), Quasi-likelihood functions, generalized linear models, and Gauss-Newton method, *Biometrika* **61**, 439-447.

TABLE 1

Asymptotic efficiency comparison: exponential distribution cases

mean $\theta = 1.5$	interval size h	0.10	0.25	0.50	0.75	1.00	1.25
Competed sample	SMLE $\hat{\theta}_2$	99.6%	99.2%	97.9%	95.9%	93.3%	90.4%
	uSMLE $\hat{\theta}_3$	99.4%	98.4%	94.8%	88.8%	81.1%	71.9%
10% censoring	SMLE $\hat{\theta}_2$	88.4%	88.6%	89.5%	87.1%	85.3%	92.3%
	uSMLE $\hat{\theta}_3$	44.1%	44.2%	43.8%	41.8%	38.9%	37.0%
20% censoring	SMLE $\hat{\theta}_2$	77.4%	77.8%	79.2%	85.0%	78.2%	93.9%
	uSMLE $\hat{\theta}_3$	36.8%	36.9%	36.8%	36.3%	33.6%	32.7%
30% censoring	SMLE $\hat{\theta}_2$	64.4%	64.4%	65.6%	66.3%	72.2%	72.5%
	uSMLE $\hat{\theta}_3$	34.9%	35.0%	35.0%	34.1%	33.1%	30.6%

Table 2

Asymptotic efficiency comparison: log-normal distribution cases

mean $\theta = 1.5$	interval size h	0.1	0.2	0.3	0.4	0.5
Competed sample	SMLE $\hat{\theta}_2$	80.8%	80.5%	80.0%	79.3%	78.6%
	uSMLE $\hat{\theta}_3$	70.7%	70.6%	70.2%	69.9%	69.5%
	misMLE $\hat{\theta}_4$	71.4%	64.5%	62.8%	68.8%	66.0%
10% censoring	SMLE $\hat{\theta}_2$	61.2%	63.1%	59.1%	62.7%	58.6%
	uSMLE $\hat{\theta}_3$	60.0%	62.5%	57.1%	62.2%	57.3%
	misMLE $\hat{\theta}_4$	10.2%	15.6%	6.30%	6.70%	6.72%
20% censoring	SMLE $\hat{\theta}_2$	48.0%	48.0%	46.7%	49.8%	45.1%
	uSMLE $\hat{\theta}_3$	38.9%	39.0%	37.1%	43.4%	35.4%
	misMLE $\hat{\theta}_4$	11.5%	12.4%	14.3%	2.56%	2.72%
30% censoring	SMLE $\hat{\theta}_2$	42.9%	43.9%	43.8%	41.9%	41.1%
	uSMLE $\hat{\theta}_3$	29.8%	31.6%	31.7%	29.0%	28.1%
	misMLE $\hat{\theta}_4$	10.2%	10.3%	10.2%	2.35%	2.38%

Data Table: HIV Data and the Estimates of Survival Probability with SMLE and Kaplan-Meier Methods

ξ_i	d_i	c_i	n_i	$\hat{P}_{i,KM}$	$\hat{S}(\xi_i)_{KM}$	SD_{KM}	$\hat{P}_{i,Smle}$	$\hat{S}(\xi_i)_{Smle}$	SD_{Smle}
1	15	2	100	0.1500	0.8500	0.0425	0.1095	0.8905	0.0308
2	5	5	83	0.0512	0.7988	0.0479	0.0383	0.8522	0.0349
3	10	2	73	0.1094	0.6894	0.0564	0.0828	0.7694	0.0416
4	4	1	61	0.0452	0.6442	0.0588	0.0350	0.7344	0.0436
5	7	0	56	0.0805	0.5637	0.0616	0.0641	0.6703	0.0465
6	2	1	49	0.0230	0.5407	0.0621	0.0190	0.6513	0.0471
7	6	1	46	0.0705	0.4702	0.0627	0.0597	0.5916	0.0484
8	4	0	39	0.0482	0.4220	0.0625	0.0419	0.5497	0.0489
9	3	0	35	0.0362	0.3858	0.0619	0.0326	0.5171	0.0490
10	3	1	32	0.0362	0.3496	0.0609	0.0337	0.4834	0.0489
11	3	0	28	0.0375	0.3121	0.0596	0.0358	0.4476	0.0485
12	2	2	25	0.0250	0.2871	0.0584	0.0248	0.4228	0.0481
13	1	0	21	0.0137	0.2734	0.0579	0.0137	0.4091	0.0479
14	1	0	20	0.0137	0.2597	0.0572	0.0142	0.3949	0.0476
15	2	0	19	0.0273	0.2324	0.0556	0.0295	0.3654	0.0470
19	0	1	17	0	0.2324	0.0556	0	0.3654	0.0470
22	1	0	16	0.0145	0.2179	0.0548	0.0202	0.3452	0.0468
24	0	1	15	0	0.2179	0.0548	0	0.3452	0.0468
30	1	0	14	0.0156	0.2023	0.0539	0.0282	0.3170	0.0470
31	1	0	13	0.0155	0.1868	0.0529	0.0290	0.2880	0.0471
32	1	0	12	0.0156	0.1712	0.0516	0.0297	0.2583	0.0471
34	1	0	11	0.0155	0.1557	0.0501	0.0308	0.2275	0.0472
35	1	0	10	0.0156	0.1401	0.0484	0.0313	0.1962	0.0471
36	1	0	9	0.0155	0.1246	0.0465	0.0317	0.1645	0.0468
43	1	0	8	0.0156	0.1090	0.0442	0.0311	0.1334	0.0472
53	1	0	7	0.0155	0.0935	0.0416	0.0237	0.1097	0.0478
54	1	0	6	0.0156	0.0779	0.0386	0.0229	0.0868	0.0483
56	0	1	5	0	0.0779	0.0386	0	0.0868	0.0483
57	1	0	4	0.0195	0.0584	0.0352	0.0266	0.0602	0.0493
58	1	0	3	0.0195	0.0389	0.0302	0.0252	0.0350	0.0497
60	0	2	2	0	0.0389	0.0302	0	0.0350	0.0497
65	2	0	0	0.0389	0		0.0350		

Note: ξ_i is survival time (in month). At time point ξ_i, d_i is the number of the deaths, c_i is the number of censoring and n_i is the number of subjects at risk.