On the generalized maximum likelihood estimator of survival function under Koziol-Green model

By: Haimeng Zhang, M. Bhaskara Rao, Rupa C. Mitra

Zhang, H., Rao, M.B., and Mitra, R.C. (2006). On the generalized maximum likelihood estimator of survival function under Koziol-Green model. *Journal of Statistical Planning and Inference*, 136(9), 3032 – 3051. doi: 10.1016/j.jspi.2004.12.001.

Made available courtesy of Elsevier: http://www.dx.doi.org/10.1016/j.jspi.2004.12.001

Abstract:

In this paper, we derive the generalized maximum likelihood estimator (GMLE) of the survival function of a random variable for censored data under the Koziol-Green model. Its small sample properties are compared with those of Abdushukurov-Cheng-Lin (ACL), Kaplan-Meier, and Nelson-Aalen type estimators. The large sample analysis indicates that the GMLE, ACL, and Nelson-Aalen type estimators are asymptotically equivalent.

Keywords: ACL estimator | Asymptotic equivalence | Censored data | Generalized maximum likelihood estimator | Koziol–Green model | Mean squared error | Nelson–Aalen type estimator

Article:

1. Introduction

Let T and C be two independent random variables with unknown distribution functions F and G, respectively. Let $\bar{F}$ and $\bar{G}$ be the corresponding survival functions. The random variables are said to have proportional hazards if $\bar{G} = (\bar{F})^{\beta_0}$ for all t⩾0 for some $\beta_0>0$, or equivalently, the hazard functions of T and C are proportional. This is the so-called Koziol-Green model. Let $Y=\min\{T,C\}=T\wedge C$ and $\Delta=I(T\leqslant C)$, where I(A) is the indicator function of the set *A*. Examples of data sets are available in the literature for which the Koziol‑Green model holds. For example, Hollander et al. (2001) presents a data set, originally considered by Fleming and Harrington (1991), on a liver study conducted by the Mayo Clinic from 1974 to 1984 for which the assumption of proportional hazards seems to hold. See also Henze (1993) andde Uña-Álverez et al. (1997).

The main objective of this paper is the nonparametric estimation of F and β based on n independent copies $(Y_1,\Delta_1),\ldots,(Y_n,\Delta_n)$ of (Y,Δ). Let $Y_{(1)}\leqslant\cdots\leqslant Y_{(n)}$ be the order statistics of Y 's and $\Delta_{[i]}$ the censoring indicator corresponding to $Y_{(i)}$. Note that $\Delta_{[1]},\ldots,\Delta_{[n]}$ are not the order statistics of $\Delta_1,\ldots,\Delta_n$. If we ignore the information that T and C have proportional hazards, Kaplan‑Meier estimator (Kaplan and Meier, 1958) can be used for estimating F . If $Y_1,\ldots,Y_n$ are distinct, which will be the case when F and G are continuous, the Kaplan‑Meier estimator is given by

equation(1)

$$\hat{\bar{F}}_{\mathrm{KM}}(t)=\prod_{i=1}^{nK_n(t)}\left(\frac{n-i}{n-i+1}\right)^{\Delta_{[i]}},\quad t>0,$$

where $K_n(t)=(1/n)\sum_{j=1}^{n}I(Y_j\leqslant t)$ is the empirical distribution of the *Y*s. Large sample properties of the Kaplan‑Meier estimator are well known. See Efron (1967), Gill (1983), and Breslow and Crowley (1974), among others. The small sample properties of the Kaplan‑Meier estimator have been documented by Chen et al. (1982) and Wellner (1985).

Abdushukurov (1984) and Cheng and Lin (1987) proposed a different estimator of F exploiting the fact that$T$ and C have a Koziol‑Green model. Here is their approach. Note that $P(\Delta=1)=p\equiv1/(1+\beta)$. The maximum likelihood estimator $\hat{p}_{\mathrm{ACL}}$ of p is $\hat{p}_{\mathrm{ACL}}=(1/n)\sum_{i=1}^{n}\Delta_i$. Consequently, the maximum likelihood estimator of β, denoted as $\hat{\beta}_{\mathrm{ACL}}$ from now on, is given by

equation(2)

$$\hat{\beta}_{\mathrm{ACL}}=\frac{1}{\hat{p}_{\mathrm{ACL}}}-1=\frac{n}{\sum_{i=1}^{n}\Delta_i}-1.$$

Let H be the distribution function of Y and $\bar{H}$ the associated survival function. Then $\bar{H}(t)=\bar{F}(t)\bar{G}(t)=(\bar{F})^{\beta+1}$ for all $t\geqslant0$, from which we have $\bar{F}(t)=(\bar{H})^{1/(1+\beta)},t\geqslant0$. The generalized maximum likelihood estimator (GMLE) of $\bar{H}(t)$ is given by $\hat{\bar{H}}(t)=(1/n)\sum_{i=1}^{n}I(Y_i>t)$. Stringing together the GMLE of $\bar{H}(t)$ and the maximum likelihood estimator of β, the Abdushukurov‐Cheng‐Lin (ACL) estimator of $\bar{F}(t)$, denoted as $\hat{\bar{F}}_{\mathrm{ACL}}(t)$, is given by

equation(3)

$$\hat{\bar{F}}_{\mathrm{ACL}}(t)=\left[\frac{1}{n}\sum_{i=1}^{n} I\left(Y_i>t\right)\right]^{\frac{1}{n}\sum_{i=1}^{n}\Delta_i}.$$

The small and large sample properties of the ACL estimator have been studied by Cheng and Lin (1987)and Cheng and Chang (1985). There are other estimators discussed in the literature, for example, see Ebrahimi (1985) and Pawlitschko (1999).

The generalized maximum likelihood method following Keifer and Wolfowitz (1956) theory has been a standard staple in nonparametric estimation. See Groeneboom and Wellner (1992) for a comprehensive exposition. One of the goals of the paper is to derive the GMLE of $\bar{F}$ and β, denoted by $\hat{\bar{F}}_{\mathrm{GMLE}}(t)$ and $\hat{\beta}_{\mathrm{GMLE}}$, respectively, simultaneously. In addition, we consider a variation of the GMLE, which we call GMLE1 and denote it $\hat{\bar{F}}_{\mathrm{GMLE1}}$. The variation stems when $\hat{\beta}_{\mathrm{GMLE}}$ is replaced by $\hat{\beta}_{\mathrm{ACL}}$ in the expression for $\hat{\bar{F}}_{\mathrm{GMLE}}$ (See Section 2).

Following the suggestion of one of the referees, we also derive a Nelson‐Aalen type estimator (NA estimator, for short) of the hazard function of T, from which we set out an estimator $\hat{\bar{F}}_{\mathrm{NA}}$ of the survival function of T (See Section 2). The estimators $\hat{\bar{F}}_{\mathrm{GMLE}}$, $\hat{\bar{F}}_{\mathrm{GMLE1}}$, $\hat{\bar{F}}_{\mathrm{NA}}$ and $\hat{\bar{F}}_{\mathrm{ACL}}(t)$ are all different.

Another goal is to compare the performance of the GMLE, GMLE1, NA, and ACL estimators in small as well as large samples. Our derivation of the GMLE presented in this paper was first outlined by Mitra (1991) in her dissertation work. Hollander et al. (2001) discuss computational aspects of the generalized maximum likelihood estimates of $\bar{F}$ and β. They have observed that the ACL estimate and GMLEs are different and the difference is small for large risk sets, which was based on some empirical experience.

In our paper, we explicitly give the GMLE in the special case when there are no ties in the observations, which is helpful in studying small and large sample properties of the GMLE.
In Section 3, we compare small sample properties of the Kaplan‐Meier, GMLE, GMLE1, ACL and NA estimators. Of the four estimators (GMLE, GMLE1, ACL, and NA) considered, GMLE gives largest likelihood of the data. Hollander et al. (2001) has presented likelihood calculations in an example for GMLE and ACL estimators. In this paper, we use the same example in our likelihood calculations. In small samples, we show that the NA estimator has the least mean

square error for a substantial range of *t* values. In bias comparisons, GMLE and GMLE1 fare better than the others. In Section 4, we focus on the large sample properties of the GMLE and NA estimator. Our study shows that the GMLE, GMLE1, and NA estimators are asymptotically equivalent to the ACL estimator, and therefore, share the same large sample properties as the ACL estimator. Furthermore, it should be noted that with the asymptotic efficiency of ACL estimator (See Hollander et al. (2001)) and the asymptotic equivalence of these estimators, GMLE, GMLE1, and NA estimators are asymptotically efficient as well.

The method of deriving influence functions for estimators is another way of comparing the performance of estimators in large samples. Lanius and Pawlitschko (2001) pursued this approach in great detail for comparing KM and ACL estimators. For the problem on hand, our work adequately compares the performance of the estimators in both small and large samples. A clear picture emerges, which is good enough in practice.

2. Generalized maximum likelihood estimator

We first derive the GMLEs of β and F. Suppose $(Y_1,\Delta_1),\ldots,(Y_n,\Delta_n)$ are n independent copies of (Y,Δ). Let $0<U_1<\cdots<U_k$ be the distinct values among $Y_1,\ldots,Y_n$. For j=1,…,k, define $D_j = \sum_{i=1}^{n} I(\Delta_i = 1, Y_i = U_j)$, and $C_j = \sum_{i=1}^{n} I(\Delta_i = 0, Y_i = U_j)$, which are the number of failure items and censored items at U_j, respectively.
Obviously, $\sum_{j=1}^{k} C_j + \sum_{j=1}^{k} D_j = n$ (the sample size).

In principle, the generalized maximum likelihood method involves maximizing the likelihood of the data over all distributions F of T and G of C. See Keifer and Wolfowitz (1956) and Johansen (1978). Suppose F and G are two probability distributions of T and C under which the data have a positive probability of occurrence. For i=1,…,k, let

$$p_i = P_F(T = U_i), \quad p_i' = P_F(U_{i-1} < T < U_i), \quad p_{k+1} = P_F(T > U_k),$$
$$q_i = P_G(C = U_i), \quad q_i' = P_G(U_{i-1} < C < U_i), \quad q_{k+1} = P_G(C > U_k).$$

Note that $\sum_{i=1}^{k}(p_i + p_i') + p_{k+1} = 1$, and $\sum_{i=1}^{k}(q_i + q_i') + q_{k+1} = 1$. Then, the generalized likelihood of the data is given by

$$L(F,G) = L = \prod_{i=1}^{k} p_i^{D_i} (P_G(C \geqslant U_i))^{D_i} q_i^{C_i} (P_F(T > U_i))^{C_i}$$

$$= \prod_{i=1}^{k} \left[p_i^{D_i} \left(q_i + \sum_{j=i+1}^{k} (q_j + q'_j) + q_{k+1} \right)^{D_i} \times q_i^{C_i} \left(\sum_{j=i+1}^{k} (p_j + p'_j) + p_{k+1} \right)^{C_i} \right].$$

Without loss of generality, we can assume that $p'_i = 0$ and $q'_i = 0$ for all i . Our problem reduces to maximizing the likelihood L over all distributions F and G which must satisfy an additional condition:

equation(4)

$$(P_F(T > t))^{\beta} = P_G(C > t), \quad \text{for all } t \geqslant 0 \text{ and for some } \beta > 0.$$

Therefore, from (4), the likelihood of the data can be simplified as

$$L = \prod_{i=1}^{k} \left[p_i^{D_i} (P(T \geqslant U_i))^{\beta D_i} (P(C \geqslant U_i) - P(C > U_i))^{C_i} (P(T > U_i))^{C_i} \right]$$

$$= \prod_{i=1}^{k} p_i^{D_i} \left(\sum_{j \geqslant i} p_j \right)^{\beta D_i} \left(\left(\sum_{j \geqslant i} p_j \right)^{\beta} - \left(\sum_{j \geqslant i+1} p_j \right)^{\beta} \right)^{C_i} \left(\sum_{j \geqslant i+1} p_j \right)^{C_i}.$$

Applying the transformation $a_i = p_i (\sum_{j \geqslant i} p_j)^{-1}, i=1,\dots,k$, we obtain

equation(5)

$$p_i = a_i \sum_{j \geqslant i} p_j = a_i \prod_{j=1}^{i-1} (1 - a_j), \quad \sum_{j \geqslant i} p_j = \prod_{j=1}^{i-1} (1 - a_j), \; i = 1, \dots, k+1$$

and hence the likelihood L is

equation(6)

$$L=\prod_{i=1}^{k} a_i^{D_i}\left(\prod_{j=1}^{i-1}(1-a_j)\right)^{(C_i+D_i)(1+\beta)} (1-a_i)^{C_i}\left[1-(1-a_i)^{\beta}\right]^{C_i}$$

$$=\prod_{i=1}^{k} a_i^{D_i}(1-a_i)^{C_i+n_i^*}\left[1-(1-a_i)^{\beta}\right]^{C_i}.$$

Here $n_i^*=\sum_{j\geqslant i+1}(D_j+C_j)(1+\beta)$. Note that we use $\prod_{i=1}^{k}\left(\prod_{j=1}^{i-1}(1-a_j)\right)^{(C_i+D_i)(1+\beta)}=\prod_{i=1}^{k}(1-a_i)^{n_i^*}$ in the last equality.

We have to maximize L over $0\leqslant a_i\leqslant 1$ and $\beta>0$. It is easy to see that L will be maximized at $\hat{a}_i$ and $\hat{\beta}$, which are the solutions of a system of non-linear equations

equation(7)

$$\frac{\partial \log(L)}{\partial a_i}=\frac{D_i}{a_i}-\frac{C_i+n_i^*}{1-a_i}+\frac{\beta C_i(1-a_i)^{\beta-1}}{1-(1-a_i)^{\beta}}=0, \quad i=1,\dots,k,$$

equation(8)

$$\frac{\partial \log(L)}{\partial \beta}=\sum_{i=1}^{k}\left[\log(1-a_i)\sum_{j\geqslant i+1}(D_j+C_j)-\frac{C_i(1-a_i)^{\beta}\log(1-a_i)}{1-(1-a_i)^{\beta}}\right]=0.$$

These equations can be solved by iterative procedures such as Newton‑Raphson method. Once the solution is obtained, we evaluate $p_1,\dots,p_{k+1}$ based on (5). The resultant estimator of β is denoted by $\hat{\beta}_{\mathrm{GMLE}}$. Also, the generalized maximum likelihood estimate $\hat{\bar{F}}_{\mathrm{GMLE}}(t)$ is given by

equation(9)

$$\hat{\bar{F}}_{\mathrm{GMLE}}(t)=P(T>t)=\sum_{j\geqslant i} p_j$$

$$=\prod_{j=1}^{r-1}(1-\hat{a}_j), \quad \textbf{if } U_{r-1}\leqslant t<U_r, \quad r=1,\dots,k,$$

$$=\prod_{j=1}^{k}(1-\hat{a}_j), \quad \textbf{if } t\geqslant U_k.$$

Note that the GMLE could be either a proper or improper survival function depending on whether or not the largest observation is censored. See also Hollander et al. (2001). We could make $\hat{\bar{F}}_{\mathrm{GMLE}}(t)$ as a proper survival function by redefining the estimator by

equation(10)

$$\hat{\bar{F}}_{\mathrm{GMLE}}(t) = \prod_{j=1}^{r-1}(1-\hat{a}_j), \quad \textbf{if } U_{r-1} \leqslant t < U_r, \ \ r = 1, \ldots, k,$$
$$= 0, \quad \textbf{if } t \geqslant U_k.$$

As per the suggestion of one of the referees, we work only with the version (9) of GMLE.

The GMLE is more tractable if $Y_1,\ldots,Y_n$ are distinct. In such a case, k=n and $D_i+C_i=1$ for any i . Let$Y_{(1)}<\cdots<Y_{(n)}$ be the order statistics of $Y_1,\ldots,Y_n$. Then $D_i=\Delta_{[i]}$, $C_i = 1 - \Delta_{[i]}, \ i = 1, \ldots, n$, and $n_i^* = \sum_{j \geqslant i+1}(D_i + C_i)(1+\beta) = (1+\beta)(n-i)$. Now the likelihood function (6) can be written as

equation(11)

$$\tilde{L} = \prod_{i=1}^{n} a_i^{\Delta_{[i]}}(1-a_i)^{1-\Delta_{[i]}+n_i^*}\left[1-(1-a_i)^{\beta}\right]^{1-\Delta_{[i]}}.$$

Note that the i th parameter a_i occurs only in the i th product in (11), and is free to range in (0,1). Therefore, (11) can be maximized overall by maximizing each term in (11) separately over a_i, and then maximizing over β. The explicit solution $\hat{a}_i$ is given by

equation(12)

$$\hat{a}_i = \begin{cases} (1+(n-i)(1+\beta))^{-1}, & \textbf{if } \Delta_{[i]} = 1, \\ 1 - \left(\dfrac{1+(n-i)(1+\beta)}{(n-i+1)(1+\beta)}\right)^{1/\beta}, & \textbf{if } \Delta_{[i]} = 0. \end{cases}$$

The maximized loglikelihood in β is

equation(13)

$$\log \tilde{L}(\beta) = \sum_{i=1}^{n}(1-\Delta_{[i]})\left[\frac{1+(n-i)(1+\beta)}{\beta}\log\frac{1+(n-i)(1+\beta)}{(n-i+1)(1+\beta)} + \log\frac{\beta}{(n-i+1)(1+\beta)}\right] + \sum_{i=1}^{n}\Delta_{[i]}\left[\log\frac{1}{1+(n-i)(1+\beta)} + (n-i)(1+\beta)\log\left(1-\frac{1}{1+(n-i)(1+\beta)}\right)\right].$$

Therefore, the GMLE $\hat{\beta}_{\text{GMLE}}$ of β is the solution of the following estimating equation

equation(14)

$$\sum_{i=1}^{n}\left[(n-i)\Delta_{[i]}\log\frac{(n-i)(1+\beta)}{1+(n-i)(1+\beta)} - (1-\Delta_{[i]})\frac{(n-i)+1}{\beta^2} \times \log\frac{1+(n-i)(1+\beta)}{(n-i+1)(1+\beta)}\right]=0.$$

After plugging $\beta=\hat{\beta}_{\text{GMLE}}$ in (12), the GMLE of $\bar{F}(t)$ under no ties is then given by

$$\hat{\bar{F}}_{\text{GMLE}}(t)=P(T>t)=\sum_{j\geqslant i} p_j=\prod_{j=1}^{i-1}(1-\hat{a}_j), \quad \text{if } Y_{(i-1)}\leqslant t<Y_{(i)}, \quad i=1,\ldots,n,$$
$$=\prod_{j=1}^{n}(1-\hat{a}_j), \quad \text{if } t>Y_{(n)}$$

with the convention that empty product=1. Following the notation in Section 1, we can rewrite the estimator as

equation(15)

$$\hat{\bar{F}}_{\text{GMLE}}(t)=\prod_{i=1}^{nK_n(t)}(1-\hat{a}_i), \quad t\geqslant 0.$$

We also look at a variation $\hat{\bar{F}}_{\text{GMLE1}}(t)$ of $\hat{\bar{F}}_{\text{GMLE}}(t)$, which is obtained by letting $\beta=\hat{\beta}_{\text{ACL}}$ in (12), i.e., in the case of no ties,

equation(16)

$$\hat{\hat{F}}_{\mathrm{GMLE1}}(t) = \prod_{i=1}^{n K_n(t)} (1 - \hat{b}_i), \quad t \geqslant 0,$$

where $\hat{b}_i$ is given by

$$\hat{b}_i = \begin{cases} (1 + (n-i)(1+\hat{\beta}_{\mathrm{ACL}}))^{-1}, & \text{if } \Delta_{[i]} = 1, \\ 1 - \left(\dfrac{1 + (n-i)(1+\hat{\beta}_{\mathrm{ACL}})}{(n-i+1)(1+\hat{\beta}_{\mathrm{ACL}})} \right)^{1/\hat{\beta}_{\mathrm{ACL}}}, & \text{if } \Delta_{[i]} = 0. \end{cases}$$

Next we consider the GMLE $\hat{\beta}_{\mathrm{GMLE}}$ of β from (14). It is well known that, under the assumption of Koziol–Green model, if T and C have continuous distribution functions and $0<P(T\leqslant C)<1$, then Y and Δ are independently distributed. See Allen (1963) and Chen et al. (1982). Under this environment, we have observed that $\Delta_{[1]},\ldots,\Delta_{[n]}$ are independently identically distributed with success probability $1/(1+\beta)$. This seems to be a new result. Consequently, in the estimating equation (14), $\Delta_{[i]}$ can be replaced by Δ_i. This fact is instrumental in studying asymptotic properties of $\hat{\beta}_{\mathrm{GMLE}}$ successfully.

Note that the estimator $\hat{\beta}_{\mathrm{GMLE}}$ is not defined under the following data scenarios.

1. $\Delta_1=\cdots=\Delta_n=0$;

2. $\Delta_1=\cdots=\Delta_n=1$;

3. $\Delta_1=\cdots=\Delta_{n-1}=0,\Delta_n=1$.

These scenarios have also been noted by Hollander et al. (2001).

It is clear that the solution $\hat{\beta}_{\mathrm{GMLE}}$ of (14) depends on $\Delta_1,\ldots,\Delta_n$. In order to emphasize the dependence, especially in (18), we will write the solution as $\hat{\beta}_{\mathrm{GMLE}}(\Delta_1,\ldots,\Delta_n)$. As has been pointed earlier, Eq. (14) has no solution under data scenarios 1–3. We need to define $\hat{\beta}_{\mathrm{GMLE}}$ for these data scenarios. We identify the following data scenarios

1′. $\Delta_1=1,\Delta_2=\cdots=\Delta_n=0$;

2′. $\Delta_1=\cdots=\Delta_{n-1}=1,\Delta_n=0$;

3′. $\Delta_1=1,\Delta_2=\cdots=\Delta_n=0$,

as the closest to the data scenarios 1–3, respectively, for each of which $\hat{\beta}_{\mathrm{GMLE}}$ is defined. We will use these $\hat{\beta}_{\mathrm{GMLE}}$s to cover scenarios 1–3. We still use the same notation $\hat{\beta}_{\mathrm{GMLE}}$ after the modifications. Note that $\hat{\beta}_{\mathrm{ACL}}$ is also not defined for scenario 1 and a similar adjustment is made accordingly.

As per the suggestion from one of the referees, we derive Nelson‑Aalen type of estimator of the survival function under the model on hand. More specifically, the total likelihood can be rewritten as

$$
\begin{aligned}
L &\equiv \prod_{i=1}^{n} [\bar{G}(Y_i)\,\mathrm{d}F(Y_i)]^{\Delta_i}\,[\bar{F}(Y_i)\,\mathrm{d}G(Y_i)]^{1-\Delta_i} \\
&= \prod_{i=1}^{n} \bar{F}^{1+\beta}(Y_i)\beta^{1-\Delta_i}\,\mathrm{d}F(Y_i)/\bar{F}(Y_i).
\end{aligned}
$$

Let $\lambda(t) = \mathrm{d}F(t)/\bar{F}(t)$ and $\Lambda(t) = \int_0^t \lambda(u)\,\mathrm{d}u$ be the hazard density and the cumulative hazard function. Rewrite the likelihood to obtain

$$
L = \prod_{i=1}^{n} \exp\{(1-\Delta_i)\log\beta - (1+\beta)\Lambda(Y_i) + \log\lambda(Y_i)\}.
$$

Assume no ties and let $Y_{(1)}<Y_{(2)}<\cdots<Y_{(n)}$ be the order statistics of $Y_1,Y_2,\ldots,Y_n$ and $\Delta_{[i]}$ the censoring indicator corresponding to $Y_{(i)},\ i = 1, 2, \ldots, n$. We also let the hazard mass at each observation $Y_{(i)}$ be λ_i, then we have $\Lambda(Y_{(i)}) = \sum_{j=1}^{i} \lambda_j$. Therefore, the likelihood is

$$
L = \prod_{i=1}^{n} \exp\left\{(1-\Delta_{[i]})\log\beta - (1+\beta)\sum_{j=1}^{i}\lambda_j + \log\lambda_i\right\}.
$$

The loglikelihood will then be

$$
\begin{aligned}
\log L &= \sum_{i=1}^{n}\left[(1-\Delta_{[i]})\log\beta - (1+\beta)\sum_{j=1}^{i}\lambda_j + \log\lambda_i\right] \\
&= \sum_{i=1}^{n}(1-\Delta_{[i]})\log\beta - (1+\beta)\sum_{i=1}^{n}\sum_{j=1}^{i}\lambda_j + \sum_{i=1}^{n}\log\lambda_i \\
&= \sum_{i=1}^{n}(1-\Delta_i)\log\beta - (1+\beta)\sum_{i=1}^{n}(n-i+1)\lambda_i + \sum_{i=1}^{n}\log\lambda_i.
\end{aligned}
$$

Taking the derivative of logL over β and λ_i's, respectively, we then have

$$\beta = \frac{n - \sum_{i=1}^{n} \Delta_i}{\sum_{i=1}^{n} (n-i+1)\lambda_i}, \quad \lambda_i = \frac{1}{(1+\beta)(n-i+1)},$$

which gives $\hat{\beta}_{\mathrm{NA}} = (1-\bar{\Delta})/\bar{\Delta} = \hat{\beta}_{\mathrm{ACL}}$, and $\hat{\lambda}_i = \bar{\Delta}/(n-i+1)$, where $\bar{\Delta} = (1/n)\sum_{i=1}^{n} \Delta_i$. Therefore, the survival function can be estimated by

equation(17)

$$\hat{\bar{F}}_{\mathrm{NA}}(t) = \prod_{\{i|Y_i \leqslant t\}} (1-\hat{\lambda}_i) = \prod_{\{i|Y_i \leqslant t\}} \left(1 - \frac{\bar{\Delta}}{n-i+1}\right), \quad t \geqslant 0.$$

It is worthwhile to notice that under general random censorship model, KM estimator and NA estimator are the same. In Section 4, we will prove ACL, GMLE, GMLE1, and NA estimators are asymptotically equivalent. The introduction of GMLE1 helps us to establish the asymptotic equivalence. See Theorem 3 in Section 4.

3. Small sample properties of estimators

Now we study the small sample properties of estimators, $\hat{\bar{F}}_{\mathrm{KM}}(t)$, $\hat{\bar{F}}_{\mathrm{ACL}}(t)$, $\hat{\bar{F}}_{\mathrm{GMLE}}(t)$, and $\hat{\bar{F}}_{\mathrm{GMLE1}}(t)$ for given distribution functions *F* and *G*, which we assume are continuous and have proportional hazards.

Under the assumption of Koziol–Green model, Chen et al. (1982) calculated the *m* th moment (m>0) of the Kaplan–Meier estimator (1) (with slight modification since $\hat{\bar{F}}_{\mathrm{KM}}(t)$ could be an improper survival function here) as follows.

$$E(\hat{\bar{F}}_{\mathrm{KM}}(t))^m = \sum_{j=0}^{n} \binom{n}{j} (\bar{H}(t))^{n-j} (H(t))^j \prod_{i=1}^{j} \left(p\left(\frac{n-i}{n-i+1}\right)^m + (1-p)\right),$$

where p=P(Δ=1)=1/(1+β) and $\bar{H}(t) = 1 - H(t)$. The same idea can be applied to calculate the*m*th moment of the ACL estimator and the GMLE estimators. It is easy to verify (Cheng and Chang, 1985) that

$$E(\hat{\bar{F}}_{\mathrm{ACL}}(t))^m = \sum_{j=0}^{n} \binom{n}{j} (\bar{H}(t))^j (H(t))^{n-j} (p(j/n)^{m/n} + (1-p))^n.$$

For our estimators, we have the following result.

Proposition 1.

equation(18)

$$E(\hat{\hat{F}}_{\mathrm{GMLE}}(t))^m = \sum_{j=1}^{n} \binom{n}{j} (\bar{H}(t))^{n-j} (H(t))^j \times \sum \prod_{i=1}^{j} \Big[(\theta_i(\delta_1, \delta_2, \ldots, \delta_n))^{m(1/\hat{\beta}_{\mathrm{GMLE}}(\delta_1,\ldots,\delta_n))^{1-\delta_i}} \times p^{\sum_{i=1}^{n} \delta_i} (1-p)^{n-\sum_{i=1}^{n} \delta_i} \Big],$$

where the summation $\sum$ *is taken over all* $(\delta_1,\ldots,\delta_n)$ *with* $\delta_i=0$ *or* $1,\ i=1,\ldots,n$ *and* $\theta_i(\delta_1, \delta_2, \ldots, \delta_n) = \left(\frac{1-\delta_i+(n-i)(1+\hat{\beta}_{\mathrm{GMLE}}(\delta_1,\ldots,\delta_n))}{(n-i)(1+\hat{\beta}_{\mathrm{GMLE}}(\delta_1,\ldots,\delta_n))+1+\hat{\beta}_{\mathrm{GMLE}}(\delta_1,\ldots,\delta_n)(1-\delta_i)} \right)$*. Similarly* , $E(\hat{\hat{F}}_{\mathrm{GMLE1}}(t))^m$ *is given by* (18) *above with* $\hat{\beta}_{\mathrm{GMLE}}$ *replaced by* $\hat{\beta}_{\mathrm{ACL}}$*. Finally,*

equation(19)

$$E(\hat{\hat{F}}_{\mathrm{NA}}(t))^m = \sum_{j=0}^{n} \binom{n}{j} (\bar{H}(t))^{n-j} (H(t))^j \sum_{k=0}^{n} \binom{n}{k} \times \prod_{i=1}^{j} \left(1 - \frac{k}{n(n-i+1)}\right)^m p^k (1-p)^{n-k}.$$

Proof.

Under the Koziol‑Green model, $\underline{Y} = (Y_1, \ldots, Y_n)$ is independent of $(\Delta_1,\ldots,\Delta_n)$. Therefore, letting R_i denote the rank of Y_i in the joint ranking of $Y_1,\ldots,Y_n$, we have, for $t>0$,

$$
\begin{aligned}
E(\hat{\bar{F}}_{\mathrm{GMLE}}(t))^m &= E\left[E[\hat{\bar{F}}^m_{\mathrm{GMLE}}(t)]|\underline{Y}\right] \\
&= \left[E\left[\prod_{i=1}^{n}\left(\frac{1-\Delta_i+(n-R_i)(1+\hat{\beta}_{\mathrm{GMLE}})}{(n-R_i)(1+\hat{\beta}_{\mathrm{GMLE}})+1+\hat{\beta}_{\mathrm{GMLE}}(1-\Delta_i)}\right)^{m(1/\hat{\beta}_{\mathrm{GMLE}})^{1-\Delta_i} I(Y_i\leqslant t)}\middle|\underline{Y}\right]\right] \\
&= E\left[\sum_{(\delta_1,\ldots,\delta_n),\delta_i=0,\text{ or } 1}\prod_{i=1}^{n}\left(\frac{1-\delta_i+(n-R_i)(1+\hat{\beta}_{\mathrm{GMLE}})}{(n-R_i)(1+\hat{\beta}_{\mathrm{GMLE}})+1+\hat{\beta}_{\mathrm{GMLE}}(1-\delta_i)}\right)^{m(1/\hat{\beta}_{\mathrm{GMLE}})^{1-\delta_i} I(Y_i\leqslant t)}\right. \\
&\quad \left.\times\; p^{\sum_{i=1}^{n}\delta_i}(1-p)^{n-\sum_{i=1}^{n}\delta_i}\right] \\
&= E\left[\sum_{(\delta_1,\ldots,\delta_n),\delta_i=0,\text{ or } 1}\prod_{i=1}^{nK_n(t)}(\theta_i(\delta_1,\delta_2,\ldots,\delta_n))^{m(1/\hat{\beta}_{\mathrm{GMLE}})^{1-\delta_i}}\right. \\
&\quad \left.\times\; p^{\sum_{i=1}^{n}\delta_i}(1-p)^{n-\sum_{i=1}^{n}\delta_i}\right],
\end{aligned}
$$

where $\hat{\beta}_{\mathrm{GMLE}} = \hat{\beta}_{\mathrm{GMLE}}(\delta_1,\ldots,\delta_n)$ or $=\hat{\beta}_{\mathrm{GMLE}}(\Delta_1,\ldots,\Delta_n)$ as appropriate, from (14). Therefore, the result follows from the fact that $nK_n(t)=\sum_{j=1}^{n} I(Y_j\leqslant t)$ is a binomial random variable with parameters n and H(t).

For the m th moment of NA estimator, we have

$$
\begin{aligned}
E(\hat{\bar{F}}_{\mathrm{NA}}(t))^m &= E\left[E\left[\hat{\bar{F}}^m_{\mathrm{NA}}(t)\right]|\underline{Y}\right] \\
&= E\left[E\left[\prod_{i=1}^{n}\left(1-\frac{\bar{\Delta}}{n-R_i+1}\right)^{mI(Y_i\leqslant t)}\middle|\underline{Y}\right]\right] \\
&= E\left[\sum_{k=0}^{n}\binom{n}{k}\prod_{i=1}^{n}\left(1-\frac{k}{n(n-R_i+1)}\right)^{mI(Y_i\leqslant t)} p^k(1-p)^{n-k}\right] \\
&= E\left[\sum_{k=0}^{n}\binom{n}{k}\, p^k(1-p)^{n-k}\prod_{i=1}^{nK_n(t)}\left(1-\frac{k}{n(n-i+1)}\right)^m\right] \\
&= \sum_{j=0}^{n}\binom{n}{j}(\bar{H}(t))^{n-j}(H(t))^j\sum_{k=0}^{n}\binom{n}{k}\prod_{i=1}^{j}\left(1-\frac{k}{n(n-i+1)}\right)^m \\
&\quad \times\; p^k(1-p)^{n-k},
\end{aligned}
$$

which is (19). □

We now focus on comparing small sample performances of the five estimators. We look at only the case when T and C have exponential distributions. As has been pointed out by Chen et al. (1982, p. 144), other cases of Koziol–Green models can be brought into the framework of exponential Koziol–Green model by appropriate adjustment in calculations. We take T to have the standard exponential distribution and C to have exponential (β_0) distribution, i.e., the survival function of C is given by

$$\bar{G}(t) = \exp\{-\beta_0 t\}, \quad t \geqslant 0.$$

We have calculated biases and mean squared errors of the estimators for a number of choices of$t=0.5(0.5)2.0$,$\beta_0=0.5(0.5)2.0$ and sample sizes $n=10(5)30$. We present the results in the form of Figs. 1 and 2 for the cases $\beta_0=1$ and $n=20$ with $t\in[0,2.5]$. The results for all other cases are similar.

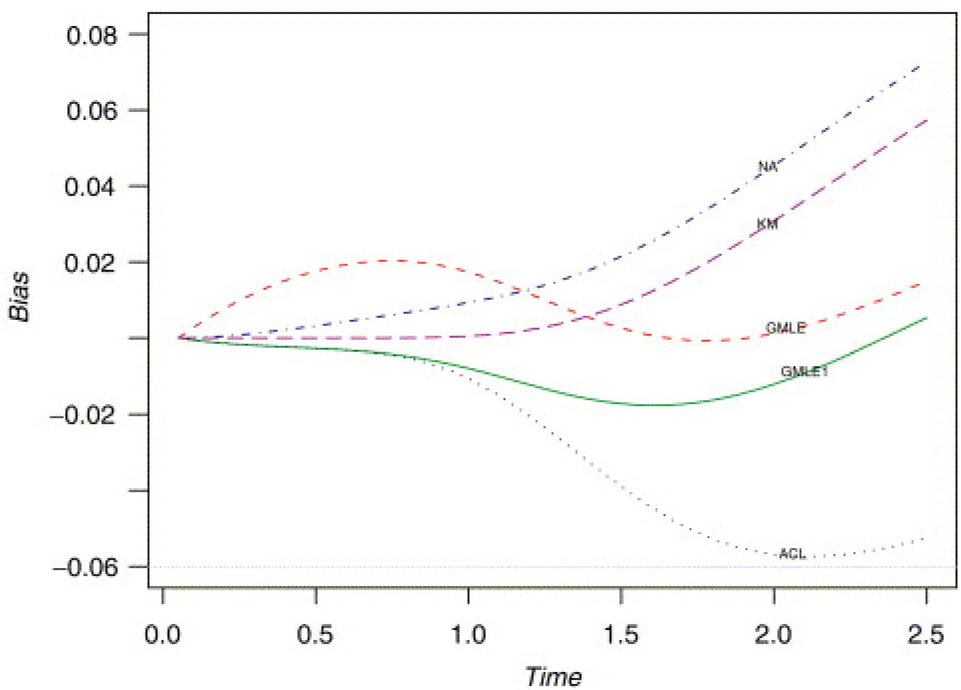

Fig. 1. Bias curves for GMLE, GMLE1, ACL, NA, and KM estimators, when $\beta_0=1$ and $n=20$.

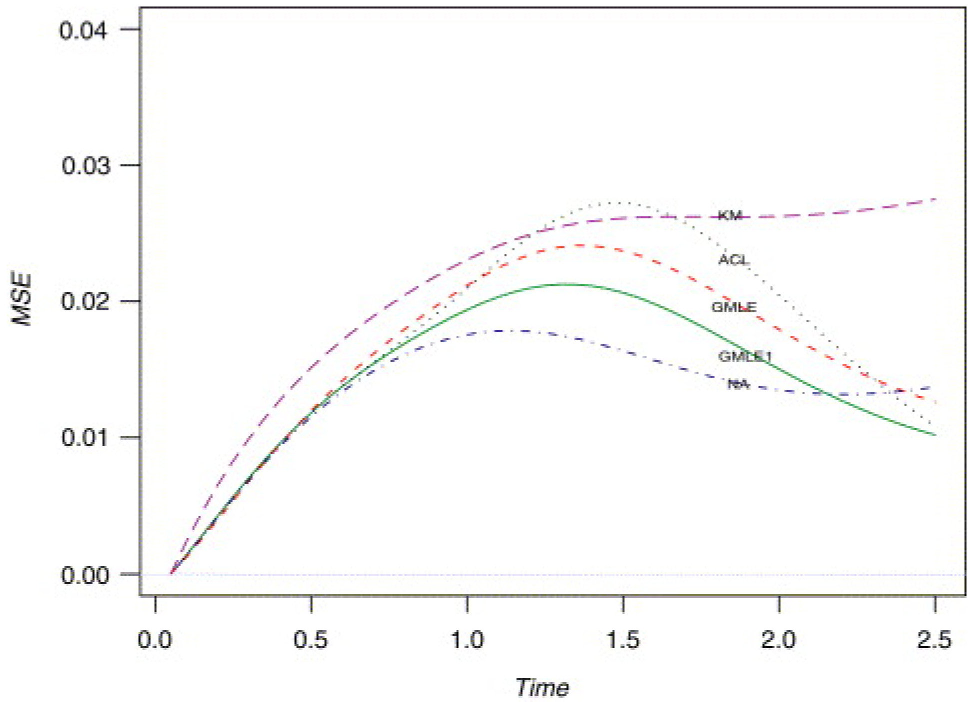

Fig. 2. MSE curves for GMLE, GMLE1, ACL, NA, and KM estimators, when β_0=1 and n=20.

From Figs. 1 and 2, we note the following features. For a wide range of t values, NA estimator is the best with respect to mean square error. The next best estimator is GMLE1 followed by GMLE. When biases are considered, GMLE and GMLE1 remain closer to the axis Bias=0 for a wide range of t values than other estimators.

Finally we calculate the likelihoods for GMLE, GMLE1, NA, and ACL estimators using the data example in Hollander et al. (2001) given below.

y	δ
4	0
6	1
8	0
9	1

The estimates and likelihoods for GMLE, GMLE1, NA, and ACL are given in Table 1. The calculation of the estimates and likelihoods of GMLE and ACL is given by Hollander et al. (2001). The same idea can be applied in the calculation of others. From Table 1, one can notice that GMLE gives the largest likelihood value out of four estimates considered, as expected. In addition, if $\bar{\Delta} \neq 1$, NA estimator is always an improper survival function even if the last observation is a failure.

Table 1. ACL, GMLEs, and NA estimates of survival function with their likelihoods

Time	ACL estimate	GMLE estimate	GMLE1 estimate	NA estimate
$0\leqslant t<4$	1	1	1	1
$4\leqslant t<6$	0.866	0.924	0.875	0.875
$6\leqslant t<8$	0.707	0.806	0.7	0.73
$8\leqslant t<9$	0.5	0.678	0.525	0.548
$t\geqslant 9$	0	0	0	0.274
Likelihood	0.00041	0.00057	0.00042	0.00021

4. Asymptotic equivalence of estimators

In this section, we focus on the large sample properties of estimators $\hat{\bar{F}}_{\mathrm{GMLE}}(t)$, $\hat{\bar{F}}_{\mathrm{GMLE1}}(t)$, $\hat{\bar{F}}_{\mathrm{NA}}(t)$, and $\hat{\bar{F}}_{\mathrm{ACL}}(t)$. Our study will show the asymptotic equivalence of $\hat{\bar{F}}_{\mathrm{GMLE}}(t)$, $\hat{\bar{F}}_{\mathrm{GMLE1}}(t)$, $\hat{\bar{F}}_{\mathrm{NA}}(t)$, and $\hat{\bar{F}}_{\mathrm{ACL}}(t)$. Let F(t) be a fixed continuous distribution function. In addition, we assume $0<\beta_0<\infty$. Let$p_0=1/(1+\beta_0)$, then $0<p_0<1$. The following facts are trivial, but they are important in the proofs of our theorems.

Lemma 1.

1. $0<-\log(1-(x+1)^{-1})-(x+1)^{-1}<(x(x+1))^{-1}$ *for* $x>0$. (*See Breslow and Crowley* (1974), *p*. 445)

2. $\sum_{i=1}^{n} 1/i = \mathrm{O}(\log n)$.

4.1. Asymptotic equivalence of $\hat{p}_{\mathrm{GMLE}}$ and $\hat{p}_{\mathrm{ACL}}$

Instead of examining asymptotic properties of $\hat{\beta}_{\mathrm{GMLE}}$, we will work with asymptotic properties of $\hat{p}_{\mathrm{GMLE}} = 1/(1+\hat{\beta}_{\mathrm{GMLE}})$. From (13) and following the discussion at the end of Section 2, the maximized loglikelihood of $p=1/(1+\beta)$ is given by

$$\log L_{\mathrm{GMLE}}(p) = \sum_{i=1}^{n} \Delta_i \left[\log\frac{p}{n-i+p} + \frac{n-i}{p}\log\left(1-\frac{p}{n-i+p}\right)\right] + \sum_{i=1}^{n}(1-\Delta_i)\left[\frac{n-i+p}{1-p}\log\left(1-\frac{1-p}{n-i+1}\right) + \log\frac{1-p}{n-i+1}\right].$$

We want to keep the subscript GMLE for L to indicate the underlying method of estimation. Therefore, the estimator $\hat{p}_{\mathrm{GMLE}}$ is the solution of the following estimating equation

$$U_{\mathrm{GMLE}}(p) = \sum_{i=1}^{n} \left[-\Delta_i \frac{n-i}{p^2} \log\left(1 - \frac{p}{n-i+p}\right) + (1-\Delta_i)\frac{n-i+1}{(1-p)^2} \log\left(1 - \frac{1-p}{n-i+1}\right)\right] = 0.$$

Let

$$V_{\mathrm{GMLE}}(p) = \frac{\mathrm{d}U_{\mathrm{GMLE}}(p)}{\mathrm{d}p} = \sum_{i=1}^{n} \left[\Delta_i \frac{n-i}{p^2} \left[\frac{2}{p} \log\left(1 - \frac{p}{n-i+p}\right) + \frac{1}{n-i+p}\right]\right] + \sum_{i=1}^{n} \left[(1-\Delta_i) \frac{n-i+1}{(1-p)^2} \left[\frac{2}{1-p} \times \log\left(1 - \frac{1-p}{n-i+1}\right) + \frac{1}{n-i+p}\right]\right].$$

Recall $\hat{p}_{\mathrm{ACL}} = \sum_{i=1}^{n} \Delta_i / n$. Note that $\hat{p}_{\mathrm{ACL}}$ is the solution of the following estimating equation

$$U_{\mathrm{ACL}}(p) = \sum_{i=1}^{n} \left[\frac{\Delta_i}{p} - \frac{1-\Delta_i}{1-p} \right] = 0.$$

Let

$$V_{\mathrm{ACL}}(p) = \frac{\mathrm{d}U_{\mathrm{ACL}}(p)}{\mathrm{d}p} = - \sum_{i=1}^{n} \left[\frac{\Delta_i}{p^2} + \frac{1-\Delta_i}{(1-p)^2} \right].$$

First, we present the following lemma.

Lemma 2.

1. $n^{-1/2}|U_{GMLE}(p)-U_{ACL}(p)| \rightarrow_{a.s.} 0$.

2. $n^{-1}|V_{GMLE}(p)-V_{ACL}(p)| \rightarrow_{a.s.} 0$.

Proof.

From Lemma 1, we can see that

$$n^{-1/2}U_{\mathrm{GMLE}}(p)=n^{-1/2}\sum_{i=1}^{n}\left(-\Delta_i\frac{n-i}{p^2}\left[-\frac{p}{n-i+p}+\mathrm{O}((n-i+p)^{-2})\right]\right)$$

$$+n^{-1/2}\sum_{i=1}^{n}\left((1-\Delta_i)\frac{n-i+1}{(1-p)^2}\left[-\frac{1-p}{n-i+1}\right.\right.$$

$$\left.\left.+\ \mathrm{O}((n-i+1)^{-2})\right]\right)$$

$$\cong n^{-1/2}\sum_{i=1}^{n}\left[\Delta_i\frac{1}{p}\frac{n-i}{n-i+p}-(1-\Delta_i)\frac{1}{1-p}\right]$$

$$+\mathrm{O}(n^{-1/2}\log n)$$

$$\cong n^{-1/2}U_{\mathrm{ACL}}(p)+\mathrm{O}(n^{-1/2}\log n).$$

Here asymptotic equivalence ≅ is in the almost sure sense. The second part can be proved analogously. □

Now we present the main theorems in this subsection.

Theorem 1 Consistency of $\hat{p}_{\mathrm{GMLE}}$.

$\hat{p}_{\mathrm{GMLE}}\rightarrow_{\mathrm{a.s.}}p_0$, *where* p_0 *is the true value.*

Proof.

For $0<p,p_0<1$, one can notice that $-n^{-1}V_{\mathrm{ACL}}(p)\rightarrow_{\mathrm{a.s.}}\frac{p_0}{p^2}+\frac{1-p_0}{(1-p)^2}>0$ by the Strong Law of Large Numbers. Combining with Lemma 2, we have

equation(20)

$$-n^{-1}V_{\mathrm{GMLE}}(p)\rightarrow_{\mathrm{a.s.}}\frac{p_0}{p^2}+\frac{1-p_0}{(1-p)^2}>0.$$

Therefore, $\frac{1}{n}\log L_{\mathrm{GMLE}}(p)$ is a sequence of strictly concave functions, and thus, $\hat{p}_{\mathrm{GMLE}}$, the maxima of the loglikelihood $\log L_{\mathrm{GMLE}}$, exists and is unique. To prove consistency, we exploit the technique used in Andersen and Gill (1982). Consider

$$
\begin{aligned}
X_n(p) &\equiv \frac{1}{n}(\log L_{\mathrm{GMLE}}(p) - \log L_{\mathrm{GMLE}}(p_0)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i \log\frac{p}{p_0} + (1-\Delta_i)\log\frac{1-p}{1-p_0}\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\Delta_i\left[\log\left(1-\frac{p-p_0}{n-i+p}\right) + \frac{n-i}{p}\log\left(1-\frac{p}{n-i+p}\right)\right. \\
&\quad \left. - \frac{n-i}{p_0}\log\left(1-\frac{p_0}{n-i+p_0}\right)\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\left[\frac{n-i+p}{1-p}\log\left(1-\frac{1-p}{n-i+1}\right)\right. \\
&\quad \left. - \frac{n-i+p_0}{1-p_0}\log\left(1-\frac{1-p_0}{n-i+1}\right)\right] \\
&\cong \frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i \log\frac{p}{p_0} + (1-\Delta_i)\log\frac{1-p}{1-p_0}\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\Delta_i\left[-\frac{n-i}{n-i+p} + \frac{n-i}{n-i+p_0}\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\left[-\frac{n-i+p}{n-i+1} + \frac{n-i+p_0}{n-i+1}\right] + O(n^{-1}\log(n)) \\
&\cong p_0(\log p - \log p_0) + (1-p_0)[\log(1-p) - \log(1-p_0)] \equiv f(p),
\end{aligned}
$$

say. Here asymptotic equivalence $\cong$ is in the almost sure sense, and is obtained by applying Lemma 1. Since $f'(p)|_{p=p0}=0$ and $f''(p)|_{p=p0}=-1/p_0-1/(1-p_0)<0$ for $0<p_0<1$. Therefore, $X_n(p)$ converges almost surely to a concave function of p with a unique maximum at $p=p_0$. In addition, since $\hat{p}_{\mathrm{GMLE}}$ maximizes the random concave function $X_n(p)$, it follows that $\hat{p}_{\mathrm{GMLE}} \rightarrow_{\mathrm{a.s.}} p_0$. This result follows from Rockafellar (1970), Theorem 10.8. See also Appendix II of Andersen and Gill (1982). □

Theorem 2 Asymptotic equivalence of estimators of *p*.

$n^{1/2}(\hat{p}_{\mathrm{GMLE}} - \hat{p}_{\mathrm{ACL}}) \rightarrow_{\mathrm{a.s.}} 0.$

Proof.

With Lemma 2 above, consistency of $\hat{p}_{\mathrm{GMLE}}$ and $\hat{p}_{\mathrm{ACL}}$, and (20), the "almost sure" version of Theorem 2 follows along the lines of that of Theorem 1 in Bailey (1984). □

4.2. Asymptotic equivalence of $\hat{\bar{F}}_{\mathrm{GMLE}}(t)$ and $\hat{\bar{F}}_{\mathrm{GMLE1}}(t)$

To prove the asymptotic equivalence, let $\tau = \sup\{t, \bar{F}(t) > 0\}$. We focus on the time interval $[0,T_0]$, where $T_0<\tau$ throughout this subsection. Note that the assumption here implies $\frac{1}{n}\sum_{i=1}^{n} I(T_i > T_0) \geqslant M > 0$ almost surely for sufficiently large n and for some constant M. Here $T_1, T_2, \ldots$, are independent copies of T.

From Theorem 2, we immediately have the following result.

Theorem 3.

The estimators $\hat{\bar{F}}_{\mathrm{GMLE}}(t)$ *and* $\hat{\bar{F}}_{\mathrm{GMLE1}}(t)$ *in* (15) *and* (16), *respectively, are asymptotically equivalent, i.e.*,

$$\sup_{0 \leqslant t \leqslant T_0} n^{1/2} |\hat{\bar{F}}_{\mathrm{GMLE1}}(t) - \hat{\bar{F}}_{\mathrm{GMLE}}(t)| \rightarrow_{\mathrm{a.s.}} 0.$$

Proof.

Note that it is sufficient to prove

$$\sup_{0 \leqslant t \leqslant T_0} n^{1/2} |\hat{\Lambda}_{\mathrm{GMLE1}}(t) - \hat{\Lambda}_{\mathrm{GMLE}}(t)| \rightarrow_{\mathrm{a.s.}} 0,$$

where $\hat{\Lambda}_{\mathrm{GMLE1}}(t) = -\log \hat{\bar{F}}_{\mathrm{GMLE1}}(t)$ is the estimator of the cumulative hazard function $\Lambda(t) = -\log \bar{F}(t)$. Similar notation applies to $\hat{\Lambda}_{\mathrm{GMLE}}(t)$.

Recall from Section 1, $nK_n(t) = \sum_{i=1}^{n} I(Y_i \leqslant t)$ denotes the number of observations of Y 's up to time t. For notational simplicity, we suppress the subscript ACL from $\hat{p}_{\mathrm{ACL}}$ throughout the proof, i.e., $\hat{p} = \hat{p}_{\mathrm{ACL}}$. Similarly we let $\tilde{p} = \hat{p}_{\mathrm{GMLE}}$. First notice that, from (16),

$$\hat{\bar{F}}_{\text{GMLE1}}(t) = \prod_{i=1}^{nK_n(t)} \left(1 - \frac{\hat{p}}{n-i+\hat{p}}\right)^{\Delta_i} \times \prod_{i=1}^{nK_n(t)} \left(1 - \frac{1-\hat{p}}{n-i+1}\right)^{(1-\Delta_i)\hat{p}/(1-\hat{p})}. \text{ Therefore,}$$

$$\hat{\Lambda}_{\text{GMLE1}}(t) = -\sum_{i=1}^{nK_n(t)} \Delta_i \log\left(1 - \frac{\hat{p}}{n-i+\hat{p}}\right) - \frac{\hat{p}}{(1-\hat{p})} \sum_{i=1}^{nK_n(t)} (1-\Delta_i) \log\left(1 - \frac{1-\hat{p}}{n-i+1}\right), \text{ and}$$

$$\hat{\Lambda}_{\text{GMLE}}(t) = -\sum_{i=1}^{nK_n(t)} \Delta_i \log\left(1 - \frac{\tilde{p}}{n-i+\tilde{p}}\right) - \frac{\tilde{p}}{(1-\tilde{p})} \sum_{i=1}^{nK_n(t)} (1-\Delta_i) \log\left(1 - \frac{1-\tilde{p}}{n-i+1}\right).$$

From Lemma 1, we have, for $t \leqslant T_0$,

$$\sum_{i=1}^{nK_n(t)} \Delta_i \log\left(1 - \frac{\hat{p}}{n-i+\hat{p}}\right) \cong -\sum_{i=1}^{nK_n(t)} \frac{\Delta_i \hat{p}}{n-i+\hat{p}} + \mathrm{O}\left(\sum_{i=1}^{nK_n(t)} \frac{\Delta_i}{(n-i+p)^2}\right) \cong -\sum_{i=1}^{nK_n(t)} \frac{\Delta_i \hat{p}}{n-i+\hat{p}} + \mathrm{O}(n^{-1}).$$

The asymptotic equivalence $\cong$ is in the almost sure sense, and the second $\cong$ above uses the fact $\sum_{i=1}^{nK_n(t)} \frac{\Delta_i}{(n-i+p)^2} = \mathrm{O}(n^{-1})$ almost surely for all $t \leqslant T_0$. Similarly,

$$\sum_{i=1}^{nK_n(t)} (1-\Delta_i) \log\left(1 - \frac{1-\hat{p}}{n-i+1}\right) \cong -\sum_{i=1}^{nK_n(t)} (1-\Delta_i) \frac{1-\hat{p}}{n-i+1} + \mathrm{O}(n^{-1}).$$

Hence,

$$\hat{\Lambda}_{\mathrm{GMLE1}}(t) \cong \sum_{i=1}^{nK_n(t)} \left[\Delta_i \frac{\hat{p}}{n-i+\hat{p}} + (1-\Delta_i)\frac{\hat{p}}{n-i+1} \right] + \mathrm{O}(n^{-1})$$

$$\cong \hat{p} \sum_{i=1}^{nK_n(t)} \left[\Delta_i \frac{1}{n-i+1} + (1-\Delta_i)\frac{1}{n-i+1} \right] + \mathrm{O}(n^{-1})$$

$$= \hat{p} \sum_{i=1}^{nK_n(t)} \frac{1}{n-i+1} + \mathrm{O}(n^{-1}),$$

i.e., the cumulative hazard function estimator based on GMLE method is asymptotically equivalent almost surely to the product of $\hat{p}$ and the usual Nelson estimator $\sum_{i=1}^{nK_n(t)} \frac{1}{n-i+1}$ for the cumulative hazard function (Nelson, 1969 and Nelson, 1972) in the case of uncensored observations. Similarly, we have

$$\hat{\Lambda}_{\mathrm{GMLE}}(t) \cong \tilde{p} \sum_{i=1}^{nK_n(t)} \frac{1}{n-i+1} + \mathrm{O}(n^{-1}).$$

Furthermore, simple calculation shows that $\sum_{i=1}^{nK_n(t)} \frac{1}{n-i+1} = \mathbf{const.} + \mathrm{O}(n^{-1})$ since $K_n(t) \rightarrow_{a.s.} P(Y \leqslant t) \leqslant P(Y \leqslant T_0) < 1$. By Theorem 2, we have the result. □

4.3. Asymptotic equivalence of $\hat{\bar{F}}_{\mathrm{GMLE}}(t)$ and $\hat{\bar{F}}_{\mathrm{ACL}}(t)$

Theorem 4 Asymptotic equivalence of estimators of $\bar{F}(t)$.

$\hat{\bar{F}}_{\mathrm{GMLE1}}(t)$ *and* $\hat{\bar{F}}_{\mathrm{ACL}}(t)$ *are asymptotically equivalent , i.e. ,*

$$\sup_{0 \leqslant t \leqslant T_0} n^{1/2} |\hat{\bar{F}}_{\mathrm{GMLE1}}(t) - \hat{\bar{F}}_{\mathrm{ACL}}(t)| \rightarrow_{\mathrm{a.s.}} 0.$$

Consequently , $\hat{\bar{F}}_{\mathrm{GMLE}}(t)$ *and* $\hat{\bar{F}}_{\mathrm{ACL}}(t)$ *are asymptotically equivalent, i.e.,*

$$\sup_{0 \leqslant t \leqslant T_0} n^{1/2} |\hat{\bar{F}}_{\mathrm{GMLE}}(t) - \hat{\bar{F}}_{\mathrm{ACL}}(t)| \rightarrow_{\mathrm{a.s.}} 0.$$

Proof.

From Theorem 3, it is sufficient to prove

$$\sup_{0 \leqslant t \leqslant T_0} n^{1/2} |\hat{\Lambda}_{\mathrm{GMLE1}}(t) - \hat{\Lambda}_{\mathrm{ACL}}(t)| \rightarrow_{\mathrm{a.s.}} 0,$$

where $\hat{\Lambda}_{\mathrm{ACL}}(t) = -\log\ \hat{\bar{F}}_{\mathrm{ACL}}(t)$ is the estimator of the cumulative hazard function $\Lambda(t) = -\log\ \bar{F}(t)$.

Recall that $nK_n(t) = \sum_{i=1}^{n} I(Y_i \leqslant t)$ denotes the number of observations of Y 's up to time t . For notational simplicity, we write $\hat{p} = \hat{p}_{\mathrm{ACL}}$. First notice that, from (3),

$$\hat{\Lambda}_{\mathrm{ACL}}(t) = -\hat{p} \cdot \log(1 - K_n(t)).$$

From the proof of Theorem 3, we have, for $t \leqslant T_0$,

$$\hat{\Lambda}_{\mathrm{GMLE1}}(t) \cong \hat{p} \sum_{i=1}^{nK_n(t)} \frac{1}{n - i + 1} + \mathrm{O}(n^{-1}).$$

Therefore, our theorem follows. □

From Theorems 2 and 3 of Cheng and Lin (1987) and our Theorem 6 above, we immediately have the following asymptotic results for GMLEs.

Theorem 5.

1 (*Law of the iterated logarithm*). *With probability one*,

$$\sup_{0 \leqslant t \leqslant T_0} |\hat{\bar{F}}_{\mathrm{GMLE}}(t) - \bar{F}(t)| = \mathrm{O}(n^{-1/2}(\log\ \log n)^{1/2}).$$

2 (*Weak convergence*). *The sequence of random processes* $\{n^{1/2}(\hat{\bar{F}}_{\mathrm{GMLE}}(t) - \bar{F}(t)), 0 \leqslant t \leqslant T_0\}$ *converges weakly to the Gaussian process* $W(t)$ *with mean* $\mathbf{E}W(t)=0$, *and for* $0 \leqslant s \leqslant t \leqslant T_0$,

$$\begin{aligned}\mathrm{Cov}\,(W(s), W(t)) = {} & (1+\beta_0)^{-2}\bar{F}(s)^{-\beta_0}\bar{F}(t)(1 - \bar{F}^{\beta_0+1}(s)) \\ & + \beta_0\bar{F}(t)\log(\bar{F}(t))\bar{F}(s)\log(\bar{F}(s)).\end{aligned}$$

4.4. Asymptotic equivalence of $\hat{\bar{F}}_{\mathrm{GMLE}}(t)$ and $\hat{\bar{F}}_{\mathrm{NA}}(t)$

Theorem 6 Asymptotic equivalence of estimators of $\bar{F}(t)$.

$\hat{\bar{F}}_{\mathrm{GMLE}}(t)$ *and* $\hat{\bar{F}}_{\mathrm{NA}}(t)$ *are asymptotically equivalent, i.e.*,

$$\sup_{0 \leqslant t \leqslant T_0} n^{1/2}|\hat{\bar{F}}_{\mathrm{GMLE}}(t) - \hat{\bar{F}}_{\mathrm{NA}}(t)| \to_{\mathrm{a.s.}} 0.$$

Proof.

Notice that

$$\hat{\Lambda}_{\mathrm{NA}}(t) = -\log \hat{\bar{F}}_{\mathrm{NA}}(t) = -\sum_{i=1}^{nK_n(t)} \log\left(1 - \frac{\bar{\Delta}}{n-i+1}\right),$$
$$\cong \sum_{i=1}^{nK_n(t)} \frac{\bar{\Delta}}{n-i+1} \cong p \sum_{i=1}^{nK_n(t)} \frac{1}{n-i+1}$$
$$\cong \hat{\Lambda}_{\mathrm{GMLE}}(t) + \mathrm{O}(n^{-1}).$$

Our result follows. □

Theorem 7.

1 (*Law of the iterated logarithm*). *With probability one,*

$$\sup_{0 \leqslant t \leqslant T_0} |\hat{\bar{F}}_{\mathrm{NA}}(t) - \bar{F}(t)| = \mathrm{O}(n^{-1/2}(\log \log n)^{1/2}).$$

2 (*Weak convergence*). *The sequence of random processes* $\{n^{1/2}(\hat{\bar{F}}_{\mathrm{NA}}(t) - \bar{F}(t)), 0 \leqslant t \leqslant T_0\}$ *converges weakly to the Gaussian process* $W(t)$ *with mean* $\mathbf{E}W(t)=0$, *and for* $0 \leqslant s \leqslant t \leqslant T_0$,

$$\mathrm{Cov}(W(s), W(t)) = (1+\beta_0)^{-2}\bar{F}(s)^{-\beta_0}\bar{F}(t)(1-\bar{F}^{\beta_0+1}(s)) + \beta_0\bar{F}(t)\log(\bar{F}(t))\bar{F}(s)\log(\bar{F}(s)). \quad \square$$

Acknowledgements

The authors sincerely appreciate the reviews provided by the two anonymous referees. The inclusion of the Nelson–Aalen type estimator of the survival function in this paper owes to one of the referees. The suggestions and comments of the referees have strengthened the paper significantly.

References

Abdushukurov, A., 1984. On some estimates of the distribution function under random censorship. Conference of Young Scientists, VINITI No. 8756-V, Mathematical Institute and Academy of Sciences, Uzbek SSR, Tashkent.

Allen, W.R., 1963. A note on conditional probability of failure when hazards are proportional. Oper. Res. 11, 658–659.

Andersen, P.K., Gill, R.D., 1982. Cox's regression model for counting processes: a large sample study. Ann. Statist. 10, 1100–1120.

Bailey, K., 1984. Asymptotic equivalence between the Cox estimator and the general ML estimators of regression and survival parameters in the Cox model. Ann. Statist. 12, 730–736.

Breslow, N., Crowley, J., 1974. A large sample study of the life table and product limit estimates under random censorship. Ann. Statist. 2, 437–453.

Chen, Y.Y., Hollander, M., Langberg, N.A., 1982. Small sample results for the Kaplan–Meier estimator. J. Amer. Statist. Assoc. 77, 141–144.

Cheng, P.E., Chang, Y.C., 1985. PLE versus MLE for survivor functions under a simple proportional hazards model. J. Chinese Statist. Assoc. 23 (special issue), 57–73

Cheng, P.E., Lin, G.D., 1987. Maximum likelihood estimator of a survival function under the Koziol–Green proportional hazards model. Statist. Probab. Lett. 5, 75–80.

Ebrahimi, N., 1985. Nonparametric estimation of survival functions for incomplete observations when the lifetime distribution is proportionally related to the censoring time distribution. Comm. Statist. Theory Methods 14, 2887–2898.

Efron, B., 1967. The two sample problem with censored data. Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 4, pp. 831–853.

Fleming, T.R., Harrington, D.P., 1991. Counting Processes and Survival Analysis. Wiley, New York.

Gill, R.D., 1983. Large sample behavior of the product limit estimator on the whole line. Ann. Statist. 11, 49–58.

Groeneboom, P., Wellner, J., 1992. Information bounds and nonparametric maximum likelihood estimation. DMV Seminar, vol. 19, Birkhäuser, Basel.

Henze, N., 1993. A quick omnibus test for the proportional hazards model of random censorship. Statistics 24, 253–263.

Hollander, M., Laird, G., Song, K.-S., 2001. Maximum likelihood estimation in the proportional hazards model of random censorship. Statistics 35, 245–258.

Johansen, S., 1978. The product limit estimator as maximum likelihood estimator. Scand. J. Statist. 5, 195–199.

Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457–481.

Keifer, J., Wolfowitz, J., 1956. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann. Math. Statist. 27, 887–906.

Lanius, V., Pawlitschko, J., 2001. Influence functions in the Koziol–Green model. In: Kunert, J., Trenkler, G. (Eds.), Mathematical Statistics with Applications in Biometry, Anniversary Publication in Honour of Siegfried Schach, Eul-Verlag, Köln, pp. 239–252.

Mitra, R.C., 1991. Properties of some nonparametric estimators under proportional hazards in survival analysis. Unpublished Ph.D. dissertation. Department of Statistics, North Dakota State University.

Nelson, W., 1969. Hazard plotting for incomplete failure data. J. Quality Technol. 1, 27–52.

Nelson, W., 1972. Theory and application of hazard plotting for censored failure data. Technometrics 14, 945–966.

Pawlitschko, J., 1999. A comparison of survival function estimations in the Koziol–Green model. Statistics 32, 277–291.

Rockafellar, R.T., 1970. Convex Analysis. Princeton University Press, Princeton.

de Uña-Álverez, J., González-Manteiga, W., Cadarso-Suárez, C., 1997. Bootstrap selection of the smoothing parameter in density estimation under the Koziol–Green model. IMS Lecture Notes—Monogr. Ser. 31, 385–398.

Wellner, J.A., 1985. A heavy censoring limit theorem for the product limit estimator. Ann. Statist. 13, 150–162.