A note on the generalized maximum likelihood estimator in partial Koziol‑Green model

By: Haimeng Zhang, M. Bhaskara Rao

Zhang, H. and Rao, M.B. (2006). A note on the generalized maximum likelihood estimator in partial Koziol-Green model. *Statistics and Probability Letters, 76(8),* 813 - 820. doi: 10.1016/j.spl.2005.10.032

Made available courtesy of Elsevier: http://www.dx.doi.org/10.1016/j.spl.2005.10.032

Abstract:

The proportional hazards model with partially informative censoring has been studied by Gather and Pawlitschko [1998. Metrika 48, 189‑207] and their partial Abdushukurov‑Cheng‑Lin (PACL) estimator is a nonparametric estimator of the underlying survival function of the model. Zhang and Rao [2004. Metrika 59, 125‑136] have derived the generalized maximum likelihood estimator of the underlying survival function using Kiefer‑Wolfowitz theory. In this note, we show that these two estimators are asymptotically equivalent.

Keywords: Asymptotic properties | Generalized maximum likelihood estimator | Informative censoring | PACL estimator | Partial Koziol–Green model

Article:

1. Introduction

Let X,Y, and Z be independent non-negative random variables with unknown distribution functions F, G_1, and G_2, respectively, and with respective survival functions $\overline{F}, \overline{G}_1$, and $\overline{G}_2$ with the following property:

equation(1)

$$\overline{G}_1(t) = (\overline{F}(t))^{\beta}, \quad t \geqslant 0$$

for some $\beta>0$. The entity X is called a lifetime or failure random variable and the estimation of its distribution is one of the main goals in Survival Analysis. The entity X is partially observable, i.e. we observe $U=\min\{X,Y,Z\}=X\wedge Y\wedge Z$ and

$$\Delta = \begin{cases} 1 & \text{if } X \leqslant Y \wedge Z, \\ 0 & \text{if } Y \leqslant X \wedge Z, \\ -1 & \text{if } Z \leqslant X \wedge Y. \end{cases}$$

The random variable Δ indicates which one of X,Y, and Z is the minimum. In order to avoid ambiguities in the definition of Δ, certain conventions are adopted. If $\min\{X,Y,Z\}=X$ and even if Y or Z equals to X, we take $\Delta=1$. If $\min\{X,Y,Z\}=Y$ and even if $Y=Z$, we take $\Delta=0$. The lifetime variable X could be censored by Y on the right, which is the case of informative censoring in view of (1) above, or be censored by Z on the right, which is the case of uninformative censoring. This model is called a proportional hazards model with partially informative censoring (Gather and Pawlitschko, 1998).

Gather and Pawlitschko (1998) proposed a nonparametric estimator of the survival function of X. Zhang and Rao (2004) derived the generalized maximum likelihood estimator of the survival function of X using Kiefer–Wolfowitz theory (Kiefer and Wolfowitz, 1956). In this note, we show that these two estimators have the same asymptotic distribution.

2. The estimators

Let $(U_1,\Delta_1),(U_2,\Delta_2),\ldots,(U_n,\Delta_n)$ be n independent realizations of (U,Δ), $U_{(1)}\leqslant U_{(2)}\leqslant\cdots\leqslant U_{(n)}$the ordered statistics of $U_1,U_2,\ldots,U_n$, and $\Delta_{[1]},\Delta_{[2]},\ldots,\Delta_{[n]}$ the concomitant Δ-values of the ordered U_i's. Non-parametric estimation of the distribution function of X based on the above data has been discussed by Gather and Pawlitschko (1998) and Csörgö (1998). The non-parametric estimator of $\bar{F}$, namely partial ACL or PACL for short, given by Gather and Pawlitschko (1998), is an analogue of Abdushukurov–Cheng–Lin (ACL) estimator (Abdushukurov, 1984; Cheng and Lin, 1987). It is given by

equation(2)

$$\hat{\bar{F}}_{PACL}(t) = (\hat{\bar{K}}(t))^{\hat{p}_{PACL}}, \quad t \geqslant 0,$$

where

$$\hat{\bar{K}}(t) = \prod_{i=1}^{n} \left(1 - \frac{\varepsilon_{[i]}}{n-i+1}\right)^{I[U_{(i)} \leqslant t]}, \quad t \geqslant 0$$

with I[A] being the indicator function of the event A, and

equation(3)

$$\hat{p}_{PACL} = \frac{\sum_{i=1}^{n} I[\Delta_i = 1]}{\sum_{i=1}^{n} I[\Delta_i \neq -1]},$$
$$= 0, \text{ if the denominator above is } 0$$

which is a reasonable estimator of $p = P_r(\Delta = 1 \mid \Delta \neq -1) = 1/(1+\beta)$. Here $\varepsilon_i = I[\Delta_i \neq -1]$ and $\varepsilon_{[1]},\ldots,\varepsilon_{[n]}$ denote the concomitant ε-values of the ordered U_i's. The estimator of β is given by

equation(4)

$$\hat{\beta}_{PACL} = (1 - \hat{p}_{PACL})/\hat{p}_{PACL}.$$

If one ignores the partially informative censoring by Y, the proportional hazards model with partially informative censoring can also be treated as a general random censorship model with right censoring by combining Y and Z. Consequently, the Kaplan–Meier estimator (Kaplan and Meier, 1958) can be used to estimate the survival function $\bar{F}$

equation(5)

$$\hat{\bar{F}}_{KM}(t) = \prod_{i=1}^{n} \left(1 - \frac{\eta_{[i]}}{n-i+1}\right)^{I[U_{(i)} \leqslant t]}, \quad t \geqslant 0,$$

where $\eta_i = I[\Delta_i = 1]$ with the concomitant η-values $\eta_{[1]},\ldots,\eta_{[n]}$ corresponding to the ordered U_i's.

A generalized maximum likelihood estimator (GMLE) has been derived by Zhang and Rao (2004). In principle, the GMLE involves maximizing the likelihood of the data over all distributions F of X and G_2 of Z. See Johansen (1978) and Miller (1981) for an exposition of the generalized maximum likelihood method. In this paper, we focus on the GMLE based on the situation when $0=U_{(0)}<U_{(1)}<U_{(2)}<\cdots<U_{(n)}$ are distinct. It is sufficient to consider distinct U_is in view of the fact that asymptotics of the GMLE are derived under the assumption of continuous F and G_2. For notational simplicity, we define $D_j=I[\Delta_{[j]}=1], C_j=I[\Delta_{[j]}=0]$, and $E_j=I[\Delta_{[j]}=-1]$. Obviously, $C_j+D_j+E_j=1, j=1,2,\ldots,n$.

Under this provision, the generalized likelihood of the data is given by

$$L(F, G_1, G_2) = L = \prod_{i=1}^{n} p_i^{D_i} (P_r(Y \geqslant U_{(i)}))^{D_i} (P_r(Z \geqslant U_{(i)}))^{D_i} q_i^{C_i} (P_r(X > U_{(i)}))^{C_i}$$
$$\times (P_r(Z \geqslant U_{(i)}))^{C_i} r_i^{E_i} (P_r(X > U_{(i)}) P_r(Y > U_{(i)}))^{E_i},$$

where for $i=1,2,\ldots,n$,

$$p_i = P_r(X = U_{(i)}), \quad q_i = P_r(Y = U_{(i)}), \quad r_i = P_r(Z = U_{(i)}).$$

Therefore, our problem reduces to maximizing the likelihood *L* over all distributions *F* and G_2 which must satisfy proportional hazards condition (1). Applying the transformation $a_i = p_i(\sum_{j \geqslant i} p_j)^{-1}, i = 1, 2, \ldots, n,$ the likelihood of the data, up to a constant independent of unknown parameters, is given by (see a detailed derivation in Zhang and Rao, 2004)

equation(6)

$$L = \prod_{i=1}^{n} a_i^{D_i}(1 - a_i)^{C_i + (n-i)(1+\beta) + (1+\beta)E_i}[1 - (1 - a_i)^{\beta}]^{C_i}$$

with unknown scales $0 \leqslant a_i \leqslant 1$ and $\beta > 0$. It can be shown (see Zhang and Rao, 2004), that the GMLE is given by

equation(7)

$$\hat{\bar{F}}_{\mathrm{GMLE}}(t) = \begin{cases} \prod_{j=1}^{i-1}(1 - \hat{a}_j) & \text{if } U_{(i-1)} \leqslant t < U_{(i)},\ i = 1, 2, \ldots, n, \\ \prod_{j=1}^{n}(1 - \hat{a}_j) & \text{if } t \geqslant U_{(n)} \end{cases}$$

with the convention that the empty product is equal to 1. Here $\hat{a}_i$ is given by

equation(8)

$$\hat{a}_i = \begin{cases} (1 + (n-i)(1 + \hat{\beta}_{\mathrm{GMLE}}))^{-1} & \text{if } D_i = 1,\ C_i = E_i = 0, \\ 1 - \left(\dfrac{1 + (n-i)(1 + \hat{\beta}_{\mathrm{GMLE}})}{(n-i+1)(1 + \hat{\beta}_{\mathrm{GMLE}})}\right)^{1/\hat{\beta}_{\mathrm{GMLE}}} & \text{if } C_i = 1,\ D_i = E_i = 0, \\ 0 & \text{if } E_i = 1,\ D_i = C_i = 0. \end{cases}$$

Note that the estimator is proper, i.e. $\prod_{j=1}^{n}(1 - \hat{a}_j) = 0$ only if the last observation $U_{(n)}$ is uncensored. In our asymptotic analysis, we set deliberately $\hat{\bar{F}}_{\mathrm{GMLE}}(t) = 0$ if $t \geqslant U_{(n)}$.
Plugging (8) in the likelihood *L* (6), one can see that the maximized loglikelihood in β is given by

equation(9)

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

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

equation(10)

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

It has been shown that the PACL estimator (2) is more efficient than the KM estimator (5) (Gather and Pawlitschko, 1998). Simulations based on small and large samples have suggested that the PACL estimator and the GMLE are asymptotically equivalent (Zhang and Rao, 2004). In this note, we provide a theoretical justification of this equivalence.

3. Main results

Let $F(t)$ and $G_2(t)$ be fixed continuous distribution functions and $\beta=\beta_0$ be fixed. Let $\alpha_i=P_r(\Delta=i), i=0,1$ and $p_0=1/(1+\beta_0)=\alpha_1/(\alpha_1+\alpha_0)$. We also let $\tau_1=\sup\{t,\bar{F}(t)>0\}$ and $\tau_2=\sup\{t,\bar{G}_2(t)>0\}$. We focus on the time interval $[0,T_0]$, where $T_0<\min\{\tau_1,\tau_2\}$. Our main results are given by the following theorems. The proof of Theorem 1 is given in the Appendix.

Theorem 1 Asymptotic equivalence of estimators of $\bar{F}(t)$.

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

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

From Theorems 4.3 and 4.4 of Gather and Pawlitschko (1998) and our Theorem 1 above, we immediately have the following asymptotic results for GMLE.

Theorem 2.

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

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

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* $s,t\in[0,T_0]$,

$$\mathrm{Cov}(W(s),W(t))=p_0\bar{F}(s)\bar{F}(t)\left(C(s,t)+\frac{1-p_0}{\alpha_0+\alpha_1}\log(\bar{F}(t))\log(\bar{F}(s))\right)$$

with

$$C(s,t)=\int_0^{s\wedge t}\frac{1}{(\bar{F}(u))^{1/p_0}\bar{G}_2(u)}\frac{dF(u)}{\bar{F}(u)}.$$

Remarks.

A variant of the GMLE (7), referred as GMLE1, was also considered in Zhang and Rao (2004). This estimator is given by

equation(11)

$$\hat{\bar{F}}_{GMLE1}(t)=\begin{cases}\prod_{j=1}^{i-1}(1-\hat{b}_j) & \text{if } U_{(i-1)}\leqslant t<U_{(i)},\ i=1,2,\ldots,n,\\ \prod_{j=1}^{n}(1-\hat{b}_j) & \text{if } t\geqslant U_{(n)}\end{cases}$$

with

$$\hat{b}_i=\begin{cases}(1+(n-i)(1+\hat{\beta}_{PACL}))^{-1} & \text{if } D_i=1,\ C_i=E_i=0,\\ 1-\left(\dfrac{1+(n-i)(1+\hat{\beta}_{PACL})}{(n-i+1)(1+\hat{\beta}_{PACL})}\right)^{1/\hat{\beta}_{PACL}} & \text{if } C_i=1,\ D_i=E_i=0,\\ 0 & \text{if } E_i=1,\ D_i=C_i=0.\end{cases}$$

The essential difference between GMLE and GMLE1 is that in GMLE (7) $\hat{\beta}_{GMLE}$ is used while in GMLE1 $\hat{\beta}_{PACL}$ is used. One can also prove that the PACL estimator and the GMLE1 are asymptotically equivalent. Therefore, the same asymptotic properties as those in Theorem 2 hold for GMLE1 as well.

Computationally, PACL estimator is easier to compute than the GMLE and GMLE1. However, when assessing the performance of the estimators, GMLE and GMLE1 estimators do better than the PACL estimator for certain regions of *t*-values. See Zhang and Rao (2004). One has to take a balanced view of computational simplicity and efficiency.

Appendix

First from (9), the loglikelihood of $p=1/(1+\beta)$ from GMLE method is given by

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

We want to keep the subscript GMLE for L to indicate the underlying method of estimation. Recall $\hat{p}_{PACL} = \sum_{i=1}^{n} I[\Delta_i = 1]/[\sum_{i=1}^{n} (I[\Delta_i = 1] + I[\Delta_i = 0])]$. Therefore, the loglikelihood of p from PACL method is given by

$$\log L_{PACL}(p) = \sum_{i=1}^{n} [I[\Delta_i = 1] \log p + I[\Delta_i = 0] \log(1-p)] = \sum_{i=1}^{n} [D_i \log p + C_i \log(1-p)].$$

Let $U_{GMLE}(p)$,$V_{GMLE}(p)$,$U_{PACL}(p)$, and $V_{PACL}(p)$ denote the first and second derivatives of$\log L_{GMLE}(p)$ and $\log L_{PACL}(p)$, respectively. Then

Lemma 1.

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

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

Proof.

From simple algebra, one can have

$$\begin{aligned}
n^{-1/2} U_{GMLE}(p) &= n^{-1/2} \sum_{i=1}^{n} \left[-D_i \frac{n-i}{p^2} \log\left(1 - \frac{p}{n-i+p}\right) + C_i \frac{n-i+1}{(1-p)^2} \log\left(1 - \frac{1-p}{n-i+1}\right) \right] \\
&= n^{-1/2} \sum_{i=1}^{n} \left[-D_i \frac{n-i}{p^2} \left[-\frac{p}{n-i+p} + O((n-i+p)^{-2}) \right] \right] \\
&\quad + n^{-1/2} \sum_{i=1}^{n} \left[C_i \frac{n-i+1}{(1-p)^2} \left[-\frac{1-p}{n-i+1} + O((n-i+1)^{-2}) \right] \right] \\
&\cong n^{-1/2} \sum_{i=1}^{n} \left[D_i \frac{1}{p} \frac{n-i}{n-i+p} - C_i \frac{1}{1-p} \right] + O(n^{-1/2} \log n) \\
&\cong n^{-1/2} U_{PACL}(p) + O(n^{-1/2} \log n).
\end{aligned}$$

The second equality is from the Taylor expansion of $\log(1-x)$ for small x>0 and the fact that $\sum_{i=1}^{n} (1/i) = O(\log n)$, and the asymptotic equivalence $\cong$ is in the almost sure sense. The second part can be proved analogously. □

We need the following lemma for the proof of Theorem 1.

Lemma 2.

1. *Strong consistency of* $\hat{p}_{GMLE}$: $\hat{p}_{GMLE} \rightarrow_{a.s.} p_0$.

2. *Asymptotic equivalence of* $\hat{p}_{GMLE}$ *and* $\hat{p}_{PACL}$:

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

Proof.

To prove the strong consistency, we exploit the technique used in Andersen and Gill (1982) (AG82 for short) in proving the consistency of regression parameter in Cox's model. Let

$$X_n(p) \equiv \frac{1}{n}(\log L_{GMLE}(p) - \log L_{GMLE}(p_0)).$$

The technique of AG82 involves the following steps:

1. $\hat{p}_{GMLE}$, the maxima of $\log L_{GMLE}(p)$, or equivalently the maxima of $X_n(p)$, exists and is unique.

2. $X_n(p)$ converges almost surely to a concave function of *p* with a unique maximum at $p=p_0$.

Then it follows, from Theorem 10.8 in Rockafellar (1970) or the discussion in Appendix II of AG82, that $\hat{p}_{GMLE} \rightarrow_{a.s.} p_0$.

To prove the first part, noting that for $0<p,p_0<1$, $-n^{-1}V_{PACL}(p) = n^{-1}\sum_{i=1}^{n}[(D_i/p^2) + (C_i/(1-p)^2)] \rightarrow_{a.s.} \alpha_1/p^2 + \alpha_0/(1-p)^2 > 0$ by strong law of large numbers, we have, from Lemma 1, that

equation(12)

$$-n^{-1}V_{GMLE}(p) \rightarrow_{a.s.} \alpha_1/p^2+\alpha_0/(1-p)^2>0.$$

Therefore, $(1/n)\log L_{GMLE}(p)$ is a sequence of strictly concave functions, which gives Part 1. Now for the second part,

$$\begin{aligned}
X_n(p) &= \frac{1}{n}\sum_{i=1}^{n}\left[D_i \log\frac{p}{p_0} + C_i \log\frac{1-p}{1-p_0}\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} D_i\left[\log\left(1-\frac{p-p_0}{n-i+p}\right) + \frac{n-i}{p}\log\left(1-\frac{p}{n-i+p}\right) - \frac{n-i}{p_0}\log\left(1-\frac{p_0}{n-i+p_0}\right)\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} C_i\left[\frac{n-i+p}{1-p}\log\left(1-\frac{1-p}{n-i+1}\right) - \frac{n-i+p_0}{1-p_0}\log\left(1-\frac{1-p_0}{n-i+1}\right)\right] \\
&\cong \frac{1}{n}\sum_{i=1}^{n}\left[D_i \log\frac{p}{p_0} + C_i \log\frac{1-p}{1-p_0}\right] + \frac{1}{n}\sum_{i=1}^{n} D_i\left[-\frac{n-i}{n-i+p} + \frac{n-i}{n-i+p_0}\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} C_i\left[-\frac{n-i+p}{n-i+1} + \frac{n-i+p_0}{n-i+1}\right] + O(n^{-1}\log(n)) \\
&\cong (\alpha_1+\alpha_0)[p_0(\log p - \log p_0) + (1-p_0)(\log(1-p) - \log(1-p_0))] \equiv f(p)
\end{aligned}$$

say. Here asymptotic equivalence ≅ is in the almost sure sense.
Since $f'(p)|_{p=p0}=0$ and $f''(p)|_{p=p0}=(\alpha_1+\alpha_0)[-1/p_0-1/(1-p_0)]<0$ for $0<p_0<1$, the second part follows.

Finally, with Lemma 1 above, consistency of $\hat{p}_{GMLE}$ and $\hat{p}_{PACL}$, and (12), the "almost sure" version of Part 2 in Lemma 2 follows along the lines of the proof of Theorem 1 in Bailey (1984), where the asymptotic equivalence of the GMLE and the partial likelihood estimator of the regression parameter in Cox's model was given. □

To prove Theorem 1, we note that if $X_1,X_2,\ldots,$ are independent copies of *X*, the assumption of $t\in[0,T_0]$ implies $(1/n)\sum_{i=1}^{n} I[X_i > T_0] \geqslant M > 0$ almost surely for sufficiently large *n* and for some constant *M*.

Proof of Theorem 1.

Obviously, it is sufficient to prove

equation(13)

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

where $\hat{\Lambda}_{GMLE}(t)=-\log \hat{\hat{F}}_{GMLE}(t)$ and $\hat{\Lambda}_{PACL}(t)=-\log \hat{\hat{F}}_{PACL}(t)$. First, let $nK_n(t)=\sum_{i=1}^{n} I[U_i\leqslant t]$ denote the number of observations of *U* 's up to time *t* . For notational simplicity, we let $\hat{p}=\hat{p}_{PACL}$ and $\tilde{p}=\hat{p}_{GMLE}$ throughout the proof.
Then (7) and (2) can be rewritten as, respectively, for $t\leqslant T_0$

equation(14)

$$\hat{\hat{F}}_{GMLE}(t)=\prod_{i=1}^{nK_n(t)}\left(1-\frac{\tilde{p}}{n-i+\tilde{p}}\right)^{D_i}\times\prod_{i=1}^{nK_n(t)}\left(1-\frac{1-\tilde{p}}{n-i+1}\right)^{C_i\tilde{p}/(1-\tilde{p})},$$

equation(15)

$$\hat{\hat{F}}_{PACL}(t)=\prod_{i=1}^{nK_n(t)}\left(1-\frac{\varepsilon_{[i]}}{n-i+1}\right)^{\hat{p}}.$$

Therefore,

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

$$\hat{\Lambda}_{PACL}(t)=-\hat{p}\sum_{i=1}^{nK_n(t)}\log\left(1-\frac{\varepsilon_{[i]}}{n-i+1}\right)=\hat{p}\sum_{i=1}^{nK_n(t)}\frac{\varepsilon_{[i]}}{n-i+1}+O(n^{-1}).$$

Hence, from Lemma 2 and the fact that $\sum_{i=1}^{nK_n(t)} 1/(n-i+\tilde{p})^2 = O(n^{-1})$ uniformly and almost surely over $[0,T_0]$,

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

$$\cong -\sum_{i=1}^{nK_n(t)} \frac{D_i\tilde{p}}{n-i+\tilde{p}} + O(n^{-1}) \cong -\sum_{i=1}^{nK_n(t)} \frac{D_i\tilde{p}}{n-i+1} + O(n^{-1})$$

The asymptotic equivalence $\cong$ above is in the uniform and almost sure sense over the compact interval$[0,T_0]$. Similarly,

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

Therefore,

$$\hat{\Lambda}_{GMLE}(t) \cong \sum_{i=1}^{nK_n(t)} \left[D_i \frac{\tilde{p}}{n-i+1} + C_i \frac{\tilde{p}}{n-i+1}\right] + O(n^{-1}) \cong \tilde{p} \sum_{i=1}^{nK_n(t)} \frac{C_i + D_i}{n-i+1} + O(n^{-1}).$$

We have

$$n^{1/2}|\hat{\Lambda}_{GMLE}(t) - \hat{\Lambda}_{PACL}(t)| \cong (n^{1/2}|\tilde{p} - \hat{p}|)O(1) + O(n^{-1/2})$$

due to the fact that $\sum_{i=1}^{nK_n(t)} (C_i + D_i)/(n-i+1) = O(1)$ uniformly and almost surely over $[0,T_0]$. Therefore, (13) follows from Part 2 of Lemma 2 and the proof is completed. □

References

Abdushukurov, A., 1984. On some estimates of the distribution function under random censorship. Conference of Young Scientists, VINITI No. 8756-V. Math. Inst. Acad. Sci., Uzbek SSR, Tashkent.

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.

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.

Csörgö, S., 1998. Testing for the partial proportional hazards model of random censorship. In: Hušková, M., Lachout, P., Víšek, J.A. (Eds.), Prague Stochastics’ 98 Proceedings. Union of Czech Mathematicians and Physicists, Prague, pp. 87–92.

Gather, U., Pawlitschko, J., 1998. Estimating the survival function under a generalized Koziol–Green model with partially informative censoring. Metrika 48, 189–207

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.

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

Miller, R., 1981. Survival Analysis. Wiley, New York.

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

Zhang, H., Rao, M.B., 2004. On generalized maximum likelihood estimation in the proportional hazards model with partially informative censoring. Metrika 59, 125–136.