Open Journal of Statistics, 2016, 6, 209-219
Published Online April 2016 in SciRes. http://www.scirp.org/journal/ojs
http://dx.doi.org/10.4236/ojs.2016.62018

Equivalence between the Dependent Right Censorship Model and the Independent Right Censorship Model

Qiqing Yu[1], Kai Yu[2]

[1]Department of Mathematical Sciences, SUNY, Binghamton, USA
[2]Department of Mathematics, University of Mississippi, Oxford, USA
Email: qyu@math.binghamton.edu, kai0456@gmail.com

Received 8 February 2016; accepted 9 April 2016; published 12 April 2016

Abstract

Yu *et al.* (2012) considered a certain dependent right censorship model. We show that this model is equivalent to the independent right censorship model, extending a result with continuity restriction in Williams and Lagakos (1977). Then the asymptotic normality of the product limit estimator under the dependent right censorship model follows from the existing results in the literature under the independent right censorship model, and thus partially solves an open problem in the literature.

Keywords

Constant-Sum Models, Right-Censoring, Dependent Censoring, Necessary and Sufficient Condition

1. Introduction

In this paper we study various dependent right censorship (RC) models and their relation to the independent RC model in the literature. The definitions of these RC models are given in Definition 1.

Right censored data occur quite often in industrial experiments and medical research. A typical example in medical research is a follow-up study; a patient is enrolled and has a certain treatment within the study period. If the patient dies within the study period, we observe the exact survival time T; otherwise, we only know that the patient survives beyond the censoring time R. Thus the observable random vector is (V,δ), where $V=\min\{T,R\}$ $(=T\wedge R)$ and $\delta=\mathbf{1}_{(T\le R)}$, the indicator function of the event $\{T\le R\}$. Let $(V_1,\delta_1),\cdots,(V_n,\delta_n)$ be i.i.d. copies of (V,δ). Let $F_T(t)(=P(T\le t))$ be the cumulative distribution function (cdf) of T and $S_T=1-F_T$. Denote F_R, F_V and $F_{T,R}$ the cdf's of R, V and (T,R), respectively, and $F_{R|T}$ the conditional cdf

How to cite this paper: Yu, Q.Q. and Yu, K. (2016) Equivalence between the Dependent Right Censorship Model and the Independent Right Censorship Model. *Open Journal of Statistics*, **6**, 209-219. http://dx.doi.org/10.4236/ojs.2016.62018

of R given T and $S_{R|T} = 1 - F_{R|T}$. Let f_T (f_R or $f_{R|T}$) be the density function of T (R or $R|T$) (with respect to (w.r.t.) some measure). The common right censorship model assumes T and R are independent ($T \perp R$). Then the likelihood (function) for RC data is often defined as

$$\mathcal{L} = \mathcal{L}(F) = \prod_{i=1}^{n} \left(f(V_i) \right)^{\delta_i} \left(S(V_i) \right)^{1-\delta_i}, \; F \in \mathcal{F}, \tag{1}$$

(see [1]), where $\mathcal{F}$ is a collection of all cdf's if under the non-parametric set-up, or a parametric cdf family with a parameter, say $\theta \in \Theta$ and Θ is the parameter space, $S = 1 - F$ and f is the density of F. Recall that the formal definition of the likelihood (function) for a sample $\mathbf{X} = \mathbf{x}$ is $\Lambda = \Lambda(\theta) = f_{\mathbf{X}}(\mathbf{x}|\theta)$, $\theta \in \Theta$. Moreover, if $\Lambda_2 = c(\mathbf{x})\Lambda(\theta)$ $\forall \theta \in \Theta$, where $c(\cdot)$ does not depend on θ, and Λ_2 is also called a likelihood. We shall call Λ the full likelihood and Λ_2 a simplified one. Since our sample $\mathbf{x} = ((V_1, \delta_1), \cdots, (V_n, \delta_n))$,

$$\begin{aligned}
\Lambda &= \prod_{i=1}^{n} \left(\int_{r \ge V_i} \mathrm{d}F_{T,R}(V_i, r) \right)^{\delta_i} \left(\int_{t > V_i} \mathrm{d}F_{T,R}(t, V_i) \right)^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left[f_T(V_i) \int_{r \ge V_i} \mathrm{d}F_{R|T}(r|V_i) \right]^{\delta_i} \left[S_T(V_i) \frac{\int_{t > V_i} \mathrm{d}F_{T,R}(t, V_i)}{S_T(V_i)} \right]^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left(f_T(V_i) \right)^{\delta_i} \left(S_T(V_i) \right)^{1-\delta_i} \left(Q(V_i) \right)^{\delta_i} \left(G(V_i) \right)^{1-\delta_i},
\end{aligned} \tag{2}$$

where

$$Q(t) = \int_{r \ge t} \mathrm{d}F_{R|T}(r|t), \; t > 0, \; G(r) = \frac{\int_{t > r} \mathrm{d}F_{T,R}(t, r)}{S_T(r)}, \tag{3}$$

and the integrals are Lebesgure integrals. We say that a function $H(\cdot)$ is non-informative about the function $F_T(\cdot|\theta)$ with $\theta \in \Theta$ if it is not assumed that H is a function of F_T or θ. We shall further clarify what "non-informative" means in the next example.

Example 1.1. *Consider* 3 *cases of right censoring*:

Case (1). $T \perp R$ *with the parameter space* $\Theta = \{(F_T, F_R) : F_T \in \mathcal{F}, F_R \in \mathcal{F}\}$ (*see Equation* (1)).

Case (2). $T \perp R$ *with* $\Theta = \{(F_T, F_R) : F_T \in \mathcal{F}, F_R = F_T\}$.

Case (3). $R = 3 - T$ *with parameters* $\theta_i = P(T = i)$, $i \in \{1,2,3,4,5\}$ $\left(\sum_{i=1}^{5} \theta_i = 1\right)$. *$F_R$ is informative about F_T in cases* (2) *and* (3), *as it is a function of F_T in case* (2) *and a function of* $(\theta_1, \theta_2, \theta_3, \theta_4)$ *in case* (3). *However, F_R is non-informative* (*not informative*) *about F_T in case* (1), *as F_T and F_R are both independent parameters*.

If $T \perp R$, Λ in Equation (2) may be simplified as $\mathcal{L}$ as in Equation (1) due to the non-informative property by the well-known result as follows.

Proposition 1.1. *The full likelihood* Λ *can be simplified as* $\mathcal{L}(F)$ (*see Equation* (1)) *iff*

$$Q(\cdot) \text{ and } G(\cdot) \text{ given in Equation } (3) \text{ are non-informative about } F_T. \tag{4}$$

Example 1.1 (continued). *In case* (1), $\mathcal{L}(F)$ *is a likelihood function, as Equation* (4) *holds. In case* (2), $F_R = F_T$ *is informative about F_T and condition Equation* (4) *fails*.

$$\Lambda(F) = \prod_{i=1}^{n} f(V_i) \prod_{i=1}^{n} \left[\left(S(V_i) \right)^{1-\delta_i} \left(S(V_i+) \right)^{\delta_i} \right] = \mathcal{L}(F) \prod_{i=1}^{n} \left[\left(f(V_i) \right)^{1-\delta_i} \left(S(V_i+) \right)^{\delta_i} \right].$$

$\mathcal{L}(F)$ *is not a likelihood, but can be viewed as a partial likelihood. The generalized maximum likelihood* (*GMLE*) $\hat{S}$ *of S_T based on* $\mathcal{L}(F)$ *is still the PLE, i.e.,* $\hat{S}(t) = \prod_{i: V_{(i)} \le t} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}}$, *where* $V_{(i)}$ *is the i-th order statistic of* V_j*'s and* $\delta_{(i)}$ *is the δ_j that is associated with* $V_{(i)}$. *The variance satisfies* $n\mathrm{Var}\left(\hat{S}(t)\right) \to \frac{1 - S_T^2(t)}{2}$ (*if S_T is continuous*), *while the GMLE based on* Λ *is* $\tilde{S}(t) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(V_i > t)}}$ (*as* $S_V = S_T^2$ *and*

$\Lambda(F_T)=\prod_{i=1}^n f_V(V_i)$ *and* $n\mathrm{Var}(\tilde S(t))\to\frac{1-S_T^2(t)}{4}$ *by the delta method. Thus the PLE is not efficient. In case* (3), $\mathcal{L}(F)$ *is not a likelihood function. The full likelihood is* $\Lambda=\prod_{i=1}^n\left[(f_T(V_i))^{\delta_i}(f_T(3-V_i))^{1-\delta_i}\right]$ *(as* $T_i=\begin{cases}V_i\le 3-V_i=R_i & \text{if } \delta_i=1\\ 3-V_i>V_i=R_i & \text{if } \delta_i=0\end{cases}$*). If one treats* $\mathcal{L}(F)$ *as a (partial) likelihood, then its GMLE of* S_T *is the PLE* $\hat S$*. Let* $f_T(t)=\frac{1}{4}$, $t\in\{1,2,4,5\}$*. Then* $\hat S(2)\xrightarrow{a.s.}\left(1-\frac14\right)^1\left(1-\frac13\right)^0\left(1-\frac12\right)^1=\frac38\ne\frac12=S_T(2)$*, i.e., the PLE is not consistent at* 2*. The GMLE* $\tilde S$ *based on* Λ *is* $\tilde S(t)=\frac1n\sum_{i=1}^n\left[V_i^{\delta_i}(3-V_i)^{1-\delta_i}\right]\to\xrightarrow{a.s.}S_T(t)$.

Remark 1.1. *Example* 1.1 *indicates that if Equation* (4) *is not valid then the MLE based on so-called "likelihood"* $\mathcal{L}(F)$ *as in Equation* (1) *can be inconsistent, or can be less efficient than the MLE based on* Λ *due to loss of information on* $\theta\in\Theta$*. However, it is difficult to verify Equation* (4) *in practical applications, thus people propose some sufficient conditions. A typical sufficient condition of Equation* (4) *is that* $T\perp R$ *and* F_R *is non-informative about* F_T.

Williams and Lagakos (W&L) [2] point out that $T\perp R$ is often un-realistic. They further propose a constant-sum model (which allows $T\not\perp R$) as follows.

$$a(u)+B(u)=1\text{ a.e. w.r.t. }\mu_{F_T}\ (\text{the measure induced by }F_T), \tag{5}$$

where

$$\begin{aligned}&a(u)=P(\delta=1|T=u),\ B(u)=\int_{t\le u}\mathrm{d}B(t),\text{ and}\\ &\mathrm{d}B(u)=P(V\in(u-\mathrm{d}u,u],\delta=0|T\ge u).\end{aligned} \tag{6}$$

In the literature, there are many studies on the asymptotic properties of the PLE by weakening the assumptions in the independent RC model over the years (see, e.g., [3]-[10]). It is conceivable that the asymptotic properties of the PLE is difficult under the continuous constant-sum model in Equation (5). However, the next theorem makes it trivial.

W&L Theorem (Theorem 3.1 in [2]). W&L (1977)). *Suppose that* (T,R) *is a continuous random vector. Then Equation* (5) *holds iff* $\exists$ *a random vector* (Z,Y) *such that* (1) $F_Z=F_T$, $F_Y=B$ *(see Equation* (6)) *and* $Z\perp Y$*, where* (2) $Z=T\le Y$ *if* $T\le R$*, and* (3) $Y=R<Z$ *if* $R<T$.

By the W&L Theorem, one can easily make use of the existing results about the PLE under the assumption $T\perp R$ to establish asymptotic properties of the PLE under the continuous constant-sum model. Indeed, by (2) and (3) of the W&L Theorem,

$$V=T\wedge R=Z\wedge Y\text{ and }\delta=\mathbf{1}_{(T\le R)}=\mathbf{1}_{(Z\le Y)}. \tag{7}$$

Since $Z\perp Y$, the PLE $\hat F$ based on $(V_i,\delta_i),\cdots,(V_n,\delta_n)$ from (V,δ) (see Equation (7)) satisfies $\sup_{t\le\tau}\left|\hat F(t)-F_Z(t)\right|\to 0$ a.s., where $\tau=\sup\{t:F_V(t)<1\}$ (see [10]). By the W&L Theorem, $F_T=F_Z$ and Equation (7) holds, so $\sup_{t\le\tau}\left|\hat F(t)-F_T(t)\right|\xrightarrow{a.s.}0$ under the continuous RC model given in Equation (5), even if $T\not\perp R$.

On the other hand, case (3) in Example 1.1 shows that the PLE can be inconsistent for $t<\tau$ under a dependent RC model. Hence the W&L Theorem is quite significant. Yu *et al.* [11] show that the PLE is consistent under the dependent RC model considered in [12]-[15], etc., which assumes A1 and A2 as follows.

A1 $\int_{(t>r,\ F_{R|T}(r|t)\ne F_{R|T}(r|\tau_T))}\mathrm{d}F_T(t)=0$ for all *r*, or equivalently, $F_{R|T}(r|t)=F_{R|T}(r|\tau_T)$ a.e. in *t* (w.r.t. μ_{F_T}) on the set $\{x:f_T(x)>0,x>r\}$.

Notice that $F_{R|T}(r|\tau_T)$ is well defined if $P(T=\tau_T)>0$ and undefined if $P(T=\tau_T)=0$. We define $F_{R|T}(r|\tau_T)=F_{R|T}(r|\tau_T-)$ if $P(T=\tau_T)=0$. Notice that A1 says that $F_{R|T}(r|t)$ is constant in $t>r$, thus $F_{R|T}(r|\tau_T-)$ is well defined if $P(T=\tau_T)=0$.

A2 $G_{rc}(\cdot)$ is non-informative about F_T, with $G_{rc}(x) = F_{R|T}(x|\tau_T)$.

Definition 1. If $T \perp R$ and F_R is non-informative about F_T, then we call the RC model the independent RC model. The dependent RC model considered in this paper assumes that A1 and A2 hold.

Next example and Example 3.1 in Section 3 are examples that satisfies A1 but $T \not\perp R$.

Example 1.2. $P(R=1|T=0)=1$, $f_{R|T}(r|1) = (r \in (0,1))$ and T has a binomial distribution ($T \sim bin(1,p)$) with parameter $p \in (0,1)$.

Yu *et al*. [11] show that A1 and A2 are the necessary and sufficient (N&S) condition of Equation (4) under the non-parametric set-up. Then we may ask the following questions:

1) Are A1 and A2 the N&S condition of Equation (4) under the parametric set-up?

2) What is the relation between the constant-sum model (5) and A1?

3) Can the W&L Theorem be extended by eliminating the continuity restriction?

We give answers to the 3 questions. In Section 2, we show that A1 and A2 are a sufficient condition for Equation (4) under both non-parametric set-up and non-parametric set-up (see Theorem 2.1). Our study suggests that the constant sum model (5) is a special case of A1. In Section 3, we extend the W&L Theorem to the case that A1 holds (rather than the case that Equation (5) holds), which allows (T,R) being discontinuous. As a consequence, we establish the asymptotic normality of the PLE under the dependent RC model and under certain regularity conditions, making use of the existing results in the literature about the PLE under the independent RC model. In Section 4, we show that under the parametric set-up, A1 and A2 are not a necessary condition of Equation (4). Section 5 is a concluding remark. Some detailed proofs are relegated to Appendix.

2. The Relation between Equation (4), Equation (5) and A1

We shall first show that A1 and A2 are a sufficient condition of Equation (4), extending a result in [11] under the non-parametric set-up. Then we shall show that if (T,R) is continuous, the constant sum model is the same as A1; otherwise, these two models are different.

Theorem 2.1. *Equation* (4) *holds if* A1 *and* A2 *hold*.

Proof. Since $Q(t) = \int_{r\ge t} \mathrm{d}F_{R|T}(r|t) = 1 - \int_{r<t} \mathrm{d}F_{R|T}(r|t) = 1 - \int_{r<t} \mathrm{d}F_{R|T}(r|\tau_T)$, it is non-informative about F_T by A2. Moreover, by A1 $G(r) = \frac{\int_{t:t>r} \mathrm{d}F_{T,R}(t,r)}{S_T(r)} = \frac{\int_{t>r} f_{R|T}(r|t)\mathrm{d}F_T(t)}{S_T(r)} = \frac{\iint_{t>r} f_{R|T}(r|\tau_T)\mathrm{d}F_T(t)}{S_T(r)} = f_{R|T}(r|\tau_T)$. Thus $G(\cdot)$ is non-informative about F_T by A1 and A2, as $f_{R|T}$ and $F_{R|T}$ are equivalent. Then Equation (4) holds. □

The next example and lemma help us to understand the constant-sum model (5).

Example 2.1. *Suppose* $T \sim bin(n,p)$, $P(R=0)=1$ *and* $T \perp R$. *Then* A1 *holds, but not the constant-sum model assumption* (5), *as Equation* (6) *yields* $a(0) = P(T=0 \le R|T=0) = 1$, *and* $B(0) = \int_{u\le 0} \mathrm{d}B(u) = P(R=0<T) = p$. *Thus* $a(0)+B(0)>1$, *violating Equation* (5).

Lemma 2.1. $a(u) = S_{R|T}(u-|u)$ *and* $B(u) = \int_{t\le u} F_{R|T>t}(t)$, *if* (T,R) *is continuous, where* $\mathrm{d}F_{R|T>t}(t) = P(R \in (t-\mathrm{d}t,t] \| T>t)$.

Theorem 2.2. *If* (T,R) *is continuous, then* A1 *and Equation* (5) *are equivalent*.

The proofs of Lemma 2.1 and Theorem 2.2 are very technical but not difficult. For a better presentation, we relegate them to Appendix (see Section A.1 and Section A.2).

Remark 2.1. *Example* 2.1 *shows that* A1 *is not a special case of Equation* (5) (*or the constant-sum model*). *However, if* (T,R) *is continuous,* A1 *and the constant-sum model are equivalent. Thus the continuous constant-sum model is a special case of* A1. *Since Yu et al*. [11] *show that under the non-parametric set-up,* A1 *and* A2 *are the N&S condition that Equation* (4) *holds, it is desirable to extend W&L Theorem to the model that assumes* A1 *rather than the constant-sum model by eliminating the continuity assumption*.

3. Extension of the W&L Theorem

In the next theorem, we extend the W&L Theorem from the continuous constant-sum model to A1.

Theorem 3.1. A1 *holds iff there exist extended random variables Z and Y such that* 1) $Z \perp Y$ *and* $F_Y(\cdot) = B(\cdot)$, *where* $B(r) = \begin{cases} F_{R|T}(r|\tau_T) & \text{if } r<\infty \\ 1 & \text{if } r=\infty \end{cases}$, 2) $T = Z \le Y$ *if* $T \le R$, *and* 3) $T > Y = R$ *if* $T > R$.

In our theorem, there are two modifications to the W&L Theorem.

1) Equation (5) with continuous (T,R) is replaced by A1 without continuity assumptions.

2) The random vector is replaced by the extended random vector.

In fact, W&L Theorem is not accurate as stated, unless a random variable is allowed to take "values" $\pm\infty$ (see Examples 3.1 and 3.2 below). However, by the common definition of a random variable, it does not take values $\pm\infty$. Thus the random variables in their theorem should be referred to the extended random variables.

Example 3.1. *Suppose that* $f_{R|T}(r|t)=\begin{cases}1-t^2te^{1-r/t} & \text{if } 0<t\le 1/2, t\le r\\ 2r & \text{if } 0<r<t\le 1/2\end{cases}$ *and T has a uniform distribution* $(T\sim U(0,1/2))$, *then A*1 *holds and* (T,R) *is a continuous random vector, but* $T\not\perp R$. *By Theorem* 2.2, *it satisfies the constant-sum model. Consequently, the assumptions in the W&L Theorem are satisfied. In particular, R does not take the value* ∞. *If the W&L Theorem were correct, according to their definition, there would be a random variable Y with a cdf* $B(\cdot)$ *defined in Equation* (6). *However,* $B(u)=1/4<1$ *for* $u>1/2$ (*the proof is given in Appendix* (*see Section A*.3)).

Thus $B(\cdot)$ *is not a proper cdf as claimed in the W&L Theorem. Y should be modified to be an extended random variable such that* $P(Y\le r)=\begin{cases}B(r) & \text{if } r\in(0,\infty)\\ 1 & \text{if } r=\infty.\end{cases}$

Example 3.2. *A random sample of complete data* $T_1,\cdots,T_n$ *from T which has the exponential distribution* $T\sim Exp(\mu)$ *can be viewed as a special case of the RC data. But* $f_{R|T}$ *is not even defined for a random variable R. However, if we consider extended random variables in A*1, *that is, R may take values* $\pm\infty$, *then we can define* $R=\infty$. *Since* $P(R=\infty)=1$, $R\perp T$. *Thus Theorem* 3.1 *is trivially true in such case.*

Proof of Theorem 3.1. It suffice to show ($\Rightarrow$) part. Since $F_{R|T}(\cdot\,|\,\tau_T)$ is a conditional distribution, $B(r)$ defines a "cdf" on $(-\infty,\infty]$. Denote $\Omega=(-\infty,\infty]^2$ and let $\mathcal{B}$ be the Borel σ-field on Ω. Without loss of generality (WLOG), one can assume that $(\Omega,\mathcal{B},P)$ is the probability space such that $P(A)=\int_A dF_{T,R}(t,r)$ $\forall A\in\mathcal{B}$. Let W be the joint cdf defined by $W(t,y)=F_T(t)B(y)$ $\forall(t,y)\in\Omega$. By the Kolmogorov consistency theorem, $W(\cdot,\cdot)$ induces a random vector (Z,Y) on Ω by $(Z,Y)(\omega)=\omega$ $\forall\omega=(t,y)\in\Omega$. Note that $P((T,R)\in A)=P((Z,Y)\in A)$ $\forall A\subset\{(t,r):t>r\}$ and $A\in\mathcal{B}$. Verify that $Z(\omega)=T(\omega)\le Y(\omega)$ if $T(\omega)\le R(\omega)$; $T(\omega)=Z(\omega)>Y(\omega)=R(\omega)$ if $T(\omega)>R(\omega)$. Verify that $Z\perp Y$ as $F_Y=B(\cdot)$. Thus conditions (1), (2) and (3) hold. □

Remark 3.1. *In the previous proof, let* $\mathcal{S}$ *be the support of* $F_{T,R}$ *and* $A=\Omega\setminus\mathcal{S}$. *Then* (T,R) *may not be defined on A, but* (Z,Y) *may be defined on A. Thus it is necessary to create a new random variable Z.*

Corollary 3.1. *If A*1 *holds then* $V=T\wedge R=Y\wedge Z$ *and* $\delta=\mathbf{1}_{(T\le R)}=\mathbf{1}_{(Z\le Y)}$.

The asymptotic properties of the PLE under the continuous constant-sum model are obtained by making use of the W&L Theorem and the existing results in the literature on the PLE under the continuous independent RC model. Denote $D_V^*=\begin{cases}\{t:t<\tau\} & \text{if } P(T=\tau\le R)=0\\ \{t:t\le\tau\} & \text{otherwise.}\end{cases}$ The consistency of the PLE under assumption A1 is established in the literature as follows.

Theorem 3.2 (Yu *et al.* [11]). *Under A*1, $\sup_{t\in D_{V^*}}\left|\hat F(t)-F_T(t)\right|\xrightarrow{a.s.}0$.

Now by Theorem 3.1 and Corollary 3.1, we can construct another proof of the consistency of the PLE as follows.

Corollary 3.2. *Under A*1, $\sup_{t\in D_{V^{**}}}\left|\hat F(t)-F_T(t)\right|\xrightarrow{a.s.}0$ *where*

$$D_{V^{**}}=\begin{cases}\{t:t<\tau\} & \text{if } F_T(\tau)>F_T(\tau-) \text{ and } F_Y(\tau-)=1\\ \{t:t\le\tau\} & \text{otherwise.}\end{cases}$$

Proof. Yu and Li [10] show that if $T\perp R$, then $\sup_{t\in D_V}\left|\hat F(t)-F_T(t)\right|\xrightarrow{a.s.}0$, where $D_V=\begin{cases}\{t:t<\tau\} & \text{if } F_T(\tau)>F_T(\tau-) \text{ and } F_R(\tau-)=1\\ \{t:t\le\tau\} & \text{otherwise.}\end{cases}$ Under A1, $T\perp R$ may not be true, but by Theorem 3.1, $Z\perp Y$, $V=T\wedge R=Z\wedge Y$, $\delta=\mathbf{1}(T\le R)=\mathbf{1}(Z\le Y)$. Thus the observation (V_i,δ_i)'s are i.i.d. from (V,δ), which can be viewed as being generated from (T,R), as well as (Z,Y). Thus replacing D_V and (F_T,F_R)

by $D_{V^{**}}$ and (F_Z, F_Y), respectively, in the previous equation yields $\sup_{t\in D_{V^{**}}}\left|\hat{F}(t)-F_Z(t)\right| \xrightarrow{a.s.} 0$. Now since $F_Z = F_T$, the proof is done. □

Remark 3.2. *Notice that the statements in Theorem* 3.2 *is slightly different from the statements in Corollary* 3.2. *One is based on* (D_{V^*}, R), *and the other is based on* $(D_{V^{**}}, Y)$.

The asymptotic normality of the PLE under A1 without continuity assumption has not been established in the literature. It can be done now by making use of Theorem 3.1 and the existing results in the literature on the PLE under the independent RC model. In particular, assuming T is continuous, Breslow and Crowley [3] and Gill [6] show that

$$\sqrt{n}\left(\hat{F}(t)-F_T(t)\right) \xrightarrow{D} N\left(0,\sigma^2\right) \text{ for } t<\tau, \tag{8}$$

$$\lim_{n\to\infty} nCov\left(\hat{F}(t),\hat{F}(s)\right) = S_T(t)S_T(s)\int_0^{t\wedge s}\frac{1}{S_V(x-)S_T(x)}\mathrm{d}F_T(x),\ \ t,s<\tau. \tag{9}$$

Without continuity assumptions, Gu and Zhang [16] and Yu and Li [17] among others established asymptotic normality of the GMLE under the double censorship (DC) model. Since the independent RC model is a special case of the DC model, their results imply that (8) and (9) also hold if $\int_{t\in(0,\tau-\epsilon)}\frac{-1}{S_R(t)}\mathrm{d}S_R(t)<\infty \quad \forall\epsilon>0$, and if either $P(T>\tau)=0$ or $P(V=\tau)>0$. The next result follows from Theorem 3.1 and Corollary 3.1, which partially solves the open problem in [11] about the asymptotic normality of the PLE under the dependent RC model.

Theorem 3.3. *Equations* (8) *and* (9) *are valid if A*1 *holds and if either T is continuous or* (1) $\int_{t\in(0,b)}\frac{-1}{S_Y(t)}\mathrm{d}S_Y(t)<\infty \quad \forall b\in(0,\tau)$ *and* (2) *either* $P(T>\tau)=0$ *or* $P(V=\tau)>0$, *where the random variable Y is defined in Theorem* 3.1.

4. Are A1 and A2 the N&S Condition of Equation (4) under the Parametric Set-Up?

The answer to the question is "No" in general. We shall explain through several examples.

Example 4.1. *Suppose that* $P(T\in\{2,3,4\})=1$, $f_T(i)=p_i$, $\boldsymbol{p}=(p_2,p_3,p_4)\in\Theta$, *and*

$$\begin{array}{c} \\ f_{R|T}(1|t) \\ f_{R|T}(2|t) \\ f_{R|T}(3|t)\end{array}\begin{array}{c} t=2\ \ t=3\ \ t=4 \\ \begin{pmatrix} 3/6 & 2/6 & 1/6 \\ 2/6 & 3/6 & 3/6 \\ 1/6 & 1/6 & 2/6 \end{pmatrix}\end{array}$$

where $\Theta_0=\{\boldsymbol{p}: p_i\ge 0, p_2+p_3+p_4=1\}$ and $\Theta=\{\boldsymbol{p}\in\Theta_0: p_2=p_4\}$. This defines a parametric family of discrete distribution functions F_T. One can verify that possible observations I_i's are $(1,\infty)$, $\{2\}$, $(2,\infty)$, $\{3\}$, $(3,\infty)$. Write $f_I(I_i)=\mu_F(I_i)W(I_i)$, then $W(I_i)$ is either Q or G in Equation (3). In particular, $W(I_1)=f_I(I_1)/\mu_F(I_1)=\sum_i f_{R|T}(1|i)p_i/1=2/6$, $\cdots$, which lead to

	I_1	I_2	I_3	I_4	I_5
$W(I_i)$:	$2/6$	$S_{R\vert T}(2-\vert 2)$	$f_{R\vert T}(2\vert\tau_T)$	$S_{R\vert T}(3-\vert 3)$	$f_{R\vert T}(3\vert\tau_T)$
$W(I_i)$:	$2/6$	$3/6$	$3/6$	$1/6$	$2/6$

Thus the parametric model satisfies the N&S condition Equation (4). *But in view of* $f_{R|T}$, A1 *fails. Note that the PLE maximizes* $\mathcal{L}(\boldsymbol{p})=p_2^{n_2}(1-p_2)^{n_3}(p_3)^{n_4}(1-p_2-p_3)^{n_5}$ *over* $\boldsymbol{p}\in\Theta_0$. *It is important to notice that* $\mathcal{L}(\boldsymbol{p})$ *with* $\boldsymbol{p}\in\Theta_0$ *is not a likelihood. However,* $\mathcal{L}(\boldsymbol{p})$ *with* $\boldsymbol{p}\in\Theta$ *is a likelihood by Proposition* 1.1. *Verify that the PLE of* $\hat{F}(2)=\frac{n_2}{n-n_1}\xrightarrow{a.s.}\frac{P(T=2\le R)}{1-P(R=1)}=\frac{3p_2/6}{1-2/6}=\frac{3p_2}{4}\ne F_T(2)$, *thus the PLE is not consistent, but the MLE which maximizes* $\mathcal{L}(\boldsymbol{p})$ *over* $\boldsymbol{p}\in\Theta$ *is consistent, as expected. In fact,*

$$\mathcal{L} = p_2^{n_2}(1-p_2)^{n_3}(1-2p_2)^{n_4}(p_2)^{n_5} = p_2^{n_2+n_5}(1-p_2)^{n_3}(1-2p_2)^{n_4}.$$

$$\frac{\partial \log L}{\partial p_2} = 0 \ \textit{yields} \ \frac{n_2+n_5}{p_2} - \frac{n_3}{1-p_2} - \frac{2n_4}{1-2p_2} = 0.$$

$$\Rightarrow (n_2+n_5)(1-3p_2+2p_2^2) - n_3(p_2-2p_2^2) - 2n_4(p_2-p_2^2) = 0.$$

$$\Rightarrow n_2+n_5 - p_2(3n_2+3n_5+n_3+2n_4) + p_2^2(2n_2+2n_5+2n_3+2n_4) = 0.$$

$$\Rightarrow n_2+n_5 - p_2(3n_2+3n_5+n_3+2n_4) + p_2^2(2(n-n_1)) = 0 \ \textit{which is of the form} \ ap_2^2+bp_2+c=0.$$

$$\tilde{p}_2 = \frac{3n_2+3n_5+n_3+2n_4 \pm \sqrt{(3n_2+3n_5+n_3+2n_4)^2 - 4(n_2+n_5)(2n-2n_1)}}{4(n-n_1)}.$$

Then the MLE is the one that $\tilde{p}_2 \in [0,1/2]$. *Verify that*

$$b/n \to 3p_2\left(\frac{3}{6}+\frac{2}{6}\right) + (1-p_2)\frac{3}{6} + 2(1-2p_2)/6 = \frac{1}{6}\left(p_2(15-3-4)+5\right) = \frac{8p_2+5}{6} \quad a.s.;$$

$$2a/n \to 4(1-2/6) = 4^2/6 \quad a.s.;$$

$$4ac/n^2 \to \frac{16}{6}\frac{10p_2}{6} \quad a.s.$$

$$\tilde{p}_2 \to \frac{8p_2+5 \pm \sqrt{(8p_2+5)^2 - 160p_2}}{4^2} = \frac{8p_2+5 \pm \sqrt{(8p_2-5)^2}}{4^2} = p_2 \ \text{ or } 5/8 \ a.s.$$

Since $5/8 \notin [0,1/2]$, *the MLE* $\tilde{p}_2 \to p_2$ *a.s. as expected. That is, the MLE of* $\boldsymbol{p}$ *based on* $\mathcal{L}(\boldsymbol{p})$ $(\boldsymbol{p} \in \Theta)$ *is consistent.*

Example 4.2. *Suppose that* $f_T(i) = p_i$ *and* $p_2+p_3+p_4=1$. *This specifies a parametric family of discrete distributions with parameter* (p_2, p_3, p_4) *subject to the constraint* $p_i \ge 0$ *and* $\sum_{i=2}^{4} p_i = 1$. *Then A1 and A2 are the N&S condition of Equation* (4) (*see Section A.4 in Appendix*).

Remark 4.1. *In Example* 4.1, *since A1 fails, the W&L Theorem does not hold*.

Both Examples 4.1 and 4.2 are parametric cases, but A1 and A2 are the N&S condition of Equation (4) only in one case. In both cases the MLE's based on the simplified likelihood $\mathcal{L}$ as in Equation (1) are consistent. They indicate that in general under the parametric set-up, A1 is not the necessary condition of Equation (4). Since the two examples are discrete case, we also discuss two continuous examples.

Example 4.3. *Suppose that T is continuous,*

$$f_T(t) = \begin{cases} p & \text{if } t \in (0,1) \\ 1-p & \text{if } t \in (1,2) \end{cases}, \ \textit{and} \ \begin{matrix} & t \in (0,0.5)\cup(1,1.5) & t \in (0.5,1)\cup(1.5,2) \\ f_{R|T}(1|t) & 1/6 & 3/6 \\ f_{R|T}(2|t) & 5/6 & 3/6 \end{matrix}.$$

This defines a parametric family of a continuous random variable with parameter p. The possible observations I_i*'s are* $\{t\}$ *and* $(1,\infty)$. *A1 is violated due to the table for* $f_{R|T}$. *The* $G(\cdot)$ *and* $Q(\cdot)$ *in Equation* (3). *satisfy*

$$G((1,\infty)) = \frac{P(R=1<T)}{S_T(1)} = \frac{P(R=1<T)}{\frac{p}{6}+(1-p)\frac{3}{6}} = \frac{\frac{1}{6}p/2 + \frac{3}{6}(1-p)/2}{\frac{p}{6}+(1-p)\frac{3}{6}} = 1/2,$$

$$Q(\{t\}) = S_{R|T}(t-|t) = \begin{cases} 1 & \text{if } t<1 \\ 5/6 & \text{if } t \in (1,1.5) \\ 3/6 & \text{if } t \in (1.5,2). \end{cases}$$

Thus both Q and G in Equation (4) *are not functions of p or* F_T *and Equation* (4) *holds. Hence in this example,* *A*1 *is not a necessary condition of Equation* (4).

Example 4.4. *Suppose that* $T^* \sim Exp(\mu)$ *and* $T = 2\left(1-\exp\left(-T^*/\mu\right)\right) \sim U(0,2)$. *Define* $f_{R|T}$ *as in Example* 4.3, *then A*1 *fails and Equation* (4) *holds for the random vector* (T, R). *Now define* $f_{R^*|T^*}(r|y) = f_{R|T}\left(2\left(1-e^{-r/\mu}\right)\middle|2\left(1-e^{-y/\mu}\right)\right)$, $y, r > 0$. *Then* (T^*, R^*) *does not satisfy A*1 *but Equation* (4) *holds. It shows that if* $T \sim Exp(\mu)$, *A*1 *is not a necessary condition of Equation* (4) *though Equation* (4) *can hold under proper assumptions on* $f_{R^*|T^*}$. *The idea can be extended to the other continuous parametric families e.g.,* $N(\mu, \sigma^2)$, *Weibull, Gamma etc.*

5. Concluding Remark

We have established the equivalence between the standard RC model and the dependent RC model. The result simplifies the study on the properties of the estimators under the dependent RC model. The results in this paper may have applications in linear regression with right-censored data. For instance, the model assumption considered in [18] can be weekend. It is also of interest to study whether the result can be extended to the double censorship model [17] and the mixed interval censorship model [19].

Acknowledgements

We thank the Editor and the referee for their valuable comments.

References

[1] Cox, D.R. and Oakes, D. (1984) Analysis of Survival Data. Chapman & Hall, New York, 70-71.

[2] Williams, J.S. and Lagakos, S.W. (1977) Models for Censored Survival Analysis: Constant-Sum and Variable-Sum Models. *Biometrika*, **64**, 215-224. http://dx.doi.org/10.2307/2335687

[3] Breslow, N.E. and Crowley, J. (1974) A Large-Sample Study of the Life Table and Product Limit Estimates under Random Censorship. *Annals of Statistics*, **2**, 437-453. http://dx.doi.org/10.1214/aos/1176342705

[4] Peterson, A.V. (1977) Expressing the Kaplan-Meier Estimator as a Function of the Empirical Subsurvival Functions. *Journal of the American Statistical Association*, **7**, 854-858.

[5] Phadia, E.G. and Van Ryzin, J. (1980) A Note on the Convergence Rates for the Product Limit Estimator. *Annals of Statistics*, **8**, 673-678. http://dx.doi.org/10.1214/aos/1176345017

[6] Gill, R. (1983) Convergence of the Product Limit Estimator on the Entire Half Line. *Annals of Statistics*, **11**, 49-59. http://dx.doi.org/10.1214/aos/1176346055

[7] Shorack, G.R. and Wellner, J.A. (1986) Empirical Processes with Applications to Statistics. Wiley, New York.

[8] Stute, W. and Wang, J.L. (1993) The Strong Law under Random Censorship. *Annals of Statistics*, **21**, 1591-1607. http://dx.doi.org/10.1214/aos/1176349273

[9] Wang, J.G. (1987) A Note on the Uniform Consistency of the Kaplan-Meier Estimator. *Annals of Statistics*, **15**, 1313-1316. http://dx.doi.org/10.1214/aos/1176350507

[10] Yu, Q.Q. and Li, L.X. (1994) On the Strong Consistency of the Product Limit Estimator. *Sankhya A*, **56**, 416-430.

[11] Yu, Q.Q., Ai, X.S. and Yu, K. (2012) Asymptotic Properties of the Product-Limit-Estimator with Dependent Right Censoring. *International Journal of Statistics and Management System*, **7**, 84-104.

[12] Self, S.G. and Grossman, E.A. (1986) Linear Rank Tests for Interval-Censored Data with Application to PCB Levels in Adipose Tissue of Transformer Repair Workers. *Biometrics*, **42**, 521-530. http://dx.doi.org/10.2307/2531202

[13] Heitjan, D.F. and Rubin, D.B. (1991) Ignorability and Coarse Data. *Annals of Statistics*, **19**, 2244-2253. http://dx.doi.org/10.1214/aos/1176348396

[14] Gill, R., van der Laan, M.J. and Robins, J.M. (1997) Coarsening at Random: Characterization, Conjectures, Counter-Examples. In: Lin, D.Y. and Fleming, T.R., Eds., *Proceedings of the First Seattle Symposium in Biostatistics*: *Survival Analysis*, *Lecture Notes in Statistics* 123, Springer-Verlag, New York, 255-294. http://dx.doi.org/10.1007/978-1-4684-6316-3_14

[15] Gómez, R., Oller, G. and Calle, M.L. (2004) Interval Censoring: Model Characterizations for the Validity of the Simplified Likelihood. *The Canadian Journal of Statistics*, **32**, 315-326. http://dx.doi.org/10.2307/3315932

[16] Gu, M.G. and Zhang, C.H. (1993) Asymptotic Properties of Self-Consistent Estimator Based on Doubly Censored Data.

Annals of Statistics, **21**, 611-624. http://dx.doi.org/10.1214/aos/1176349140

[17] Yu, Q.Q. and Li, L.X. (2001) Asymptotic Properties of the GMLE of Self-Consistent Estimators with Doubly-Censored Data. *Acta Mathematica Sinica*, **17**, 581-594. http://dx.doi.org/10.1007/s101140000039

[18] Wang, Y.S., Chen, C.X. and Kong, F.H. (2010) Variance Estimation of the Buckley-James Estimator under Discrete Assumptions. *Journal of Statistical Computation and Simulation*, **81**, 481-496. http://dx.doi.org/10.1080/00949650903421085

[19] Yu, Q.Q., Wong, G.Y.C. and Li, L.X. (2001) Asymptotic Properties of Self-Consistent Estimators with Mixed Interval-Censored Data. *Annals of The Institute of Statistical Mathematics*, **53**, 469-486. http://dx.doi.org/10.1023/A:1014656726982

Appendix

We shall give the proofs of Lemma 2.1 and Theorem 2.2 and the proofs in some examples of the paper here.

A1. Proof of Lemma 2.1

WLOG, one can assume that u satisfies $P(T>u)>0$. By Equation (6),

$$a(u)=P(\delta=1|T=u)=P(R\geq T|T=u)=P(R\geq u|T=u)=S_{R|T}(u-|u).$$

$$\begin{aligned}
\mathrm{d}B(u) &= P(V\in(u-\mathrm{d}u,u],\delta=0\,|\,T\geq u) \quad \text{(by Equation (6))}\\
&= P(R\in(u-\mathrm{d}u,u],T>R\,|\,T\geq u)\\
&= P(R\in(u-\mathrm{d}u,u],T>R,T\geq u)P(T\geq u)\\
&= \frac{P(R\in(u-\mathrm{d}u,u],T>R,T>u)}{P(T\geq u)}\frac{P(T>u)}{P(T>u)}+\frac{P(R\in(u-\mathrm{d}u,u],T>R,T=u)}{P(T\geq u)}\\
&= P(R\in(u-\mathrm{d}u,u]\|T>u)\frac{P(T>u)}{P(T\geq u)}+\frac{P(R\in(u-\mathrm{d}u,u),T=u)}{P(T\geq u)}\\
&= \mathrm{d}F_{R|T>u}(u)\frac{P(T>u)}{P(T\geq u)}+\frac{P(R\in(u-\mathrm{d}u,u),T=u)}{P(T\geq u)}.
\end{aligned}$$

If T is a continuous random variable, then the previous equation and Equation (6) yield

$$B(u)=\int_{t\leq u}\frac{P(T>t)}{P(T\geq t)}\mathrm{d}F_{R|T>t}(t)=\int_{t\leq u}\mathrm{d}F_{R|T>t}(t). \quad \square$$

A2. Proof of Theorem 2.2

Assume that (T,R) is continuous. Then by Lemma 2.1, Equation (5) holds iff $S_{R|T}(u-|u)+\int_{r\leq u}\mathrm{d}F_{R|T>r}(r)=1$ a.e. in u w.r.t. μ_{F_T}, iff $F_{R|T}(u-|u)=\int_{r\leq u}\mathrm{d}F_{R|T>r}(r)$ a.e. in u w.r.t. μ_{F_T}.

Since (T,R) is continuous, $F_{R|T}(u-|u)=\int_{r<u}f_{R|T}(r|u)\mathrm{d}r$. By Lemma 2.1,

$$\int_{r\leq u}\mathrm{d}F_{R|T>r}(r)=\int_{r<u}\mathrm{d}F_{R|T>r}(r)=\int_{r<u}P(R\in(r-\mathrm{d}r,r]\|T>r)=\int_{r<u}f_{R|T>r}(r)\mathrm{d}r,$$

where $f_{R|T>r}(r)=\frac{\int_{t>r}f_{R|T}(r|t)f_T(t)\mathrm{d}t}{P(T>r)}$. Thus, Equation (5) holds iff

$\int_{r<u}f_{R|T}(r|u)\mathrm{d}r=\int_{r<u}\frac{\int_{t>r}f_{R|T}(r|t)f_T(t)\mathrm{d}t}{P(T>r)}\mathrm{d}r$ a.e. in u (w.r.t. μ_{F_T});

iff $\mu_{F_{R|T}(\cdot|u)}\left(\left\{r:r<u,f_{R|T}(r|u)\neq\frac{\int_{t>r}f_{R|T}(r|t)f_T(t)\mathrm{d}t}{P(T>r)}\right\}\right)=0$ a.e. in u (w.r.t. μ_{F_T});

iff $\mu_{F_{R|T}(\cdot|u)}\left(\left\{r:r<u,\int_{t>r}f_{R|T}(r|u)f_T(t)\mathrm{d}t\neq\int_{t>r}f_{R|T}(r|t)f_T(t)\mathrm{d}t\right\}\right)=0$ a.e. in u;

iff $\mu_{F_{T,R}}\left(\left\{(t,r):r<u,r<t,f_{R|T}(r|t)\neq f_{R|T}(r|u)\right\}\right)=0$ a.e. in u (w.r.t. μ_{F_T});

iff for almost all r (w.r.t. μ_{F_R}), $f_{R|T}(r|t)$ is constant in t a.e. w.r.t. μ_{F_T} on $\{t:t>r\}$;

iff $\forall r$, $F_{R|T}(r|t)$ is constant in t a.e. w.r.t. μ_{F_T} on $\{t : t > r\}$ (which is A1). □

A3. Proof of the Equation $B(u) = 1/4 < 1$ for $u > 1/2$ in Example 3.1

$$
\begin{aligned}
B(u) &= \int_{r<u} \mathrm{d}F_{R|T>r}(r) \quad (\text{by Lemma 2.1}) \\
&= \int_{r<1/2} \mathrm{d}F_{R|T>r}(r) \quad (\text{as } T \sim U(0,1/2)) \\
&= \int_{r<1/2} f_{R|T}(r|\tau_T)\mathrm{d}r \quad (\text{by Theorem 2.2, where } \tau_T = 1/2) \\
&= \int_0^{1/2} 2r\mathrm{d}r \quad (\text{by the given expression of } f_{R|T}) \\
&= 1/4 < 1. \ \square
\end{aligned}
$$

A4. Proof of Example 4.2

If $f_{R|T}(r|t)$ is not constant a.e. (w.r.t. μ_{F_T}) in $\{t : t > r\}$, $\exists i, j > r$ such that

1) $f_{R|T}(r|i) = a < b = f_{R|T}(r|j)$,
2) $p_i > 0$ and $p_j > 0$.
3) $p_i + p_j = S_T(r)$ is fixed.

$$
\begin{aligned}
S_T(r)G(r) &= \int_{t>r} f_{R|T}(r|t)\mathrm{d}F_T(r) \\
&\geq \int_{t=j} \left(f_{R|T}(r|j) - c\right)\mathrm{d}F_T(t) \left(\text{if } p_i \approx 0 \text{ and } c = \frac{b-a}{4}\right) \\
&\approx S_T(r)(b-c) \text{ (by (3))}.
\end{aligned}
$$

$$
\begin{aligned}
S_T(r)G(r) &= \int_{t>r} f_{R|T}(r|t)\mathrm{d}F_T(t) \\
&\leq \int_{t=i} \left(f_{R|T}(r|i) + c\right)\mathrm{d}F_T(t) \left(\text{if } p_j \approx 0\right) \\
&\approx S_T(r)(a+c) \text{ (by (3))}
\end{aligned}
$$

Thus $S_T(r)(a+c) \geq S_T(r)(b-c)$, contradicting $a + c < b - c$ by assumption (2). □