Ann. Inst. Statist. Math.
Vol. 43, No. 3, 579–596 (1991)

MINIMAXITY AND ADMISSIBILITY OF THE PRODUCT LIMIT ESTIMATOR*

E. G. PHADIA[1] AND QIQING YU[2]**

[1] *Department of Mathematics, William Paterson College of New Jersey, Wayne, NJ 07470, U.S.A.*
[2] *Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook, NY 11794, U.S.A.*

(Received October 26, 1989; revised May 31, 1990)

Abstract. In this article we examine the minimaxity and admissibility of the product limit (PL) estimator under the loss function

$$L(F, \hat{F}) = \int (F(t) - \hat{F}(t))^2 F^{\alpha}(t)(1 - F(t))^{\beta} dW(t).$$

To avoid some pathological and uninteresting cases, we restrict the parameter space to $\Theta = \{F\colon F(y_{\min}) \geq \epsilon\}$, where $\epsilon \in (0, 1)$ and $y_1, \dots, y_n$ are the censoring times. Under this set up, we obtain several interesting results. When $y_1 = \cdots = y_n$, we prove the following results: the PL estimator is admissible under the above loss function for $\alpha, \beta \in \{-1, 0\}$; if $n = 1$, $\alpha = \beta = -1$, the PL estimator is minimax iff $dW(\{y\}) = 0$; and if $n \geq 2$, $\alpha, \beta \in \{-1, 0\}$, the PL estimator is not minimax for certain ranges of ϵ. For the general case of a random right censorship model it is shown that the PL estimator is neither admissible nor minimax. Some additional results are also indicated.

Key words and phrases: Minimaxity, censored data, admissibility, nonparametric estimation, product limit estimator.

1. Introduction

Nonparametric minimax estimators of a cumulative distribution function (cdf), F, were proposed in Phadia (1973) for the Cramer-von Mises type loss function,

$$L(F, \hat{F}) = \int (F(t) - \hat{F}(t))^2 h(F(t)) dW(t) \tag{1.1}$$

* Partially supported by the Governor's Challenge Grant.
** Part of the work was done while the author was visiting William Paterson College.

where $h(t) = t^\alpha(1-t)^\beta$ with α, $\beta \in \{-1, 0\}$ and W is a known weight function. No assumptions were made about F. It was shown that these estimators were step functions taking jumps at the observed values and were independent of W. In particular, it was shown that when $\alpha = \beta = -1$, the sample distribution function is minimax.

When the data is censored on the right, Kaplan and Meier (1958) proposed the product limit (PL) estimator as an analog of the sample distribution function to estimate F nonparametrically. Since the publication of this pioneer paper, a considerable interest was focused on constructing, among others, Bayesian (Susarla and Van Ryzin (1976), Ferguson and Phadia (1979)), empirical Bayesian (Susarla and Van Ryzin (1978), Phadia (1980)) and minimax estimators based on censored data. For further references in this direction, see Ferguson *et al.* (1989). The natural question that arises is whether the PL estimator enjoys similar properties as the sample cdf. Recently Meeden *et al.* (1989) showed the admissibility of the PL estimator when $\alpha = \beta = -1$. Still, the whole area of decision theoretic consideration in the presence of censoring is open and needs to be addressed.

Our objective in this paper is to make an attempt, as a beginning, to address this question in some specific cases. Specifically, we consider the problem of the minimaxity and admissibility of the PL estimator under the above loss function. Since the PL estimator is undefined beyond the largest observation if it happens to be a censored one, we shall define it to be zero, consistent with the usual practice. Also, if it is not zero, the risk may not be finite (e.g., consider the case where $\alpha = \beta = -1$ or $\alpha = 0$ and $\beta = -1$) in which case the minimaxity is of no interest. This modified estimator is known to be a self-consistent (Efron (1967)) estimator. Hereafter, the PL estimator refers to this self-consistent estimator. Also, the problem of finding minimax and admissible estimators based on right censored data is considered.

The difficulty that arises in treating the censored data is that the equalizer rule approach used in the uncensored data case is no longer applicable here. For example, consider the case of $n = 1$. Let $X \sim F$, $Y \sim G$, X and Y be independent and assume that we observe $Z = X \wedge Y$, $\delta = 1(X \le Y)$, where $1(A)$, $1[A]$ or $1_{(A)}(t)$ stands for the indicator function of the set A. The Bayes estimator (Susarla and Van Ryzin (1976)) with respect to the Dirichlet process prior with parameter α, is

$$
\begin{aligned}
(1.2) \qquad \hat{S}(t) &= \frac{\alpha((t, +\infty)) + 1_{(-\infty, Z)}(t)}{\alpha(R) + 1}\left(1 + \frac{1_{[Z,\infty)}(t)1(X > Y)}{\alpha((Y, \infty))}\right) \\
&= a + b1_{(-\infty,Z)}(t) + ac1_{[Z,\infty)}1(X > Y),
\end{aligned}
$$

say, where a, b and c are constants.

Now computing the risk under the loss function, L, with $\alpha = \beta = 0$, it can be seen that it is almost impossible to show the existence of an equalizer rule unless the loss function is suitably modified (one such modification is indicated by (1.5) in Phadia and Yu (1989)). Even then, it is not easy to specify a suitable prior or a sequence of priors with respect to which the proposed equalizer rule will be a Bayes rule, an essential step in proving the minimaxity.

In view of these difficulties, the approach adopted in this paper is based on the basic definitions of the minimaxity and admissibility.

In Section 2, we consider the case of observations censored on the right by $y_1 = \cdots = y_n = y$, and establish some minimaxity and admissibility results. In Section 3, we consider the general case of right censorship, i.e. we no longer assume $y_1 = \cdots = y_n = y$. The Appendix contains some proofs of lemmas and propositions referred to in the text.

2. The case $y_1 = \cdots = y_n = y$

In keeping with the context of the survival data analysis, we assume the domain of F to be $[0, +\infty)$, $S(t) = \Pr(X > t) = 1 - F(t)$, and $F(t)$ and $S(t)$ will be used interchangeably. In our treatment, we assume the parameter space

$$\Theta_\epsilon = \{F\colon\ F(y_{\min}) \geq \epsilon\} \quad \text{for some positive } \epsilon, \tag{2.1}$$

where $y_{\min} = \min\{y_1, \ldots, y_n\}$ and $y_1, \ldots, y_n$ are fixed known censoring times. This is logical, since if $F(y_{\min}) = \epsilon = 0$, we have a pathological situation that all observations are censored. This is, furthermore, consistent with the Bayesian point of view that if a priori the distribution function is known to belong to some restricted parameter space, there is no reason to go beyond this restricted parameter space. Indeed, in practice, who will try to estimate a survival function based on data that, say, out of 30 observations, all but one, are censored? (this would imply that, in practice, $\epsilon \geq 1/30$). Obviously, the anticipated proportion of real observations, based on past experience, will provide a reasonable value to take for ϵ.

Our estimator of $S(t)$ will belong to the family of all non-increasing functions with range within $[0, 1]$. The risk of d will be denoted by $R(S, d) = E\{L(F, 1-d)\}$ and we use the following abbreviations for the loss function defined in (1.1).

$$\begin{aligned} L_1 &= L(F, \hat{F}) \quad \text{where } \alpha = \beta = 0; \\ L_2 &= L(F, \hat{F}) \quad \text{where } \alpha = \beta = -1; \\ L_3 &= L(F, \hat{F}) \quad \text{where } \alpha = -1,\ \beta = 0; \\ L_4 &= L(F, \hat{F}) \quad \text{where } \alpha = 0,\ \beta = -1. \end{aligned} \tag{2.2}$$

Our main interest is in L_2, since the PL estimator is minimax and admissible under L_2 in the uncensored data case, whereas it is not minimax under L_1, L_3 and L_4. Thus a natural question is whether this is true for the censored data case as well.

As indicated earlier we consider in this section the censoring points $y_1 = \cdots = y_n = y > 0$. Thus, we are assuming that $F \in \Theta_\epsilon = \{F\colon\ F(y) \geq \epsilon\}$, where $\epsilon > 0$. This restriction on the space of F highlights the distinction of this case in contrast to the case of uncensored data. Note that if $\epsilon = 1$, $F(y) = 1$ and it reduces to the uncensored data case (provided $dW(\{y\}) = 0$ since $d_{PL}(y) = 0$). Therefore we assume hereafter that $\epsilon < 1$ unless it is specified otherwise. We assume that

$X_1 \cdots X_n \overset{\text{iid}}{\sim} F$, and $Z_i = X_i \wedge y$ and $\delta_i = 1\ (X_i \leq y)$ are the observable random variables. Based on (Z_i, δ_i), $i = 1, 2, \ldots, n$, we thus consider the problem of estimation of F nonparametrically. However, in our treatment since y is assumed to be fixed, this is equivalent to considering X_i's and y instead of Z_i's and δ_i's.

Unlike in the uncensored case, we have to devote special attention to the case, $t > y$. For example, if $\alpha = \beta = -1$, we see immediately that in order that the risk be finite, the estimator has to be zero for $t > y$. Also, at times we are forced to consider the estimators which give a positive value to the singleton $\{y\}$. Thus, we are considering W which may or may not be continuous.

Recall that in the uncensored case, the sample cdf is minimax under the loss function L_2. The natural question is, whether a similar conclusion holds for PL estimator in the censored case? We provide some partial answers in this section. For the sample size $n = 1$, we show in Theorem 2.1 that, under certain conditions on the weight function W, d_{PL} is minimax. On the other hand for $n \geq 2$, we show in Theorem 2.2 that for a certain range of ϵ, d_{PL} is not minimax irrespective of the nature of W. Also in this section, we extend the minimax property of the PL estimator for the cases L_1, L_3 and L_4 from the uncensored case (i.e. $\epsilon = 1$) to the censored case. Finally, we construct the minimax estimator under L_2 for the case $n = 1$.

Meeden *et al.* (1989) proved the admissibility of d_{PL} under L_1, assuming the random right censorship model. If the censoring times are assumed to be equal and known fixed value, we show a stronger result in Theorem 2.4, namely, that d_{PL} is not only admissible for a smaller parameter space Θ_ϵ under L_1, but also under the arbitrary L as (1.1) with $\alpha, \beta \in [-1, 1)$.

THEOREM 2.1. *Let $n = 1$, $L = L_2$ and $F \in \Theta_\epsilon$. Let D_c denote the class of estimators with form as follows,*

$$d(t) = d_b(t) = \begin{cases} b & \text{if} \quad t = y < X, \\ d_{PL} & \text{otherwise}, \end{cases} \tag{2.3}$$

where $0 \leq b \leq 1 - \epsilon$. Then (i) *$D_c$ is the class of all admissible estimators with finite risks (for any $F \in \Theta_\epsilon$);* (ii) *$d_{1-\epsilon}(t)$ is minimax;* (iii) *d_{PL} is minimax iff $dW\{y\} = 0$, and d_{PL} is admissible.*

We need the following lemma to prove Theorem 2.1.

LEMMA 2.1. *Let $n = 1$, $L = L_2$ and $F \in \Theta_\epsilon$. Then the estimators of $S(t)$ having finite risks are necessarily of the form as* (2.3) *with $0 \leq b < 1$.*

See the Appendix for the proof of this lemma.

Let D denote the class of estimators of form (2.3), where $0 \leq b < 1$.

PROOF OF THEOREM 2.1.

Case (i): By Lemma 2.1, it follows that the estimators with finite risks are of the form d_b. Given $d_b \in D$, consider an estimator $d_a \in D$, with $a \neq b$ and $R(S, d_a) - R(S, d_b) \leq 0$ for all F. That is

$$
\begin{aligned}
(2.4)\quad & R(S,\, d_a) - R(S,\, d_b) \\
&= \int_{\{y\}} [(1 - F(t) - a)^2 - (1 - F(t) - b)^2] \\
&\quad \cdot (1 - F(t))/[F(t)(1 - F(t))] dW(t) \\
&= dW(\{y\})(b - a)(2 - 2F(y) - (a + b))/F(y) \leq 0 \qquad \forall F \in \Theta_\epsilon .
\end{aligned}
$$

If $a > b$, it implies that $2 - 2F(y) - (a + b) \geq 0$ for $F \in \Theta_\epsilon$. However, this will lead to a contradiction if we take an $F_0 \in \Theta_\epsilon$ such that $F_0(y) = 1$.

If $a < b$, then (2.4) implies that $2 - 2F(y) - (a + b) \leq 0$ for $F \in \Theta_\epsilon$, or $F(y) \geq 1 - (a+b)/2$ for all $F \in \Theta_\epsilon$. Since $F(y) \geq \epsilon$, this will be so iff $\epsilon \geq 1 - (a+b)/2$. That is, given d_b, there is a d_a which is better than d_b, iff a and b satisfy $a + b \geq 2(1 - \epsilon)$ and $a < b$. (It is easy to see that the strict inequality holds if $F(y) = 1$.) This means that d_b is admissible iff $b \in [0,\, 1 - \epsilon]$. (Note that the case $b = 0$ yields the PL estimator.)

Case (ii): By Lemma 2.1, $\inf_d \sup_F R(S,\, d) = \inf_{d \in D} \sup_F R(S,\, d)$. Thus, it suffices to show that $\sup_F R(S,\, d_{1-\epsilon}) = \inf_{d \in D} \sup_F R(S,\, d)$. Since $d \in D$ and d_{PL} differ only on the set $\{t = y < X\}$, we have

$$
\begin{aligned}
(2.5)\qquad R(S,\, d) &= R(S,\, d_{PL}) - \int_{\{y\}} E[L(F,\, 1 - d_{PL})1(y < X)] dW(t) \\
&\quad + \int_{\{y\}} E[L(F,\, 1 - d)1(y < X)] dW(t) \\
&= \left\{ \int_0^{y^-} + \int_{\{y\}} \frac{1 - F(t)}{F(t)} + \int_{y^+}^{\infty} \frac{1 - F(t)}{F(t)} \right. \\
&\quad - \int_{\{y\}} \frac{1 - F(t)}{F(t)} \cdot (1 - F(y)) \\
&\quad \left. + \int_{\{y\}} \frac{(1 - F(t) - b)^2}{F(t)(1 - F(t))} \cdot (1 - F(y)) \right\} dW(t) \\
&\leq \int_0^{y^-} dW(t) + \int_{(y,\infty)} \frac{1 - F(y)}{F(y)} dW(t) \\
&\quad + \left[(1 - F(y)) + \frac{(1 - F(y) - b)^2}{F(y)} \right] dW(\{y\}),
\end{aligned}
$$

by simplifying and using the monotonicity of F. Writing $F(y) = x$, $W\{(0,\, y)\} = w_1$, $W\{(y,\, \infty)\} = w_2$ and $W\{y\} = w_0$, we have

$$
(2.6)\qquad R(S,\, d) \leq [w_2 + (1 - b)^2 w_0]/x + 2bw_0 + w_1 - w_2 - w_0 .
$$

Define the RHS of (2.6) to be $\phi(b,\, x)$, which is a decreasing function of $x \in [\epsilon,\, 1)$. It is easy to see (for details, refer to Phadia and Yu (1989))

$$
(2.7)\qquad \sup_{\epsilon \leq F(y) \leq 1} R(S,\, d) = [w_2 + (1 - b)^2 w_0]/\epsilon + 2bw_0 + w_1 - w_2 - w_0 .
$$

Now minimizing $\phi(b, \epsilon)$ with respect to b, we have the solution $b = 1 - \epsilon$ and the supremum of the risk of $d_{1-\epsilon}$ is

$$\phi(1-\epsilon, \epsilon) = \int_0^{y^-} dW + \int_{\{y\}} (1-\epsilon)dW + \int_{y^+}^{\infty} (1/\epsilon - 1)dW.$$

To prove (iii) we see that if $dW\{y\} = 0$, the risk of $d \in D$ is the same as the risk of d_{PL}. Hence d_{PL} is minimax. On the other hand if $dW\{y\} > 0$, then the solution, $b = 1 - \epsilon$, that minimizes (2.7) with respect to b, is unique. Therefore, d_{PL} is not minimax since $\epsilon < 1$ by assumption. The admissibility of d_{PL} follows from (i). □

Remark 2.1. Theorem 2.1 can be extended to the case that the censoring time is random under certain conditions. For example, we have the following statement:

Suppose that $n = 1$ and the parameter space $\Theta_\epsilon = \{F\colon\ F(y_0) \geq \epsilon\}$. Let Y be the random censoring time satisfying $P\{Y \geq y_0\} = 1$. Then

$$d_{1-\epsilon} = \begin{cases} 1-\epsilon & \text{if} \quad t = Y < X \\ d_{PL} & \text{otherwise,} \end{cases}$$

is minimax.

Recall that under the uncensored model, the sample cdf is minimax under L_2 for all $n \geq 1$. The similar result is not true for the PL estimator, as the following theorem shows.

THEOREM 2.2. *Let $n \geq 2$, $L = L_2$ and $F \in \Theta_\epsilon$. Then d_{PL} is not minimax for $0 < \epsilon < 1 - 1/n$.*

PROOF. The proof is in demonstrating suitable estimators that beat d_{PL} in the sense of smaller supremum risk. Let

$$d(t) = \begin{cases} b & \text{for} \quad X_{(n-1)} \leq t < y \wedge X_{(n)} \\ d_{PL} & \text{otherwise,} \end{cases} \tag{2.8}$$

where $1/n < b = b(\epsilon) < 2/n$. Let $0 < \delta = b - 1/n < 1/n$. We have to show that

$$\sup_{F \in \Theta_\epsilon} R(S, d) - \sup_{F \in \Theta_\epsilon} R(S, d_{PL}) < 0.$$

However, for L_2,

$$\sup_{F \in \Theta_\epsilon} R(S, d_{PL}) = \int_{[0,y)} (1/n)dW(t) + \int_{[y,\infty)} [(1-\epsilon)/\epsilon]dW(t)$$

and

$$\int_{[0,y)} (1/n)dW(t) = E\int_{[0,y)} (S(t) - d_{PL})^2/[S(t)(1-S(t))]dW(t),$$

which is independent of $S(t)$. Thus, it suffices to show

$$\sup_{F\in\Theta_\epsilon}\left\{E\int_{[0,y)}[(S(t)-d)^2/[S(t)(1-S(t))] \right.$$
$$-(S(t)-d_{PL})^2/[S(t)(1-S(t))]]dW(t)$$
$$\left.+\int_{[y,\infty)}[(1-F(t))/F(t)-(1-\epsilon)/\epsilon]dW(t)\right\}<0. \tag{2.9}$$

The quantity in the brace for a fixed $F(=1-S)\in\Theta_\epsilon$ is equal to

$$E\int_{[0,y)}[(S(t)-b)^2-(S(t)-1/n)^2]/[S(t)(1-S(t))] \tag{2.10}$$
$$\cdot 1(X_{(n-1)}\le t<y\wedge X_{(n)})dW(t)$$
$$+\int_{[y,\infty)}[(1-F(t))/F(t)-(1-\epsilon)/\epsilon]dW(t)$$
$$=\int_{[0,y)}[2(b-1/n)(F(t)-(2-b-1/n)/2)$$
$$\cdot nF^{n-1}(t)S(t)]/[F(t)S(t)]dW(t)$$
$$+\int_{[y,\infty)}[1/F(t)-1/\epsilon]dW(t)$$
$$\le\int_{[0,y)}[2(b-1/n)(F(t)-(2-b-1/n)/2)\cdot nF^{n-2}(t)]dW(t)$$
$$+\int_{[y,\infty)}[1/F(y)-1/\epsilon]dW(t)\qquad(\text{since } F(t)\ge F(y) \text{ if } t>y).$$

Without loss of generality (WLOG), we can assume that the maximum of the integrand of the first integral of (2.10) or (2.9) is achieved at $F(t)=c$, where $c\le F(y)$.

Thus, the quantity in braces in (2.9) is less than and equal to the supremum of

$$\int_{[0,y)}[2(b-1/n)(c-(2-b-1/n)/2)\cdot nF^{n-2}(t)]dW(t)+\int_{[y,\infty)}\left[\frac{1}{c}-\frac{1}{\epsilon}\right]dW(t)$$

since $c\le F(y)$ if $t>y$. Furthermore, $\exists\, F_0\in\Theta_\epsilon$ such that $F_0(y)=c$. Thus, the quantity in braces in (2.9) is less than or equal to the supremum of

$$\int_{[0,y)}2\delta(F(y)-1+\delta/2+1/n)\cdot nF^{n-2}(y)dW(t)+\int_{[y,\infty)}[1/F(y)-1/\epsilon]dW(t). \tag{2.11}$$

To find the maximum with respect to $F(y)\in[\epsilon,\,1]$, we take the derivative of (2.11) with respect to $F(y)$ yielding,

$$\int_{[0,y)}2n\delta F^{n-3}(y)\left[F(y)+(n-2)\left(F(y)-1+\frac{\delta}{2}+\frac{1}{n}\right)\right]dW(t)$$
$$+\int_{[y,\infty)}-F^{-2}(y)dW(t).$$

Since $[F(y) + (n-2)(F(y) - 1 + \delta/2 + 1/n)] \leq 2n$, by choosing δ sufficiently small, the derivative (2.11) can be made negative. This would imply that (2.11) is a decreasing function of $F(y)$ and hence the maximum occurs at $F(y) = \epsilon$. Substituting $F(y) = \epsilon$ in (2.11), we see that (2.11) is negative and (2.9) holds, whenever $\epsilon < 1 - 1/n - \delta/2$. That is, given the weight function W and ϵ, it is possible to choose δ ($\in (0, 2(1-\epsilon-1/n))$) sufficiently small such that (2.9) holds. □

Remark 2.2. The above result for L_2 is surprising since in the uncensored case the sample cdf is minimax for L_2 loss function. It shows that d_{PL} is not minimax for $0 < \epsilon < 1 - 1/n$ regardless of W. On the other hand, if $\epsilon = 1$, then $F(y) = 1$, which implies that all observations are uncensored and hence d_{PL}, which is the same as the empirical survival function (except for $t = y$), is minimax if $dW(\{y\}) = 0$. Thus it is suspected that for $n \geq 2$, d_{PL} may be minimax for $1 - 1/n \leq \epsilon \leq 1$ if $dW(\{y\}) = 0$ (recall that, if $n = 1$, it was shown in Theorem 2.1 that d_{PL} is minimax iff $dW\{y\} = 0$ for any $\epsilon > 0$). But we are unable to establish it as a fact. However, if $dW(\{y\}) > 0$, we have the following result.

PROPOSITION 2.1. *Suppose that* $L = L_2$, $F \in \Theta_\epsilon$, *where* $\epsilon \in (0, 1)$, $n \geq 2$ *and* $dW(\{y\}) > 0$. *Then* d_{PL} *is not minimax. Furthermore,* $\sup_F R(S, d_1) < \sup_F R(S, d_{PL})$, *where*

$$d_1 = \begin{cases} 1-\epsilon & \text{if} \quad t = y < X_{(1)}, \\ d_{PL} & \text{otherwise.} \end{cases}$$

The proof can be found in Phadia and Yu (1989). This result however, is not of much real interest, since usually $dW(\{y\}) = 0$.

THEOREM 2.3. *Let* $n \geq 1$ *and* $F \in \Theta_\epsilon$. *Then* d_{PL} *is not minimax for the following cases*:

(i) $L = L_1$ *and* $\epsilon > (1 - \sqrt{2\sqrt{n}+1}/(\sqrt{n}+1))/2$,

(ii) $L = L_3$ *and* $\epsilon > 0$,

(iii) $L = L_4$ *and* $\epsilon > n/(\sqrt{n}+1)^2$.

In the proof of Theorem 2.3, we seek suitable estimators that beat d_{PL}. For all these cases, the estimator which achieves the objective is given by

$$d(t) = \begin{cases} 1 - \hat{F}_m(t) & \text{for} \quad t < y, \\ 0 & \text{for} \quad t \geq y, \end{cases}$$

where $\hat{F}_m(t)$ is the minimax estimator for the uncensored data (under respective loss function), obtained in Phadia (1973). Note that d is essentially the truncated version of $\hat{F}_m(t)$. As $y \to \infty$, it reduces to the uncensored case. For details, see Phadia and Yu (1989).

Remark 2.3. Note that when $\epsilon = 1$, it reduces to the uncensored case and it is clear that the sample cdf is not minimax for L_1, L_3 and L_4. The above theorem shows that this is true for the PL estimator even in a smaller parameter space Θ_ϵ, with $\epsilon < 1$.

The next theorem establishes the admissibility of the PL estimator.

THEOREM 2.4. *Let $n \geq 1$, $F \in \Theta_\epsilon$ and loss function L given in* (1.1). *Then d_{PL} is admissible.*

PROOF. We want to show that, for any estimator d if $R(S, d) \leq R(S, d_{PL})$ for $\forall F \in \Theta_\epsilon$, then $d(t) = d_{PL}(t)$ except for $t \in [y, t_0]$, where t_0 is a fixed point and $t_0 > y$ with $dW([y, t_0]) = 0$. Once we show this, then it follows that $R(S, d) = R(S, d_{PL})$ for all $F \in \Theta_\epsilon$ (whether $dW([y, t_0]) = 0$ or not) and the admissibility follows.

Let $\Theta_1 = \{F\colon F(y) = 1\}$. Clearly $\Theta_1 \subset \Theta_\epsilon$; then for $F \in \Theta_1$,

$$R(S, d) = \int_{[0,y)} E(S(t) - d(t))^2 h(F(t)) dW(t) + \int_{[y,\infty)} E(d(t))^2 h(F(t)) dW(t)$$

and $R(S, d_{PL}) = \int_{[0,y)} E(S(t) - d_{PL}(t))^2 h(F(t)) dW(t)$. If $R(S, d) \leq R(S, d_{PL})$ for $\forall F \in \Theta_\epsilon$, then clearly,

$$\int_{[0,y)} E(S(t)-d(t))^2 h(F(t)) dW(t) \leq \int_{[0,y)} E(S(t)-d_{PL}(t))^2 h(F(t)) dW(t) \tag{2.12}$$

for $\forall F \in \Theta_1$. Furthermore, given a distribution $F \notin \Theta_1$, i.e., $F(y) \in [0, 1)$ (note that F might not belong to Θ_ϵ), then there is

$$F_1(t) = 1 - S_1(t) = \begin{cases} F(t) & \text{if } \quad t < y \\ 1 & \text{otherwise} \end{cases} \in \Theta_1$$

such that, for any estimator d,

$$\int_{[0,y)} (S_1(t) - d(t))^2 h(F_1(t)) dW(t) = \int_{[0,y)} (S(t) - d(t))^2 h(F(t)) dW(t).$$

It follows that (2.12) is true for all distribution functions (i.e. Θ_0 instead of Θ_ϵ for $\epsilon > 0$). Hence by using the argument of Cohen and Kuo (1985) (since for $t < y$, $1 - d_{PL}$ is exactly the same as the sample cdf), we conclude that $d = d_{PL}$ for $t < y$. Thus we establish the equality (instead of inequality) in (2.12).

Now if $R(S, d) \leq R(S, d_{PL})$ for $\forall F \in \Theta_\epsilon$, then clearly,

$$\int_{[y,\infty)} E(S(t) - d(t))^2 h(F(t)) dW(t) \leq \int_{[y,\infty)} E(S(t))^2 h(F(t)) dW(t)$$

for all $F \in \Theta_\epsilon$. It can be shown (see Phadia and Yu (1989)) that this implies that

$$d = 0 \text{ for } t \geq y \text{ provided that } dW\{[y, t]\} > 0. \tag{2.13}$$

Note that for some $t_1 > y$, if $dW\{[y, t_1)\} = 0$, d may not be equal to $0 = d_{PL}$. But the admissibility is not effected since on the interval $[y, t_1)$ the integrals $\int_{[y,t_1)}(d(t)-S(t))^2 h(F(t))dW(t)$ and $\int_{[y,t_1)}(d_{PL}(t)-S(t))^2 h(F(t))dW(t)$ are both zero (thus (2.13) is slightly stronger than the admissibility result as in Cohen and Kuo (1985)). The admissibility of d_{PL} follows from (2.13) and the fact that $d = d_{PL}$ for $t < y$. □

3. The general case

In this section we remove the restriction of equal censoring times, i.e., $y_1, \ldots, y_n$ are assumed to be arbitrary but fixed censoring times. In addition, we also consider the random censorship model, i.e., Y_i's need not be fixed constants. First, we assume that $Z_i = X_i \wedge y_i$, $i = 1, \ldots, n$, $y_{\min} = \min_i y_i < \max_i y_i = y_{\max}$, $\Theta_\epsilon = \{F\colon F(y_{\min}) \geq \epsilon > 0\}$. Let $\int_{y_{\min}}^{y_{\max}} dW(t) > 0$; otherwise it reduces to the case $y_1 = \cdots = y_n = y$ treated in Section 2.

THEOREM 3.1. *Let $L = L_i$, $i = 1, 2$ and 3 and $y_1, \ldots, y_n$ $(n \geq 2)$ be the censoring times associated with $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$, $F \in \Theta_\epsilon$. Let Z_i, $i = 1, \ldots, n$ be the observable variables. Then d_{PL} is neither minimax nor admissible.*

It is interesting to note that under the random right censorship model, Meeden *et al.* (1989) showed that the PL estimator is admissible under L_1 for Θ_ϵ with $\epsilon = 0$. However, the results in Theorems 3.1 and 3.2 show that the PL estimator is no longer admissible if the parameter space is restricted to Θ_ϵ with $\epsilon > 0$ and for L as in (1.1). This is understandable since the PL estimator takes on values outside the parameter space. In fact, by taking the special advantage of the particular nature of the parameter space (with its dependence on ϵ), we find an estimator which is better than d_{PL}.

PROOF. WLOG, we assume that $y_1 < y_2 < \cdots < y_n$ and $dW\{[y_1, t)\}$ is a strictly increasing function of t in (y_1, s), $s > y_1$. Let $X = (X_1, \ldots, X_n)$. We define an improved estimator as follows.

$$d = d(X, t) = \begin{cases} 1-\delta & \text{if} \quad (X, t) \in A, \\ d_{PL} & \text{otherwise,} \end{cases} \tag{3.1}$$

where $0 < \delta < \epsilon/n$, $A = \{y_1 \leq t < X_{(1)} \wedge y_n\}$ and $X_{(1)}$ is the smallest among the observed uncensored observations. Note that for fixed $t \in (y_i, y_{i+1})$, A is the event such that the first i observations, $Z_1, \ldots, Z_i$ are censored. Therefore $E1(A) = \Pr\{X_j > y_j,\ j \leq i;\ X_k > t,\ k > i\}$ if $t \in (y_i, y_{i+1})$. Note also that d and d_{PL} are the same except on the set A. Thus the difference of risks of these two estimators is

$$\begin{aligned} &R(S, d) - R(S, d_{PL}) \\ &= \sum_{i=1}^{n-1} \int_{y_i}^{y_{i+1}} [(F(t)-\delta)^2 - F^2(t)] h(F(t)) E1(A) dW(t) \end{aligned} \tag{3.2}$$

$$= \sum_{i=1}^{n-1} \int_{y_i}^{y_{i+1}} [(F(t)-\delta)^2 - F^2(t)]h(F(t))$$
$$\cdot \prod_{j=1}^{i}(1-F(y_j))[1-F(t)]^{n-i}dW(t)$$
$$\leq \int_{y_1}^{y_2} [(F(t)-\delta)^2 - F^2(t)]h(F(t))(1-F(y_1))[1-F(t)]^{n-1}dW(t)$$

(the remaining terms are nonpositive anyway)

which is < 0 $\forall F \ni$: $F(y_1) \neq 1$. This implies the inadmissibility of d_{PL}.

To prove the minimaxity result, we first note that

$$R(S,\, d) - R(S,\, d_{PL}) < 0 \ \forall F \ni:\ F(y_1) \neq 1;$$
$$R(S,\, d) - R(S,\, d_{PL}) = 0 \ \forall F \ni:\ F(y_1) = 1.$$

Therefore $\sup_F R(S,\, d) \leq \sup_F R(S,\, d_{PL})$. By Lemma 3.1 below, we have $\sup_{F:F(y_1)=1} R(S,\, d_{PL}) < \sup_F R(S,\, d_{PL})$, i.e. the $\sup_F R(S,\, d_{PL})$ is not achieved at $F(y_1) = 1$. There are two cases: (a) $\exists\, F_s$ such that $R(S_s,\, d) = \sup_F R(S,\, d)$ or (b) otherwise. If (a) is true, then it is easy to see that

$$\sup_F R(S,\, d) < \sup_F R(S,\, d_{PL}). \tag{3.3}$$

On the other hand if (a) is false, it can be shown that there is a sequence of $\{F_m\} \subset \Theta_\epsilon$ such that $\lim_{m\to\infty} R(S_m,\, d) = \sup_F R(S,\, d)$ (where $S_m = 1 - F_m$). Let us denote $G(t) = \lim_{m\to\infty} F_m(t)$, $u = G(y_1)$ and $\lim_{t\downarrow y_1} \lim_{m\to\infty} F_m(t) = v = \lim_{t\downarrow y_1} G(t)$. Since $G(t)$ might not be right continuous at y_1, we have two possibilities: (1) $u \neq v$ and $dW(\{y_1\}) > 0$; or (2) $u = v$ or $dW(\{y_1\}) = 0$.

If (1) is true,

$$\lim_{m\to\infty} [R(S_m,\, d) - R(S_m,\, d_{PL})]$$
$$\leq \lim_{m\to\infty} \int_{\{y_1\}} [(F_m(t)-\delta)^2 - F_m^2(t)]$$
$$\cdot h(F_m(t))(1-F_m(t))[1-F_m(t)]^{n-1}dW(t)$$

(by (3.2) the other terms are nonpositive)

$$= \int_{\{y_1\}} [(u-\delta)^2 - u^2]h(u)(1-u)[1-u]^{n-1}dW(t) < 0, \qquad (u < v \leq 1),$$

and (3.3) is true.

If (2) is true, we proceed as follows. Take a sequence $\{t_m\}$ such that $t_m \downarrow y_1$ and $F_m(t_m) \to v$ as $m \to \infty$. Let

$$F_{m1}(t) = \begin{cases} F_m(t_m) & \text{if} \quad t \in [y_1,\, t_m] \\ F_m(t) & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned}
&\lim_{m\to\infty} R(S_{m1}, d) - \lim_{m\to\infty} R(S_m, d) \\
&\quad = \lim_{m\to\infty} \int_{[y_1, t_m]} [E(S_{m1} - d)^2 h(F_{m1}) - E(S_m - d)^2 h(F_m)] dW(t) \\
&\quad = \int_{\{y_1\}} [(1 - v - d)^2 h(u) - (1 - u - d)^2 h(u)] dW(t) = 0 \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if} \quad dW(\{y_1\}) = 0.
\end{aligned}$$

i.e. if (2) is true, $\lim_{m\to\infty} R(S_{m1}, d) = \lim_{m\to\infty} R(S_m, d) = \sup_F R(S, d)$. This means that we can assume that $u = v$ simply. Then, either $v = 1$ or $v < 1$. If $v = 1$ ($u = v = 1$), $G(t)$ is right continuous at y_1 and equals 1. Also, for $t < y_1$ we note that both $E(d_{PL} - S(t))^2 h(F(t))$ and $E(d - S(t))^2 h(F(t))$ are constant for any $F \in \Theta_\epsilon$. Thus G can be taken to be any distribution function for $t < y_1$ ($E(d_{PL} - (1 - G(t)))^2 h(F(t))$ has the same value anyway). Therefore we can assume that $G(t)$ is a proper distribution and case (a) is again applicable. On the other hand, if $v < 1$, $\exists\, y > y_1$ such that $\lim_{m\to\infty} F_m(y)$ is very close to $\lim_{m\to\infty} F_m(y_1)$, for simplicity, say, equal. Furthermore, without loss of generality, one can assume that $F_m(y) = v$ for all m. Thus, $\exists\, \eta > 0$ such that for all m,

$$\begin{aligned}
&R(S_m, d) - R(S_m, d_{PL}) \\
&\quad \le \int_{y_1}^{y} [(F_m(t) - \delta)^2 - F_m^2(t)] h(F_m(t))(1 - v)[1 - v]^{n-1} dW(t) < -\eta.
\end{aligned}$$

i.e. (3.3) holds if (2) is true. In either case, (3.3) is true, so d_{PL} is not minimax. □

LEMMA 3.1. $\sup_{F:F(y_1)=1} R(S, d_{PL}) < \sup_F R(S, d_{PL})$.

PROOF. It suffices to show that the following statement holds.

(ST) Given $F_2 \in \Theta_\epsilon$ such that $F_2(y_1) = 1$, $\exists$ an $x \in (\epsilon, 1)$ and $F_1 \in \Theta_\epsilon$ such that $F_1(t) = F_1(y_1) = x$ for $t \in [y_1, y_n]$ and $R(S_1, d_{PL}) - R(S_2, d_{PL})$ is a positive constant, independent of the choice of F_2 with $F_2(y_1) = 1$.

WLOG, we assume hereafter, that $F(t) = F(y_1)$ for $\forall t \in [y_1, y_n]$ since we need to demonstrate the above inequality for only one F_1. This would imply that no uncensored observation will be observed with positive probability in the interval $[y_1, y_n]$.

$$\begin{aligned}
(3.4)\qquad R(S, d_{PL}) = &\int_0^{y_1} [F(t)(1 - F(t)) h(F(t))/n] dW(t) \\
&+ \sum_{i=1}^{n-1} E \int_{y_i}^{y_{i+1}} (S(t) - d_{PL})^2 h(F(t)) dW(t) \\
&+ \int_{y_n}^{\infty} (1 - F(t))^2 h(F(t)) dW(t).
\end{aligned}$$

For L_2, first note that $R(S, d_{PL})$ is a constant if $F(y_1) = 1$. Also, the first term on the RHS of (3.4) is a constant and, the third term is zero when $F(y_1) = 1$ and positive otherwise. By Lemma 3.2 below, the second term has the derivative (w.r.t. $F(y_1)$ when $F(t) = F(y_1)\ \forall t \in [y_1, y_n]$), $\sum_{i=1}^{n-1} \int_{y_i}^{y_{i+1}} -i(3n-2i)/n^2 dW(t)$, which is < 0 at $F(y_1) = 1$. Thus, the maximum is achieved for $F(y_1) = x < 1$, i.e. given $F_2 \in \Theta_\epsilon$ and $F_2(y_1) = 1$, $\exists$ an $x \in (\epsilon, 1)$ and $F_1 \in \Theta_\epsilon$ such that $F_1(t) = F_1(y_1) = x$ for $t \in [y_1, y_n]$ and $R(S_1, d_{PL}) - R(S_2, d_{PL})$ is a positive constant, independent of the choice of F_2 with $F_2(y_1) = 1$. Thus the statement (ST) holds.

Similarly, we can show that the statement (ST) holds for the other two loss functions, though the first term of (3.4) is no longer constant there. This completes the proof. □

LEMMA 3.2. *Suppose that $F(t) = F(y_1)$ for $t \in [y_1, y_n]$. The derivative of*

$$E \int_{y_i}^{y_{i+1}} (S(t) - d_{PL})^2 h(F(t)) dW(t) \tag{3.5}$$

with respect to $F(y_1)$ at $F(y_1) = 1$ (i.e., left hand side derivative) is

$$\begin{cases} \int_{y_i}^{y_{i+1}} -(n-i)/n^2 dW(t) & \text{under} \quad L_1, L_3; \\ \int_{y_i}^{y_{i+1}} -i(3n-2i)/n^2 dW(t) & \text{under} \quad L_2; \\ \int_{y_i}^{y_{i+1}} -[(2i-1)(n-i)+ni]/n^2 dW(t) & \text{under} \quad L_4; \end{cases}$$

where $i = 1, \ldots, n-1$.

For the proof of the lemma, see the Appendix.

Remark 3.1. The result of admissibility in Theorem 3.1 can be extended to the case $L = L_4$. However, for the minimaxity result of d_{PL} an additional assumption is needed, for example, $\int_{[0,y_1)} 1/n dW(t) < \int_{[y_1,\infty)} dW(t)$, since $\int_{[0,y_1)} F(t)/n dW(t)$ is increasing in $F(t)$.

THEOREM 3.2. *Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$, $Y_i \sim G_i$ $(i = 1, \ldots, n)$, where $G_i(t)$'s have the support $[y_0, \infty)$ $(y_0 > 0)$ and are continuous, Y_i's are independent of each other and of X_i's. Let $Z_i = X_i \wedge Y_i$ and $\delta_i = 1(X_i \le Y_i)$ $i = 1, 2, \ldots, n$ be the observable random variables. If $F \in \Theta_\epsilon = \{F\colon\ F(y_0) \ge \epsilon > 0\}$ and $L = L_i$, $i = 1, 2$ and 3, then the d_{PL} is neither admissible nor minimax.*

Remark 3.2. The assumption that $G_i(t) = 0$ for $t < y_0$, $i = 1, \ldots, n$ is justified since we want zero probability of observations being censored for values of t close to zero.

PROOF. For the inadmissibility result we need to show that $\exists\, d$ such that $R(S, d) \le R(S, d_{PL})$ for all $F \in \Theta_\epsilon$ and with strict inequality holding for at least one F. For simplicity, assume that the Y_i's are arranged such that $Y_1 \leq Y_2 \leq \cdots \leq Y_n$ and the X_i's correspond to these rearranged Y_i's. Let Y be the random vector induced by the Y_i's. Define d as in (3.1) except that the Y_i's are now random. The difference of the risks is given by

$$
\begin{aligned}
(3.6)\qquad & R(S,\, d) - R(S,\, d_{PL}) \\
&= E_Y\left(\sum_{i=1}^{n}\int_{Y_i}^{Y_{i+1}} [(F(t)-\delta)^2 - F^2(t)]h(F(t)) \right. \\
&\qquad \left. \cdot \Pr\{X_k > Y_k,\, k=1,\ldots,\, i,\, t < X_{(1)}\}dW(t)\right) \\
&= E_Y \sum_{i=1}^{n}\int_{Y_i}^{Y_{i+1}} [(F(t)-\delta)^2 - F^2(t)]h(F(t)) \\
&\qquad \cdot \prod_{k=1}^{i}(1-F(Y_k))(1-F(t))^{n-i}dW(t) \\
&= E_Y \sum_{i=1}^{n}\prod_{k=1}^{i}(1-F(Y_k))\int_{Y_i}^{Y_{i+1}} [(F(t)-\delta)^2 - F^2(t)] \\
&\qquad \cdot h(F(t))(1-F(t))^{n-i}dW(t)
\end{aligned}
$$

which is ≤ 0 for all $F \in \Theta_\epsilon$ and < 0 for all $F(t)$ such that $1 - F(Y_k) \neq 0$ for some $k \in \{1, \ldots, i\}$ and $1 - F(t) > 0$. As in the proof of Theorem 2.1 it is sufficient to consider for verification the case:

(C1) $\quad 1 - F(Y_i) \neq 0$ with positive probability and $1 - F(t) \neq 0$ for $t \geq y_0$.

For this purpose, take

$$
F(t) = \begin{cases} 0 & \text{if} \quad t < y_0, \\ \epsilon + (1-\epsilon)(1-e^{-(t-y_0)}) & \text{otherwise.} \end{cases}
$$

Then clearly $1 - F(t) > 0$ for $t \geq y_0$ and the rest of the verification will be accomplished if we show that $E_{Y_1,Y_2}[1 - F(Y_1)] > 0$. However, $E_{Y_1,Y_2}[1 - F(Y_1)] = \int_0^\infty \int_0^\infty \int_0^{x\wedge y_2} dG_1(y_1)dG_2(y_2)dF(x)$ which is > 0 since G_1 and G_2 are nondegenerate as per our assumption. Thus, we have shown the inadmissibility of d_{PL}.

For the minimaxity, we have additional complexity since the Y_i's are random. However, it can be handled in view of the right continuity of F (i.e. there is a $y_\epsilon > y_0$ such that if $F(y_0) < 1$, then $F(y_\epsilon) < 1$ also) and the continuity of G_i's (i.e. G_i gives positive mass to $(y_0,\, y_\epsilon)$). In this case, with positive probability (G_i's), $1 - F(Y_1) > 0$ and hence (3.6) is < 0. Therefore, as in Theorem 3.1 it is sufficient to show that the $\sup_F R(S,\, d_{PL})$ is not achieved for $F(y_0) = 1$. To show this we proceed as follows.

Consider a realization of $Y = \vec{y}_0$ with the first coordinate y_1. In the proof of Lemma 3.1 it is shown that given F_2 where $F_2(y_1) = 1$, there is an $x \in [\epsilon,\, 1)$ and $F_1 \in \Theta_\epsilon$ such that $F_1(t) = F_1(y_1) = x$ for $t \in [y_1,\, y_n]$ and

$$(3.7)\qquad E(L(F_1,\, 1 - d_{PL}) \mid Y = \vec{y}_0) - E(L(F_2,\, 1 - d_{PL}) \mid Y = \vec{y}_0) = c > 0,$$

where c is a constant, independent of $F_2(y_1)$ with $F_2(y_1) = 1$. There is certainly an $F_1 \in \Theta_\epsilon$ such that $F_1(t) = x$ for $t \in (y_0, y_\epsilon)$. For such an F_1, (3.7) is true for $y_1 \in (y_0, y_\epsilon)$ though c might depend on y_1. Therefore, $R(S_1, d_{PL}) > R(S_2, d_{PL})$. Taking expectation w.r.t. the joint distribution function of $(Y_1, \ldots, Y_n)$ we can show that $\sup_F R(S_1, d_{PL}) > \sup_{F_2:F(y_0)=1} R(S_2, d_{PL})$, concluding the proof. □

Acknowledgements

We are greatly indebted to the editor and referees for their invaluable suggestions and comments.

Appendix

PROOF OF LEMMA 2.1. The risk of d may be expressed as

$$\begin{aligned} R(S,\, d) &= \int_0^{y^-} E\{(S(t) - d(t))^2(1(X \le t] + 1[X > t))/[S(t)(1 - S(t))]\}dW(t) \\ &\quad + \int_{\{y\}} E\{(S(t) - d(t))^2(1(X \le y) + 1(X > y))/[S(t)(1 - S(t))]\}dW(t) \\ &\quad + \int_{y^+}^{\infty} E\{(S(t) - d(t))^2/[S(t)(1 - S(t))]\}dW(t). \end{aligned}$$

In order that the risk be finite, it is clear that for $t > y$, $d(t)$ must be zero, in view of the factor $S(t)$ in the denominator.

Now consider the case, $x \le t < y$. Suppose $\exists\, t_1 < t_2$ such that, $d(x, t_0) > \delta > 0$ if $t_1 = x \le t_0 \le t_2 < y$. Define $F_0(u) = 0$ for $u < t_1$ and $F_0(u) = 1$ for $u \ge t_1$, say, then clearly $E\{[1 - F_0(u) - \delta]^2/[F_0(u)(1 - F_0(u))]\}1(X \le u) = \infty$ for $t_1 < u < t_2$. Thus, in the above case, $d(t)$ should be 0. Similarly, we can show that $d(t) = 1$ if $t < x$ and $t < y$.

For the case, $x \le y = t$, if $d(t) \ne 0$, define $F_2(u) =$ arbitrary for $u < y$ and $F_2(u) = 1$ for $u \ge y$. Then

$$\int_{\{y\}} E\{(S_2(t) - d(t))^2(1(X \le y))/[S_2(t)(1 - S_2(t))]\}dW(t) = \infty \qquad \text{if} \quad dW(\{y\}) > 0.$$

Thus the only case remaining is the case $t = y < x$ where no such "blowing up" occurs as long as $d(y) = b < 1$, proving the characterization. □

PROOF OF LEMMA 3.2. Consider a fixed t such that $y_i \le t < y_{i+1}$, we evaluate

$$E\int_{y_i}^{y_{i+1}} (S(t) - d_{PL})^2 h(F(t))dW(t). \tag{A.1}$$

For $y_i \le t < y_{i+1}$, there are at least i observations $Z_1, \ldots, Z_i$ below t, some of which are censored. Assume that there are exactly k observations among $Z_1, \ldots, Z_i$ that are censored and the remaining $i-k$ are uncensored. Also, some of the observations among $Z_{i+1}, \ldots, Z_n$ may lie below t. Thus, (A.1) is equal to

$$
\begin{aligned}
&E \int_{y_i}^{y_{i+1}} (S(t) - d_{PL})^2 h(F(t)) \\
&\quad \cdot \sum_{k=0}^{i} \sum_{j=i-k}^{n-k} 1\{\text{exactly } k \text{ observations among } Z_1, \ldots, Z_i \text{ are censored} \\
&\quad \text{and } j-(i-k) \text{ observations among } Z_{i+1}, \ldots, Z_n \text{ lie below } t\} dW(t) \\
&\quad = \int_{y_i}^{y_{i+1}} \sum_{k=0}^{i} \sum_{j=i-k}^{n-k} E(S(t) - d_{PL})^2 h(F(t)) 1\{\text{exactly } k \text{ observations among} \\
&\quad Z_1, \ldots, Z_i \text{ are censored}\} 1[X_{(j)} \le t < X_{(j+1)}] 1[X_{(j)} \le y_1] dW(t) \\
&\quad \text{(for details, see Phadia and Yu (1989). Note that } E1[X_{(j)} \le t < X_{(j+1)}] \\
&\quad 1[X_{(j)} > y_1] = 0 \text{ since } F(t) = F(y_1) \forall t \in [y_1, y_n]) \\
&\quad = \int_{y_i}^{y_{i+1}} \sum_{k=0}^{i} \binom{i}{k} (1 - F(y_1))^k \\
&\quad \cdot \left[\sum_{j=i-k}^{n-k-1} (j/n - F(t))^2 \binom{n-i}{j-(i-k)} F^j(y_1)(1 - F(y_1))^{n-k-j} \right. \\
&\quad \left. + (1 - F(t))^2 \binom{n-i}{(n-k)-(i-k)} F^{n-k}(y_1)(1 - F(t))^{n-k-(n-k)} \right] \\
&\quad \cdot h(F(t)) dW(t)
\end{aligned}
$$

(where for the last term in the bracket, we use the fact that $d_{PL} = 0$ when $j = n-k$, i.e. all the mass has been accounted for). Thus, (A.1) equals

$$
\text{(A.2)} \quad \int_{y_i}^{y_{i+1}} \sum_{k=0}^{i} \binom{i}{k} \left[\sum_{j=i-k}^{n-k-1} (j/n - F(t))^2 \binom{n-i}{j-(i-k)} F^j(y_1)(1 - F(y_1))^{n-j} \right.
$$
$$
\left. + F^{n-k}(y_1)(1 - F(t))^{k+2} \right] h(F(t)) dW(t).
$$

(A.2) is true for any $h(t)$. For the sake of saving space, we just give the proof of the lemma for the case $h(t) = t^{-1}(1-t)^{-1}$. In the other three cases of $h(t)$, the similar argument can be used.

Now assume that $h(t) = t^{-1}(1-t)^{-1}$. Then (A.1) or (A.2) equals

$$
\text{(A.3)} \quad \int_{y_i}^{y_{i+1}} \sum_{k=1}^{i} \binom{i}{k} \Bigg[\sum_{j=i-k}^{n-k-1} (j/n - F(t))^2 \cdot \binom{n-i}{j-(i-k)} F^{j-1}(y_1)(1-F(y_1))^{n-j-1} + F^{n-k-1}(y_1)(1-F(t))^{k+1} \Bigg]
$$

$$
+ \Bigg[\sum_{j=i}^{n-2} (j/n - F(t))^2 \binom{n-i}{j-(i)} F^{j-1}(y_1)(1-F(y_1))^{n-j-1} + ((n-1)/n - F(t))^2 (n-i) F^{n-2}(y_1) + F^{n-1}(y_1)(1-F(t))^1 \Bigg] dW(t).
$$

To show that the maximum of (A.1) is not achieved at $F(y_1) = 1$, we take the derivative w.r.t. $F(y_1)$, evaluate at $F(y_1) = 1$ and show that the quantity is negative. Thus the function is strictly decreasing at $F(y_1) = 1$ and Lemma 3.2 follows. Also in taking the derivative it is obvious that we can discard expressions involving higher than one powers of $(1 - F(y_1))$, since they would vanish anyway upon substituting $F(y_1) = 1$. These considerations lead to taking the derivative for $k = 1$ when $j = n - k - 1$ and for $k = 0$ when $j = n - k - 2$ $(\geq i)$, and of the last two terms of (A.3). The resulting expression for the derivative at $F(y_1) = 1$ yields $\int_{y_i}^{y_{i+1}} -i(3n - 2i)/n^2 dW(t)$. $\square$

References

Cohen, M. P. and Kuo, L. (1985). The admissibility of the empirical distribution function, *Ann. Statist.*, **13**, 262–271.

Efron, B. (1967). Two sample problem with censored data, *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, Vol. 4, 831–854, Univ. of California Press, Berkeley.

Ferguson, T. S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data, *Ann. Statist.*, **8**, 226–229.

Ferguson, T. S., Phadia, E. G. and Tiwari, R. C. (1989). Bayesian nonparametric inference, Current Issues in Statistical Inference: Essays in honor of D. BASU, *IMS Lecture Notes-Monograph Ser.*, **17** (1991) (eds. M. Ghosh and P. K. Pathak).

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

Meeden, G., Ghosh, M., Srinivasan, C. and Vardeman, S. (1989). The admissibility of the Kaplan-Meier and other maximum likelihood estimators in the presence of censoring, *Ann. Statist.*, **17**, 1509–1531.

Phadia, E. G. (1973). Minimax estimator of a cumulative distribution function, *Ann. Statist.*, **1**, 1149–1157.

Phadia, E. G. (1980). A note on empirical Bayesian estimation of a distribution function based on censored data, *Ann. Statist.*, **8**, 226–229.

Phadia, E. G. and Yu, Q. (1989). Minimaxity and admissibility of the product limit estimator, Tech. Report No. 89/1, Department of Mathematics, William Paterson College of New Jersey.

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations, *J. Amer. Statist. Assoc.*, **71**, 897–902.

Susarla, V. and Van Ryzin, J. (1978). Empirical Bayesian estimation of a distribution (survival) function from right censored observations, *Ann. Statist.*, **6**, 740–754.