The Annals of Statistics
1980, Vol. 8, No. 3, 673–678

A NOTE ON CONVERGENCE RATES FOR THE PRODUCT LIMIT ESTIMATOR

BY E. G. PHADIA[1] AND J. VAN RYZIN[2]

William Paterson College and Columbia University

In this note, we give a lemma which shows that the expected squared difference between the Bayes estimator with a Dirichlet process prior and the Kaplan-Meier product limit (PL) estimator for a survival function based on censored data is $O(n^{-2})$. This lemma, together with already proven pointwise consistency properties of the Bayes estimator, is used to establish two properties of the PL estimator; namely, the mean square consistency of the PL estimator with rate $O(n^{-1})$ and strong consistency of the PL estimator with rate $o(n^{-\frac{1}{2}} \log n)$.

1. Introduction. The purpose of this note is to establish, as a tool for arriving at the main results, that, based on censored data, the Bayes estimator $\hat{F}_\alpha(u)$ of the survival curve $F(u) \equiv P(X > u)$ given by Susarla and Van Ryzin [9] is pointwise asymptotically close in expected squared distance to the familiar product limit (PL) estimator $\hat{F}_0(u)$ of Kaplan and Meier [7] as modified by Efron [3]; more precisely that $E(\hat{F}_\alpha(u) - \hat{F}_0(u))^2 = O(n^{-2})$. This lemma, together with the known asymptotic properties of $\hat{F}_\alpha(u)$, is used to show two asymptotic properties of the PL estimator: $E(\hat{F}_0(u) - F(u))^2 = O(n^{-1})$ (Theorem 1 in Section 2) and $\hat{F}_0(u) - F(u) = o(n^{-\frac{1}{2}} \log n)$ with probability one (Theorem 2 in Section 2). These highly desirable large sample properties complement and extend the asymptotic normality and weak consistency for the PL estimator established by Breslow and Crowley [2] and the strong consistency of the PL estimator shown by Peterson [8].

Also related to this work are two recent articles called to our attention by a referee. Aalen ([1] Theorem 1) has proved the strong consistency of the PL estimator uniformly on a closed and bounded interval $[0, a]$, $a > 0$ of the real line, with rate $o(n^{-\frac{1}{2}} \log n)$ when the hazard rate $f(u)/(1 - F(u))$ is continuous. This uniform result on $[0, a]$ obviously subsumes our Theorem 2 (Section 2) on $[0, a]$ in the continuous hazard rate case, but does not in the case when the hazard rate is discontinuous since we make no such assumption on F. Földes, Rejtö, and Winter [6] have proven uniform strong convergence results of orders $o(n^{-\frac{1}{4}}(\log n)^{\frac{1}{2}})$ for the PL estimator on the interval $[0, a]$, $a > 0$ and $o(n^{-\frac{1}{8}}(\log n)^{\frac{1}{4}})$ on the whole real line respectively under various restrictions on the underlying distributions of the

---

Received March 1978; revised February 1979.
[1]The work of this author is partially supported by NSF Grant MCS77-01653.
[2]The work of this author is supported by DHEW, PHS, National Institute of Health under Grant 5 R01 GM 23129.
AMS 1970 *subject classifications.* Primary 62G05; secondary 60F99.
Key words and phrases. Product limit estimator, survival distribution, Bayes estimator, rates of convergence, strong consistency, mean square consistency, censored data.

survival and censoring times. Although their results are stronger than ours uniformity wise (ours being a pointwise result), the rates of convergences are considerably slower than ours. Moreover, neither of these two papers addresses the mean square consistency topic.

The proofs of Theorems 1 and 2 are immediate applications of the lemma in Section 2 by using already proven analogous large sample results for the Bayes estimator under squared error loss given in Susarla and Van Ryzin [10, 11]. Thus the essence of this note is to state and prove the lemma, and to use the lemma to show that the PL and Bayes estimators have essentially identical large sample properties in terms of convergence rate for mean squared error and strong consistency.

2. Results. Let $X_1, \cdots, X_n$ be a random sample from a right-sided distribution function (df) $F(u) \equiv P(X > u)$ (also known as the survival function) on $(0, \infty)$, and let $Y_1, \cdots, Y_n$ be another random sample (of censoring variables) from a right-sided continuous df $G(u) \equiv P(Y > u)$ on $(0, \infty)$ such that $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ are mutually independent. The df's F and G may be defined on the real line. However, in keeping with the context of survival analysis, we prefer to restrict them to the positive half of the real line. Also, independence of X_i's and Y_i's may be replaced by a weaker condition $P(\min(X_i, Y_i) > t) = F(t) \cdot G(t)$. Let $\delta_i \equiv [X_i \leqslant Y_i]$, where [ ] stands for the indicator function. Define $Z_i = \min(X_i, Y_i)$, for $i = 1, 2, \cdots, n$. Based on the observable random variables δ_i and Z_i, consider two estimators. First, the familiar PL estimator of F due to Kaplan and Meier [7], as modified by Efron [3] to be "self-consistent," may be written as

$$(1)\qquad \hat{F}_0(u) = \frac{n-i}{n}\prod\left(\frac{n-j+1}{n-j}\right)^{[\delta_{(j)}=0]} \quad \text{if } Z_{(i)} \leqslant u < Z_{(i+1)}, \quad i = 0, 1, 2, \cdots, n-1$$
$$= 0 \qquad \text{if } u \geqslant Z_{(n)},$$

where the symbol $\prod$ (and hereafter) stands for the product taken over $j = 1, 2, \cdots, i$, $Z_{(i)}$ for the ith order statistic of the Z_i's, and $\delta_{(i)}$ for the indicator corresponding to $Z_{(i)}$. Next, consider the Bayes estimator (under an integrated squared error loss function) obtained by Susarla and Van Ryzin [9], where 1-F is assumed to be distributed with a Dirichlet process prior [4] with parameter $\alpha(\cdot)$:

(2)
$$\hat{F}_\alpha(u) = \frac{n-i+\alpha(u)}{n+\alpha(0)}\prod\left(\frac{n-j+\alpha[Z_{(j)}) + 1}{n-j+\alpha[Z_{(j)})}\right)^{[\delta_{(j)}=0]} \quad \text{if } Z_{(i)} \leqslant u < Z_{(i+1)}, \quad i = 0, 1, \cdots n$$

where $\alpha(u) = \alpha((u, \infty)) > 0$, $\alpha[u) = \alpha([u, \infty))$, and $Z_{(n+1)} = \infty$. This result has also been extended to a more general class of priors, processes neutral to the right, by Ferguson and Phadia [5], and they have shown that the estimator (2) is also the posterior mode using the Dirichlet process prior.

In [10], $\hat{F}_\alpha(u)$ was shown to be mean square consistent with rate $O(n^{-1})$ and strongly consistent with rate $o(n^{-\frac{1}{2}} \log n)$ (see also [11]), provided u is such that $\alpha(u) > 0$, $F(u) > 0$ and $G(u) > 0$. Our strategy is to show that under these same conditions $E|\hat{F}_\alpha(u) - \hat{F}_0(u)|^2$ is $O(n^{-2})$ in the lemma which follows and whose proof is in the Appendix. With this lemma, the similar properties for the PL estimator $\hat{F}_0(u)$ requiring only that $G(u) > 0$ follow easily and are given as Theorems 1 and 2.

Before stating and proving our results in detail we discuss the implications of our assumptions.

The assumption of continuity of G rules out ties among the $Z_{(i)}$'s for which $\delta_{(i)} = 0$ in the definitions (1) and (2). For a more general definition of (2) allowing ties, see [9]. The main motivation for not allowing ties is to use the large sample results of Susarla and Van Ryzin [10, 11] wherein they assume G is continuous.

The assumption that $G(u) > 0$ in Theorems 1 and 2 is necessary to identify $F(u)$ uniquely nonparametrically from the data. To see this, observe that if u is such that $G(u) = 0$, then $0 = F(u)G(u) = P(Z_i > u)$, $i = 1, \cdots, n$, no matter what the value of $F(u)$.

The assumption that $\alpha(u) > 0$ in the definition of $\hat{F}_\alpha(u)$ in (2) and in the proof of the lemma is required to guarantee that $\hat{F}_\alpha(u)$ is well defined whenever $Z_{(n)} \leqslant u$ and $\delta_{(n)} = 0$. However, if the convention were made to define $\hat{F}_\alpha(u) = 0$ whenever $Z_{(n)} \leqslant u$ and $\delta_{(n)} = 0$, then $\alpha(u)$ could be taken as zero in (2) and the proof of the lemma (see Appendix). This convention would be similar to the suggestion of Efron [3] for the PL estimator whenever the largest observation is censored. However, we retain the assumption $\alpha(u) > 0$ in our lemma, since this condition was imposed by Susarla and Van Ryzin ([10] page 757) to keep the logarithmic term in their expansion (2.1) well-defined for the largest observation when it is censored. If we adopted the above convention for $\hat{F}_\alpha(u)$, it is clear from the results of this note that $\alpha(u)$ could be taken as zero in the rate results given in Theorems 2.1 and 2.2 of [10, 11]. Furthermore, in regard to Theorem 1 and 2 of this note whose statements do not involve α, it does not matter how strong an assumption is made on α since α only enters in their proofs through the intermediate use of the following lemma.

LEMMA. *For every u for which $\alpha(u) > 0$, $F(u) > 0$ and $G(u) > 0$,*

$$E|\hat{F}_\alpha(u) - \hat{F}_0(u)|^2 = O(n^{-2}).$$

From this lemma, we easily establish the following two theorems.

THEOREM 1. *For every u for which $G(u) > 0$, the* PL *estimator $\hat{F}_0(u)$ is mean square consistent with rate $O(n^{-1})$.*

PROOF. With $F(u) > 0$ and $\alpha(u) > 0$, the result follows immediately from the above lemma and the fact that $\hat{F}_\alpha(u)$ is mean square consistent with rate $O(n^{-1})$ (Theorem 2.2 of [10]) by applying the Minkowski inequality. If $F(u) = 0$, then $Z_{(n)} \leqslant u$ and $\hat{F}_0(u) = 0$ with probability one and the result is trivially true.

THEOREM 2. *For every u for which $G(u) > 0$, the* PL *estimator $\hat{F}_0(u)$ is almost surely consistent with rate $o(n^{-\frac{1}{2}} \log n)$.*

PROOF. By applying the Chebyshev inequality and the above lemma we obtain, for $\varepsilon > 0$,

$$P\left[\frac{n^{\frac{1}{2}}}{\log n}|\hat{F}_\alpha(u) - \hat{F}_0(u)| > \varepsilon\right] \leqslant \frac{C^*}{n(\log n)^2}$$

where C^* is a constant. Now since $\sum_{n=1}^{\infty} n^{-1}(\log n)^{-2} < \infty$, $(\log n)^{-1} n^{\frac{1}{2}}|\hat{F}_\alpha(u) - \hat{F}_0(u)| \to 0$ almost surely by the Borel-Cantelli lemma. The theorem now follows from the triangle inequality and the fact that $\hat{F}_\alpha(u)$ is almost surely consistent with rate $o(n^{-\frac{1}{2}} \log n)$ whenever $F(u) > 0$ and $\alpha(u) > 0$ as shown in Theorem 3.1 of [10] and the note [11]. The case $F(u) = 0$ follows trivially as in Theorem 1.

REMARK. It is easy to see that when all the observations are uncensored, the PL estimator reduces to the empirical distribution function (EDF). Thus it would be natural to compare the asymptotic results obtained here for the censored case with that for the uncensored case. It can easily be verified that the EDF is mean square consistent with best rate $O(n^{-1})$. Also it is well known by the law of the iterated logarithm that the EDF is almost surely pointwise consistent with best rate $O(n^{-\frac{1}{2}}(\log \log n)^{\frac{1}{2}})$. Thus in the case of mean square consistency we have the same rates for the censored as in the uncensored case. However, it is not known whether the same is true for almost sure pointwise consistency. The rate obtained here for the PL estimator is slightly slower than the best corresponding rate for the EDF. Whether our results for the PL estimator and the Bayes estimator can be improved to satisfy the law of the iterated logarithm rate is an open question.

APPENDIX

PROOF OF LEMMA. For $Z_{(i)} \leqslant u < Z_{(i+1)}$, $i = 0, 1, 2, \cdots, n-1$, with $Z_{(0)} = 0$,

$$|\hat{F}_0(u) - \hat{F}_\alpha(u)|^2 = \left|\frac{n-i}{n}\Pi A_j - \frac{n-i+\alpha(u)}{n+\alpha(0)}\Pi B_j\right|^2$$

where

$$A_j = \left(\frac{n-j+1}{n-j}\right)^{[\delta_{(j)}=0]} \quad \text{and} \quad B_j = \left(\frac{n-j+\alpha[Z_{(j)}) + 1}{n-j+\alpha[Z_{(j)})}\right)^{[\delta_{(j)}=0]}$$

for $j = 1, 2, \cdots, i$. Hence,

(3)

$$|\hat{F}_0(u) - \hat{F}_\alpha(u)|^2 \leqslant 2\left|\frac{n-i}{n}(\Pi A_j - \Pi B_j)\right|^2 + 2\left(|\Pi B_j|\left|\frac{n-i}{n} - \frac{n-i+\alpha(u)}{n+\alpha(0)}\right|\right)^2.$$

Using the facts that $\frac{n-i}{n}\Pi A_j$ and $\frac{n-i+\alpha(u)}{n+\alpha(0)}\Pi B_j$ are $\leqslant 1$ for $i = 0, 1, \cdots, n-1$; $B_j \leqslant A_j$ for $j = 1, 2, \cdots, i$; $\Pi A_j - \Pi B_j = \Sigma_{j=1}^{i}\Pi_{l=1}^{j-1}B_l(A_j - B_j)\Pi_{l=j+1}^{i}A_l$; and $(\Sigma_{j=1}^{i}a_j)^2 \leqslant i\Sigma_{j=1}^{i}a_j^2$, we obtain after simplification of (3),

$$(4) \qquad |\hat{F}_0(u) - \hat{F}_\alpha(u)|^2 \leqslant 2i\Sigma_{j=1}^{i}|A_j - B_j|^2 + 2\cdot\left|\frac{(n-i)\alpha(0) - n\alpha(u)}{n(n-i+\alpha(u))}\right|^2.$$

But

$$(5) \qquad i\Sigma_{j=1}^{i}|A_j - B_j|^2 = i\Sigma_{j=1}^{i}[\delta_{(j)} = 0]\left(\frac{\alpha[Z_{(j)})}{(n-j)(n-j+\alpha[Z_{(j)})}\right)^2$$

$$\leqslant \frac{2i}{(n-i)^2}\Sigma_{j=1}^{i}\frac{\alpha^2(0)}{(n-j)(n-j+1)} \leqslant \frac{2i\alpha^2(0)}{(n-i)^3}$$

for $i = 1, 2, \cdots, n-1$. In the last inequality we have used the fact that the sum forms a telescoping series. Therefore, from (4) and (5) and the fact that

$$\left(\frac{(n-i)\alpha(0) - n\alpha(u)}{n(n-i+\alpha(u))}\right)^2 \leqslant \frac{2\alpha^2(0)}{(n-i)^2},$$

we have

$$(6) \qquad |\hat{F}_0(u) - \hat{F}_\alpha(u)|^2 \leqslant \frac{4\alpha^2(0)}{(n-i)^2}\left[\frac{i}{n-i} + 1\right] = \frac{4n\alpha^2(0)}{(n-i)^3}.$$

Now, with $E_i = \{Z_{(i)} \leqslant u < Z_{(i+1)}\}$ and $Z_{(n+1)} = +\infty$, we have

$$(7) \qquad E\{|\hat{F}_0(u) - \hat{F}_\alpha(u)|^2\} = \Sigma_{i=0}^{n-1}E\{|\hat{F}_0(u) - \hat{F}_\alpha(u)|^2|E_i\}\cdot P(E_i)$$
$$+ E\{|\hat{F}_0(u) - \hat{F}_\alpha(u)|^2|E_n\}\cdot P(E_n).$$

The necessity of considering the case $Z_{(n)} \leqslant u$ separately follows from the fact that the PL estimator is zero for $u \geqslant Z_{(n)}$. However, since

$$E\{|\hat{F}_0(u) - \hat{F}_\alpha(u)|^2|E_n\} \leqslant E|\hat{F}_\alpha(Z_{(n)})|^2 \leqslant 1,$$

the second term in (7) is bounded by $P(Z_{(n)} \leqslant u) = (1 - F(u)G(u))^n = \exp[-n\log(1 - F(u)G(u))^{-1}]$. Under the assumption that $F(u)G(u) > 0$, this term goes to zero exponentially fast. Here we have used the fact that $Z_1, \cdots, Z_n$ are i.i.d. with $P(Z_i > u) = F(u)G(u)$.

Also, it follows that the first term in (7) using (6) is

$$\leq \Sigma_{i=0}^{n-1}\left[\frac{4n\alpha^2(0)}{(n-i)^3}\cdot\binom{n}{i}(1-F(u)G(u))^i(F(u)G(u))^{n-i}\right]$$

$$(8)\qquad \begin{aligned} &\leq \frac{4\cdot 4!\,n\alpha^2(0)}{(n+1)(n+2)(n+3)(F(u)G(u))^3} \\ &\quad\times \Sigma_{i=0}^{n-1}\binom{n+3}{i}(1-F(u)G(u))^i(F(u)G(u))^{n-i+3} \\ &\leq \frac{4\cdot 4!\,\alpha^2(0)}{(n+2)(n+3)(F(u)G(u))^3}. \end{aligned}$$

With $F(u)G(u) > 0$, the inequality (8) implies that the first term in (7) is $O(n^{-2})$. Hence the lemma is proved.

REFERENCES

[1] Aalen, Odd (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. *Ann. Statist.* **6** 534–545.

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

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

[4] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

[5] Ferguson, Thomas S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist.* **7** 163–186.

[6] Földes, A., Rejtö, L. and Winter, B. B. (1977). Strong consistency properties of nonparametric estimators for randomly censored data. Part I: The product-limit estimator. Unpublished manuscript.

[7] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–581.

[8] Peterson, Arthur V., Jr. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *J. Amer. Statist. Assoc.* **72** 854–858.

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

[10] Susarla, V. and Van Ryzin, J. (1978). Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples. *Ann. Statist.* **6** 755–768.

[11] Susarla, V. and Van Ryzin, J. (1980). Addendum to "Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples." *Ann. Statist.* **8** 693.

Department of Mathematics
William Paterson College of New Jersey
Wayne, New Jersey 07470

Division of Biostatistics
Columbia University
New York, New York 10032