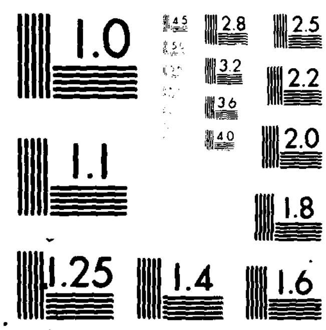
AD-A089 685 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
A NEW METHOD FOR ESTIMATING LIFE DISTRIBUTIONS FROM INCOMPLETE --ETC(U)
MAY 80 J KITCHIN, N A LANGBERG, F PROSCHAN AFOSR-78-3678
UNCLASSIFIED FSU-STATISTICS-M548 AFOSR-TR-80-0932 NL

1 OF 1

END
DATE
FILMED
10-80
DTIC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LEVEL II

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 80-0932	2. GOVT ACCESSION NO. AD-A089 685	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A NEW METHOD FOR ESTIMATING LIFE DISTRIBUTIONS FROM INCOMPLETE DATE		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) John Kitchin, Naftali Langberg, Frank Proschan		8. CONTRACT OR GRANT NUMBER(s) AFOSR 78-3678
9. PERFORMING ORGANIZATION NAME AND ADDRESS Florida State University, Department of Statistics, Tallahassee, FL 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332		12. REPORT DATE May 1980
		13. NUMBER OF PAGES 19
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTE SEP 29 1980 S A

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Piecewise Exponential Estimator, Product Limite Estimator, estimating life distributions, censoring random variables, stron consistency

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We construct a new estimator for a continuous life distribution from incomplete data, the Piecewise Exponential Estimator (PEXE). We show that the PEXE is strongly consisten under a mild restriction on the distribution of the censoring random variables (possible non-identical and non-continuous). Then we consider the Product Limit Estimator (PLE), introduced by Kaplan and Meier (1958). We prove the strong consistency of the PLE under a mild regularity condition on the distributions of the censoring random variables. This result extends previos ones obtained by various researchers. Finally we compare the

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 400277

AD A089685

DDC FILE COPY

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20. Abstract cont.

new PEXE and traditional PLE.

unclassified

18 19
AFOSR-TR-80-0932

6
A NEW METHOD FOR ESTIMATING LIFE DISTRIBUTIONS FROM INCOMPLETE DATA

by

10
John Kitchin, Naftali A. Langberg[1], and Frank Proschan[1]

FSU Statistics Report-M548,
AFOSR Technical Report No. 78-109

12 23

11
May, 1980

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

Dist | Special
A

15
[1]Research ~~supported by the~~ Air Force Office of Scientific Research, AFSC, USAF, under Grant AFOSR-78-3678

AMS Subject Classification Numbers: 62N05, 62G05

80 9 22 103

A New Method for Estimating Life Distributions from Incomplete Data

by

John Kitchin, Naftali A. Langberg, and Frank Proschan

ABSTRACT

We construct a new estimator for a continuous life distribution from incomplete data, the Piecewise Exponential Estimator (PEXE). We show that the PEXE is strongly consistent under a mild restriction on the distribution of the censoring random variables (possibly non-identical and non-continuous). Then we consider the Product Limit Estimator (PLE), introduced by Kaplan and Meier (1958). We prove the strong consistency of the PLE under a mild regularity condition on the distributions of the censoring random variables. This result extends previous ones obtained by various researchers. Finally we compare the new PEXE and traditional PLE.

Key Words: Piecewise Exponential Estimator, Product Limit Estimator, estimating life distributions, censoring random variables, strong consistency.

1. Introduction and Summary.

Let X_1, X_2, ... be independent identically distributed (i.i.d.) lifelengths with a common continuous life distribution G and let L_1, L_2, ... be nonnegative censoring random variables (r.v.'s) not necessarily continuous or i.i.d. We assume that $[X_1, L_1]$, $[X_2, L_2]$, ... is a sequence of independent pairs defined on a common probability space $(\Omega, \mathcal{B}, P)$. Without loss of generality we assume that for each $\omega \in \Omega$ the sequence $X_1(\omega)$, $X_2(\omega)$, ... consists of distinct positive real numbers.

Let I denote the indicator function, let $T_q = \min\{X_q, L_q\}$ be the removal time of the qth item on test, and let $\zeta_q = I(X_q \le L_q)$ denote the cause of removal of the qth item, q = 1, 2, ... We consider the problem of estimating $\bar{G} \equiv 1 - G$, the underlying survival function, from $[T_1, \xi_1]$, ..., $[T_n, \xi_n]$, n = 1, 2, ... This is the incomplete data problem as formulated by Kaplan and Meier (1958). We note that our assumptions on the lifelengths and censoring r.v.'s are less restrictive than those assumed in the theory of competing risks, the theory of life tables, and the usual treatments in biostatistics. Traditionally one or more of the following assumptions have been made: i.i.d. censoring r.v.'s, continuous lifelengths, continuous censoring r.v.'s, and the independence of the lifelength and the corresponding censoring r.v. of each item on test. [Kaplan and Meier (1958), Breslow and Crowley (1974), Peterson (1977), Langberg, Proschan, and Quinzi (1980)].

In Section 2 we construct a new estimator for the underlying survival function $\bar{G}$, the Piecewise Exponential Estimator (PEXE), and denote it by $\bar{E}_n(t)$. In Section 3 we use a theorem proven in Section 4 to show that the PEXE is a strongly consistent estimator of $\bar{G}$ under mild regularity conditions. In particular, we obtain the strong consistency of the PEXE when X_q, L_q are independent r.v.'s, q = 1, 2, ... and under most of the "traditional" assumptions discussed in the previous paragraph. In Section 5 we consider the Product Limit Estimator (PLE), introduced by Kaplan and Meier (1958), that serves as the principal nonparametric estimator to date of the survival function $\bar{G}$. We show that under a variety of conditions the PLE is a strongly consistent estimator of $\bar{G}$. These results extend those obtained by Peterson (1977), and Langberg, Proschan, and Quinzi (1980). Finally, in Section 6 we compare the new PEXE and the traditional PLE.

2. Piecewise Exponential Estimator.

In this section we introduce a new estimator for a continuous life distribution from incomplete data: the Piecewise Exponential Estimator (PEXE).

We approach the incomplete data problem from the viewpoint of reliability and life testing. An item at age zero is placed on life test. It eventually leaves the test either because it fails, yielding a complete life length, or because it is withdrawn while still functioning, yielding a censored lifelength. Thus if an item on test is an observed failure, $\xi_q = 1$, and if it is a withdrawal, $\xi_q = 0$, $q = 1, \ldots, n$. Starting with a sample of initial size n, the number of items at time t is denoted by $N_n(t) = \sum_{q=1}^{n} I(T_q > t)$. Let $\tau(n) = \sum_{q=1}^{n} \xi_q$ denote the number of observed failures and let $Z_{n:1} < \ldots < Z_{n:\tau(n)}$ denote the consecutive observed failure times, with $Z_{n:0} \equiv 0$.

Let $R(t) = -\ln \bar{G}(t)$ denote the hazard function of the life distribution G, $t \in [0, \operatorname{Sup}\{u: \bar{G}(u) > 0\})$. On the interval $(Z_{n:q-1}, Z_{n:q}]$, we estimate $[Z_{n:q} - Z_{n:q-1}]^{-1}\{R(Z_{n:q}) - R(Z_{n:q-1})\}$ by $r_{n,q} = \left[\int_{Z_{n:q-1}}^{Z_{n:q}} N_n(u)du\right]^{-1}$, the number of observed failures per unit time in the interval $(Z_{n:q-1}, Z_{n:q}]$, $q = 1, \ldots, \tau(n)$. We note that $\int_{Z_{n:q-1}}^{Z_{n:q}} N_n(u)du$ is the total time on test in the interval $(Z_{n:q-1}, Z_{n:q}]$, $q = 1, \ldots, \tau(n)$. These hazard slope estimators: $r_{n,q}$, $q = 1, \ldots, \tau(n)$, define a piecewise linear estimator of the hazard function R, which in turn leads to a piecewise exponential estimator of the underlying survival function $\bar{G}$, given explicitly by the following definition.

Definition 2.1. For $\tau(n, \omega) \geq 1$ let $\Delta_{n,q}(\omega) = (Z_{n:q}(\omega) - Z_{n:q-1}(\omega))r_{n:q}(\omega)$, $q = 1, \ldots, \tau(n, \omega)$, $\omega \in \Omega$. Then the Piecewise Exponential Estimator (PEXE) of the survival function $\bar{G}$, denoted by $\bar{F}_n(t, \omega)$, is equal to 1 on the set $\{\tau(n, \omega) = 0$ or $t \in (-\infty, 0]\}$, is equal to $\exp\{-\sum_{j=1}^{q-1} \Delta_{n,j}(\omega) - (t - Z_{n:q-1}(\omega))r_{n,q}(\omega)\}$ on the set $\{\tau(n, \omega) \geq 1, t \in (Z_{n:q-1}(\omega), Z_{n:q}(\omega)]$, $q = 1, \ldots, \tau(n, \omega)\}$, and is equal to $\exp -\sum_{j=1}^{\tau(n,\omega)} \Delta_{n,j}\}$ on the set

$\{\tau(n, \omega) \geq 1,\ t \in (Z_{n:\tau(n,\omega)}(\omega), \infty)$.

For the sake of simplicity we supress the argument ω in $\bar{E}_n(t, \omega)$.

3. Strong Consistency of the PEXE.

In this section we use Theorem 3.2, stated below and proven in Section 4, to obtain the strong consistency of the PEXE under various conditions. We need a definition and some notation.

Definition 3.1. Let K be a function defined on $(-\infty, \infty)$. We say that K is a subdistribution function (s.d.f.) if K is nondecreasing, right continuous, and assumes values in $[0, 1]$.

For a s.d.f. K, let $\bar{K}(x) = \lim_{y\to\infty} K(y) - K(x)$ be the subsurvival function corresponding to K, let $C(K)$ be the set of all continuity points of K, and let $\alpha(K) = \sup\{t: \bar{K}(t) > 0\}$.

We are now ready to state Theorem 3.2.

Theorem 3.2. Assume the following:

(3.1) There is a s.d.f. $F(t)$ such that $\lim_{n\to\infty} n^{-1}\sum_{q=1}^{n} I(T_q > t) = \bar{F}(t)$ for $t \in C(F)$,

and

(3.2) There is a s.d.f. $F(t, 1)$ such that

$$\lim_{n\to\infty} n^{-1}\sum_{q=1}^{n} P\{T_q \le t, \xi_q = 1\} = F(t, 1) \text{ for } t \in (-\infty, \infty).$$

Then there is a set $\Omega_1 \in \mathcal{B}$, $P\{\Omega_1\} = 1$, such that for all $\omega \in \Omega_1$:

$$\lim_{n\to\infty} \bar{E}_n(t) = \exp\{-\int_0^t [\bar{F}(u)]^{-1} dF(u, 1)\} \text{ for } t \in [0, \alpha(F(\cdot, 1))).$$

Now we use Theorem 3.2 to obtain the strong consistency of the PEXE under various conditions. For simplicity, we denote throughout $\alpha(F(\cdot, 1))$ by α_1.

First we prove the strong consistency when the lifelength of each item on test and its censoring r.v. are independent.

<u>Theorem 3.3</u>. Assume the following:

(i) The r.v.'s X_q, L_q are independent for $q = 1, 2, \ldots$,

and

(ii) There is a s.d.f. $H(t)$ such that $\lim_{n\to\infty} n^{-1} \sum_{q=1}^{n} P\{L_q \le t\} = H(t)$ for $t \in C(H)$.

Then there is a set Ω_1, $P\{\Omega_1\} = 1$, such that for all $\omega \in \Omega_1$:

$$\lim_{n\to\infty} \bar{E}_n(t) = \bar{G}(t) \text{ for } t \in [0, \alpha_1).$$

<u>Proof</u>. Let $\bar{F}(t) = \bar{G}(t)\bar{H}(t)$ and $F(t, 1) = \int_0^t \bar{H}(u)dG(u)$ for $t \in [0, \infty)$. To obtain the desired result it suffices, by Theorem 3.2, to verify Conditions (3.1), (3.2) and to show that:

$$\bar{G}(t) = \exp\{-\int_0^t [\bar{F}(u)]^{-1} dF(u, 1)\} \text{ for } t \in [0, \alpha_1). \tag{3.3}$$

First we verify Conditions (3.1), (3.2). Let $t \in (-\infty, \infty)$ and $n = 1, 2, \ldots$. Then by Assumption (i):

$$\lim_{n\to\infty} n^{-1}\sum_{q=1}^{n} P\{T_q > t\} = \bar{G}(t)[n^{-1} \sum_{q=1}^{n} P\{L_q > t\}], \text{ and} \tag{3.4}$$

$$\lim_{n\to\infty} n^{-1}\sum_{q=1}^{n} P\{T_q \le t, \xi_q = 1\} = \int_0^t [n^{-1} \sum_{q=1}^{n} P\{L_q > u\}]dG(u). \tag{3.5}$$

By Assumption (i) and the definition of $\bar{F}$, $C(\bar{F}) = C(H)$. Thus Condition (3.1) follows by Assumptions (i), (ii), and Equation (3.4).

Condition (3.2) follows from Assumptions (i), (ii), Equation (3.5), and the dominated convergence theorem.

Finally we verify Equation (3.3). By the definitions of $\bar{F}(t)$ and $F(t, 1)$, $\int_0^t [\bar{F}(u)]^{-1} dF(u, 1) = \int_0^t [\bar{G}(u)]^{-1} dG(u)$ for $t \in [0, \alpha(F))$. Consequently Equation (3.3) follows by the continuity of G. ||

Next we obtain from Theorem 3.3 a corollary of practical importance.

Assume that most of the "traditional" conditions presented in Section 1 hold (but not necessarily the continuity of the censoring r.v.'s). Then Assumptions (i), (ii) of Theorem 3.3 hold and we obtain:

Corollary 3.4. Assume that the r.v.'s X_q, L_q are independent, $q = 1, 2, \ldots$, and that the r.v.'s L_q, $q = 1, 2, \ldots$, are i.i.d. Then there is a set $\Omega_1 \in \mathcal{B}$, $P\{\Omega_1\} = 1$, such that for all $\omega \in \Omega_1$:

$$\lim_{n\to\infty} \bar{E}_n(t) = \bar{G}(t) \text{ for } t \in [0, \alpha_1).$$

Now we derive the strong consistency of $\bar{E}_n(t)$ when the censoring r.v.'s are i.i.d. but not necessarily independent of the corresponding lifelengths.

Theorem 3.5. Assume that the pairs $[X_q, L_q]$, $q = 1, 2, \ldots$ are i.i.d. Then there is a set $\Omega_1 \in \mathcal{B}$, $P\{\Omega_1\} = 1$, such that for all $\omega \in \Omega_1$, $\lim_{n\to\infty} \bar{E}_n(t) = \bar{G}(t)$ for $t \in [0, \alpha_1)$ iff

$$(3.6) \qquad P\{L_1 > t | X_1 = t\} = P\{L_1 > t | X_1 > t\} \text{ for } t \in [0, \alpha_1).$$

Proof. To obtain the desired result it suffices to show that (a) Conditions (3.1), (3.2) hold and that (b) Conditions (3.3) and (3.6) are equivalent.

(a) Let $F(t) = P\{T_1 \le t\}$ and $F(t, 1) = P\{T_1 \le t, \xi_1 = 1\}$, $t \in (-\infty, \infty)$. Then Conditions (3.1), (3.2) hold trivially.

(b) Let A be a Borel set contained in $[0, \alpha_1)$. Then:

$$(3.7) \qquad P\{T_1 \in A, \xi_1 = 1\} = \int_A P\{L_1 > u | X_1 = u\} dG(u).$$

First we show that Condition (3.3) implies Condition (3.6). By the continuity of G and by Equation (3.7):

$$\int_0^t [\bar{G}(u)]^{-1} dG(u) = -\ln \bar{G}(t) = \int_0^t [\bar{F}(u)]^{-1} dF(u, 1)$$

$$= \int_0^t [\bar{F}(u)]^{-1} P\{L_1 > u | X_1 = u\} dG(u) \text{ for } t \in [0, \alpha_1).$$

Consequently Condition (3.6) follows since $\alpha_1 \leq \alpha(G)$.

Finally we show that Condition (3.6) implies Condition (3.3). By Condition (3.6) and Equation (3.7), we have:

$$(3.8) \qquad F(t, 1) = \int_0^t [\bar{G}(u)]^{-1} \bar{F}(u) dG(u) \text{ for } t \in [0, \alpha_1), \text{ and}$$

$$(3.9) \qquad \int_0^t [\bar{F}(u)]^{-1} dF(u, 1) = \int_0^t [\bar{G}(u)]^{-1} dG(u) \text{ for } t \in [0, \alpha_1).$$

Consequently Condition (3.3) follows from Equation (3.9) and the continuity of G. ||

Note that $\bar{E}_n(t)$ is a strongly consistent estimator of the survival function $\bar{G}$ whenever Conditions (3.1) through (3.3) hold. Finally we provide an example where Conditions (3.1) through (3.3) hold. First we need the following definition.

Definition 3.6. [Marshall and Olkin (1967)]. Let λ_1, λ_2, and $\lambda_{1,2}$ be nonnegative real numbers, $\lambda_1 + \lambda_2 + \lambda_{1,2} > 0$. Then the random pair $[U, V]$ with nonnegative components has a Marshall Olkin Bivariate Exponential Distribution (MOBVED) with parameter $[\lambda_1, \lambda_2, \lambda_{1,2}]$ if for all $t, s \in [0, \infty)$:

$$P\{U > t, V > s\} = \exp\{-\lambda_1 t - \lambda_2 s - \lambda_{1,2} \max(t, s)\}.$$

We now construct the example.

Example 3.7. Let $\lambda_1, \lambda_{1,2}, \gamma_1, \gamma_2, \ldots$ be nonnegative real numbers, $\Lambda_q = \lambda_1 + \lambda_{1,2} + \gamma_q > 0$, $q = 1, 2, \ldots$, and let $[X_q, L_q]$, $q = 1, 2, \ldots$, be a sequence of random pairs having MOBVED's with parameters $[\lambda_1, \gamma_q, \lambda_{1,2}]$ respectively. Assume that $\lim_{n\to\infty} \gamma_n = \gamma \in [0, \infty)$, and that $\lambda_1 + \lambda_{1,2} + \gamma > 0$. We show that under these assumptions, $\bar{E}_n(t)$ is a strongly consistent estimator.

Note that for $t \in [0, \infty)$ and $q = 1, 2, \ldots$:

$$P\{X_q > t\} = \exp\{-(\lambda_1 + \lambda_{1,2})t\},\ P\{T_q > t\} = \exp\{-\Lambda_q t\}, \text{ and}$$

$$P\{T_q > t,\ \xi_q = 1\} = (\lambda_1 + \lambda_{1,2})\Lambda_q^{-1}\exp\{-\Lambda_q t\},$$

Let $\overline{F}(t) = \exp\{-\Lambda t\}$ and $\overline{F}(t, 1) = (\lambda_1 + \lambda_{1,2})\Lambda^{-1}\exp\{-\Lambda t\}$, $t \in [0, \infty)$. Clearly Conditions (3.1) through (3.3) hold. Thus, by Theorem 3.2, $\overline{E}_n(t)$ is a strongly consistent estimator of $\overline{G}$ for $t \in [0, \infty)$.

In particular, if the random pairs $[X_q, L_q]$, $q = 1, 2, \ldots$, are i.i.d. with a MOBVED, then $\overline{E}_n(t)$ is a strongly consistent estimator of $\overline{G}$ for $t \in [0, \infty)$.

4. Proof of Theorem 3.2.

In this section we present a proof of Theorem 3.2. Let $\bar{F}_n(t, \omega) = n^{-1}\sum_{q=1}^{n} I(T_q(\omega) > t)$ and $F_n(t, 1, \omega) =$

$$n^{-1}\sum_{q=1}^{n} I(T_q(\omega) \le t,\ \xi_q(\omega) = 1),\ n = 1, 2, \ldots,\ t \in (-\infty, \infty),\ \omega \in \Omega.$$

For the sake of simplicity in notation we supress throughout the argument ω.

Note that the continuity of the function $\exp\{-\int_0^t [\bar{F}(u)]^{-1} dF(u, 1)\}$ in $[0, \alpha_1)$ follows from the continuity of $F(t, 1)$, which in turn, follows from the continuity of G. Thus to prove the result of Theorem 3.1, it suffices to show by a standard argument [Chung (1974), pp. 132-133] that:

$$\lim_{n\to\infty} \bar{F}_n(t) = \exp\{-\int_0^t [\bar{F}(u)]^{-1} dF(u, 1)\},\ \text{w.p.1. for } t \in [0, \alpha_1). \tag{4.1}$$

First we prove that to obtain Statement (4.1), it suffices to show that:

$$\lim_{n\to\infty} \int_0^t [\bar{F}_n(u)]^{-1} dF_n(u, 1) = \int_0^t [\bar{F}(u)]^{-1} dF(u, 1),\ \text{w.p.1. for } t \in [0, \alpha(F)). \tag{4.2}$$

Then we complete the proof of Theorem 3.2 by verifying Statement (4.2).

We need some more notation and two lemmas. Let $a(n, t) = \max\{k\colon\ k = 0, \ldots, \tau(n),\ Z_{n:k} \le t\}$, $t \in [0, \alpha_1)$, $n = 1, 2, \ldots$, and let $[x]$ denote the largest integer less than or equal to x, $x \in (-\infty, \infty)$.

<u>Lemma 4.1</u>. Assume that Conditions (3.1), (3.2) hold. Then

$$\lim_{n\to\infty} \bar{F}_n(t) = \bar{F}(t),\ \text{w.p.1. for } t \in C(F), \tag{4.3}$$

and

$$\lim_{n\to\infty} F_n(t, 1) = F(t, 1),\ \text{w.p.1. for all } t \in [0, \infty). \tag{4.4}$$

Proof. To prove Statements (4.3) and (4.4) it suffices to show by Conditions (3.1), (3.2) that for all $t \in (-\infty, \infty)$:

$$\lim_{n\to\infty}[F_n(t) - n^{-1}\sum_{q=1}^{n} P\{T_q > t\}] = 0, \text{ w. p. 1., and that}$$

$$\lim_{n\to\infty}[F_n(t, 1) - n^{-1}\sum_{q=1}^{n} F\{T_q \le t, \xi_q = 1\}] = 0, \text{ w.p.1.}$$

The preceding two statements follow by the strong law of large numbers. ||

Lemma 4.2. Assume that Conditions (3.1), (3.2) hold. Then

(4.5) $$\underline{\lim}_{n\to\infty} Z_{n:\tau(n)} > t, \text{ w.p.1. for } t \in [0, \alpha_1).$$

Proof. Let $\underline{\lim}_{n\to\infty} Z_{n:\tau(n)} = a$. For $a = \infty$, Statement (4.5) follows trivially. Assume $a < \infty$; then by Statement (4.4):

$$F(s, 1) = \lim_{m\to\infty} m^{-1} F_m(s, 1) \le \underline{\lim}_{m\to\infty} m^{-1} \sum_{q=1}^{m} p\{\xi_q = 1\}$$

$\le F(a, 1)$ for all $s \in [0, \infty)$, where $\{m\}$ is a subsequence of the positive integers.

Thus $a \ge \alpha_1$. Consequently Statement (4.5) follows. ||

We show now that to prove Statement (4.1), it suffices to verify Statement (4.2).

Lemma 4.3. Assume that Conditions (3.1), (3.2) hold. Then:

(4.6) $$\lim_{n\to\infty} |\ln E_n(t) + \int_0^t [F_n(u)]^{-1} dF_n(u, 1)| = 0, \text{ w.p.1. for } t \in [0, \alpha_1).$$

Proof. Let $\Omega_2 \in \mathcal{B}$, $P\{\Omega_2\} = 1$, and let $t \in [0, \alpha_1)$. Then by Statements (4.3) through (4.5) for every $\omega \in \Omega_2$, there is a positive integer $n(\omega)$ such that $F_n(t) > 0$, $\tau(n) \ge 1$, $\int_0^t [F_n(u)]^{-1} dF_n(u, 1) < \infty$, and $Z_{n:\tau(n)} > t$ for $n \ge n(\omega)$. Let $\omega \in \Omega_2$. Then:

$$(4.7) \qquad [n\bar{F}_n(Z_{n:q-1})]^{-1} \le (Z_{n:q} - Z_{n:q-1})r_{n,q} \le$$

$$[n\bar{F}_n(Z_{n:q})]^{-1} \text{ for } q = 1, \ldots, \tau(n),\ n \ge n(\omega).$$

By Inequality (4.7):

$$-\ell n\bar{E}_n(t) \ge \sum_{q=1}^{a(n,t)} [n\bar{F}_n(Z_{n:q-1})]^{-1} \ge$$

$$\int_0^t [\bar{F}_n(u)]^{-1} dF_n(u, 1) - [n\bar{F}_n(t)]^{-1} \text{ for } n \ge n(\omega).$$

Consequently:

$$(4.8) \qquad \overline{\lim_{n\to\infty}}[\ell n\bar{E}_n(t) + \int_0^t [\bar{F}_n(u)]^{-1} dF_n(u, 1)] \le 0.$$

Further, by Inequality (4.7):

$$-\ell n\bar{E}_n(t) \le \sum_{q=1}^{a(n,t)+1} [n\bar{F}_n(Z_{n:q})]^{-1} \le \int_0^t [\bar{F}_n(u)]^{-1} dF_n(u, 1)$$

$$+ [n\bar{F}_n(Z_{n:a(n,t)+1})]^{-1}.$$

By Condition (3.2) there is a $\delta \in (0, \alpha_1 - t)$ and a positive integer n_1 such that $a(n, t) + 1 \le nF(t + \delta, 1)$ for $n \ge n_1$. Thus:

$$[n\bar{F}_n(Z_{n:a(n,t)+1})]^{-1} \le [n(1 - n^{-1}(a(n, t) + 1))]^{-1}$$

$$\le [n(1 - F(t + \delta, 1))]^{-1} \le [n\bar{F}(t + \delta, 1)]^{-1} \text{ for } n \ge n_1.$$

Consequently:

$$(4.9) \qquad \underline{\lim_{n\to\infty}} [\ell n\bar{E}_n(t) + \int_0^t [\bar{F}_n(u)]^{-1} dF_n(u, 1)] \ge 0.$$

Statement (4.6) follows now from Inequalities (4.8) and (4.9). ||

Let $\beta = \text{Sup}\{F(t, 1): t \in [0, \infty)\}$, $t \in [0, \alpha(F))$, and $b(n, t) = \max\{\tau(n), a(n, t) + 1\}$. Assume that the r.v.'s $L_1, \ldots, L_2, \ldots$ are i.i.d. Then by considering the cases $F(t, 1) = \beta$ and $F(t, 1) < \beta$, it follows that $\underline{\lim}_{n\to\infty} \bar{F}_n(Z_{n:b(n,t)}) > 0$. Thus if the r.v.'s $L_1, L_2, \ldots$ are i.i.d. Statement (4.6) holds for $t \in [0, \alpha(F))$.

We complete the proof of Theorem 3.2 by verifying Statement (4.2).

Lemma 4.4. Assume that Conditions (3.1), (3.2) are satisfied. Then Statement (4.2) holds.

Proof. Let $F_n(u-, 1) = n^{-1}\sum_{q=1}^{n} P\{T_q < t, \xi_q = 1\}$, $u \in (-\infty, \infty)$, and let $t \in [0, \alpha(F))$. Upon integration by parts:

$$\int_0^t [\bar{F}_n(u)]^{-1} dF_n(u, 1) = -\int_0^t F_n(u-, 1) d[\bar{F}_n(u)]^{-1}$$

$$+ [\bar{F}_n(t)]^{-1} F_n(t, 1) - [\bar{F}_n(0)]^{-1} F_n(0, 1) =$$

$$-\int_0^t \{F_n(u-, 1) - F(u, 1)\} d[\bar{F}_n(u)]^{-1} - \int_0^t F(u, 1) d[\bar{F}_n(u)]^{-1}$$

$$+ [\bar{F}_n(t)]^{-1} F_n(t, 1) - [\bar{F}_n(0)]^{-1} F_n(0, 1).$$

Thus, by the continuity of $F(\cdot, 1)$, and upon integration by parts:

$$(4.10) \qquad \int_0^t [\bar{F}_n(u)]^{-1} dF_n(u, 1) = \int_0^t [\bar{F}_n(u)]^{-1} dF(u, 1) +$$

$$[\bar{F}_n(t)]^{-1}\{F_n(t, 1) - F(t, 1)\} - [\bar{F}_n(0)]^{-1}\{F_n(0, 1) - F(0, 1)\}$$

$$- \int_0^t \{F_n(u-, 1) - F(u, 1)\} d[\bar{F}_n(u)]^{-1}.$$

Next note that by Statement (4.3), the continuity of $F(u, 1)$, and the dominated convergence theorem:

$$(4.11) \qquad \lim_{n\to\infty} \int_0^t [\bar{F}_n(u)]^{-1} dF(u, 1) = \int_0^t [\bar{F}(u)]^{-1} dF(u, 1), \text{ w.p.1.},$$

that by Statements (4.3), (4.4):

(4.12) $$\lim_{n\to\infty}[\bar{F}_n(t)]^{-1}\{F_n(t, 1) - F(t, 1)\} = 0, \text{ and}$$

$$\lim_{n\to\infty}[\bar{F}_n(0)]^{-1}\{F_n(0, 1) - F(0, 1)\} = 0, \text{ w.p.1.},$$

and that by the continuity of $F(\cdot, 1)$:

(4.13) $$\lim_{n\to\infty} \text{Sup}\{|F_n(u-, 1) - F(u, 1)|, u \in [0, t]\} = 0, \text{ w.p.1.}$$

Consequently the desired result follows by Statements (4.10) through (4.13). ||

Finally, assume that the r.v.'s $L_1, L_2, \ldots$, are i.i.d. Then by the remark following Lemma 4.3, and by Lemma 4.4, Statement (4.1) holds for $t \in [0, \alpha(F))$. Consequently if the r.v.'s $L_1, L_2, \ldots$ are i.i.d. then Corollary 3.4, holds for $t \in [0, \alpha(F))$.

5. Strong consistency of the PLE.

In this section we first prove that as $n \to \infty$ the PLE converges to $\exp\{-\int_0^t [\bar{F}(u)]^{-1} dF(u, 1)\}$ for $t \in [0, \alpha(F))$ Then we use this result to obtain the strong consistency of the PLE under various conditions.

Let $\bar{M}_n(t)$ denote the PLE. Then by the continuity of G, $\bar{M}_n(t)$ is given by

$$\bar{M}_n(t) = \begin{cases} 1 & , \tau(n) = 0 \text{ or } t \in (-\infty, 0), \\ \prod_{j=1}^{q-1} [n\bar{F}_n(Z_{n:j})][n\bar{F}_n(Z_{n:j}) + 1]^{-1}, & \\ & t \in [Z_{n:q-1}, Z_{n:q}),\ q = 1, \ldots, \tau(n), \\ \prod_{j=1}^{\tau(n)} [n\bar{F}_n(Z_{n:j})][n\bar{F}_n(Z_{n:j}) + 1]^{-1}, & \tau(n) \geq 1,\ t \in [Z_{n:\tau(n)}, \infty). \end{cases}$$

Note that originally Kaplan and Meier (1958) left the PLE undetermined on the set $(\max\{T_q, q = 1, \ldots, n\}, \infty)$.

We prove now that $\bar{M}_n(t)$ converges.

Theorem 5.1. Assume that Conditions (3.1), (3.2) hold. Then there is a set $\Omega_1 \in \mathcal{B}$, $P\{\Omega_1\} = 1$, such that for all $\omega \in \Omega_1$:

$$\lim_{n\to\infty} \bar{M}_n(t) = \exp\{-\int_0^t [\bar{F}(u)]^{-1} dF(u, 1)\} \text{ for } t \in [0, \alpha(F)).$$

Proof. Note that the continuity of the function $\exp\{-\int_0^t [\bar{F}(u)]^{-1} dF(u, 1)\}$ in $[0, \alpha(F))$ follows from the continuity of $F(t, 1)$ which, in turn, follows from the continuity of G. Thus to prove the desired result it suffices to show by a standard argument [Chung (1974), pp. 132-133] that:

$$\text{(5.1)} \qquad \lim_{n\to\infty} \bar{M}_n(t) = \exp\{-\int_0^t [\bar{F}(u)]^{-1} dF(u, 1)\}, \text{ w.p.1. for } t \in [0, \alpha(F)).$$

To prove Statement (5.1), it suffices to show by Lemma 4.4 that:

$$(5.2) \qquad \lim_{n\to\infty} \left| \ln \overline{M}_n(t) + \int_0^t [F_n(u)]^{-1} dF_n(u, 1) \right| = 0 \text{ w.p.1. for } t \in [0, \alpha(F)).$$

We prove now Statement (5.2). Note that:

$$(5.3) \qquad x^{-1} \geq -\ln x(1 + x)^{-1} \geq (x + 1)^{-1} \text{ for } x \in (0, \infty).$$

By the definition of $\overline{M}_n(t)$ and Inequality (5.3), for $t \in [0, \alpha(F))$:

$$-\ln \overline{M}_n(t) \leq \sum_{q=1}^{a(n,t)} [nF_n(Z_{n:q})]^{-1} = \int_0^t [F_n(u)]^{-1} dF_n(u, 1), \text{ w.p.1.}$$

Thus

$$(5.4) \qquad \underline{\lim}_{n\to\infty} [\ln \overline{M}_n(t) + \int_0^t [F_n(u)]^{-1} dF_n(u, 1)] \geq 0, \text{ w.p.1.}$$

Further, by the definition of $\overline{M}_n(t)$ and Inequality (5.3), for $t \in [0, \alpha(F))$:

$$-\ln \overline{M}_n(t) \geq \sum_{q=1}^{a(n,t)} [nF_n(Z_{n:q}) + 1]^{-1} \geq$$

$$\sum_{q=1}^{a(n,t)} [nF_n(Z_{n:q-1})]^{-1} \geq \int_0^t [F_n(u)]^{-1} dF_n(u, 1)$$

$$-[nF_n(t)]^{-1}, \text{ w.p.1.}$$

Thus, by Condition (3.1):

$$(5.5) \qquad \overline{\lim}_{n\to\infty} [\ln \overline{M}_n(t) + \int_0^t F_n(u)]^{-1} dF_n(u, 1)] \leq 0, \text{ w.p.1.}$$

Statement (5.2) follows now by Inequalities (5.4) and (5.5). ||

Note that for $t \in [0, \alpha(F))$, $\underline{\lim}_{n\to\infty} \max\{T_q, q = 1, \ldots, n\} > t$. Thus Statement 5.2 holds regardless how we define $\overline{M}_n(t)$ on the set: $(\max\{T_q, q = 1, \ldots, n\}, \infty)$.

It follows from Theorem 5.1 that if we replace in Theorems 3.3, 3.5, and Corollary 3.4, $E_n(t)$ by $\bar{M}_n(t)$ and α_1 by $\alpha(F)$, the results remain valid. Thus we obtain the strong consistency of the PLE under a variety of Conditions. In particular, we obtain the strong consistency of the PLE when X_q, L_q are independent r.v.'s, $q = 1, 2, \ldots$, and under most of the "traditional" assumptions.

Finally we note that these results extend those obtained by Peterson (1977), and by Langberg, Proschan, and Quinzi (1980).

6. A Comparison of the Piecewise Exponential Estimator and the Product-Limit Estimator.

In this final section we point out some differences and similarities between the PEXE and the PLE.

The most obvious difference between the two estimators is that the PEXE is continuous and strictly decreasing on $[0, Z_{n:\tau(n)})$ while the PLE is a step function with jumps at the observed failures. Since in most life testing situations the survival function is anticipated to be decreasing smoothly over time, the PEXE seems the more appropriate estimator of a life distribution.

Another difference which favors the PEXE is its dependence on the actual withdrawal times in each interval between consecutive observed failures (through the total time on test) compared with the PLE's dependence on only the <u>number</u> of withdrawals in each of the intervals. The PEXE uses more information from the incomplete data than does the PLE.

It is clear from Sections 4 and 5 that the PEXE and the PLE have the same strong (w.p.1) limiting function. Also the PEXE has the same weak limiting process as that given for the PLE by Breslow and Crowley (1974). [A derivation of this result is forthcoming.] Consequently, finite sample comparisons of the PEXE and the PLE will be important in determining whether the differences cited in the previous paragraph (which disappear in the limit) result in quantitative advantages for the PEXE over the PLE. Chen, Hollander, and Langberg (1980) are conducting such a study. They assume that the restrictive assumptions discussed in Section 1 hold and that, in addition, $P\{X_1 > t\} = [P\{L_1 > t\}]^\rho$ for $t \in [0, \infty)$, where ρ is a positive real number.

REFERENCES

(1) Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censoring. Ann. Statist. 2, 437-453.

(2) Chen, Y. Y., Hollander, M., and Langberg, N. A. (1980). The small sample properties of the Product Limit and the Piecewise Exponential Estimators. [In process].

(3) Chung, K. L. (1974). A Course in Probability Theory. Academic Press, Inc., New York, Second Edition.

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

(5) Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. J. Amer. Statist. Assoc., 62, 30-44.

(6) Langberg, N. A., Proschan, F., and Quinzi, A. J. (1980). Estimating dependent life lengths, with applications to the theory of competing risks. Ann. Statist. (to appear in Nov. 1980).

(7) Peterson, A. V. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. J. Amer. Statist. Assoc. 72, 854-858.

(8) Rao, C. R. (1973). Linear Statistical Inference and Its Application. Second Edition. John Wiley and Sons, New York.

UNCLASSIFIED
Security Classification of this Page

REPORT DOCUMENTATION PAGE

1. REPORT NUMBERS FSU No. M548 AFOSR No. 78-109	2. GOVT. ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE A New Method for Estimating Life Distributions from Incomplete Data		5. TYPE OF REPORT AND PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M548
7. AUTHOR(s) John Kitchin, Naftali Langberg, Frank Proschan		8. CONTRACT OR GRANT NUMBER(s) AFOSR-78-3678
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Florida State University Department of Statistics Tallahassee, Florida 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA AND WORK UNIT NOS.
10. CONTROLLING OFFICE NAME AND ADDRESS The U.S. Air Force Air Force Office of Scientific Research Bolling Air Force Base, D.C. 20332		12. REPORT DATE May, 1980 13. NUMBER OF PAGES 19
14. MONITORING AGENCY NAME AND ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Piecewise Exponential Estimator, Product Limit Estimator, estimating life distributions, censoring random variables, strong consistency

20. ABSTRACT

We construct a new estimator for a continuous life distribution from incomplete data, the Piecewise Exponential Estimator (PEXE). We show that the PEXE is strongly consistent under a mild restriction on the distribution of the censoring random variables (possibly non-identical and non-continuous). Then we consider the Product Limit Estimator (PLE), introduced by Kaplan and Meier (1958). We prove the strong consistency of the PLE under a mild regularity condition on the distributions of the censoring random variables. This result extends previous ones obtained by various researchers. Finally we compare the new PEXE and traditional PLE.