The Annals of Statistics
1985, Vol. 13, No. 1, 142–149

A COMBINATORIC APPROACH TO THE KAPLAN–MEIER ESTIMATOR

BY DAVID MAURO

Trinity College

The paper considers the Kaplan–Meier estimator F_n^{KM} from a combinatoric viewpoint. Under the assumption that the estimated distribution F and the censoring distribution G are continuous, the combinatoric results are used to show that $\int |\theta(z)| \, dF_n^{\text{KM}}(z)$ has expectation not larger than $\int |\theta(z)| \, dF(z)$ for any sample size n. This result is then coupled with Chebychev's inequality to demonstrate the weak convergence of the former integral to the latter if the latter is finite, if F and G are strictly less than 1 on $\mathscr{R}$ and if θ is continuous.

1. Introduction. Suppose $\{X_i, C_i \mid i \in \mathscr{N}\}$ is a set of independent, real valued random variables on some complete probability space $(\Omega, \mathscr{F}, P)$. Suppose also that the C_i's and the X_i's have unknown cumulative distribution functions G and F, respectively. If θ is a known Borel measurable function on $\mathscr{R}$ such that the expected value of $\theta(X_1)$ (hereafter denoted $E[\theta(X_1)]$) is finite, the statistician often estimates this expectation by observing $X_i = x_i$, $1 \le i \le n$, then constructing the unbiased estimate

$$\sum_{i=1}^n \theta(x_i)/n \equiv \sum_{i=1}^n \theta(x_i) \, d\hat{F}_n(x_i) \tag{1.1}$$

where $\hat{F}_n$ is the usual empirical cumulative distribution function of X_1. However, under various circumstances (such as those described by Chen, Hollander and Langberg, 1982), the statistician may be constrained to observe only the right censored sample

$$M_i \equiv \min(X_i, C_i), \quad \delta_i \equiv \begin{Bmatrix} 1 & X_i \le C_i \\ 0 & X_i > C_i \end{Bmatrix}, \quad 1 \le i \le n.$$

Since some X_i's may be censored (i.e. have associated indicator 0), alternative estimators of F and $E[\theta(X_1)]$ must be employed.

One such estimator of F was introduced in 1958 by Kaplan and Meier. If $M_{(i)}$ is defined to be the ith order statistic of the n minima and if $\delta_{(i)}$ is the indicator associated with $M_{(i)}$, then the Kaplan–Meier estimator of F at x is given by

$$F_n^{\text{KM}}(x) \equiv \begin{cases} 1 - \prod_{i: M_{(i)} \le x} ((n-i)/(n-i+1))^{\delta_{(i)}} & \text{if } x \le M_{(n)} \\ 1 & \text{if } x > M_{(n)}, \quad \delta_{(n)} = 1 \\ \text{undefined} & \text{if } x > M_{(n)}, \quad \delta_{(n)} = 0. \end{cases}$$

The estimator F_n^{KM} is not well defined if there are censored observations tied with uncensored observations. While the convention under this condition is to treat the censored observations as infinitesimally larger than the uncensored

Received April, 1983; revised September, 1983.
AMS 1980 *subject classifications.* Primary 60C05, 62G05, 62G30; secondary 62G99.
Key words and phrases. Censored, Kaplan–Meier estimator.

ones with which they are tied, this issue will be avoided by assuming that F and G are continuous.

The estimator F_n^{KM} is everywhere (P) a right continuous, increasing step function having nonzero jumps at x if and only if x is an uncensored observation. The magnitude of the jump at x, denoted $dF_n^{\mathrm{KM}}(x)$ and referred to as the Kaplan–Meier weight of x, is $1/n$ if x equals $M_{(1)}$ and x is uncensored. Otherwise, if x is an uncensored observation, the Kaplan–Meier weight of x can be seen to depend only upon the order of the censored and uncensored minima that are smaller than x. It is also readily verifiable that (i) any two consecutive uncensored minima have the same Kaplan–Meier weight, and (ii) the sum of the n Kaplan–Meier weights is one or less than one as per whether $\delta_{(n)}$ is one or zero.

These facts, in turn, spawn other properties as well as the useful recursion formula

$$dF_n^{\mathrm{KM}}(M_{(j)}) = \frac{1 - \sum_{i=1}^{j-1} dF_n^{\mathrm{KM}}(M_{(i)})}{n - j + 1} \quad \text{if} \quad \delta_{(j)} = 1 \quad \text{and} \quad j \geq 2.$$

Large sample properties of the Kaplan–Meier estimator have come under study by various authors. Peterson (1977) has established its strong consistency. Breslow and Crowley (1974) as well as Meier (1975) have established the weak convergence of the estimator regarded as a stochastic process. Miller (1976) has considered this estimator in the context of censored regression.

In the case of F and G having support in the right tail of $\mathscr{R}$, properties of $\int \theta(z)\, dF_n^{\mathrm{KM}}(z)$ have remained elusive. Susarla and Van Ryzin (1980) have shown the strong consistency of the mean estimator $\int_0^{A_n} 1 - \hat{F}(z)\, dz$, where $\hat{F}$ is their Bayes generalization of F_n^{KM} and $\langle A_n \rangle$ is a sequence of constants requiring calculation. Gill (1983) has studied whole line properties of the Kaplan–Meier estimator from a stochastic calculus stance, applying the results to q functions and positive random variable mean estimation.

In Section 4 of this paper, it is shown that the censored data analogue of (1.1)

$$\sum_{i=1}^{n} \theta(M_i)\, dF_n^{\mathrm{KM}}(M_i) \left(\equiv \int \theta(z)\, dF_n^{\mathrm{KM}}(z) \right) \tag{1.2}$$

converges (P) to the correct value if support for F and G can be found in the right tail of $\mathscr{R}$ and if θ is continuous. Sections 2 and 3 (which require neither the support nor the continuity constraints) respectively establish the main combinatoric result and the bound on the expectation of (1.2).

2. Combinatoric aspects of the Kaplan–Meier weight function.

Begin with the establishment of some notation and definitions.

DEFINITION 2.1. The modified Kaplan–Meier weight of a real number x is

$$d\tilde{F}_n^{\mathrm{KM}}(x) \equiv \begin{cases} 1 - \sum_{i=1}^{n-1} dF_n^{\mathrm{KM}}(M_{(i)}) & \text{if } x = M_{(n)} \text{ and } \delta_{(n)} = 0 \\ 0 & \text{if } x > M_{(n)} \\ dF_n^{\mathrm{KM}}(x) & \text{elsewhere.} \end{cases}$$

Besides being defined at every real x, this weight function differs from dF_n^{KM} in that positive mass is assignd to $M_{(n)}$ whether or not that observation is censored. In particular, $d\tilde{F}_n^{\text{KM}}(M_{(n)})$ can be found by calculating $dF_n^{\text{KM}}(M_{(n)})$ under the pretense that $M_{(n)}$ is uncensored. This, in turn, implies that the sum of the n modified Kaplan–Meier weights is 1. (Equivalently, $d\tilde{F}_n^{\text{KM}}$ can be considered the weight function of a modified Kaplan–Meier estimator $\tilde{F}_n^{\text{KM}}$ whose value if $\delta_{(n)} = 0$ and $x \geq M_{(n)}$ is defined to be 1.)

DEFINITION 2.2. For $1 \leq i \leq n$, let Y_i denote X_i and Y_{i+n} denote C_i. Let $\{\sigma^j \mid 1 \leq j \leq (2n)!\}$ be the set of permutations of the first $2n$ positive integers. Then

$$\sigma^j(\Omega) \equiv \{\omega \in \Omega \mid Y_{\sigma^j(1)} < Y_{\sigma^j(2)} < \cdots < Y_{\sigma^j(2n)}\}.$$

Since the ordering of the $2n$ Y_k's is fixed on each $\sigma^j(\Omega)$, $dF_n^{\text{KM}}(M_i)$ and $d\tilde{F}_n^{\text{KM}}(M_i)$ are constant on $\sigma^j(\Omega)$. For the sake of notational convenience, these will be denoted $W_i(\sigma^j)$ and $\tilde{W}_i(\sigma^j)$, respectively.

Now consider the $(2n)_n$ n-tuples whose distinct entries are elements of $\{1, 2, \cdots, 2n\}$. Arbitrarily label them 1 through $(2n)_n$.

DEFINITION 2.3. If the yth such n-tuple is $(t_1, t_2, t_3, \cdots, t_n)$, then $S_y \equiv \{\sigma^j \mid \sigma^j(t_1) = 1, \sigma^j(t_2) = 2, \cdots, \sigma^j(t_n) = n\}$ and $S_y(\Omega) \equiv \{\sigma^j(\Omega) \mid \sigma^j \in S_y\}$.

Simply put, $S_y(\Omega)$ is a set of subsets $\sigma^j(\Omega)$ of Ω such that the positions of the C_i's vary from subset to subset, while the positions of the X_i's remain fixed.

EXAMPLE. Suppose $n = 3$. Suppose the yth such 3-tuple is (1, 6, 4). Then $S_y = \{\sigma^j \mid \sigma^j(1) = 1, \sigma^j(6) = 2, \sigma^j(4) = 3\} = \{(1, 4, 5, 3, 6, 2), (1, 5, 4, 3, 6, 2), (1, 4, 6, 3, 5, 2), (1, 6, 4, 3, 5, 2), (1, 5, 6, 3, 4, 2), (1, 6, 5, 3, 4, 2)\}$. Further, $S_y(\Omega) = \{\{\omega \mid Y_1 < Y_4 < Y_5 < Y_3 < Y_6 < Y_2\}, \{\omega \mid Y_1 < Y_5 < Y_4 < Y_3 < Y_6 < Y_2\}, \{\omega \mid Y_1 < Y_4 < Y_6 < Y_3 < Y_5 < Y_2\}, \{\omega \mid Y_1 < Y_6 < Y_4 < Y_3 < Y_5 < Y_2\}, \{\omega \mid Y_1 < Y_5 < Y_6 < Y_3 < Y_4 < Y_2\}, \{\omega \mid Y_1 < Y_6 < Y_5 < Y_3 < Y_4 < Y_2\}\}$.

Clearly, $|S_y| = n!$ and $\{S_y \mid 1 \leq y \leq (2n)_n\}$ partitions $\{\sigma^j \mid 1 \leq j \leq (2n)!\}$.

Sufficient notation now exists that the main combinatoric result can be easily stated. The proof (by induction) will follow a small body of lemmas and propositions dealing with the two weight formulae.

THEOREM 2.1. *For each* y, $1 \leq y \leq (2n)_n$, *and each* i, $1 \leq i \leq n$, $\sum_{\sigma^j \in S_y} \tilde{W}_i(\sigma^j) = (n-1)!$.

This theorem states that the sum of the modified Kaplan–Meier weights (associated with observation M_i) taken over the orderings of an arbitrary set S_y in the partition of $\{\sigma^j \mid 1 \leq j \leq (2n)!\}$ is $(n-1)!$. That this sum is invariant to S_y should, in fact, be expected if the random variable $\sum_{i=1}^n \theta(X_i)\, d\tilde{F}_n^{\text{KM}}(M_i)$ is to have expectation equal to $E[\theta(X_1)]$ in our distribution-free context (the proofs of Lemma 3.2 and Theorem 3.1 will bear this out.) In fact, Lemma 3.2 and its application in the proof of Theorem 3.1 will hold true under any censored data weight function that satisfies Theorem 2.1.

Theorem 2.1 can be illustrated via the last example: select i, $1 \le i \le 3$, calculate the modified Kaplan–Meier weight of M_i for each element of $S_y(\Omega)$ and note that the sum of the six weights is $(3-1)! = 2$.

LEMMA 2.1. *Select y such that on any $\sigma^j(\Omega)$ in $S_y(\Omega)$, X_u is the pth smallest and X_v is the $(p+1)$th smallest among the Y_k's (that is, $\sigma^j(p) = u$ and $\sigma^j(p+1) = v$). Then $\sum_{\sigma^j \in S_y} \tilde{W}_u(\sigma^j) = \sum_{\sigma^j \in S_y} \tilde{W}_v(\sigma^j)$.*

PROOF. Partition S_y into four sets:

$$s^1 \equiv \{\sigma^j \in S_y \mid M_u = X_u \text{ and } M_v = X_v \text{ on } \sigma^j(\Omega)\}$$

$$s^2 \equiv \{\sigma^j \in S_y \mid M_u = X_u \text{ and } M_v = C_v \text{ on } \sigma^j(\Omega)\}$$

$$s^3 \equiv \{\sigma^j \in S_y \mid M_u = C_u \text{ and } M_v = X_v \text{ on } \sigma^j(\Omega)\}$$

$$s^4 \equiv \{\sigma^j \in S_y \mid M_u = C_u \text{ and } M_v = C_v \text{ on } \sigma^j(\Omega)\}.$$

For any permutation σ^j in s^1, M_u and M_v are consecutive and uncensored in the ordering of the n minima on $\sigma^j(\Omega)$. Hence, $\tilde{W}_u(\sigma^j) = \tilde{W}_v(\sigma^j)$, giving

PROPOSITION 2.1. $\sum_{\sigma^j \in s^1} \tilde{W}_u(\sigma^j) = \sum_{\sigma^j \in s^1} \tilde{W}_v(\sigma^j)$.

Now let g be the bijection from s^2 to s^3 such that on $g(\sigma^j)(\Omega)$, the ordering of the Y_k's is identical to the ordering of the Y_k's on $\sigma^j(\Omega)$ but for a reversal of C_u and C_v. Then the configuration of censored and uncensored minima which precede the uncensored M_u on $\sigma^j(\Omega)$ is identical to the configuration of censored and uncensored minima which precede the uncensored M_v on $g(\sigma^j)(\Omega)$. Thus $\tilde{W}_u(\sigma^j) = \tilde{W}_v(g(\sigma^j))$, giving

$$\sum_{\sigma^j \in s^2} \tilde{W}_u(\sigma^j) = \sum_{\sigma^j \in s^2} \tilde{W}_v(g(\sigma^j)) = \sum_{\sigma^j \in s^3} \tilde{W}_v(\sigma^j). \tag{2.1}$$

Moreover, $M_{(n)}$ is neither C_u nor C_v on $\sigma^j(\Omega)$ for σ^j in $s^2 \cup s^3$. Hence $\sum_{\sigma^j \in s^3} \tilde{W}_u(\sigma^j) = \sum_{\sigma^j \in s^2} \tilde{W}_v(\sigma^j) = 0$, which together with (2.1) gives

PROPOSITION 2.2. $\sum_{\sigma^j \in s^2 \cup s^3} \tilde{W}_u(\sigma^j) = \sum_{\sigma^j \in s^2 \cup s^3} \tilde{W}_v(\sigma^j)$.

Finally, consider the following partition of s^4:

$$s_1^4 \equiv \{\sigma^j \in s^4 \mid M_{(n)} = C_u \quad \text{on} \quad \sigma^j(\Omega)\}$$

$$s_2^4 \equiv \{\sigma^j \in s^4 \mid M_{(n)} = C_v \quad \text{on} \quad \sigma^j(\Omega)\}, \quad s_3^4 \equiv s^4 - s_1^4 - s_2^4.$$

Then

$$\sum_{\sigma^j \in s_2^4 \cup s_3^4} \tilde{W}_u(\sigma^j) = \sum_{\sigma^j \in s_1^4 \cup s_3^4} \tilde{W}_v(\sigma^j) = 0. \tag{2.2}$$

Now extend the domain of g to include s_1^4. It is easily seen that g is bijective from s_1^4 to s_2^4. Further, if σ^j is in s_1^4, the configuration of censored and uncensored minima which precede $M_{(n)} \equiv C_u$ on $\sigma^j(\Omega)$ is identical to the configuration of said minima which precede $M_{(n)} \equiv C_v$ on $g(\sigma^j)(\Omega)$. Thus

$$\sum_{\sigma^j \in s_1^4} \tilde{W}_u(\sigma^j) = \sum_{\sigma^j \in s_1^4} \tilde{W}_v(g(\sigma^j)) = \sum_{\sigma^j \in s_2^4} \tilde{W}_v(\sigma^j).$$

By (2.2), then, we have

PROPOSITION 2.3. $\sum_{\sigma^j\in s^4} \tilde{W}_u(\sigma^j) = \sum_{\sigma^j\in s^4} \tilde{W}_v(\sigma^j)$.

The lemma closes with the combining of Propositions 2.1, 2.2 and 2.3.

PROOF OF THEOREM 2.1. Suppose that y is arbitrarily fixed, $1 \leq y \leq (2n)_n$. With no loss of generality, suppose $X_1 < X_2 < \cdots < X_n$ on $\sigma^j(\Omega)$ in $S_y(\Omega)$, with X_i being the p_ith smallest score among the Y_k's. Then it suffices to prove the initializing claim.

(i) $\sum_{\sigma^j\in S_y} \tilde{W}_1(\sigma^j) = (n-1)!$

and the conclusion to the inductive hypothesis

(ii) if $\sum_{\sigma^j\in S_y} \tilde{W}_i(\sigma^j) = (n-1)!$ for $1 \leq i \leq m-1 < n-1$, then $\sum_{\sigma^j\in S_y} \tilde{W}_m(\sigma^j) = (n-1)!$

PROOF OF (i). By hypothesis, $p_1 - 1$ censoring variables precede X_1 in the ordering of the Y_k's on $\sigma^j(\Omega)$ in $S_y(\Omega)$.

Case 1. If there exists σ^j in S_y such that $M_{(n)} = C_1$ on $\sigma^j(\Omega)$, then $X_1, X_2, \cdots, X_n$ must be the $(n+1)$th through $(2n)$th smallest of the Y_k's, respectively. Hence the X_i's are consecutive on $\sigma^j(\Omega)$ and thus on all elements of $S_y(\Omega)$. By Lemma 2.1, then,

$$\sum_{\sigma^j\in S_y} \tilde{W}_1(\sigma^j) = \sum_{\sigma^j\in S_y} \tilde{W}_2(\sigma^j) = \cdots = \sum_{\sigma^j\in S_y} \tilde{W}_n(\sigma^j).$$

Since

$$\sum_{i=1}^{n} \tilde{W}_i(\sigma^j) = 1 \quad \text{and} \quad |S_y| = n! \tag{2.3}$$

each of the above summations must be $(n-1)!$.

Case 2. Suppose there exists no σ^j in S_y such that $M_{(n)} = C_1$ on $\sigma^j(\Omega)$. Then, for each σ^j in S_y, $\tilde{W}_1(\sigma^j) = W_1(\sigma^j)$.

Since $p_1 - 1$ censoring variables C_i precede X_1 on any $\sigma^j(\Omega)$ in $S_y(\Omega)$, then on exactly $(p_1 - 1)[(n-1)!]$ such sets $\sigma^j(\Omega)$, C_1 precedes X_1. The Kaplan–Meier weight of $M_1 = C_1$ on these sets is thus 0.

On the remaining $n! - (p_1 - 1)[(n-1)!]$ elements of $S_y(\Omega)$, $M_1 = X_1$ is the p_1th smallest minimum, preceded by $p_1 - 1$ censored minima. The Kaplan–Meier weight of M_1 is therefore $1/(n - p_1 + 1)$. Hence

$$\sum_{\sigma^j\in S_y} \tilde{W}_1(\sigma^j) = [n! - (p_1 - 1)[(n-1)!]][1/(n - p_1 + 1)] = (n-1)!$$

PROOF OF (ii).

Case 1. Suppose $m = n$. Since $\sum_{i=1}^{n-1} \sum_{\sigma^j\in S_y} \tilde{W}_i(\sigma^j) + \sum_{\sigma^j\in S_y} \tilde{W}_m(\sigma^j) = n!$, the inductive hypothesis gives the result.

Case 2. Let $m < n$. Let y' be such that on any $\sigma^j(\Omega)$ in $S_y'(\Omega)$, X_i is the p_ith smallest score of the Y_k's for $1 \leq i \leq m$ and $X_m, X_{m+1}, \cdots, X_n$ are consecutive. Define a bijection h from S_y to S_y' such that, if the C_i's are ordered in a given way on $\sigma^j(\Omega)$ in $S_y(\Omega)$, then those variables are ordered in that same way on $h(\sigma^j)(\Omega)$. The order of the Y_k's less than or equal to X_m on $\sigma^j(\Omega)$ in $S_y(\Omega)$ is thus identical to the ordering of the Y_k's less than or equal to X_m on $h(\sigma^j)(\Omega)$, giving

$$\sum_{\sigma^j \in S_y} \tilde{W}_i(\sigma^j) = \sum_{\sigma^j \in S_{y'}} \tilde{W}_i(\sigma^j) \quad \text{for} \quad 1 \le i \le m. \tag{2.4}$$

By the inductive hypothesis, therefore,

$$\sum_{\sigma^j \in S_{y'}} \tilde{W}_i(\sigma^j) = (n-1)! \quad \text{for} \quad 1 \le i \le m-1. \tag{2.5}$$

However, by Lemma 2.1,

$$\sum_{\sigma^j \in S_{y'}} \tilde{W}_m(\sigma^j) = \sum_{\sigma^j \in S_{y'}} \tilde{W}_{m+1}(\sigma^j) = \cdots = \sum_{\sigma^j \in S_{y'}} \tilde{W}_n(\sigma^j).$$

The above summations must thus be $(n-1)!$ by (2.3) and (2.5). Hence $\sum_{\sigma^j \in S_y} \tilde{W}_m(\sigma^j) = (n-1)!$ by (2.4).

3. An application. Theorem 2.1 is here used in proving

THEOREM 3.1. *$E[|\theta(X_1)|] \ge E[\sum_{i=1}^n |\theta(M_i)| \, dF_n^{\mathrm{KM}}(M_i)]$.*

LEMMA 3.1. *If σ^q and σ^r are in S_y, then $\int |\theta(X_i)| I(\sigma^q(\Omega)) \, dP$ equals $\int |\theta(X_i)| I(\sigma^r(\Omega)) \, dP$.*

The proof of this lemma is a matter of the independence of the Y_k's, the respective common distribution functions of the X_i's and the C_i's, and Fubini's theorem. Details are omitted.

LEMMA 3.2. *Let u and v be arbitrarily selected integers, $1 \le u, v \le n$. Then $E[|\theta(X_u)| \, d\tilde{F}_n^{\mathrm{KM}}(M_u)] = E[|\theta(X_u)| \, d\tilde{F}_n^{\mathrm{KM}}(M_v)]$.*

PROOF.

$$\begin{aligned}
E[|\theta(X_u)| \, d\tilde{F}_n^{\mathrm{KM}}(M_u)] &\equiv \int |\theta(X_u)| \, d\tilde{F}_n^{\mathrm{KM}}(M_u) \, dP \\
&= \sum_{y=1}^{(2n)_n} \sum_{\sigma^j \in S_y} \int |\theta(X_u)| I(\sigma^j(\Omega)) \, d\tilde{F}_n^{\mathrm{KM}}(M_u) \, dP \\
&= \sum_{y=1}^{(2n)_n} \sum_{\sigma^j \in S_y} \tilde{W}_u(\sigma^j) \int |\theta(X_u)| I(\sigma^j(\Omega)) \, dP \\
&\qquad \text{(since } d\tilde{F}_n^{\mathrm{KM}}(M_u) \text{ is constant on } \sigma^j(\Omega)) \\
&= \sum_{y=1}^{(2n)_n} \left[\int |\theta(X_u)| I(\sigma^j(\Omega)) \, dP \sum_{\sigma^j \in S_y} \tilde{W}_u(\sigma^j) \right] \\
&\qquad \text{(by Lemma 3.1)} \\
&= \sum_{y=1}^{(2n)_n} \left[\int |\theta(X_u)| I(\sigma^j(\Omega)) \, dP \sum_{\sigma^j \in S_y} \tilde{W}_v(\sigma^j) \right] \\
&\qquad \text{(by Theorem 2.1)} \\
&= \int |\theta(X_u)| \, d\tilde{F}_n^{\mathrm{KM}}(M_v) \, dP \equiv E[|\theta(X_u)| \, d\tilde{F}_n^{\mathrm{KM}}(M_v)].
\end{aligned}$$

Before proceeding with the proof of Theorem 3.1, the reader is urged to keep in mind that $|\theta(M_i)|\, d\tilde{F}_n^{\mathrm{KM}}(M_i)$ and $|\theta(X_i)|\, d\tilde{F}_n^{\mathrm{KM}}(M_i)$ are two different things. The former is always observable. The latter is not observable if and only if $M_i = M_{(n)}$ and M_i is censored.

PROOF OF THEOREM 3.1. Select t, $1 \le t \le n$. Then

$$E[|\theta(X_t)|] = E[|\theta(X_t)| \textstyle\sum_{i=1}^n d\tilde{F}_n^{\mathrm{KM}}(M_i)] = \sum_{i=1}^n E[|\theta(X_t)|\, d\tilde{F}_n^{\mathrm{KM}}(M_i)]$$

$$= \textstyle\sum_{i=1}^n E[|\theta(X_t)|\, d\tilde{F}_n^{\mathrm{KM}}(M_t)] \quad \text{(by Lemma 3.2)}$$

$$= nE[|\theta(X_t)|\, d\tilde{F}_n^{\mathrm{KM}}(M_t)\},$$

giving $(1/n)E[|\theta(X_t)|] = E[|\theta(X_t)|\, d\tilde{F}_n^{\mathrm{KM}}(M_t)]$. By the arbitrariness of t, then,

$$(1/n) \textstyle\sum_{i=1}^n E[|\theta(X_i)|] = \sum_{i=1}^n E[|\theta(X_i)|\, d\tilde{F}_n^{\mathrm{KM}}(M_i)].$$

However, the X_i's are identically distributed. The left side of the above equality is thus $E[|\theta(X_1)|]$, giving

$$E[|\theta(X_1)|] = \textstyle\sum_{i=1}^n E[|\theta(X_i)|\, d\tilde{F}_n^{\mathrm{KM}}(M_i)]. \tag{3.1}$$

Moreover, $0 \le dF_n^{\mathrm{KM}}(M_i) \le d\tilde{F}_n^{\mathrm{KM}}(M_i)$ everywhere (P). This implies

$$E[|\theta(X_1)|] \ge \textstyle\sum_{i=1}^n E[|\theta(X_i)|\, dF_n^{\mathrm{KM}}(M_i)]$$

The latter expectation is, though, $\sum_{i=1}^n E[|\theta(M_i)|\, dF_n^{\mathrm{KM}}(M_i)]$ since $X_i = M_i$ if and only if $dF_n^{\mathrm{KM}}(M_i)$ does not equal zero.

4. THEOREM 4.1. *Suppose Γ is a continuous function on $\mathscr{R}$ such that $E(\Gamma(X_1)) < \infty$. Suppose that the suprema of points of support of F and G are ∞. Then*

$$\textstyle\sum_{i=1}^n \Gamma(M_i)\, dF_n^{\mathrm{KM}}(M_i) \to_P E(\Gamma(X_1)).$$

PROOF. Select ε, $\delta > 0$. Select real b such that $E(|\Gamma(X_1)I(b < X_1 < \infty)|) < \varepsilon\delta/6$. Since $b < \infty$ and Γ is continuous, it is fairly easy to show

$$\textstyle\sum_{i=1}^n \Gamma(M_i)I(-\infty < M_i \le b)\, dF_n^{\mathrm{KM}}(M_i) \to_{\mathrm{a.e}} \displaystyle\int_{-\infty}^{b} \Gamma(z)\, dF(z). \tag{4.1}$$

Hence it so converges in probability, giving

$$P\Big(\Big|\textstyle\sum_{i=1}^n \Gamma(M_i)I(-\infty < M_i \le b)\, dF_n^{\mathrm{KM}}(M_i) - \displaystyle\int_{-\infty}^{b} \Gamma(z)\, dF(z)\Big| < \varepsilon/3\Big) > 1 - \delta/2 \tag{4.2}$$

for all n greater than some positive integer N_0. Further, we have

$$\varepsilon\delta/6 > E(|\Gamma(X_1)I(b < X_1 < \infty)|)$$

$$\ge E(\textstyle\sum_{i=1}^n |\Gamma(M_i)I(b < M_i < \infty)|\, dF_n^{\mathrm{KM}}(M_i)) \qquad \text{(by Theorem 3.1)}$$

$$\ge E(|\textstyle\sum_{i=1}^n \Gamma(M_i)I(b < M_i < \infty)\, dF_n^{\mathrm{KM}}(M_i)|).$$

This is true for any n. By Chebychev, then, for any n,

$$\begin{aligned} & P(|\textstyle\sum_{i=1}^{n} \Gamma(M_i)I(b < M_i < \infty)\, dF_n^{\text{KM}}(M_i)| \leq \varepsilon/3) \\ \text{(4.3)} \qquad & \geq 1 - E(|\textstyle\sum_{i=1}^{n} \Gamma(M_i)I(b < M_i < \infty)\, dF_n^{\text{KM}}(M_i)|)/(\varepsilon/3) \\ & > 1 - \delta/2. \end{aligned}$$

Thus, by (4.2) and (4.3), for $n > N_0$,

$$P(|\textstyle\sum_{i=1}^{n} \Gamma(M_i)\, dF_n^{\text{KM}}(M_i) - \displaystyle\int_{-\infty}^{b} \Gamma(z)\, dF(z)| < 2\varepsilon/3) > 1 - \delta.$$

But $\varepsilon/3 > \delta\varepsilon/6 > E(|\Gamma(X_1)I(b < X_1 < \infty)|) \geq |E(\Gamma(X_1)I(b < X_1 < \infty))| \equiv |\int_b^\infty \Gamma(z)\, dF(z)|$. So, for $n > N_0$,

$$P(|\textstyle\sum_{i=1}^{n} \Gamma(M_i)\, dF_n^{\text{KM}}(M_i) - \displaystyle\int_{-\infty}^{\infty} \Gamma(z)\, dF(z)| < \varepsilon) > 1 - \delta.$$

5. Final comments. It is felt that the results of this paper could be extended to the case of noncontinuous F and/or G. The combinatoric arguments would be virtually identical in spirit to those given, but would necessarily include the messy bookkeeping chore of accounting for ties.

Also, it is noted that (i) Theorem 4.1 holds for any function Γ (continuous or not) such that (4.1) holds for any finite b, and (ii) Lemma 3.1, Lemma 3.2 and result (3.1) hold in the absence of absolute value signs.

REFERENCES

CHEN, Y., HOLLANDER, M. and LANGBERG, N. (1982). Small sample results for the Kaplan-Meier estimator. *J. Amer. Statist. Assoc.* **77** 141–144.

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

GILL, R. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

KAPLAN, E. and MEIER, P. (1958). Non-parametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

MEIER, P. (1975). Estimation of a distribution function from incomplete observations. *Perspectives in Probability and Statistics* 67–87. Ed. J. Gani. Applied Probability Trust, Sheffield, England.

MILLER, R. (1976). Least squares regression with censored data. *Biometrika* **63** 449–464.

PETERSON, A. (1977). Expression the Kaplan-Meier estimator as a function of empirical subsurvival functions. *J. Amer. Statist. Assoc.* **72** 854–858.

SUSARLA, V. and VAN RYZIN, J. (1980). Large sample theory for an estimator of the mean survival time from censored samples. *Ann. Statist.* **8** 1002–1016.

MATHEMATICS DEPARTMENT
TRINITY COLLEGE
HARTFORD, CONNECTICUT 06106