This Week's Citation Classic

Kaplan E L & Meier P. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assn.* **53**:457-81, 1958.
[Univ. California Radiation Laboratory, CA and University of Chicago, IL]

The product-limit formula estimates the proportion of organisms or physical devices surviving beyond any age t, even when some of the items are not observed to die or fail, and the sample is rather small. The actuarial and reduced-sample methods are also studied. [The *Science Citation Index*® (*SCI*®) and the *Social Sciences Citation Index*® (*SSCI*®) indicate that this paper has been cited in over 1,495 publications since 1961.]

Edward L. Kaplan
Department of Mathematics
Oregon State University
Corvallis, OR 97331

April 15, 1983

"This paper began in 1952 when Paul Meier at Johns Hopkins University (now at the University of Chicago) encountered Greenwood's paper[1] on the duration of cancer. A year later at Bell Telephone Laboratories I became interested in the lifetimes of vacuum tubes in the repeaters in telephone cables buried in the ocean. When I showed my manuscript to John W. Tukey, he informed me of Meier's work, which already was circulating among some of our colleagues. Both manuscripts were submitted to the *Journal of the American Statistical Association*, which recommended a joint paper. Much correspondence over four years was required to reconcile our differing approaches, and we were concerned that meanwhile someone else might publish the idea.

"The nonparametric estimate specifies a discrete distribution, with all the probability concentrated at a finite number of points, or else (for a large sample) an actuarial approximation thereto, giving the probability in each of a number of successive intervals. This paper considers how such estimates are affected when some of the lifetimes are unavailable (*censored*) because the corresponding items have been lost to observation, or their lifetimes are still in progress when the data are analyzed. Such items cannot simply be ignored because they may tend to be longer-lived than the average.

"The result is that every item r has an age t_r' associated with it, but some of the t_r' correspond to deaths (or failures), and some to *losses* (from observation). Now let the t_r' be listed and labeled in order of increasing magnitude, so that one has $0 \leqslant t_1' \leqslant t_2' \leqslant \ldots t_N'$. Then the product-limit estimate of the proportion surviving beyond the age t is the product $\hat{P}(t) = \Pi_r[(N-r)/(N-r+1)]$, where r assumes those values (from 1 to N) such that $t_r' \leqslant t$ and t_r' corresponds to a death.

"This means that to each age t′ at which a death is observed, there is assigned the probability $\hat{P}(t'-0) - \hat{P}(t'+0)$, the amount by which the step-function $\hat{P}(t)$ decreases at the point t′. The most intuitive derivation of the formula is as a product of the conditional survival probabilities $(N-r)/(N-r+1)$, which would be obtained by the actuarial method if the intervals were made so short that each contained only one death.

"Meier's 1975 paper[2] obtained asymptotic properties for $\hat{P}(t)$ considered as a stochastic process. Three more recent publications[3-5] are listed below.

"Presumably this paper is frequently cited because it gives a good presentation of a simple solution to a problem often encountered by researchers. (It has also been used in a seminar intended to introduce students to the use of the literature.) Similar objectives have motivated my recent book, *Mathematical Programming and Games*."[6]

1. **Greenwood M.** *The natural duration of cancer.* London, England: His Majesty's Stationery Office, 1926. Reports on Public Health and Medical Subjects, No. 33.
2. **Meier P.** Estimation of a distribution function from incomplete observations. (Gani J, ed.) *Perspectives in probability and statistics.* New York: Academic Press, 1975. p. 67-87.
3. **Kalbfleisch J D & Prentice R L.** *The statistical analysis of failure time data.* New York: Wiley, 1980. 321 p.
4. **Efron B.** Censored data and the bootstrap. *J. Amer. Statist. Assn.* **76**:312-19, 1981.
5. **Chen Y Y, Hollander M & Langberg N A.** Small-sample results for the Kaplan-Meier estimator. *J. Amer. Statist. Assn.* 77:141-4, 1982.
6. **Kaplan E L.** *Mathematical programming and games.* New York: Wiley, 1982. 588 p.