The Annals of Statistics
1983, Vol. 11, No. 3, 989–993

THE ESTIMATION OF THE HAZARD FUNCTION FROM RANDOMLY CENSORED DATA BY THE KERNEL METHOD[1]

BY MARTIN A. TANNER AND WING HUNG WONG

University of Wisconsin and University of Chicago

By convolution smoothing of the empirical hazards, a kernel estimate of the hazard function from censored data is obtained. Small and large sample expressions for the mean and the variance of the estimator are given. Conditions for asymptotic normality are investigated using the Hajek projection method.

1. Introduction. In life testing, medical follow up and other studies, the observation of the occurrence of the event of interest (called a failure) may be prevented by the previous occurrence of another event (called a censoring event). Thus, if $T_1, \cdots, T_n$ are the lifetimes for the n items under study, and $C_1, \cdots, C_n$ the corresponding censoring times, then it is not possible to observe both T_i and C_i. Instead, we can only observe $X_i = \min(T_i, C_i)$ and $\delta_i = I_{[T_i \le C_i]}$ (here I_A denotes the indicator of the event A). In this paper we assume the random censorship model: $T_1, \cdots, T_n$ are i.i.d. with cdf F_T, independent of $C_1, \cdots, C_n$ which are i.i.d. with cdf F_C. The cdf and density of X_i will be denoted by F and f respectively (without any subscript).

Let $X_{(1)}, \cdots, X_{(n)}$ be the ordered X's, $\delta_{(1)}, \cdots, \delta_{(n)}$ be the corresponding indicators, and R_j be the rank of X_j. The purpose of this paper is to investigate the properties of

$$\hat{\lambda}_T(x) = \sum_{j=1}^n (n-j+1)^{-1}\delta_{(j)}K_h(x - X_{(j)}) = \sum (n - R_i + 1)^{-1}\delta_i K_h(x - X_i)$$

as an estimator of the failure rate $\lambda_T(x) = -(d/dx)\log(1 - F_T(x))$ which is assumed to be continuous. Here K is a symmetric nonnegative kernel, $K(t) = o(t^{-1})$ as $t \to \infty$, $\int K(t)\,dt = 1$, $K_h(y) = h^{-1}K(y/h)$. The point of interest x is assumed fixed throughout the study and satisfying $0 < F(x) < 1$.

The estimate $\hat{\lambda}_T$ can be regarded as a convolution smoothing of the formal derivative of the empirical cumulative hazards $\hat{H}(x) = \sum_{X_i \le x} a_i$ where $a_i = \delta_i/$(No. of items at risk at time X_i) $= \delta_i/(N - R_i + 1)$. In Section 2 we derive the expectation and variance of $\hat{\lambda}_T$, and provide asymptotic expressions for them. These results generalize those of Watson and Leadbetter (1964) (henceforth referred to as WL) for the uncensored case. In Section 3, conditions for asymptotic normality of $\hat{\lambda}_T(x)$ is established by the Hajek projection technique.

The above estimator has also been considered, independently of the present work and of each other, by Ramlau-Hansen (1983) and Yandell (1983). The former, based on the theory of multiplicative intensity counting process, is perhaps the most general. The latter, while dealing only with the random censorship model, provides simultaneous confidence bands. Both analyses, however, require that all the observations fall into a certain compact interval, and that the kernel has compact support. Such requirements are not necessary in our approach. It is hoped that the result of the present analysis, which is based on more elementary arguments, will complement those of the other two papers.

Received March, 1982; revised December, 1982.

[1] Support for this research was provided by National Science Foundation Grant No. MCS-8101836, and for Tanner, the research was also supported in part by the United States Army under Contract No. DAAG29-80-C-0041.

AMS 1980 *subject classification.* 62G05, 62P10.

Key words and phrases. Censored data, hazard function, kernel method, Hajek's projection method.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

2. Expressions for mean and variance. Let $m(y) = f_T(y)(1 - F_C(y))/f(y)$ if $f(y) > 0$, then

LEMMA 1. *$E(\delta_{(j)} | X_{(j)} = y) = m(y)$ for all j; and*

$$E(\delta_{(r)}\delta_{(s)} | X_{(r)} = y, X_{(s)} = z) = m(y)m(z) \quad \text{for all} \quad r < s, \quad y < z.$$

THEOREM 1. $E\hat{\lambda}_T(x) = \int (1 - F^n(y))\lambda_T(y)K_h(x - y)\, dy,$

$$\mathrm{Var}(\hat{\lambda}_T(x)) = \int I_n(F(y))\lambda_T(y)K_h^2(x - y)\, dy + 2\int\int_{y\le z} \Big\{F^n(z) - F^n(y)F^n(z)$$

$$- \frac{1 - F(y)}{F(z) - F(y)}[F^n(z) - F^n(y)]\Big\} \lambda_T(y)\lambda_T(z)K_h(x - y)K_h(x - z)\, dy\, dz$$

where

$$I_n(F) = \textstyle\sum_0^{n-1} (n - k)^{-1} \binom{n}{k} F^k(1 - F)^{n-k}.$$

PROOF. The lemma can be verified by direct calculation. Using the lemma, the calculation of $E\hat{\lambda}_T$ and Var $\hat{\lambda}_T$ proceeds in essentially the same line as in WL for the uncensored case. To illustrate the idea,

$$E\hat{\lambda}_T(x) = \textstyle\sum_{j=1}^n \displaystyle\int E(\delta_{(j)} | X_{(j)} = y)(n - j + 1)^{-1} f_{X_{(j)}}(y)K_h(x - y)\, dy$$

$$= \int \left[\sum \frac{1}{(n - j + 1)} \frac{n!}{(j - 1)!(n - j)!} F^{j-1}(y)(1 - F(y))^{n-j}\right] f(y)m(y)K_h(x - y)\, dy$$

$$= \int (1 - F^n(y))\lambda_T(y)K_h(x - y)\, dy.$$

The verification of the variance formula is similar, though involving longer calculation. □

As $n \to \infty$ the dominant part of $E\hat{\lambda}_T(x)$ is the convolution $\lambda_T * K_h(x)$ which can be regarded as an approximation of $\lambda_T(x)$ by the weighted average $\int \lambda_T(y)K_h(x - y)\, dy$. For this to be a good approximation, it is necessary that the values $\lambda_T(y)$ for y far away from x must not be so large that the down-weighting of the kernel is insufficient. Since $\lambda_T(y) = f_T(y)(1 - F_T(y))^{-1}$ this amounts to a compatibility condition on the tail behavior of K_h and $1 - F_T$.

DEFINITION. K is said to be compatible with a cdf F if for any $M > 0$, there exists h small enough such that $h^{-1}K(h^{-1}(y - x))/(1 - F(y))$ is uniformly bounded for $|y - x| > M$. This uniform bound will henceforth be denoted by G_M.

THEOREM 2. *Let $n \to \infty$, $h \to 0$ and $nh \to \infty$;*
a) *if K is compatible with F_T then $E\hat{\lambda}_T(x) \to \lambda_T(x)$;*
b) *if K is compatible with both F_T and F_C then*

$$\mathrm{Var}(\hat{\lambda}_T(x)) = (nh)^{-1}\left(\int K^2(t)\, dt\right)\lambda_T(x)(1 - F(x))^{-1} + o((nh)^{-1}).$$

PROOF. Since the arguments are quite similar to those in WL, we will only outline the main ideas. For part (a),

$$E\hat{\lambda}_T(x) - \lambda_T(x) = \int [(1 - F^n(y))\lambda_T(y) - \lambda_T(x)]K_h(x - y)\, dy.$$

Using the compatibility of K to F_T, it is easy to show that the integrand is dominated by an integrable function. Furthermore, the integrand clearly converges to zero for all y. Hence (a) follows from the dominated convergence theorem.

For part (b), it is readily verified (Lemma 6 of WL) that $n\, I_n(F(y)) \to (1 - F(y))^{-1}$ provided $F(y) < 1$. It then follows from a dominated convergence argument that

$$nh\left(\int K^2(t)\, dt\right)^{-1} \int I_n(F(y))\lambda_T(y)K_h^2(x-y)\, dy \to \lambda_T(x)(1-F(x))^{-1}.$$

This is the dominant term in $nh(\int K^2(t)\, dt)^{-1} \operatorname{Var} \hat{\lambda}_T(x)$, the other term can be shown to be convergent to zero by the same argument as in WL. □

REMARKS.

(i) As an immediate consequence of the theorem, $\hat{\lambda}_T(x)$ is mean square consistent for $\lambda_T(x)$.

(ii) The compatibility of K to F_C can be relaxed, if some further conditions are imposed on K and h_n.

3. Asymptotic normality. In this section the projection method (Hajek, 1968) is applied to investigate the asymptotic normality of $\hat{\lambda}_T(x)$. Suppose $Y_1, \cdots, Y_n$ are i.i.d. and W is a statistics based on $\mathbf{Y}$. The key idea of Hajek's method is that even though the central limit theorem is concerned with sums of independent r.v.'s, its scope may be extended to statistics asymptotically equivalent to such sums. Thus we can try to approximate W by its projection $\hat{W}$ on the subspace of all such sums of independent terms. Hajek gave the following formulae which are easy to verify:

$$\hat{W} = \textstyle\sum_{i=1}^n E(W \mid Y_i) - (n-1)EW, \quad E\hat{W} = EW,$$

$$E(W - \hat{W})^2 = \operatorname{Var}(W) - \operatorname{Var}(\hat{W}).$$

In our present problem, let $Y_i = (X_i, \delta_i)$, and

$$W = \hat{\lambda}_T(x) = \textstyle\sum_1^n W_j$$

where

$$W_j = (n - R_j + 1)^{-1}\delta_j K_h(x - X_j).$$

LEMMA 2. *$E(W_i \mid Y_i) = n^{-1}V_n(Y_i)$; and for $j \neq i$,*

$$E(W_j \mid Y_i) = (n-1)^{-1}\int (1-F(y))^{-1}(1-F^{n-1}(y))m(y)K_h(x-y)f(y)\, dy + (n(n-1))^{-1}U_n(Y_i)$$

where

$$V_n(Y_i) = (1-F(X_i))^{-1}(1-F^n(X_i))\delta_i K_h(x-X_i),$$

$$U_n(Y_i) = -\int (1-F(y))^{-2}[1-F^n(y) - nF^{n-1}(y)(1-F(y))]I_{(y\le X_i)}m(y)f(y)K_h(x-y)\, dy.$$

PROOF. Using the facts that:

$$\text{given } (X_i, \delta_i), \quad R_i \sim 1 + \text{Binomial}(n-1; F(X_i)), \text{ and}$$

$$\text{given } (X_i, \delta_i), (X_j, \delta_j), R_j \sim \begin{cases} 1 + \text{Bin}(n-2, F(X_j)) & \text{if } X_j \le X_i \\ 2 + \text{Bin}(n-2, F(X_j)) & \text{if } X_j > X_i, \end{cases}$$

the formulae follow from direct calculation. □

Using Hajek's formula for $\hat{W}$, Lemma 2 and Theorem 1, we can write

$$\hat{W} - E\hat{W} = \sum_{i=1}^{n} [E[W_i | Y_i) + (n-1)E(W_j | Y_i) - EW]$$
$$= \sum_{i=1}^{n} [n^{-1}V_n(Y_i) + n^{-1}U_n(Y_i) + \Delta_n].$$

where $\Delta_n = -\int F^{n-1}(y)m(y)K_h(x-y)f(y)\,dy$.

THEOREM 3. *If K is compatible with both F_T and F_C, then the standardized form of $W = \hat{\lambda}_T(x)$ has an asymptotic normal distribution, as $n \to \infty$, $h \to 0$, $nh \to \infty$.*

PROOF. The main steps are
(i) $|U_n| = O(\log n)$, $|\Delta_n| = O((n(n+1))^{-1})$.

PROOF. Choose M such that $F_X(x+M) < 1$, then

$$0 \le -U_n(Y_i) \le d_n + G_M \int_{|y-x|>M} (1 + F + \cdots + F^{n-1} - nF^{n-1})\,dF(y)$$

where $d_n = O(1)$. Similarly,

$$0 \le -\Delta_n \le O(F^{n-1}(x+M)) + G_M \int F^{n-1}(1-F)\,dF. \qquad \Box$$

(ii) $E|V_n(Y_i)|^r = \alpha_{r,h}(1-F(x))^{-r}m(x)f(x) + o(\alpha_{r,h})$ where

$$\alpha_{r,h} = \int K_h^r(y)\,dy = h^{-(r-1)} \int K^r\,dy, \quad r = 1, 2, \cdots.$$

PROOF.

$$E|V_n|^r = \int [(1-F(y))^{-r}(1-F^n(y))^r m(y)K_h^r(x-y)f(y)]\,dy$$
$$\le \int_{|y-x|\le M} [\cdots]\,dy + G_M^r \int_{|y-x|>M} (1-F^n)^r\, mf\,dy.$$

Since $K_h^r/\alpha_{r,h}$ is also a peaking kernel, and M is chosen such that $F(x+M) < 1$, the result follows. $\Box$

(iii) The standardized version of $\hat{W}$ and W have the same asymptotic distribution.

PROOF. Using (i) and (ii),

$$\mathrm{Var}(\hat{W}) = n\,\mathrm{Var}(V_n/n + U_n/n + \Delta_n) = n^{-1}\alpha_{2,h}\lambda_T(x)(1-F(x))^{-1} + o(nh)^{-1}).$$

Comparing with the variance expression in Theorem 2, it follows that $\mathrm{Var}(\hat{W})/\mathrm{Var}(W) \to 1$. Hence

$$E[(\hat{W} - E\hat{W})/\sqrt{\mathrm{Var}(\hat{W})} - (W - EW)/\sqrt{\mathrm{Var}(\hat{W})}]^2$$
$$= \mathrm{Var}(\hat{W})^{-1}E(\hat{W} - W)^2$$
$$= \mathrm{Var}(\hat{W})^{-1}[\mathrm{Var}(W) - \mathrm{Var}(\hat{W})] \to 0. \qquad \Box$$

(iv) $(\hat{W} - E\hat{W})/(\mathrm{Var}(\hat{W}))^{1/2} \to N(0,1)$.

PROOF. Since Δ_n is negligible, by Lyapounov's theorem a sufficient condition for (iv) is that $(\mathrm{Var}(\hat{W}))^{-3/2} nE|V_n/n + U_n/n|^3$ converges to zero. It is readily established using the bounds in (i) and (ii) that the above quantity is $O((nh)^{-1/2})$, hence completing the proof. $\Box$

Acknowledgment. We are grateful to Professors R. R. Bahadur and S. M. Stigler for bringing Hajek's projection method to our attention.

REFERENCES

HAJEK, J. (1968). Asymptotic normality of simple linear statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.

WATSON, G. S. and LEADBETTER, M. R. (1964). Hazard analysis II. *Sankhya Ser. A* **26** 101–116.

YANDELL, B. S. (1983). Non-parametric inference for rates and densities with censored serial data. *Ann. Statist.* (to appear).

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637