The Annals of Statistics
1983, Vol. 11, No. 3, 994–998

A NOTE ON THE VARIABLE KERNEL ESTIMATOR OF THE HAZARD FUNCTION FROM RANDOMLY CENSORED DATA[1]

BY MARTIN A. TANNER

University of Wisconsin

In a recent paper (Tanner and Wong, 1983b), a family of data-based nonparametric hazard estimators was introduced. Several of these estimators were studied in an extensive simulation experiment. The estimator which allows for variable bandwidth was found to have a superior performance. In this note, sufficient conditions for the variable kernel estimator to be strongly consistent are presented.

1. Introduction. Let (T_i, C_i), $i = 1, \cdots, n$ be independent and identically distributed pairs of non-negative random variables. Assume that T_i and C_i are independent for all i. Denote by $S_T(f_T)$ and $S_C(f_C)$ the survivor (density) functions of T_i and C_i, respectively. (See Miller, 1981.) In the random censorship model we observe the pairs $(Y_i, \tilde{\delta}_i)$, $i = 1, \cdots, n$ where

$$Y_i = \min(T_i, C_i)$$

$$\tilde{\delta}_i = I(T_i < C_i)$$

$$\ell = \sum_{i=1}^{n} \tilde{\delta}_i.$$

The problem is to estimate the hazard function $h(z) = f_T(z)/S_T(z)$.

Define R_k as the distance from the point z to the kth nearest of $Y_{i_1}, \cdots, Y_{i_\ell}$, where $\tilde{\delta}_{i_1} = \tilde{\delta}_{i_2} = \cdots = \tilde{\delta}_{i_\ell} = 1$ (assume $k \leq \ell$). R_k, then is the distance to the kth closest failure neighbor from z. Let δ_i be the indicator random variable associated with $Y_{(i)}$.

The *variable kernel estimator* of $h(z)$ is defined as

$$\hat{h}_n(z) = \frac{1}{2R_k} \sum_{i=1}^{n} \frac{\delta_i}{n - i + 1} K\left(\frac{z - Y_{(i)}}{2R_k}\right). \tag{1}$$

This estimator has the appealing feature that the configuration of the data plays a role in determining the degree of smoothing. In data sparse (dense) regions, R_k will be large (small) and the kernel will be flat (peaked).

In an extensive simulation study, Tanner and Wong (1983b) compare a data-based 3-parameter nonparametric estimator, which incorporates the kth nearest failure neighbor distance, to a data-based 1-parameter nonparametric estimator with constant bandwidth. (The theoretical properties of the 1-parameter estimator are discussed in detail in Tanner and Wong (1983a), while Yandell (1983) and Ramlau-Hansen (1983) examine a truncated 1-parameter kernel estimator.) The performance of the data-based 3-parameter estimator is shown to be superior to that of the 1-parameter estimator. Our ultimate goal is to establish the theoretical properties of this fully data-adaptive estimator. However, this is a difficult problem. We regard the present paper as solving a significant component problem. One must understand how these estimators behave when the parameters are chosen deterministically as a prerequisite to the analysis of the behavior of the data-adaptive procedure.

Received March 1982; revised December 1982.

[1] Support for this research has been provided in part by Research Grant No. MCS-8101836 of the National Science Foundation and by the United States Army under Contract No. DAAG29-80-C-0041. The author wishes to express his sincere appreciation to Professors R. R. Bahadur, S. M. Stigler, and W. H. Wong for their valuable suggestions.

AMS 1980 *subject classifications.* Primary 62P10; secondary, 62G05, 65D10.

Key words and phrases. Hazard function, variable kernel, nearest neighbor, random censoring, nonparametric estimation.

Several authors (Fix and Hodges, 1951; Loftsgaarden and Quesenberry, 1965; Wagner, 1975; Moore and Yackel, 1977; and Mack and Rosenblatt, 1979) have discussed the theoretical properties of the variable kernel estimator of the density function and the special case nearest neighbor estimator. We point out that the estimation of the hazard is a somewhat more difficult problem, since formula (1) depends on both the order statistics of the sample and the ordering induced by estimating the hazard at a point and sorting the data to obtain the kth nearest failure neighbor of this point. For this reason, direct application of previous techniques yields intractable formulas.

2. Consistency of $4\tilde{h}_n$. We assume that the survivor and density functions are continuous in a neighborhood around the point of interest. We begin with some lemmas. In Lemma 2.1, we present the density of R_k. We use this result in Lemma 2.2 to show that R_k converges almost surely to zero. Lemma 2.3 enables us to use Proposition 3i of Aalen (1978) to prove almost sure convergence of $\tilde{h}_n(z)$.

LEMMA 2.1. *Let R_k represent the distance between the point x and its kth nearest failure point. Let $p = P(T_i > C_i)$,*

$$G(r) = \int_{|x-y|<r} f_T(y)S_C(y)\,dy, \quad F(r) = (1-p)G(r),$$

$$G'(r) = f_T(x-r)S_C(x-r) + f_T(x+r)S_C(x+r) \quad \text{and} \quad F'(r) = (1-p)G'(r).$$

Then the density of R_k is

$$f_{R_k} = n\binom{n-1}{k-1}F(r)^{k-1}(1-F(r))^{n-k}F'(r).$$

PROOF. The probability of m censored observations in a sample of size n is given as

$$P(m) = \binom{n}{m}p^m(1-p)^{n-m}.$$

In addition, given that m observations in a sample of size n have been censored, the density of R_k is given as

$$P(r \mid m) = (n-m)\binom{n-m-1}{k-1}G(r)^{k-1}(1-G(r))^{n-m-k}G'(r).$$

The result now follows by direct calculation.

LEMMA 2.2. *Let $k = k(n) = [n^\alpha]$, $0 < \alpha < 1$, and let R_k be defined as above. Then $R_k \to_{\text{a.s.}} 0$.*

PROOF. Given $\delta' > 0$, by Lemma 2.1 and repeated application of integration by parts it is easy to show that

$$P(R_k > \delta') \le \sum_{i=0}^{k-1}\binom{n}{i}\delta^i(1-\delta)^{n-i}.$$

From Chernoff (1952), it can be shown that this quantity is bounded by $2^{-nA(n)}$, where $A(n)$ equals

$$-[\log_2(\delta^\delta(1-\delta)^{1-\delta})] + \log_2[p^p(1-p)^{1-p}] + \log_2\left(\frac{1-\delta}{\delta}\right)^p - \log_2\left(\frac{1-\delta}{\delta}\right)^\delta,$$

with $p = k/n$. It is now straightforward to show that

$$-nA(n) = -n\left\{\left[\varepsilon - \left|\log_2\left(1-\frac{1}{n^{1-\alpha}}\right)\right|\right] - \frac{1}{n^{1-\alpha}}\left[(1-\alpha)\log_2(nc) - \left|\log_2\left(1-\frac{1}{n^{1-\alpha}}\right)\right|\right]\right\}.$$

For n sufficiently large we have $-nA(n) < -n\varepsilon'$, for some positive $\varepsilon' < \varepsilon$, and the result follows.

LEMMA 2.3. *Let R_k and $k = k(n)$ be defined as above, with $\frac{1}{2} < \alpha < 1$. Then*

$$\frac{n^{1/2}}{\log(n)} R_{k(n)} \to_{\text{a.s.}} \infty.$$

PROOF. The result will follow if we can show that for all $\varepsilon > 0$,

$$\sum_{n=2}^{\infty} P\left(\frac{n^{1/2}}{\log(n)} R_{k(n)} \leq \varepsilon\right) < +\infty.$$

Now

$$P\left(\frac{n^{1/2}}{\log(n)} R_{k(n)} \leq \varepsilon\right) = \int_0^{F(\varepsilon_n)} n\binom{n-1}{k-1} t^{k-1}(1-t)^{n-k}\, dt,$$

where $\varepsilon_n = \varepsilon \dfrac{\log(n)}{n^{1/2}}$. One can show that the result will follow if

$$\sum_{n=2}^{\infty} \int_0^{\varepsilon_n} n\binom{n-1}{k-1} t^{k-1}(1-t)^{n-k}\, dt < +\infty.$$

Proceeding analogously to Lemma 2.2

$$\int_0^{\varepsilon_n} n\binom{n-1}{k-1} t^{k-1}(1-t)^{n-k}\, dt = \sum_{i=k}^{n}\binom{n}{i} \varepsilon_n^i (1-\varepsilon_n)^{n-i} \leq 2^{-nA(n)},$$

where, for $p = k/n$,

$$A(n) = -\log_2 \varepsilon_n^p - \log_2(1-\varepsilon_n)^{1-p} + \log_2(p)^p + \log_2(1-p)^{1-p}$$

$$= \frac{1}{n^{1-\alpha}} \log_2\left(\frac{n^{1/2+\alpha-1}}{\varepsilon \log(n)}\right) + \left(1 - \frac{1}{n^{1-\alpha}}\right)\log_2\left(\frac{1 - \dfrac{1}{n^{1-\alpha}}}{1 - \dfrac{\varepsilon \log n}{n^{1/2}}}\right).$$

Hence for $1/2 < \alpha < 1$ and sufficiently large n, $-nA(n) < -n^{\alpha'}$, where $0 < \alpha' < \alpha$, and the result follows.

THEOREM 2.1. *Let $k = k(n) = [n^{\alpha}]$, $1/2 < \alpha < 1$, and R_k be defined as above, let $K(\cdot)$ be a function of bounded variation with compact support on the interval $[-1, +1]$, let h be continuous as z, then $\tilde{h}_n(z) \to_{\text{a.s.}} h(z)$.*

PROOF. Let $A_n = \{\sup_{m \geq n} |\tilde{h}_m(z) - h(z)| > \varepsilon\}$ for $\varepsilon > 0$. Now choose δ such that $(z - 2\delta) \geq 0$, then

$$A_n = \{A_n \cap \{\sup_{m \geq n} R_{k(m)} > \delta\}\} \cup \{A_n \cap \{\sup_{m \geq n} R_{k(m)} < \delta\}\}.$$

Now

$$P(A_n) \leq P(A_n \cap \{\sup_{m \geq n} R_{k(m)} > \delta\}) + P(A_n \cap \{\sup_{m \geq n} R_{k(m)} < \delta\})$$

and by Lemma 2.2, $R_k \to_{\text{a.s.}} 0$. Hence we need only consider the event $\{A_n \cap \{\sup_{m \geq n} R_{k(m)} < \delta\}\}$. Now by the triangle inequality one can show that

$$\{A_n \cap \{\sup_{m \geq n} R_{k(m)} < \delta\}\} \subseteq \{A_n' \cap \{\sup_{m \geq n} R_{k(m)} < \delta\}\} \cup \{A_n'' \cap \{\sup_{m \geq n} R_{k(m)} < \delta\}\},$$

where,

$$A'_n = \left\{ \sup_{m \geq n} \left(\left| \frac{1}{2R_{k(m)}} \int_{|u| \leq 1} K(u)\, d\hat{H}_m(z - 2R_{k(m)}u) - \frac{1}{2R_{k(m)}} \int_{|u| \leq 1} K(u)\, dH(z - 2R_{k(m)}u) \right| \right) \geq \frac{\varepsilon}{2} \right\},$$

$$A''_n = \left\{ \sup_{m \geq n} \left(\left| \frac{1}{2R_{k(m)}} \int_{|u| \leq 1} K(u)\, dH(z - 2R_{k(m)}u) - h(z) \right| \right) \geq \frac{\varepsilon}{2} \right\},$$

$H(t)$ is the cumulative hazard function and $\hat{H}_n(t)$ is the empirical cumulative hazard function discussed in Nelson (1972).

Regarding the first event, using an application of integration by parts, one can show

$$\{A'_n \cap \{\sup_{m \geq n} R_{k(m)} < \delta\}\} \subseteq \left\{ \sup_{m \geq n} \left(\frac{c}{2R_{k(m)}} \sup_{y \in [z-2\delta, z+2\delta]} | \hat{H}_m(y) - H(y) | \right) \geq \frac{\varepsilon}{4} \right\},$$

since $K(\cdot)$ is assumed to be a function of bounded variation with compact support. But by Proposition 3i of Aalen (1978) and Lemma 2.3, we have that

$$\lim_{n \to \infty} P(\sup_{m \geq n} (c \sup_{y \in [z-2\delta, z+2\delta]} | \hat{H}_m(y) - H(y) | / 2R_{k(m)}) \geq \varepsilon/4) = 0.$$

Regarding the second event, it is immediate that

$$\{A''_n \cap \{\sup_{m \geq n} R_{k(m)} < \delta\}\} \subseteq \{A''_n\}.$$

Therefore, if the function

$$f(\alpha) = \begin{cases} 0 & \alpha = 0 \\ \left| \int_{|u| \leq 1} K(u) h(z - 2\alpha u)\, du - h(z) \right| & \alpha > 0 \end{cases}$$

is continuous at z, then $\lim_{n \to \infty} P(A''_n) = 0$, since $R_k \to_{\text{a.s.}} 0$. Now $f(\alpha)$ can be shown to be dominated by

$$\max_{|u| \leq 1} | h(z - 2\alpha u) - h(z) | \int_{|u| \leq 1} | K(u) |\, du.$$

If we let $\alpha \to 0$, the result follows.

REFERENCES

AALEN, O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. *Ann. Statist.* **6** 534–545.

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. *Ann. Math. Statist.* **23** 493–507.

FIX, E., and HODGES, J. L. (1951). Discriminatory analysis. Nonparametric discrimination: consistency properties, Report 4, Project Number 21-49-004, USAF School of Aviation Medicine, Randolf Field, Texas.

LOFTSGAARDEN, D. O., and QUESENBERRY, C. P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.* **36** 1049–1051.

MACK, Y. P., and ROSENBLATT, M. (1979). Multivariate k-nearest neighbor density estimates. *J. Multivariate Anal.* **9** 1–15.

MILLER, R. G. (1981). *Survival Analysis.* Wiley, New York.

MOORE, D. S., and YACKEL, J. W. (1977). Large sample properties of nearest neighbor density function estimators. In *Statistical Decision Theory and Related Topics* (S. S. Gupta and D. S. Moore, Eds.) Academic, New York.

NELSON, W. B. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14** 945–966.

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

TANNER, M. A. and WONG, W. H. (1983a). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist.* **11** 989–993.

TANNER, M. A. and WONG, W. H. (1983b). Data-based nonparametric estimation of the hazard function with applications to exploratory analysis and model diagnostics. *J. Amer. Statist. Assoc.* (to appear).

WAGNER, T. J. (1975). Nonparametric estimates of probability densities. *IEEE Trans. Inform. Theory* IT-21, 438–440.

YANDELL, B. S. (1983). Non-parametric inference for rates and densities with censored serial data. *Ann. Statist.* **11** (to appear).

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706