The Annals of Statistics
1985, Vol. 13, No. 2, 592–605

A HISTOGRAM ESTIMATOR OF THE HAZARD RATE WITH CENSORED DATA[1]

BY REGINA Y. C. LIU AND JOHN VAN RYZIN

Rutgers University and Columbia University

A histogram type estimator of the hazard rate of lifetimes which are randomly right-censored is studied. This estimator is based on random spacings of the order statistics of uncensored observations and extends the idea of the histogram density estimation suggested by Van Ryzin (1973) to censored data.

In this paper we establish pointwise large sample properties of the estimator including strong consistency, strong uniform consistency on a bounded interval, and two asymptotic distribution results. Of particular interest is the distribution result obtained by imposing extra "symmetry" conditions on the interval covering the given point. In fact, it yields the best attainable rate of convergence among all nonnegative estimators.

Comparisons of our results with the kernel type estimators proposed in the literature are also given.

1. Introduction. In many lifetime studies, some of the subjects under study are censored on the right by a prior censoring time. Let $X_1, \cdots, X_n$ denote lifetimes (times to failure) for the n subjects under study, and $C_1, \cdots, C_n$ be the corresponding censoring times. The observed random variables are then Z_i and δ_i where

$$Z_i = \min(X_i, C_i) \quad \text{and} \quad \delta_i = I_{(X_i \le C_i)}, \tag{1.1}$$

where $I_{(A)}$ is the indicator function defined to be 1 if event A occurs and 0 otherwise.

We assume throughout the paper that:

(i) $X_1, \cdots, X_n$ are nonnegative and iid with common continuous d.f. F and continuous density f,
(ii) $C_1, \cdots, C_n$ are nonnegative and iid with common continuous d.f. G and continuous density g, and
(iii) lifetimes and censoring times are independent.

The problem considered here is estimation of the hazard rate function given by

$$\lambda(x) = f(x)/(1 - F(x)), \quad F(x) < 1. \tag{1.2}$$

Mathematically, we see that

$$\lambda(x) = (d/dx)[-\log(1 - F(x))]. \tag{1.3}$$

Received February 1984; revised November 1984.

[1] Research supported by NIH Grant 7 R01-ES-02557.

AMS 1980 *subject classifications.* Primary 62G05; secondary 62P10, 62E20.

Key words and phrases. Censored data, hazard function, histogram method, consistency, asymptotic normality.

In the censoring case if $G(x) < 1$, we have

$$\lambda(x) = \frac{f(x)(1 - G(x))}{(1 - F(x))(1 - G(x))}. \tag{1.4}$$

If we let

$$L(x) = P(Z_i \le x), \tag{1.5}$$

we see that

$$(1 - L(x)) = (1 - F(x))(1 - G(x)). \tag{1.6}$$

Let

$$f^*(x) = f(x)(1 - G(x)), \tag{1.7}$$

and substitute (1.6) and (1.7) into (1.4) to obtain

$$\lambda(x) = f^*(x)/(1 - L(x)). \tag{1.8}$$

Nonparametric estimation of the hazard rate has received much attention in recent statistical literature because of its practical importance in survival analysis. There have been three approaches to estimation of the hazard function $\lambda(x)$ due to the three different expressions of $\lambda(x)$ in (1.2), (1.3) and (1.8). Expression (1.8), which is only useful in the random censorship model, will be the motivation for our approach. Blum and Susarla (1980) were the first to use expression (1.8) in the random censorship model by an approach using the kernel method. Subsequent papers in the censored data case with the kernel method include Tanner and Wong (1983), Yandell (1983) and Burke (1983). The paper by Földes, Rejtö and Winter (1981) gives consistency results for kernel-type and fixed partition histogram-type hazard rate estimates. For approaches using (1.2) and (1.3) via the histogram method of this paper in the case of uncensored data, see Prakasa Rao and Van Ryzin (1983).

2. Estimation of the hazard rate function. Before introducing our estimator we need some preliminaries:

Let D_n denote the number of observed deaths (uncensored observations) when the sample size is n, i.e., $D_n = \sum_{i=1}^n \delta_i$. We will write D instead of D_n to simplify notation. Let $\phi(0) = 0$ and let $\phi(m) = \inf\{\ell: \sum_{i=1}^{\ell} \delta_i = m\}$ if $m \ge 1$. Let $T_m = Z_{\phi(m)}$ for $m = 0, 1, \cdots, D$, where we define $T_0 = Z_0 = -\infty$, (i.e., T_m is the mth uncensored observation in the sequence of $Z_0, Z_1, \cdots, Z_n$). Let U_j be the jth order statistic of $T_0, T_1, \cdots, T_D$ and let $U_{D+1} = +\infty$, then $-\infty = U_0 < U_1 < \cdots < U_D < U_{D+1} = +\infty$.

Defining $F^*(x) = P(Z_i \le x, \delta_i = 1)$, we have $F^*(x) = \int_0^x (1 - G(s))\, dF(s)$. Note that F^* is the subdistribution of the uncensored lifetimes with density $f^*(x) = (d/dx)F^*(x) = (1 - G(x))f(x)$. Let the subempirical distribution function be denoted as F_n^*, i.e.,

$$F_n^*(x) = (1/n) \sum_{i=1}^n I_{(Z_i \le x, \delta_i = 1)}. \tag{2.1}$$

Let

$$H(x) = P(X_i \le x \mid \delta_i = 1) \tag{2.2}$$

and

$$p = P(\delta_i = 1), \quad 0 < p \le 1. \tag{2.3}$$

Note that from (2.2) and (2.3),

$$F^*(x) = P(Z_i \le x,\ \delta_i = 1) = H(x)p. \tag{2.4}$$

Our estimator of $\lambda(x)$ is defined to be

$$\lambda_n(x) = f_n^*(x)/(1 - L_n(x)), \tag{2.5}$$

where

$$L_n(x) = (1/n) \textstyle\sum_{i=1}^n I_{(Z_i \le x)}, \tag{2.6}$$

and $f_n^*(x)$ is an estimator of $f^*(x)$ obtained by using the density estimation scheme proposed by Van Ryzin (1973) applied to the uncensored observations. That is, for a fixed point $x \in \mathbf{R}$, we first choose positive-integer valued random variables $A_n(x)$ which are measurable w.r.t the σ-field generated by $X_1, \cdots, X_n$ and $C_1, \cdots, C_n$ such that

$$P(0 \le A_n(x) \le D + 1 - k, \quad U_{A_n(x)} \le x < U_{A_n(x)+k}) = 1 \tag{2.7}$$

and

$$A_n(x) = \begin{cases} 0 & \text{if} \quad x < U_1 \quad \text{or} \quad D + 1 - k < 0 \\ D + 1 - k & \text{if} \quad x > U_D, \end{cases}$$

where $k = k_n$ is a sequence of positive integers satisfying:

$$\text{(i) } k/n \to 0, \quad \text{and} \quad \text{(ii) } (\log n)/k \to 0 \quad \text{as} \quad n \to \infty. \tag{2.8}$$

We simply write $A_n(x) = A$ when convenient, and estimate $f^*(x)$ by

$$f_n^*(x) = \frac{F_n^*(U_{A+k}) - F_n^*(U_A)}{U_{A+k} - U_A}. \tag{2.9}$$

Note that in the case where $A_n(x)$ is a constant in x between any two consecutive order statistics, the estimator (2.9) is a histogram (although not of the classical type). Computationally such a choice of $A_n(x)$ has the advantage that one need only compute the estimator once between pairs of order statistics.

To prove large sample properties for $\lambda_n(x)$, we first investigate the distribution of the uncensored observations $T_1, \cdots, T_D$, conditional on D.

LEMMA 2.1. *Conditional on $D = d \ge 1$, $T_1, \cdots, T_d$ are iid with the common d.f. H, where H is given in* (2.2).

PROOF. We prove that

$$P(T_i \le x_i,\ i = 1, \cdots, D \mid D = d) = \prod_{i=1}^{d} P(T_i \le x_i \mid D = d)$$

in order to show the independence. That $P(T_i \le x_i \mid D = d) = H(x_i)$, for $i = 1, \cdots, d$ follows by similar arguments which are omitted. Let $x_i \in \mathbf{R}$, $i = 1, \cdots, d$. Since

$$\begin{aligned} P(T_1 \le x_1, \cdots, T_d \le x_d \mid D = d) & \\ &= P(X_{\phi(1)} \le x_1, \cdots, X_{\phi(d)} \le x_d \mid D = d) \\ &= \textstyle\sum^* P(X_{i_1} \le x_1, \cdots, X_{i_d} \le x_d \mid \phi(1) = i_1, \cdots, \phi(d) = i_d) \\ &\quad \cdot P(\phi(1) = i_1, \cdots, \phi(d) = i_d \mid D = d). \end{aligned} \tag{2.10}$$

where $\sum^*$ is the sum over all $\binom{n}{d}$ subsets of distinct integers $(i_1, \cdots, i_d)$ from the integers $1, \cdots, n$, the lemma follows from the independence of δ_i's in the definition of $\phi(\cdot)$'s.

3. Consistency of $\lambda_n(x)$. Since $\lambda_n(x) = f_n^*(x)/(1 - L_n(x))$, to prove consistency of $\lambda_n(x)$, we first prove consistency of $f_n^*(x)$. The following lemmas will be useful for the proofs of the consistency of $f_n^*(x)$.

LEMMA 3.1. *$A_n/n \to F^*(x)$ w.p.1, $(A_n + k)/n \to F^*(x)$ w.p.1 as $n \to \infty$. Furthermore, if $F^*(x - \varepsilon) < F^*(x) < F^*(x + \varepsilon)$ for any $\varepsilon > 0$, then as $n \to \infty$, $U_{A_n} \to x$ w.p.1, $U_{A_n+k} \to x$ w.p.1 and $U_{A_n+k} - U_{A_n} \to 0$ w.p.1.*

PROOF. The proof is similar to Lemma 1 in Van Ryzin (1973) with the replacement of F by F^* and by use of the Glivenko-Cantelli lemma applied to subdistributions.

From Lemma 2.1 we see that conditional on $D = d$, $T_1, \cdots, T_d$ and hence $H(T_1), \cdots, H(T_d)$ are iid. Furthermore, we have

LEMMA 3.2. Given $D = d \ge 1$,

(a) *$H(T_1), \cdots, H(T_d)$ are iid with common uniform d.f. on $(0, 1)$*
(b) *$H(U_1), H(U_2) - H(U_1), \cdots, 1 - H(U_d)$ have the same joint distribution as $Q_1/S_{d+1}, Q_2/S_{d+1}, \cdots, D_{d+1}/S_{d+1}$, where $S_j = \sum_{i=1}^{j} Q_i$, and the Q_i's are iid with a common exponential d.f. with mean 1. Furthermore, $H(U_{j+k}) - H(U_j)$ and $(S_{j+k} - S_j)/S_{d+1}$ are identically distributed and hence $H(U_{j+k}) - H(U_j)$ has a beta distribution with parameters k and $d - k + 1$ for $j = 0, 1, \cdots, d - k + 1$.*

PROOF. (a) follows from the definition of H; while (b) results from (a), the monotonicity of H, and a well-known theorem on the joint distribution of coverages (see, e.g., Breiman, 1968, Proposition 13.15).

The following representation of $f_n^*(x)$ will be useful. If $F^*(U_{A+k}) > F^*(U_A)$, we write

$$f_n^*(x) = V_n(x)W_n(x), \tag{3.1}$$

where

$$V_n(x) = \frac{F_n^*(U_{A+k}) - F_n^*(U_A)}{F^*(U_{A+k}) - F^*(U_A)} = \frac{k/n}{F^*(U_{A+k}) - F^*(U_A)},$$

and

$$W_n(x) = \frac{F^*(U_{A+k}) - F^*(U_A)}{U_{A+k} - U_A}.$$

The following two lemmas follow from Lemma 3.2 by conditional arguments on $D = d$ which are similar to those of Lemmas 2, 3 and 4 of Van Ryzin (1973). Readers are referred to the technical report by Liu and Van Ryzin (1983) for details.

LEMMA 3.3. (a) *If* $F^*(x - \varepsilon) = F^*(x)$ *or* $F^*(x) = F^*(x + \varepsilon)$ *for some* $\varepsilon > 0$, *then* $f_n^*(x) \to 0$ *w.p.1 as* $n \to \infty$.

(b) *If* $F^*(x - \varepsilon) < F^*(x) < F^*(x + \varepsilon)$ *for any* $\varepsilon > 0$ *and* $x \in C(f)$, *the continuity set of* f, *then* $W_n(x) \to f^*(x)$ *w.p.1 as* $n \to \infty$.

LEMMA 3.4. *If* $k = k_n \to \infty$ *as* $n \to \infty$, *then for any* $\varepsilon > 0$, *and* $m \geq 1$, *we have* (i):

$$P((H(U_{j+k}) - H(U_j)) > k/(np(1 - \varepsilon)) \mid D = d)$$
$$\leq \exp\left(m\left\{\log\left[p(1 - \varepsilon)\frac{n}{d + 1}\right] + \frac{m - 1}{2k}\right\}\right)$$

and (ii):

$$P((H(U_{j+k}) - H(U_j)) < k/(np(1 + \varepsilon)) \mid D = d)$$
$$\leq \exp\left(m\left\{\log\frac{d}{n} - \log\left[p(1 + \varepsilon)\left(1 - \frac{m}{k}\right)\right]\right\}\right),$$

$j = 0, 1, \cdots, d - k + 1$, *if* $d - k + 1 \geq 0$.

THEOREM 3.1 (Strong consistency of $f_n^*(x)$). *Let* $A_n(x)$ *satisfy* (2.7) *and let* $k = k_n$ *satisfy* (2.8) *and* $(k \log n) = o(n)$. *If* $x \in C(f^*)$, *then* $f_n^*(x) \to f^*(x)$ *w.p.1 as* $n \to \infty$.

PROOF. *If* $F^*(x + \varepsilon) = F^*(x)$ or $F^*(x - \varepsilon) = F^*(x)$ for some $\varepsilon > 0$, then $f_n^*(x) = 0$ and the result follows from Lemma 3.3(a).

Suppose $F^*(x + \varepsilon) > F^*(x) > F^*(x - \varepsilon)$ for every $\varepsilon > 0$. Let $f_n^*(x) = V_n(x)W_n(x)$ as in (3.1). Then the result follows from Lemma 3.3(b) provided we show that

$V_n(x) \to 1$ w.p.1 as $n \to \infty$. But

$$
\begin{aligned}
&P(|V_n(x) - 1| > \varepsilon) \\
&= P\left(\left|\frac{k/n}{F^*(U_{A+k}) - F^*(U_A)} - 1\right| > \varepsilon\right) \\
&= \sum_{d=0}^{n} P\left(U_{j=0}^{d-k+1}\left\{\left|\frac{k/n}{p(H(U_{A+k}) - H(U_A))} - 1\right| > \varepsilon, A = j\right\}\middle| D = d\right)P(D = d) \qquad (3.2)\\
&\le \sum_{d=0}^{n} \sum_{j=1}^{d-k+1} P\left(\left|\frac{k/n}{p(H(U_{j+k}) - H(U_j))} - 1\right| > \varepsilon \,\middle|\, D = d\right)P(D = d) \\
&\le n \sum_{d=0}^{n} P((H(U_{1+k}) - H(U_1)) > k/(np(1 - \varepsilon)) \mid D = d)P(D = d) \\
&\quad + n \sum_{d=0}^{n} P((H(U_{1+k}) - H(U_1)) < k/(np(1 + \varepsilon)) \mid D = d)P(D = d).
\end{aligned}
$$

From Lemma 3.4, the first term of (3.2) is

$$
\begin{aligned}
&\le n \sum_{d=0}^{n} \exp\left(m\left\{\log\left[p(1 - \varepsilon)\frac{n}{d + 1}\right] + \frac{m - 1}{2k}\right\}\right)P(D = d) \\
&= n \sum_{\{d: d > np(1-\varepsilon')-1\}} \exp\left(m\left\{\log\left[p(1 - \varepsilon)\frac{n}{d + 1}\right] + \frac{m - 1}{2k}\right\}\right)P(D = d) \\
&\quad + n \sum_{\{d: d \le np(1-\varepsilon')-1\}} \exp\left(m\left\{\log\left[p(1 - \varepsilon)\frac{n}{d + 1}\right] + \frac{m - 1}{2k}\right\}\right)P(D = d), \\
&= \text{(a)} + \text{(b)}, \quad \text{say, where } \varepsilon' \text{ is such that } 0 < \varepsilon' < \varepsilon.
\end{aligned}
$$

By choosing $m = m_n$ such that $\lim_{n\to\infty} m_n/k = \delta$, where

$$\delta = \delta_{\varepsilon',p} = -\log(1 - \varepsilon)/(1 - \varepsilon') > 0,$$

it can be shown that

$$\text{(a)} = nO(e^{-(\delta^2/2)k}).$$

For (b), we have

$$\text{(b)} \le n[p(1 - \varepsilon)n]^m e^{m(m-1)/2k} P(D \le np(1 - \varepsilon) - 1).$$

Then by the central limit theorem for binomial random variables, we have

$$P(D \le np(1 - \varepsilon) - 1) \doteq 1 - \Phi\left(\frac{np\varepsilon + 1}{\sqrt{np(1 - p)}}\right),$$

where Φ is the d.f. of a standard normal distribution and $\doteq$ means approximately equivalent up to an error of $O(n^{-1/2})$. Apply the fact that for all $x > 0$,

$$\frac{x}{1 + x^2} e^{-x^2/2} < \int_x^\infty e^{-y^2/2}\, dy < \frac{1}{x} e^{-x^2/2} \quad \text{to} \quad \left[1 - \Phi\left(\frac{np\varepsilon + 1}{\sqrt{np(1 - p)}}\right)\right],$$

we have $P(D \le np(1 - \varepsilon) - 1) = O(e^{-cn})$ for some $c > 0$. Substituting this result into (b) we have

$$\text{(b)} = O(\exp[(m(1 + \delta/2)(\log n)/n) - c]n).$$

Since $k \log n = o(n)$ and since m is chosen to satisfy $m/k = \delta > 0$, we see (b) $= O(e^{-\alpha n})$ for some $\alpha > 0$. Hence (a) + (b) $= O(ne^{-\delta^2/2)k}) + O(e^{-an})$, which equals $O(e^{-\beta k})$ for some $\beta > 0$ since $\log n = o(k)$. Note that also

$$\sum_{n=1}^{\infty} O(e^{-\beta k}) = O(\sum_{n=1}^{\infty} (1/n^{(\beta k)/\log n})),$$

which converges since $k/(\log n) \to \infty$, as $n \to \infty$.

Similar arguments hold for the second term of (3.2) and thus we have shown that $\sum_{n=1}^{\infty} P(|V_n(x) - 1| > \varepsilon)$ converges. Therefore by the Borel-Cantelli Lemma,

$$V_n(x) \to 1 \quad \text{w.p.1 as} \quad n \to \infty. \tag{3.3}$$

THEOREM 3.2. (Strong uniform consistency of $f_n^*(x)$). *Suppose f is uniformly continuous on the support of f, $S(f) = (a, b)$ with $b < T$, where $T = \inf\{x: G(x) = 1\}$. Suppose $k = k_n$ satisfies, in addition to* (2.8), $\sum_{n=1}^{\infty} n\nu^k < +\infty$ *for all ν, $0 < \nu < 1$. Then as $n \to \infty$, $\sup_{x<b} |f_n^*(x) - f^*(x)| \to 0$ w.p.1.*

The set of lemmas below will be needed in the proof of Theorem 3.2 given below.

LEMMA 3.5. *If $k = k_n$ satisfies the conditions in Theorem 3.2, then as $n \to \infty$, $\sup_{x<b} |F^*(x) - A_n(x)/n| \to 0$ w.p.1.*

PROOF. Note that

$$\sup_{x<b} |F^*(x) - A_n(x)/n| \le \sup_{x<b} |F^*(x) - F_n^*(x)| + \sup_{x<b} |F_n^*(x) - A_n(x)/n|.$$

By the Glivenko-Cantelli Lemma for subdistributions, we have

$$\sup_{x<b} |F^*(x) - F_n^*(x)| \to 0 \quad \text{w.p.1 as} \quad n \to \infty.$$

The proof will be completed provided we show $\sup_{x<b} |F_n^*(x) - A_n(x)/n| \to 0$ w.p.1 as $n \to \infty$. Since from (2.7), $U_{A_n(x)} \le x < U_{A_n(x)+k}$ w.p.1 for all x, we have $A_n(x) \le nF_n^*(x) < A_n(x) + k$ w.p.1 for all x and thus

$$\sup_{x<b} |F_n^*(x) - A_n(x)/n| \le k/n \to 0 \quad \text{as} \quad n \to \infty.$$

From Lemma 3.2(b) we can choose a proper set of iid exponentially distributed r.v.'s $\{Q_i\}$ with mean 1 and perform the following transformations: $H(U_j) = S_n/S_{D+1}$, $j = 0, 1, \cdots, D + 1$, where $S_m = \sum_{i=1}^{m} Q_i$ and $S_0 \equiv 0$. Then from Lemma 3.2(b), we have

$$H(U_{A(x)+k}) - H(U_{A(x)}) = (S_{A(x)+k} - S_{A(x)})/S_{D+1}. \tag{3.4}$$

LEMMA 3.6. (Kim and Van Ryzin (1975). *If $k = k_n$ satisfies conditions in Theorem 3.2, then, as $n \to \infty$,*

$$\sup_{x<b} |(1/k)(S_{A_n(x)+k} - S_{A_n(x)})| \to 1 \quad \textit{w.p.1}.$$

LEMMA 3.7. *Let k satisfy conditions in Theorem 3.2. Then, as $n \to \infty$, $\sup_{x<b} | F^*(U_{A(x)+k}) - F^*(U_{A(x)}) | \to 0$ w.p.1.*

PROOF. By the fact that

$$F^*(U_{A(x)+k}) - F^*(U_{A(x)}) = p[H(U_{A(x)+k}) - H(U_{A(x)})], \tag{3.5}$$

the result follows by an argument similar to that in Kim and Van Ryzin (1975) with the representation (3.4).

LEMMA 3.8. *If f is uniformly continuous on (a, b), the support of f, with $b < T$, $T = \inf\{x: G(x) = 1\}$, then $f^* o (F^*)^{-1}$ is uniformly continuous on $[0, F^*(b)]$.*

PROOF. The proof can be carried out as in Kim and Van Ryzin (1975) with F replaced by F^*.

PROOF OF THEOREM 3.2. Recall that by (3.1),

$$W_n(x) = \frac{F^*(U_{A(x)+k}) - F^*(U_{A(x)})}{U_{A(x)+k} - U_{A(x)}}.$$

Since

$$\sup_{x<b} | f_n^*(x) - f^*(x) | \le \sup_{x<b} | f_n^*(x) - W_n(x) | + \sup_{x<b} | W_n(x) - f^*(x) |,$$

it suffices to prove (i) $\sup_{x<b} | f_n^*(x) - W_n(x) | \to 0$ w.p.1 and (ii)

$$\sup_{x<b} | W_n(x) - f^*(x) | \to 0 \quad \text{w.p.1 as} \quad n \to \infty.$$

For (i) we have, by (3.4)

$$\begin{aligned} f_n^*(x) &= W_n(x) \left[\frac{k}{np} \cdot \frac{1}{H(U_{A(x)+k}) - H(U_{A(x)})} \right] \\ &= W_n(x) \left[\frac{1}{k}(S_{A(x)+k} - S_{A(x)}) \right]^{-1} \left[\frac{S_{D+1}}{np} \right]. \end{aligned}$$

Then

$$\begin{aligned} &\sup_{x<b} | f_n^*(x) - W_n(x) | \\ &\qquad \le [\sup_{x<b} W_n(x)] \sup_{x<b} \left| \left[\frac{1}{k}(S_{A(x)+k} - S_{A(x)}) \right]^{-1} \left(\frac{S_{D+1}}{np} \right) - 1 \right|. \end{aligned}$$

Lemma 3.6 and $S_{D+1}/(np) \to 1$ w.p.1 by SLLN, imply the second part on the right tends to zero as $n \to \infty$. Applying the mean value theorem to $W_n(x)$, we have $W_n(x) = f^*(U_x^n)$ where $U_{A(x)} < U_x^n < U_{A(x)+k}$, and hence

$$\sup_{x<b} W_n(x) = \sup_{x<b} f^*(U_x^n) \le \sup_{x<b} f^*(x) < \infty.$$

Therefore (i) is proved. For (ii), we write

$$f^*(x) = \begin{cases} 0 & \text{if } x \le a \text{ and } F^*(x) = 0 \\ f^* o (F^*)^{-1}(F^*(x)) & \text{if } a < x < b \\ 0 & \text{if } x \ge b \text{ and } F^*(x) = F^*(b). \end{cases}$$

By the mean value theorem, there exists U_x^n such that $W_n(x) = f^*(U_x^n)$ w.p.1 where $U_{A(x)} < U_x^n < U_{A(x)+k}$. Also, by (2.7) we have $A(x) = 0$ if $x < U_1$ and then $U_x^n = a$, and $A(x) = D - k + 1$ if $x \geq U_D$ and then $U_x^n = b$. The above facts and the monotonicity of F^* imply that $W_n(x) = f^* o(F^*)^{-1}(F^*(U_x^n))$ w.p.1. Therefore, $\sup_{x<b} | W_n(x) - f^*(x) | \to 0$ w.p.1, if we verify that $\sup_{x<b} | F^*(U_x^n) - F^*(x) | \to 0$ w.p.1 as $n \to \infty$, since

$$\sup_{x<b} | W_n(x) - f^*(x) | = \sup_{x<b} | f^*(U_x^n) - f^*(x) |$$
$$= \sup_{x<b} | f^* o(F^*)^{-1}(F^*(U_x^n)) - f^* o(F^*)^{-1}(F^*(x)) |$$

and $f^* o(F^*)^{-1}$ is proved uniformly continuous on $[0, F^*(b)]$ in Lemma 3.8. To verify $\sup_{x<b} | F^*(U_x^n) - F^*(x) | \to 0$ w.p.1, we note that $F^*(U_{A(x)}) < F^*(U_x^n) < F^*(U_{A(x)+k})$ and apply the result by Lemma 3.1 to complete the proof.

THEOREM 3.3. (Strong consistency of $\lambda_n(x)$). *Let x be such that $L(x) < 1$. Let $A_n(x)$ satisfy* (2.7) *and k satisfy* (2.8) *and* $(k \log n) = o(n)$. *If $x \in C(f^*)$, then $\lambda_n(x) \to \lambda(x)$ w.p.1* as $n \to \infty$.

PROOF. Since $L(x) < 1$, by the SLLN it is easy to see that $(1 - L_n(x))^{-1} \to (1 - L(x))^{-1}$ w.p.1 and the result follows from Theorem 3.1.

THEOREM 3.4. (Strong uniform consistency of $\lambda_n(x)$). *Let u be such that $L(u) < 1$. Suppose f is continuous on $[0, u]$. Let $\{A_n(x)\}$ and $k = k_n$ satisfy conditions in Theorem* 3.2, *then* $\sup_{0 \leq x \leq u} | \lambda_n(x) - \lambda(x) | \to 0$ *w.p.1 as $n \to \infty$.*

PROOF. Note that

$$\sup_{0 \leq x \leq u} | \lambda_n(x) - \lambda(x) |$$
$$= \sup_{0 \leq x \leq u} \left| \frac{f_n^*(x)}{1 - L_n(x)} - \frac{f^*(x)}{1 - L(x)} \right|$$
$$\leq (\sup_{0 \leq x \leq u} f_n^*(x))(\sup_{0 \leq x \leq u} | (1 - L_n(x))^{-1} - (1 - L(x))^{-1} |)$$
$$+ (\sup_{0 \leq x \leq u} (1 - L(x))^{-1})(\sup_{0 \leq x \leq u} | f_n^*(x) - f^*(x) |).$$

The second term on the right of the inequality tends to zero directly by the result of Theorem 3.2 and by the fact that

$$\sup_{0 \leq x \leq u} (1 - L(x))^{-1} = (1 - L(u))^{-1} < +\infty.$$

Hence the result will follow if we prove the first term goes to zero. Consider

$$\sup_{0 \leq x \leq u} f_n^*(x) \leq \sup_{0 \leq x \leq u} | f_n^*(x) - f^*(x) | + \sup_{0 \leq x \leq u} f^*(x).$$

But $\sup_{0 \leq x \leq u} | f_n^*(x) - f^*(x) | \to 0$ w.p.1 as $n \to \infty$ by Theorem 3.2, and $\sup_{0 \leq x \leq u} f^*(x) < +\infty$ imply $\sup_{0 \leq x \leq u} f_n^*(x) < +\infty$ w.p.1. Thus we need only show that

$$\sup_{0 \leq x \leq u} | (1 - L_n(x))^{-1} - (1 - L(x))^{-1} | \to 0 \quad \text{w.p.1.}$$

However,

$$\sup_{0\le x\le u}|(1-L_n(x))^{-1}-(1-L(x))^{-1}|$$
$$\le ((1-L(u))(1-L_n(u)))^{-1}(\sup_{0\le x\le u}|L_n(x)-L(x)|)$$

and the desired statement follows by the Glivenko-Cantelli lemma.

4. Asymptotic normality of $\lambda_n(x)$. To derive the asymptotic distribution of $\lambda_n(x)$, we first give two theorems on the asymptotic distribution of $f_n^*(x)$. We indicate convergence in distribution in this section by "$\to_d$". Let $S(f^*) = \{x: f^*(x) > 0\}$ be the support of f^*.

THEOREM 4.1. *Let $x \in S(f^*)$ be a continuity point of $(f^*)'$, the first derivative of f^*. Let $\{A_n(x)\}$ satisfy* (2.7) *and $A_n(x) = A_n(x; X_1, \cdots, X_n, C_1, \cdots, C_n)$ be invariant under the permutation of* $\min(X_1, C_1), \cdots, \min(X_n, C_n)$. *Let $k = k_n$ satisfy* (2.8) *and additionally let k satisfy the condition $k^{3/2} = o(n)$. Then,*

$$k^{1/2}(f_n^*(x) - f^*(x)) \to_d N(0, (f^*(x))^2).$$

THEOREM 4.2. *Let $x \in S(f^*)$ be a continuity point of $(f^*)''$, the second derivative of f^*. Let k satisfy* (2.8). *Let $A_n(x)$ satisfy* (2.7),

$$R_n(x) = \max\{j;\ U_j \le x\}$$

and let $A_n(x) = A_n(x; X_1, \cdots, X_n, C_1, \cdots, C_n)$ be invariant under the permutation of $\min(X_1, C_1), \cdots, \min(X_n, C_n)$. *Suppose*

(i) $k^{1/2}[2A_n(x) + k - 2R_n(x)] = o_p(n)$, *where $X_n = o_p(a_n)$ if and only if $X_n/a_n \to 0$ in p. as $n \to \infty$,*
(ii) $A_n(x) + k - R_n(x) \to \infty$ *in p. as* $n \to \infty$,
(iii) $R_n(x) - A_n(x) \to \infty$ *in p. as* $n \to \infty$, *and*
(iv) $k^{1/2}n^{-2}[A_n(x) + k - R_n(x)][R_n(x) - A_n(x)] \to d$ *in p. as* $n \to \infty$,

where d is a nonnegative constant. Then

$$k^{1/2}(f_n^*(x) - f^*(x)) \to_d N(b(x), (f^*(x))^2),$$

where $b(x) = d(f^*)''(x)/6(f^*)^2(x)$.

To prove these theorems, we again use the representation $f_n^*(x) = W_n(x) \cdot V_n(x)$ defined in (3.1) to obtain

$$\begin{aligned} &k^{1/2}(f_n^*(x) - f^*(x)) \\ &\qquad = k^{1/2}V_n(x)(W_n(x) - f^*(x)) + k^{1/2}f^*(x)(V_n(x) - 1). \end{aligned} \tag{4.1}$$

Lemma 4.2 below states that the second term on the right-hand side of (4.1) converges in distribution to $N(0, (f^*(x))^2)$, and Lemma 4.3 below proves that the first term converges to zero in probability. Therefore, the proof of Theorem 4.1 will be completed by Lemmas 4.2 and 4.3. The following result quoted from the literature will be useful for proving Lemmas 4.2 and 4.3.

LEMMA 4.1 (Van Ryzin, 1977). *Let* $\{\xi_i\}$ *be iid with* $E\xi_1 = 0$, $E\xi_1^2 = 1$. *For each* n *and* i, $i = 1, \cdots, n$, *let* $I_n(i) = I_n(i; r_1, \cdots, r_n)$ *be a Borel measurable function on n-dimensional Euclidean space taking on values* 0 *and* 1 *which is symmetric under the permutation of* $r_1, \cdots, r_n$. *If* $T_n = \sum_{i=1}^n I_n(i)$ *and* $S_n' = \sum_{i=1}^n I_n(i)\xi_i$, *then* S_n' *and* $S_{T_n}^*$ *have the same limiting distribution, where* $S_j^* = \sum_{i=1}^j \xi_i$. *Furthermore, if* $T_n/ET_n \to 1$ *in p. and* $E(T_n) \to \infty$ *as* $n \to \infty$, *then* $T_n^{-1/2}S_n' \to_d N(0, 1)$.

LEMMA 4.2. *Under conditions of Theorem* 4.1, $k^{1/2}(V_n(x) - 1) \to_d N(0, 1)$.

PROOF. Recall that $V_n(x) = (k/n)/(p[H(U_{A+k}) - H(U_A)])$. Then by (3.4), we have

$$V_n(x) = (k/(S_{A+k} - S_A))(S_{D+1}/np)$$

and thus

$$\begin{aligned} k^{1/2}(V_n(x) - 1) &= \left[k^{1/2}\left(\frac{k}{S_{A+k}} - 1\right)\right]\frac{S_{D+1}}{np} + k^{1/2}\left(\frac{S_{D+1}}{np} - 1\right) \\ &= (I_1) \cdot (I_2) + (I_3), \quad \text{say.} \end{aligned}$$

Note that $(I_2) = (S_{D+1}/(D + 1))((D + 1)/np) \to 1$ w.p.1, by the SLLN with random index and by the regular SLLN. By noting that

$$(I_1) = (k/(S_{A+k} - S_A))((k - (S_{A+k} - S_A))/k^{1/2}),$$

we see (I_1) converges in distribution to $N(0, 1)$ since by Lemma 4.1 both $(k - (S_{A+k} - S_A))/k^{1/2} \to_d N(0, 1)$ and $((S_{A+k} - S_A)/k) \to 1$ in p.. Properly decomposing (I_3) and applying Lemma 4.1 and Slutsky's Theorem, it can be shown that (I_3) converges to zero in p., thus completing the proof.

LEMMA 4.3. *Under the conditions of Theorem* 4.1, $k^{1/2}(W_n(x) - f^*(x)) \to 0$ *in p. as* $n \to \infty$.

PROOF. Applying Taylor's expansion to $F^*(U_{A+k})$ and $F^*(U_A)$ in $W_n(x)$ about x to the second-order terms, we have

$$W_n(x) - f^*(x) = \frac{(U_{A+k} - x)^2(f^*)'(U_{1n}) - (U_A - x)^2(f^*)'(U_{2n})}{2(U_{A+k} - U_A)},$$

with U_{1n} and U_{2n} are r.v.s between U_A and U_{A+k} w.p.1. Since $U_{A+k} \to x$ and $U_A \to x$ w.p.1, from Lemma 3.2(a), and since $(f^*)'$ is continuous at x, we have $(f^*)'(U_{1n}) \to (f^*)'(x)$ and $(f^*)'(U_{2n}) \to (f^*)'(x)$ w.p.1. Therefore,

$$W_n(x) - f^*(x) = (\tfrac{1}{2}(U_A + U_{A+k}) - x)(f^*)'(x) + o_p(U_{A+k} - U_A).$$

Note that, from (3.1),

$$|\tfrac{1}{2}(U_A + U_{A+k}) - x| \le U_{A+k} - U_A = (k/n)/V_n(x)W_n(x).$$

Recall that $W_n(x) \to f^*(x)$ w.p.1 from Lemma 3.3(b) and $V_n(x) \to 1$ w.p.1 from (3.3) and the conclusion then follows since $k^{3/2} = o(n)$.

For the proof of Theorem 4.2, we observe that Lemma 4.2 has taken care of the second term of the expression in (4.1), and therefore what remains to be shown is the following result, which can be proven in a manner similar to Lemma 3 of Kim and Van Ryzin (1980) with the replacement of $f(x)$ *by* $f^*(x)$.

LEMMA 4.4. *Under the conditions of Theorem* 4.2, $k^{1/2}(W_n(x) - f^*(x)) \to b(x)$ *in p. as* $n \to \infty$.

Theorems 4.3 and 4.4 on the limiting distribution of $\lambda_n(x)$ are direct applications of Theorems 4.1 and 4.2, respectively.

THEOREM 4.3. *Suppose that the hypotheses of Theorem* 4.1 *hold. Then,*

$$k^{1/2}(\lambda_n(x) - \lambda(x)) \to_d N(0, \lambda^2(x)), \quad \text{if} \quad L(x) < 1. \tag{4.2}$$

PROOF. Rewrite

$$\begin{aligned} k^{1/2}(\lambda_n(x) - \lambda(x)) &= k^{1/2}n^{-1/2}f_n^*(x) \cdot n^{1/2}[(1 - L_n(x))^{-1} - (1 - L(x))^{-1}] \\ &\quad + (1 - L(x))^{-1} \cdot k^{1/2}(f_n^*(x) - f^*(x)). \end{aligned} \tag{4.3}$$

Applying the result of Theorem 4.1 to (4.3), the proof will be completed if we show that the first term tends to zero in probability. But this follows from Theorem 3.1, the fact that $(k/n)^{1/2} \to 0$, and the result that

$$n^{1/2}[(1 - L_n(x))^{-1} - (1 - L(x))^{-1}] \to_d N(0, L(x)/(1 - L(x))^3).$$

THEOREM 4.4. *Suppose that the hypotheses of Theorem* 4.2 *are satisfied, then,*

$$\begin{aligned} k^{1/2}(\lambda_n(x) - \lambda(x)) \to_d N(b(x)/(1 - L(x)), \lambda^2(x)), \quad &\text{if} \quad L(x) < 1, \\ &\text{where} \quad b(x) = d(f^*)''(x)/6(f^*)^2(x). \end{aligned} \tag{4.4}$$

PROOF. The result follows by the proof of Theorem 4.3 and directly applying the result of Theorem 4.2 to (4.3).

5. Concluding remarks.

REMARK 5.1. To illustrate the differences between Theorems 4.3 and 4.4 and their respective conclusions (4.2) and (4.4), we have the following remark. When $k^{3/2} = o(n)$, conclusions (4.2) of Theorem 4.3 and (4.4) of Theorem 4.4 are identical. However, if $\lim_{n\to\infty}(k^{3/2}n^{-1}) = c > 0$, Theorem 4.3 provides no answer, while Theorem 4.4 yields an answer provided f^* is twice continuously differentiable at x and the additional symmetry conditions about $R_n(x)$ for $A_n(x)$ and $(A_n(x) + k)$, as given by (i)–(iv) in Theorem 4.2, are satisfied. The conclusion in this case can be shown to be (4.4) with $\lim_{n\to\infty}(4n^{-2}k_n^{5/2}) = d$. For example if $k_n = n^{4/5}d_n^2$, the rate of convergence of the normalizing term in the limiting distribution in (4.4) is now $n^{2/5}d_n$. If $d_n \to d > 0$, then the bias term $b(x)/(1 - L(x))$ appears in the limiting distribution. If $d = 0$, then the limiting distribution has mean 0 (i.e., no bias term). It is worth noting that in Theorem

4.4 if $d = 0$ and d_n is properly chosen, the rate of convergence is the best attainable rate for nonnegative estimators (see Farrell, 1972).

REMARK 5.2. By using kernel function methods of density estimation, Blum and Susarla (1980) and Tanner and Wong (1983) in the censored data case derived kernel function methods of estimating $\lambda(x)$. Although the respective methods are slightly different, both estimators yield a $\lambda_n^*(x)$ satisfying

$$(nh_n)^{-1}(\lambda_n^*(x) - \lambda(x)) \rightarrow_d N(0, c_0\lambda(x)/L(x)),$$

where $c_0 = \int K^2(y)\, dy$, $K(y) \geq 0$ is a symmetric kernel satisfying certain conditions, provided $(nh_n)n^{-5/4} = o(1)$ and $(f^*)''$ is continuous at x. We can compare the asymptotic efficiency of their estimators with ours by taking $nh_n \sim k_n$. Comparing the ratio of the asymptotic variances (theirs to ours), we get the asymptotic relative efficiency from the above results and Theorem 4.4 to be

$$\text{As eff. } (\lambda_n^*(x), \lambda_n(x)) = (c_0\lambda(x)/L(x))/\lambda^2(x) = c_0/f^*(x),$$

$f^*(x) = f(x)(1 - G(x))$. Note that our estimator is more efficient in the tail than the kernel estimator whenever the subdensity function $f^*(x)$ is decreasing in the tail, a common situation in survival analysis.

REMARK 5.3. It is also natural to consider the limiting distribution of the maximal deviation of the hazard rate estimator λ_n over intervals. This is done in Liu and Van Ryzin (1984).

Acknowledgement. The authors are grateful to the referees for their helpful comments and suggestions.

REFERENCES

BLUM, J. R. and SUSARLA, V. (1980). Maximal deviation theory of density and failure rate function estimates based on censored data. In *Multivariate Analysis* V, (ed. by P. R. Krishnaiah) 213–222. North-Holland, Amsterdam.

BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.

BURKE, M. D. (1983). Approximations of some hazard rate estimators in a competing risks model. *Stochastic Process. Appl.* **14** 157–174.

FÖLDES, A., REJTÖ, L. and WINTER, B. Nonparametric estimators for randomly censored data, II: Estimation of density and failure rate. *Period. Math. Hungar.* **1** 15–19.

FARRELL, R. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **48** 170–180.

KIM, B. K. and VAN RYZIN, J. (1975). Uniform consistency of a histogram density estimator and model estimation. *Comm. Statist.* **4** 303–315.

KIM, B. K. and VAN RYZIN, J. (1980). On the asymptotic distribution of a histogram density estimator. *Colloquia Societatis Janos Bolyai 32, Nonparametric Statistical Inference*, Budapest. 483–499.

LIU, R. and VAN RYZIN, J. (1983). A histogram estimator of the hazard rate with censored data. Columbia University Stat. Report B-30.

LIU, R. and VAN RYZIN, J. (1984). The asymptotic distribution of the normalized maximal deviation of a hazard rate estimator under random censoring. Columbia University Stat. Report B-41.

PRAKASA RAO, B. L. S. and VAN RYZIN, J. (1985). Asymptotic theory for two estimators of the generalized failure rate. In *Statistical Theory and Data Analysis*, (ed by K. Matusita). 547–563.

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

VAN RYZIN, J. (1973). A histogram method of density estimation. *Comm. Statist.* **2** 493–506.

VAN RYZIN, J. (1977). A central limit theorem for randomly selected summands. *Bull. Int. Stat. Inst.* 438–441.

YANDELL, B. S. (1983). Nonparametric inference for rates with censored data. *Ann. Statist.* **11** 1119–1135.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903

DIVISION OF BIOSTATISTICS
AND DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10032