The Annals of Statistics,
1983, Vol. 11, No. 4, 1119–1135

NONPARAMETRIC INFERENCE FOR RATES WITH CENSORED SURVIVAL DATA[1]

BY BRIAN S. YANDELL

University of Wisconsin

This paper concerns nonparametric inference for hazard rates with censored serial data. The focus is upon "delta sequence" estimators of the form $h_n(x) = \int K_b(x, y)\, dH_n(y)$ with K_b integrating to 1 and concentrating mass near x as $b \to 0$. H_n is the Nelson-Aalen empirical cumulative hazard. Strong approximation and simultaneous confidence bands are derived for Rosenblatt-Parzen estimators, with $K_b(x, y) = w((x - y)/b)/b$, $b = o(n^{-1})$, and $w(\cdot)$ a well-behaved density. This work generalizes global deviation and mean square deviation results of Bickel and Rosenblatt and others to censored survival data. Simulations with exponential survival and censoring indicate the effect of censoring on bias, variance, and maximal absolute deviation. Data from a survival experiment with serial sacrifice are analysed.

1. Introduction. This paper concerns nonparametric inference for hazard rates with censored survival data. A sample of n individuals is observed from birth to the time of death or censoring, with the censoring process independent of the survival process. The problem is to estimate the hazard rate and to infer certain properties of this rate. For instance, one may wish to test whether the rate follows some parametric form, or whether it differs substantially from that of another population. One may want to graph the rate function without making assumptions about the unknown survival process, and to visually compare rate functions from different samples.

Inference about rates has been studied in survival analysis, demography, reliability, and other fields. For recent reviews see David and Moeschberger (1978), Prentice *et al.* (1978), and Hoem (1976). Rate estimation is closely tied to density estimation through the relation $f(x) = h(x)S(x)$. Recent reviews of density estimation include Bean and Tsokos (1980) and Wertz (1978). Wertz and Schneider (1978) have other density and rate references. Nonparametric techniques are characterized by choosing an estimate from a broad class which cannot be represented by a finite-dimensional parameter. For example, the maximum likelihood estimate of the death rate among all distributions with continuous non-negative rates is a right-continuous step function with jumps at every death. This rough estimator may be smoothed by reducing the class in various ways (see Barlow *et al.* 1972 and reviews).

The method adopted in this paper uses "delta sequences" (Susarla and Walter, 1981) and the kernel approach introduced by Rosenblatt (1956) and Parzen (1962). Földes *et al.* (1981) and Yandell (1981) generalized the density estimators to randomly right-censored data. Watson and Leadbetter (1964a, b) proposed three rate estimators based on this kernel. Yandell (1981) generalized these rate estimators to randomly right-censored survival data. Section 2 contains definitions and properties of one estimator presented in detail here.

Received December 1981; revised May 1983.

[1] This research was part of the author's dissertation for the Biostatistics Program, University of California, Berkeley. Partial support came from National Institute of Enviromental Health Sciences Grant 2-R01-ES01299, Environmental Protection Agency Grant R805379-01-6, School of Public Health Federal Traineeship 748, and University of California Regents Fellowships in 1975–76 and 1977–78.

AMS 1980 *subject classifications.* Primary 62G05; secondary 62G15, 62G10, 62P10.

Key words and phrases. Delta sequence, hazard rate, kernel estimation, maximal deviation, simultaneous confidence band, survival rate.

The following works use the Rosenblatt-Parzen kernel. McNichols and Padgett (1982) examined the rate under a proportional hazards model. Ramlau-Hansen (1983) considered the more general multiplicative intensity model (Aalen, 1978). Tanner and Wong (1982) studied point-wise properties of the kernel estimator. Clevenson and Zidek (1977) studied histogram and uniform kernel window Bayesian rate estimators for a Poisson process. Bartoszyński *et al.* (1981) examined uniform kernel and other rate estimators in studying cancer metastasis.

One can obtain a global measure of deviation for Rosenblatt-Parzen estimators which leads to simultaneous confidence bands and graphical tests. Denote the maximal deviation by $M_n = \| (nb/V_f(x))^{1/2}(f_n(x) - f(x)) \|$, with $(nb)^{-1}V_f$ the asymptotic variance process of f_n, and $\| \cdot \|$ the supremum over x in $[0, T]$. Bickel and Rosenblatt (1973) showed in the noncensored case for suitable r_n and d_n and appropriate conditions, that for all u,

$$P\{r_n(M_n - d_n) < u\} \to \exp(-2e^{-u}).$$

Rosenblatt (1976) extended this result to multivariate densities. Simultaneous confidence bands for some kernel estimators of hazard rates in the absence of censoring were derived by Rice and Rosenblatt (1976) and Sethuraman and Singpurwalla (1981). Burke and Horváth (1982) derived simultaneous confidence bands for hazard rates in a competing risks model. Bounds on the rate of convergence of the distribution of maximal deviation to the limiting distribution and a second order correction term were obtained by Konakov and Piterbarg (1979) for the univariate kernel density estimator.

The maximal deviation and simultaneous confidence band results are extended to Rosenblatt-Parzen kernel estimators of rates and densities for censored survival data in Section 4 of the present paper using strong approximation results derived in Section 3. Section 5 shows how to use these bands for graphical testing. Monte Carlo simulations in Section 6 indicate the effect of censoring and bandwidth, and the slow convergence to the limiting distribution for maximal deviations. Section 7 examines data from a survival experiment with serial sacrifice. Section 8 presents conclusion and remarks about fixed censorship.

2. Definitions and properties. Let $(X_1, D_1), \cdots, (X_n, D_n)$ be independent and identically distributed random pairs, with $X_i \geq 0$ being the "lifetime" and D_i the indicator of death $(D = 1)$ or censoring $(D = 0)$ for the ith individual. Denote the number of deaths in $[0, x]$ by $N(x) = \#\{X_i \leq x \mid D_i = 1\}$, and let $R(x) = \#\{X_i \geq x\}$ denote the number at risk of death or censoring at time $x \geq 0$. Let $R_i = R(X_i)$. Let $J(x) = I[R(x) > 0]$, and $J_i = J(X_i)$.

The hazard rate, denoted by $h(x)$, $x \geq 0$, is defined for small dx by

$$\Pr\{X_i < x + dx, D_i = 1 \mid X_i \geq x\} = h(x)\,dx + o(dx), \quad x \geq 0.$$

Let H denote the cumulative hazard, that is $H(x) = \int_0^x h(t)\,dt$. The survival curve is denoted by $S(x) = 1 - F(x) = \exp(-H(x))$. In addition, denote the survival sub-distribution function by $\tilde{F}(x) = \Pr\{X_1 \leq x \mid D_i = 1\}$ and the sub-density by $\tilde{f} = \tilde{F}'$. If C, the censoring curve, is differentiable, one can define the censoring rate $g(x)$ such that for small dx,

$$Pr\{X_i < x + dx, D_i = 0 \mid X_i \geq x\} = g(x)\,dx + o(dx), \quad x \geq 0,$$

with cumulative censoring hazard $G = -\log(C)$.

The kernel rate estimator is based on the empirical cumulative rate (Nelson 1972; Aalen 1978)

$$H_n(x) = \int_0^x \frac{J}{R}\,dN = \sum_{\{X_i \leq x\}} \frac{D_i J_i}{R_i},$$

with the convention that $J/R = 0$ if $J = 0$. The hazard rate estimator for $x \in (0, T)$, with some T such that $S(T)C(T) > 0$, is

$$h_n(x) = \int_0^T K_b(x, y)\, dH_n(y) = \textstyle\sum_{i=1}^n I(X_i \le T) K_b(x, X_i) \dfrac{D_i J_i}{R_i}.$$

This generalizes an estimator of Watson and Leadbetter (1964a, b). T is estimated consistently in practice.

Several assumptions are needed. The distributional assumptions are as follows:

S1. S, C and L are continuous and non-increasing, and $L = SC$.
S2. There is a $\delta > 0$ such that $1 - \delta \ge L(T) \ge \delta$.
S3. The density f is continuous, positive, and bounded on $[0, T]$.
S4. The second derivative f'' of f exists and is bounded on $[0, T]$.
S5. $f^{1/2}$ is absolutely continuous and has bounded derivative on $[0, T]$.
S6. The censoring cumulative rate $G = -\log C$ is absolutely continuous and has bounded derivative g over $[0, T]$.

S1–S2 are needed to use strong uniform consistency results (Földes and Rejtö, 1981). Further, the hazard rate must be reasonably behaved (S3). S4 helps insure that the bias is asymptotically negligible. S5 is needed for the strong approxmations of Section 4 (Bickel and Rosenblatt, 1973). S6 is used to modify strong approximation results for censored data.

Kernel assumptions follow:

K1. $\displaystyle\int_0^T K_b(x, y) h(y)\, dy \to h(x)$ as $b \to 0$ for each bounded $h(\cdot)$, $x \in (0, T)$.

K2. $K_b \ge 0$ and for all $x \in (0, T)$,

(i) $\displaystyle\int K_b(x, y)\, dy = 1$

(ii) $\displaystyle\sup_{r>0} r \int_{|x-y|>r} K_b(x, y)\, dy = O(b)$

(iii) $\| K_b(\cdot, \cdot) \| = O(b^{-1})$

(iv) for all $r > 0$, $\sup\{K_b(x, y);\, |x = y| > r\} \to 0$ as $b \to 0$.

K3. $K_b(x, y) = w((x - y)/b)/b$. The weight function w is absolutely continuous with derivative w' on $[-J, J]$ and $w = 0$ off $[-J, J]$, some $J < \infty$.
K4. The bandwidth $b \to 0$ and $nb \to \infty$ as $n \to \infty$.
K5. $w(\cdot)$ is symmetric (about 0) and $z^2 w(z)$ is integrable.
K6. The following integral over $\{|z| \ge 3\}$ is bounded:

$$\int |z|^{3/2} [\log \log |z|]^{1/2} \{|w'(z)| + |w(z)|\}\, dz$$

K7. $nb/\log n \to \infty$ as $n \to \infty$.

Kl–K2 define delta sequences of positive type (Susarla and Walter, 1981). K3–K4 restrict attention to the kernels of Rosenblatt (1956) and Parzen (1962). K5 helps insure asymptotic unbiasedness, while K6 was introduced by Bickel and Rosenblatt (1973) for the strong approximations. K7 allows us to conclude that a kernel density estimator is uniformly bounded if f is, based on uniform convergence results (Collomb, 1978; Silverman, 1978).

Properties of the estimator h_n have been studied by various authors recently. The bias and covariance arise as special cases of Ramlau-Hansen (1983). See Tanner and Wong (1982) for an exact expression of the finite variance.

THEOREM 2.1. *Let* $x, t \in (0, T)$.

a) *If* K1–K2 *and* S1–S3 *hold, then*

$$\mathrm{Bias}(h_n(x)) = \int_0^T K_b(x, y)\, dH(y) - h(y) + \int_0^T K_b(x, y)[1 - L(y)]^n\, dH(y)$$

If K3–K5 *and* S4 *also hold, then*

$$\mathrm{Bias}(h_n(x)) = \frac{1}{2} h''(x) b^2 \int w(y) y^2\, dy + o(b^2) + O((1 - \delta)^n).$$

b) *If* K1–K2 *and* S1–S3 *hold, then*

$$\mathrm{Cov}(h_n(x), h_n(t)) = \int_0^T K_b(x, y)\, K_b(t, y) E\left\{\frac{J(y)}{R(y)}\right\} dH(y).$$

If K3–K4 *hold, then*

$$nb\, \mathrm{Cov}(h_n(x), h_n(t)) \to 0 \quad \text{if} \quad x \neq t$$

$$nb\, \mathrm{Var}(h_n(x)) \to h(x) L^{-1}(x) \int w^2(z)\, dz = V_h(x).$$

Ramlau-Hansen (1983) obtained mean square and uniform consistency. Yandell (1981) and Burke and Horváth (1982) derived rates of strong uniform consistency.

Ramlau-Hansen (1983) proved pointwise asymptotic normality, as did Tanner and Wong (1982) under different assumptions, for the Rosenblatt-Parzen estimator.

3. Strong approximation. Under suitable conditions on the censored survival process and the Parzen (1962) kernel, the pivot process

$$W_n(x) = (nb/V_h(x))^{1/2}(h_n(x) - Eh_n(x)), \quad 0 \le x \le T,$$

can be approximated by a Gaussian process

$${}_3W_n(x) = \left(b \int w^2\right)^{-1/2} \int_0^T w\left(\frac{x - y}{b}\right) dZ(y), \quad 0 \le x \le T,$$

in which Z is a version of Brownian motion. Similar approximations for other kernel density and rate estimators are possible (Yandell, 1981; Burke and Horváth, 1982). The approximations parallel Bickel and Rosenblatt (1973). Throughout this section, for fixed x and n let $K(y) = w((x - y)/b)/b$ with assumptions K1–K3 and S1–S2 satisfied. Let $\| Y \|$ denote the sup $| Y(x) |$ over $[0, T]$.

The sub-distribution $\tilde{F}$ agrees on the range $[0, T + J)$ with the distribution of the random variable $\tilde{X}$ defined to be X if $D = 1$, and $X + T + 2J$ if $D = 0$, in which $2J$ is the window width of assumption K3. $\tilde{f}_n = \int K\, d\tilde{F}_n$ is a kernel estimator for the density of $\tilde{X}$ on $[0, T]$. One may therefore use results for the non-censored estimator. Let $Z_n^0(t) = n^{1/2}(\tilde{F}_n(\tilde{F}^{-1}(t)) - t)$ and

$$Y_n(x) = \left(\frac{b}{\tilde{f}(x)}\right)^{-1/2} \int_0^T K\, dZ_n^0(\tilde{F}).$$

Define ${}_0Y_n$ and ${}_1Y_n$ by replacing Z_n^0 by Z^0 and Z, respectively. Here Z^0 and Z are versions of the Brownian bridge and Brownian motion, respectively. Let ${}_2Y_n = (b/\tilde{f}(x))^{1/2} \cdot \int K\tilde{f}^{1/2}\, dZ$, and ${}_3Y_n = b^{1/2} \int K\, dZ$. The following theorem and lemma are central to the next propositions.

THEOREM 3.A. Kolmos, Major and Tusnady, 1975). *Let* S1–S3 *hold. A sequence of Brownian bridges can be constructed so that*

$$\sup_{0\le x\le 1} | Z_n^0 - Z^0 | = O_p(n^{-1/2}\log n).$$

LEMMA 3.1. (Bickel and Rosenblatt, 1973; Rosenblatt, 1976). *Suppose the Y processes are as above. Let* S1 *and* K1–K3 *hold.*

(i) *If* K3 *and* S3 *hold, then* $\| Y_n - {}_0Y_n \| = O_p((nb)^{-1/2}\log n)$.
(ii) *If* S2–S3 *hold, then* $\| {}_0Y_n - {}_1Y_n \| = O_p(b^{1/2})$.
(iii) *If* S2–S3, S5 *and* K6 *hold, then* $\| {}_2Y_n - {}_3Y_n \| = O_p(b^{1/2})$.

If S1 *holds, then* $h = \tilde{f}/L$, *and* $dH_n = L_n^{-1}\, d\tilde{F}_n$. *Thus,*

$$h_n(x) = \int_0^T K\, dH_n = \int_0^T KL_n^{-1}\, d\tilde{F}_n.$$

The strategy below involves replacing L_n^{-1} by L^{-1} and proceeding by a series of approximations using Lemma 3.1. Let

$$h_n^*(x) = \int_0^T K(y)L^{-1}(y)\, d\tilde{F}_n(y).$$

PROPOSITION 3.1. *Let* K1–K4, K7 *and* S1–S3 *hold. Then*

$$\| (nb/V_h)^{1/2}(h_n - h_n^*) \| = O_p((b \log n)^{1/2}).$$

PROOF. Clearly, S2 implies

$$| h_n(x) - h_n^*(x) | \le \int_0^T K\, | L_n^{-1} - L^{-1} |\, d\tilde{F}_n \le \tilde{f}_n(x)\, \| L^{-1} - L_n^{-1} \|.$$

By Földes and Rejtö (1981), $\| L_n - L \| = O_p((\log n/n)^{1/2})$. Therefore, since $L(T) \ge \delta > 0$, the same rate holds for $\| L_n^{-1} - L^{-1} \|$. K1–K4, K7 and S3 imply that $\| \tilde{f}_n - \tilde{f} \| \to 0$ a.s. (Collomb, 1978; Silverman, 1978). Hence $\| \tilde{f}^{1/2} - \tilde{f}_n/\tilde{f}^{1/2} \| \to 0$ a.s. The proposition obtains by combining terms and noting that $\tilde{f}^{1/2}$ is bounded.

The pivot process W_n, with L_n replaced by L, may be written as

$$W_n^*(x) = \left(\frac{b}{V_h(x)}\right)^{1/2} \int_0^T K(y)L^{-1}(y)\, dZ_n^0(\tilde{F}(y)).$$

Define ${}_0W_n$ and ${}_1W_n$ by replacing Z_n^0 by Z^0 and Z, respectively. Denote by

$${}_2W_n(x) = \left(\frac{b}{V_h(x)}\right)^{1/2} \int_0^T K\tilde{f}^{1/2}L^{-1}\, dZ = \left(\frac{b}{V_h(x)}\right)^{1/2} \int_0^T K\left(\frac{h}{L}\right)^{1/2} dZ$$

$${}_3W_n(x) = \left(\frac{b}{\int w^2}\right)^{1/2} \int_0^T K\, dZ$$

in which Z is a version of Brownian motion on $[0, 1]$ for ${}_1W_n$ and on $(-\infty, \infty)$ for ${}_2W_n$ and ${}_3W_n$. Consider the following:

PROPOSITION 3.2. *Let the W processes be as above, and* K1–K3, S1–S3 *hold. Then*
(i) *If* S6 *holds, then* $\| W_n^* - {}_0W_n \| = O_p((nb)^{-1/2}\log n)$.
(ii) $\| {}_0W_n - {}_1W_n \| = O_p(b^{1/2})$.
(iii) *If* K6, S5 *and* S6 hold, then $\| {}_2W_n - {}_3W_n \| = O_p(b^{1/2})$.

PROOF. (i) By S1–S2, $C \geq \delta$. Thus

$$| W_n^*(x) - {}_0W_n(x) | \leq \delta^{-1} | Y_n(x) - {}_0Y_n(x) | \left(\int w^2 \right)^{-1/2} + \delta^{-1} \| Z_n^0 - Z^0 \| (bV_h(x))^{-1/2} \int_0^T \left| g(y) w\left(\frac{x-y}{b}\right) \right| dy.$$

The first term is of proper order by Lemma 3.1 and K1–K3. S6 insures that g exists and is bounded. The variance V_h is continuous and positive since L and h are, by S1 and S3. Together with Theorem 3.A, this gives the order for part (i).

(ii) Note that $| {}_0W_n(x) - {}_1W_n(x) | = (L(x)/\int w^2)^{1/2} | {}_0Y_n(x) - {}_1Y_n(x) |$. Thus (ii) follows from Lemma 4.1, part (ii), and $L \leq 1$.

(iii) In a fashion similar to (i),

$$| {}_2W_n(x) - {}_3W_n(x) | \leq \delta^{-1} | {}_2Y_n(x) - {}_3Y_n(x) | \left(\int w^2 \right)^{-1/2} + \left(\frac{b}{4V_h(x)} \right)^{1/2} \int_0^T | Z(yb + x) | g(yb + x) \left(\frac{h(yb + x)}{L(yb + x)} \right)^{1/2} | w(y) | \, dy.$$

S1–S2 and Lemma 3.1 make the first term of the proper order. S1–S3 and S6 bound uniformly the term $g(h/L)^{1/2}$ in the integral. K6 and the law of the itegrated logarithm for Brownian motion ensure that the second term is $O_p(b^{1/2})$.

The Gaussian processes ${}_1W_n$ and ${}_2W_n$ have the same covariance structure and hence have the same law. Hence by Propositions 3.1 and 3.2,

$$\| W_n - {}_3W_n \| = O_p(b \log n + b^{1/2} + (nb)^{-1/2} \log n).$$

If in addition K5 and S4 hold, $Eh_n(\cdot)$ may be replaced by $h(\cdot)$ in W_n.

One may then substitute W_n for ${}_3W_n$ in a sequence of functionals, such as maximal deviation or mean square deviation, provided $b = b(n)$ converges to 0 at the correct rate. Let M_n be a sequence of functionals satisfying Lipschitz condition such that, for some $J_n > 0$,

$$| M_n(x) - M_n(y) | \leq J_n | x - y |.$$

If $\| W_n - {}_3W_n \| = o_p(1/J_n)$, then $M_n(W_n)$ converges in law if and only if $M_n({}_3W_n)$ does, and to the same limit.

REMARK. Censoring has been assumed to be random right-censoring for this work. The results of Meier (1975) and the comments of Breslow and Crowley (1974) and Aalen (1978) indicate how one could proceed in the case of fixed right-censoring. If one knows the censoring form, then it may be incorporated directly. The assumption S6 that $G = -\log C$ has a continuous derivative may be replaced by

S6′. The jumps in G are uniformly bounded over $[0, T]$. That is, $| dG | < M$ for some $M < \infty$.

Then one needs to replace $| g(y) | \, dy$ by $| dG(y) |$ at every occurrence in the proof of Proposition 3.2.

4. Simultaneous confidence bands. The strong approximation results of the previous section show that we have the same approximating process for the censored case as for the noncensored case. Therefore we obtain the same limit as Bickel and Rosenblatt (1973). This is inverted to derive simultaneous confidence bonds. We extend Konakov and Piterbarg (1979) to allow refinement of the asymptotic approximations, adding a second-order term to the limiting probability.

Let $T_n = X_{(i)}$, the ith ordered lifetime, with $i = [n(1-\delta)]$. Thus $T_n \to T = L^{-1}(\delta) < \infty$, converging in probability. The following is due to Bickel and Rosenblatt (1973) (see Rice and Rosenblatt, 1976).

THEOREM 4.B. *Let* $r_n = (2\log(T_n/b))^{1/2}$, *and* $d_n = r_n + (\log w^*)/r_n$, *with*

$$w^* = r_n[w^2(J) + w^2(-J)](8\pi)^{-1/2}\left(\int w^2\right)^{-1} \quad \text{if} \quad w(J) > 0$$

$$= \left[\int w'^2 \Big/ \int w^2\right]^{1/2} \Big/ 2\pi \qquad \text{if} \quad w(J) = 0$$

in which J *is the constant of assumption* K3. *Let* $M_n = \| {}_3W_n \|$, *the* sup *over* $[0, T_n]$. *Let* K1–K4 *hold. Then for all* x,

$$\Pr\{r_n(M_n - d_n) > x\} \to \exp(-2e^{-x}).$$

The next result for censored rates follows directly:

THEOREM 4.1. *Let* K1–K7 *and* S1–S6 *hold. Let* $M_n = \| (nb/V_h)^{1/2}(h_n - h) \|$, *in which the* sup *is over* $[0, T_n]$. *Let* r_n *and* d_n *be defined as in Theorem* 4.B. *Then, for* $-\infty < x < \infty$,

$$\Pr\{r_n(M_n - d_n) < x\} \to \exp(-2e^{-x}).$$

Simultaneous confidence bands arise from these results using Slutsky's Theorem on V_h. First, estimate L by L_n. Replace h by h_n in V_h to yield the symmetric form

$$h_n \pm k(h_n/L_n)^{1/2}$$

with $k = (d_n + x/r_n)(\int w^2/nb)^{1/2}$ and $x = \log(-2/\log(1-\alpha))$. The asymmetric band arises by inverting a quadratic in h:

$$h_n + \frac{k^2}{2L_n} \pm k\left(\frac{h_n}{L_n}\right)^{1/2}\left[1 + \frac{k^2}{4h_nL_n}\right]^{1/2}.$$

The latter band is wider for all x, reflecting the more conservative substitution. This band is above 0 unless $h_n(x) = 0$, when it becomes $[0, k^2/L_n(x)]$. The symmetric band may have a negative lower bound, which in pratice is usually truncated to 0. Similar results for competing risks models may be found in Burke and Horváth (1982).

Konakov and Piterbarg (1979, Theorems 3 and 4) obtained a second order expansion of the limiting distribution which depends on n through the term t_n. One needs additional assumptions on the smoothness of the kernel window $w(\cdot)$, and on the bandwidth b. The strong approximation argument presented in Section 3 extends their result to the censored estimators.

THEOREM 4.2. *Let* K1–K7 *and* S1–S6 *hold. Let* $b = n^{-p}$, $1/3 < p < 1/2$. *Let* r_n *and* w^* *be as in Theorem* 4.B. *Suppose* $w(J) = w(-J) = 0$, *and* $\int (w'')^2 < \infty$. *Then for* $-\infty < y < \infty$,

$$\Pr\{r_n(M_n - d_n) < y\} = \exp(-2\exp(-y - (y + \log w^*)^2/2r_n^2)) + L_1(n, y).$$

If $y > r_n(1 - d_n)$ *and* $b/T_n < (6/\pi w^*)^{p^2}$, *then*

$$| L_1(n, x) | < un^{-v}e^{-2x}x^2$$

for some $u > 0$ *and* $v > 0$.

PROOF. The proof rests on the substitution, for fixed n, $d_n + y/r_n = l_n + x/l_n$, with $l_n^2 = r_n^2 + 2 \log w^*$. See Konakov and Piterbarg (1979) and Yandell (1981).

REMARKS. (1) Pointwise asymptotic confidence bands arise in similar form to the above, with $k = z(\int w^2/nb)^{1/2}$, in which z is the upper $(1 - \alpha/2)$ point of the normal distribution. However, Sacks and Ylvisaker (1981) show that the optimal bandwidth b depends on f for pointwise density estimation.

(2) One can approximate the variance $V_h(\cdot)$ by

$$V_{h;n}(x) = b^{-1} \int_0^T w^2\left(\frac{x - y}{b}\right) dV_{H;n}(y),$$

in which $V_{H;n}(x)$ is an estimate of $V_H(x) = \int_0^x (J/R)\, dH$. This may more accurately depict the variance for small n, particularly at the boundaries of the data. The asymmetric band can be formulated with this variance estimate.

(3) In order to get x for a particular significance level α, one chooses

$$x = \log(-2/\log(1 - \alpha))$$

for Theorem 4.B and Theorem 4.1, or

$$x = -(r_n^2 + \log w^*) + r_n[r_n^2 + 2 \log w^* + 2 \log(-2/\log(1 - \alpha))]^{1/2}$$

for Theorem 4.2.

5. Testing. One may perform goodness-of-fit tests for the composite hypothesis $\mathbf{H}: h = h_0(\cdot, \theta)$, in which θ is an unknown real-valued parameter. See Bickel and Rosenblatt (1973) for goodness-of-fit tests based on maximal absolute deviation and mean square deviation which work here because of the strong approximation. A two-sample test of the hypothesis $\mathbf{H}: h_1 = h_2$ is possible using the next theorem.

THEOREM 5.1. *Let* K1–K7 *and* S1–S6 *hold for survival processes with censoring in two independent populations. Let h_i, $i = 1, 2$, represent the hazard rates. Let V_h be the variance process for $h(\cdot)$, the common (unknown) rate under the null hypothesis. For samples of size n_1 and n_2, $n = n_1 + n_2$, let $h_{i;n}$ be the kernel estimate of h_i, $i = 1, 2$, using the same b and $w(\cdot)$. Let r_n and d_n be defined in terms of b and w as in Theorem* 4.B. *For $0 \le x \le T_n$, $t_n = \min(T_{n_1}, T_{n_2})$, let*

$$W(x) = (n_1 n_2 b/(n V_h(x)))^{1/2}(h_{1;n}(x) - h_{2;n}(x)).$$

If $n_1/n \to \lambda$, for some $0 < \lambda < 1$, then under $\mathbf{H}$, *for any x,*

$$\Pr\{r_n(\| W \| - d_n) < x\} \to \exp(-2e^{-x}),$$

in which the sup is over $[0, T_n]$.

PROOF. Substitute n_1/n for λ by Slutsky's theorem. Write $W = (1 - \lambda)^{1/2} W_1 - \lambda^{1/2} W_2$ in which W_i is the deviation statistic for sample i as in Theorem 4.1. The strong approximation results imply that W, W_1 and W_2 have the same asymptotic distribution.

REMARKS. (1) Without assuming equal variance, the process

$$(n_1 n_2 b/(n_2 V_1 + n_1 V_2))^{1/2}(h_{1;n} - h_{2;n}),$$

with V_i the variance of h_i, has the same asymptotic distribution as W.

(2) One may choose $b = (n_1 b_1 + n_2 b_2)/n$. If $b_i = k_i n_i^{-p}$, $i = 1, 2$, $1/3 < p < 1/2$, then $b = kn^{-p}$, with $k = k_1 \lambda^{1-p} + k_2(1 - \lambda)^{1-p}$. Note that $k \ge k_1, k_2$.

(3) A graphical test arises (Al Wiggins, 1981) by plotting $h_{1;n}$ and $h_{2;n}$ and surrounding

them with a simultaneous confidence band which is an adjustment of the individual bands, the relative widths being proportional to $(V_i(x)/n_i)^{1/2}$ or to $n_i^{-1/2}$. The test rejects if any gap appears in the band between the two curves. Alternatively one could plot the difference $h_{1;n} - h_{2;n}$ with a simultaneous confidence band.

6. Monte Carlo simulation. Monte Carlo simulation studies are presented for pseudo-random samples of exponential survival and censoring. The exponential parameters were 1 for survival and several values between 0 and 2 for censoring. Numbers came from a log transformation of ranm, a uniform generator on the MathStat PDP 11/45 UNIX (a trademark of Bell Laboratories) computer system. The effects of censoring on the form of the confidence bands are investigated for individual trials of sample size 200. Empirical 80% pointwise confidence intervals with Monte Carlo trials and sample size 200 are compared to theoretical pointwise confidence intervals. Empirical distributions with 100 Monte Carlo trials of the maximal deviation statistic $r_n(M_n - d_n)$ are investigated at length, with sample sizes from 50 to 500. Kernel bandwidths derived from normal theory are scaled down by factors of .5, .25 and .125 to compare empirical distributions of the maximal deviation with the first and second-order theoretical curves. Results are plotted as theoretical vs. empirical significance level.

The kernel bandwidth and window were chosen to minimize the mean integrated square error (Rosenblatt, 1971). The window was the quadratic $w(x) = 1.5 - 6x^2$, $-0.5 \le x \le 0.5$, with $w(x) = 0$ outside this interval. The bandwidth is taken from the classical case of normal density and no censoring, i.e. $b = 4.483sn^{-0.2}$ with s^2 the sample variance of the mean survival time (Rosenblatt, 1971; Yandell, 1981). T_n was chosen as $S_n^{-1}(0.2)$.

Theory suggests estimating the asymptotic variance V_h by $h_n(x)L_n^{-1}(x) \int w^2$. However, for finite samples, the estimate $b^{-1} \int_0^t w^2((x-y)/b)L_n^{-1}(y)\, dH_n(y)$ better reflects the bias at the boundaries (0 and $T = -\ln(0.2)$). Simulations show that the asymptotic variance is too small near the boundary. The empirical 80% bands in Figure 1 show this underestimate near 0 by the asymptotic 80% pointwise confidence bands. The empirical bands were derived by dividing the time axis into 32 equal probability intervals of 0.025 and deter-

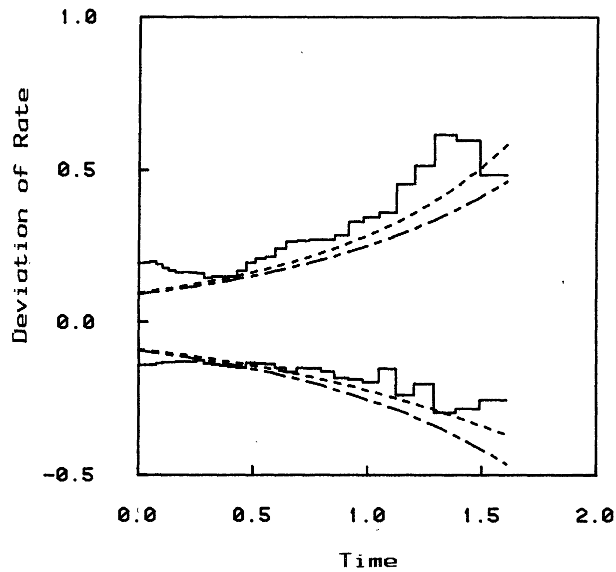

FIG. 1. 80% *simultaneous confidence bands for* $n = 200$. *Solid line: Monte Carlo bands for* 100 *trials (see text). Dot-dash: symmetric theoretical bands. Short-dash: asymmetric theoretical bands.*

mining the empirical distribution of maxima and minima of 100 simulations of the rate in each interval. All simulations presented use the finite sample variances.

Simultaneous 80% confidence bands are shown in Figure 2 for a constant rate with symmetric and asymmetric form. Note the greater dispersion in Figure 2(b) with 50% censoring over no censoring in Figure 2(a), reflecting the effect of censoring on the variance. The slight asymmetry of the empirical bands in Figure 1 suggests using asymmetrical bands in practice. The constant rate estimates are drawn, showing that a test (Section 5) would not reject a constant hazard rate.

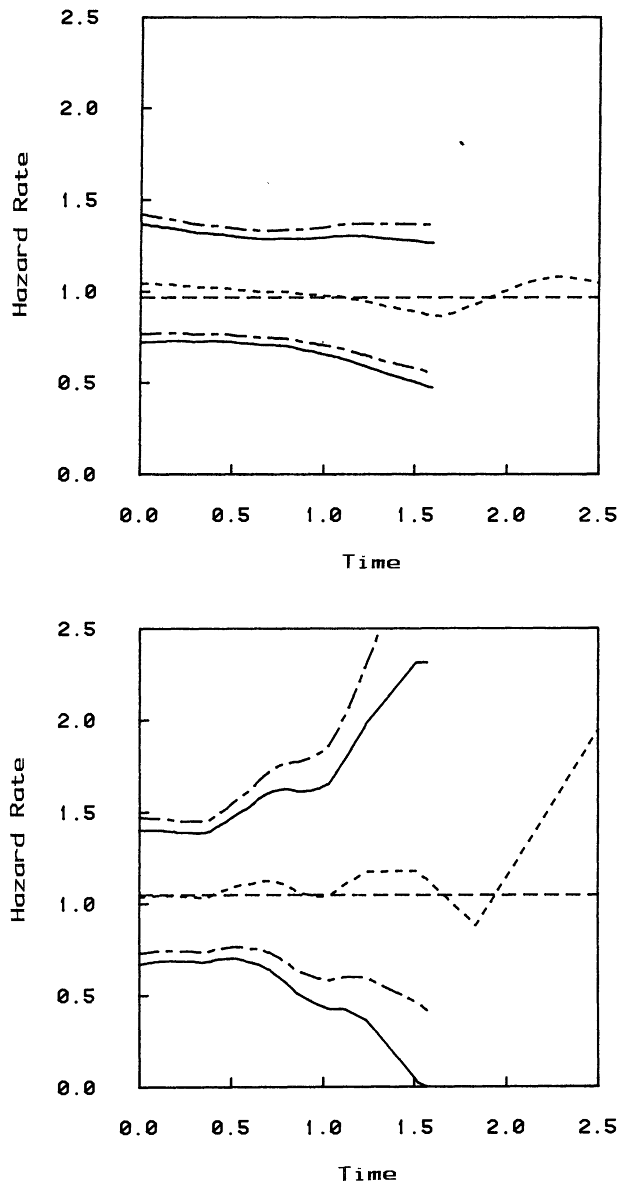

FIG. 2. *Rate estimate for* $n = 200$: (a) *no censoring*: (b) 50% *censoring* (*rate* 1). *Short-dash* = *kernel estimate*; *long-dash* = *constant rate estimate*; *solid line* = *symmetric* 80% *bands*; *dot-dash* = *asymmetric* 80% *bands*.

Empirical distribution functions (EDFs) of the maximal deviation statistic were computed from 100 Monte Carlo trials. The EDF for the rate deviation is compared with the theoretical curve $\exp(-2e^{-x})$ in Figure 3 for sample sizes 50, 200, and 500 and no censoring. Note the slow convergence. Thus 80% confidence bands of Figure 2 are wider than necessary, i.e. conservative. The times of maximal deviation were fairly evenly distributed. Figure 4 investigates the rate of convergence of the EDFs of maximal deviations for h_n as a function of bandwidth, reducing b by a factor of 2 for each curve, with $n = 200$ and no

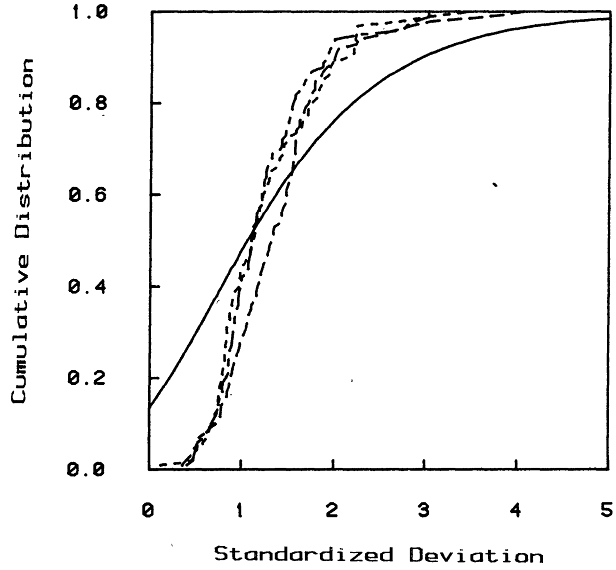

FIG. 3. *Empirical distribution of maximal deviation statistic. Solid line = theoretical* $\exp(-2e^{-x})$. *Other curves are* EDF *for* $n = 50$ (*short-dash*), 200 (*dot-dash*), *and* 500 (*long-dash*).

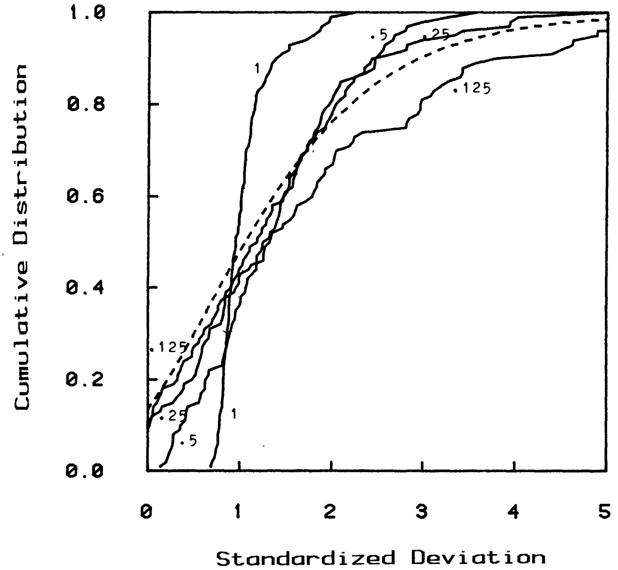

FIG. 4. EDF *of maximal deviation as bandwidth b varies* (n = 200, *trials* = 100). *Short-dash = theoretical. Solid EDFs are* $b = 4.4835n^{-0.2} \times 1.0$, $\times$ 0.5, $\times$ 0.25, *and* $\times$ 0.125. *No censoring.*

censoring. The convergence to $\exp(-2e^{-x})$ is fairly rapid compared to the convergence in n. However, Figure 5 shows that the bias increases markedly as the bandwidth is reduced. Figure 6 investigates the effect of censoring on the distribution of maximal deviation, with $n = 200$ and reduced bandwidth $b = 1.121sn^{-0.2}$. Note that the curves for 50% censoring ($h = g = 1$, long dash lines) are slightly above those for the noncensored data, but the characterstic shape remains. A detailed analysis of effect of censoring awaits more thorough Monte Carlo studies.

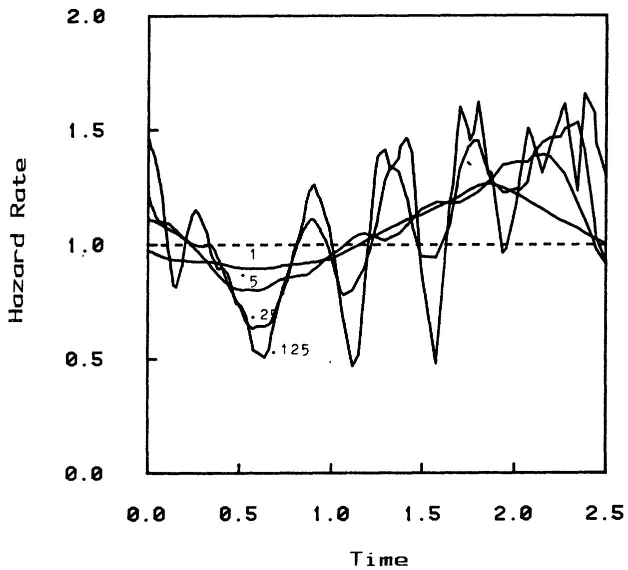

FIG. 5. *Hazard rate estimate as bandwidth b varies* ($n = 200$). *Short-dash = constant rate* (1). *See Figure 4 for solid curves. No censoring.*

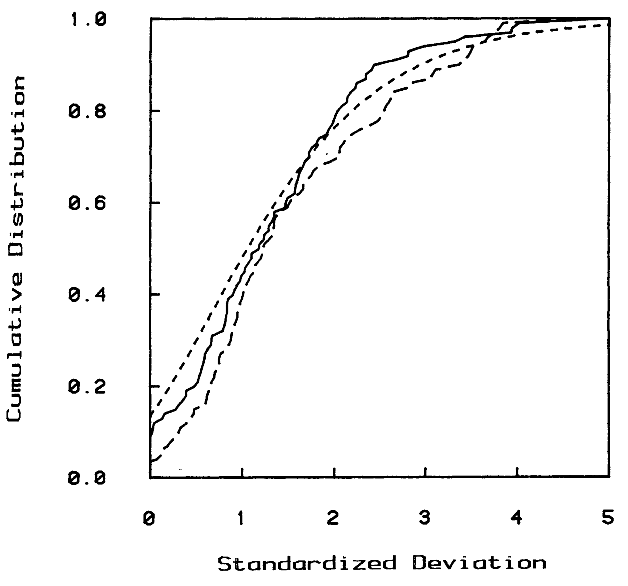

FIG. 6. EDF *of maximal deviation with censoring* ($n = 200$, *trials* = 100). *Short-dash = theoretical; solid = no censoring; long-dash = 50% censoring* (*rate* 1). *Bandwidth* $b = 1.1215n^{-0.2}$.

The relative difference between the theoretical level and empirical significance levels for the rate estimator is shown in Figure 7, for the first-order (a) and second-order (b) theory. Thus in practice one may want to have a second-order, and maybe even a third-order term to ensure some semblance of the correct significance probability for a simultaneous confidence band.

7. Data analysis. This section examines data from a survival experiment with serial sacrifice, designed to investigate the effect of a 300 rad dose of gamma irradiation on mice (Upton *et al.*, 1969). Animals in two groups, treated and control, died naturally or were

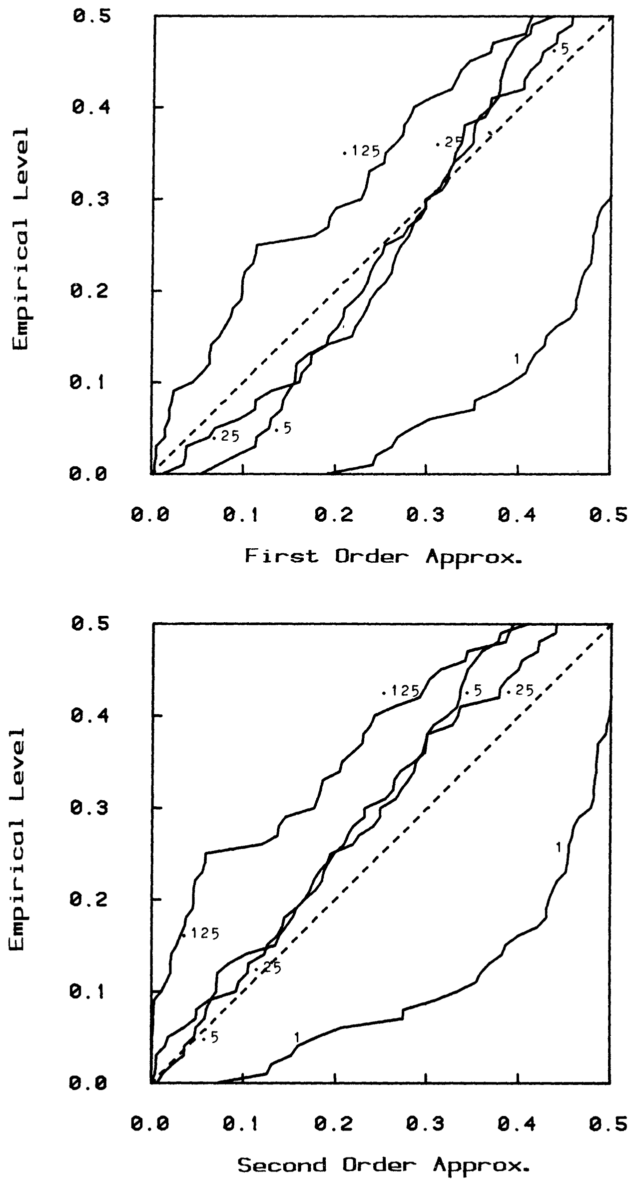

FIG. 7. *Significance levels from empirical study vs.* (a) *first order Bickel-Rosenblatt and* (b) *second order Konakov-Piterbarg. Short-dash = unity line. See Figure 4 for solid curves. No censoring.*

sacrificed. The data consist of time and mode of death; data on pathologies were not used here. There were 1080 control mice and 1454 treated mice, in which 361 control and 343 treated mice were sacrificed. We investigate the following questions: (1) By treatment group, is the death rate constant over the experiment? What do the death rates look like? (2) Do death rates differ between treated and control groups? The estimates use the quadratic window and bandwidth computed by normal theory as in Section 6. Simultaneous 80% confidence bands were derived using the finite sample variance and the results of Theorem 6.1.

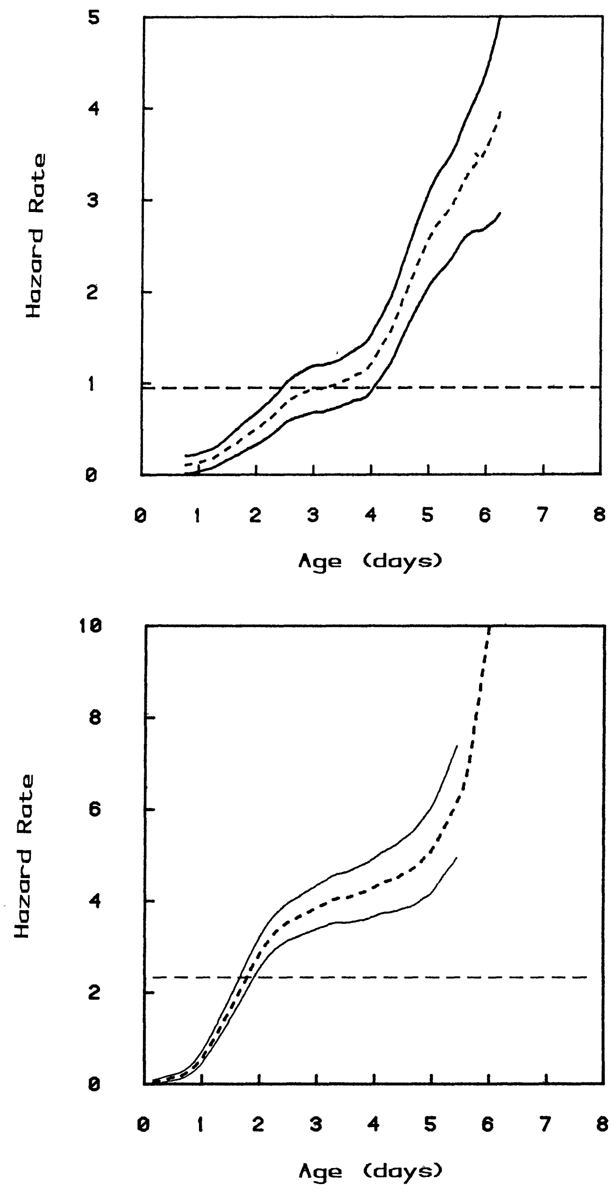

FIG. 8. *Kernel rates for mouse survival*: (a) *control group*, $n = 1080$; (b) *treated group*, $n = 1454$. *Short-dash = kernel rate with solid line* 80% *symmetric confidence bands. Long-dash = constant rate estimate,* 0.000945 *for control,* 0.00233 *for treated mice.*

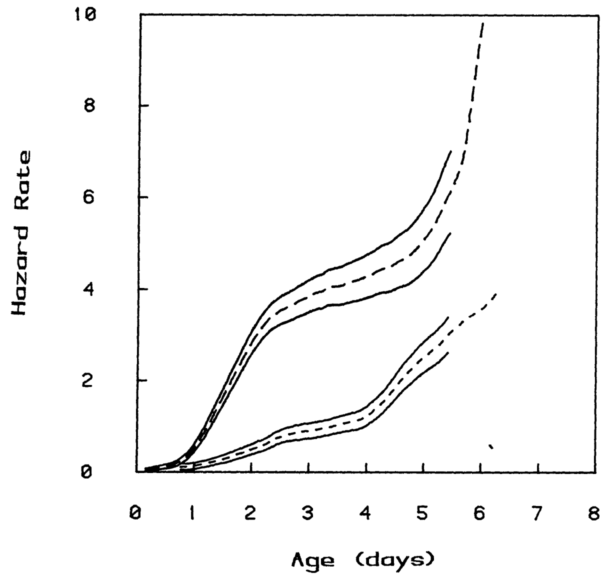

FIG. 9. *Kernel rate graphical test of equal rates. Short-dash = control; long-dash = treated. Solid line = simultaneous confidence band of Section* 6, *remark* (3).

Kernel estimates of control and treated group survival rates are plotted in Figure 8(a–b). The control death rate is definitely increasing with age, as compared with the estimated constant rate of $.945 \times 10^{-3}$. One could perhaps fit a Gompertz-Makeham rate ($h(t) = a + b^t$) or a Weibull rate ($h(t) = \mu a t^{a-1}$), but the general shape of the curve is fairly well determined, without parametric assumptions. The kernel bandwidth for the control group is 153.6 days. The treated group death rate also increase, clearly rejecting the hypothesis of constant rate (estimated at 2.33×10^{-3}). The treated kernel bandwidth is 181.2 days. One sees that it would be difficult to fit any of the standard parametric models because of the flat section of the treated death rate between 250 and 400 days. One might postulate an early death rate increase due to radiation, leveling off in middle age, and picking up again as the mice get old and susceptible to a variety of pathologies. These fine points might be missed if one only studied the survival curve.

Treated and control group rate estimates are shown together in Figure 9, computed for common bandwidth 169.4 days and $T_n = 544$, with a simultaneous confidence band weighted proportional to $n_i^{-1/2}$ about each curve. Clearly they are different. One sees that the net additive effect of gamma irradiation to the death rate increases except during the middle age range. Thus the hazard rates are probably not proportional between treated and control groups.

8. Conclusion. We have presented a class of kernel estimators of the rate functions of a survival process in the presence of censoring. These generalize estimators proposed by Watson and Leadbetter (1964a, b) and are seen to be asymptotically unbiased, strongly consistent, and asymptotically normal (Ramlau-Hansen, 1983). Through a series of strong approximations, the asymptotic distribution of the maximal deviation of an estimate from its true value was derived, leading to simultaneous confidence bands and graphical tests. Theory and simulations indicate the desirability of second- and perhaps third-order expansions of the limiting distribution, due to the slow convergence rate of these maximal deviations. Data from a survival experiment with serial sacrifice was briefly analysed, indicating a large treatment effect without making unrealistic assumptions about the form of the survival distributions. The simulations and data analysis demonstrate that for moderate sample sizes (say 1000 or more) one can do rather well with inference about rates with right censoring. A program called "kernel" to compute estimators and simultaneous confidence bands is available as part of the ISP system distributed by the Statistics Department, University of California, Berkeley.

Acknowledgements. Kjell Doksum offered considerable help in formulating these ideas. Peter Guttorp assisted at several stages of this work. The reviewers offered important suggestions for reorganization.

REFERENCES

AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression.* Wiley, New York.

BARTOSZYŃSKI, R., BROWN, B. W., MCBRIDE, C., and THOMPSON, J. R. (1981). Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process. *Ann. Statist.* **9** 1050–1060.

BEAN, S. J., TSOKOS, C. P. (1980). Developments in nonparametric density estimation. *Internat. Statist. Rev.* **48** 267–287.

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095. Correction, *Ann. Statist.* **3** 1370.

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

BURKE, M. D., and HORVÁTH, L. (1982). Density and failure rate estimation in a competing risks model. Unpublished manuscript, University of Calgary, Canada.

CLEVENSON, M. L., and ZIDEK, J. V. (1977). Bayes linear estimators of the intensity function of the nonstationary Poisson process. *J. Amer. Statist. Assoc.* **72** 112–120.

COLLOMB, G. (1978). Conditions necessaires et suffisantes de convergence uniforme d'un estimateur de la regression, estimation des derivees de la regression. *Comptes Rendus Acad. Sci. Paris* **288A** 161–164. (In French).

DAVID, H. A. and MOESCHBERGER, M. L. (1978). *The Theory of Competing Risks.* Griffin's Statistical Monographs and Courses No. 39. Griffin, London.

FÖLDES, A., and REJTÖ, L. (1981). Strong uniform consistency for nonparametric survival curve estimators from randomly censored data. *Ann. Statist.* **9** 122–129.

FÖLDES, A., REJTÖ, L., and WINTER, B. B. (1981). Strong consistency properties of nonparametric estimators for randomly censored data. Part II: estimation of density and failure rate. *Period. Math. Hungar.* **12** 15–29.

HOEM, J. M. (1976). The statistical theory of demographic rates. *Scand. J. Statist.* **3** 169–185.

KOMLOS, J., MAJOR, P., and TUSNADY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. verw. Gebiete* **32** 111–131.

KONAKOV, V. D. and PITERBARG, V. I. (1979). Large excursions of Gaussian processes and empirical densities. *Theory Probab. Appl.* **24** 660–662.

MCNICHOLS, D. T. and PADGETT, W. G. (1982). Hazard rate estimation under the Koziol-Green model of random censorship. Dept. Math. and Statistics, Univ. South Carolina.

MEIER, P. (1975). Estimation of a distribution function from incomplete observations. In *Perspectives in Probability and Statistics: Papers in Honour of M. S. Bartlett* (J. Gani, ed.). Academic, New York, 67–87.

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

PARZEN, E. (1962). On estimation of a probability density and mode. *Ann. Math. Statist.* **35** 1065–1076.

PRENTICE, R. L., KALBFLEISCH, J. D., PETERSON, A. V., JR., FLOURNOY, N., FAREWELL, V. T., and BRESLOW, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34** 541–554.

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

RICE, J. and ROSENBLATT, M. (1976). Estimation of the log survivor function and hazard function. *Sankhyā, Ser. A* **38** 60–78.

ROSENBLATT, M. (1956). Remarks on some non-parametric estimators of a density function. *Ann. Math. Statist.* **27** 642–669.

ROSENBLATT, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815–1842.

ROSENBLATT, M. (1976). On the maximal deviation of k-dimensional density estimates. *Ann. Probab.* **4** 1009–1015.

SACKS, J. and YLVISAKER, D. (1981). Asymptotically optimum kernels for density estimation at a point. *Ann. Statist.* **9** 334–346.

SETHURAMAN, J. and SINGPURWALLA, N. D. (1981). Large sample estimates and uniform confidence bounds for the failure rate function based on a naive estimator. *Ann. Statist.* **9** 628–632.

SILVERMAN, B. W. (1978a). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6** 177–184.

SUSARLA, V. and WALTER, G. (1981). Estimation of a multivariate density function using delta sequences. *Ann. Statist.* **9** 347–355.

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

UPTON, A. C., ALLEN, R. C., BROWN, R. C., CLAPP, N. K., CONLIN, J. W., COSGROVE, G. E., DARDEN, E. B., JR., KASTENBAUM, M. A., O'DELL, P. T., JR., SERRANO, L. J., TYNDALL, R. L., and WALBURG, H. A., JR. (1969). Quantitative experimental study of low-level radiation carcinogenesis. In *Radiation Induced Cancer*, I.A.E.A., Vienna, 425–438.

WATSON, G. S. and LEADBETTER, M. R. (1964a). Hazard analysis. I. *Biometrika* **51** 175–184.

WATSON, G. S. and LEADBETTER, M. R. (1964b). Hazard analysis. II. *Sankhyā* **26A** 110–116.

WERTZ, W. (1978). *Statistical Density Estimation: A Survey.* Vandenhoeck and Ruprecht, Gottingen.

WERTZ, W. and SCHNEIDER, B. (1979). Statistical density estimation: a bibliography. *Internat. Statist. Rev.* **47** 155–175.

WIGGINS, A. D. (1981). Personal communication. University of California, Davis.

YANDELL, B. S. (1981). Non-parametric inference for rates and densities with censored serial data. Ph.D. dissertation, Biostatistics Program, University of California, Berkeley, Dec., 1981.

STATISTICS DEPARTMENT
UNIVERSITY OF WISCONSIN
1210 W. DAYTON ST.
MADISON, WISCONSIN 53706