Non-Parametric Inference for Rates and Densities
with Censored Serial Data

By

Brian Stuart Yandell

B.S. (California Institute of Technology) 1974
M.A. (University of California) 1978

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biostatistics

in the

GRADUATE DIVISION

OF THE

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved: Kjell A. Doksum 11-4-81
Chairman .... Date
.... 11/13/81
.... Nov. 9, 1981

....

Non-Parametric Inference for Rates and Densities with Censored Serial Data

by Brian S. Yandell

Biostatistics Program, University of California, Berkeley

Abstract

This thesis concerns non-parametric inference for density and rate functions with censored serial data. The focus is upon "delta sequence" curve estimators of the form $a_n(x) = \int K_m(x,y)dA_n(y)$ with K_m integrating to 1 and concentrating mass near x as m->oo. Typically, A_n is either the Kaplan-Meier product-limit estimator of the cumulative distribution or the Nelson-Aalen empirical cumulative rate. Bias, covariance, expected mean square error convergence, and uniform consistency are presented. Asymptotic normality and simultaneous confidence bands are derived for Rosenblatt-Parzen estimators, with $K_m(x,y) = mw(m(x-y))$, $m=o(n)$, and $w(.)$ a well-behaved density. This work generalizes global deviation and mean square deviation results of Bickel and Rosenblatt. and others to censored serial data. Simulations with exponential survival and censoring indicate the effect of censoring on bias, variance, and maximal absolute deviation. Results extend to a multiple decrement/competing risks model. Death rates and sacrifice frequencies are analysed with data from a survival experiment with serial sacrifice.

Kjell A Doksum

Kjell A. Doksum

Thesis Advisor

To Jerzy Neyman

Mr. Neyman helped me through some hard times as I was formulating the ideas that finally turned into this thesis. He introduced me in the spring of 1979 to the important paper of Kaplan and Meier on survival curves for censored data, and to the serial sacrifice experiment. He helped me write my first paper, causing me endless grief with every rewrite. Each time I visited him, I learned something new. Usually we discussed nonidentifiability and illness-death models, but sometimes he talked of weather modification, carcinogenesis, a Latin phrase, a Belgian poet, or a political cartoon in a recent New Yorker. Though we had our differences--he was a stubborn individual--I remember his intense desire to understand the chance mechanism of a biological or physical process, and to teach young people like me to do the same. Yes, Mr. Neyman, "life is complicated, but not uninteresting". Thank you for enriching my life.

Brian S. Yandell
13 November 1981

Acknowledgements

Kjell Doksum kept me on the track of my thesis, encouraged me in many ways, and helped to improve my soccer playing as well. Chin Long Chiang introduced me to stochastic processes and life tables and sparked my interest in applying these to biological problems. Elizabeth L. Scott served as my Graduate Advisor, allowing me the freedom to make mistakes, and guiding me through many challenges of my graduate career. Bland Ewing first showed me the potentials and pitfalls of mathematics as a tool for biological problems, while I worked with him and David L. Wood. Bland also suggested I enter the Biostatistics Program so long ago. Sheldon Ross kindly agreed to serve on my thesis committee late in the game.

At one time I thought I would write about plant ecology, when Herbert Baker shared many marvels of that field. In a way my thesis topic grew out of problems of monitoring herbivory, plant-eating by animals, while visiting Suzanne Koptur in Costa Rica. That work brought me back to Berkeley and Jerzy Neyman.

I can't hope to remember everyone here. Tom Permutt, Ross Ihaka, Steve Hubert, Mark Abrahams, Carl Smith and Jim Reeds helped create a pleasant and personalized computer system. Lucien LeCam in his quiet way provided helpful direction. Peter Guttorp offered his notes and sound advice. Kris Burdzy helped me translate some Russian. Thanks to Linda Lipner, Ed Guerrant, Ralph Price, Seong Ju Kim, Rob Colwell, Lisa Wagner, Sara Henning, Jan Rader, Myoungshic Jhun, Ken Lindahl, Amy Edwards, Michael Gursky, Dave Eberth, Carol Langhauser, Ellen Bradley, Bonnie Hutchings, Susan Justice, Janet Fesl, Monica Jackson, Shryl

Dawkins, Jeanne Lovasich, and my parents for being there.

Partial support for this thesis came from National Institute of Environmental Health Sciences Grant 2-R01-ES01299, Environmental Protection Agency Grant R805379-01-6, School of Public Health Federal Traniee-ship 748, and University of California Regents Fellowships in 1975-76 and 1977-78. I wish to thank the granting agencies and the principal investigators for their assistance.

Table of Contents

Non-parametric Inference for Rates and Densities with Censored Serial Data

1. Introduction

This paper concerns non-parametric inference for density and rate functions with censored serial data. In survival experiments, the form of the rate function is of primary importance to understanding the underlying process. 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 death 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 in addition want to estimate and graph the rate function without making assumptions about the unknown survival process, and to visually compare rate functions from different samples. Efforts at non-parametric inference for rates naturally lead to similar questions for densities. Here a broad class of smooth estimators is investigated, with more refined results on simulataneous confidence bands for the special class of kernel estimators introduced by Rosenblatt (1956) and Parzen (1962).

This problem arose in an effort to draw inference about pathology-related death rates and prevalence of pathologies (diseases) in a survival experiment with serial sacrifice. This experiment was designed to investigate the effect of a treatment, such as radiation or chemical exposure, on animals in terms of the time course of pathological states (Upton *et al.* 1969). Jerzy Neyman introduced me to this problem, pointing out aspects of non-identifiability and statistical inference that plague such experiments (Neyman 1980, 1982). We will return to this later to examine the overall death rates and sacrifice frequencies in treated and control groups for a particular experiment.

Inference about rates has been the subject of study in survival analysis, demography, reliability, and other fields for a long time. The first investigation of a death rate in the presence of competing risks focused upon the merits of smallpox vaccination. Daniel Bernoulli's mathematical treatment in 1760 inspired critical review by D'Alembert, Laplace and others (see Chiang 1978; David and Moeschberger 1978). Many methods of inference about rates in competing risks or multiple decrement models have been developed since. For recent reviews see David (1974), Hoem (1976), Prentice et al. (1978), and David and Moeschberger (1978). Additional references appear in Wertz and Schneider (1979).

Rate estimation is closely tied to density estimation through the relation $f(x)=h(x)(1-F(x))$. Work on density estimation seems to have begun with efforts of Karl Pearson around the turn of the century (see Wegman 1972). Recent reviews of density estimation include Bean and Tsokos (1980, Wertz and Schneider (1979), Tapia and Thompson (1978), Wertz (1978) and Fryer (1977), Wegman (1972) and Rosenblatt (1971). Density and rate estimates are special cases of non-parametric regression estimates (see Collomb 1981 for review).

This paper focuses upon non-parametric rate and density estimation in the presence of competing risks, namely random right-censoring. Non-parametric techniques are characterized by choosing an estimate from a broad class which cannot be easily characterized by a finite-dimensional parameter. For example, the non-parametric maximum likelihood estimate of the death rate among distributions with continuous non-negative rates is a right-continuous step-function with step values being the ratio of the number of deaths in an interval to the length of life lived in that interval (cf. Barlow _et al._ 1972). This estimator tends to be very rough. Grenander (1956) focused attention upon the sub-class of all non-decreasing rates, obtaining a smooth estimator which was later generalized to IFRA and star-shaped distributions; this approach is intimately related to the total-time-on-test statistic (cf. Barlow _et al._ 1972). Other efforts have constricted the class of rates in a variety of ways, such as introducing penalty functions (e.g. Scott, Tapia and Thompson 1980), assuming proportional hazard rates over the observation span (Lehmann 1953; Cox 1959) or separately for each age-specific interval (Chiang 1961; 1968); or incorporating concomitant information about an individual (Cox 1972; Lagakos 1981).

Another approach to rate estimation, the one considered here, draws upon ideas of the kernel density estimator introduced by Rosenblatt (1956) and studied by Parzen (1962) and many others since. Watson and Leadbetter (1964ab) proposed three rate estimators, based on the Rosenblatt-Parzen kernel without censoring.

The Rosenblatt-Parzen kernel method and many other methods are special cases of the "delta sequence" approach recently studied in a general form (Walter and Blum 1979; Susarla and Walter 1981; Lo 1980ab). The estimators considered in this paper are of the form

$$a_n(x) = \int K_m(x,y)\,dA_n(y)$$

in which A_n estimates a cumulative function and the sequence $\{K_m\}$, $m=m(n)$, converges to the Dirac delta function. The Rosenblatt-Parzen kernel estimators satisfy the relations $K_m(x,y) = mw(m(x-y))$, $m=1/b$, $b\to 0$ and $nb\to\infty$, in which b is the "bandwidth" and w(.) is some nicely behaved density. The density estimate, denoted by f_n, arises when $1-A_n=S_n$, the product-limit survival curve (Kaplan and Meier 1958). The three rate estimators are denoted by $h_n^{(i)}$, i=1,2,3. $h_n^{(1)}$ is the ratio f_n/S_n; the other two arise when A_n is either the empirical cumulative rate (Nelson 1972; Aalen 1976, 1978) or $-\log(S_n)$. Details of notation appear in section 2.

The problem of non-parametric estimation and inference for rates and densities when the data are censored has received little attention. Foldes, Rejto and Winter (1981) proved strong consistency for f_n and $h_n^{(1)}$ with histogram type and Rosenblatt-Parzen type kernels. Guttorp (1978) proved pointwise consistency and asymptotic normality for $h_n^{(2)}$ with a Rosenblatt-Parzen kernel in a general random censorship model. Lo (1980b) considered nonparametric inference for the rate function of a multivariate counting process studied by Aalen (1978), a generalization of $h_n^{(2)}$, in a Bayesian context. McNichols and Padget (1981) studied the mean, variance and limiting behavior of f_n with a Rosenblatt-Parzen kernel and with the censoring rate proportional to the death rate.

The Rosenblatt-Parzen kernel estimators have been the subject of some criticism, leading some authors to consider bandwidths depending on location (Abramson 1981; Sacks and Ylvisaker 1981; Breiman, Meisel and Purcell 1977). However, it is possible to arrive at some global measures of deviation for Rosenblatt-Parzen estimators which lead to simultaneous confidence bands and graphical tests.

Bickel and Rosenblatt (1973; 1975) introduced a global measure of deviation leading to simultaneous confidence bands for f_n in the non-censored case. Let $M_n = ||(nb/V_f(x))^{1/2}(f_n(x)-f(x))||$, with $(nb)^{-1}V_f$ the variance process of a_n, and $||.||$ the sup over $[0,T]$. They showed for suitable r_n and d_n, and under appropriate conditions, that for all x,

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

Rosenblatt (1976) extended this result to multivariate densities, while Johnston (1981) extended it to bivariate density estimators with normalized weights (Nadaraya 1964; Watson 1964). Simultaneous confidence bands for kernel estimtors of death rates in the absence of censoring were derived by Rice and Rosenblatt (1976), Guttorp (1978), and Sethuraman and Singpurwalla (1981). Major (1973) and Revesz (1977) obtained maximal deviation results for the non-parametric regression estimators with histogram type and Rosenblatt-Parzen type kernels, respectively. Bounds on the rate of convergence of the distribution of maximal deviation to the limiting distribution were given by Konakov and Piterbarg (1979) for the univariate kernel density estimate. Under stronger conditions on the kernel window and bandwidth, Konakov and Piterbarg (1979) obtained a second order correction term for finite sample size.

In section 2, notation and the estimates are introduced. Section 3 details the assumptions used in later sections. In section 4 we derive the mean and covariance structure of the estimates; section 5 concerns strong consistency. The deviation of an estimate from its expectation, suitably normalized by the variance, is strongly approximated by a Gaussian process in section 6, allowing extension of maximal deviation results to obtain simultaneous confidence bands in section 7. Section 8 turns these bands into graphical tests of goodness-of-fit when parameters must be estimated, and tests of equality of rates in two populations. Section 9 addresses some questions of kernel choice. Monte Carlo simulations in section 10 indicate the effect of censoring and bandwidth, and the slow convergence to the limiting distribution for maximal deviations. Section 11 analyses data from a survival experiment with serial sacrifice, testing for constant death rate and the effect of radiation treatment. This section also tests whether or not the censoring rate is constant. Section 12 presents conclusion and remarks about possible extensions to multiple decrement/competing risks models, fixed censorship, and other issues.

Notation

Let $(X_1,D_1),\ldots,(X_n,D_n)$ 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 i-th individual. For convenience let $N(x) = \#\{X_i \leq x \mid D_i=1\}$, i.e. the number of deaths in $[0,x]$, and let $R(x) = \#\{X_i \geq x\}$, the number at risk of death or censoring at time $x \geq 0$. The process is observed over a finite interval $[0,T_n]$, with T_n tending in probability to some $T<\infty$.

The death rate, denoted by $h(x)$, $x \geq 0$, has been variously called the intensity, force of mortality, and instantaneous failure rate. One assumes for infinitessimal dx,

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

Let H denote the cumulative death rate, that is $H(x) = \int_0^x h(t)dt$. The survival curve is denoted by $S(x)=\exp(-H(x))$, with cumulative distribution function $F=1-S$ and survival density $f=F'=hS$. This paper concerns inference about h and f in the presence of censoring.

Censoring may in practice be fixed or random. This paper concentrates on random right-censoring which is independent of the survival process. Thus, letting C(.) denote the censoring curve and L(.) denote the "life" curve, $L(x)=\Pr\{X_i>x\}$, one finds that $L(x)=S(x)C(x)$. If the censoring curve C is sufficiently smooth, as will be assumed later, one can define the censoring rate $g(x)$ such that

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

with cumulative censoring rate $G=-\log(C)$. In addition, the sub-distribution function $\tilde{F}(x)=\Pr\{X_i \leq x \mid D_i=1\}$ and the sub-density $\tilde{f}=\tilde{F}'$ are used in this paper.

All functions and stochastic processes are left-continuous, unless otherwise noted. For a stochastic process Y, let dY denote the jump process and the differential for integration. Integrals are left-continuous, with the limits of integration understood to be (-oo, oo) unless otherwise noted. A subscript Y_n denotes sample estimate. $Y_n = o_p(n)$ means $\Pr\{Y_n/n \to 0 \text{ as } n\to\infty\} = 1$, and $Y_n = O_p(n)$ means $\Pr\{\lim\sup |Y_n/n| \le c\} = 1$ for some $c>0$. $||Y||$ denotes the $\sup\{|Y(t)|;\ 0<t<T\}$, unless otherwise noted.

The density estimator considered here is the following:

$$f_n(x) = \int_0^T K_m(x,y)dF_n(y) = n^{-1} \sum_{i=1}^n K_m(x,X_i)D_i/C_n(X_i); \quad 0\le x\le T.$$

Here, $S_n=1-F_n$ and C_n are product-limit (PL) estimates (Kaplan and Meier 1958) of the survival and censoring distribution functions, respectively. Namely, the PL estimate of S is

$$S_n(x) = \prod (1-1/R_i)^{D_i} \qquad \text{if } x \le X_n,$$

$$= 0 \qquad \text{if } x > X_n.$$

in which $R_i=R(X_i)$, the rank of X_i in decreasing order, and the product is over $\{i|X_i\le x\}$. C_n is defined similarly, with D_i replaced by $(1-D_i)$.

The three kernel rate estimators considered here are

$$h_n^{(1)}(x) = f_n(x)/S_n(x)$$

$$h_n^{(2)}(x) = \int_0^T K_m(x,y)dH_n(y) = \sum_{i=1}^n K_m(x,X_i)D_i/R_i$$

$$h_n^{(3)}(x) = \int_0^T K_m(x,y)d(-\log S_n^*(y)) = \sum_{i=1}^n K_m(x,X_i)D_i\log(1+1/R_i)$$

in which S_n^* is S_n with R_i replaced by R_i+1, and $H_n(x) = \sum D_i/R_i$, with the sum over $\{i|X_{i\le x}\}$, is the empirical cumulative rate introduced by Nelson

(1972; Aalen 1976, 1978). The form of $h_n^{(1)}$ and $h_n^{(3)}$ generalize those of Rice and Rosenblatt (1976). Watson and Leadbetter (1964ab), and Foldes, Rejto and Winter (1981) among others, investigated

$$h_n^{(4)}(x) = f_n(x) / \int_x^{\infty} f_n(t)dt.$$

Watson and Leadbetter (1964a) defined $h_n^{(3)}$ in an hueristic fashion, as the "graphical derivative" of $-\log(S_n)$ without explicit representation. Sethuraman and Singpurwalla (1981) studied $h_n^{(3)}$ and a kernel estimate based on the maximum likelihood rate estimator mentioned in the introduction. Only f_n and $h_n^{(i)}$, i=1,2,3, are studied in the present paper.

3. Assumptions

This section details the assumptions used in this paper. They are referred to as K0-K6 and S1-S8 in the text. Kernel assumption K0 was formulated by Walter and coworkers (Walter and Blum 1979; Susarla and Walter 1981) to assure consistency of a broad class of estimates. Kernel assumptions K1-K2 are needed for consistency of the Rosenblatt-Parzen estimates (Rosenblatt 1956; Parzen 1962). K3 is used in conjunction with the law of the iterated logarithm in approximation results (Bickel and Rosenblatt 1973). K4 insures that the bias and MSE are manageable (Rosenblatt 1971), while K5 implies uniform consistency (Collomb 1978; Silverman 1978a). K6 is introduced in this paper to allow approximation of the censoring distribution.

K0. (a) $\{K_m(x,y); 0 \le x,y \le T\}$ is a "delta sequence". That is, for each bounded function a with support in (0,T) and for all x in (0,T),

$$\int_0^T K_m(x,y)a(y)dy \to a(x) \text{ as } m \to \infty.$$

Typically m=m(n,x) or simply m=m(n).

(b) $\{K_m\}$ is of "positive type" if $K_m \ge 0$ and for all x in (0,T),

(i) $\int_0^T K_m(x,y)dy = 1.$

(ii) $\sup_{r>0} \int_{|x-y|>r} K_m(x,y)dy = O(m^{-1}).$

(iii) $||K_m(x,.)|| = O(m).$

(iv) for all r>0, $\sup\{K_m(x,y) ; |x-y|>r\} \to 0$ as $m \to \infty$.

K1. The weight function w has integral 1, and either (a) is absolutely continuous with derivative w' on [-A,A] and vanishes off [-A,A], or (b) is absolutely continuous with derivative w' on the real line such that, for j=1,2,

$\int |w'(x)|^j dx < \infty$.

K2. The bandwidth $b=b(n)$ tends to 0 and $nb \to \infty$ as $n \to \infty$.

K3. The following integral over $\{|z| \geq 3\}$ is bounded:

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

K4. w is symmetric (about 0) and $z^2 w(z)$ is integrable.

K5. $nb/\log n \to \infty$ as $n \to \infty$.

K6. $b \log n \to 0$.

K7. $nb^5 \log b \to 0$.

The censored survival assumptions are a combination of those for product-limit estimators and those for kernel estimation. S1-S3 are needed for consistency of the PL estimates S_n and C_n. S4 insures strong uniform consistency (Foldes and Rejto 1981) of the PL estimates. S5 is needed for existence and convergence of the kernel estimate (Parzen 1962; Rosenblatt 1956). S6 is used in approximations. S7 makes the bias tend to 0 at an acceptable rate. S8 is used to extend the Bickel and Rosenblatt (1973) approximations.

S1. The survival distribution S is continuous and positive over $[0,T]$.

S2. The censoring distribution C is continuous and positive over $[0,T]$.

S3. Survival and censoring act independetly. Thus the life distribution is $L=SC$.

S4. L is bounded away from 0 on $[0,T]$, in which T is finite. There is a $\delta>0$ such that $1-\delta \geq L(T) \geq \delta$.

S5. The density f is continuous, positive, and bounded on $[0,T]$.

S6. The function $f^{1/2}$ is absolutely continuous and has derivative of bounded absolute value on $[0,T]$.

S7. The second derivative f'' of f exists and is bounded on $[0,T]$.

S8. The censoring cumulative rate $G=-\log C$ is absolutely continuous and has bounded derivative g over $[0,T]$.

4. Bias and Variance

In this section, exact expressions and approximations for bias and variance are derived.

Let A be a bounded, positive continuous non-decreasing function on (0,T) with continuous bounded derivative a. Let A_n be an estimate of A of the form $\int Y dN$ in which N is a counting process for the number of "deaths" (uncensored events) and Y is a predictable process with respect to the history of deaths and censoring. In particular we are interested in the cumulative distribution F and the cumulative rate H, with respective estimators, $0 \leq x \leq T$,

$$F_n(x) = n^{-1} \int_0^x C_n^{-1}(y) dN(y),$$

$$H_n(x) = n^{-1} \int_0^x L_n^{-1}(y) dN(y),$$

in which C_n and L_n are product limit estimators of the censoring and life (death + censoring) distributions, respectively. Estimates considered in this paper are of the form

$$a_n = \int_0^T K_m(x,y) dA_n(y), \ 0<x<T,$$

in which $\{K_m\}$ is a delta sequence satisfying assumption K0. See Walter and Blum (1979) for examples of delta sequences. Of particular interest is the positive type Parzen (1962) kernel with $b=m^{-1}$, $K_m(x,y) = mw(m(x-y))$, satisfying assumptions K1(a) and K2.

The stochastic integrals in this paper coincide with the Lebesgue-Stieltjes integrals for almost every realization. The following result on integration by parts is needed.

Theorem A. (cf. Liptser and Shiryayev, 1978, ch. 18) If X and Y are left continuous functions bounded variation, then the Lebesgue-Stieltjes integration by parts is

$$X(x)Y(x) = X(0)Y(0) + \int_0^x X_+ dY + \int_0^x YdX.$$

where $X_+=X+dX$, the right-continuous version of X.

In order to specialize results to density and rate estimators, one needs to know the covariance structure of H_n and F_n. These were determined, for random right-censoring, by Breslow and Crowley (1974), among others:

Theorem B. For $0<x\le y<T$, if S1-S4 hold then

$$nCov(H_n(x),H_n(y)) \to V(x),$$
$$nCov(F_n(x),F_n(y)) \to S(x)S(y)V(x),$$

in which $V(x) = \int_0^x L^{-1}(y)dH(y)$.

The following lemma is usefull here and for later approximations.

Lemma 4.1. If S1-S4 hold, then

$$||Y-Y_n|| = o_p((\log n/n)^{1/2}),$$
$$||Y^{-1}-Y_n^{-1}|| = o_p((\log n/n)^{1/2}),$$

in which Y is replaced by either S, C, or L.

Proof: The result for $||S-S_n||$ is Theorem 3.2 of Foldes and Rejto (1981). The result for C follows by interchanging the roles of death and censoring. L follows as the special case of no censoring (though a stronger rate is possible, see Foldes and Rejto (1981)). Note that if $||S-S_n||<\delta/2$ then $S_n>\delta/2$ by assumption S4. Thus for sufficiently large n, the rate for $||S^{-1}-S_n^{-1}||$ is obtained. A similar argument yields the result for C and L.

The main result of this section concerns bias and covariance.

Theorem 4.1. Let $K_m(x,y)$ be a delta sequence of positive type. Then

1) $\text{Bias}(a_n(x)) = B_a(x) = \int_0^T K_m(x,y)dA(y) - a(x) + \int_0^T K_m(x,y)dB_A(y)$

in which $B_A = EA_n - A = $ bias of A_n.

2) $$n\text{Cov}(a_n(x),a_n(t)) = \int_0^T K_m(x,y)K_m(t,y)[V_A(dy,y) - V_A(y,dy)]$$
$$+ \int_0^T \int_0^z [K_m(x,y)K_m(t,z) + K_m(x,z)K_m(t,y)]V_A(dy,dz)$$

in which $V_A(y,z) = n\text{Cov}(A_n(y),A_n(z))$.

Proof. For fixed n, m, x and t let $K(y) = K_m(x,y)$ and $J(z) = K_m(t,z)$. The mean of a_n is determined using integration by parts:

$$Ea_n(x) = K(T)E\{A_n(T)\} - \int_0^T \{EA_n\}dK$$
$$= K(T)A(T) - \int_0^T AdK + K(T)B_A(T) - \int_0^T B_A dK$$
$$= \int_0^T KdA + \int_0^T KdB_A$$

The asymptotic covariance is found as follows:

$$n\text{Cov}(a_n(x),a_n(t)) = K(T)J(T)V_A(T,T) - \int_0^T V_A(y,T)[K(T)dJ(y)+J(T)dK(y)]$$
$$+ \int_0^T \int_0^z V_A(y,z)[dK(y)dJ(z) + dJ(y)dK(z)]$$

Integrate by parts once on the second term and twice on the third, with a change of order of integration. Recombine terms and note a last integration by parts which uses the relation

$$K(T)V_A(T,T) - \int_0^T V_A(y,y)d[K(y)] = \int_0^T K(y)[V_A(dy,y) + V_A(y,dy)],$$

to arrive at the covariance.

For density and rate estimators we obtain

Theorem 4.2. Let $\{K_m\}$ be a positive-type delta sequence. Let f_n and $h_n^{(i)}$, i=1,2,3, be defined as in the introduction and let S1-S5 hold.

Then

1) Expectations: For x in (0,T),

$$Ef_n(x) = \int_0^T K_m(x,y)[1-(1-L(y))^n]dF$$

$$Eh_n^{(1)}(x) = S^{-1}(x)Ef_n(x) + O((\log n / n)^{1/2})$$

$$Eh_n^{(2)}(x) = \int_0^T K_m(x,y)[1-(1-L(y))^n]dH$$

$$Eh_n^{(3)}(x) = Eh_n^{(2)}(x) + O(1/n)$$

Suppose K_m is a Parzen (1962) kernel, $m=b^{-1}$, satisfying K1(a), K2 and K4, and suppose S7 holds. Then

$$Ef_n(x) = f(x) + 0.5f''(x)b^2 \int w(y)y^2dy + o(b^2) + O((1-\delta)^n).$$

$$Eh_n^{(2)}(x) = h(x) + 0.5h''(x)b^2 \int w(y)y^2dy + o(b^2) + O((1-\delta)^n).$$

in which $B = \int(1-L)^n dF$ is the bias of S_n.

2) Covariances: For x,t in (0,T),

$$nCov(f_n(x),f_n(t)) = \int_0^T K_m(x,y)K_m(t,y)C^{-1}(y)dF(y)$$

$$+ \int_0^T \int_0^z [K_m(x,y)K_m(t,z)+K_m(x,z)K_m(t,y)](V(y)-L^{-1}(y))dF(y)dF(z)$$

$$\to 0 \qquad \text{if } x \neq t$$

$$\to \int_0^T K_m^2(x,y)C^{-1}(y)dF(y) \qquad \text{if } x=t$$

$$nCov(h_n^{(2)}(x),h_n^{(2)}(t)) = \int_0^T K_m(x,y)K_m(t,y)dV(y)$$

$$\to 0 \qquad \text{if } x \neq t$$

Suppose K_m is a Parzen (1962) kernel, $m=b^{-1}$, satisfying K1(a) and K2. Then

$$nbVarf_n(x) \to V_f(x) = f(x)C^{-1}(x) \int w^2,$$

$$nbVarh_n^{(i)}(x) \to V_h(x) = h(x)L^{-1}(x) \int w^2, \qquad i=1,2,3,$$

uniformly in x on (0,T).

<u>Proof:</u>

$$EH_n(x) = \int_0^x \Pr\{R(y)>0\}dH(y) = H(x) - \int_0^x (1-L(y))^n dH(y).$$

$$EF_n(x) = \int_0^x S(y)\Pr\{R(y)>0\}dH(y) = F(x) - \int_0^x (1-L(y))^n dF(y).$$

(cf. Watson and Leadbetter 1964b; Aalen 1976). These together with Theorem 4.1 yield Ef_n and $Eh_n^{(2)}$. f_n is uniformly bounded by K1(a), K2, K5, and S5 (Collomb 1978; Silverman 1978a). Applying Lemma 4.1 for the survival curve S, $h_n^{(1)}=f_n/S + O_p((\log n/n)^{1/2})$. Since f_n and $1/S_n$ are uniformly bounded, the order of the approximation of $h_n^{(1)}$ extends to $Eh_n^{(1)}$. One sees that

$$h_n^{(3)}=h_n^{(2)}+\sum_{i=1}^n K_m(x,X_i)D_i(\log(1+1/R_i) - 1/R_i).$$

Following Rice and Rosenblatt (1976), one notices that $|x-\log(1+x)| \leq x^2/2$, $0\leq x\leq 1$. Letting $K(y)=w((x-y)/b)/b$, for fixed x and n,

$$E|h_n^{(2)}(x)-h_n^{(3)}(x)| \leq \int_0^T \sum_{i=1}^n (1-L(y))^{i-1}L(y)^{n-i} \frac{1}{n-i+1} K(y)dF(y).$$

In an analogous fashion to the argument following (2.17) in Rice and Rosenblatt (1976), with L(.) replacing 1-F(.), one finds

$$Eh_n^{(3)}(x) = Eh_n^{(2)}(x) + O(1/n)$$

Parzen kernel approximations for bias in the non-censored case were derived by Rosenblatt (1956; 1971), Watson and Leadbetter (1964ab) and Rice and Rosenblatt (1976). The only difference here is the addition of a last term due to the bias of F_n or H_n, which are both bounded a.s. by a term decreasing exponentially with n.

2) The covariances follow from Theorems 1 and B. The Rosenblatt-Parzen kernel convergence follows for f_n and $h_n^{(2)}$ from Parzen (1962, Theorem 1A) and the fact that the partial derivatives of S(x)S(y)V(x) and V(x), $0<x\leq y<T$, are uniformly bounded by S1-S5. For $h_n^{(1)}$, use the approximation to f_n/S and the relation $V_h = S^{-2}V_f$. The result for $h_n^{(3)}$ follows

from its uniform closeness to $h_n^{(2)}$.

Remark. As noted by Rice and Rosenblatt (1976), $f''/S = h''-3hh'+h^3$ and $h'' = S^{-1}(f''+3hf'+2h^2f)$. Thus the bias of $h_n^{(1)}(x)$ is smaller than those of $h_n^{(2)}(x)$ and $h_n^{(3)}(x)$ if f has a local minimum at x, and larger if h has a local minimum at x.

5. Consistency

In this section consistency is explored in two senses, expected mean square error and maximum absolute deviation. These are extensions of similar results in Susarla and Walter (1981).

Theorem 5.1. Let $\{K_m\}$ be a delta sequence of positive type with $m = o(n)$. Suppose the first and second order derivatives of V_A are uniformly bounded and that B_A tends to 0 uniformly in x. Then for each x in $(0,T)$, $E[(a_n(x) - a(x))^2] \to 0$ a.s.

Proof:

$$nVar(a_n(x)) \leq ||K_m(x,.)|| \int_0^T K_m(x,y)|V_A(dy,y) - V_A(y,dy)| + o(1)$$

$$= O(m) + o(1) = o(n).$$

The definition of delta sequence insures that $\int K_m dA \to a$. Since $B_A \to 0$ uniformly in x by assumption, the theorem obtains.

Theorem 5.2. Let $\{K_m\}$ be a delta sequence of positive type such that $\int_0^T |K_m(x,dy)| = cO(m)$, with the constant c possibly depending on x. Then $|a_n(x) - a(x)| < ||A_n - A||O(m)$ a.s.

If c does no depend on x, then the bound is uniform a.s.

Proof. Similar to Theorem 4.4 of Susarla and Walter (1981).

Foldes, Rejto and Winter (1981) proved strong uniform consistency for f_n and $h_n^{(4)}$ with Rosenblatt-Parzen kernels under the stronger condition that the bandwidth $b=o((\log n/n)^{1/8})$. Guttorp (1978) proved strong uniform consistency of $h_n^{(2)}$ with a Rosenblatt-Parzen kernel and certain side conditions, for the non-censored case. Here we show strong uniform consistency for delta sequence estimators of density and rate with censored serial data.

Corollary 5.1. If S1-S4 and K0 hold, and if the constant c of Theorem 5.2 is independent of x, then

(i) $||f_n - f|| = O_p(m (\log n)^{1/2}/n)$ a.s.

(ii) $||h_n^{(1)} - h|| = O_p((\log n/n)^{1/2} + m (\log n)^{1/2}/n)$ a.s.

(iii) $||h_n^{(2)} - h|| = O_p(m (\log n)^{1/2}/n)$ a.s.

(iv) $||h_n^{(3)} - h|| = O_p(n^{-1} + m (\log n)^{1/2}/n)$ a.s.

Proof. (i) follows from Foldes and Rejto (1981, Theorem 3.2) and Theorem 5.2. (ii) and (iv) follow from (i) and (iii), respectively, and the approximation shown in the proof of Theorem 4.2. (iii) follows from the next result, which was suggested by Peter M. Guttorp and does not appear to be in the literature (Sethuraman and Singpurwalla (1981) prove a similar result for $-\log(S_n)$ in the case of no censoring).

Theorem 5.3 (Guttorp and Yandell). Let S1-S4 hold. Then

$$||H_n - H|| = O_p((\log n)^{1/2}/n).$$

Proof: We can write

$$H_n(x) = H_n^*(x) + \int_0^x (L_n^{-1} - L^{-1}) dF_n.$$

in which $H_n^*(x) = \int_0^x L^{-1} dF_n$. By Lemma 4.1 (for L), the second term is of proper order. Foldes and Rejto (1981, Lemma 3.1) show for all u in $[0,T]$, $v>0$,

$$\Pr\{|H_n^*(u) - H(u)| > v\} < 2\exp(-2nv^2\delta^2).$$

The result follows by analogy to the proof of Theorems 3.1 and 3.2 of Foldes and Rejto (1981), replacing the survival curve and its estimate (denoted by them as F_n^* and F, respectively) by H_n^* and H.

6. Strong Approximation

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

$$W_n(x) = (nb/V_f(x))^{1/2}\,(f_n(x)-Ef_n(x)),\qquad 0\le x\le T,$$

can be approximated by a Gaussian process

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

in which Z is a version of Brownian motion. Similar approximations for the rate estimators $h_n^{(i)}$, $i=1,2,3$, are indicated at the end of the section. The series of approximations parallels the work of Bickel and Rosenblatt (1973; 1975). Throughout this section, for fixed x and n let $K(y)=w(t/b)/b$ with assumptions K1(a) and K2 satisfied.

The empirical sub-distribution function $\tilde{F}_n$ agrees on the range $[0,T+A)$ with the empirical distribution function for the random variable

$$\tilde{X} = X \qquad \text{if } D=1,$$
$$= X+T+2A \qquad \text{if } D=0,$$

in which 2A is the window width of assumption K1(a). The function $\tilde{f}_n = \int Kd\tilde{F}_n$ is a non-censored kernel estimator for the density of $\tilde{X}$ on $[0,T]$. One may therefore employ existing results for the non-censored case; this is done without further comment below. Let

$$Y_n(x) = (b/\tilde{f}(x))^{1/2}\int_0^T KdZ_n^o(\tilde{F}).$$

Define $_0Y_n$ and $_1Y_n$ by replacing Z_n^o by Z^o and Z, respectively. Here Z^o is a version of the Brownian bridge and Z is a version of Brownian motion. Let $_2Y_n = (b/\tilde{f}(x))^{1/2}\int K\tilde{f}^{1/2}dZ$, and let $_3Y_n = b^{1/2}\int KdZ$. The following theorem and lemma are central to the next proposition. The lemma is a restatement of Propositions 2.1 and 2.2 of Bickel and Rosenblatt (1973), using the result of Komlos, Major and Tusnady (1975). See Rosenblatt (1976).

Theorem C. (Komlos, Major and Tusnady 1975; Revesz 1976) Let K1(a) and S1-S3 hold. A sequence of Brownian bridges can be constructed so that

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

This refined earlier convergence rate results by Brillinger (1969) and Breiman (1968) in which they independently obtained the rate $O_p((\log n)^{1/2}(\log\log n/n)^{1/4})$.

Lemma 6.1. Suppose the Y processes are as defined above. Let S1-S3 hold.

(i) If K1(a) and S5 hold, then $||Y_n - {}_0Y_n|| = O_p((nb)^{-1/2} \log n)$.

(ii) If S4-S5 hold, then $||{}_0Y_n - {}_1Y_n|| = O_p(b^{1/2})$.

(iii) If S4-S6 and K3 hold, then $||{}_2Y_n - {}_3Y_n|| = O_p(b^{1/2})$.

If S1-S3 hold, then $f=\tilde{f}/C$. In addition $dF_n = C_n^{-1} d\tilde{F}_n$. Thus the density estimator of interest, f_n, may be written as

$$f_n(x) = \int_0^T K dF_n = \int_0^T K C_n^{-1} d\tilde{F}_n.$$

The strategy below involves replacing C_n^{-1} by C^{-1} and proceeding by a series of approximations using Lemma 6.1. Let

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

Proposition 6.1: Let K1(a), K2, K5 and S1-S5 hold. Then

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

Proof: Clearly, S4 implies

$$|f_n(x) - f_n^*(x)| \le \int_0^T K|C_n^{-1} - C^{-1}| d\tilde{F}_n$$

$$\le \tilde{f}_n(x)\ ||C^{-1} - C_n^{-1}||.$$

By Lemma 4.1, $||(C^{-1}-C_n^{-1}|| = O_p((\log n/n)^{1/2})$. K1(a), K2, K5, and S5 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 C_n replaced by C, may be written as

$$W_n^*(x) = (b/V_f(x))^{1/2} \int_0^T K(y)C^{-1}(y)dZ_n^o(\tilde{F}(y)).$$

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

$$ {}_2W_n(x) = (b/V_f(x))^{1/2} \int_0^T K\tilde{f}^{1/2}C^{-1}dZ$$
$$ = (b/V_f(x))^{1/2} \int_0^T K(f/C)^{1/2}dZ$$
$$ {}_3W_n(x) = (b/\int w^2)^{1/2} \int_0^T KdZ$$

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

<u>Proposition 6.2</u>: Suppose the W processes are as above, and that K1(a) and S1-S5 hold. Then

(i) If S8 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 K3, S6 and S8 hold, then $||{}_2W_n - {}_3W_n|| = O_p(b^{1/2})$.

<u>Proof</u>: (i) By S3-S4, $C \geq \delta$. Thus

$$|W_n^*(x)-{}_0W_n(x)| <= \delta^{-1}|Y_n(x)-{}_0Y_n(x)|(\int w^2)^{-1/2}$$
$$+ \delta^{-1}||Z_n^o-Z^o|| \ (bV_f(x))^{-1/2} \int_0^T |g(y)w((x-y)/b)|dy$$

The first term is of proper order by Lemma 6.1, part (i), and K1(a). S8 insures that g exists and is bounded. V_f is continuous and positive since C and f are, by S2 and S5. Together with Theorem C, this gives the order for part (i).

(ii) Note that $|{}_0W_n(x) - {}_1W_n(x)| = (C(x)/\int w^2)^{1/2} |{}_0Y_n(x) - {}_1Y_n(x)|$. Thus (ii) follows from Lemma 6.1, part (ii), and $C \leq 1$.

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

$$|{}_2W_n(x) - {}_3W_n(x)| \leq \delta^{-1} |{}_2Y_n(x) - {}_3Y_n(x)| (\int w^2)^{-1/2}$$
$$+ (b/4V_f(x))^{1/2} \int_0^T |Z(yb+x)| g(yb+x)(f(yb+x)/C(yb+x))^{1/2} |w(y)| dy.$$

S3-S4 and Lemma 6.1 make the first term of the proper order. S1-S5 and S8 bound uniformly the term $g(f/C)^{1/2}$ in the integral. K3 and the law of the iterated logarithm for Brownian motion insure that the second term is $O_p(b^{1/2})$.

${}_1W_n$ and ${}_2W_n$ are Gaussian processes with the same covariance structure and hence have the same law. Hence by Propositions 6.1 and 6.2,

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

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, as indicated below.

The strong approximation result for $h_n^{(2)}$ follows in an analogous fashion. First one approximates $h_n^{(2)}$ by

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

Proceed by replacing C by L and V_f by V_h in subsequent formulae and propositions. Similar results obtain for $h_n^{(1)}$ and $h_n^{(3)}$ but the rate for proposition 1 for $h_n^{(3)}$ must be replaced by $O_p((b/n)^{1/2})$ due to the approximation indicated in the estimation section. Better rates may well be attainable. To summarize,

<u>Proposition 6.3</u>. Let K1(a), K2, K5, S1-S6 and S8 hold. Let

$$W_a(x) = (nb/V_a(x))^{1/2} (a_n(x) - Ea_n(x)), \quad 0<x<T,$$

Then for $a_n=f_n$, $a_n=h_n^{(1)}$, and $a_n=h_n^{(2)}$,

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

For $a_n=h_n^{(3)}$,

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

If in addition K4 and S7 hold, $Ea_n(.)$ may be replaced by $a(.)$ in W_a. Let M_n be a sequence of functionals satisfying Lipschitz condition such that, for $J_n > 0$,

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

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

7. 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 non-censored case. Therefore we obtain the same limiting double-exponential probability as did Bickel and Rosenblatt (1973). This is inverted to derive simultaneous confidence bands. We present extensions of work due to Konakov and Piterbarg (1979) which allow refinement of the asymptotic approximations, adding a second and third order term to the limiting probability.

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

Theorem D. Let $r_n = (2\log(T_n/b))^{1/2}$, and $d_n = r_n + (\log w^*)/r_n$, in which

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

$$= [\int w'^2/\int w^2]^{1/2}/2\pi \quad \text{if } w(A) = 0.$$

Let $M_n = ||{}_3W_n||$, in which the sup is of x over $[0,T_n]$. Let K1-K2 hold. Then, for $-\infty<x<\infty$,

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

The next result for censored densities and rates follows directly:

Theorem 7.1. Suppose K1(a), K2-K6, and S1-S8 hold. Let h_n represent any one of $h_n^{(1)}$, $h_n^{(2)}$, $h_n^{(3)}$. Let

$$M_f = ||(nb/V_f)^{1/2}(f_n - f)||,$$

$$M_h = ||(nb/V_h)^{1/2}(h_n - h)||,$$

in which the sup is over $[0,T_n]$. Let r_n and d_n be defined as in Theorem

D. Then, for $-\infty<x<\infty$,

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

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

From these probability statements one may derive bands in at least two ways using Slutsky's Theorem. If C is known, one can use it directly. For the following let C be estimated by C_n. For the density, if one replaces f by f_n in V_f, then the band is of the symmetric form

$$f_n \pm k(f_n/C_n)^{1/2}$$

in which $k = (d_n + x/r_n)(\int w^2/nb)^{1/2}$ and $x = \log(-2/\log(1-\alpha))$. Inverting the probability statement without substituting in V_f yields the asymmetric band

$$f_n + \frac{k^2}{2C_n} \pm k(\frac{f_n}{C_n})^{1/2}[1 + \frac{k^2}{4f_nC_n}]^{1/2}.$$

The latter band is wider for all x, reflecting the more conservative substitution. Empirical evidence from simulations presented below indicates that this may be the appropriate band. Writing this band as

$$(f_n + \frac{k^2}{2C_n}) \pm [(f_n + \frac{k^2}{2C_n})^2 - f_n^2]^{1/2}$$

shows that this band remains above 0 unless $f_n(x)=0$. At this point, the band becomes $[0,k^2/C_n(x)]$. Note that the symmetric band may have a negative lower bound, which in practice is usually truncated to 0.

Similarly for the rate one may replace L by L_n and may or may not replace h by $h_n=h_n^{(i)}$, i=1,2,3, obtaining the two band forms

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

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

Konakov and Piterbarg (1979, Theorems 3 and 4) improved upon

Theorem D. They obtained a second order term which depends on sample size through the standardizing term r_n. In order to use these results one needs additional assumptions on the smoothness of the kernel window w(.), and the bandwidth must be of the form $b=kn^{-p}$, $1/3<p<1/2$. The strong approximation argument presented in Section 6 extends their result to the censored estimators considered here. However, as stated by them, the normalizing terms are different. We state their theorem and then modify it as a corollary.

<u>Theorem E</u>. (Konakov and Piterbarg 1979) Let K1(a), K2-K6 and S1-S8 hold, with no censoring. Let $b=n^{-p}$, $1/3<p<1/2$. Let r_n and w^* be as in Theorem D. Suppose $w(A)=w(-A)=0$, and $\int(w'')^2<\infty$. Let M_f be defined as in Theorem 7.1, but with the sup over [0,1]. Then there are constants $0<u<\infty$, $v>0$ such that, for any x in $(-\infty, \infty)$,

$$\Pr\{l_n(M_f-l_n)<x\} = \exp(-2\exp(-x-x^2/2l_n^2)) + L_1(n,x),$$

with $l_n = (r_n^2 + 2\log w^*)^{1/2}$. If $x>l_n(1-l_n)$ and $n > (6/\pi w^*)^p$, then

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

<u>Theorem 7.2</u>. Let the assumptions of Theorem 7.1 hold. Let $b=kn^{-p}$, $1/3<p<1/2$, and let $w(A)=w(-A)=0$ and $\int(w'')^2<\infty$. Then for $-\infty<y<\infty$,

$$\Pr\{r_n(M_a-d_n)<y\} = \exp(-2\exp(-y-(y+\log w^*)^2/2r_n^2)) + L_2(n,y).$$

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

$$L_2(n,y) = O(L_1(n,y)),$$

in which L_1 is the same function as in Theorem E.

<u>Proof</u>: The proof rests on the substitution, for fixed n,

$$d_n + y/r_n = l_n + x/l_n.$$

The following need to be justified:

(i) $e^{-2x}x^2 = O(e^{-2y}y^2)$.

(ii) $x + x^2/2l_n^2 = y + (y + \log w^*)^2/2r_n^2$.

To prove (i) note that $y=xr_n/l_n+r_n(l_n-d_n)$. As $n\to\infty$, $r_n\to\infty$ and $r_n/l_n\to 1$. Also,

$$r_n(l_n-d_n) = (\log w^*)^2(l_n+d_n)/r_n \to 2(\log w^*)^2.$$

Thus $y-x \to 2(\log w^*)^2$ as $n \to \infty$, with $\log w^* < \infty$, implying (i). The second equality follows from the expansion

$$\begin{aligned} x + x^2/2l_n^2 &= 0.5(l_n + x/l_n)^2 - 0.5l_n^2 \\ &= 0.5(d_n + y/r_n)^2 - 0.5r_n^2 - \log w^* \\ &= y + (y + \log w^*)^2/2r_n^2 \end{aligned}$$

Unfortunately the paper of Konakov and Piterbarg (1979) contains several typographical errors and no proofs, being the summary of a symposium talk. The necessary tools for proof are indicated as references. The second-order approximation may be obtained directly from Berman (1971) where it is shown that if X is a stationary Gaussian process with 0 mean and covariance $r(t)=1-t^2/2+o(t)$, then

$$\Pr\{||X|| \leq u\} \to \exp(-2\sigma_1),$$

in which $u=(2\log(T/2\pi\sigma_1))^2$ and the sup is over $[0,T]$. Berman (1971) suggested setting $\sigma_1=\exp(-x)$ and approximating u by the first two terms of a binomial expansion, a course apparently followed by Bickel and Rosenblatt (1973, Theorem A.1). Konakov and Piterbarg (1979) appear to set $\sigma_1=\exp(-x-x^2/2l_T^2)$, in which $l_T=(2\log(T/2\pi))^{1/2}$. Then $u=l_T + x/l_T$. A similar argument leads to the expansion in Theorem 7.2.

A further refinement is possible, due to the Edgeworth expansion of the maximal deviation in the neighborhood of 0 by Piterbarg (1978, Theorems 3 and 4). This result supercedes Lemma 2.2 of Berman (1971) and leads to a third order expansion of the maximal deviation distribution, if one can obtain an 8-degree polynomial approximation of r(t), by modifying the proof to Theorem 7.1 of Berman (1971). The expansion of the limiting probability becomes

$$\exp(-\sigma_1[1-q(\sigma_1/T)^p u^{-5}])$$

in which q and p are positive constants depending on polynomial coefficients of the approximation to r(.), and u and σ_1 are as above. The point is that the sign is negative, indicating that this correction places the probability between the first- and second- order expansions. See Appendix for details. This is explored empirically in the simulation section.

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, see Sacks and Ylvisaker (1981) and Abramson (1981) regarding optimal kernels for a point.

(2) One can approximate the variance $V_a(.)$ by

$$V_{a;n}(x) = b^{-1}\int_0^T w^2((x-y)/b)dV_{A;n}(y),$$

in which $V_{A;n}$ is an estimate of V_A. This may more accurately depict the variance for moderate sample sizes, particularly in regard to the edge effect of the kernel estimate. This is compared with the estimate of the asymptotic variance process in the simulation section, and is used for data analysis. Though the asymmetric band can be formulated with this variance estimate, it is not clear how to interpret it.

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

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

for Theorem D and Theorem 7.1, 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 7.2.

8. Testing

This section concerns goodness-of-fit tests for the composite hypothesis H: $f=f_o(.,\Theta)$ or H: $h=h_o(.,\Theta)$, in which Θ is an unknown real-valued parameter, and two-sample tests of the form H: $f_1=f_2$ and H: $h_1=h_2$. Bickel and Rosenblatt (1973) derived goodness-of-fit tests based on maximal absolute deviation and mean square deviation. Sections 3 and 4 of their paper apply in the current situation for testing a composite hypothesis about a density or a rate with censored serial data.

One may substitute an estimate $\hat{\Theta}$ for Θ in the maximal statistic M_a (a=h or f) provided the following assumptions are met for every Θ:

(a) $\left\| \int_0^T w_n(x-y)dA_o(.,\hat{\Theta}) - \int_0^T w_n(x-y)dA_o(.,\Theta)\right\| = o_p(n\ (\log b)^{-1/2})$.

(b) $\|a_o(.,\hat{\Theta}) - a_o(.,\Theta)\| = o_p((\log b)^{-1})$.

These conditions are satisfied if $|\hat{\Theta} - \Theta| = O_p(n^{-1/2})$ and the partial derivatives of the density (rate) are bounded in a neighborhood of Θ. One can then graphically test the hypothesis by drawing the estimate f_n with confidence bands and seeing if $f_o(.,\hat{\Theta})$ lies within the bands.

As pointed out by Bickel and Rosenblatt (1973), the maximal deviation test is asymptotically inadmissable relative to the mean square deviation test. For the latter one needs the second partials bounded in a neighborhood of Θ, and $|\hat{\Theta} - \Theta| = o_p((nb)^{-1/2})$. The mean square statistic does not readily translate into a graphical tool. They show that this test may be better than the classical chi-square test, at least asymptotically. See their paper for details.

Bickel and Rosenblatt (1973) did not treat the two-sample testing problem. One may use their maximal deviation results, with the approximations of the current paper, to derive the asymptotic distribution of the maximal deviation of rate (or density) estimates from two populations. This suggests a convenient graphical test, which is employed in the data analysis section.

Theorem 8.1. Let K1(a), K2-K6 hold, and let S1-S8 hold for survival processes with censoring in two independent populations. Let a_i, i=1,2, represent the death rate functions (a=h) or survival densities (a=f) for the two populations. Let V_a be the variance process for a(.), the common (unknown) function under the null hypotheses H: $a_1=a_2$. For a samples of size n_1 and n_2, $n=n_1+n_2$, let $a_{i;n}$ be the Rosenblatt-Parzen kernel estimate of a_i, i=1,2, using a common bandwidth b and kernel window w(.). Let

$$M = ||(n_1 n_2 b/(nV_a))^{1/2}(a_{1;n} - a_{2;n})||$$

in which the sup is over $[0,T_n]$, with $T_n < T_{n_i}$, i=1,2, If $n_1/n \to \lambda$, for some $0 < \lambda < 1$, then under the null hypothesis,

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

in which r_n and d_n are defined as in Theorem D, functions of b, w and T_n.

Proof: Let $Y_i(x) = (n_i b/V_a(x))^{1/2} (a_{i;n}(x)-a_i(x))$, i=1,2. Y_1 and Y_2 can be strongly approximated by independent Gaussian processes, say W_1 and W_2, which have the same distribution as ${}_3W_n$, defined in the strong approximation section. Let $W = (1-\lambda)^{1/2}W_1 - \lambda^{1/2}W_2$. W has the same distribution as ${}_3W_n$, as can be seen by computing its covariance. Hence one can apply Theorem D to W, and to the process $Y = (1-\lambda)^{1/2}Y_1 - \lambda^{1/2}Y_2$, by way of the strong approximation results. Rewriting M and substituting n_1/n for λ shows the correspondence.

Corollary 8.1. The expansion in Theorem 7.2. is valid for M under the same conditions as in Theorem 7.2.

Remarks. (1) One might not wish to assume equal variance. One can instead consider the maximal deviation of the process

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

in which V_i is the variance process for a_i, i=1,2.

(2) A possible choice choice for b is $b=(n_1 b_1 + n_2 b_2)/n$. 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 may be greater than k_1 and k_2.

(3) Following a suggestion of Al Wiggins (pers. comm.), one can plot the two curves $a_{1;n}$ and $a_{2;n}$ and place a simultaneous confidence band about them. Ideally the band should appear as an adjustment of the individual confidence bands, with the relative widths remaining 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 $a_{1;n}-a_{2;n}$ and place a simultaneous confidence band about this curve.

(4) If one is interested primarily in testing it may be adviseable to concentrate upon the cumulative distribution or cumulative rate function. See Nair (1981) for recent developments with maximal and mean square deviations.

9. Choice of Kernel

The optimal choice of kernel window and bandwidth presents interesting challenges. See Rosenblatt (1956, 1971), and references in Rudemo (1981) for a theoretical treatment. Rudemo (1981), Davis (1981), Silverman (1978b), and Tapia and Thompson (1978) suggested empirical methods. Schuster and Gregory (1981) showed that the cross-validation empirical approach to bandwidth selection is not consistent for Rosenblatt-Parzen kernel estimates of long-tailed distributions.

We use the criterion of minimizing the mean square error (MSE) or mean integrated square error (MISE) suggested by Epanechnikov (1969; Rosenblatt 1971). From the estimation section, the MSE for density estimation is

$$E[f_n(x)-f(x)]^2 = V_f(x)/nb + 0.25\,[f''(x)]^2 b^4 [\int w(y)y^2 du]^2 + o((nb)^{-1}+b^2) + O((1-\delta)^n).$$

This is minimized by choosing $b=k_f(x)n^{-0.2}$, with

$$[k_f(x)]^5 = V_f(x)\ (f''(x))^{-2}\ [\int w(y)y^2 dy]^{-2}.$$

The MISE is minimized by choosing $b=k_f n^{-0.2}$, with

$$[k_f]^5 = [\int_0^T V_f(y)dy]\ [\int (f''(y))^2 dy]^{-1}\ [\int w(y)y^2 dy]^{-2}.$$

Similar formulae obtain for rate estimation, with f replaced by h at every occurrence. In order to evaluate the constant $k_a(x)$ or k_a, a=f,h, one needs to know the form of the kernel window w(.), the derivative a", and the variance process V_a. We take w(.) to be the "optimal" quadratic window, $w(x) = 1.5-6x^2$, $-0.5\leq x\leq 0.5$, with $w(x) = 0$ outside this interval. Thus $\int w^2(y)dy = 1.2$, $\int w(y)y^2 dy = 0.05$. For comparison purposes below, the uniform kernel, $w(x)=1$, $-0.5\leq x\leq 0.5$, is also used. For the remainder of this discussion, suppose $L(T)=\delta$ and $S(T) = \gamma \geq \delta > 0$.

Censoring enters the formulae for k_a through the variance. If one makes the proportionality assumption (cf. Chiang 1968) $g(x)=g_o h(x)$ for some $g_o \geq 0$, then

$$\int_0^T V_f = (\int w^2)(g_o-1)^{-1}\,(\gamma^{(g_o-1)}-1) \quad \text{if } g_o \neq 1,$$

$$= (\int w^2)(-\log(\gamma)) \quad \text{if } g_o=1.$$

$$\int_0^T V_h = (\int w^2)(1+g_o)^{-1}\,(\gamma^{-(g_o+1)}-1)$$

If $g_o=0$, then both integrals coincide with $(\int w^2)(\gamma^{-1}-1)$. Thus $\int V_f$ increases as g_o goes from 0 to $1+\gamma$, and then decreases; $\int V_h$ decreases monotonically as g_o increases. The effect of non-proportional censoring will depend on the form of $g(.)$ and $h(.)$ and is not investigated further here.

Classically, people have investigated the normal case, in which

$$f''(x) = \sigma^{-5}(x^2-\sigma^2)(2\pi)^{-1/2}\exp(-x^2/2\sigma^2), \quad 0 \leq x \leq T,$$

yielding pointwise constant $k_f(0)=3.247\sigma$ in the case of a uniform window (compare with Bickel and Doksum 1977, p. 385). For the MISE, one needs

$$\int (f'')^2 = 0.375\pi^{-1/2}\sigma^{-5},$$

yielding $k_f=3.686\sigma$ for the uniform kernel and $k_f=4.483\sigma$ for the optimal kernel with no censoring. σ^2 is usually estimated from the data as the sample variance of the mean survival time. Simulations and data analysis presented in later sections use the optimal window and the MISE bandwidth computed from normal theory.

The normal model does not lend itself readily to survival analysis. Alternatively, we investigate a model with exponential survival and censoring distributions. In other words, for this discussion, h and g are time-homogeneous. Thus $S(x) = \exp(-\lambda x)$ and $C(x) = \exp(-g_o\lambda x)$. For the density, this yields $\int(f'')^2 = 0.5h^5(1-\gamma^2)$. Thus,

$$k_f^5 = 960\lambda^{-5}(1-\delta)^{-1}\ (-0.5\log(\delta)) \qquad \text{if } g_o = 1,$$

$$= 960\lambda^{-5}(1-\gamma^2)^{-1}\ (1-g_o)^{-1}(1-\gamma^2/\delta) \qquad \text{if } g_o \neq 1.$$

which reduces to $800\lambda^{-5}$ in the case of no censoring when $\gamma=0.2$. In this case the bandwidth becomes $b = 3.807\lambda^{-1}n^{-0.2}$, with λ estimated from the data by the exponential case MLE (cf. Aalen 1978) $\lambda_n = N(T)/\int_0^T R(y)dy$. The following table indicates how k changes with amount of censoring:

g_o	0	.1	.2	.5	.8	1.0
$k\lambda$	3.807	3.854	3.902	4.062	4.244	4.379

Allowing $\delta\to0$, $\gamma^2/\delta\to0$, $T\to\infty$, one finds that the optimal bandwidth tends to $4.569\lambda^{-1}n^{-0.2}$, rather close to that found with normal theory.

The choice of bandwidth for rate estimation appears to be an open question. Note that one cannot simply use the earlier approach with constant rate! We decided to simply use the optimal density bandwidth for rate and density estimation. Another approach would be to determine the bandwidth with Parzen's (1979) "weighted spacings." The advantage of this would be that the weighted spacings are constant at 1 under the null hypothesis H: $f=f_o$ for any f_o. One could also optimize with respect to some parametric family of rate functions, such as Weibull.

10. Monte Carlo Simulation

Monte Carlo simulation studies are presented for samples from an exponential survival distribution with exponential censoring. Numbers came from a log transformation of pseudo-random numbers generated by ranm, a function developed and implemented on the MathStat UNIX (a trademark of Bell Laboratories) computer system. All computing was done on their PDP 11/45. The exponential parameters were 1 for survival and several values between 0 and 2 for censoring. The effects of censoring on bias, variance, and the form of the confidence bands are investigated for individual trials of sample size 200. Empirical distributions with 100 Monte Carlo trials of the maximal deviation statistic $r_n(M_a-d_n)$ are investigated at length, with sample sizes ranging from 50 to 500. Kernel bandwidths derived from normal theory ($4.483\sigma n^{-0.2}$) 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. Selected results are transformed to plots of theoretical vs. empirical significance level. Empirical 80% pointwise confidence intervals with 100 Monte Carlo trials and sample size 200 are compared to theoretical pointwise confidence intervals.

Figure 1(a-b) show the bias of the density estimator f_n and of the rate estimator $h_n^{(2)}$ for a random sample of size 200 with censor exponential parameter ranging over 0, .5, 1, and 2. Figure 1(c) shows the rate estimate bias for the 3 estimators at 50% censor (g=1). Note that $h_n^{(2)}$ and $h_n^{(3)}$ are very similar, and $h_n^{(1)}$ has bias properties like f_n, as Theorem 4.2 suggests. The remainder of simulations focus upon f_n and $h_n^{(2)}$.

Figures 2(a-b) show the difference between the theoretical asymptotic variance (TAV), its estimate (EAV), and the estimated sample variance (ESV) for f_n and $h_n^{(2)}$, respectively, with a random sample of size 200. The variances are transformed to log $(V_a/\int w^2)$ in order that the TAV becomes a straight line through 0. The formuale for these variances are

$$\mathrm{TAV} = f(x)C^{-1}(x) \int w^2$$

$$\mathrm{EAV} = f_n(x)C_n^{-1}(x) \int w^2$$

$$\mathrm{ESV} = b^{-1} \int_0^x w^2((x-y)/b)C_n^{-1}(y)dF_n(y)$$

for f_n and

$$\mathrm{TAV} = h(x)L^{-1}(x) \int w^2$$

$$\mathrm{EAV} = h_n^{(2)}(x)L_n^{-1}(x) \int w^2$$

$$\mathrm{ESV} = b^{-1} \int_0^x w^2((x-y)/b)L_n^{-1}(y)dH_n(y)$$

for $h_n^{(2)}$. Note that the ESV is much larger than the TAV near 0, reflecting the edge effect of the kernel estimate. That is, the variance near 0 is much larger than asymptotics suggests. EAV underestimates TAV near 0 for f due to the bias of f_n. The ESV remains higher than the TAV and the EAV over much of the range, regardless of the amount of censoring, as seen in Figure 2(c) for exponential censoring parameter values 0, 0.5, 1, and 2. (The variance in Figure 2(c) is transformed as log(ESV/TAV).) The remainder of the simulations use the ESV to estimate the variance.

Simultaneous confidence bands are shown in Figure 3(a-d) for density and rate with symmetric and asymmetric form. Note the increased dispersion implied with increased censoring, reflecting the earlier results about the variance. The constant rate estimates are drawn for comparison, indicating that the goodness-of-fit test (Section 8) would not reject a constant death rate.

Empirical distribution functions (EDFs) of the maximal deviation statistic were computed from 100 Monte Carlo trials. These were random samples of non-censored data, except in the case of Figure 7 below. The EDF for the rate deviation is compared with the theoretical curve $\exp(-2e^{-x})$ in Figure 4(a) for sample sizes 50, 200, and 500. Note the slow convergence. Figure 4(b) shows these same curves transformed to significance level; the theoretical curve is the diagonal. The lower triangular area corresponds to "conservative" theory, that is theory which yield a larger significance level than reality. Thus one supposes that the confidence bands of Figure 3 are wider than necessary, although the effect of the deviation depends on the relative magnitude of d_n and x/r_n (see Section 8). Now $d_n = r_n + \log w^*/r_n$, with $\log w^* = -.6866$. The following table shows that r_n does not change much over a wide range of n:

n	200	500	1000	10000	10^6
r_n	.2655	.6611	.8452	1.2788	4.158

Figure 4(c) shows the EDF of times of maximal deviation, indicating a slight buildup near 0, but fairly even distribution elsewhere.

Figures 5(a-f) investigate the rate of convergence of f_n and $h_n^{(2)}$ as a function of bandwidth, reducing the bandwidth by a factor of 2 for each new curve with the sample size remaining at 200. Note that the convergence to $\exp(-2e^{-x})$ is fairly rapid compared to the convergence in n. It should be noted that the normal theory bandwidth $b=4.483\sigma n^{-0.2}$ with n=200 is practically as large as the interval $[0,T_n]$, with

$T=L^{-1}(0.2)=1.6$. This is reflected in the EDF of times of maximal deviation, particularly for the density (Figure 5(f)). However, Figure 6(a-b) show what happens to the bias as the bandwidth is reduced--the compromise of greater agreement with the asymptotic distribution for greater bias is probably not a desireable one.

Figures 7(a-b) investigate the effect of censoring on the distribution of maximal deviation. These Monte Carlo simulations were performed with sample sizes of 200 and bandwidths 1/4 normal theory, that is $b=1.121\sigma n^{-0.2}$. Note that the curves for 50% censoring (h=g=1, long dash lines) are slightly above those for the non-censored data. They may be significantly different with a Kolmogorov-Smirnov test; however, a detailed analysis of any such difference awaits more in-depth Monte Carlo studies with several sample sizes.

Figures 8(a-b) show the shape of the Konakov-Piterbarg (1979) second-order expansion of the limiting probability and its relation to the double exponential. The curves depend only on bandwidth and window parameters. A comparison of Figure 8(a) or 8(b) with those of Figure 6 suggests that the expansion may overcorrect, at least for exponentially distributed data. Figures 8(c-d) show the relative difference between the Konakov-Piterbarg theoretical significance level and the empirical significance level for density and rate, respectively. Thus in practice one should have the second-order correction, and maybe even a third-order one to insure some semblance of the correct significance probability for a simultaneous confidence band.

Empirical 80% pointwise confidence bands were derived by dividing the time axis into 32 equal probability intervals, 0.025 to each using upper cutoff point T=-log(0.2). Compare the empirical bands in Figure 9 with the symmetric and asymmetric normal theory pointwise 80% confidence intervals indicated by the dashed lines on the figure. The "liberal" nature of the theoretical bands may reflect two things: (1) the TAV was used, thus underestimating the variance; (2) the empirical bands are not precisely pointwise, as they represent maxima over small intervals. Note the slight asymmetry of the empirical bands, suggesting that the asymmetrical bands might in fact be preferred.

11. Data Analysis

This section addresses certain inference questions regarding the survival of mice exposed to radiation. The data come from a survival experiment with serial sacrifice, designed to investigate the effect of a treatment, in this case 300 rads of gamma irradiation, on animals in terms of the time course of pathological states (Upton *et al.* 1969). Animals in two groups, treated and control, died naturally or were sacrificed. The data consist of time and mode of death, and the presence or absence of a finite number of pathologies. There were 1080 control mice and 1454 treated mice, in which 361 control and 343 treated mice were sacrificed. The pathological states are not used in the present study. 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?

(3) Is the frequency of sacrifices constant over the experiment?

The last question arises because in some experiments the frequency of sacrifices in the future may depend on the history of deaths up to the present. Although this is not indicated in the current experiment, the methodology of choosing mice for sacrifice is not explicitly stated in the documents at our disposal. Some people have suggested that the sacrifice times appear nearly gamma distributed (J. Neyman and Estie Sid Hudes, pers. comm.).

The estimates used are Rosenblatt-Parzen kernel estimates, either

f_n or $h_n^{(2)}$ defined earlier, with the quadratic window and bandwidth computed by normal theory. Simultaneous 80% confidence bands were derived using the estimated sample variance (ESV) and the results of Theorem 7.1. Estimates are plotted as dashed lines, with solid line confidence bands. Constant rate estimates, if present, are long-dashed lines.

Kernel estimates of control and treated group survival rates are plotted in Figure 10(a-b). One sees that the control death rate is definitely increasing with age, as compared with the estimated constant rate of $.945\times10^{-3}$. 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 increases, 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 constant, or possibly decreasing, 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 (see Figure 10(c)), for the changes in slope of the rate function appear as only slight bends in these curves, except for the sudden decrease in treated survival.

Treated and control group rate estimates are shown together in Figure 11, 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 would undoubtedly find a difference with any test, but non-parametric tests such as this have the advantage that the difference detected is tied to much weaker assumptions about the unknown survival curve. One sees that the net additive effect of gamma irradiation to the death rate increases except during the middle age range.

Some question arises as to whether serial sacrifices can be considered as random right-censoring. As the author understands, sacrifices for this experiment were scheduled to occur with a roughly uniform distribution over the entire course of the experiment. The data indicate that sacrifices occurred every few days, with usually 1 and at most 9 sacrifices on any given day. My understanding is that on a sacrifice day a mouse was chosen at random from among those living and sacrificed. However, if one views sacrifice as fixed censoring, see the concluding remarks of this paper.

Figure 12(a-b) investigate the frequency of sacrifices (i.e. censoring densities) of control and treated groups. The frequency of sacrifices in the control group is not significantly different from a constant frequency. However, it does appear to dip down and then increase in the last 150 days. For the treated group, the frequency of sacrifices appears constant for 3-400 days, after which it increases significantly. Thus it appears possible that the frequency of sacrifices increases in the last 150 days or so, perhaps a clean-up effort by the experimenters to examine the long-lived mice before they died of some complication of pathologies. However, only 145 control and 13 treated mice lived beyond 700 days. Similar results obtain by examining the censoring rates.

Many questions of inference remain unanswered with this analysis. Although one cannot identify the transition probabilities between pathological states (Clifford 1977, 1982) or the lethality of pathologies (Neyman 1982; Yandell 1981), it is possible to measure the prevalence of pathologies, P_i, and the pathology-related mortality rates μ_i. However, these do not exactly look like densities or rates, and require a more sophisticated approach than is possible in this paper. The author intends to pursue this problem further in the near future.

12. Conclusion

We have presented a class of "delta sequence", or "kernel", estimators of the density and rate functions of a survival process in the presence of censoring. These generalize estimators proposed by Rosenblatt (1956) and Watson and Leadbetter (1964ab) and are seen to be asymptotically unbiased, strongly consistent, and asymptotically normal. 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 desireability 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. These tests demonstrate that for moderate sample sizes (say 1000 or more) one can do rather well with inference about rates or densities with right censoring. We conclude with a few remarks about extensions and further work.

1. In a multiple decrement or competing risks model, one wishes to draw inference about the rates h_i, $i=1,\dots,I$. A natural estimate would be the generalization of $h_n^{(2)}$,

$$h_{i;n}(x) = \int_0^T K_m(x,y)\,dH_{i;n}(y), \quad i=1,\dots,I$$

in which $H_{i;n}$ is the empirical cumulative rate for event i. Aalen (1976, 1978) showed that $\{H_{i;n}, i=1,\dots,I\}$ are asymptotically independent under assumptions that h_i are continuous and that the events $i=1,\dots,I$ are mutually exclusive. Thus, since $h_{i;n}$ depends only on $H_{i;n}$

and the sure function K_m, one has asymptotic independence of $\{h_{i;n}, i=1,\ldots,I\}$. The properties of $h_n^{(2)}$ extend to these estimators in a natural manner.

2. 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. The main point where this presents a problem is estimating C consistently by C_n in Proposition 6.1 (Foldes and Rejto 1981 assume continuous C). However, if one knows the censor form, then it may be incorporated directly. The assumption S8 that $G=-\log C$ has a continuous derivative may be replaced by

> S8'. 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 6.2.

3. The case of rate inference for a multivariate counting process with arbitrary censoring deserves some attention. Guttorp (1978) began work on this problem, and joint efforts continue with this author to generalize results presented here. Lo (1980b) investigated the problem from a Bayesian standpoint with gamma priors. Comparison of the classical and Bayesian approach seems in order.

4. The stages of disease model presented by Chiang (1979) appears to be a useful and identifiable model for biological applications. Here the risk set for stage i+1 consists of those survivors in stage i. This model begs generalization from proportional transition rates to the case

of arbitrary (continuous) transition rates. The work of Fleming (1978ab) is relevant to this problem. The author plans to investigate this in the context of a "stages of herbivory" model of damage to leaves by insects, joint work with Suzanne Koptur.

13. References

Aalen, O. (1976) Nonparametric inference in connection with multiple decrement models. Scand. J. Statist. 3: 15-27.

Aalen, O. (1978) Nonparametric inference for a family of counting processes. Ann. Statist. 6: 701-26.

Abramson, I.S. (1981) On kernel estimates of probability densities. Ph.D. dissertation, Dept. of Statist., Univ. of Cal., Berkeley, June 1981.

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. John Wiley & Sons, New York: 388 pp.

Bean, S.J., and Tsokos, C.P. (1980) Developments in nonparametric density estimation. Int. Stat. Rev. 48: 267-87.

Berman, S.M. (1971) Asymptotic independence of the number of high and low level crossings of stationary Gaussian processes. Ann. Math. Statist. 42: 927-945.

Bickel, P.J., and Doksum, K.A. (1977) Mathematical Statistics: basic ideas and selected topics. Holden-Day, Inc., San Francisco, 493 pp.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. Ann. Statist. 1: 1071-95.

Bickel, P. J. and Rosenblatt, M. (1975) Corrections to "On some global measures of the deviations of density function estimates." 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: 437-53.

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

Breiman, L., Meisel, W. and Purcell, E. (1977) Variable kernel estimates of multivariate densities. Technometrics 19: 135-144.

Brillinger, D.R. (1969) An asymptotic representation of the sample distribution function. Bull. Amer. Math. Soc. 75: 545-7.

Chiang, C.L. (1961) A stochastic study of the life table and its applications: III. The follow-up study with the consideration of competing risks. Biometrics 17: 57-78.

Chiang, C.L. (1968) Introduction to Stochastic Processes in Biostatistics. Wiley, New York.

Chiang, C.L. (1978) Life Tables and Mortality Analysis. World Health Organization, 399 pp.

Chiang, C.L. (1979) Survival and stages of disease. Math. Biosci. 43: 159-171.

Chiang, C.L. (1980) An Introduction to Stochastic Processes and Their Applications. R.E. Kreiger Publ. Co., Huntington, N.Y., 517 pp.

Clifford, P. (1977) Nonidentifiability in stochastic models of illness and death. Proc. Natl. Acad. Sci. 74: 1338-40.

Clifford, P. (1982) The limits of nonidentifiability in time-dependent compartment models with applications to serial-sacrifice experiments. In Statistical Methods and Cancer: Proceedings of an Interdisciplinary Cancer Study Conference, Berkeley, July 6-8, 1981. 9 pp.

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)

Collomb, G. (1981). Estimation non-parametrique de la regression: revue bibliographique. Intl. Statist. Rev. 49: 75-93. (in French)

Cox, D.R. (1959) The analysis of exponentially distributed life-times with two types of failure. J. Royal Stat. Soc. Ser B 21: 411-21.

Cox, D.R. (1972) Regression models and life tables. J. Royal Stat. Soc. Ser B 34: 187-220.

David, H.A. (1974) Parametric approaches to the theory of competing risks. In Reliability and Biometry: Statistical Analysis of Lifelength, SIAM, Philadelphia: 275-90.

David, H.A. and Moeschberger, M.L. (1978) The Theory of Competing Risks. Griffin's Statistical Monographs & Courses No. 39, Charles Griffin & Company LTD, 103 pp.

Davis, K.B. (1981) Estimation of the scaling parameter for a kernel-type density estimate. J.A.S.A. 76: 632-6.

Epanechnikov, V.A. (1969) Nonparametric estimates of a multivariate probability density. Theor. Prob. Appl. 14: 153-8.

Fleming, T.R. (1978a) Nonparametric estimation for nonhomogeneous Markov processes in the problem of competing risks. _Ann. Statist._ _6_: 1057-70.

Fleming, T.R. (1978b) Asymptotic distribution results in competing risks estimation. _Ann. Statist._ _6_: 1071-9.

Foldes, A., and Rejto, L. (1981) Strong uniform consistency for non-parametric survival curve estimators from randomly censored data. _Ann. Statist._ _9_: 122-9.

Foldes, A., Rejto, 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. Hung._ _12_: 15-29.

Fryer, M.J. (1977) A review of some non-parametric methods of density estimation. _J. Inst. Math. Applic._ _20_: 335-54.

Grenander, U. (1956) On the theory of mortality measurement. _Skand. Akt. Tidskr._ _39_: 70-96 and 125-53.

Guttorp, Peter M. (1978) A rate estimator for a family of multivariate counting processes. Unpublished manuscript, 21 pp.

Hoem, J.M. (1976) The statistical theory of demographic rates. _Scand. J. Statist._ _3_: 169-85.

Johnston, G.J. (1980) Probabilities of maximal deviations for non-parametric regression function estimates. Institute of Statistics Mimeo Series #1271, March 1980, Department of Statistics, Chapel Hill, No. Carolina, 18pp.

Kaplan, E.L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. *J. Amer. Stat. Assoc.* *53*: 457-81.

Komlos, J., Major, P., and Tusnady, G. (1975) An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahr. verw. Gebiete* *32*: 111-31.

Konakov, V.D. and Piterbarg, V.I. (1979) Large excursions of Gaussian processes and empirical densities. *Theo. Prob. Appl.* *24*: 660-2.

Lagakos, S.W. (1981) The graphical evaluation of explanatory variables in proportional hazards regression models. *Biometrika* *68*: 93-8.

Lehmann, E.I. (1953) The power of rank tests. *Ann. Math. Statist.* *24*: 23-43.

Liptser, R.S. and Shiryayev, A.N. (1978) *Statistics of Random Processes II: Applications*. New York: Springer-Verlag.

Lo, A.Y. (1980) On a class of Bayesian nonparametric estimates: I. density estimates. To appear in *Ann. Statist.*

Lo, A.Y. (1980) On a class of Bayesian nonparametric estimates: II. rate function estimates. To appear in *Ann. Statist.*

Major, P. (1973) On non-parametric estimation of the regression function. *Studia Scientiarum Math. Hungarica* *8*: 347-61.

McNichols, D.T. and Padgett, W.G. (1981) A kernel density estimator from arbitrarily right-censored data. Abstract 178-19, *Bull. I.M.S.* *10*(*6*): 261.

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 Press, New York: 67-87.

Nadaraya, E.A. (1964) On estimating regression. *Theo. Prob. Appl.* *9*: 141-2.

Nair, V.N. (1981) Plots and tests for goodness of fit with randomly censored data. *Biometrika* *68*: 99-103.

Nelson, W. (1972) Theory and applications of hazard plotting for censored failure data. *Technometrics* *14*: 945-66.

Neyman, J. (1980) Some memorable incidents in probabilistic/statistical studies. In *Asymptotic Theory of Statistical Tests and Estimation*, I.M. Chakravarti, ed., Academic Press, San Francisco: 1-32.

Neyman, J. (1982) Avenue to understanding the mechanism of radiation effects: extended serial sacrifice experimental methodology. In *Statistical Methods and Cancer: Proceedings of an Interdisciplinary Cancer Study Conference*, Berkeley, July 6-8, 1981. 12 pp.

Parzen, E. (1962) On estimation of a probability density and mode. *Ann. Math. Statist.* *35*: 1065-76.

Parzen, E. (1979) Nonparametric statistical data modeling. *J.A.S.A.* *74*: 105-31.

Piterbarg, V.I. (1978) Asymptotic expansions for the probabilities of large excursions of Gaussian processes. *Soviet Math. Dokl.* *19*: 1279-83.

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: 541-54.

Revesz, P. (1976) On strong approximation of the multidimensional empirical process. Ann. Prob. 4: 729-43.

Revesz, P. (1977) How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. Math. Operationsforsch. Statist. Ser. Statist. 8: 119-26.

Rice, J. and Rosenblatt, M. (1976). Estimation of the log survivor function and hazard function. Sankhya, 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. Prob. 4: 1009-15.

Rudemo, M. (1981) Empirical choice of histograms and kernel estimators. to appear in Scand. J. Statist., 46 pp.

Sacks, J. and Ylvisaker, D. (1981) Asymptotically optimum kernels for density estimation at a point. Ann. Statist. 9: 334-346.

Schuster, E.F. and Gregory, G.G. (1981) On the non-consistency of maximum likelihood nonparametric density estimators. To appear in Computer Science and Statistics: 13th Symposium on the Interface. Springer-Verlag, 4 pp.

Scott, D.W., Tapia, R.A. and Thompson, J.R. (1980) Nonparametric probability density estimation by discrete maximum penalized-likelihood criteria. Ann. Statist. 8: 820-32.

Silverman, B.W. (1978a) Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. Ann. Statist. 6: 177-84.

Silverman, B.W. (1978b) Choosing the window width when estimating a density. Biometrika 65: 1-11.

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(3): 628-32.

Susarla, V. and Walter, G. (1981) Estimation of a multivariate density function using delta sequences. Ann. Statist. 9: 347-55.

Tapia, R.A. and Thompson, J.R. (1978) Non-parametric Probability Density Estimation. Baltimore, London: Johns Hopkins Univ. Press.

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.

Walter, G. and Blum, J. (1979) Probability density estimation using delta sequences. Ann. Statist. 7: 328-40.

Watson, G.S. (1964) Smooth regression analysis. Sankhya 26 A: 359-72.

Watson, G.S. and Leadbetter, M.R. (1964a) Hazard analysis I. Biometrika

51: 175-84.

Watson, G.S. and Leadbetter, M.R. (1964b) Hazard analysis II. Sankhya 26A: 110-6.

Wegman, E.J. (1972) Nonparametric probability density estimation: I. A summary of available methods. Technometrics 14: 533-46.

Wertz, W. (1978) Statistical Density Estimation: a Survey. Vandenhoeck and Ruprecht, Gottingen.

Wertz, W. and Schneider, B. (1979) Statistical density estimation: a bibliography. Int. Statist. Review 47: 155-75.

Yandell, B.S. (1981) Nonidentifiability of lethality in the survival experiment with serial sacrifice. Manuscript in review, 11 pp.

A. Appendix

A theorem of Piterbarg (1978) on the asymptotic expansion of maximal deviations is used in conjunction with the proof of a theorem of Berman (1971) to derive a third-order correction to the distribution of large excursions of stationary Gaussian processes.

Let $X(t)$, $0 \le t \le T$, be a stationary Gaussian process with zero mean and unit variance. Let $\int |X'|^2 < \infty$, $\int |X''|^2 < \infty$. Define the covariance function $r(t) = EX(s)X(s+t)$, with $r(0)=1$. Let $N(.)$ be the standard normal distribution and $n(.)$ the standard normal density.

Theorem F (Piterbarg 1978). Let X be defined as above, with the following conditions satisfied:

(1) The determinant of the covariance matrix of $\{X(t), X'(t), X''(t)\}$ is uniformly bounded away from 0.

(2) For all $t_1 \neq t_2$, the distribution of the six-dimensional vector $\{X(t_i),\ X'(t_i),\ X''(t_i),\ i=1,2\}$ has a density.

(3) There exist $c>0$ such that for all t, $|t|<c$ implies $r(t) \ge \cos(t)$.

Suppose $r(.)$ has the form

$$r(t) = 1 - t^2/2 + Ct^4/4! - Dt^6 + Et^8/8! + o(t^8), \quad t \to 0.$$

Then there exists $T_o > 0$ such that for all $T \le T_o$,

$$P_X(u,T) = N(u) - T(2\ )^{-1/2} n(u) + k(C,D)Tu^{-5} n(u(1-1/C)^{-1/2})(1+o(1))$$

as $u \to \infty$, in which $k(C,D) = [27(C-1)^9/32\pi^3]^{1/2}\ (D-C^2)^{-1} \ge 0$, and

$$P_X(u,T) = \Pr\{\max_{0 \le t \le T} X(t) \le u\}.$$

Note that Theorem F is only valid for $T \ll u$. To extend to large T one needs an argument such as that given in Berman (1971). We adapt his results in the following

Theorem A.1. Let X be as in Theorem F. Let $\sigma_1 = Tn(u)(2\pi)^{-1/2}$ for fixed u. Then

$$P_X(u,T) \to \exp\{-\sigma_1[1-(\sigma_1/T)^{1/(C-1)} u^{-5} k_2(C,D)]\} \text{ as } T \to \infty$$

in which $k_2(C,D)$ is a constant times $k(C,D)$.

Proof: (sketch)

The proof follows along the lines of that of Theorem 7.1 in Berman (1971). One divides $[0,T]$ into disjoint intervals of length t, $0<t<1$. Each of these intervals is subdivided into two intervals of lengths tb and $t(1-b)$, $0<b<1$. It is shown (Berman 1971, equation 7.4) that the maximum over half of these intervals, those of the form

$$I_j = [jt,(j+b)t], \quad j=1,\ldots,[T/t],$$

is close to the desired maximum over $[0,T]$. Berman shows that one may effectively treat the variables

$$M_j = \max\{X(s)\ ;\ jt \le s \le (j+b)t\}, \quad j=1,\ldots,[T/t],$$

as independent. They are identically distributed by stationarity of X, and have distribution given by Theorem F (which supercedes Lemma 2.2 of Berman 1971). Thus,

$$\begin{aligned} P_X(u,T) &= \{P_X(u,tb)\}^{[T/t]} \quad \text{(approximately)} \\ &= \{N(u) - (t/T)b\sigma_1[1-(T/\sigma_1)^{1/C} Tu^{-5} k_2(C,D)(1+o(1))]\}^{[T/t]} \\ &\to \exp\{-b\sigma_1[1-(T/\sigma_1)^{1/C} Tu^{-5} k_2(C,D)]\} \quad \text{as } T \to \infty. \end{aligned}$$

Now as b is arbitrary, let $b \to 1$. One finds that

$$k_2(C,D) = k(C,D)(2\pi)^{1.5-1/C},$$

which can be seen by equating $k(C,D)n(u(1-1/C)^{-1/2})$ and $k_2(C,D)(\sigma_1/T)^{1-1/C}$.

Figure 1(a)

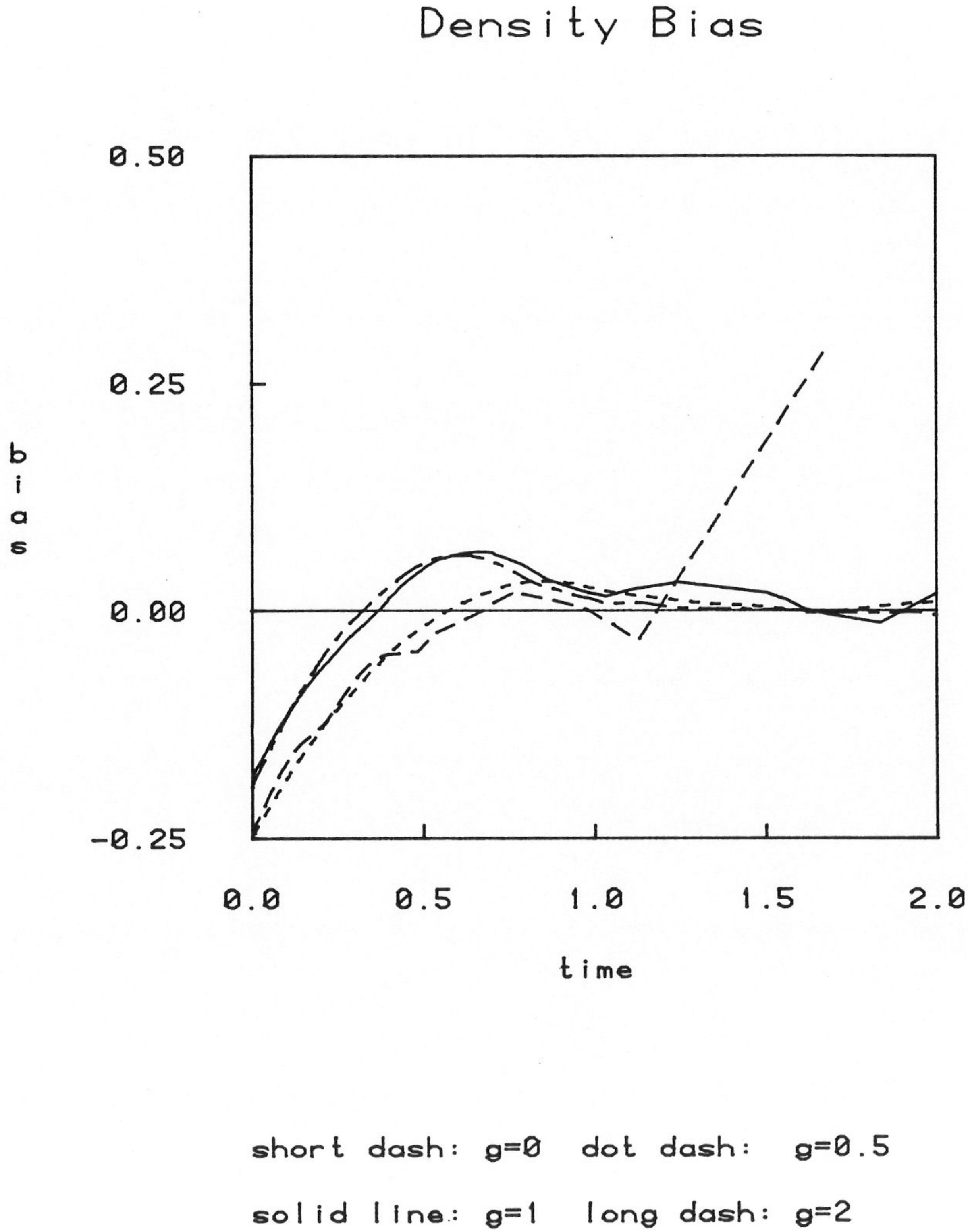
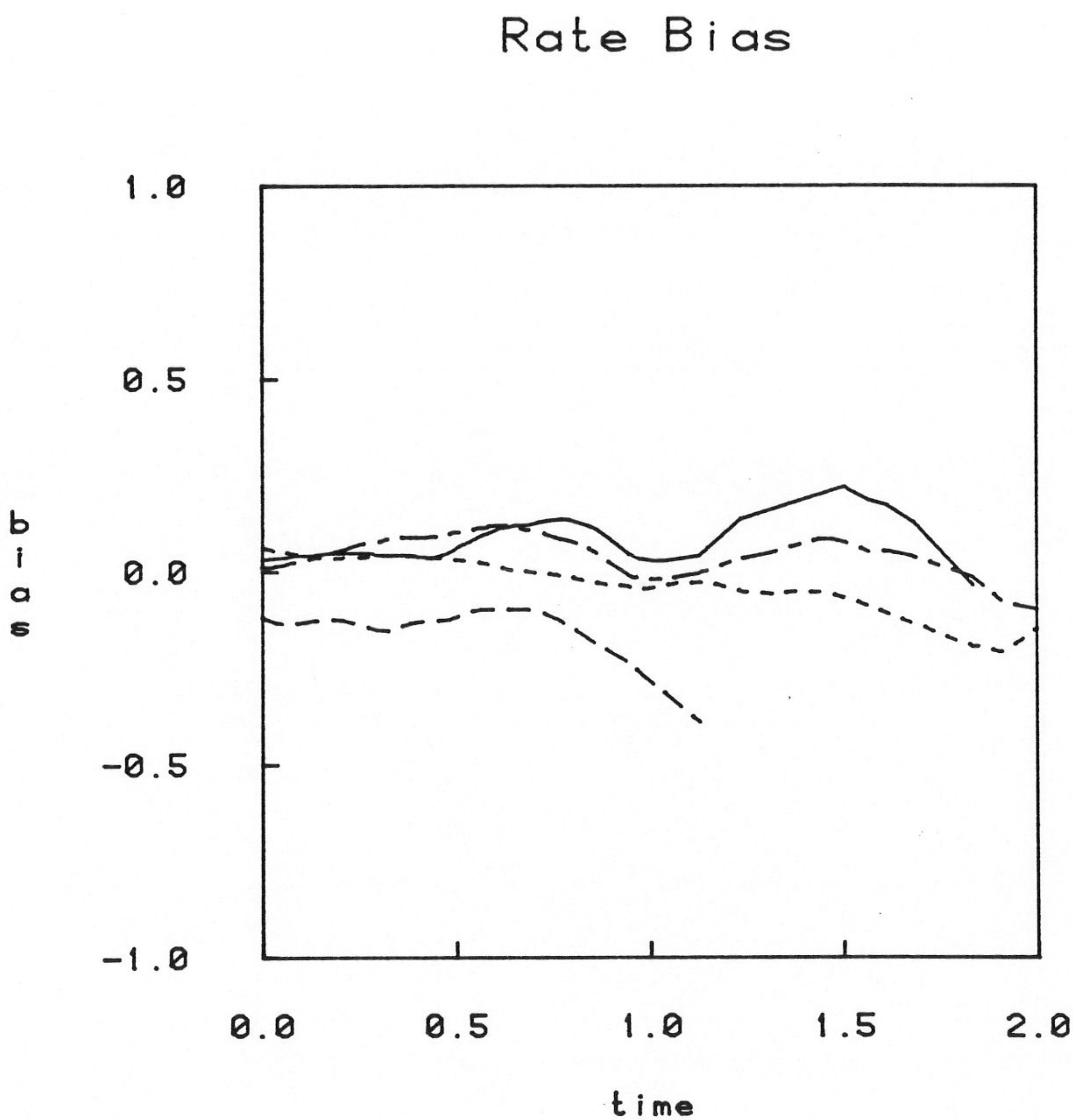

Density Bias
0.50
0.25
0.00
-0.25
b
i
a
s
0.0
0.5
1.0
1.5
2.0
time
short dash: g=0 dot dash: g=0.5
solid line: g=1 long dash: g=2

Figure 1(b)



short dash: g=0   dot dash: g=0.5

solid line: g=1   long dash: g=2

Figure 1(c)

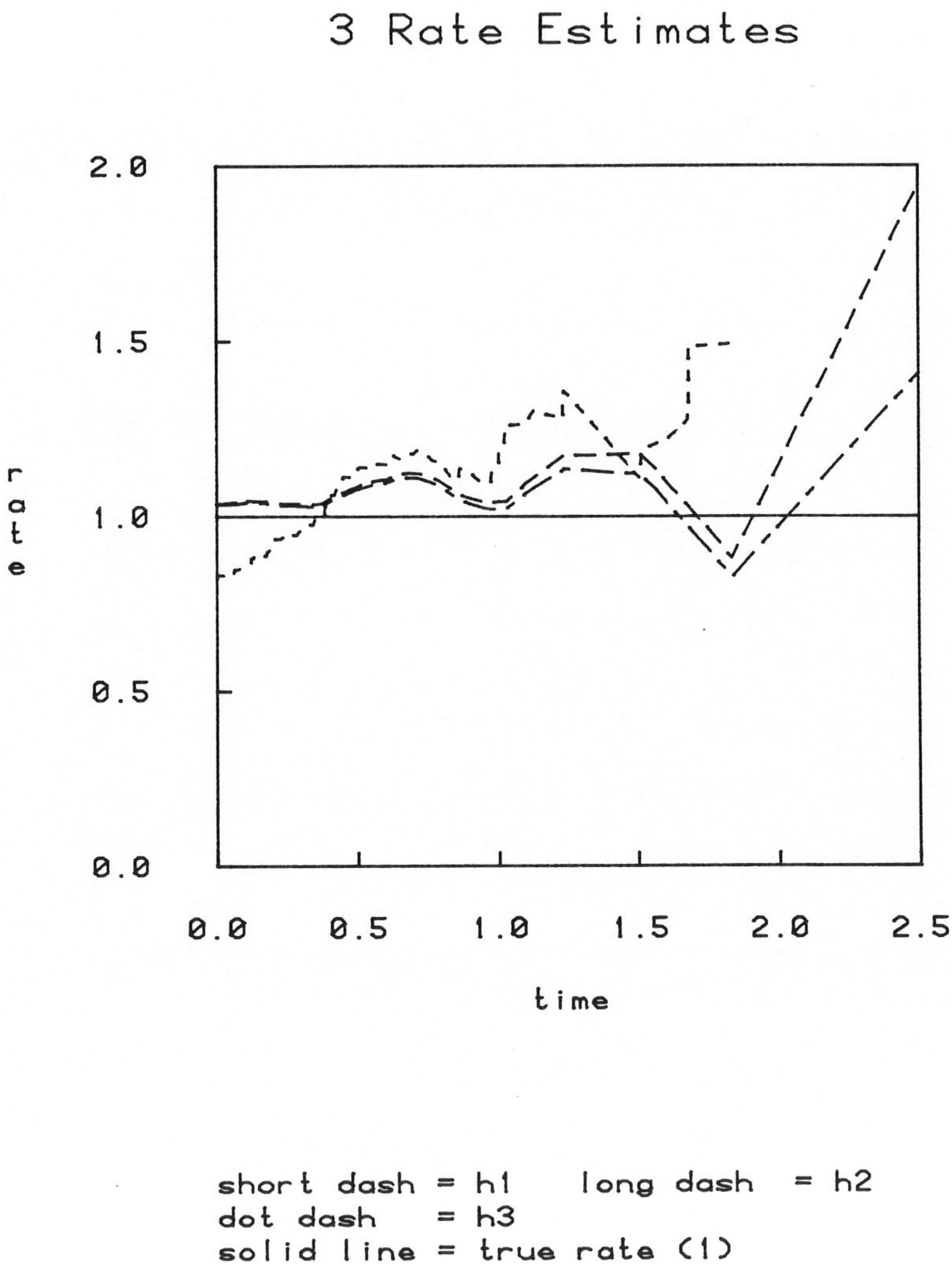
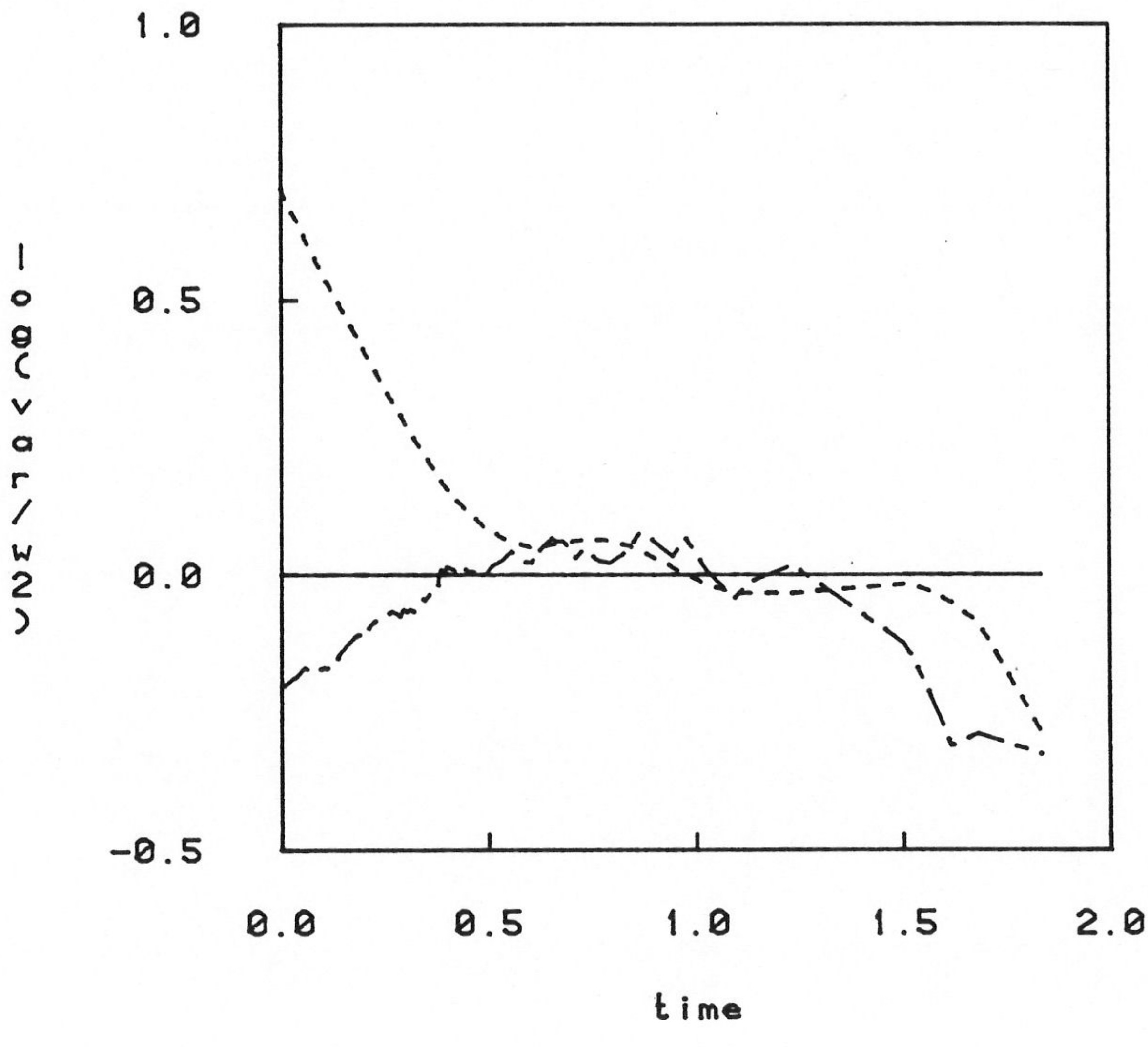

3 Rate Estimates
2.0
1.5
1.0
0.5
0.0
rate
0.0
0.5
1.0
1.5
2.0
2.5
time
short dash = h1 long dash = h2
dot dash = h3
solid line = true rate (1)

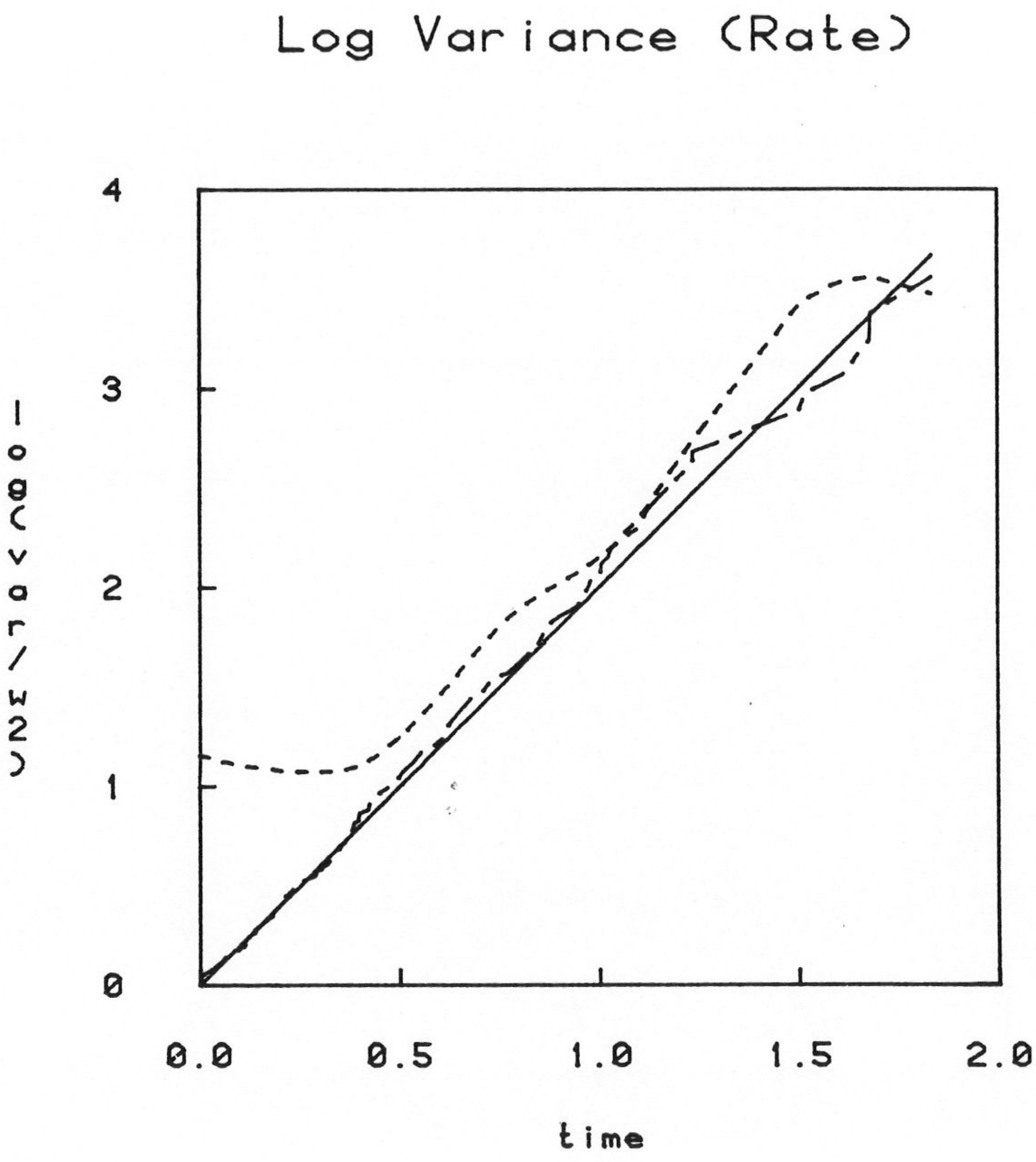
Figure 2(a)

Log Variance (Density)



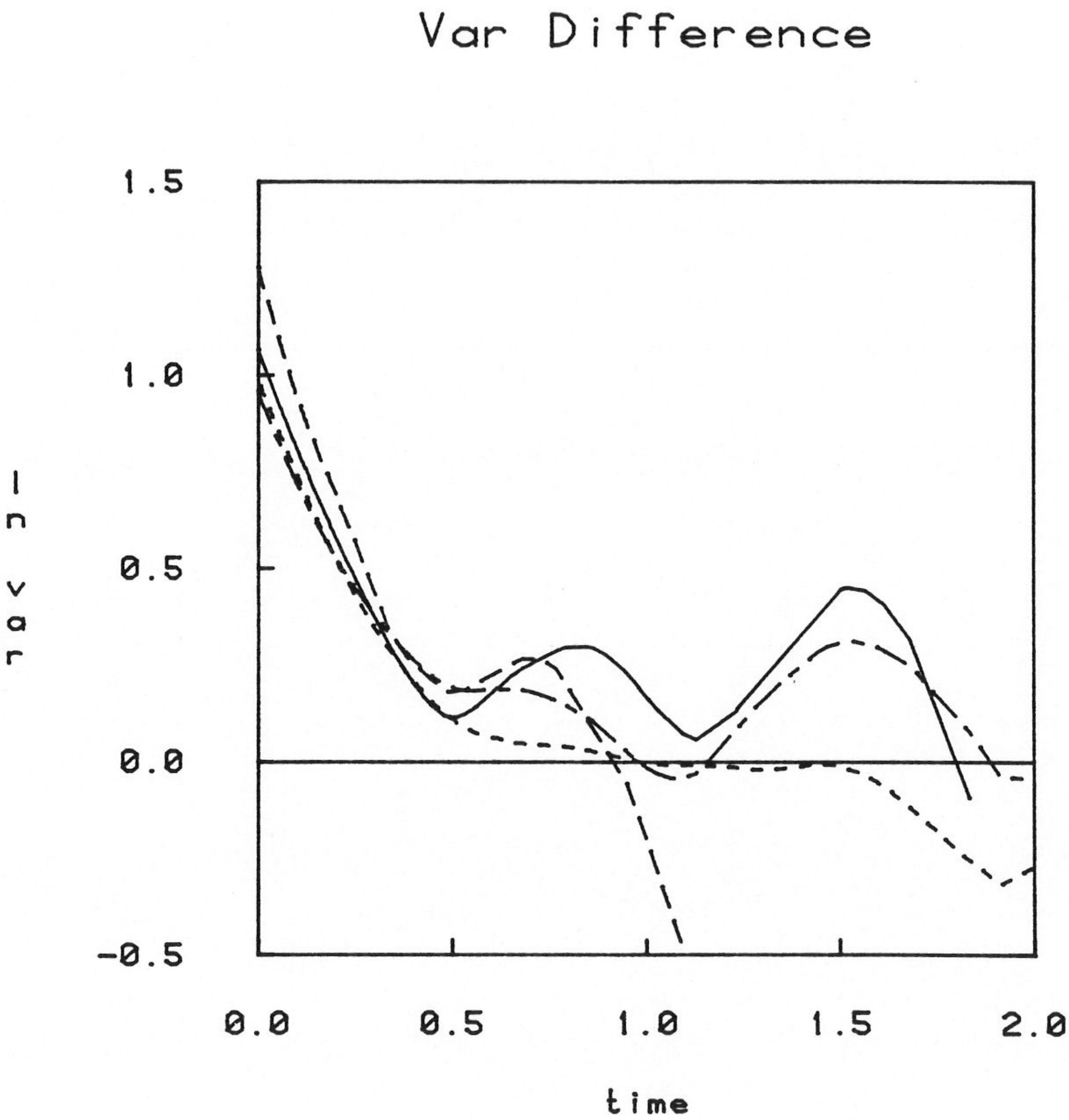
dash = estimate (ESV)
dot-dash = estimated asymptotic (EAV)
solid = theoretical asymptotic (TAV)

Figure 2(b)



dash = estimate (ESV)
dot-dash = estimated asymptotic (EAV)
solid = theoretical asymptotic (TAV)

Figure 2(c)



Censoring parameters are

short dash: g=0   dot dash: g=0.5

solid line: g=1   long dash: g=2

Figure 3(a)

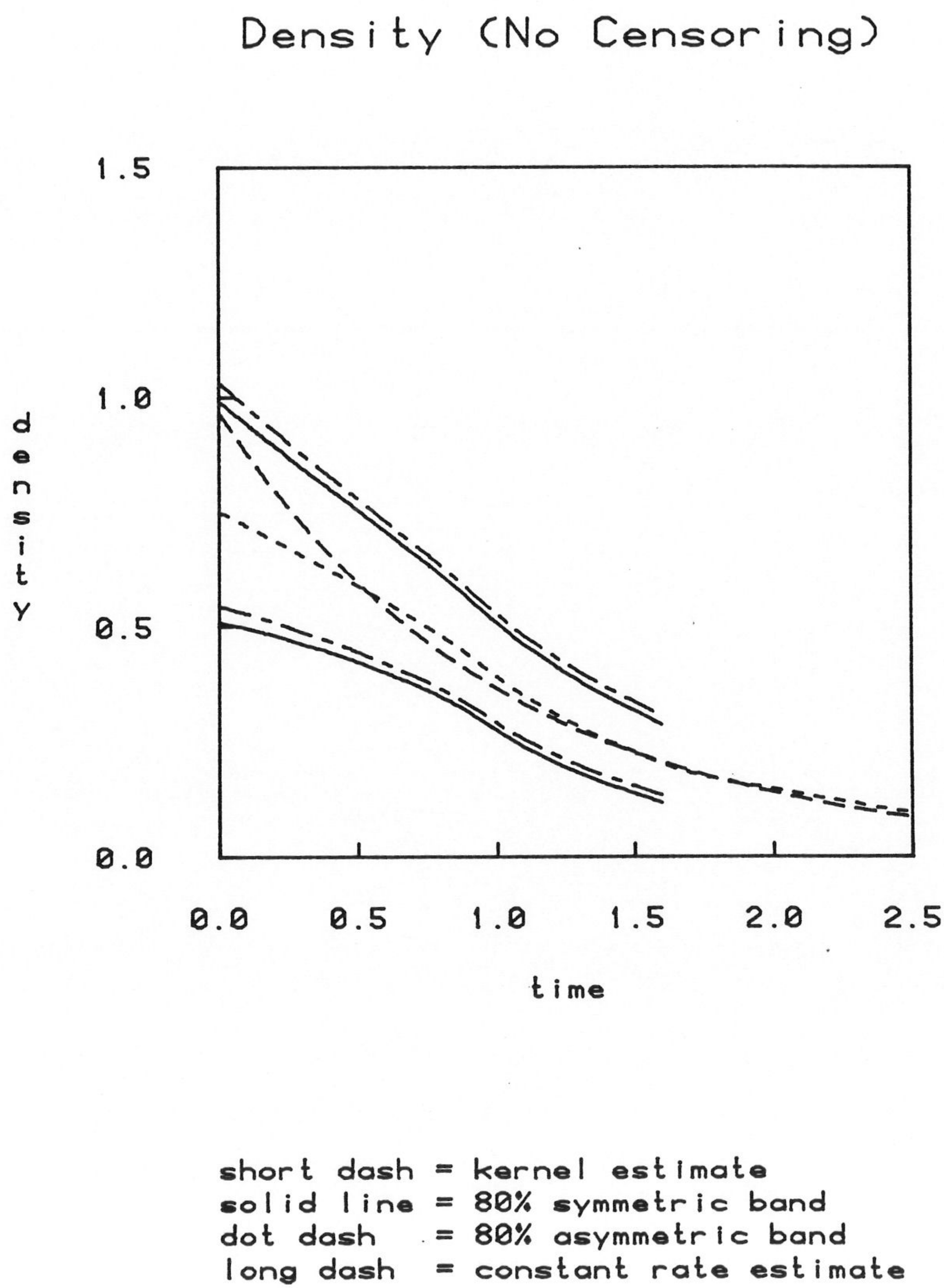
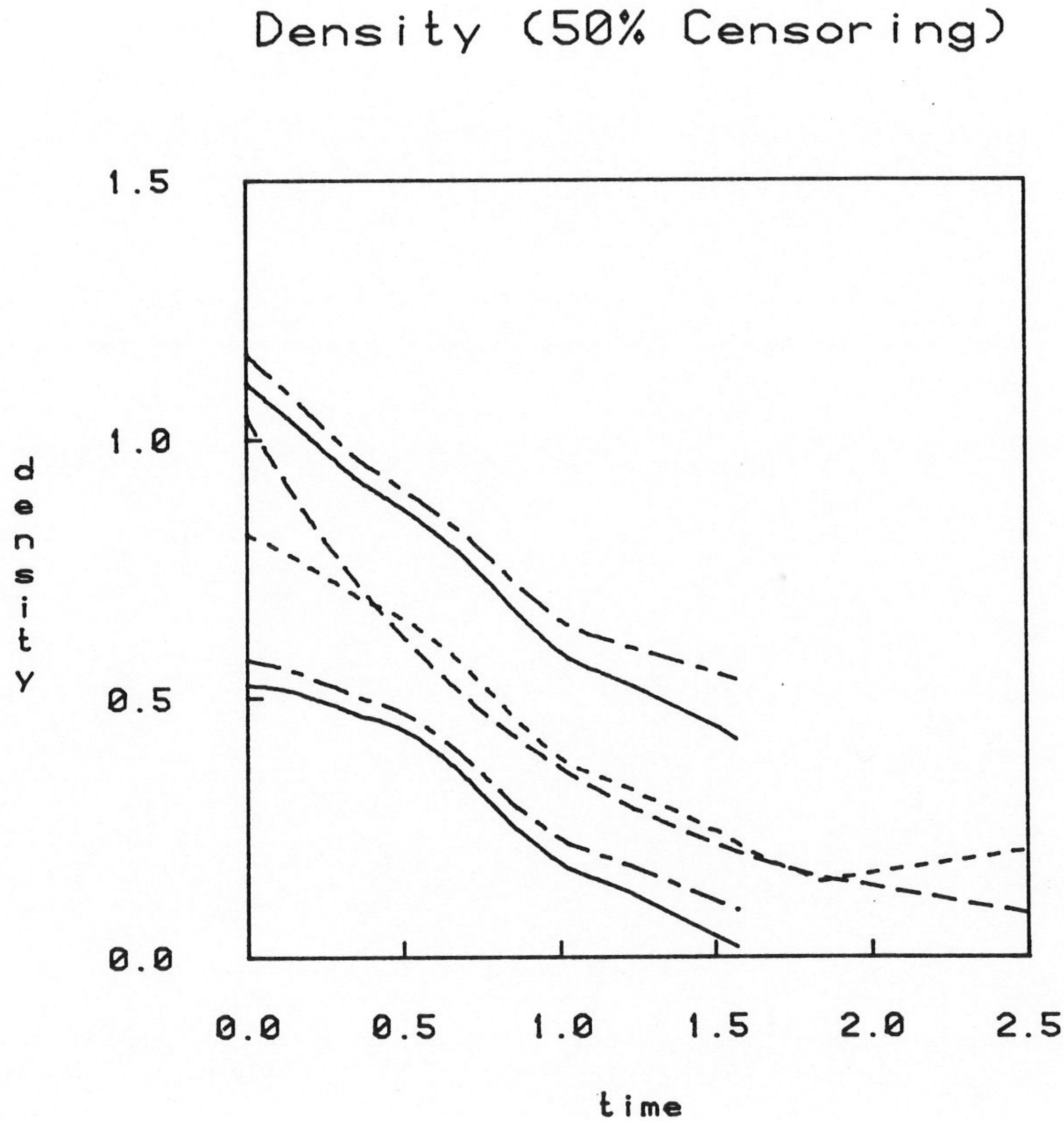
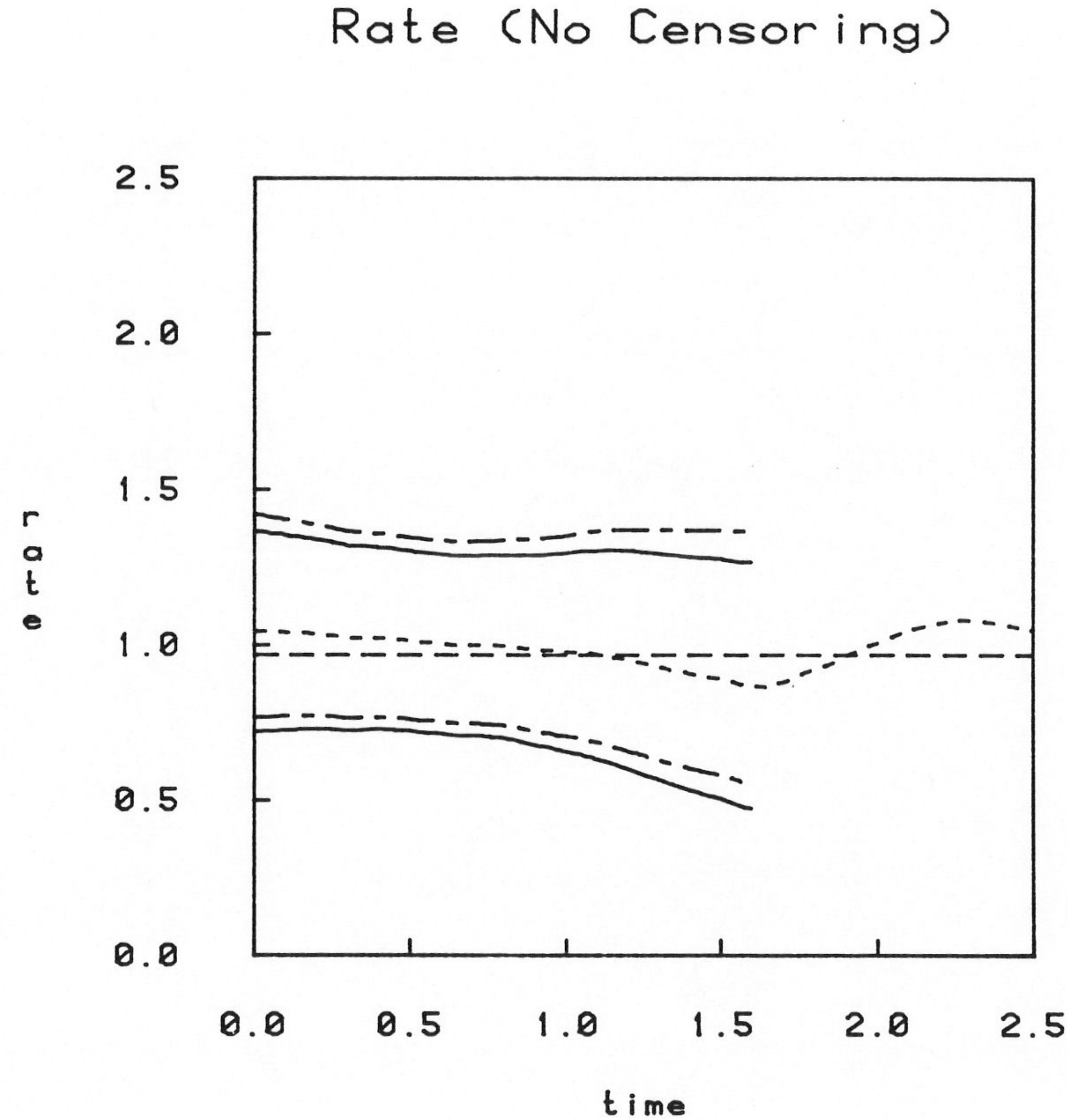

Density (No Censoring)
1.5
1.0
0.5
0.0
density
0.0
0.5
1.0
1.5
2.0
2.5
time
short dash = kernel estimate
solid line = 80% symmetric band
dot dash = 80% asymmetric band
long dash = constant rate estimate

Figure 3(b)



short dash = kernel estimate
solid line = 80% symmetric band
dot dash = 80% asymmetric band
long dash = constant rate estimate

Figure 3(c)



short dash = kernel estimate
solid line = 80% symmetric band
dot dash = 80% asymmetric band
long dash = constant rate estimate

Figure 3(d)

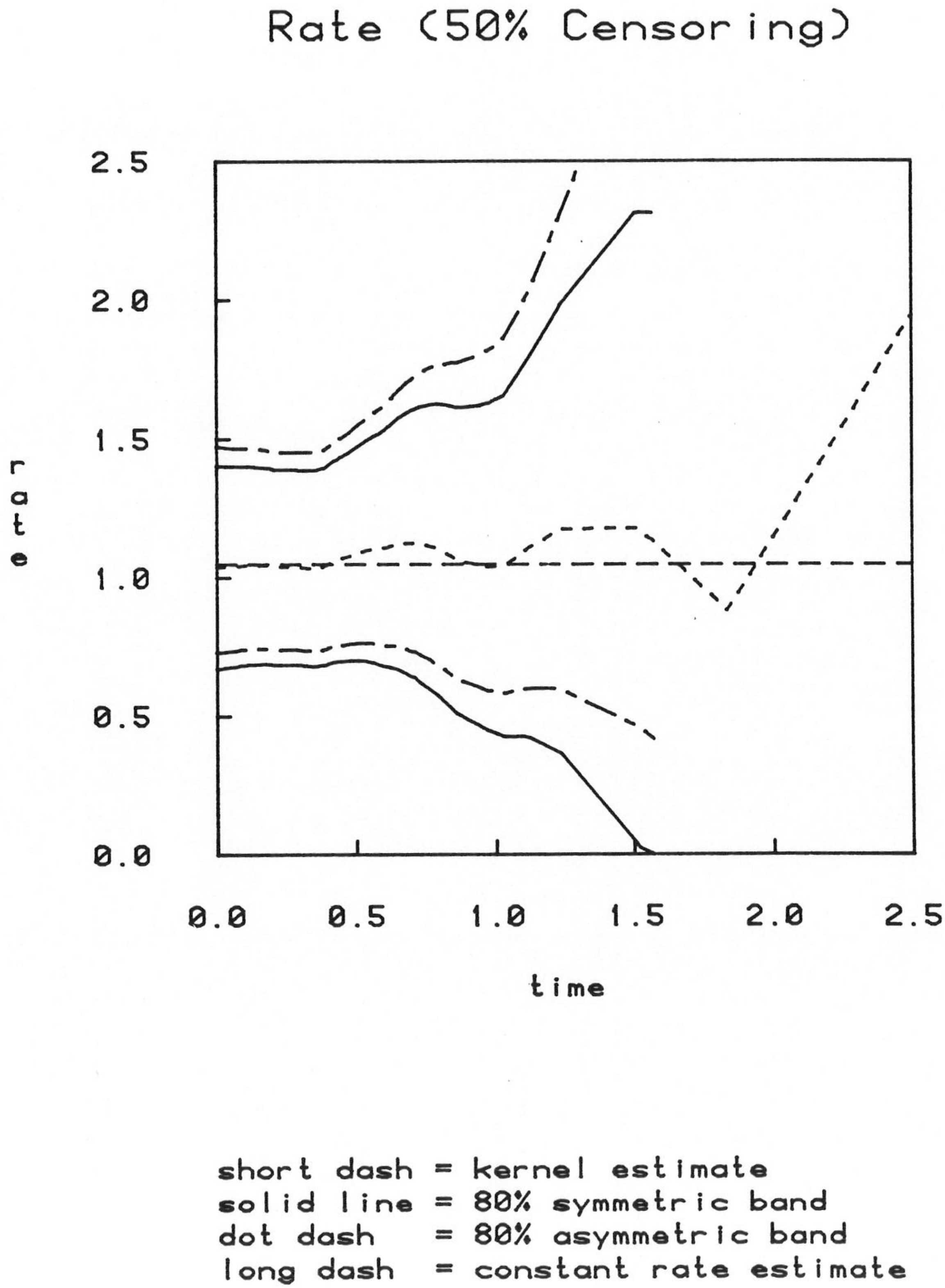

Rate (50% Censoring)
2.5
2.0
1.5
1.0
0.5
0.0
rate
0.0
0.5
1.0
1.5
2.0
2.5
time
short dash = kernel estimate
solid line = 80% symmetric band
dot dash = 80% asymmetric band
long dash = constant rate estimate

Figure 4(a)

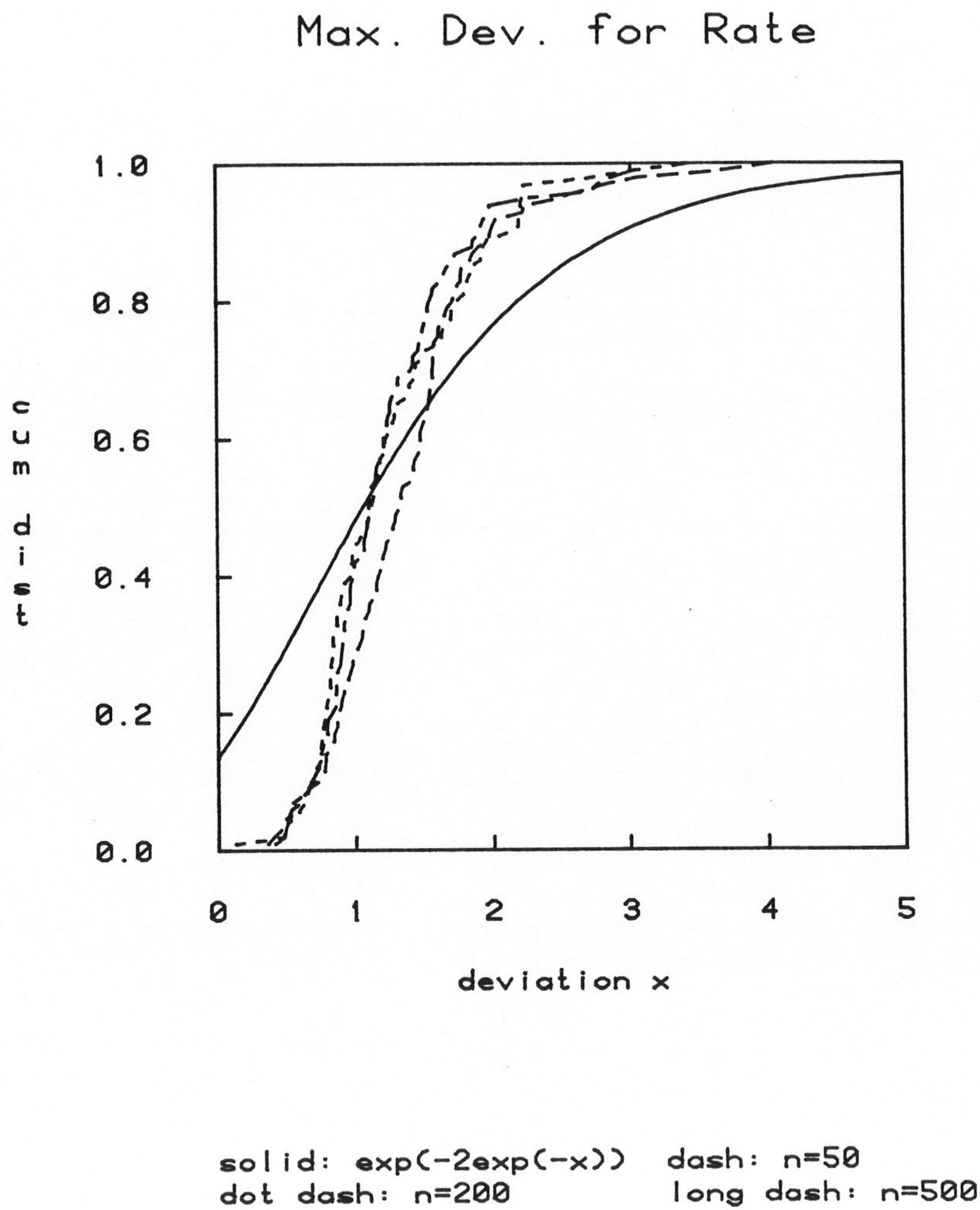

Max. Dev. for Rate
1.0
0.8
0.6
0.4
0.2
0.0
cum dist
0
1
2
3
4
5
deviation x
solid: exp(-2exp(-x))   dash: n=50
dot dash: n=200   long dash: n=500

Figure 4(b)

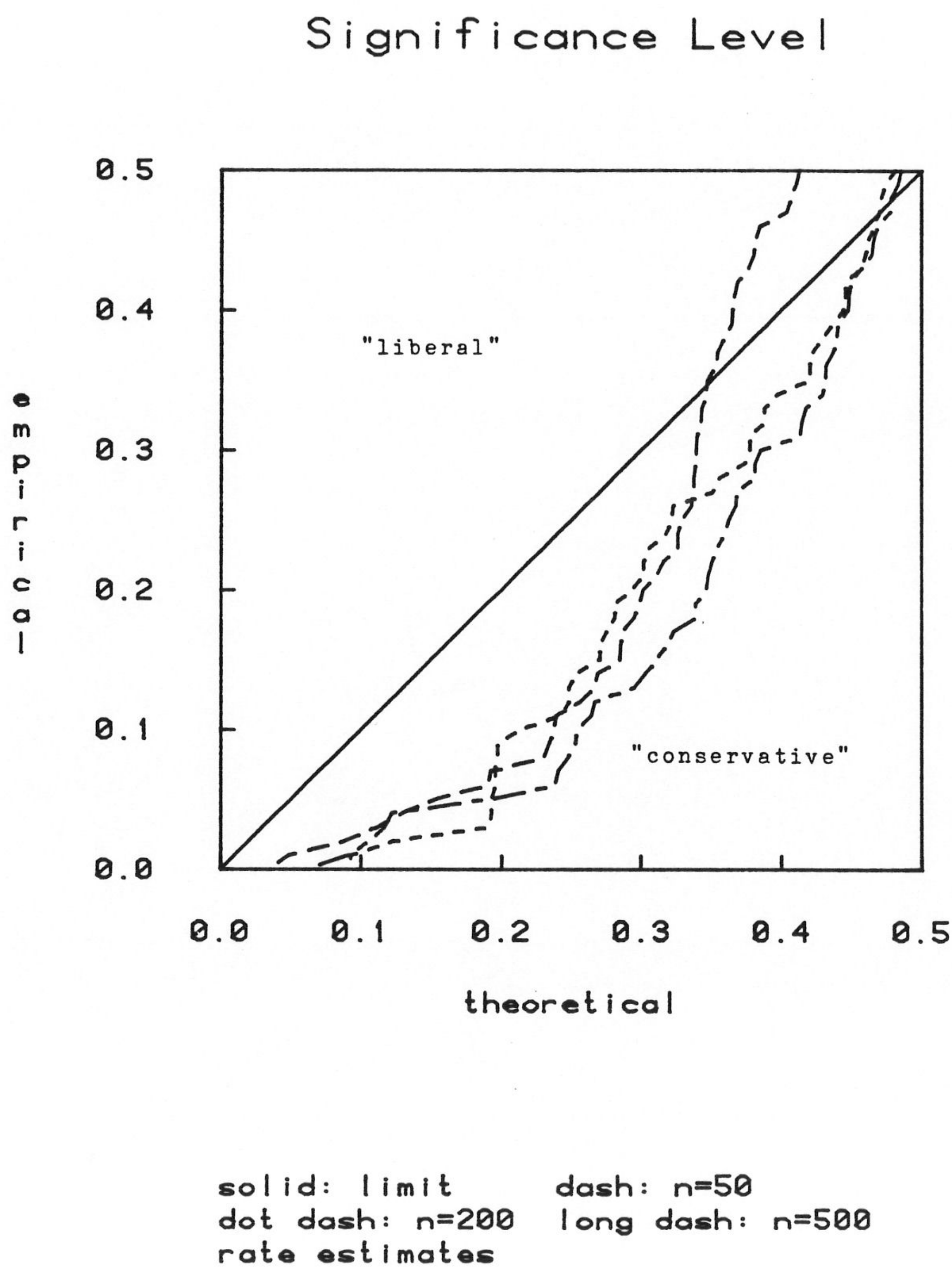

Significance Level
0.5
0.4
0.3
0.2
0.1
0.0
"liberal"
"conservative"
empirical
0.0 0.1 0.2 0.3 0.4 0.5
theoretical
solid: limit dash: n=50
dot dash: n=200 long dash: n=500
rate estimates

Figure 4(c)

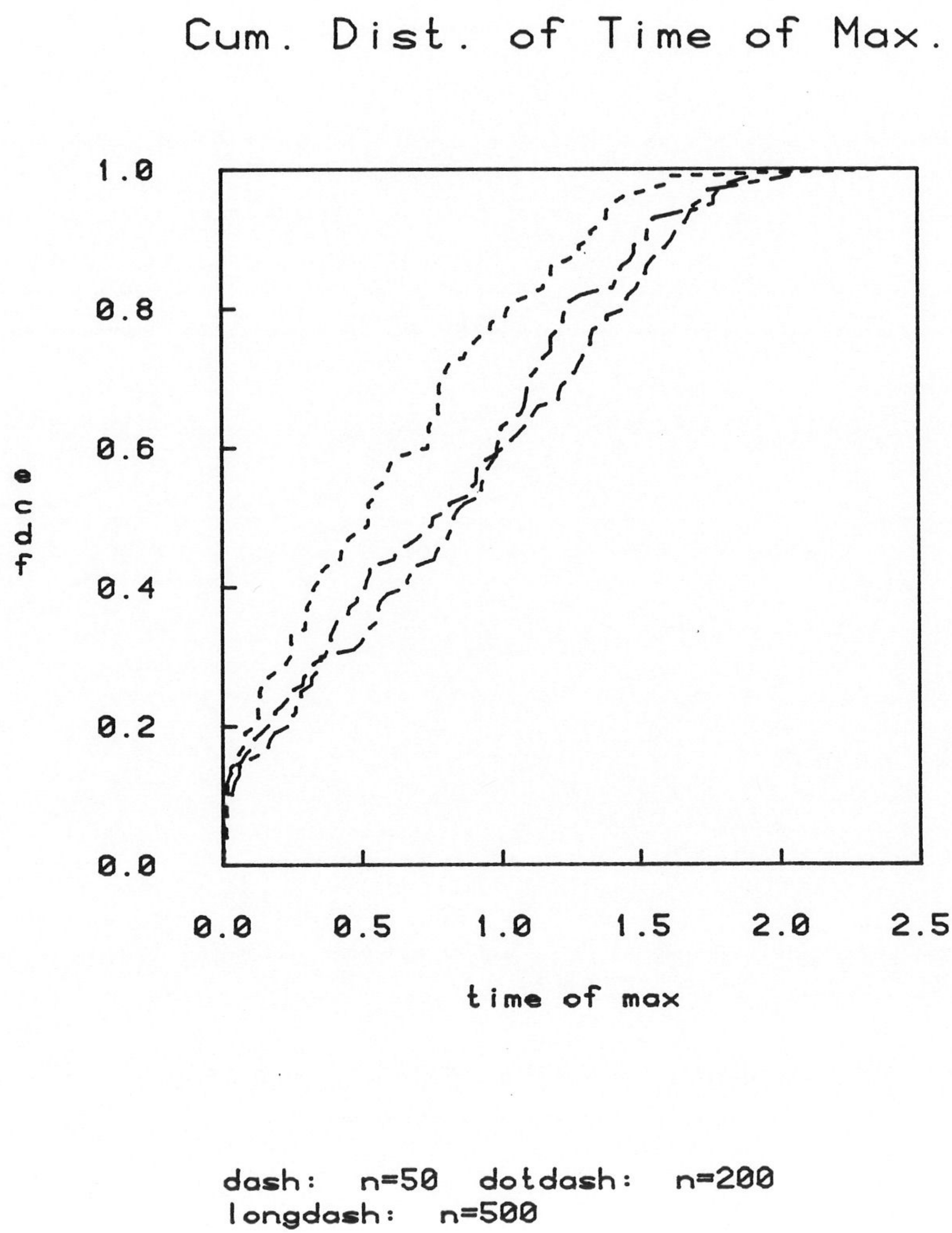
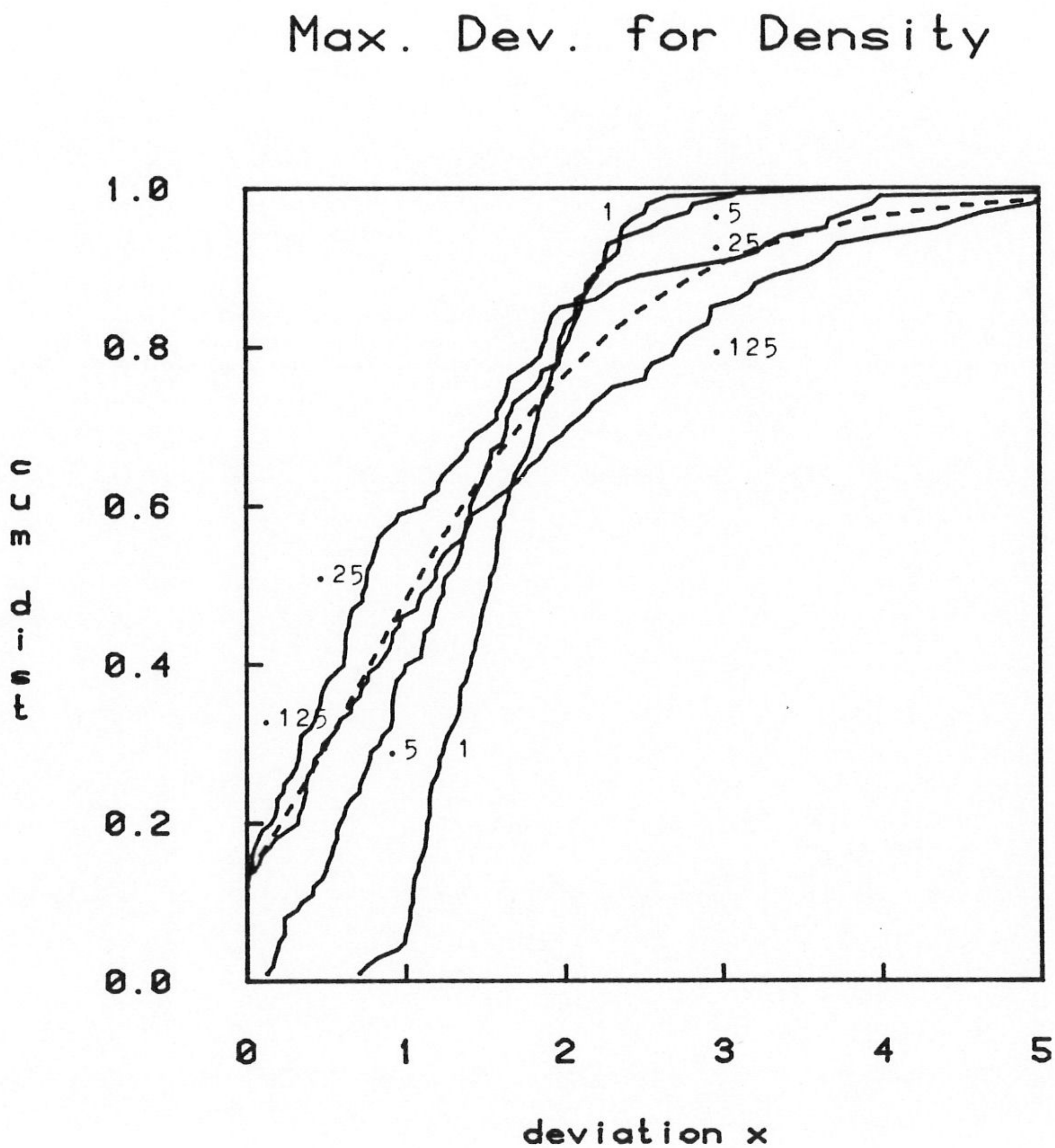
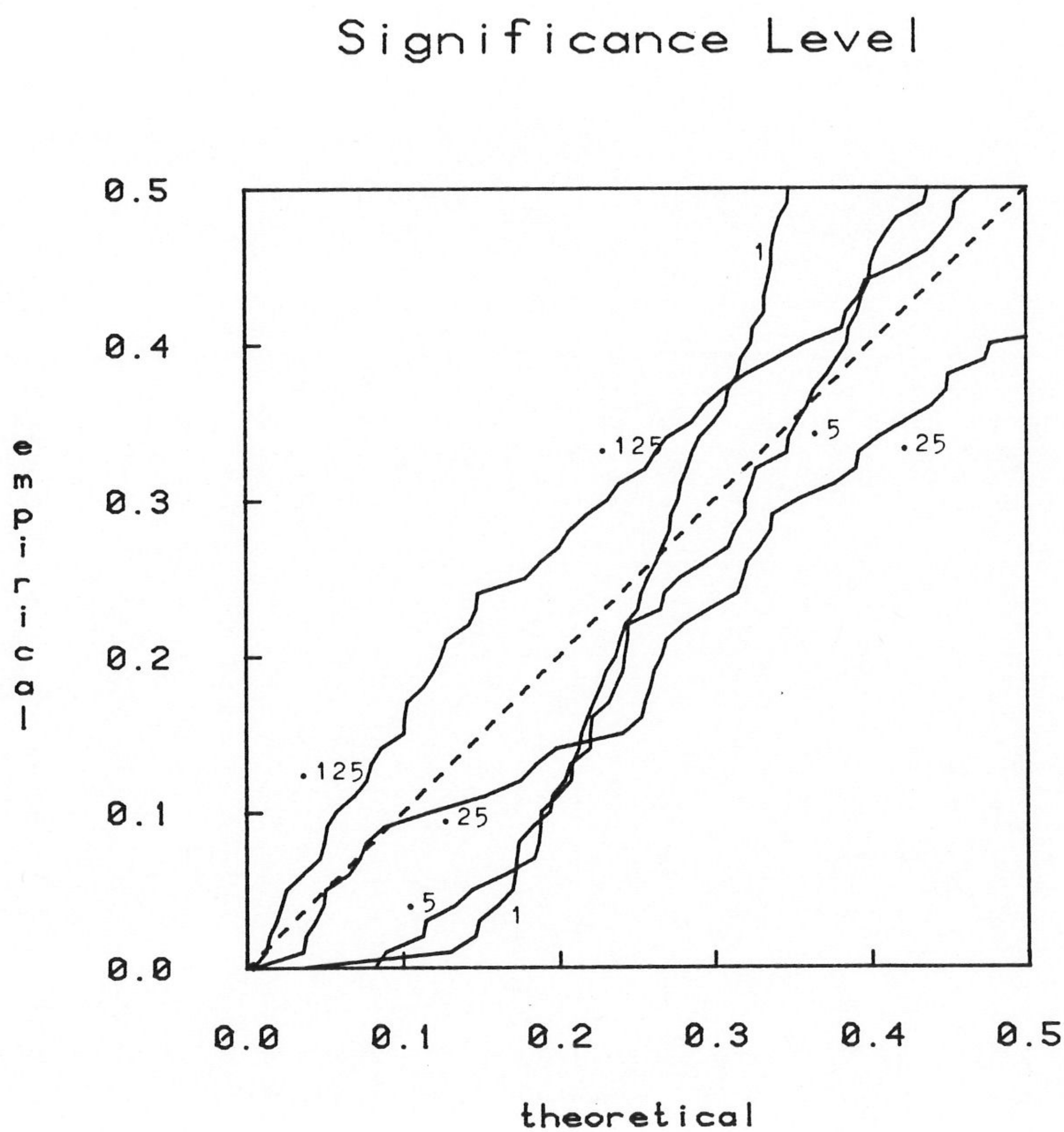

Cum. Dist. of Time of Max.
1.0
0.8
0.6
0.4
0.2
0.0
ecdf
0.0
0.5
1.0
1.5
2.0
2.5
time of max
dash: n=50 dotdash: n=200
longdash: n=500

Figure 5(a)



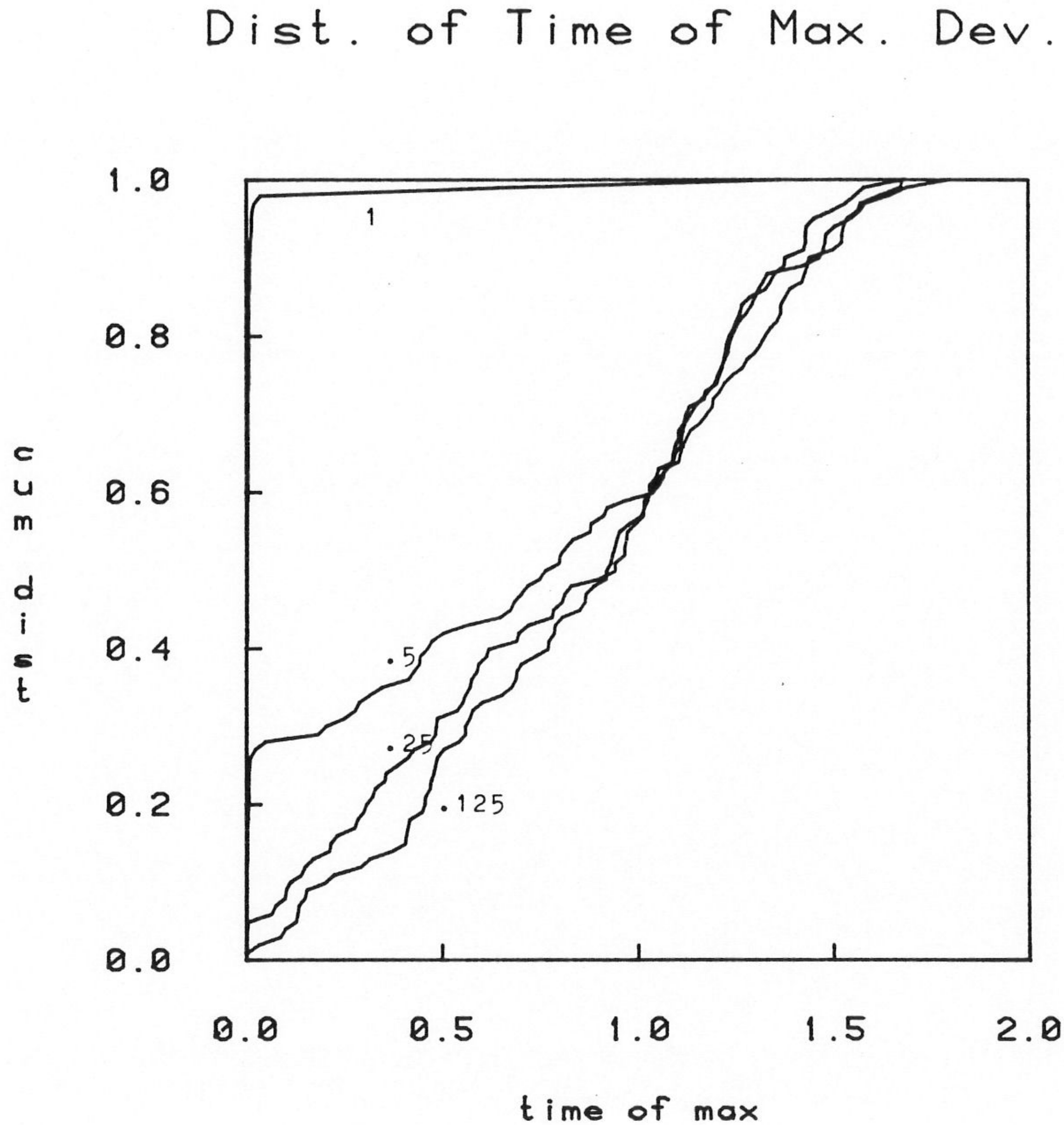
dash = exp(-2exp(-x))
n=200, no censor
bandwidth scale = 1, .5, .25, .125

Figure 5(b)



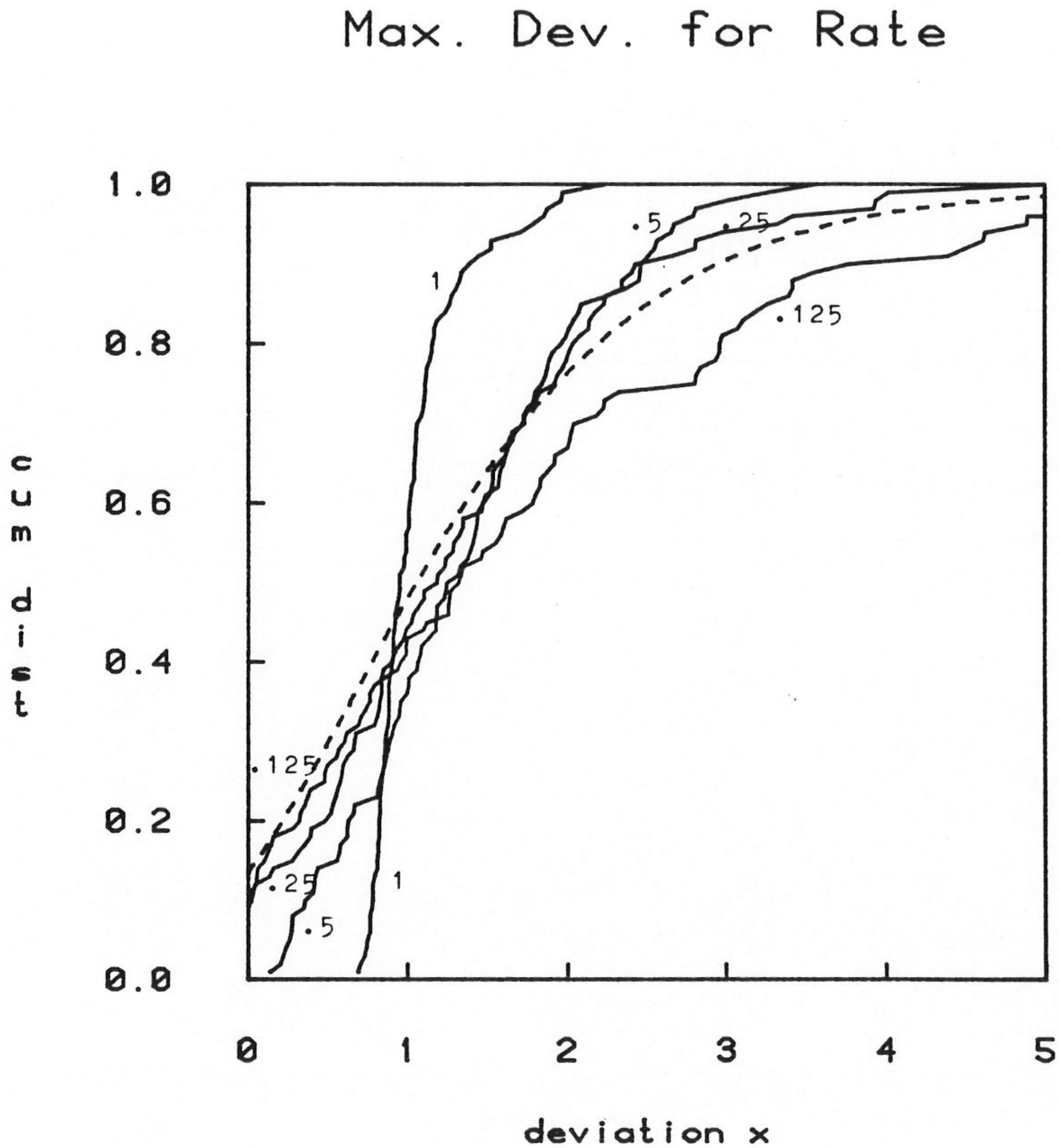
n=200, no censor
bandwidth scale = 1, .5, .25, .125
density estimates

Figure 5(c)



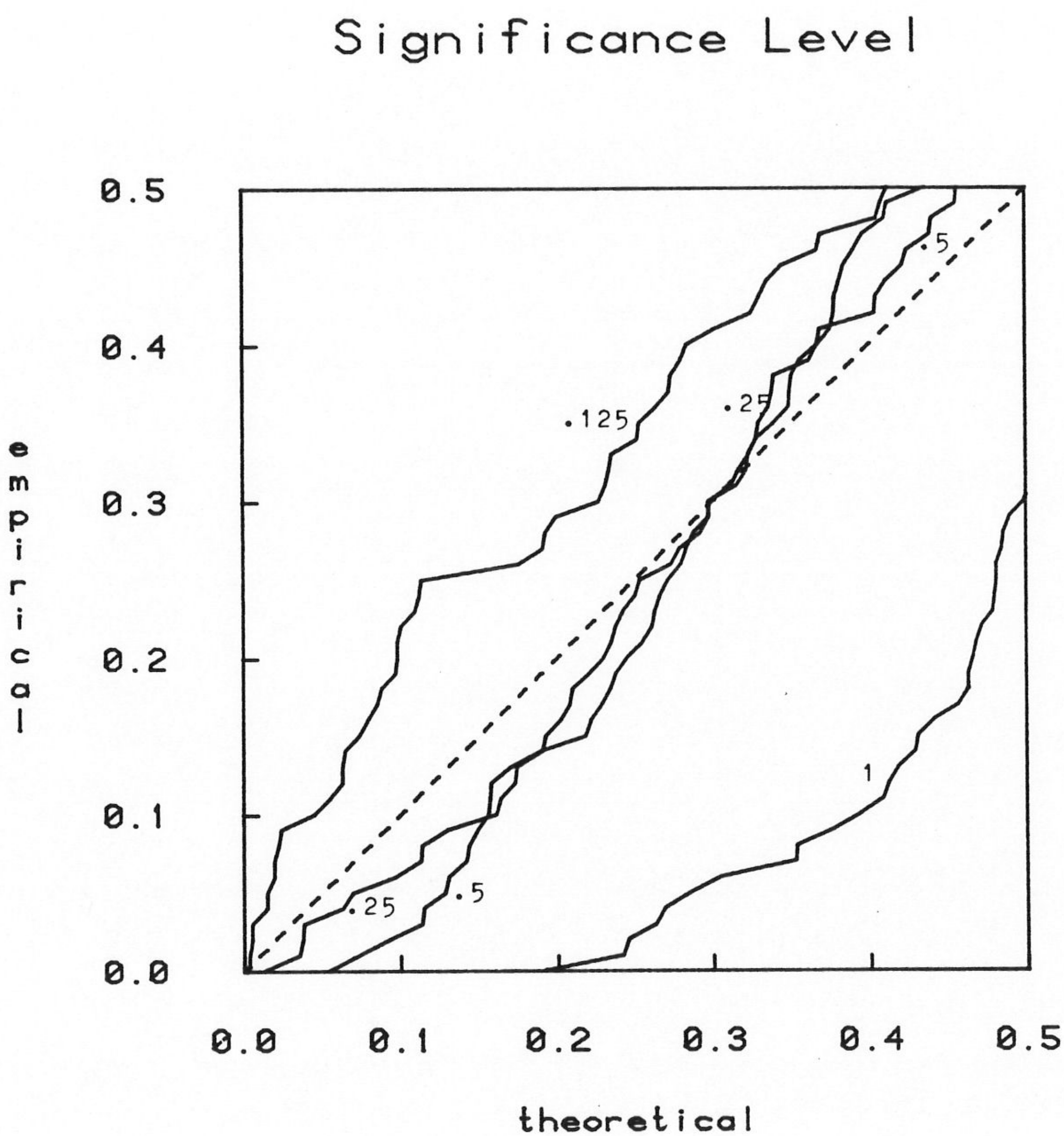
n=200, no censor
bandwidth sscale = 1, .5, .25, .125
rate estimates

Figure 5(d)



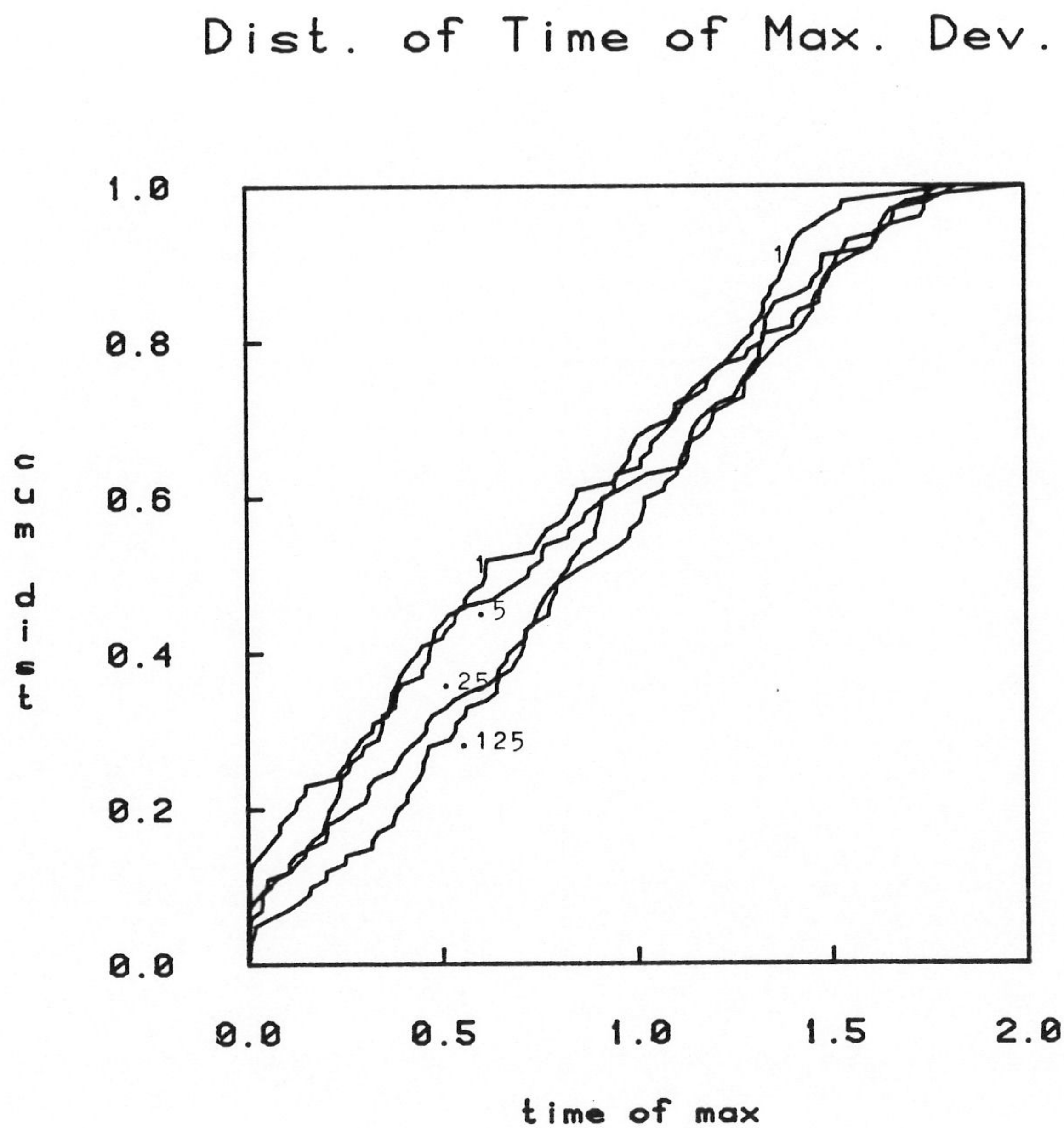
dash = exp(-2exp(-x))
n=200, no censor
bandwidth scale = 1, .5, .25, .125

Figure 5(e)



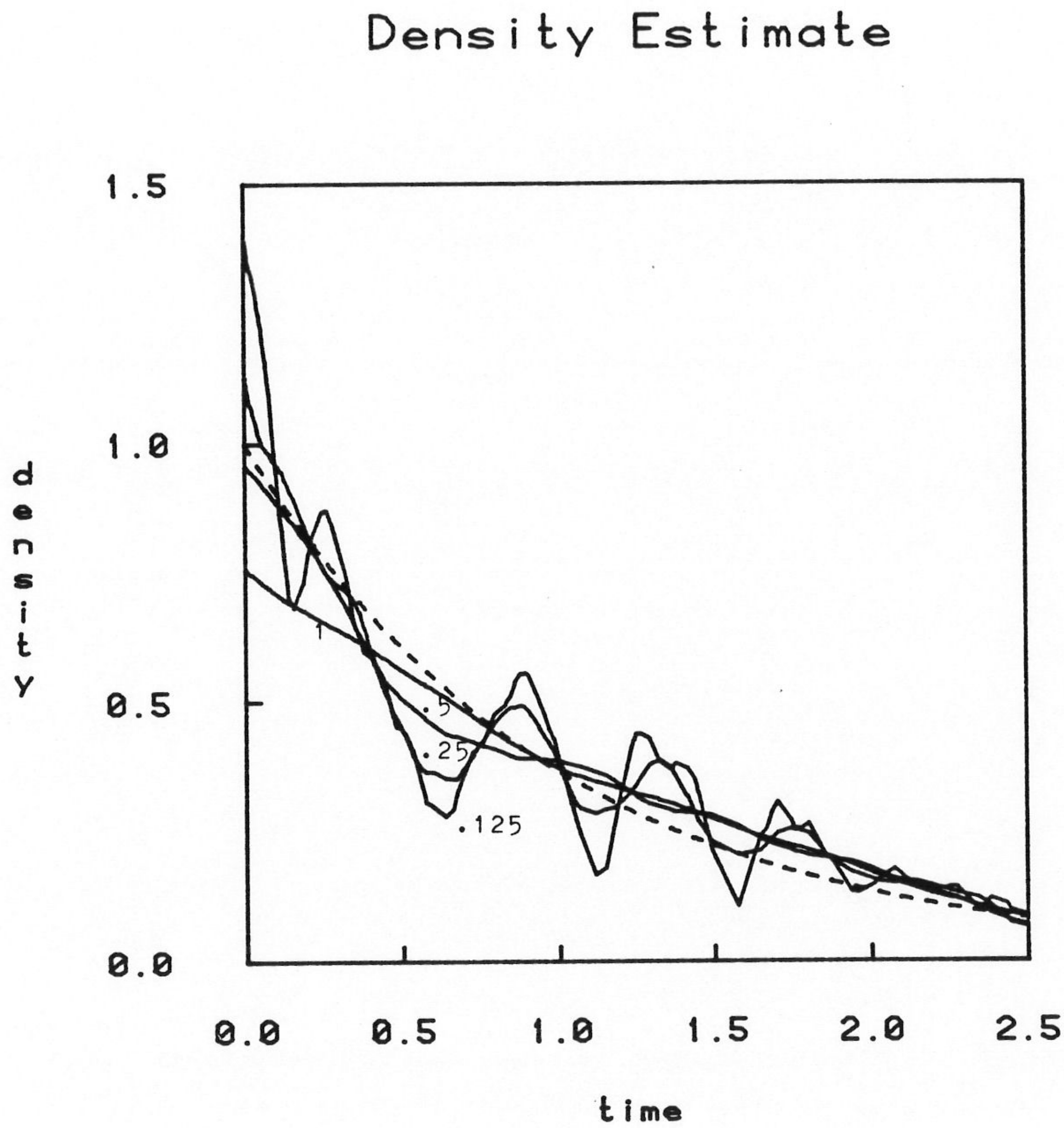
n=200, no censor
bandwidth scale = 1, .5, .25, .125
rate estimates

Figure 5(f)



n=200, no censor
bandwidth scale = 1, .5, .25, .125

Figure 6(a)



n=200, no censor
bandwidth scale = 1, .5, .25, .125

Figure 6(b)

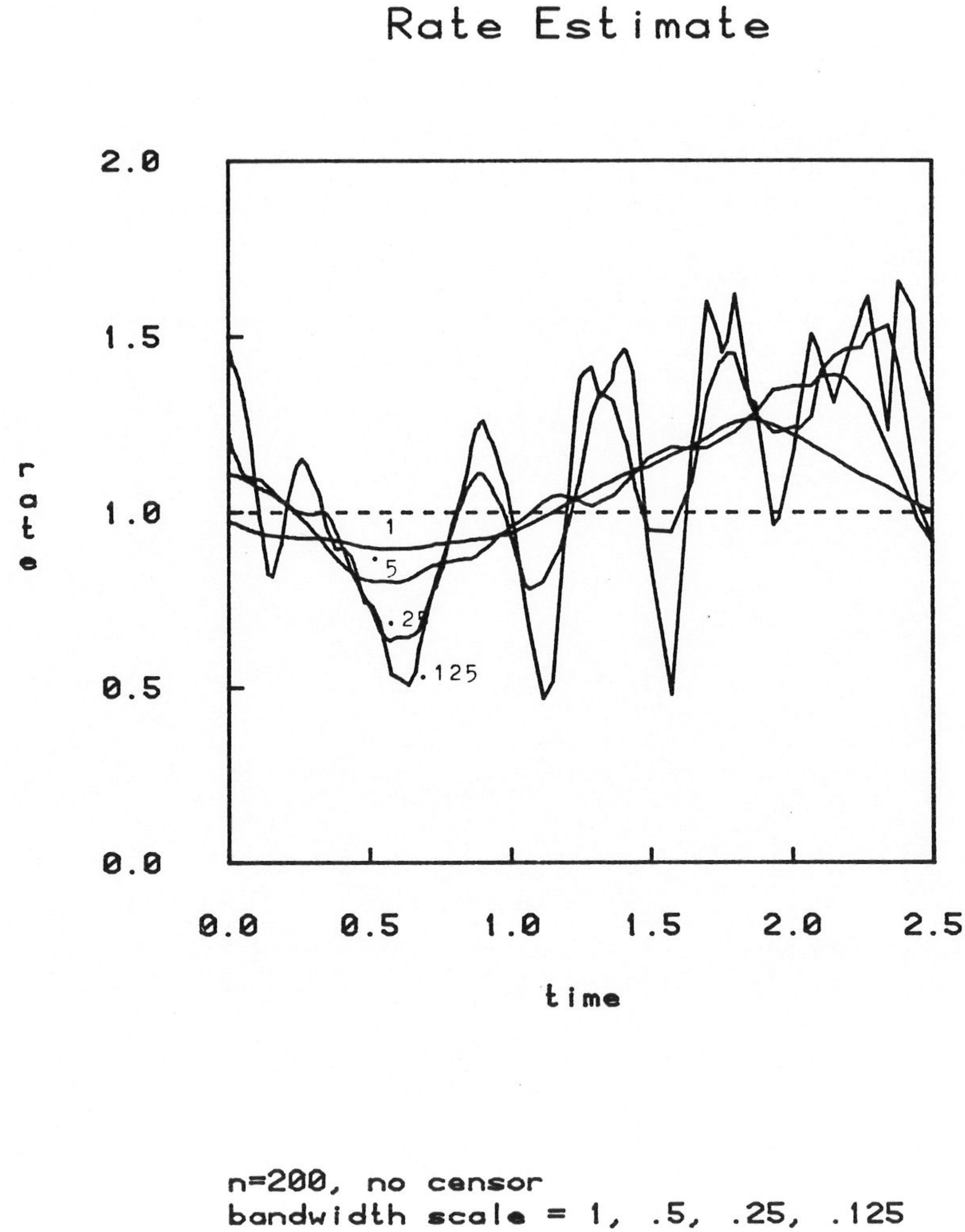

Rate Estimate
2.0
1.5
1.0
0.5
0.0
rate
1
.5
.25
.125
0.0
0.5
1.0
1.5
2.0
2.5
time
n=200, no censor
bandwidth scale = 1, .5, .25, .125

Figure 7(a)

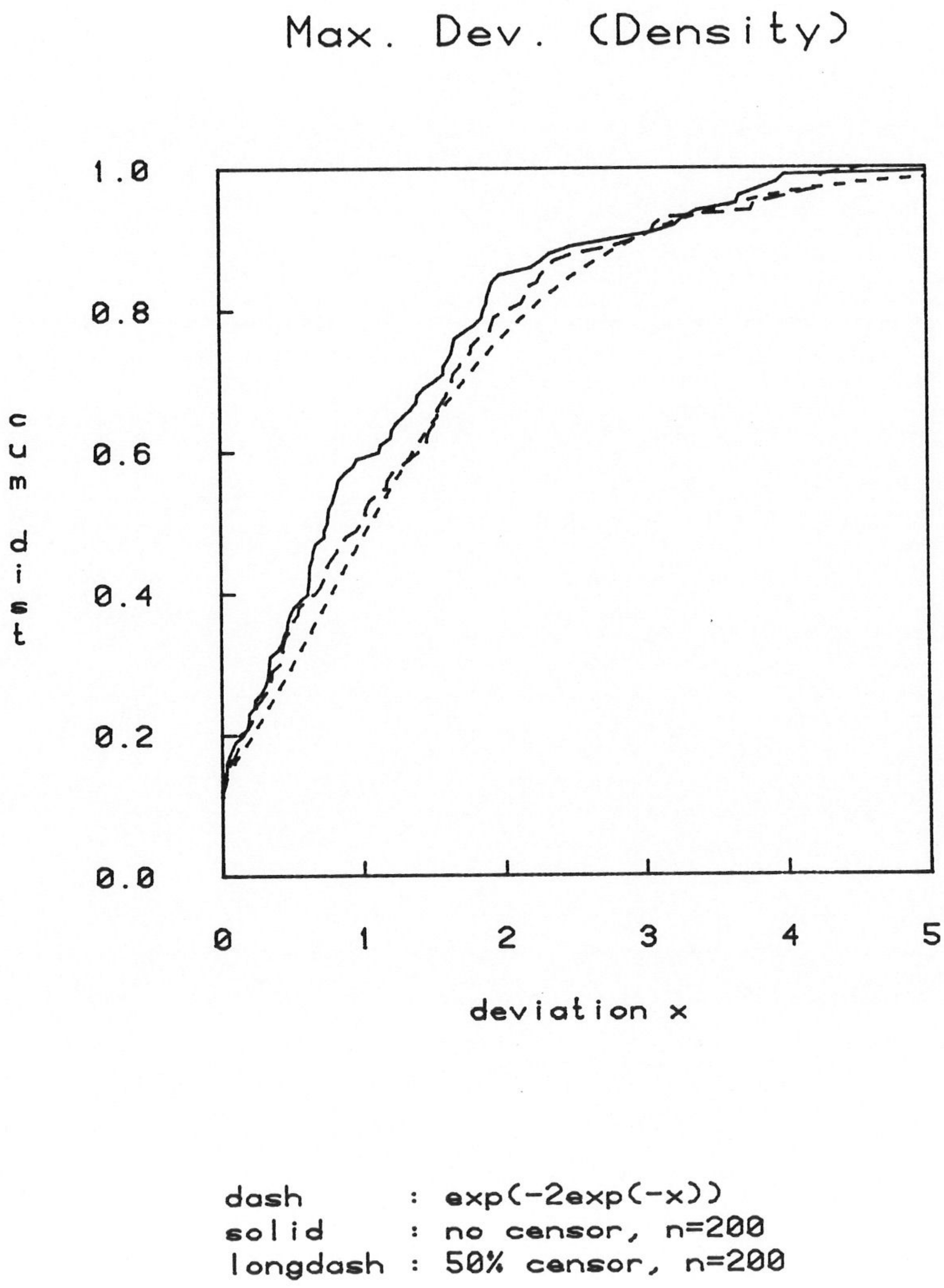

Max. Dev. (Density)
cum dist
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
deviation x
dash : exp(-2exp(-x))
solid : no censor, n=200
longdash : 50% censor, n=200

Figure 7(b)

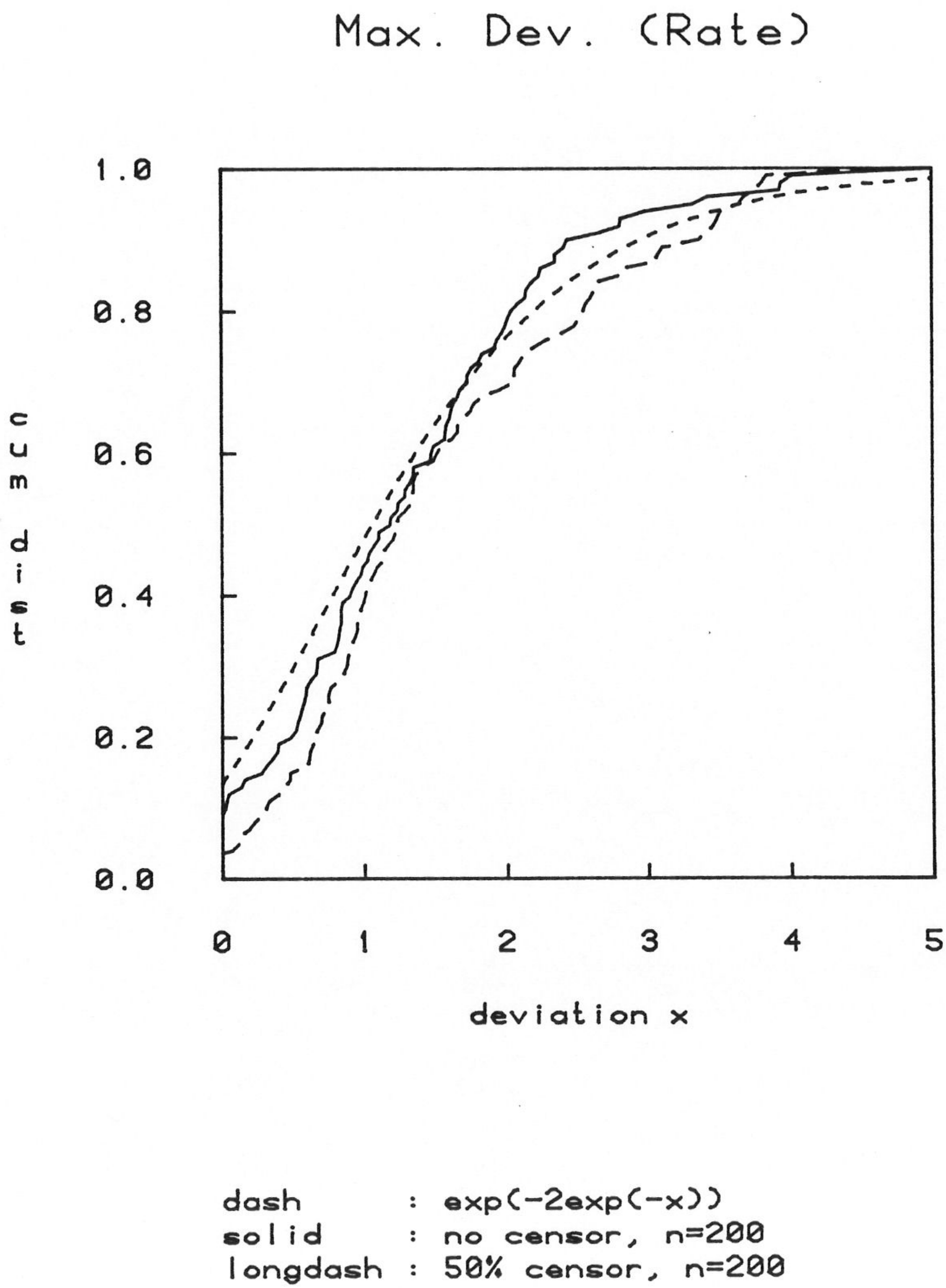
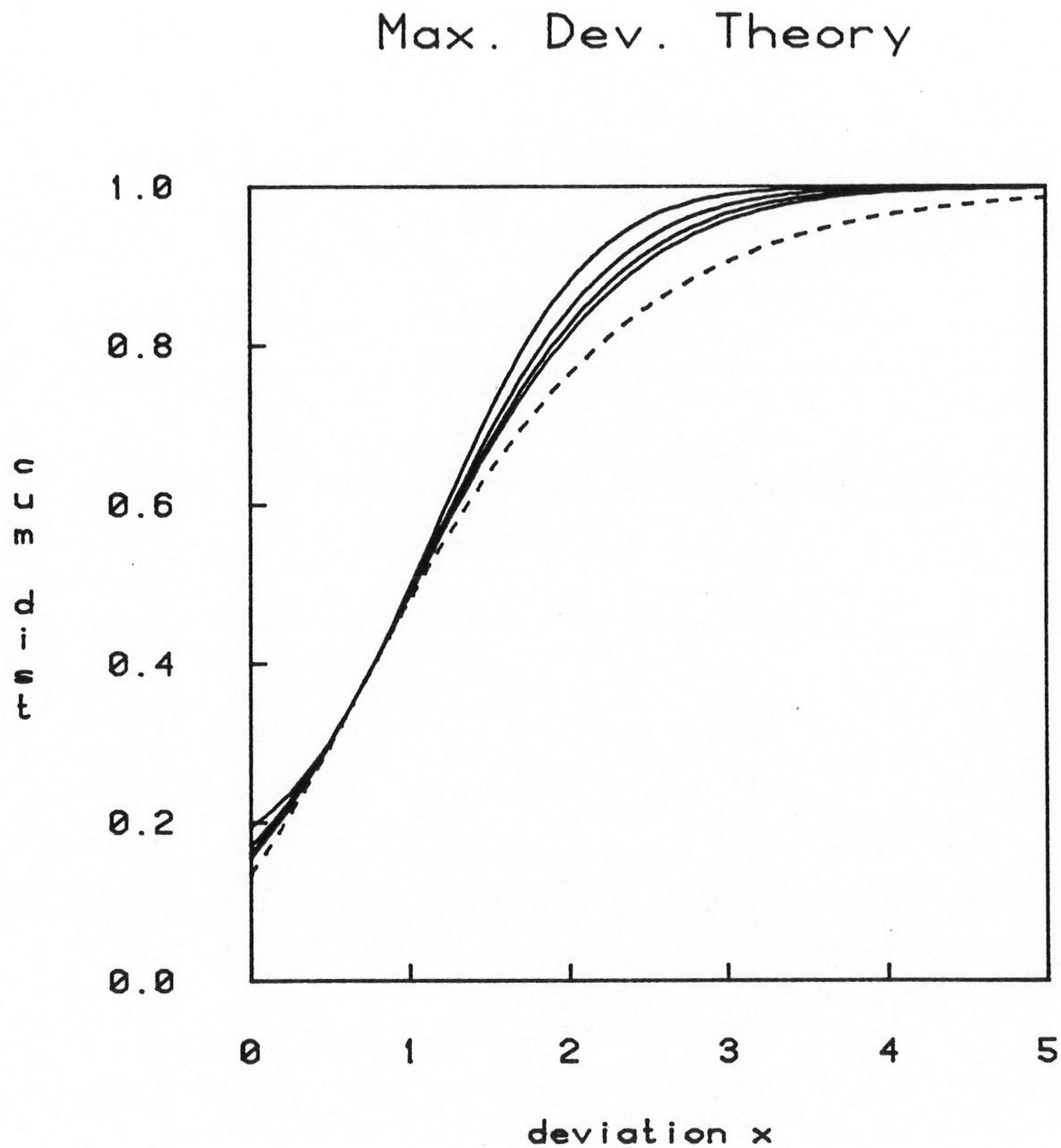

Max. Dev. (Rate)
1.0
0.8
0.6
0.4
0.2
0.0
cum dist
0
1
2
3
4
5
deviation x
dash : exp(-2exp(-x))
solid : no censor, n=200
longdash : 50% censor, n=200

Figure 8(a)



dash: exp(-2exp(-x))
solid: bandwidth varies
by factor of 2 from 0.8 to 0.1

Figure 8(b)

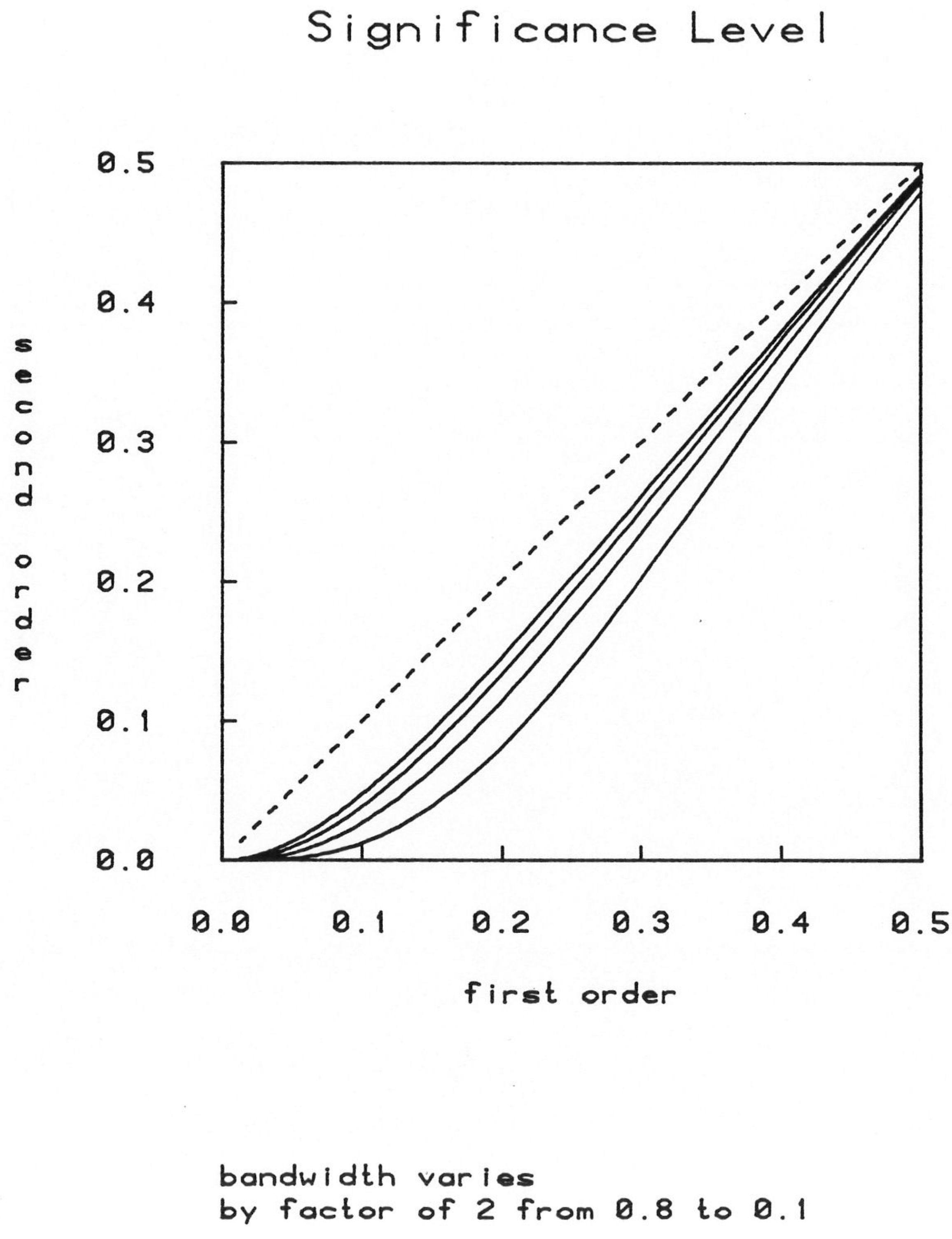
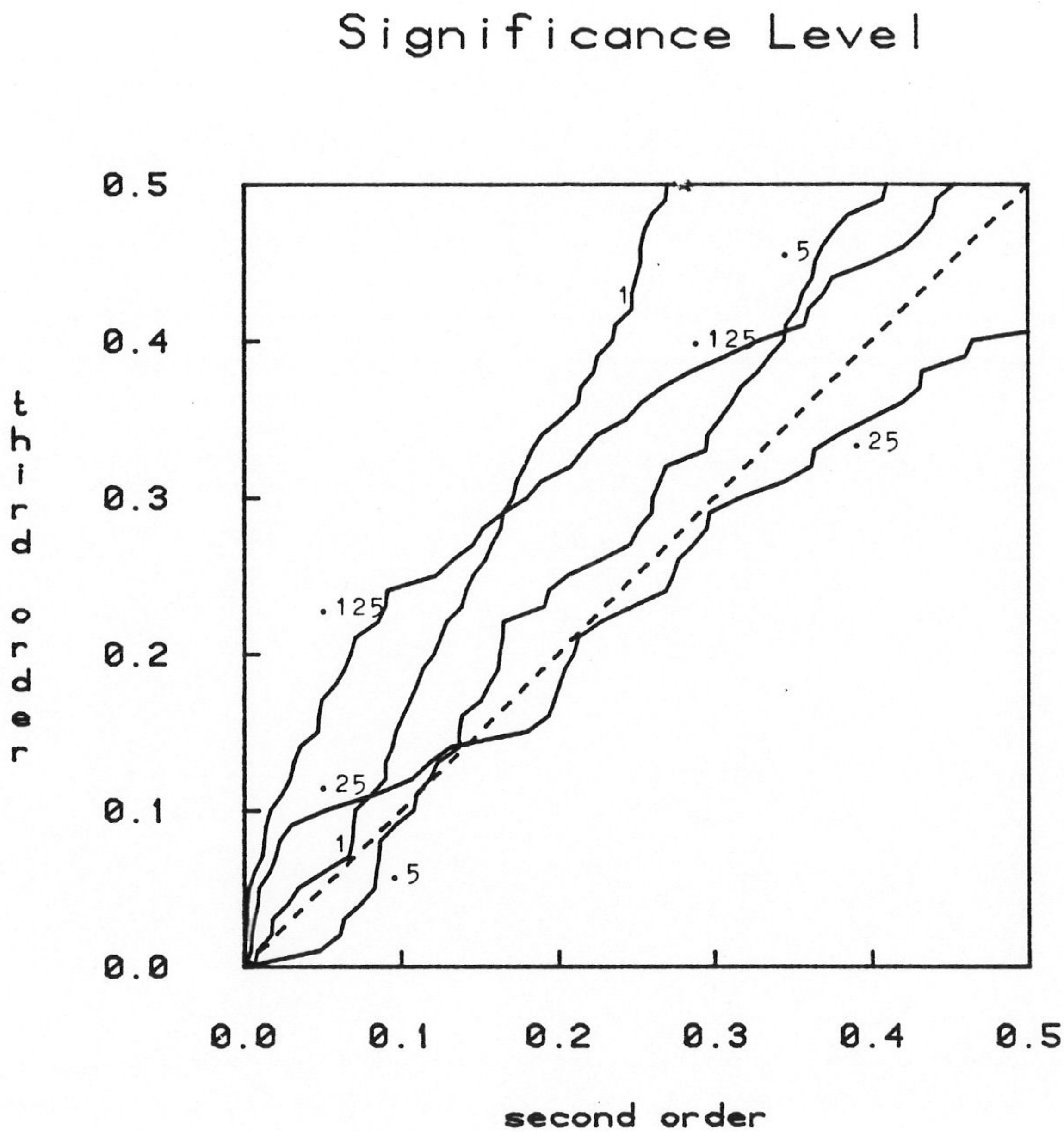
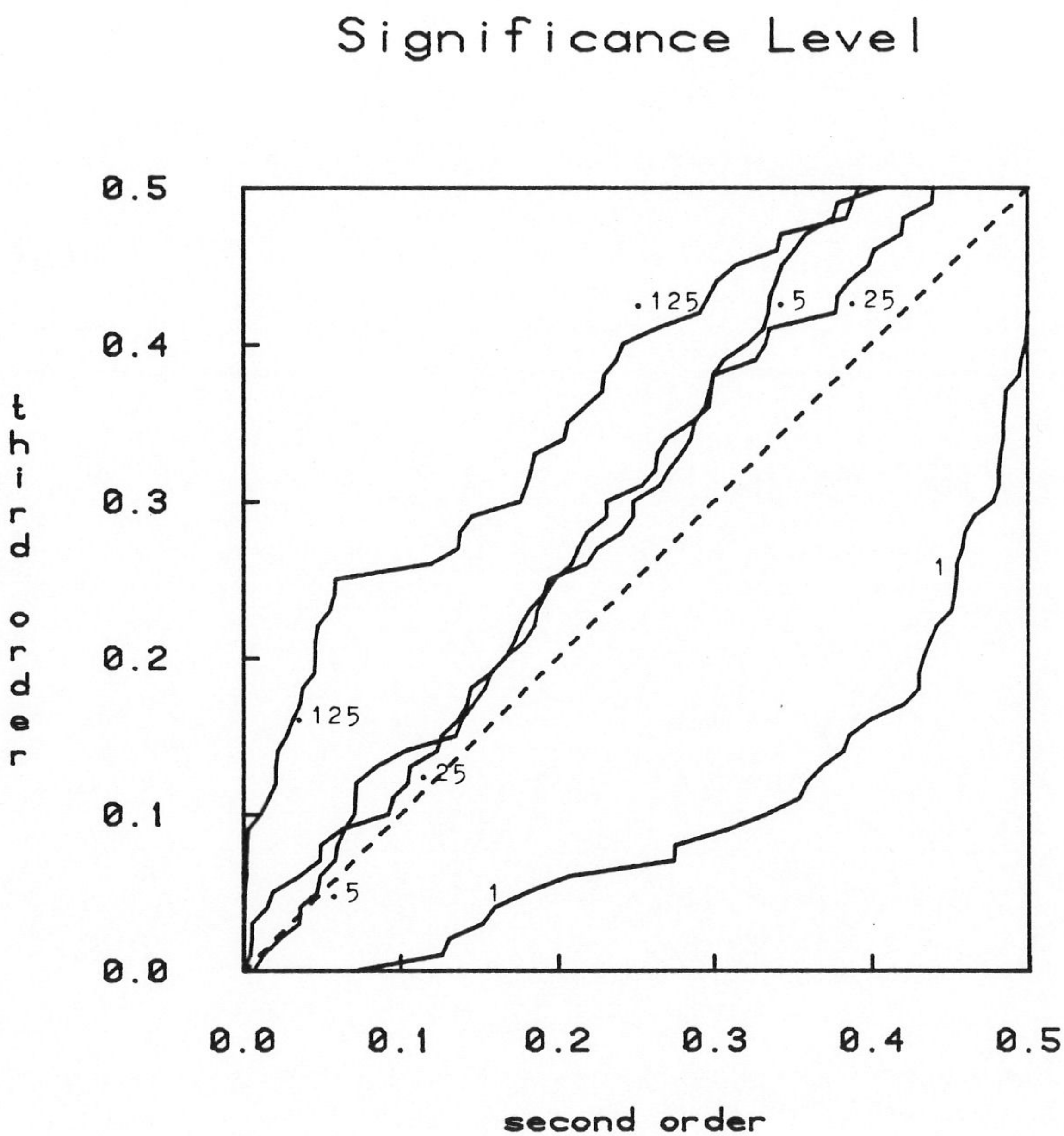

Significance Level
0.5
0.4
0.3
0.2
0.1
0.0
second order
0.0
0.1
0.2
0.3
0.4
0.5
first order
bandwidth varies
by factor of 2 from 0.8 to 0.1

Figure 8(c)



n=200, no censoring
bandwidth scale = 1, .5, .25, .125
density estimates

Figure 8(d)



n=200, no censoring
bandwidth scale = 1, .5, .25, .125
rate estimates

Figure 9

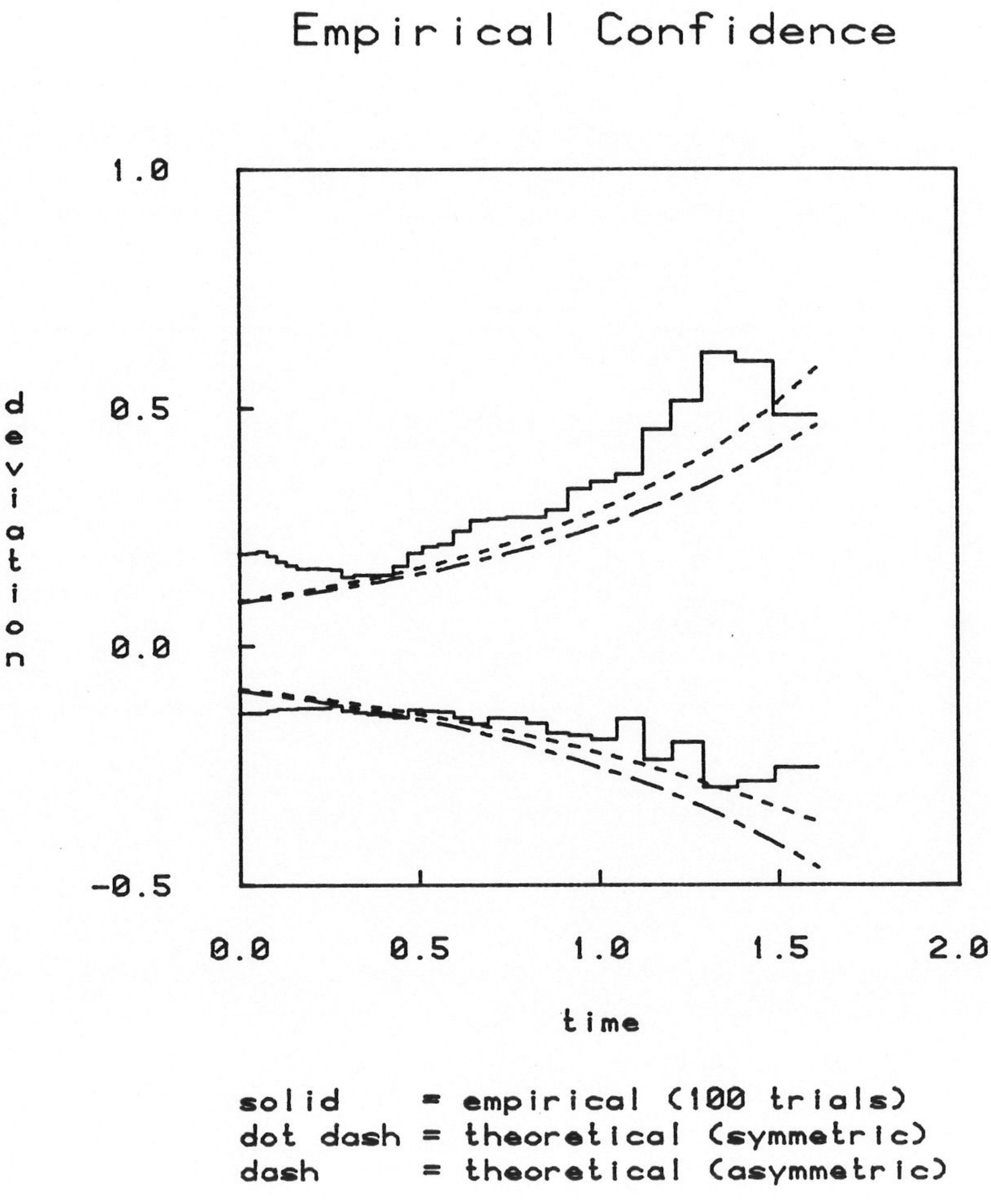

Empirical Confidence
1.0
0.5
0.0
-0.5
deviation
0.0
0.5
1.0
1.5
2.0
time
solid = empirical (100 trials)
dot dash = theoretical (symmetric)
dash = theoretical (asymmetric)

Figure 10(a)

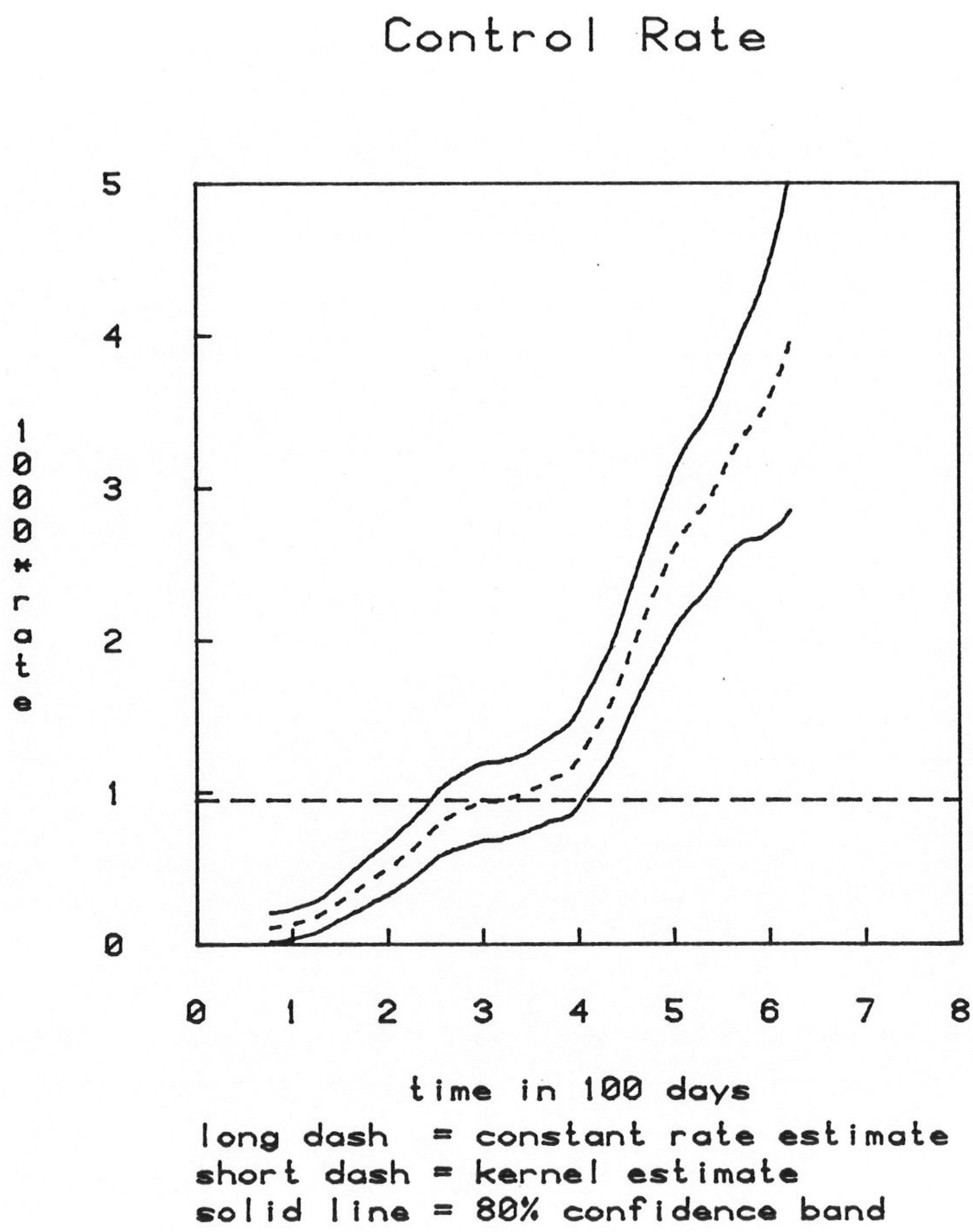

Control Rate
5
4
3
2
1
0
1000*rate
0 1 2 3 4 5 6 7 8
time in 100 days
long dash = constant rate estimate
short dash = kernel estimate
solid line = 80% confidence band

Figure 10(b)

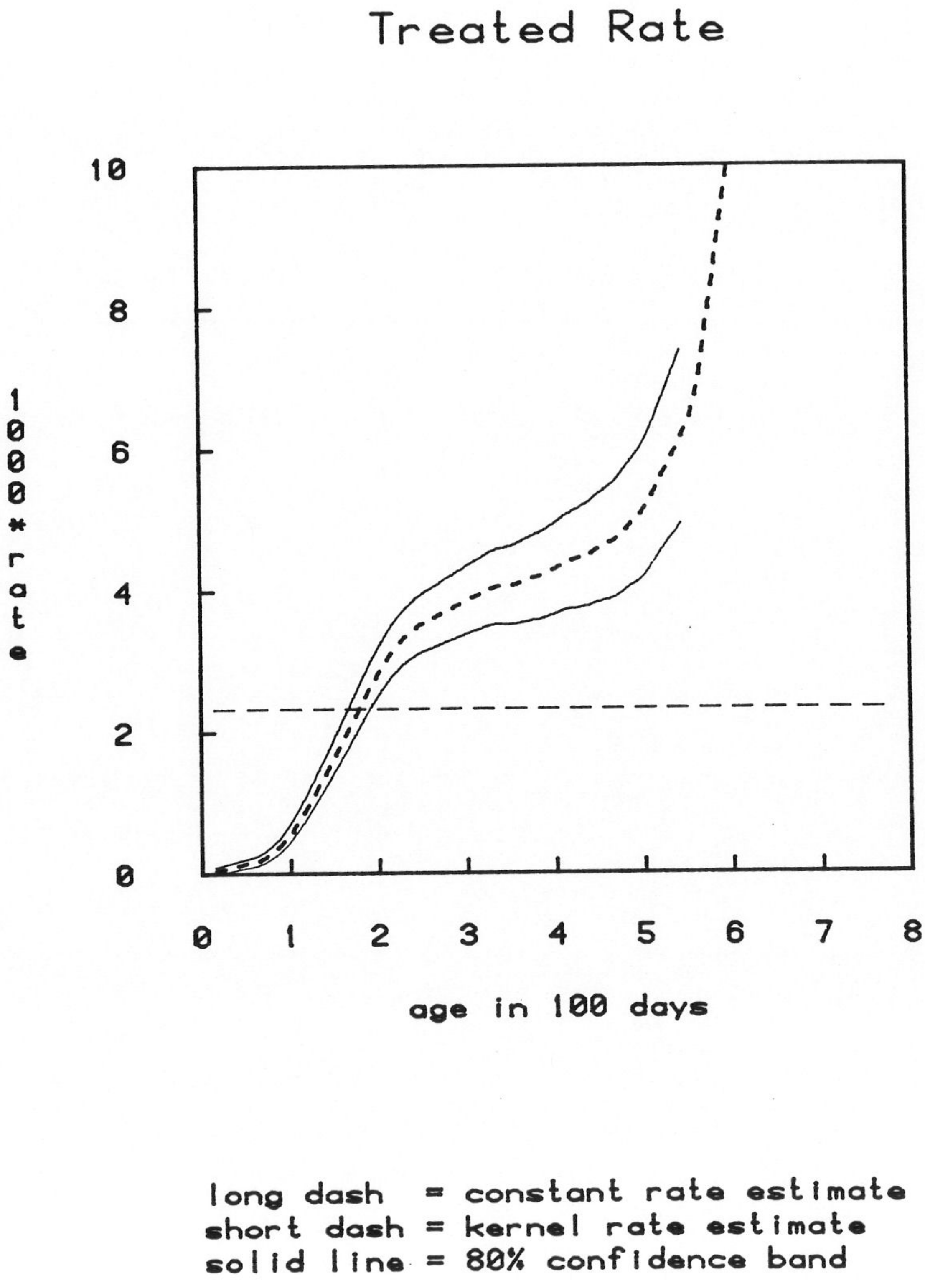

Treated Rate
10
8
6
4
2
0
1000*rate
0 1 2 3 4 5 6 7 8
age in 100 days
long dash = constant rate estimate
short dash = kernel rate estimate
solid line = 80% confidence band

Figure 10(c)

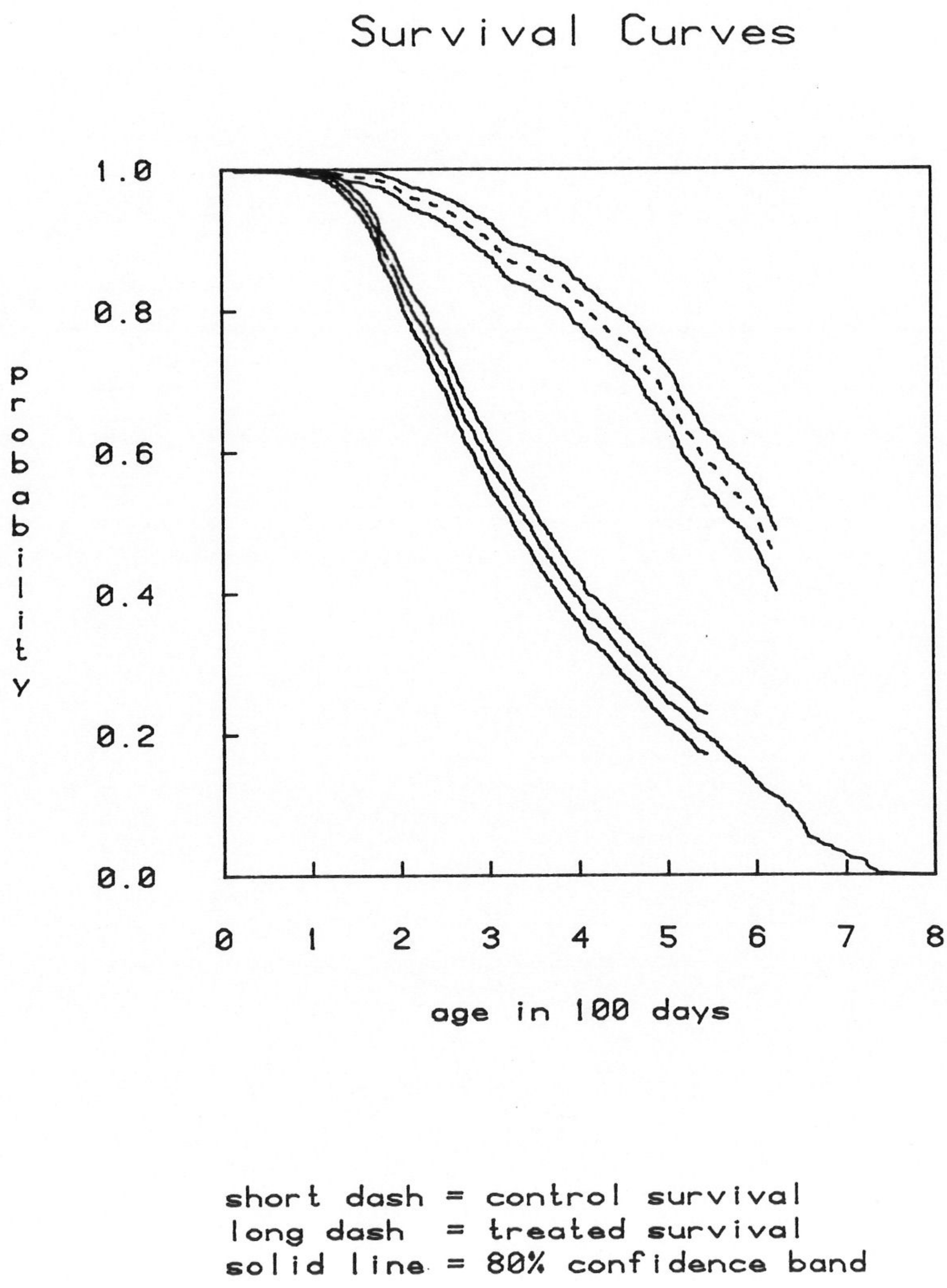

Survival Curves
1.0
0.8
0.6
0.4
0.2
0.0
probability
0
1
2
3
4
5
6
7
8
age in 100 days
short dash = control survival
long dash = treated survival
solid line = 80% confidence band

Figure 11

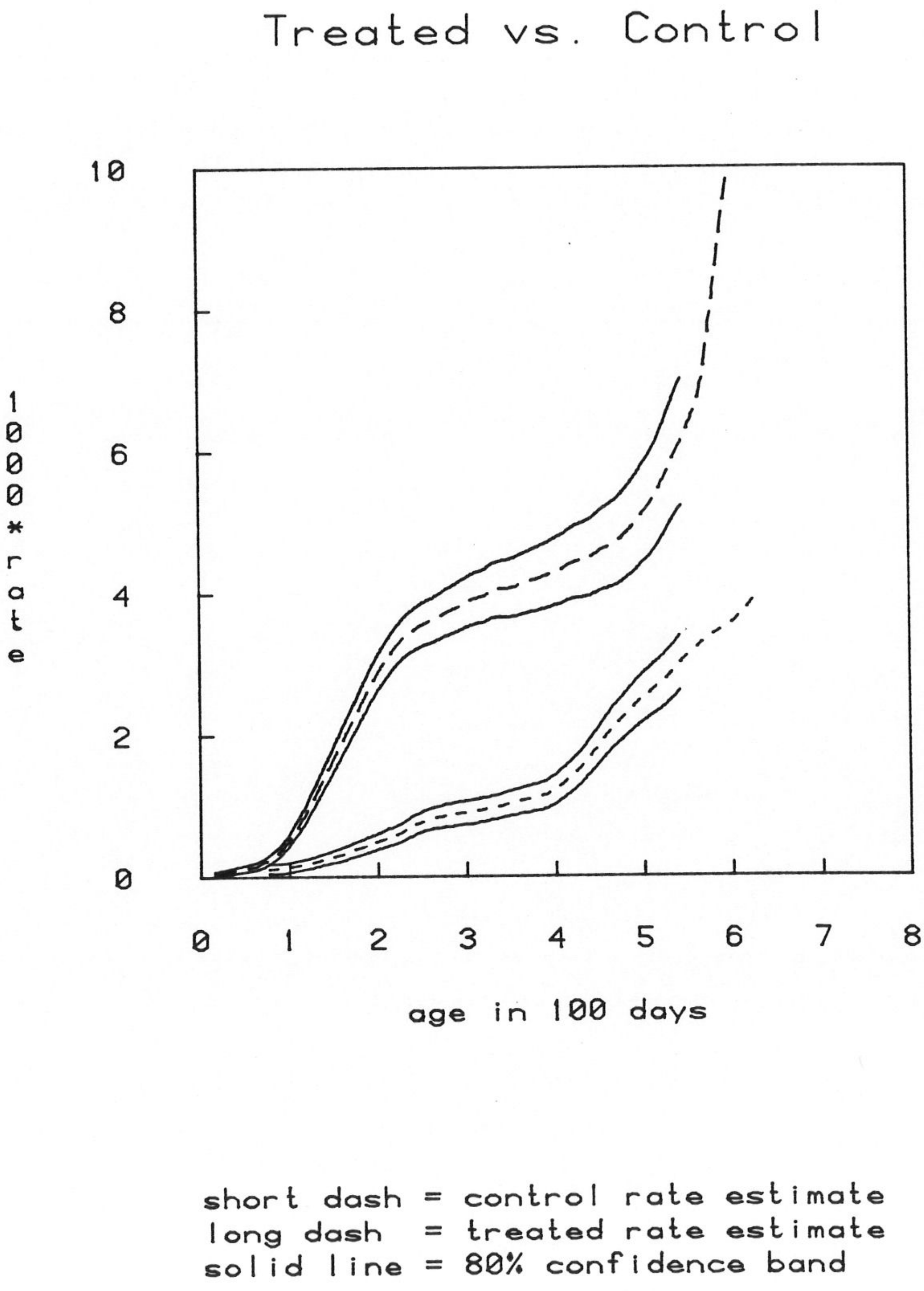

Treated vs. Control
1000*rate
10
8
6
4
2
0
0
1
2
3
4
5
6
7
8
age in 100 days
short dash = control rate estimate
long dash = treated rate estimate
solid line = 80% confidence band

Figure 12(a)

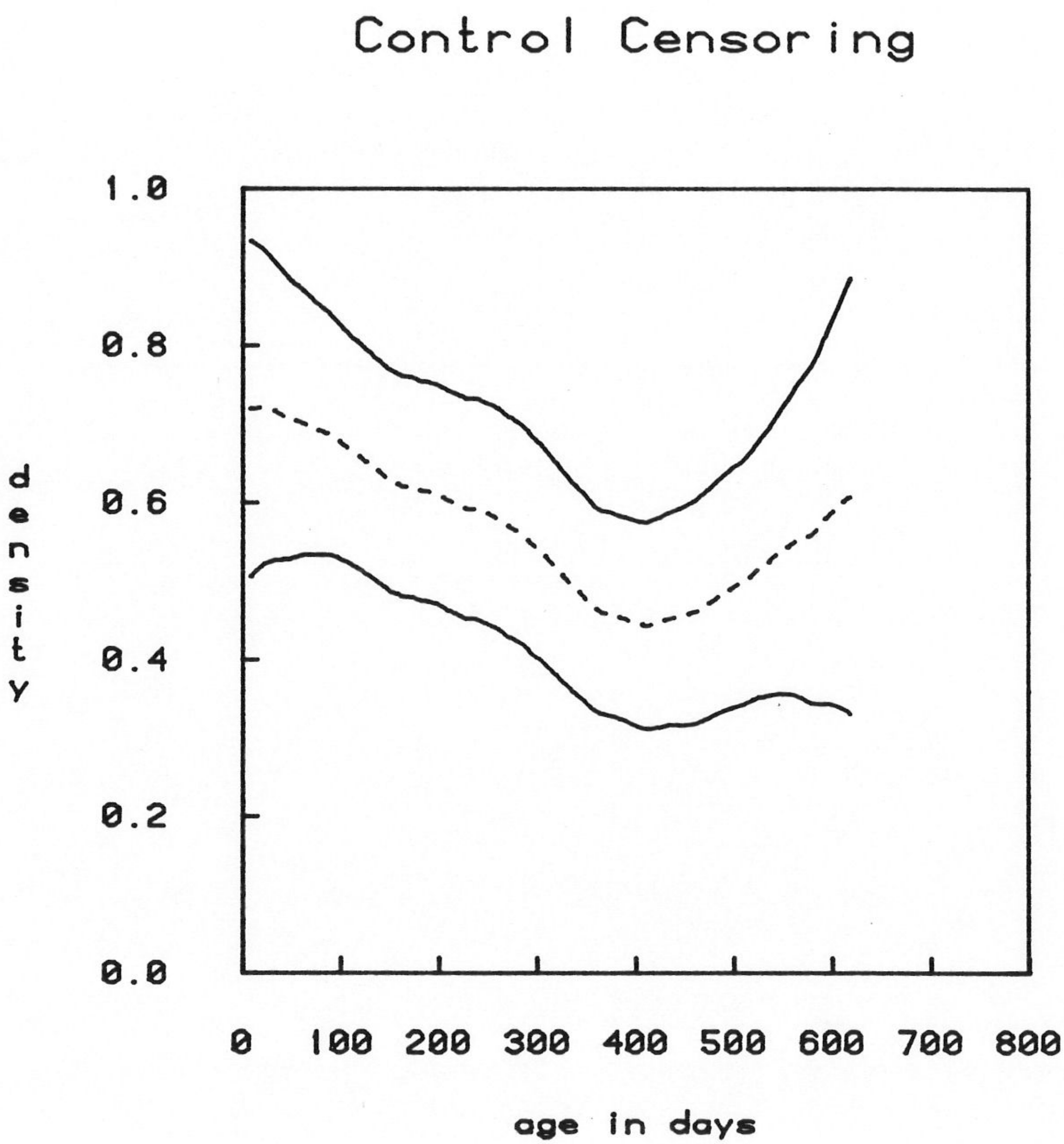
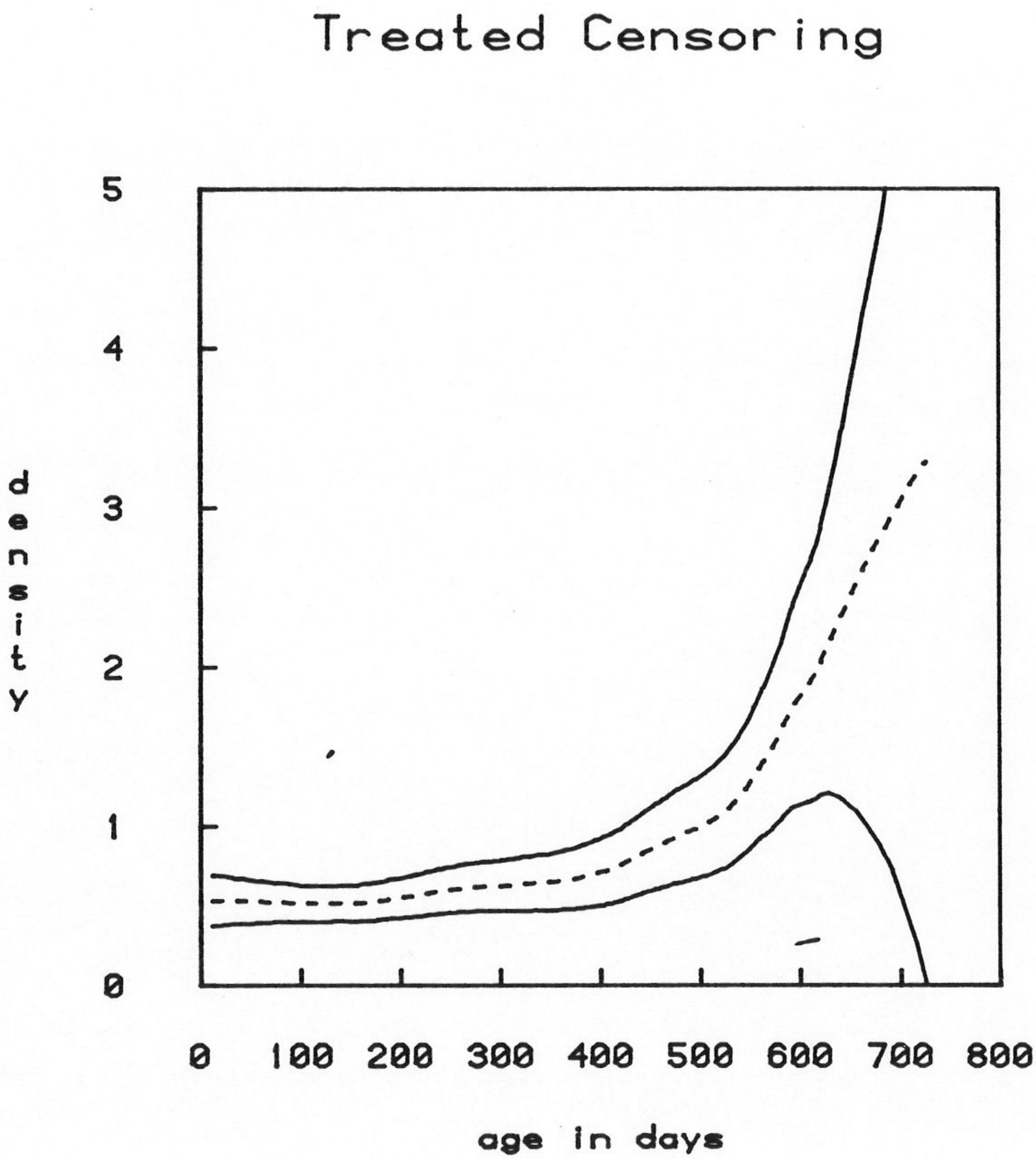

Control Censoring
1.0
0.8
0.6
0.4
0.2
0.0
density
0
100
200
300
400
500
600
700
800
age in days

Figure 12(b)


Treated Censoring
density
0
1
2
3
4
5
0
100
200
300
400
500
600
700
800
age in days