The Annals of Statistics
1981, Vol. 9, No. 3, 628–632

LARGE SAMPLE ESTIMATES AND UNIFORM CONFIDENCE BOUNDS FOR THE FAILURE RATE FUNCTION BASED ON A NAIVE ESTIMATOR

BY J. SETHURAMAN[1] AND NOZER D. SINGPURWALLA[2]

Florida State University and George Washington University

In this paper we consider a naive estimator of the failure rate function and smooth it using any band-limited window. We show that this smoothed estimate is equivalent to estimates obtainable from the sample hazard function, as in Rice and Rosenblatt (1976). We obtain the asymptotic distribution of the global deviation of the smoothed estimate from the failure rate function, which can then be used to construct uniform confidence bands.

1. Introduction and summary. Let $X_1, \cdots, X_n$ be independent and identically distributed random variables with a common distribution function $F(x)$ which has a density function $f(x)$. Assume that $F(0) = 0$, and for x with $\bar{F}(x) = 1 - F(x) > 0$, the hazard function $H(x)$ is defined to be $-\log \bar{F}(x)$ and the failure rate function $r(x)$ is defined to be $f(x)/\bar{F}(x)$, which is also equal to $(d/dx)H(x)$. Let F_n be the empirical df of $X_1, \cdots, X_n$, $\bar{F}_n(x) = 1 - F_n(x)$, and $H_n(x) = -\log \bar{F}_n(x)$, for $x < X_{(n)}$, where $0 = X_{(0)} \leqq X_{(1)} \leqq \cdots \leqq X_{(n)}$ are the order statistics of $X_1, \cdots, X_n$. The purpose of this paper is to provide estimates of $r(x)$ and to obtain the asymptotic distribution of the global deviation of these estimates.

We shall consider here a naive estimate of $r(x)$, originally proposed by Grenander (1956), and also considered by Marshall and Proschan (1965), and by Barlow, Bartholomew, Bremner, and Brunk (1972, Section 5.3). For $X_{(i-1)} \leqq x < X_{(i)}$, $i = 1, 2, \cdots, n$, the *naive estimate* $r_n(x)$ of $r(x)$ is defined as $r_n(x) = [(n - i + 1)\,(X_{(i)} - X_{(i-1)})]^{-1}$; $r_n(x) = 0$, for $x \geqq X_{(n)}$.

Let $G(\alpha; \lambda)$ denote a gamma random variable with density $\alpha^{\lambda} e^{-x\alpha} x^{\lambda-1}/\Gamma(\lambda)$, $x \geqq 0$, and let $G^{-1}(\alpha; \lambda) = 1/G(\alpha; \lambda)$ be the inverse gamma random variable. Then (as pointed out by a referee), part (i) of the following theorem follows from the fact that $(r_n(x))^{-1} \to G(r(x), 2)$ in law, whereas part (ii) of the theorem follows from Watson and Leadbetter (1964, page 110). A detailed proof is given in Sethuraman and Singpurwalla (1978).

THEOREM 1.1
(i) $r_n(x) \to G^{-1}(r(x), 2)$ *in law.*
(ii) *Let* $x_1, \cdots, x_k$ *be distinct. Then* $\{r_n(x_1), \cdots, r_n(x_k)\}$ *are asymptotically independent.*

Thus, $r_n(x)$ is not a consistent estimate of $r(x)$; indeed, $r_n(x)$ has a limiting nondegenerate distribution. While the asymptotic mean of $r_n(x)$ is $r(x)$, the fact that $E(r_n(x))^{-1} \to 2(r(x))^{-1}$ is surprising, and is reminiscent of the interarrival paradox discussed in Feller (1966, Volume 2, page 11). Since $(r_n(x_1), \cdots, r_n(x_k))$ are asymptotically independent, the graph of $\{r_n(x), x \geqq 0\}$ will exhibit wild fluctuations, and therefore deter us from using $r_n(x)$.

Received March, 1980; revised April, 1980.
[1] Supported by the U.S. Army Research Office, Grant DAAG 29-76-G0238.
[2] Supported by the Office of Naval Research, Contract N00014-77-C0263, Project NR 042 372.
AMS 1970 *subject classifications.* 60F05, 60G15, 62G05, 62N05.
Key words and phrases. Failure rate function, consistent estimate, asymptotic uniform confidence bands, maxima of stationary processes.

In Section 2, we propose a smoothed estimate of $\bar{r}_n$ obtained by averaging the naive estimate r_n using a band-limited window. We show that $\bar{r}_n$ can be approximated by an appropriate Gaussian process (Theorem 2.2), and thus obtain the asymptotic distribution of the global deviation on any finite interval. This result can be used to construct confidence bands for the failure rate function.

Rice and Rosenblatt (1976) have proposed three estimates, $h_n^{(1)}$, $h_n^{(2)}$, and $h_n^{(3)}$, of the failure rate function which are nonparametric in nature. They have applied the results of Bickel and Rosenblatt (1973), strengthened by Rosenblatt (1976), for the sample density function to obtain the asymptotic global results for $h_n^{(1)}$. They have also shown that $h_n^{(2)}$ is asymptotically close to $h_n^{(3)}$. By approximating the normalized and centered sample hazard functions H_n by a Wiener process under a monotone transform of time, we obtain, in much the same way as Bickel and Rosenblatt, the asymptotic global results for our smoothed estimate $\bar{r}_n$ and the Rice and Rosenblatt estimator $h_n^{(3)}$. In the course of this proof, we also show that $h_n^{(2)}$ and $h_n^{(3)}$ are uniformly equivalent to our smoothed estimator $\bar{r}_n$ on each finite interval. Other nonparametric estimates of the failure rate function have been studied by Ahmed (1976), Ahmed and Lin (1977), Shaked (1978), and Miller and Singpurwalla (1980).

2. Smoothing of the naive estimate. The behavior of the naive estimate is analogous to that of the periodogram which is used to estimate the spectrum of a stationary time series. We shall therefore use the standard technique of "smoothing" the naive estimate with a "window" $w(u)$, which is a *bounded* ($0 \leqq w(u) \leqq c < \infty$), *band-limited* ($w(u) = 0$, for $|u| \geqq A$), *symmetric function of integral one* to obtain a consistent and asymptotically normal estimate of the failure rate. The smoothed estimator depends on $w(u)$ and a sequence $\{b_n\}$, where $b_n \downarrow 0$ and $nb_n \uparrow \infty$, as $n \to \infty$. Without loss of generality, we may assume that $0 < b_n \leqq 1$.

We define a *smoothed estimator* $\bar{r}_n(x)$ obtained by smoothing r_n with window w and *bandwidth* $2b_nA$ as:

$$\bar{r}_n(x) = \frac{1}{b_n}\int w\left(\frac{x-s}{b_n}\right)r_n(s)ds = \int w(s)r_n(x-b_ns)\,ds. \tag{2.1}$$

Whenever w is a band-limited window, $\bar{r}_n(x)$ will be used as an estimate of $r(x)$ only for $x \geqq b_nA$.

For $x < X_{(n)}$, we may rewrite $\bar{r}_n(x)$ as

$$\bar{r}_n(x) = \frac{1}{b_n}\int w\left(\frac{x-s}{b_n}\right)dR_n(s), \qquad \text{where} \quad R_n(x) = \int_0^x r_n(y)\,dy.$$

Using the elementary inequality $|x + \log(1-x)| \leqq x^2/2(1-x)$ for $0 \leqq x \leqq 1$, we can, using some straightforward steps (see Sethuraman and Singpurwalla, 1978), show that $R_n(x)$ and $H_n(x)$ are uniformly close to each other in bounded intervals. That is, for some $K < X_{(n)}$,

$$\sup_{0\leqq x\leqq K}|R_n(x) - H_n(x)| \leqq 2/3n\bar{F}_n(K), \qquad \text{w.p.1.} \tag{2.2}$$

The three estimates of the failure rate introduced by Rice and Rosenblatt (1976) are:

$$h_n^{(1)}(x) = f_n(x)/\bar{F}_n(x), \quad \text{where} \quad f_n(x) = \frac{1}{b_n}\int w\left(\frac{x-s}{b_n}\right)dF_n(s)$$

$$h_n^{(2)}(x) = \frac{1}{b_n}\int w\left(\frac{x-s}{b_n}\right)\frac{dF_n(s)}{\bar{F}_n(s)};$$

and $h_n^{(3)}(x) = \dfrac{1}{b_n}\displaystyle\int w\left(\frac{x-s}{b_n}\right)dH_n(s).$ Under the additional assumption that

(W1) $$\int |w'(x)|\,dx \leqq c < \infty,$$

where w' is the derivative of w, we can, using (2.2), show that $h_n^{(3)}(x)$ is uniformly close to $\bar{r}_n(x)$ in bounded intervals. That is, whenever $K + A < X_{(n)}$,

(2.3) $$\sup_{0 \leqq x \leqq K} |\bar{r}_n(x) - h_n^{(3)}(x)| \leqq 9C/2nb_n\bar{F}_n(K + A), \qquad \text{w.p.1.}$$

Let $B(x) = F(x)/\bar{F}(x)$, $B'(x)$ the derivative of $B(x)$, and

$$\xi_n(x) = \frac{\sqrt{nb_n}}{\sqrt{B'(x)}}\,(h_n^{(3)}(x) - r_n^*(x)), \qquad b_nA \leqq x \leqq k,$$

where $r_n^*(x) = \frac{1}{b_n}\int w\left(\frac{x-s}{b_n}\right) dH(x)$.

In Theorem 2.2, we show that the process $(\xi_n(x);\ b_nA \leqq x \leqq K)$ is uniformly close to a stationary Gaussian process. Because of (2.3), this result also holds for the process $\{\bar{\xi}_n(x);\ b_nA \leqq x \leqq K\}$, where $\bar{\xi}_n(x)$ is $\xi_n(x)$ with $h_n^{(3)}(x)$ replaced by $\bar{r}_n(x)$. In Theorem 2.3 we obtain the asymptotic distributions of $M_n = \max_{b_nA \leqq x \leqq K} |\xi_n(x)|$ and $\bar{M}_n(x) = \max_{b_nA \leqq x \leqq K} |\bar{\xi}_n(x)|$. In order to use this result for obtaining confidence bounds for $r(x)$, $b_nA \leqq x \leqq K$, we replace $r_n^*(x)$ by $r(x)$, and also replace any unknown quantities that enter in the asymptotic theory by their estimates. This is done in Theorem 2.4. Define $Z_n(x) = \sqrt{n}\,(H_n(x) - H(x))$.

THEOREM 2.1. *There exists a Gaussian process* $\{Z(x);\ 0 \leqq x \leqq K + A\}$ *with mean function zero and* $E(Z(x)Z(y)) = B(x)$ *for* $x \leqq y$, *and such that*

$$\sup_{0 \leqq x \leqq K+A} |Z_n(x) - Z(x)| = O\left(\frac{D \log n}{\sqrt{n}}\right), \qquad \text{w.p.1,}$$

whenever $X_{(n)} > K + A$, *where* D *is a random variable with* $P(D < \infty) = 1$.

PROOF. We use the generic name D for any random variable with $P(D < \infty) = 1$. From Komlós, Major, and Tusnady (1975), there exists a Gaussian process $\{Y(x);\ 0 \leqq x\}$ with mean function zero and $EY(x)Y(y) = F(x)(1 - F(y))$ for $0 \leqq x \leqq y$, and such that

$$\sup_{0 \leqq x} |\sqrt{n}\,(F_n(x) - F(x)) - Y(x)| = O\left(\frac{D \log n}{\sqrt{n}}\right), \qquad \text{w.p.1.}$$

The theorem follows from the above, if we note that for $0 \leqq x \leqq K + A$, $Z_n(x) = -\sqrt{n}\log(\bar{F}_n(x)/\bar{F}(x))$ can be written as $Z_n(x) = Z(x) + D\log n/\sqrt{n}$, where $Z(x) = Y(x)/\bar{F}(x)$ is Gaussian with mean 0 and $E(Z(x)Z(y)) = B(x)$ for $x \leqq y$. □

Let $\{W(s), -\infty < s < \infty\}$ denote a Wiener process, and define $\zeta(\theta) = \int w(\theta - t)\,dW(t)$, $0 \leqq \theta < \infty$. Then $\{\zeta(\theta);\ 0 \leqq \theta < \infty\}$ is a stationary Gaussian process with mean 0 and $E(\zeta(\theta + \delta)\zeta(\theta)) = \int_{-\infty}^{\infty} w(\delta + t)w(t)\,dt = \rho(\delta)$, say. To prove Theorem 2.2 below, we need the following two restrictions on F:

(F1) $B''(x)$ is bounded on $0 \leqq x \leqq K$;

(F2) $\inf_{0 \leqq x \leqq K+A} B'(x) > 0$.

THEOREM 2.2. *Let* (W1), (F1), *and* (F2) *hold. Then there exists a stationary Gaussian process* $\{\zeta_n(\theta);\ A \leqq \theta\ K/b_n\}$ *which is a restriction of* $\{\zeta(\theta)\}$ *for* $A \leqq \theta \leqq K/b_n$ *such that*

$$\sup_{b_nA \leqq x \leqq K} |\xi_n(x) - \zeta_n(b_nx)| = O\left[D\left(\frac{\log n}{\sqrt{nb_n}\,\bar{F}_n(K + A)} + b_n^{1/2}\right)\right], \qquad \text{w.p.1,}$$

whenever $X_{(n)} > K + A$. *The above statement is also true when* $\xi_n(x)$ *is replaced by* $\bar{\xi}_n(x)$.

PROOF. Follows from the steps which are analogous to those of Bickel and Rosenblatt for their Propositions 2.1 and 2.2, and our Theorem 2.1 and Equation (2.3). □

To obtain the asymptotic distributions of $\bar{M}_n$ and $\bar{M}_n$, let $\lambda(w) = \int w^2(t)dt$ and $K_1(w) = (w^2(A) + w^2(-A))/2\lambda(w)$. If $K_1(w) = 0$, we set $K_2(w) = \int (w'(t))^2dt/2\lambda(w)$ and require that

(W2) $$\int (w'(t))^2\,dt < \infty;$$

notice that (W2) ⟹ (W1). Let C_n $(2\log(K/b_n))^{1/2}$ and $\beta_n = C_n/(\lambda(w))^{1/2}$. When w satisfies (W1) and $K_1(w) > 0$, define $\alpha_n = (\lambda(w))^{1/2}[C_n + \log(C_nK_1(w)/\sqrt{2\pi})]/C_n$. When w satisfies (W2) and $K_1(w) = 0$, define $\alpha_n = (\lambda(w))^{1/2}[C_n^2 + \log(K_2(w)/\pi)]/C_n$.

THEOREM 2.3. *Let* (W1) *hold with* $K_1(w) > 0$, *or* (W2) *hold with* $K_1(w) = 0$. *Let* (F1) *and* (F2) *also hold. Then for* $0 < x < \infty$,

$$P\{\beta_n(M_n - \alpha_n) \leqq z\} \to e^{-2e^{-z}}.$$

The above statement is also true when M_n *is replaced by* $\bar{M}_n$.

PROOF. Follows from Theorem 2.2 and the results on the extreme of a stationary Gaussian process with an autocorrelation function $\rho(\theta)$ given in Appendix A of Bickel and Rosenblatt. □

In order to replace $r_n^*(x)$ in $\xi_n(x)$ and $\bar{\xi}_n(x)$ by $r(x)$, we impose an additional restriction on F:

(F3) $r(x)$ is twice continuously differentiable.

Let (F3) hold; then by the uniform continuity of $r''(x)$ on $[0, K + A]$,

$$\sup_{b_nA\leqq x\leqq K} |r_n^*(x) - r(x)| \leqq Lb_n^2,$$

where L is some finite number.

Let the constants b_n satisfy the additional condition

(B1) $$nb_n^5 \log b_n \to 0.$$

THEOREM 2.4. *Under the additional assumption that* (B1) *holds, we may replace* $r_n^*(x)$ *by* $r(x)$ *in the definitions of* $\xi_n(x)$, $\bar{\xi}_n(x)$, M_n, *and* $\bar{M}_n$, *and Theorem 2.3 continues to hold for the new* M_n *and* $\bar{M}_n$.

PROOF. Note that

$$\begin{aligned}
&\beta_n\left[\frac{nb_n}{B'(x)}\right]^{1/2}(h_n^{(3)}(x) - r(x)) \\
&= \beta_n\left[\frac{nb_n}{B'(x)}\right]^{1/2}(h_n^{(3)}(x) - r_n^*(x)) + \beta_n\left[\frac{nb_n}{B'(x)}\right]^{1/2}(r_n^*(x) - r(x)) \\
&= \beta_n\xi_n(x) + O((nb_n^5\log(K/b_n))^{1/2}) = \beta_n\xi(x) + o(1)
\end{aligned}$$

uniformly for $b_n A \leqq x \leqq K$. Thus,

$$\beta_n\left[\sup_{b_nA\leqq x\leqq K}\left(\frac{nb_n}{B'(x)}\right)^{1/2}(h_n^{(3)}(x) - r(x)) - \alpha_n\right] = \beta_n[M_n - \alpha_n] + o(1).$$

This proves the theorem. □

The denominator $B'(x) = r(x)/\bar{F}(x)$ in the definitions of $\xi_n(x)$ and $\bar{\xi}(x)$ is unknown. Since $\bar{r}_n(x)/\bar{F}_n(x)$ is a uniformly consistent estimate of $B'(x)$, we may now obtain $100\alpha\%$ uniform confidence bands for $r(x)$, $b_n A \leqq x \leqq K$, as

$$\bar{r}_n(x) \pm \left(\frac{\bar{r}_n(x)}{nb_n\bar{F}_n(x)}\right)^{1/2}\left(\frac{z}{b_n} + \alpha_n\right),$$

where $z = -\log(-(½) \log \alpha)$.

In a similar manner we can also obtain uniform confidence bounds for $r(x)$ using $h_n^{(3)}(x)$.

Acknowledgment. The authors gratefully acknowledge a referee for pointing out several references and making various detailed comments which have improved this paper.

REFERENCES

AHMAD, I. A. (1976). Uniform strong convergence of the generalized failure rate estimate. *Bull. Math. Statist.* **17** 77–84.

AHMAD, I. and LIN, PI-ERH (1977). Nonparametric estimation of a vector-valued bivariate failure rate. *Ann. Statist.* **5** 1027–1038.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M., and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions.* Wiley, New York.

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

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications*, Vol. II. Wiley, New York.

GRENANDER, U. (1956). On the theory of mortality measurements, Part II. *Skand. Akt.* **39** 125–153.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent rv's and the sample df.I. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete.* **32** 111–131.

MARSHALL, A. W. and PROSCHAN, F. (1965). Maximum likelihood estimates for distributions with monotone failure rates. *Ann. Math. Statist.* **36** 69–77.

MILLER, D. R. and SINGPURWALLA, N. D. (1977). Failure rate estimation using random smoothing. To appear in *Sankhyā*, Ser. B.

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

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

SETHURAMAN, J. S. and SINGPURWALLA, N. D. (1978). Large sample estimates and uniform confidence bounds for the failure rate function based on a naive estimator. Tech. Report No. M472, Florida State Univ. Dept. Statist., Tallahassee.

SHAKED, MOSHE (1978). An estimator for the generalized failure rate function. Tech. Report No. 343, Univ. New Mexico.

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

DEPARTMENT OF STATISTICS
THE FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306

DEPARTMENT OF OPERATIONS RESEARCH,
AND DEPARTMENT OF STATISTICS
THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, D.C. 20052