Probab. Th. Rel. Fields 72, 505–516 (1986)

Probability Theory and Related Fields

The Limiting Distribution of the Maximal Deviation of a Density Estimate and a Hazard Rate Estimate

Regina Y. Liu[1] and John Van Ryzin[2]

[1] Dept. of Statistics, Rutgers University, New Brunswick, NY 08903, USA

[2] Dept. of Statistics, Columbia University, New York, NY 10027, USA

I. Introduction

Let $X_1, X_2, \ldots$ be i.i.d. r.v.'s with d.f. F having a continuous density f on R. Let $k=k_n$ be a sequence of positive integers such that

$$k \to \infty, \quad k/n \to 0 \quad \text{as} \quad n \to \infty. \tag{1.1}$$

A class of histogram type nonparametric estimates of $f(x)$ given by Van Ryzin (1973) has the form

$$f_n(x) = \frac{k/n}{Y_{A_n(x)+k} - Y_{A_n(x)}}, \tag{1.2}$$

where $Y_1 < Y_2 < \ldots < Y_n$ are order statistics of $X_1, \ldots, X_n$ and $\{A_n(x)\}$ is a preassigned sequence of positive-integer valued r.v.'s. For each $n \geqq 1$, $A_n(x)$ is measurable w.r.t. the σ-field generated by $X_1, \ldots, X_n$, and is such that, for any $x \in R$

$$\begin{aligned} &0 < A_n(x) \leqq n-k, \ Y_{A_n(x)} \leqq x \leqq Y_{A_n(x)+k}, && \text{if } Y_1 \leqq x \leqq Y_n,\\ &A_n(x) = 0 && \text{if } Y_1 > x,\\ &A_n(x) = n+1 && \text{if } Y_n < x, \end{aligned}$$

where $Y_0 = -\infty$ and $Y_{n+1} = +\infty$. Therefore,

$$P(0 \leqq A_n(x) \leqq n+1, \ Y_{A_n(x)} \leqq x \leqq Y_{A_n(x)+k}) = 1. \tag{1.3}$$

The local properties of such estimates have been discussed by Van Ryzin (1973) and Kim and Van Ryzin (1975, 1980). Our objective is to obtain global measures of how good $f_n(x)$ is as an estimate of $f(x)$. In particular, the asymptotic distribution of the function $\sup_{x \in J} |f_n(x) - f(x)|/f(x)$, i.e., the maximum of the normalized deviation of the estimate from the true density, for some compact interval J is evaluated under proper conditions as $n \to \infty$.

Bickel and Rosenblatt (1973) considered a kernel density estimate $\hat{f}_n$ and established the asymptotic distribution of the functionals

$$\sup_{0 \leqq x \leqq 1} |\hat{f}_n(x) - f(x)|/(f(x))^{(1/2)} \quad \text{and} \quad \int_0^1 \frac{[\hat{f}_n(x) - f(x)]^2}{f(x)} dx$$

under appropriate conditions, as $n \to \infty$. Mack (1982) also established the asymptotic distribution for $\sup_{x \in J} |f_n^*(x) - f(x)|/f(x)$, where f_n^* is the k_n nearest neighbor density estimate.

The study of the asymptotic distribution of the maximal normalized deviation of the density estimate from its true density over an interval leads to the construction of confidence bands for the true density and are useful for goodness-of-fit tests concerning the unknown density.

As a natural extension, we also give similar results for a hazard rate function estimate based on $f_n(x)$ given in (1.2).

II. Main Results

We define a normalized deviation process based on $f_n(x)$ as follows:

$$Q_n(x) = k^{1/2} f^{-1}(x)(f_n(x) - f(x)). \tag{2.1}$$

The limiting distribution of the process $Q_n(x)$ will be established through the following steps: We first express the order statistics in the denominator of our estimator $f_n(x)$ in terms of two intermediate order statistics, and then adapt the almost sure representation of intermediate order statistics (as later seen in Proposition 1) to those two resulting intermediate order statistics separately. Secondly, we rewrite our deviation process in terms of the empirical processes and then proceed to approximate these empirical processes through the use of related Brownian bridges and Brownian motions.

To facilitate the first step, we now introduce two sets of order statistics on the right and left of the given point $x \in J$, where $J = [a, b]$, $-\infty < a < b < \infty$, and $0 < m \leqq f(x) \leqq M < +\infty$ on $J_\delta = [a - \delta, b + \delta]$ with $\delta > 0$. For fixed $x \in J$, let

$$U_{x,i} = \begin{cases} \infty & \text{if } X_i \leqq x \\ X_i - x & \text{if } X_i > x, \; i = 1, \ldots, n, \end{cases}$$

and let $U_{x,(1)} \leqq \ldots \leqq U_{x,(n)}$ be the order of statistics of $\{U_{x,1}, \ldots, U_{x,n}\}$. Define G_x as the d.f. of $U_{x,i}$, $i = 1, \ldots, n$, i.e.,

$$G_x(t) = P(U_{x,i} \leqq t) = \begin{cases} 0 & \text{if } t \leqq 0 \\ F(x+t) - F(x) & \text{if } 0 < t < \infty. \\ 1 & \text{if } t = \infty \end{cases} \tag{2.2}$$

Let

$$W_{x,i} = \begin{cases} \infty & \text{if } X_i \geqq x \\ x - X_i & \text{if } X_i < x \quad i = 1, \ldots, n, \end{cases}$$

and let $W_{x,(1)} \leqq \dots \leqq W_{x,(n)}$ be the order of statistics of $\{W_{x,1}, \dots, W_{x,n}\}$. Define H_x as the d.f. of $W_{x,i}$, $i=1, \dots, n$, i.e.,

$$H_x(t)=P(W_{x,i} \leqq t)=\begin{cases} 0 & \text{if } t \leqq 0 \\ F(x)-F(x-t) & \text{if } 0<t<\infty. \\ 1 & \text{if } t=\infty \end{cases} \tag{2.3}$$

For the remainder of the paper we let α be a preassigned number, $0 \leqq \alpha \leqq 1$, such that when n is large $A_n(x)+k-R_n(x)=[\alpha k]$, where $R_n(x)=\max\{j\colon Y_j \leqq x\}$ and $[a]=\max\{i\colon i \text{ is an integer and } i \leqq a\}$. Since $R_n(x)=nF_n(x)$,

$$A_n(x)+k=nF_n(x)+[\alpha k] \quad \text{and} \quad A_n(x)=nF_n(x)-(k-[\alpha k]). \tag{2.4}$$

For simplicity we use $A=A_n(x)$ when there is no confusion.

Before investigating $Q_n(x)$, we begin with investigation of a slightly modified process $\hat{Q}_n(x)$,

$$\hat{Q}_n(x)=k^{1/2} f^{-1}(x) \left(f_n(x)-\frac{k/n}{c_n(x)+d_n(x)}\right), \tag{2.5}$$

where $c_n(x)$ and $d_n(x)$ are defined implicitly by

$$G_x(c_n(x))=[\alpha k]/n \quad \text{and} \quad H_x(d_n(x))=(k-[\alpha k])/n \tag{2.6}$$

respectively. Note that $\hat{Q}_n(x)$ can be rewritten as

$$\hat{Q}_n(x)=k^{1/2} f^{-1}(x) \left(\frac{k/n}{Y_{A+k}-Y_A}-\frac{k/n}{c_n(x)+d_n(x)}\right). \tag{2.7}$$

The following theorem gives the asymptotic distribution of the deviation process $\hat{Q}_n(x)$ defined in (2.5).

Theorem 1. *Let $J=[a,b]$, $a<b$, and $J_\delta=[a-\delta, b+\delta]$, $\delta>0$. Assume f' exists and is bounded on J_δ, and let $0<m \leqq f(x) \leqq M<\infty$ on J_δ. For fixed $x \in J$, let $\{A_n(x)\}$ satisfy* (1.3) *and* (2.4) *and let k satisfy*

(i) *$k/n \to 0$ and $k \to \infty$ as $n \to \infty$,*

(ii) $(\log n)^3 \left(\log \frac{n}{k}\right)^2 = o(k)$ *and* (2.8)

(iii) $k^{5/4}=o(n(\log n)^{1/4})$.

Then,

$$\lim_{n\to\infty} P\{a_n[\sup_J |\hat{Q}_n(x)|-b_n] \leqq \lambda\}=\exp\{-2e^{-\lambda}\},$$

where

$$a_n=\left\{2\log\left(\frac{n}{k} P(X \in J)\right)\right\}^{1/2},$$

and

$$b_n=a_n+(2a_n)^{-1}\left\{\log\log\left(\frac{n}{k} P(X \in J)\right)-\log \pi\right\}. \tag{2.9}$$

As an immediate corollary we have

Theorem 2. *Let k satisfy* (a) $k/n \to 0$ *and* $k \to \infty$ *as* $n \to \infty$

(b) $(\log n)^5 = o(k)$ *and assume further that*

(c) $k\left(\log \frac{n}{k}\right)^{1/5} = o(n^{1/4})$ *if* $\alpha = 1/2$

(c′) $k\left(\log \frac{n}{k}\right)^{1/3} = o(n^{4/5})$ *if* $\alpha \neq 1/2$.

Assume also that f' and f'' exist and are bounded on J_δ and $0 < m \leqq f(x) \leqq M < +\infty$ on J_δ. Then $\hat{Q}_n(x)$ in Theorem 1 can be replaced by $Q_n(x)$ in (2.1) *and more explicitly*

$$\lim_{n\to\infty} P\left\{ f_n(x) - \frac{f_n(x)}{\sqrt{k}}\left(\frac{\lambda}{a_n} + b_n\right) \leqq f(x) \leqq f_n(x) + \frac{f_n(x)}{\sqrt{k}}\left(\frac{\lambda}{a_n} + b_n\right), \text{ for all } x \in J \right\}$$
$$= \exp\{-2e^{-\lambda}\}.$$

III. Proofs

For the second step in proving Theorem 1 we investigate some related processes which arise naturally and are defined here. It will be shown step by step that $\hat{Q}_n(x)$ can be properly approximated uniformly on J through this series of processes.

Let F_n be the empirical d.f. based on X_i's and D_n the corresponding empirical process, i.e., $D_n(x) = \sqrt{n}(F_n(x) - F(x))$. Also let $D_n^*(x) = D_n(F^{-1}(x))$, i.e., $D_n^*(x) = \sqrt{n}(F_n(F^{-1}(x)) - x)$. We now define

$${}_1Q_n(x) = \tau_n(x)\left(\frac{n}{k}\right)^{1/2}[D_n(x + c_n(x)) - D_n(x - d_n(x))],$$

where

$$\tau_n(x) = \left(\frac{k}{n}\right)^2 (c_n(x) + d_n(x))^{-2}[G_x'(c_n(x))H_x'(d_n(x))]^{-1}, \tag{3.1}$$

$${}_2Q_n(x) = \left(\frac{n}{k}\right)^{1/2}[D_n^*(F(x + c_n(x))) - D_n^*(F(x - d_n x)))], \tag{3.2}$$

$${}_3Q_n(x) = \left(\frac{n}{k}\right)^{1/2}\left[D_n^*\left(F(x) + \alpha\frac{k}{n}\right) - D_n^*\left(F(x) - (1-\alpha)\frac{k}{n}\right)\right] \tag{3.3}$$

$${}_4Q_n(x) = \left(\frac{n}{k}\right)^{1/2}\left[B_n^0\left(F(x) + \alpha\frac{k}{n}\right) - B_n^0\left(F(x) - (1-\alpha)\frac{k}{n}\right)\right], \tag{3.4}$$

where B_n^0 is a sequence of Brownian bridges used in the proof of Proposition 2 below, and

$${}_5Q_n(x) = \left(\frac{n}{k}\right)^{1/2}\left[B_n\left(F(x) + \alpha\frac{k}{n}\right) - B_n\left(F(x) - (1-\alpha)\frac{k}{n}\right)\right], \tag{3.5}$$

where B_n is a sequence of Brownian motions, i.e., $B_n^0(t) \overset{d}{=} B_n(t) - tB_n(1)$, for $0 \leqq t \leqq 1$. Here $\overset{\text{"}d\text{"}}{=}$ means equivalent in distribution.

The following results quoted as propositions from the literature are needed in the proof of Theorem 1.

Proposition 1 (Watts, 1980). *Let X_n be i.i.d. random variables defined on the probability space (Ω, A, P) with d.f. F and assume F has a finite left end point x_0 such that $F(x_0)=0$. Suppose in an interval $(x_0, x_0+\delta)$, $\delta>0$, F is twice differentiable with f'' bounded and $\lim_{x\downarrow x_0} F'(x)$ exists and is positive. For sufficiently large n, we define x_n by $F(x_n)=k_n/n$, where k_n satisfies* (i) $k_n\to\infty$, (ii) $k_n/n\to 0$, *and* (iii) $k_n/\log^3 n\to\infty$ *as* $n\to\infty$. *Then for almost all* $\omega\in\Omega$,

$$X_{n,(k_n)}(\omega)=x_n+\frac{F(x_n)-F_n(x_n,\omega)}{F'(x_n)}+R_n(\omega),$$

where $R_n(\omega)=O(n^{-1}k^{1/4}\log^{3/4} n)$ as $n\to\infty$ w.p.1, and $X_{n,(k_n)}$ is the k_nth order statistic of $X_1, \ldots, X_n$.

Proposition 2 (Komlos, Major and Tusnady, 1975). *Given independent r.v.'s $U_1, U_2, \ldots, U_n$ uniformly distributed on* $(0,1)$ *with corresponding empirical d.f. H_n, there exists a sequence of Brownian bridges $\{B_n^0(t);\ 0\leqq t\leqq 1\}$ such that for all n*

$$\sup_{0\leqq t\leqq 1} |\gamma_n(t)-B_n^0(t)|=O(n^{-1/2}\log n) \qquad \text{w.p.1},$$

where $\gamma_n(t)=\sqrt{n}(H_n(t)-t)$, $0\leqq t\leqq 1$.

Proposition 3 (Stute, 1982). *Suppose β_n is a sequence of numbers satisfying*

(i) $\beta_n\downarrow 0$,

(ii) $\log\dfrac{1}{\beta_n}=o(n\beta_n)$ *and*

(iii) $\log\log n=o\left(\log\dfrac{1}{\beta_n}\right)$.

Then, $\lim_{n\to\infty}\left(2\beta_n\log\dfrac{1}{\beta_n}\right)^{-1/2}\sup_{|t-u|\leqq\beta_n}|D_n^*(t)-D_n^*(u)|=1$, *w.p.1.*

Proposition 4 (Bickel and Rosenblatt, 1973, Theorem A.1). *Let $Y_T(\cdot)$ be a sequence of separable Gaussian processes with mean $\mu_T(\cdot)$ such that $Y_T(\cdot)-\mu_T(\cdot)$ is stationary. Let $r(\cdot)$ be the covariance function of Y_T,*

$$M_T=\max\{Y_T(t)\colon 0\leqq t\leqq T\}, \qquad m_T=\min\{Y_T(t)\colon 0\leqq t\leqq T\}.$$

Let $b_T(t)=\mu_T(t)\,(2\log T)^{1/2}$. Suppose that:

(i) *$b_T(t)$ is uniformly bounded in t and T on $[0, T]$ as $T\to\infty$.*

(ii) *$b_T(t)\to b(t)$ uniformly on $[0, T]$ as $T\to\infty$.*

(iii) *$T^{-1}\lambda[t\colon b(t)\leqq x,\ 0\leqq t\leqq T]\to\eta(t)$(=the d.f. of a probability measure) as $T\to\infty$. (λ denotes Lebesgue measure.)*

(iv) *$b(\cdot)$ is uniformly continuous on R.*

(v) *$r(t)=1-c|t|^\beta+0(|t|^\beta)$, $0<\beta\leqq 2$, as $t\to 0$.*

(vi) $\int_0^\infty r^2(t)<\infty$.

Let

$$B(t)=(2\log t)^{1/2}+\frac{1}{(2\log t)^{1/2}}\left\{\left(\frac{1}{\beta}-\frac{1}{2}\right)\log\log t+\log((2\pi)^{-1}(c^{-1/2}H_\beta 2^{(2-\beta)/2})\right\},$$

where $H_\beta=\lim_{T\to\infty}\frac{1}{T}\int_0^\infty e^s P(\sup_{0\leqq t\leqq T} Y(t)>s)ds$ *and* Y *is a Gaussian process with,*

$$E(Y(t))=-|t|^\beta,\ \operatorname{cov}(Y(t_1),\ Y(t_2))=|t_1|^\beta+|t_2|^\beta-|t_1-t_2|^\beta.$$

Then,

$$U_T=(2\log T)^{1/2}(M_T-B(T)) \quad and \quad V_T=-(2\log T)^{1/2}(m_T+B(T))$$

are asymptotically independent with

$$P(U_T<z)\to e^{-\lambda_1 e^{-z}}, \qquad P(V_T<z)\to e^{-\lambda_2 e^{-z}},$$

where $\lambda_1=\int e^z d\eta(z)$ *and* $\lambda_2=\int e^{-z}d\eta(z)$.

In addition to the above propositions, the proof of Theorem 1 will be facilitated by the following set of lemmas which we prove now.

Lemma 1. *Suppose that the conditions of Theorem 1 hold, then* $\hat{Q}_n(x)={}_1Q_n(x)+O(k^{-1/4}(\log n)^{3/4})$ *w.p.1 uniformly on* J.

Proof. Observe that by using the mean value theorem repeatedly and by the definition of G_x and H_x in (2.2), (2.3) and (2.6), we have

$$[\alpha k]/n=G_x(c_n(x))=F(x+c_n(x))-F(x)=c_n(x)f(x_{1n}) \tag{3.6}$$

for some x_{1n}, $x<x_{1n}<x+c_n(x)$,

$$(k-[\alpha k])/n=H_x(d_n(x))=F(x)-F(x-d_n(x))=d_n(x)f(x_{2n}) \tag{3.7}$$

for some x_{2n}, $x-d_n(x)<x_{2n}<x$,

$$G_x'(c_n(x))=f(x+c_n(x))=f(x)+c_n(x)f'(x_n^*) \tag{3.8}$$

for some x_n^*, $x<x_n^*<x+c_n(x)$, and

$$H_x'(d_n(x))=f(x-d_n(x))=f(x)-d_n(x)f'(x_n^{**}) \tag{3.9}$$

for some x_n^{**}, $x-d_n(x)<x_n^{**}<x$.

Hence if $0<m\leqq f(x)\leqq M<\infty$ and $|f'(x)|\leqq N<\infty$, we see that

$$c_n(x)=[\alpha k]/(nf(x_{1n}))\sim\frac{k}{n}, \tag{3.10}$$

$$d_n(x)=(k-[\alpha k])/(nf(x_{2n}))\sim\frac{k}{n}, \tag{3.11}$$

where $\alpha_n\sim\beta_n$ means $0<c\leqq\left|\frac{\alpha_n}{\beta_n}\right|\leqq d<\infty$, for n sufficiently large.

Hence,

$$G'_x(c_n(x)) = f(x) + O\left(\frac{k}{n}\right), \tag{3.12}$$

and

$$H'_x(d_n(x)) = f(x) + O\left(\frac{k}{n}\right). \tag{3.13}$$

By Taylor's expansion we have

$$\hat{Q}_n(x) = \frac{k^{3/2}}{nf(x)(c_n(x)+d_n(x))^2}\left\{[c_n(x)+d_n(x)-(Y_{A+k}-Y_A)] + O\left(\frac{[c_n(x)+d_n(x)-(Y_{A+k}-Y_A)]^2}{c_n(x)+d_n(x)}\right)\right\}. \tag{3.14}$$

Observing that $Y_{A+k} - Y_A = U_{x,([\alpha k])} + W_{x,(k-[\alpha k])}$ and applying Proposition 1 to $U_{x,([\alpha k])}$ and $W_{x,(k-[\alpha k])}$, we obtain

$$\begin{aligned} c_n(x) &+ d_n(x) - (Y_{A+k} - Y_A) \\ &= \frac{G_{x,n}(c_n(x)) - G_x(c_n(x))}{G'_x(c_n(x))} + \frac{H_{x,n}(d_n(x)) = H_x(d_n(x))}{H'_x(d_n(x))} \\ &\quad + O(n^{-1}k^{1/4}(\log n)^{3/4}), \quad \text{w.p.1, uniformly on } J. \end{aligned} \tag{3.15}$$

Combine the first two terms on the right hand side of (3.15) with common denominator $G'_x(c_n(x))\, H'_x(d_n(x))$ and apply (3.8), (3.9) to $G'_x(c_n(x))$ and $H'_x(d_n(x))$ in the resulting numerator, and then apply Proposition 3 to obtain

$$\begin{aligned} c_n(x) &+ d_n(x) - (Y_{A+k} - Y_A) \\ &= f(x)[G'_x(c_n(x))H'_x(d_n(x))]^{-1}[F_n(x+c_n(x)) - F(x+c_n(x)) \\ &\quad - F_n(x-d_n(x)) + F(x-d_n(x))] + O(n^{-1}k^{1/4}(\log n)^{3/4}), \end{aligned} \tag{3.16}$$

w.p.1, uniformly on J, provided $k^{5/4} = O(n(\log n)^{1/4})$.

Applying Proposition 3 again to the first term of (3.16) we have

$$O\left(\frac{[c_n(x)+d_n(x)-(Y_{A+k}-Y_A)]^2}{c_n(x)+d_n(x)}\right) = O\left(n^{-1}\log\frac{n}{k}\right). \tag{3.17}$$

w.p.1, uniformly on J. The result follows by applying Eqs. (3.16) and (3.17) to (3.14).

Lemma 2. *If the conditions of Theorem 1 hold, then* $a_n|\tau_n^{-1}(x) - 1| \to 0$ *uniformly on* J *as* $n \to \infty$.

Proof. By applying Taylor's expansion on G_x^{-1} and H_x^{-1} respectively we have

$$c_n(x) = [\alpha k]/(nf(x)) + O[(k/n)^2] \quad \text{and} \quad d_n(x) = (k - [\alpha k])/(nf(x)) + O[(k/n)^2]$$

Together with (3.12) and (3.13) we obtain

$$\begin{aligned} a_n|\tau_n^{-1}(x)-1| &= \left(2\log c\frac{n}{k}\right)^{1/2}\left(\frac{n}{k}\right)^2\left|(c_n(x)+d_n(x))^2(G_x'(c_n(x))H_x'(d_n(x))-\left(\frac{k}{n}\right)^2\right| \\ &= \left(2\log c\frac{n}{k}\right)^{1/2}\left(\frac{n}{k}\right)^2\left|\left[\frac{k/n}{f(x)}+O\left(\frac{k}{n}\right)^2\right]^2\left(f^2(x)+O\left(\frac{k}{n}\right)\right)-\left(\frac{k}{n}\right)^2\right| \\ &= O\left[\frac{k}{n}\left(\log\frac{n}{k}\right)^{1/2}\right], \end{aligned}$$

which tends to zero uniformly on J under the conditions assumed.

Lemma 3. *(i)* $F(x+c_n(x))=F(x)+\alpha\frac{k}{n}+O\left(\left(\frac{k}{n}\right)^2\right)$ *and* (ii) $F(x-d_n(x))=F(x)-(1-\alpha)\frac{k}{n}+O\left(\left(\frac{k}{n}\right)^2\right)$, *uniformly on J.*

Proof. The desired conclusions follow immediately from Taylor's expansion and the assumed conditions on J.

Lemma 4. *For n large enough,*

$$\sup_J |{}_4Q_n(x)| \stackrel{d}{=} \sup_{0\le t\le \frac{n}{k}(\|(b)-\|(a))}\left|{}_5Q_n\left(F^{-1}\left(\frac{k}{n}t\right)\right)\right|+O_p\left(\left(\frac{k}{n}\right)^{1/2}\right).$$

Proof. Note that

$${}_5Q_n(x)=\left(\frac{n}{k}\right)^{1/2}\int_{-\infty}^{\infty} I\left(F(x)-(1-\alpha)\frac{k}{n}<s<F(x)+\alpha\frac{k}{n}\right)dB_n(s),$$

where the last integral can be interpreted in the I to sense. Let $v=\frac{n}{k}(F(t))$ $\left(\text{i.e., } t=F^{-1}\left(\frac{k}{n}v\right)\right)$ for $t\in[a,b]$, and let $V_n(v)=\left(\frac{n}{k}\right)^{1/2}\int_{-\infty}^{\infty} I_{\left(v-(1-\alpha)<\frac{n}{k}s<v+\alpha\right)}dB_n(s)$.

Due to Brownian scaling, in this case it is $y=\frac{n}{k}s$,

$$V_n(\lambda)\stackrel{d}{=}\int_{-\infty}^{\infty} I_{(v-(1-\alpha)<y<v+\alpha)}dB_n(y);\quad \frac{n}{k}F(a)\le v\le\frac{n}{k}F(b).$$

Since, if $v_1<v_2$,

$$\begin{aligned} \operatorname{Cov}(V_n(v_1),V_n(v_2)) &= E\,V_n(v_1)V_n(v_2) \\ &= \int_{-\infty}^{\infty} I_{(v_1-(1-\alpha)<y<v_1+\alpha)}\cdot I_{(v_2-(1-\alpha)<y<v_2+\alpha)}dy \\ &= \begin{cases} 1-(v_2-v_1) & \text{if } v_2-v_1\le 1 \\ 0 & \text{if } v_2-v_1>1, \end{cases} \end{aligned}$$

we conclude that the covariance function of the process

$$\left\{V_n(v): \frac{n}{k}F(a)\le v\le\frac{n}{k}F(b)\right\}$$

is

$$\gamma(u)=\begin{cases}1-|u| & \text{if } |u|\leqq 1\\ 0 & \text{if } |u|>1,\end{cases}$$

indicating stationarity. Thus, together with the fact that

$${}_4Q_n(x)\overset{d}{=}{}_5Q_n(x)-\left(\frac{k}{n}\right)^{1/2}B_n(1),$$

we have the result.

We now return to the proof of Theorem 1.

Proof of Theorem 1. The proof will be completed by examining the error incurred by approximating $\hat{Q}_n$, defined in (2.5), successively by ${}_1Q_n$ to ${}_5Q_n$, defined in (3.1) to (3.5). Observe the following:

(i) By Lemma 1, we have

$$\sup_J |\hat{Q}_n(x)|=\sup_J |{}_1Q_n(x)|+O(k^{-1/4}(\log n)^{3/4}) \quad w.p.1.$$

Note also that $a_n O(k^{-1/4}(\log n)^{3/4}\to 0$ as $n\to\infty$ is implied by condition (iii) in (2.8).

(ii) By the definition of D_n^* and D_n, we have ${}_2Q_n(x)=\tau_n^{-1}(x){}_1Q_n(x)$ and the error due to approximating ${}_1Q_n$ by ${}_2Q_n$ is taken care of by Lemma 2.

(iii) Proposition 3 and Lemma 3 give

$$\sup_J \left|D_n^*\left(F(x+c_n(x))-D_n^*\left(F(x)+\alpha\frac{k}{n}\right)\right)\right|=O\left(\frac{k}{n}\left(\log\frac{n}{k}\right)^{1/2}\right) \quad \text{w.p.1},$$

and

$$\sup_J \left|D_n^*\left(F(x-d_n(x))-D_n^*\left(F(x)-(1-\alpha)\frac{k}{n}\right)\right)\right|=O\left(\frac{k}{n}\left(\log\frac{n}{k}\right)^{1/2}\right) \quad \text{w.p.1.}$$

Hence,

$$\sup_J |{}_2Q_n(x)-{}_3Q_n(x)|=O\left(\left(\frac{k}{n}\log\frac{n}{k}\right)^{1/2}\right) \quad \text{w.p.1.}$$

(iv) By Proposition 2, we have

$$\sup_J |{}_3Q_n(x)-{}_4Q_n(x)|=O(k^{-1/2}\log n) \quad \text{w.p.1.}$$

Therefore, by Lemma 4 and by applying Proposition 4 to ${}_5Q_n\left(F^{-1}\left(\frac{k}{n}(t)\right)\right)$ in Lemma 4 with $T=\frac{n}{k}(F(b)-F(a))$, $c=1$, $\beta=1$ and $H_1=1$ (given by Pickands 1969), the proof is completed.

Proof of Theorem 2. Note that the result is determined solely on the order of the magnitude of $\sup_J\left|\dfrac{k/n}{c_n(x)+d_n(x)}-f(x)\right|$ which in turn, by applying Taylor's expansion on G_x^{-1} and H_x^{-1} w.r.t 0 respectively, is equivalent to the order of $\frac{[\alpha k]^2}{n^2}-\frac{(k-[\alpha k])^2}{n^2}$. Observe that

$$\frac{[\alpha k]^2}{n^2}-\frac{(k-[\alpha k])^2}{n^2}=\begin{cases} O\left(\frac{k^2}{n^2}\right) & \text{for } 0\leqq\alpha\leqq 1 \\ O\left(\frac{k}{n^2}\right) & \text{for } \alpha=1/2. \end{cases} \tag{3.18}$$

Thus, for any α, $0\leqq\alpha\leqq 1$

$$a_n \sup_J \frac{k^{1/2}}{f(x)}\left|\frac{k/n}{c_n(x)+d_n(x)}-f(x)\right|=O\left(\left(\log\frac{n}{k}\right)^{1/2}\frac{k^{3/2}}{n}\right) \tag{3.19}$$

from which the result follows if $\alpha\neq 1/2$ under condition (c'). When $\alpha=1/2$, we consider separately the cases $k\leqq n^{1/2}$ and $k>n^{1/2}$. If $k\leqq n^{1/2}$, the left hand side of (3.19) has order $O((\log n)^{1/2}n^{-1/4})$ which tends to zero. If $k>n^{1/2}$, we have, by (3.18), the left hand side of (3.19) is

$$O\left(\frac{k^{5/2}}{2}\left(\log\frac{n}{k}\right)^{1/2}\left(\frac{n}{k^2}\right)\right)<O\left(\frac{k^{5/2}}{n^2}\left(\log\frac{n}{k}\right)^{1/2}\right.$$

which tends to zero when (c) is satisfied.

Remark 1. When α is chosen to be 1/2, it is the symmetric case mentioned in Van Ryzin (1973). In this case Theorem 2 of Kim and Van Ryzin (1980) establishes that with $k^5/n^4\to d$ as $n\to\infty$, where $d\geqq 0$, the estimate has a better rate of convergence than if $\alpha\neq 1/2$. Similarly in our Theorem 2, if we assume f'' exists and is bounded on J_δ and $\alpha=1/2$, then we have condition (c) instead of (c'), which asserts a higher rate of k and hence a better rate of convergence and tighter confidence bands.

Remark 2. When $\alpha=1/2$, the deviation process based on f_n in (1.2) has the same limiting distribution as the deviation process based on the k-nearest-neighbor estimate (c.f., Mack, 1982). Therefore in studying the deviation processes the k-nearest-neighbor estimate is asymptotically equivalent to the symmetric case of the histogram density estimate.

Remark 3. Theorem 3.1 of Bickel and Rosenblatt (1973) derived the limiting distribution of the maximal deviation between a kernel estimate and the true density over a compact interval. In spite of the apparent similarity between this result and out Theorem 1, the resulting confidence bands are not the same. A similar remark was made by Mack (1982) for k-nearest-neighbor estimates. In particular, a_n depends on F, so the result of our Theorem 1 is not distribution-free. Nevertheless, we can make it distribution-free by replacing a_n and b_n by a'_n and b'_n respectively, where $a'_n=\left\{2\log\left[\frac{n}{k}(F_n(b)-F_n(a))\right]\right\}^{1/2}$ and b'_n is defined accordingly. It can be shown that

$$|a_n-a'_n|=O\left(\left[n^{-1}(\log\log n)\left(\log\frac{n}{k}\right)^{-1}\right]^{1/2}\right) \quad \text{w.p.1.}$$

IV. Hazard Rate Estimation

As a natural extension, the hazard rate

$$h(x)=f(x)/(1-F(x)), \tag{4.1}$$

if $F(x)<1$, can be estimated by

$$h_n(x)=f_n(x)/(1-F_n(x)), \tag{4.2}$$

where $f_n(x)$ is given in (1.2) and F_n is the empirical distribution of the X_i's. The pointwise asymptotic properties of h_n such as consistency and asymptotic normality are special cases of more general results by Prakasa Rao and Van Ryzin (1985). We now focus on the deviation process based on $h_n(x)$ given in (4.2), i.e.,

$$\sup_{x\in J}|h_n(x)-h(x)|/h(x),$$

where

$$J=[a,b], \quad 0<a<b<\infty, \quad F(b)<1.$$

Define

$$P_n(x)=k^{1/2}h^{-1}(x)(h_n(x)-h(x)), \tag{4.3}$$

the following theorem gives the asymptotic distribution for the maximum of the normalized deviation process based on $P_n(x)$:

Theorem 3. *Assume the hypotheses of Theorem 2 and assume $J=[a,b]$ with $F(b)<1$. We then have $\lim_{n\to\infty} P\{a_n[\sup_J|P_n(x)|-b_n]\leq\varepsilon\}=\exp\{-2e^{-\varepsilon}\}$, where a_n and b_n are given in* (2.9).

Proof. Apply the definitions of $h(x)$ in (4.1) and $h_n(x)$ in (4.2) and write $P_n(x)$ in terms of $Q_n(x)$ in (2.1), we see

$$\begin{aligned}P_n(x)&=k^{1/2}h^{-1}(x)(h_n(x)-h(x)),\\ &=\left(\frac{1-F(x)}{1-F_n(x)}\right)k^{1/2}f^{-1}(x)(f_n(x)-f(x))+k^{1/2}(1-F(x))\left(\frac{1}{1-F_n(x)}-\frac{1}{1-F(x)}\right)\\ &=(\mathrm{I})+(\mathrm{II}), \text{ say.}\end{aligned}$$

Note that $(\mathrm{I})=\left(1+\dfrac{F_n(x)-F(x)}{1-F_n(x)}\right)Q_n(x)$ and

$$(\mathrm{II})=k^{1/2}\frac{F_n(x)-F(x)}{1-F_n(x)}.$$

By the law of iterated logarithm for $F_n(x)$ and the fact that $1-F_n(x)$ is bounded uniformly in J w.p.1, we have

$(\mathrm{I})=Q_n(x)+Q_n(x)\cdot O((n^{-1}(\log\log n))^{1/2})$ w.p.1 and

$(\mathrm{II})=O((n^{-1}k(\log\log n))^{1/2})$ w.p.1.

Therefore,

$$a_n\sup_J|(\mathrm{I})|=a_n\sup_J|Q_n(x)|+(a_n\sup_J|Q_n(x)|)\,O((n^{-1}\log\log n)^{1/2}).$$

and

$$a_n \sup_J |(\mathrm{II})| = O\left(\left(n^{-1} k \left(\log \frac{n}{k}\right) (\log \log n)\right)^{1/2}\right).$$

The result then follows from Theorem 2.

References

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

Kim, B.K., Van Ryzin, J.: Uniform consistency of a histogram density estimator and model estimation. Comm. Statist. **4** (4), 303–315 (1975)

Kim, B.K., Van Ryzin, J.: On the asymptotic distribution of a histogram density estimator. Colloquia Societatis Janos Bolyai 32, Nonparametric Statistical Inference, Budapest. 483–499 (1980)

Komlos, J., Major, P., Tusnady, G.: An approximation of partial sums of independent r.v.s. and the sample d.f. I. Z. Wahrscheinlichkeitstheorie Verw. Geb. **32**, 111–131 (1975)

Mack, Y.P.: Rate of strong uniform convergence of $k-NN$ density estimators. J. Statist. Planning and Inf. **8**, 185–192 (1983)

Mack, Y.P.: On a goodness-of-fit problem of some nonparametric density estimates. Scand. J. Statist. **9**, 179–182 (1982)

Pickands, J.S. III: Upcrossing probability for Gaussian processes. Trans. Amer. Math. Soc. **145**, 51–73 (1969)

Prakasa Rao, B.L.S., Van Ryzin, J.: Asymptotic theory for two estimators of the generalized failure rate. Pacific Area Statistics Conference, ed. by K. Matusita. Amsterdam: North-Holland Press 547–569 (1985)

Stute, W.: The oscillation behavior of empirical processes. Ann. Probability **10**, 86–107 (1982)

Van Ryzin, J.: A histogram method of density estimation. Comm. Statist. 2, 493–506 (1973)

Watts, V.: The almost sure representation of intermediate order statistics. Z. Wahrscheinlichkeitstheorie Verw. Geb. **54**, 281–285 (1980)

Received August 20, 1984