AN ASYMPTOTICALLY OPTIMAL WINDOW SELECTION RULE FOR KERNEL DENSITY ESTIMATES[1]

BY

CHARLES J. STONE

TECHNICAL REPORT NO. 31
JUNE 1984

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

AN ASYMPTOTICALLY OPTIMAL WINDOW SELECTION

RULE FOR KERNEL DENSITY ESTIMATES[1]

By Charles J. Stone

University of California, Berkeley

June 4, 1984

Summary. Kernel estimates of an unknown multivariate density are investigated, with mild restrictions being placed on the kernel. A window selection rule is considered, which can be interpreted in terms of cross-validation. Under the mild assumption that the unknown density and its one-dimensional marginals are bounded, the rule is shown to be asymptotically optimal. This strengthens recent results of Peter Hall.

1. <u>Introduction.</u> Let X_1, X_2,... be independent $\mathbb{R}^d$-valued random variables having common unknown density p and consider the random sample $X_1,...,X_n$ of size n. In this paper we will study the asymptotic behavior as the sample size tends to infinity of a certain window selection rule for kernel estimates of the unknown density based on the random sample.

The kernel estimates are of the form

$$p_{nh}(x) = \frac{1}{n} \Sigma_1^n K_h(x-X_i) \;,$$

where $K_h(x) = v_h^{-1}K(x/h)$. Here the "window" $h = (h_1,...,h_d)$ belongs to $\mathbb{R}_+^d$, the collection of d-tuples of positive numbers; $v_h = h_1 \cdot ... \cdot h_d$ is the corresponding volume; $x/h = (x_1/h_1,...,x_d/h_d)$ for $x = (x_1,...,x_d) \in \mathbb{R}^d$; and K is a function on $\mathbb{R}^d$ having integral one and satisfying some mild restrictions, which will be described in Section 2.

The integrated squared error loss $L_{nh} = \int(p_{nh}-p)^2$ of the estimate p_{nh} can be written as

$$L_{nh} = \int p_{nh}^2 - 2\int p_{nh}p + \int p^2 \;.$$

The goal of minimizing this loss is equivalent to that of minimizing

$$L_{nh} - \int p^2 = \int p_{nh}^2 - 2\int p_{nh}p \;;$$

but this goal cannot be realized in practice, since $\int p_{nh}p$ is unknown. Observe, however, that

$$\int p_{nh}p = \frac{1}{n} \Sigma_1^n \int K_h(x-X_i)p(x)dx$$

and hence that

$$E\int p_{nh}p = \int\int p(x)p(y)K_h(x-y)dxdy = EK_h(X-Y) ,$$

where X and Y are independent random variables each having density
p . Consequently,

$$E\int p_{nh}p = E[\frac{1}{n(n-1)} \underset{i\neq j}{\Sigma\Sigma} K_h(X_i-X_j)] ,$$

where i and j are understood to range over {1,...,n} . This leads
to the unbiased estimate

$$\frac{1}{n(n-1)} \underset{i\neq j}{\Sigma\Sigma} K_h(X_i-X_j)$$

of $\int p_{nh}p$. A slight simplification leads to the estimate

$$\frac{1}{n^2} \underset{i\neq j}{\Sigma\Sigma} K_h(X_i-X_j)$$

of $\int p_{nh}p$; to the corresponding estimate

$$M_{nh} = \int p_{nh}^2 - \frac{2}{n^2} \underset{i\neq j}{\Sigma\Sigma} K_h(X_i-X_j) = \frac{1}{n^2} \underset{1}{\overset{n}{\Sigma}} \underset{1}{\overset{n}{\Sigma}} K_h^{(2)}(X_i-X_j) - \frac{2}{n^2} \underset{i\neq j}{\Sigma\Sigma} K_h(X_i-X_j)$$

of $L_{nh} - \int p^2$; and to the window selection rule, "choose the window h
to minimize the criterion M_{nh} ." This and other asymptotically equivalent
criteria have been proposed and studied by Rudemo (1982), Bowman (1984),
and Hall (1983a, 1983b). They point out that such criteria can also be
thought of in terms of cross-validation. Specifically, let p_{nih} be
the kernel estimate of p based on the random sample with the ith case
removed:

$$p_{nih}(x) = \frac{1}{n-1} \underset{j\neq i}{\Sigma} K_h(x-X_j) .$$

Then

$$\frac{1}{n} \Sigma_1^n P_{nih}(X_i) = \frac{1}{n(n-1)} \underset{i \neq j}{\Sigma\Sigma} K_h(X_i - X_j)$$

is the cross-validation estimate of $\int p_{nh} p$.

An alternative, asymptotically equivalent, cross-validation criterion (estimate of $L_{nh} - \int p^2$) considered by these authors is

$$\frac{1}{n} \Sigma_1^n \int p_{nih}^2 - \frac{2}{n} \Sigma_1^n P_{nih}(X_i) \, .$$

Hall showed that choosing $h \in H_n$ to minimize this cross-validation criterion is asymptotically optimal under certain conditions on K, H_n and p . In particular, K is assumed to be nonnegative. (If p is sufficiently smooth, then faster rates of convergence of the integrated squared error loss to zero can be obtained when the nonnegativity restriction on K is dropped.) The unknown density p is assumed to have a uniformly continuous square-integrable second derivative and to have finite second moment. Moreover, $H_n = \{(h_1,\ldots,h_1): \varepsilon \leq n^{1/(4+d)} h_1 \leq \lambda\}$, where $0 < \varepsilon < \lambda < \infty$. On the other hand, two of the restrictions imposed on K in Section 2 of this paper, compact support and Hölder continuity, are not required in Hall's results. (No serious attempt has been made here to eliminate or weaken these restrictions on K . For it is numerically more efficient to compute M_{nh} when K is a suitably chosen function with compact support; also minimizing M_{nh} by a numerical search technique is more attractive when K is at least mildly smooth. When $d = 1$ these two considerations suggest using the triangular kernel K defined by $K(x) = 1 - |x|$ for $|x| \leq 1$ and $K(x) = 0$ elsewhere; with this choice of K , after a preliminary sort of $X_1,\ldots,X_n$, the determination of M_{nh} for any given value of h requires only $O(n)$ computations.)

The purpose of this paper is to show that choosing $h \in \mathbb{R}_+^d$ to minimize M_{nh} is asymptotically optimal under a surprisingly mild assumption on p ; namely that p and its one-dimensional marginals are bounded. In this level of generality, there are no known theoretical results on the asymptotic behavior of the optimal window h or the optimal rate of convergence to zero of the integrated squared error of estimation.

The main result is described in Section 2 and proven in Section 3. The formulation of the result and the method of proof were influenced to some extent by several recent theoretical investigations of the Final Prediction Error (FPE) and other closely related model selection criteria in the regression context: Shibata (1981), Breiman and Freedman (1983), Rice (1983) and Chen (1983). The relatively long proof of Lemma 3 in Section 3 is given in Section 4. It uses "Poissonization", which has been employed by Rosenblatt (1975) and Krieger and Pickands (1981) in related contexts; interestingly, it also uses multiple stochastic integration with respect to a Poisson process.

A result similar to Theorem 1 was obtained for histogram density estimates in Stone (1983). The method of proof was also similar, except that the Poissonization argument used to prove the analog of Lemma 3 was much simpler.

Under various restrictions, Krieger and Pickands (1981) and Sacks and Ylvisaker (1981) obtained asymptotically optimal selection rules for kernel estimates of the density at a fixed point. In the later paper the entire kernel, not just the window, was optimized.

2. **Statement of the main result.** As mentioned above, the kernel K is required to have integral one. In addition, it is required to be symmetric about the origin, to have compact support, and to be Hölder continuous; that is, such that for some positive constants β and c ,

$$|K(y)-K(x)| \leq c|y-x|^\beta \quad \text{for} \quad x,y \in \mathbb{R}^d$$

(here $|x| = (x_1^2+\ldots+x_d^2)^{\frac{1}{2}}$ for $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$). The function K is not required to be nonnegative. Let $K^{(2)}$ denote the convolution of K with itself, so that $K^{(2)}(x) = \int K(x-y)K(y)dy$. Then $K^{(2)}$ satisfies the same assumptions as K ; in addition, $K^{(2)}(0) = \int K^2(y)dy > 0$. The kernel K is further restricted by requiring that $K^{(2)}(0) < 2K(0)$ (which necessarily holds if K is nonnegative and $K(0) = \max_x K(x)$).

Let h, v_h, x/h and K_h be defined as in Section 1 and note that $0 < v_h \leq |h|^d$. Also define $K_h^{(2)}$ by $K_h^{(2)}(x) = v_h^{-1}K^{(2)}(x/h)$. Then K_h and $K_h^{(2)}$ each have integral one and $K_h^{(2)}$ is the convolution of K_h with itself. Let p_{nh} and L_{nh} be defined as in Section 1, and observe that $\int p_{nh}^2$ and L_{nh} are both continuous on $\mathbb{R}_+^d$.

A window selection rule h_n is a $\mathbb{R}_+^d$-valued function of $X_1,\ldots,X_n$. Clearly

$$\frac{L_{nh_n}}{\min_h L_{nh}} \geq 1 .$$

The indicated minimum is actually taken on at some $h \in \mathbb{R}_+^d$. For it is easily seen that

$$\liminf_{h \to \partial\mathbb{R}_+^d} (L_{nh}-\int p^2) \geq 0 ;$$

also if the coordinates of h are all large, then

$$\int p_{nh}^2 \sim v_h^{-1} K^{(2)}(0) \quad \text{and} \quad \int p_{nh} p \sim v_h^{-1} K(0) ,$$

so

$$L_{nh} - \int p^2 \sim v_n^{-1}(K^{(2)}(0) - 2K(0)) < 0 .$$

(Here we have used the restriction that $K^{(2)}(0) < 2K(0)$.) The window selection rule h_n is said to be asymptotically optimal provided that

$$\lim_n \frac{L_{nh_n}}{\min_h L_{nh}} = 1 \quad \text{with probability one.}$$

Consider the window selection rule $\hat{h}_n$ defined to be a value of $h \in \mathbb{R}_+^d$ that minimizes the criterion M_{nh} introduced in Section 1. (It follows as in the previous paragraph that the minimum of M_{nh} is taken on at some $h \in \mathbb{R}_+^d$.) The one-dimensional marginals of p are defined to be the densities of the coordinates of X , where X has density p . The main result of this paper can now be stated simply as follows.

THEOREM 1. *If* p *and its one-dimensional marginals are bounded, then* $\hat{h}_n$ *is asymptotically optimal.*

Suppose p satisfies the assumptions of Theorem 1. Then, in the notation of Section 3, $\| p_h - p \| \to 0$ as $h \to 0$. Thus it follows from Theorem 1 together with Lemma 1 and Lemma 4 of Section 3 that $\hat{h}_n$ and $L_{n\hat{h}_n}$ both converge to zero with probability one as $n \to \infty$. For contrasting results when the Fourier transform of p vanishes outside a compact set C and the Fourier transform of K is the indicator function of C , see Ibragimov and Khasminskii (1982).

Burman (1984) has concurrently used the arguments of Shibata

(1980, 1981) to obtain a more general asymptotic optimality result for density estimation (with "in probability" instead of "with probability one" in the definition of asymptotic optimality). When specialized to kernel density estimation, the window h is selected from a finite set $H_n = \{h_1,\ldots,h_{N_n}\}$ subject to certain restrictions on N_n and the deterministic sequence $h_1,h_2,\ldots$; p is assumed to be bounded; and K is required to have finite 8th moment, but K is not required to be symmetric or continuous or to have compact support.

For related work in which integrated squared error loss is replaced by other measures of loss see Chow, Geman and Wu (1983); Devroye and Györfi (1983); Stone (1983); Birgé (1983); Marron (1984); and Bowman, Hall and Titterington (1984). For a recent review of a wide variety of smoothing techniques in statistics see Titterington (1984).

3. Proof of Theorem 1.

Throughout this section and the next one, it is assumed that p is bounded. Let p_h denote the convolution of K_h and p, so that

$$p_h(x) = \int K_h(x-y)p(y)dy = Ep_{nh}(x) \ .$$

Set $\| p_h - p \| = (\int (p_h - p)^2)^{\frac{1}{2}}$ and let $s \wedge t$ denote the minimum of $s, t \in \Re$.

LEMMA 1. *There are positive constants* b *and* c *such that*

$$\| p_h - p \|^2 \geq c(|h|^{bd} \wedge 1) \geq c(v_h^b \wedge 1) \quad for \quad h \in \Re_+^d \ .$$

PROOF. Let ϕ and ρ denote the Fourier transforms of K and p respectively. Then ϕ is bounded and continuous; it is real-valued since K is symmetric; it vanishes at infinity by the Riemann-Lebesgue lemma; it equals one at the origin and is not identically one on any neighborhood of the origin. The Fourier transform ϕ_h of K_h is given by $\phi_h(t) = \phi(ht)$, where $ht = (h_1 t_1, \ldots, h_d t_d)$; and the Fourier transform of p_h is $\phi_h \rho$. According to Parseval's identity and the boundness of the density p , $\int |\rho|^2 = (2\pi)^d \int p^2 < \infty$ and

$$(2\pi)^d \| p_h - p \|^2 = \int |\phi_h \rho - \rho|^2 = \int (1 - \phi_h)^2 |\rho|^2 \ .$$

Now ρ is continuous and $\rho(0) = 1$, so there is a nonempty bounded open ball C centered at the origin of $\Re^d$ such that $|\rho|^2 \geq \frac{1}{2}$ on C . Also $\| p_h - p \|^2$ is bounded away from zero for h outside any neighborhood of the origin. Suppose the desired conclusion is false. It then follows easily from the power series for the cosine function and a compactness argument that there is a unit vector $u \in \Re^d$ such that

$$\int_C dt \left(\int (ut \cdot x)^k K(x) dx \right)^2 = 0$$

for every positive even integer k. By continuity, for each such k,

$$\int (ut \cdot x)^k K(x) dx = 0 \quad \text{for all } t \in C.$$

Choose $j \in \{1,\ldots,d\}$ such that $u_j \neq 0$. By proper choice of t it follows that

$$\int x_j^k K(x) dx = 0$$

for every even integer k. By the symmetry of K, this equality holds for every positive integer k. But this is clearly impossible since K has integral one and compact support. (Suppose, say, that $j = 1$ and define K_1 by

$$K_1(x_1) = \int \cdots \int K(x_1,\ldots,x_d) dx_2 \cdots dx_d .$$

Then K_1 has integral one and compact support and $\int_{-\infty}^{\infty} x_1^k K_1(x_1) dx_1 = 0$ for every positive integer k. Consequently the Fourier transform of K_1 is identically equal to one, which contradicts the conclusion of the Riemann-Lebesgue lemma.)

Set

$$J_{nh} = \|p_h - p\|^2 + 1/nv_h ,$$

$$J_{nhr} = v_h^r \wedge 1 + 1/nv_h \quad \text{for } r > 0 ,$$

$$G_{nh} = n^{-1} \Sigma_1^n p_h(X_i) - E p_h(X) ,$$

and

$$G_n = n^{-1} \Sigma_1^n p(X_i) - E p(X) .$$

A modified form of Theorem 1 will first be proven, in which h ranges over a finite subset H_n of $\mathbb{R}_+^d$, the number of whose elements increases at most algebraically fast in n ; the original form of the theorem then follows (see the end of this section).

CONDITION 1. $\#(H_n) \leq An^a$ for $n \geq 1$, where A and a are positive constants.

LEMMA 2. *If Condition 1 holds, then*

$$\lim_{n} \max_{h \in H_n} J_{nh}^{-1} |G_{nh}-G_n| = 0 \text{ with probability one}$$

and

$$\lim_{n} \max_{h \in H_n} J_{nh}^{-1} |\int(p_{nh}-p_h)(p_h-p)| = 0 \text{ with probability one} .$$

PROOF. Set

$$Z_{ih} = p_h(X_i) - p(X_i) - (Ep_h(X)-Ep(X)) .$$

Then Z_{ih}, $i \geq 1$, are independent and identically distributed random variables each having mean zero. Since p is bounded, there is a positive constant c independent of h such that $|Z_{ih}| \leq c$ and Var $(Z_{ih}) \leq cu_h^2$, where $u_h = \| p_h-p \|$. Observe that $G_{nh} - G_n = \bar{Z}_{nh} = (Z_{1h}+...+Z_{nh})/n$. By Bernstein's inequality (see Hoeffding, 1963)

$$Pr(|\bar{Z}_{nh}| \geq t) \leq 2 \exp[-\tau\lambda/2(1+\lambda/3)] ,$$

where $0 \leq \lambda \leq t/u_h^2$ and $\tau = nt/c$. Choose $\varepsilon > 0$. Suppose that $u_h \geq n^{\varepsilon-\frac{1}{2}}$. Set $t = n^{\varepsilon-\frac{1}{2}}u_h$ and $\lambda = n^{\varepsilon-\frac{1}{2}}/u_h \leq 1$. Then $\lambda\tau = n^{2\varepsilon}/c$.

Suppose instead that $u_h < n^{\varepsilon-\frac{1}{2}}$. Set $t = n^{2\varepsilon-1}$ and $\lambda = 1$. Again, $\lambda\tau = n^{2\varepsilon}/c$. Thus in either case it follows from Bernstein's inequality that

$$Pr(|\bar{Z}_{nh}| \geq t) \leq 2 \exp(-n^{2\varepsilon}/3c) .$$

Hence by Condition 1.

$$\lim_n Pr(|\bar{Z}_{nh}| \geq n^{\varepsilon-\frac{1}{2}}u_h + n^{2\varepsilon-1} \text{ for some } h \in H_n) = 0 .$$

Thus to verify the first conclusion of Lemma 2 it is enough to show that for some $\varepsilon > 0$

$$\lim_n \max_{u>0} \frac{n^{\varepsilon-\frac{1}{2}}u+n^{2\varepsilon-1}}{u^2+1/nu^{2/b}} = 0 ,$$

where the positive number b is defined as in Lemma 1. For $0 < \varepsilon < 1/2(1+b)$, this result is easily shown by considering separately: $0 < u \leq n^{\varepsilon-\frac{1}{2}}$, $n^{\varepsilon-\frac{1}{2}} < u < n^{-b/2(1+b)}$, and $u > n^{-b/2(1+b)}$. The second conclusion of the Lemma follows from the same argument applied to

$$Z_{ih} = \int(K_h(x-X_i)-p_h(x))(p_h(x)-p(x))dx .$$

Let P_n denote the empirical distribution of $X_1,\ldots,X_n$ defined by

$$P_n(B) = n^{-1}\#\{i: 1 \leq i \leq n \text{ and } X_i \in B\} \text{ for } B \subseteq \Re^d .$$

The proof of the next result is postponed to Section 4.

LEMMA 3.   *If Condition 1 holds, then for all $r > 0$*

$$\lim_n \max_{h \in H_n} J_{nhr}^{-1}|\underset{x \neq y}{\iint} K_h(x-y)(P_n(dx)-P(dx))(P_n(dy)-P(dy))| = 0$$

   with probability one.

LEMMA 4. *If Condition 1 holds, then for all* $r > 0$

$$\lim_n \max_{h \in H_n} J_{nhr}^{-1} \left| \int (p_{nh} - p_h)^2 - K^{(2)}(0)/nv_h \right| = 0 \quad \textit{with probability one.}$$

PROOF. Observe that

$$\int (p_{nh} - p_h)^2 = \int \left(\int K_h(z-x)(P_n(dx) - P(dx)) \right)^2 dz$$

$$= \int\int K_h^{(2)}(x-y)(P_n(dx) - P(dx))(P_n(dy) - P(dy))$$

$$= \int\int_{x \neq y} K_h^{(2)}(x-y)(P_n(dx) - P(dx))(P_n(dy) - P(dy))$$

$$+ K^{(2)}(0)/nv_h ,$$

so the desired result follows from Lemma 3 (applied to $K_h^{(2)}$ instead of K_h).

Suppose now that h is constrained to lie in H_n , that $\hat{h}_n$ minimizes M_{nh} over H_n , and that Condition 1 holds. To verify that $\hat{h}_n$ is asymptotically optimal, it suffices to show that with probability one

$$\lim_n \max_{h,h' \in H_n} \frac{\left| L_{nh'} - L_{nh} - (M_{nh'} - M_{nh}) \right|}{L_{nh} + L_{nh'}} = 0 .$$

For this it is enough to show that

(3) $$\liminf_n \min_{h \in H_n} \frac{L_{nh}}{J_{nh}} > 0 \quad \text{with probability one}$$

and

(4) $$\lim_n \max_{h,h' \in H_n} \frac{\left| L_{nh'} - L_{nh} - (M_{nh'} - M_{nh}) \right|}{J_{nh} + J_{nh'}} = 0 \quad \text{with probability one.}$$

Since

$$L_{nh} = \int (p_{nh} - p)^2 = \int (p_{nh} - p_h)^2 + \| p_h - p \|^2 + 2\int (p_{nh} - p_h)(p_h - p) ,$$

(3) follows from Lemmas 1, 2 and 4. Observe next (see Section 1) that

$$L_{nh} - M_{nh} - 2G_n - \int p^2 = 2(G_{nh} - G_n) + 2 \iint_{x \neq y} K_h(x-y)(P_n(dx) - P(dx))(P_n(dy) - P(dy)) .$$

Thus (4) follows from Lemmas 1, 2 and 3.

Since K is Hölder continuous, the original form of Theorem 1 is easily reduced to the modified form based on Condition 1, the details being left to the reader. (Recall the assumption that the one-dimensional marginals of p are bounded. Accordingly, for given n , if one of the coordinates of h is very small, then $K_h^{(2)}(X_i - X_j) = K_h(X_i - X_j) = 0$ for $1 \leq i < j \leq n$ except on an event having very small probability.)

4. **Proof of Lemma 3.** The proof is based on "Poissonization."

Given a positive number λ , let $N(dx)$ be a Poisson process on $\mathbb{R}^d$ with $EN(B) = \lambda P(B)$. By definition, $N(B)$ has a Poisson distribution; and if $B_1,\ldots,B_k$ are disjoint, then $N(B_1),\ldots,N(B_k)$ are independent. Set $M(dx) = N(dx) - \lambda P(dx)$. Also, given a positive integer ℓ , let P^ℓ denote the probability measure on $\mathbb{R}^{\ell d}$ defined by

$$P^\ell(dx_1\cdots dx_\ell) = P(dx_1)\cdots P(dx_\ell) \ .$$

Let k and ℓ denote positive integers with $\ell \leq k$. Let $\Gamma^0_{k\ell}$ denote the collection of all k-tuples $i_1,\ldots,i_k$ of integers in $\{1,\ldots,\ell\}$ such that:

(a) each $i \in \{1,\ldots,\ell\}$ appears one or more times among $i_1,\ldots,i_k$;

(b) if i , $i' \in \{1,\ldots,\ell\}$ and $i < i'$, then i appears before i' among $i_1,\ldots,i_k$.

Given $x \in (x_1,\ldots,x_k) \in \mathbb{R}^k$ and $\gamma = (i_1,\ldots,i_k) \in \cup_1^k \Gamma^0_{k\ell}$, set $x_\gamma = (x_{i_1},\ldots,x_{i_k})$. Let $\Gamma_{k\ell}$ denote the subcollection of all $\gamma = (i_1,\ldots,i_k) \in \Gamma^0_{k\ell}$ such that each $i \in \{1,\ldots,\ell\}$ appears two or more times among $i_1,\ldots,i_k$. Observe that $\Gamma_{k\ell}$ is empty for $\ell > [k/2]$, where $[c]$ is the greatest integer no greater than c . By definition, $\Gamma_{k1} = \{(1,\ldots,1)\}$ for $k \geq 2$; while Γ_{42} consists of the three 4-tuples $(1,1,2,2)$, $(1,2,1,2)$ and $(1,2,2,1)$.

LEMMA 5. *Let* g *be a (Borel) function on* $\mathbb{R}^k$ *such that*

$$\sum_{\ell=1}^{k} \ \sum_{\gamma \in \Gamma^0_{k\ell}} |g(x_\gamma)| P^\ell(dx) < \infty \ .$$

Then

$$E \int\cdots\int g(x_1,\ldots,x_k) M(dx_1)\cdots M(dx_k) = \sum_{\ell=1}^{[k/2]} \lambda^\ell \sum_{\gamma \in \Gamma_{k\ell}} \int g(x_\gamma) P^\ell(dx) \ .$$

PROOF. It suffices to prove the result for functions g of the product form $g(x_1,\ldots,x_k) = \prod_1^k \psi_j(x_j)$, where ψ_j, $1 \le j \le k$, are bounded; the general result follows by the usual L^1 approximation argument. For functions of the indicated product form the desired result follows in a straightforward manner from the formula

$$Ee^{\sum_1^k t_i \int \psi_i dM} = e^\phi ,$$

where

$$\phi = \lambda \int (e^{\sum t_i \psi_i} - 1 - \sum t_i \psi_i) dP .$$

(Observe that

$$E \prod_1^k \int \psi_i dM = \frac{\partial^k e^\phi}{\partial t_1 \cdots \partial t_k} \Big|_0 ;$$

here $|_0$ means that $t_1 = \ldots = t_k = 0$. Note that $\phi|_0 = 0$ and $\partial\phi/\partial t_j |_0 = 0$. Thus it follows, for example, that

$$E \int \psi_1 dM \int \psi_2 dM = \frac{\partial^2 e^\phi}{\partial t_1 \partial t_2} \Big|_0$$

$$= \left(\frac{\partial^2 \phi}{\partial t_1 \partial t_2} + \frac{\partial\phi}{\partial t_1} \frac{\partial\phi}{\partial t_2} \right) e^\phi \Big|_0$$

$$= \frac{\partial^2 \phi}{\partial t_1 \partial t_2} \Big|_0$$

$$= \lambda \int \psi_1 \psi_2 dP$$

$$= \lambda \int g(x_1, x_1) dP$$

where $g(x_1, x_2) = \psi_1(x_1)\psi_2(x_2)$.)

For results related to Lemma 5 see Ogura (1972) and Krausz (1975).

LEMMA 6. *For each positive integer* k *there is a positive constant* c_k *such that*

$$E[(\iint_{x \neq y} K_h(x-y)M(dx)M(dy))^{2k}] \leq c_k v_h^{-2k} \sum_{\ell=2}^{2k} \lambda^\ell v_h^{[(\ell+1)/2]}$$

for $\lambda > 0$ *and* $h \in \mathbb{R}_+^d$.

PROOF. It follows from Lemma 5 that the indicated expectation is a finite linear combination of terms of the form

$$\lambda^\ell \int \cdots \int \prod_m K_h^{\nu_m}(x_{i_m} - x_{j_m})P(dx_1) \cdots P(dx_\ell) ,$$

where $1 \leq i_m < j_m \leq \ell$ and $\nu_m > 0$ for all m , $2 \leq \ell \leq 2k$, $\Sigma_m \nu_m = 2k$, and each $i \in \{1,\ldots,\ell\}$ appears at least once in the sequence $i_1,j_1,i_2,j_2,\ldots$. It follows easily from the boundedness of p and the definition of K_h that terms of this form are bounded in absolute value by a constant multiple of $\lambda^\ell v_h^{-2k} v_h^{[(\ell+1)/2]}$. The desired result now follows immediately.

Set $N = N(\mathbb{R}^d)$.

LEMMA 7. *For each positive integer* k *there is a positive constant* c_k *such that*

$$E[(\iint_{x \neq y} K_h(x-y)(N(dx)-NP(dx))(N(dy)-NP(dy))^{2k}]$$

$$\leq c_k(\lambda+\lambda^{2k}+v_h^{-2k} \sum_{\ell=2}^{2k} \lambda^\ell v_h^{[(\ell+1)/2]}) \text{ for } \lambda > 0 \text{ and } h \in \mathbb{R}_+^d .$$

PROOF. Observe first that

$$\iint_{x \neq y} K_h(x-y)(N(dx)-NP(dx))(N(dy)-NP(dy))$$

$$= \iint_{x \neq y} K_h(x-y)M(dx)M(dy) - 2(N-\lambda)\int p_h dM + (N-\lambda)^2 \int p_h p .$$

Now $|\int p_h p|$ is bounded in h and $E(N-\lambda)^{4k}$ is bounded above by a constant multiple of $\lambda + \lambda^{2k}$. Also $p_h(x)$ is bounded in h and x and

$$Ee^{t\int p_h dM} = \exp(\lambda \int (e^{tp_h}-1-tp_h)p) ,$$

so each cumulant of $\int p_h dM$ is a multiple of λ that is bounded in h . Since this random variable has mean zero, its $4k^{th}$ moment is bounded above by a constant multiple of $\lambda + \lambda^{2k}$. The desired result now follows from Lemma 6.

LEMMA 8. *For each positive integer* k *there is a positive constant* c_k *such that*

$$E[(\iint_{x \neq y} K_h(x-y)(P_n(dx)-P(dx))(P_n(dy)-P(dy)))^{2k}]$$

$$\leq c_k n^{-4k}(n^{2k}+v_h^{-2k} \sum_{\ell=2}^{2k} n^\ell v_h^{[(\ell+1)/2]}) \quad for \quad n \geq 1 \quad and \quad h \in \mathbb{R}_+^d .$$

PROOF. Set $N_n(dx) = nP_n(dx)$ and

$$Z = \iint_{x \neq y} K_h(x-y)(N_n(dx)-nP(dx))(N_n(dy)-nP(dy)) .$$

Let μ_n denote the $2k^{th}$ moment of Z and set $\mu_0 = 0$. Let $R(\lambda)$ denote the $2k^{th}$ moment of the random variable obtained through replacing n in the definition of Z by a Poisson random number N having mean λ , N being independent of X_i , $i \geq 1$. Then

$$R(\lambda) = \sum_n \Pr(N=n)\mu_n = \sum_n \frac{\lambda^n}{n!} e^{-\lambda}\mu_n$$

determines a polynomial of degree $2k$ in λ with $R(0) = 0$, and by

Lemma 7 there is a positive constant c_k' such that

$$0 \leq \sum_{j=1}^{2k} \frac{R^{(j)}(0)}{j!} \lambda^j = R(\lambda) \leq c_k'(\lambda + \lambda^{2k} + v_h^{-2k} \sum_{\ell=1}^{2k} \lambda^\ell v_h^{[(\ell+1)/2]})$$

for $\lambda > 0$ and $h \in \mathfrak{R}_+^d$. Thus there is a positive constant c_k'' such that

$$\sum_{j=1}^{2k} \frac{|R^{(j)}(0)|}{j!} \lambda^j \leq c_k''(\lambda + \lambda^{2k} + v_h^{-2k} \sum_{\ell=1}^{2k} \lambda^\ell v_h^{[(\ell+1)/2]})$$

for $\lambda > 0$ and $h \in \mathfrak{R}_+^d$. (For suppose otherwise and note that for each fixed $c > 0$, if

$$\frac{|R^{(j)}(0)|}{j!} \lambda^j \gg c_k'(\lambda + \lambda^{2k} + v_h^{-2k} \sum_{\ell=1}^{2k} \lambda^\ell v_h^{[(\ell+1)/2]})$$

(where $a \gg b > 0$ means that a/b is "very large"), then

$$\frac{R^{(j)}(0)}{j!} (c\lambda)^j \gg \sum_{j=1}^{2k} \frac{R^{(j)}(0)}{j!} (c\lambda)^j \geq 0 ;$$

by a compactness argument, there would then be a nonzero polynomial in c of degree $2k$ which equals zero at more than $2k$ distinct points.) Consequently,

$$\mu_n = \sum_{j=1}^{2k} \frac{n! R^{(j)}(0)}{(n-j)! j!} \leq \sum_{j=1}^{2k} \frac{|R^{(j)}(0)|}{j!} n^j$$

$$\leq c_k''(n + n^{2k} + v_h^{-2k} \sum_{\ell=1}^{2k} n^\ell v_h^{[(\ell+1)/2]}) ,$$

which yields the desired result.

Lemma 3 follows from Lemma 8 and a Chebychev type inequality involving the $2k^{th}$ moment by considering four cases separately: $v_h \geq 1$, $n^{-1/(r+1)} \leq v_h < 1$, $n^{-2} \leq v_h < n^{-1/(r+1)}$, and $0 < v_h < n^{-2}$.

REFERENCES

BIRGÉ, L. (1983). On estimating a density using Hellinger distance and some other strange facts. Technical Report MSRI 045-83, Mathematical Sciences Research Institute, Berkeley.

BOWMAN, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. Biometrika 71, to appear.

BOWMAN, A. W., HALL, P. and TITTERINGTON, D. M. (1984). Cross-validation in nonparametric estimation of probabilities and probability densities. Biometrika 71, to appear.

BREIMAN, L. and FREEDMAN, D. A. (1983). How many variables should be entered in a regression equation? J. Amer. Statist. Assoc. 78 131-136.

BURMAN, P. (1984). A data dependent approach to density estimation, manuscript.

CHEN, K.-W. (1983). Asymptotically optimal selection of a piecewise polynomial estimator of a regression function. Ph.D. Dissertation Dept. of Statist., Univ. of California, Berkeley.

CHOW, Y.-S., GEMAN, S. and WU, L. D. (1983). Consistent cross-validated density estimation. Ann. Statist. 11 25-38.

DEVROYE, L. and GYÖRFI, L. (1983). Nonparametric Density Estimation: The L_1 View, manuscript.

HALL, P. (1983). Large sample optimality of least squares cross-validation in density estimation. Ann. Statist. 11 1156-1174.

HALL, P. (1983). Asymptotic theory of minimum integrated square error for multivariate density estimation. Proceedings of the Sixth International Symposium on Multivariate Analysis, Pittsburg, 25-29 July 1983, to appear.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58 13-30.

IBRAGIMOV, I. A. and KHASMINSKI, R. Z. (1982). Estimation of distribution density belonging to a class of entire functions. Theory Probab. Appl. 27 551-562.

KRAUSZ, H. K. (1975). Identification of nonlinear systems using random impulse train inputs. Biol. Cybernetics 19 217-230.

KRIEGER, A. M. and PICKANDS, J. III (1981). Weak convergence and efficient density estimation at a point. Ann. Statist. 9 1066-1078.

MARRON, J. S. (1984). An asymptotically efficient solution to the bandwidth problem of kernel density estimation, manuscript.

OGURA, H. (1972). Orthogonal functionals of the Poisson Process. IEEE Trans. Information Theory IT-18 473-481.

RICE, J. (1983). Bandwidth choice for nonparametric kernel regression. Ann. Statist., to appear.

ROSENBLATT, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. Ann. Statist. 3 1-14.

RUDEMO, M. (1982). Empirical choice of histograms and kernel density estimators. Scand. J. Statist. 9 65-78.

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

SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. Ann. Statist. 8 147-164.

SHIBATA, R. (1981). An optimal selection of regression variables. Biometrika 68 45-54.

STONE, C. J. (1983). Optimal uniform rate of convergence for nonparametric estimators of a density function or its derivatives. Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday, 393-406, M. H. Rizvi, J. S. Rustagi, and D. Siegmund (eds.), Academic Press, New York.

STONE C. J. (1983). An asymptotically optimal histogram selection rule. Proceedings of the Neyman-Kiefer Meeting, to appear.

TITTERINGTON, D. M. (1984). Common structure of smoothing techniques in statistics, manuscript.

TECHNICAL REPORTS
Statistics Department
University of California, Berkeley

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.

2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosciences, (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.

3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis, (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.

4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11. No. 1, 185-214.

5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.

6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.

7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift, (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.

8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.

9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.

10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.

11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.

12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.

14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.

15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.

16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.

17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review.

18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.

19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research, (W. M. Mason and S. E. Fienberg, eds.).

20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting. 1985, Vol. 4, 251-262.

21. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.

22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.

23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics, (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984.

25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.

26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.

27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.

28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimate Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1985.

29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.

30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES. 1985 Vol 3 pp. 1-13.

31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.

32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.

33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.

34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.

35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science. Feb 1986, Vol. 1, No. 1, 3-39.

36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.

37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly. Feb 1986, Vol. 93, No. 2, 123-125.

38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré, 1985, 21, 225-287.

39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.

40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.

41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.

42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.

43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?

44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.

45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.

46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.

47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data.

48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.

49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift, 1986. D. Reidel.

50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada. January, 1986.

51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.

53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss.

54. BLACKWELL, D. (November 1985). Approximate normality of large products.

55. FREEDMAN, D. A. (June 1987). As others see us: A case study in path analysis. Journal of Educational Statistics. 12, 101-128.

56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.

57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.

58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.

59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.

60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.

61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.

62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.

63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.

64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.

65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.

66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.

67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data.

68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition).

69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.

71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.

72. LEHMANN, E.L. (July 1986). Statistics - an overview.

73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.

74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomial, and some continued fractions of Gauss.

75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.

76. O'SULLIVAN, F. (September 1986). Relative risk estimation.

77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.

78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.

79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks. To appear in Statistical Science.

80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.

81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.

82. DOKSUM, K.J. and LO, A.Y. (November 1986). Consistent and robust Bayes Procedures for Location based on Partial Information.

83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.

85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory. Ann. Inst. Henri Poincaré, 1987, 23, 397-423.

86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.

87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations.

88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.

89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987, revised January 1988). Estimating and testing in a two sample generalized odds rate model.

90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.

91. DIACONIS, P and FREEDMAN, D.A. (April 1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. To appear in the Journal of Applied Probability.

92. DABROWSKA, D.M. (April 1987, revised September 1987). Kaplan-Meier estimate on the plane.

92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.

93. DABROWSKA, D.M. (June 1987, revised Feb 1988). Nonparametric quantile regression with censored data.

94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery.

95. CANCELLED

96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in *Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley*.

97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for carcinogenesis. To appear in Environmental Health Perspectives.

98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.

99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.

100. LE CAM, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald.

101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators. Annals of Statistics, June, 1988.

102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.

103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.

104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer. IEEE Computer Graphics and applications, June, 1988.

105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.

106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.

107. CHENG, C-S. (August 1987). Some orthogonal main-effect plans for asymmetrical factorials.

108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.

109. KLASS, M.J. (August 1987). Maximizing $E \max_{1 \le k \le n} S_k^+ / ES_n^+$: A prophet inequality for sums of I.I.D. mean zero variates.

110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals. Annals of Statistics, June, 1988.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987, revised June 1988). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.

113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.

114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.

115. BICKEL, P.J. and RITOV, Y. (Sept. 1987, revised Aug 1988). Large sample theory of estimation in biased sampling regression models I.

116. RITOV, Y. and BICKEL, P.J. (Sept.1987, revised Aug. 1988). Achieving information bounds in non and semiparametric models.

117. RITOV, Y. (October 1987). On the convergence of a maximal correlation algorithm with alternating projections.

118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.

119. HESSE, C.H. (October 1987). An asymptotic expansion for the mean of the passage-time distribution of integrated Brownian Motion.

120. DONOHO, D. and LIU, R. (October 1987, revised March 1988). Geometrizing rates of convergence, II.

121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. To appear in *Statistics a Guide to the Unknown*.

122. ALDOUS, D., FLANNERY, B. and PALACIOS, J.L. (November 1987). Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains.

123. DONOHO, D.L., MACGIBBON, B. and LIU, R.C. (Nov.1987, revised July 1988). Minimax risk for hyperrectangles.

124. ALDOUS, D. (November 1987). Stopping times and tightness II.

125. HESSE, C.H. (November 1987). The present state of a stochastic model for sedimentation.

126. DALANG, R.C. (December 1987, revised June 1988). Optimal stopping of two-parameter processes on nonstandard probability spaces.

127. Same as No. 133.

128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.

129. SMITH, D.L. (December 1987). Exponential bounds in Vapnik-Červonenkis classes of index 1.

130. STONE, C.J. (Nov.1987, revised Sept. 1988). Uniform error bounds involving logspline models.

131. Same as No. 140

132. HESSE, C.H. (December 1987). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes

133. DONOHO, D.L. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean, I.

134. DUBINS, L.E. and SCHWARZ, G. (December 1987). A sharp inequality for martingales and stopping-times.

135. FREEDMAN, D.A. and NAVIDI, W. (December 1987). On the risk of lung cancer for ex-smokers.

136. LE CAM, L. (January 1988). On some stochastic models of the effects of radiation on cell survival.

137. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the uniform consistency of Bayes estimates for multinomial probabilities.

137a. DONOHO, D.L. and LIU, R.C. (1987). Geometrizing rates of convergence, I.

138. DONOHO, D.L. and LIU, R.C. (January 1988). Geometrizing rates of convergence, III.

139. BERAN, R. (January 1988). Refining simultaneous confidence sets.

140. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.

141. BRILLINGER, D.R. (January 1988). Two reports on trend analysis: a) An Elementary Trend Analysis of Rio Negro Levels at Manaus, 1903-1985 b) Consistent Detection of a Monotonic Trend Superposed on a Stationary Time Series

142. DONOHO, D.L. (Jan. 1985, revised Jan. 1988). One-sided inference about functionals of a density.

143. DALANG, R.C. (February 1988). Randomization in the two-armed bandit problem.

144. DABROWSKA, D.M., DOKSUM, K.A. and SONG, J.K. (February 1988). Graphical comparisons of cumulative hazards for two populations.

145. ALDOUS, D.J. (February 1988). Lower bounds for covering times for reversible Markov Chains and random walks on graphs.

146. BICKEL, P.J. and RITOV, Y. (Feb.1988, revised August 1988). Estimating integrated squared density derivatives.

147. STARK, P.B. (March 1988). Strict bounds and applications.

148. DONOHO, D.L. and STARK, P.B. (March 1988). Rearrangements and smoothing.

149. NOLAN, D. (March 1988). Asymptotics for a multivariate location estimator.

150. SEILLIER, F. (March 1988). Sequential probability forecasts and the probability integral transform.

151. NOLAN, D. (March 1988). Limit theorems for a random convex set.

152. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On a theorem of Kuchler and Lauritzen.

153. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the problem of types.

154. DOKSUM, K.A. (May 1988). On the correspondence between models in binary regression analysis and survival analysis.

155. LEHMANN, E.L. (May 1988). Jerzy Neyman, 1894-1981.

156. ALDOUS, D.J. (May 1988). Stein's method in a two-dimensional coverage problem.

157. FAN, J. (June 1988). On the optimal rates of convergence for nonparametric deconvolution problem.

158. DABROWSKA, D. (June 1988). Signed-rank tests for censored matched pairs.

159. BERAN, R.J. and MILLAR, P.W. (June 1988). Multivariate symmetry models.

160. BERAN, R.J. and MILLAR, P.W. (June 1988). Tests of fit for logistic models.

161. BREIMAN, L. and PETERS, S. (June 1988). Comparing automatic bivariate smoothers (A public service enterprise).

162. FAN, J. (June 1988). Optimal global rates of convergence for nonparametric deconvolution problem.

163. DIACONIS, P. and FREEDMAN, D.A. (June 1988). A singular measure which is locally uniform.

164. BICKEL, P.J. and KRIEGER, A.M. (July 1988). Confidence bands for a distribution function using the bootstrap.

165. HESSE, C.H. (July 1988). New methods in the analysis of economic time series I.

166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.

167. BREIMAN, L., STONE, C.J. and KOOPERBERG, C. (August 1988). Confidence bounds for extreme quantiles.

168. LE CAM, L. (August 1988). Maximum likelihood an introduction.

169. BREIMAN, L. (August 1988). Submodel selection and evaluation in regression-The conditional case and little bootstrap.

170. LE CAM, L. (September 1988). On the Prokhorov distance between the empirical process and the associated Gaussian bridge.

171. STONE, C.J. (September 1988). Large-sample inference for logspline models.

172. ADLER, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.

173. MILLAR, P.W. (October 1988). Optimal estimation in the non-parametric multiplicative intensity model.

174. YOR, M. (October 1988). Interwinings of Bessel processes.

175. ROJO, J. (October 1988). On the concept of tail-heaviness.

176. ABRAHAMS, D.M. and RIZZARDI, F. (September 1988). BLSS - The Berkeley interactive statistical system: An overview.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.