ASYMPTOTIC PROPERTIES OF LOGSPLINE DENSITY ESTIMATION[1]

By

Charles J. Stone

Technical Report No. 69
July 1986

[1]Research partially supported by
National Science Foundation Grants MCS83-01257 and DMS-8600409.

Department of Statistics
University of California
Berkeley, California

ASYMPTOTIC PROPERTIES OF LOGSPLINE DENSITY ESTIMATION[1]

by Charles J. Stone

University of California, Berkeley

Summary

Let f be a continuous and positive, but otherwise unknown, density function on a known compact interval I, let F denote the distribution function of f, and let $Q = F^{-1}$ denote its quantile function. An exponential family model for f is constructed having p parameters and based on a B-spline model for $\log f$. Maximum likelihood estimation of the parameters of the model based on a random sample of size n from f yields estimates $\hat{f}$, $\hat{F}$ and $\hat{Q}$ of f, F, and Q. Under mild conditions, if $p \to \infty$ appropriately as $n \to \infty$, these estimators achieve the optimal rate of convergence. The asymptotic behavior of the corresponding confidence intervals is also investigated. In particular, it is shown that the asymptotic standard errors of $\hat{F}$ and $\hat{Q}$ coincide with those of the usual empirical distribution function and empirical quantile function.

[1]This research was supported in part by National Science Foundation Grants MCS83-01257 and DMS-8600409.

AMS 1980 subject classifications. Primary 62G05; secondary 62F12.

Key words and phrases. Exponential family, spline, maximum likelihood, nonparametric density estimate, optimal rate of convergence, confidence intervals.

1. Statement of Results. Let $Y_1, Y_2 \cdots$ be independent and identically distributed random variables taking on values in a known compact interval, which is taken to be $I = [0, 1]$. These random variables are assumed to have a continuous and positive density function f on I. Let F denote the distribution function of f and let $Q = F^{-1}$ denote its quantile function. Given the positive integer n, consider the random sample $Y_1, \cdots, Y_n$ of size n. We will construct an exponential family model for f having p_n parameters and based on a B-spline model for $\log f$. Maximum likelihood estimation of these parameters yields estimators and confidence intervals for f, F, and Q. The asymptotic theory of these procedures, a blend of parametric theory and nonparametric theory, will be investigated in this paper. The main results will be described here and proven in the following sections.

Let m denote a positive integer and let K_n, $n \geq 1$, denote a sequence of positive integers. Let I be partitioned into subintervals

$$I_k = [(k-1)/K_n, k/K_n), \ 1 \leq k < K_n, \qquad \text{and} \qquad I_{K_n} = [(K_n-1)/K_n, 1].$$

Let $\mathcal{S}_n$ denote the collection of functions s on I satisfying the following two properties: s is a polynomial of order m (degree $m-1$) or less on each of the subintervals $I_1, \cdots, I_{K_n}$; and if $m \geq 2$, s is $(m-2)$-times continuously differentiable on I. Then $\mathcal{S}_n$ is a vector space of dimension $p_n = m+K_n-1$, which is referred to as the space of polynomial splines of order m with simple knots at k/K_n for $1 \leq k < K_n$. The functions in $\mathcal{S}_n$ are called piecewise constant, linear, quadratic or cubic splines according as $m = 1, 2, 3$, or 4. Let B_{nk}, $1 \leq k \leq p_n$, denote the usual B-spline basis of $\mathcal{S}_n$ (see de Boor, 1978). Then $0 \leq B_{nk} \leq 1$ and $\Sigma_1^{p_n} B_{nk} = 1$ on I. There is a fixed positive integer J, depending on m but not on n, such that the support of each B_{nk} is contained in the convex hull of J consecutive knots and if $|k-j| > J$, the supports of B_{nj} and B_{nk} are disjoint.

Given $\theta \in \Theta_n$, the collection of all p_n-dimensional vectors, set

$$|\theta| = (\sum_j \theta_j^2)^{1/2},$$

$$s_n(\cdot;\ \theta) = \sum_j \theta_j B_{nj},$$

$$C_n(\theta) = \log(\int \exp(s_n(\cdot;\ \theta))),$$

and

$$f_n(\cdot;\ \theta) = \exp(s_n(\cdot;\ \theta) - C_n(\theta)).$$

Then

$$\int f_n(\cdot;\ \theta) = 1 \qquad \text{for} \quad \theta \in \Theta_n.$$

Observe that $f_n(\cdot;\ \theta)$, $\theta \in \Theta_n$, is an exponential family having p_n parameters. Let $F_n(\cdot;\ \theta)$ and $Q_n(\cdot;\ \theta)$ denote the distribution function and quantile function corresponding to $f_n(\cdot;\ \theta)$. Set

$$\Lambda_n(\theta) = E \log f_n(Y;\ \theta) = \int \log f_n(\cdot;\ \theta) f = \int s_n(\cdot;\ \theta) f - C_n(\theta), \qquad \theta \in \Theta_n.$$

It is assumed from now on that $p_n \geq 2$ for all n. The vector θ of parameters is not identifiable; for if we add a constant to each element of θ, we do not change $f_n(\cdot;\ \theta)$. Let Θ_{n0} denote the (p_n-1)-dimensional subspace of Θ_n consisting of those vectors $\theta \in \Theta_n$ the sum of whose elements is zero.

Let $H_n(\theta)$ denote the Hessian of $C_n(\cdot)$ at θ; that is, the $p_n \times p_n$ matrix whose (j, k)th element is

$$\frac{\partial^2 C_n(\theta)}{\partial\theta_j \partial\theta_k}.$$

It is an elementary and well known property of exponential families (see Lehmann, 1983) that if $\theta, \tau \in \Theta_n$, then

(1) $\quad \tau' H_n(\theta) \tau = \int (s_n(\cdot;\ \tau) - a)^2 f_n(\cdot;\ \theta), \qquad \text{where} \quad a = \int s_n(\cdot;\ \tau) f_n(\cdot;\ \theta).$

Thus $\tau' H_n(\theta) \tau > 0$ if τ is a nonzero element of Θ_{n0}. Consequently, $C_n(\cdot)$ is a strictly convex function on Θ_{n0}. Since $-H_n(\theta)$ is the Hessian matrix of $\Lambda_n(\theta)$ at θ, $\Lambda_n(\theta)$ is strictly concave on Θ_{n0}. If $\theta \in \Theta_{n0}$ and

$\theta \neq 0$, then

$$\Lambda_n(t\theta) = t\int s_n(\cdot;\ \theta)f - \log(\int e^{ts_n(\cdot;\ \theta)})$$

and $s_n(\cdot;\ \theta)$ is not almost everywhere equal to a constant on I; so $\Lambda_n(t\theta) \to -\infty$ as $t \to \infty$. It follows that *for each* $n \geq 1$, *there is a unique* $\theta_n \in \Theta_{n0}$ *that maximizes* $\Lambda_n(\theta)$, $\theta \in \Theta_{n0}$. Set $f_n = f(\cdot;\ \theta_n)$, $F_n = F_n(\cdot;\ \theta_n)$ and $Q_n = Q_n(\cdot;\ \theta_n)$.

It follows from the assumption on f that $\log f$ is continuous and hence bounded on I. Let $\|\ \|_2$ and $\|\ \|_\infty$ denote the usual L_2 and L_∞ norms of functions on I. Set

$$\delta_n = \inf_{s \in \mathcal{S}_n} \|s - \log f\|_\infty.$$

If $p_n \to \infty$ as $n \to \infty$, then $\delta_n = o(1)$ by (2) on Page 167 of de Boor (1978). Let m_1 be a nonnegative integer, let $0 < \alpha \leq 1$ and set $q = m_1+\alpha$. If f is m_1-times differentiable and its mth derivative satisfies a Hölder condition with index α, then $\delta_n = O(p_n^{-q})$ (see de Boor, 1978).

THEOREM 1. (i) $\|f_n - f\|_\infty = O(\delta_n)$; (ii) $\|F_n - F\|_\infty = O(p_n^{-1}\delta_n)$; *and* (iii) $\|Q_n - Q\|_\infty = O(p_n^{-1}\delta_n)$.

Let ℓ_n be the log-likelihood function based on the logspline and the random sample of size n; so that

$$\ell_n(\theta) = \sum_i \log f_n(Y_i;\ \theta) = \sum_i (s_n(Y_i;\ \theta)-C_n(\theta)).$$

Then $\ell_n(\cdot)$ is a strictly concave function on Θ_{n0}. Let $\hat{\theta}_n \in \Theta_{n0}$ denote the maximum likelihood estimate (MLE) of $\theta \in \Theta_{n0}$ based on the random sample of size n. Then $\hat{\theta}_n$ is unique if it exists. (A necessary and sufficient condition for existence is given in Barndorff-Nielsen, 1978; see also Johansen, 1979.) Set $\hat{f}_n = f_n(\cdot;\ \hat{\theta}_n)$, $\hat{F}_n = F_n(\cdot;\ \hat{\theta}_n)$ and $\hat{Q}_n = Q_n(\cdot;\ \hat{\theta}_n)$. Then $\hat{f}_n$ is called a *logspline density estimate* of f since $\log \hat{f}_n = s_n(\cdot;\ \hat{\theta}_n) - C_n(\hat{\theta}_n) \in \mathcal{S}_n$. If $m = 1$, then $\hat{f}_n$ is the usual histogram density estimate.

From now on it is assumed that

(2) $$p_n = o(n^{.5-\epsilon}) \quad \text{for some} \quad \epsilon > 0.$$

THEOREM 2. (i) $\hat{\theta}_n$ *exists except on an event whose probability tends to zero with* n;

(ii) $|\hat{\theta}_n - \theta_n| = O_{pr}(n^{-1/2} p_n)$;

(iii) $\max_{1 \le j \le p_n} |\hat{\theta}_{nj} - \theta_{nj}| = O_{pr}((n^{-1} p_n \log(p_n))^{1/2})$;

(iv) $\|\hat{f}_n - f_n\|_2 = O_{pr}((n^{-1} p_n)^{1/2})$;

(v) $\|\hat{f}_n - f_n\|_\infty = O_{pr}((n^{-1} p_n \log(p_n))^{1/2})$;

(vi) $\|\hat{F}_n - F_n\|_\infty = O_{pr}(n^{-1/2})$;

and

(vii) $\|\hat{Q}_n - Q_n\|_\infty = O_{pr}(n^{-1/2})$.

Theorems 1 and 2 allow us to get the usual optimal rates of convergence under various smoothness assumptions on f; see Stone (1980, 1982, 1983). Consider a smoothness assumption that leads, as above, to a conclusion of the form $\delta_n = O(p_n^{-q})$; and suppose that $q > 1/2$. Set $\gamma = 1/(2q+1)$ and $r = q/(2q+1) = q\gamma$. Then the positive constants γ and r are both less than $1/2$. To get the optimal rate of convergence of $\|\hat{f}_n - f\|_2$ to zero, we choose $p_n \sim n^\gamma$ and obtain

$$\|\hat{f}_n - f\|_2 = O_{pr}(n^{-r}).$$

(Here $a_n \sim b_n$ means that a_n/b_n is bounded away from zero and infinity.) To get the optimal rate of convergence of $\|\hat{f}_n - f\|_\infty$ to zero, we choose $p_n \sim (n/\log(n))^\gamma$ and obtain

$$\|\hat{f}_n - f\|_\infty = O_{pr}((n^{-1}\log(n))^r).$$

To get the optimal rate of convergence of $\|\hat{F}_n - F\|_\infty$ or $\|\hat{Q}_n - Q\|_\infty$ to zero, we choose p_n so that

$$Mn^{1/(2q+2)} \le p_n = O(n^{.5-\epsilon}) \quad \text{for some} \quad M, \epsilon > 0$$

and obtain

$$\|\hat{F}_n - F\|_\infty = O_{pr}(n^{-1/2}) \quad \text{and} \quad \|\hat{Q}_n - Q\|_\infty = O_{pr}(n^{-1/2}).$$

Let $\mathcal{I}_n(\theta) = nH_n(\theta)$ denote the information function based on the random sample of size n and let $\mathcal{I}_n^{-1}(\theta)$ denote the inverse to $\mathcal{I}_n(\theta)$ viewed as a linear transformation of Θ_{n0}. Set $\mathcal{I}_n^{-1} = \mathcal{I}_n^{-1}(\theta_n)$ and $\hat{\mathcal{I}}_n^{-1} = \mathcal{I}_n^{-1}(\hat{\theta}_n)$. Let $G_n(y), \hat{G}_n(y) \in \Theta_{n0}$ denote the p_n-dimensional vectors having elements

$$G_{nj}(y) = B_{nj}(y) - \frac{\partial C_n}{\partial \theta_j}(\theta_n)$$

and

$$\hat{G}_{nj}(y) = B_{nj}(y) - \frac{\partial C_n}{\partial \theta_j}(\hat{\theta}_n).$$

respectively. Set

$$SE(\hat{f}_n(y)) = f_n(y)(G_n(y)'\mathcal{I}_n^{-1}G_n(y))^{1/2}$$

and

$$\hat{SE}(\hat{f}_n(y)) = \hat{f}_n(y)(\hat{G}_n(y)'\hat{\mathcal{I}}_n^{-1}\hat{G}_n(y))^{1/2}.$$

THEOREM 3. *Suppose that* $p_n \to \infty$ *as* $n \to \infty$. *Then uniformly in* $y \in I$,

$$SE(\hat{f}_n(y)) \sim (n^{-1}p_n)^{1/2},$$

$$\frac{\hat{SE}(\hat{f}_n(y))}{SE(\hat{f}_n(y))} = 1 + o_{pr}(1),$$

and

$$\mathcal{L}\left[\frac{\hat{f}_n(y)-f_n(y)}{SE(\hat{f}_n(y))}\right] \to \mathcal{N}(0, 1).$$

It follows from Theorem 3 that $\hat{f}_n(y) \pm z_{1-.5\alpha}\hat{SE}(\hat{f}_n(y))$ is an asymptotic $(1-\alpha)$-level confidence interval for $f_n(y)$; if $\delta_n = o((n^{-1}p_n)^{1/2})$, it is also an asymptotic $(1-\alpha)$-level confidence interval for $f(y)$. Here $\Phi(z_q) = q$, Φ being the standard normal distribution function. Set

$$SE(\hat{F}_n(y)) = (F(y)(1-F(y))/n)^{1/2} \quad \text{and} \quad SE(\hat{Q}_n(t)) = \left\{\frac{t(1-t)}{nf^2(Q(t))}\right\}^{1/2}.$$

Note that $\mathrm{Int}(I) = (0, 1)$, since $I = [0, 1]$.

THEOREM 4. *Suppose that* $p_n \to \infty$ *as* $n \to \infty$. *Then*

$$\mathcal{L}\left[\frac{\hat{F}_n(y)-F_n(y)}{SE(\hat{F}_n(y))}\right] \to \mathcal{N}(0,\ 1) \quad \textit{uniformly on compact subsets of} \quad \mathrm{int}(I)$$

and

$$\mathcal{L}\left[\frac{\hat{Q}_n(t)-Q_n(t)}{SE(\hat{Q}_n(t))}\right] \to \mathcal{N}(0,\ 1) \quad \textit{uniformly on compact subsets of} \quad (0,\ 1).$$

Theorem 4 leads in an obvious manner to asymptotic $(1-\alpha)$-level confidence intervals for $F_n(y)$ and $Q_n(t)$. It follows from Theorem 1 that under the mild condition $\delta_n/p_n = o(n^{-1/2})$, these are also asymptotic $(1-\alpha)$-level confidence intervals for $F(y)$ and $Q(t)$. It is interesting to note that the associated standard errors coincide with those for the usual nonparametric estimators of these quantities. ($\hat{F}_n$ and $\hat{Q}_n$ are much smoother than the corresponding nonparametric estimators.)

The results in this paper can be extended in two directions with essentially no change in proof: the restriction that the functions in $\mathcal{S}_n$ be $(m-2)$-times continuously differentiable on I can be weakened in an arbitrary manner; and the knot locations $1/K_n, \cdots, (K_n-1)/K_n$ can be replaced by a sequence which is σ-quasi-uniform in the sense of Page 216 of Schumaker (1981) (i. e., such that the ratios of the differences between consecutive knots are bounded away from zero and infinity).

Modifications to handle $I = (0,\ \infty)$ or $I = (-\infty,\ \infty)$ involving data-dependent knot selection, linear restrictions on the tails of the functions in $\mathcal{S}_n$, and transformations were described in Stone and Koo (1986) and illustrated on simulated data with $n = 200$ and $p_n = 5$ (4 degrees of freedom). The various estimates and confidence intervals look very reasonable and, especially those for extreme quantiles, appear to have considerable practical utility. But a thorough Monte Carlo study or a rigorous theoretical analysis that takes these modifications into account has yet to be carried out.

For recent bibliographies of nonparametric density estimation, see Wertz and Schneider (1979) and Titterington (1985). Leonard (1978) and Silverman (1982) considered various nonparametric estimators of $\log f$, an attractive feature of such an approach being that the corresponding estimate of f itself is automatically positive. Logspline density estimation corresponds to estimating $\log f$ by a normalized member of a particular flexible finite-dimensional vector space, $\mathcal{S}_n$, of functions and hence to estimating f by a member of a particular flexible exponential family. Previously, Neyman (1937) and Crain (1974, 1976a, 1976b, 1977) considered other such exponential families.

2. Proof of Theorem 1.

LEMMA 1. $\int s(f_n - f) = 0$ *for* $s \in \mathcal{S}_n$.

PROOF. Choose $s \in \mathcal{S}_n$ and define g on $\mathbb{R}$ by $\int f_n e^{ts-g(t)} = 1$. Then $g'(0) = \int sf_n$. Also $\int(\log f_n + ts - g(t))f$ is maximized at $t = 0$ and hence $g'(0) = \int sf$. Thus the conclusion of the lemma is valid.

By Lemma 1, $f_n - s$ is the orthogonal projection of $f - s$ onto $\mathcal{S}_n$ for all $s \in \mathcal{S}_n$. Thus by Theorem 1 of de Boor (1976), there is an $M > 0$ such that $\|f_n - s\|_\infty \le M\|f - s\|_\infty$ for $n \ge 1$ and $s \in \mathcal{S}_n$. If $s \in \mathcal{S}_n$ and $\|f - s\|_\infty \le 2\delta_n$, then $\|f_n - f\|_\infty \le 2(M+1)\delta_n$. This yields (i). We will now verify (ii), from which (iii) follows easily. It can be assumed that $p_n \to \infty$ as $n \to \infty$. It will be shown below that there is a positive constant M satisfying the following condition: for $n \ge 1$ and $x \in [2Mp_n^{-1}, 1-Mp_n^{-1}]$ there is an $s \in \mathcal{S}_n$ such that $-1 \le s \le 1$ on I, $s = 1$ on $[Mp_n^{-1}, x - Mp_n^{-1}]$, and $s = 0$ on $[x + Mp_n^{-1}, 1]$. According to Lemma 1,

$$0 = \int_0^1 s(f_n - f) = \int_0^x (f_n - f) + \int_0^{Mp_n^{-1}} (s - 1)(f_n - f)$$
$$+ \int_{x-Mp_n^{-1}}^{x} (s - 1)(f_n - f) + \int_x^{x+Mp_n^{-1}} s(f_n - f).$$

The desired result now follows from Theorem 1(i).

In constructing the desired function s it can be assumed that $m > 1$ (since the result is obvious when $m = 1$). Let B_1 and B_2 be elements of the usual B-spline basis with m replaced by $m-1$ such that B_1 vanishes outside $[0, Mp_n^{-1}]$ and B_2 vanishes outside $[x-Mp_n^{-1}, x+Mp_n^{-1}]$. Then B_1 and B_2 are nonnegative functions. Let c_1 and c_2 be positive constants such that $\int c_1B_1 = 1$ and $\int c_2B_2 = 1$. Then s, defined by

$$s(y) = \int_0^y (c_1B_1(z) - c_2B_2(z))dz,$$

has the desired properties. This completes the proof of Theorem 1.

3. **Proof of Theorem 2.**

LEMMA 2. *For any positive number* M *there is an* $\delta > 0$ *such that if* f *and* g *are probability density functions on* I, $\|g\|_\infty \le M$ *and* $\|\log f - \log g\|_\infty \le M$, *then*

$$\min_a \int(\log f - \log g - a)^2 \ge \delta\int(\log f - \log g)^2.$$

PROOF. The minimum value of a is $a_0 = \int\log f - \int\log g$. Set $h = \log f - \log g - a_0$. Then $|h| \le 2M$. Since f and g integrate to 1, $a_0 = -\log(1 + \int g(e^h-1))$. Thus there is a positive constant M_1 depending on M such that $a_0^2 \le M_1\int h^2$. Consequently

$$\int(\log f - \log g)^2 = a_0^2 + \int h^2 \le (M_1 + 1)\int h^2,$$

which yields the desired result.

LEMMA 3. *There is a positive number* M *such that*

$$|C_n(\theta)| \le M\|\log f_n(\cdot;\ \theta)\|_\infty \quad \textit{for all} \quad \theta \in \Theta_{n0}.$$

PROOF. Since

$$\log f_n(\cdot;\ \theta) = \sum_j(\theta_{nj}-C_n(\theta))B_{nj} \qquad \text{for} \quad \theta \in \Theta_n,$$

the desired result follows from (viii) on Page 155 of de Boor (1978).

LEMMA 4. *Let* M *be a positive constant. Then there are positive constants* M_1 *and* δ *such that if* $n \ge 1$, $\theta \in \Theta_n$ *and*

$$\|\log f_n(\cdot;\ \theta) - \log f_n\|_2 \le Mp_n^{-1/2},$$

then

$$\|\log f_n(\cdot;\ \theta_n+t(\theta-\theta_n))\|_\infty \le M_1 \qquad \textit{for} \quad 0 \le t \le 1$$

and

$$\Lambda_n(\theta) - \Lambda_n(\theta_n) \le -\delta\|\log f_n(\cdot;\ \theta) - \log f_n\|_2^2.$$

PROOF. Without loss of generality, it can be assumed that $\theta \in \Theta_{n0}$. It follows from Theorem 1(i) and from Lemma 7 of Stone (1986) that there is a positive constant M_2 depending on M such that if n and θ are as in the statement of the present lemma, then $\|\log f_n\|_\infty \le M_2$ and

$\|\log f_n(\cdot;\ \theta)\|_\infty \le M_2$. Thus by Lemma 3 there is a positive constant M_3 depending on M such that

$$\|s_n(\cdot;\ \theta_n + t(\theta - \theta_n))\|_\infty \le M_3 \quad \text{for } 0 \le t \le 1,$$

which yields the first conclusion of the lemma. Observe that

$$\frac{d}{dt} \Lambda_n(\theta_n + t(\theta - \theta_n))\big|_{t=0} = 0.$$

Thus it follows from (1) and Taylor's theorem with remainder that

$$\Lambda_n(\theta) - \Lambda_n(\theta_n) = -\frac{1}{2}\int (s_n(\cdot;\ \theta - \theta_n) - a)^2 f_n(\cdot;\ \theta_n + t(\theta - \theta_n))$$

for some $t \in (0,\ 1)$ and constant a. The second conclusion of the lemma now follows from the first conclusion and Lemma 2.

LEMMA 5. *Given* $M > 0$, *there is a* $\delta > 0$ *such that*

$$\Pr\left[\left|\frac{\ell_n(\theta) - \ell_n(\theta_n)}{n} - (\Lambda_n(\theta) - \Lambda_n(\theta_n))\right| \ge c\|\log f_n(\cdot;\ \theta) - \log f_n\|_2\right] \le 2e^{-\delta nc^2}$$

for $n \ge 1$, $\theta \in \Theta_n$, *and* $0 < c \le Mp_n^{-1/2}$.

PROOF. Write

$$\ell_n(\theta) - \ell_n(\theta_n) - n(\Lambda_n(\theta) - \Lambda_n(\theta_n)) = \Sigma_1^n Z_i,$$

where

$$Z_i = \log f_n(Y_i;\ \theta) - \log f_n(Y_i) - E(\log f_n(Y_i;\ \theta) - \log f_n(Y_i)).$$

Set $a = \|\log f_n(\cdot;\ \theta) - \log f_n\|_2$. By Lemma 7 of Stone (1986) there is a positive number M_1 such that

$$\|\log f_n(\cdot;\ \theta) - \log f_n\|_\infty \le M_1 a p_n^{1/2}.$$

Thus there is a positive constant M_2 such that $|Z_i| \le M_2 a p_n^{1/2}$ and $\operatorname{Var}(Z_i) \le M_2 a^2$. The desired result now follows from Bernstein's inequality (see Theorem 3 of Hoeffding, 1963).

The next result is an immediate consequence of the definitions of the various terms.

LEMMA 6. *If* $\theta_1, \theta_2 \in \Theta_n$, *then*

$$\left|\frac{\ell_n(\theta_2) - \ell_n(\theta_1)}{n} - (\Lambda_n(\theta_2) - \Lambda_n(\theta_1))\right| \le 2\|\log f_n(\cdot;\ \theta_2) - \log f_n(\cdot;\ \theta_1)\|_\infty.$$

Set $\mathcal{F}_n = \{f_n(\cdot;\,\theta): \theta \in \Theta_n\}$. It is convenient to define the "diameter" of a subset F of $\mathcal{F}_n$ as

$$\sup\{\|\log f_2 - \log f_1\|_\infty: f_1, f_2 \in F\}.$$

The next result, essentially Lemma 12 of Stone (1986), is a consequence of Lemma 7 of that paper and (viii) on Page 155 of de Boor (1978).

LEMMA 7. *Given* $\epsilon > 0$ *and* $\delta > 0$, *there is an* $M > 0$ *such that the following is valid*:

$$\{\theta \in \Theta_n: \|\log f_n(\cdot;\,\theta) - \log f_n\|_2 \le n^\epsilon (p_n/n)^{1/2}\}$$

can be covered by $O(\exp(Mp_n \log(n)))$ *subsets each having diameter at most* $\delta n^{2\epsilon} p_n/n$.

LEMMA 8. $\hat{f}_n$ *exists and is unique except on an event whose probability tends to zero with* n. *Moreover*,

$$\|\log \hat{f}_n - \log f_n\|_2 = O_{pr}(n^\epsilon (n^{-1}p_n)^{1/2}) \qquad \textit{for all } \epsilon > 0.$$

PROOF. Set $c_n = n^\epsilon (n^{-1}p_n)^{1/2}$ and

$$\Theta_{n1} = \{\theta \in \Theta_{n0}: \|\log f_n(\cdot;\,\theta) - \log f_n\|_2 \le c_n\}.$$

Then Θ_{n1} is a compact set whose boundary, relative to Θ_{n0}, is contained in

$$\Theta_{n2} = \{\theta \in \Theta_{n0}: \|\log f_n(\cdot;\,\theta) - \log f_n\|_2 = c_n\}.$$

In light of (2), it can be assumed that there is a positive constant M such that $c_n \le Mp_n^{-1/2}$ for $n \ge 1$. By Lemma 4 there is a $\delta > 0$ such that

$$\Lambda_n(\theta) - \Lambda_n(\theta_n) \le -\delta c_n^2 \qquad \text{for all } \theta \in \Theta_{n2}.$$

Thus by Lemmas 5-7, except on an event whose probability tends to zero with n,

$$\ell_n(\theta) < \ell_n(\theta_n) \qquad \text{for all } \theta \in \Theta_{n2}$$

and hence $\ell_n(\cdot)$ has a local maximum in the relative interior of Θ_{n1}. The desired conclusions now follow from the strict concavity of $\ell_n(\theta)$ on Θ_{n0}.

Let $S_n(\theta) \in \Theta_{n0}$ denote the score function; that is, the p_n-dimensional vector of elements

$$\frac{\partial \ell_n(\theta)}{\partial \theta_j} = \sum_i \left[B_{nj}(Y_i) - \frac{\partial C_n}{\partial \theta_j}(\theta)\right].$$

Set $S_n = S_n(\theta_n)$. Then $ES_n = 0$ and

$$E|S_n|^2 = n\sum_j \mathrm{Var}(B_{nj}(Y)) \le n\sum_j EB_{nj}^2(Y) = nE\sum_j B_{nj}^2(Y) = O(n).$$

Consequently, the following result is valid.

LEMMA 9. $|S_n|^2 = O_{pr}(n)$.

The maximum likelihood equation for $\hat{\theta}_n$ is $S_n(\hat{\theta}_n) = 0$, which can be rewritten as

$$D_n(\hat{\theta}_n - \theta_n) = S_n,$$

where D_n is the $p_n \times p_n$ matrix defined by

$$D_n = n\int_0^1 H_n(\theta_n + t(\hat{\theta}_n - \theta_n))dt.$$

LEMMA 10. *There is a* $\delta > 0$ *such that, except on an event whose probability tends to zero with* n,

$$(\hat{\theta}_n - \theta_n)'D_n(\hat{\theta}_n - \theta_n) \ge \delta n \|\log \hat{f}_n - \log f_n\|_2^2.$$

PROOF. It follows from Lemma 4, Lemma 8 and (2) that

$$\max_{0 \le t \le 1} \|\log f_n(\cdot;\, \theta_n + t(\hat{\theta}_n - \theta_n))\|_\infty = O_{pr}(1).$$

The desired result now follows from (1), Theorem 1(i) and Lemma 2.

LEMMA 11. $(\hat{\theta}_n - \theta_n)'S_n = O_{pr}((np_n)^{1/2})\|\log \hat{f}_n - \log f_n\|_2$.

PROOF. Let u denote the p_n-dimensional vector each of whose coordinates is 1. Then $u'S_n = 0$ since $S_n \in \Theta_{n0}$. Now

(5) $$\log \hat{f}_n - \log f_n = \sum_j (\hat{\theta}_{nj} - \theta_{nj} - C_n(\hat{\theta}_n) + C_n(\theta_n))B_{nj};$$

so it follows from Lemma 9 together with (12) of Stone (1986) that

$$\begin{aligned}|(\hat{\theta}_n - \theta_n)'S_n|^2 &= |(\hat{\theta}_n - \theta_n - (C_n(\hat{\theta}_n) - C_n(\theta_n))u)'S_n|^2 \\ &\le |(\hat{\theta}_n - \theta_n) - (C_n(\hat{\theta}_n) - C_n(\theta_n)u|^2 |S_n|^2 \\ &= O_{pr}(np_n)\|\log \hat{f}_n - \log f_n\|_2^2\end{aligned}$$

as desired.

LEMMA 12. (i) $\|\log \hat{f}_n - \log f_n\|_2 = O_{pr}((n^{-1}p_n)^{1/2})$;

(ii) $\|\log \hat{f}_n - \log f_n\|_\infty = O_{pr}(n^{-1/2}p_n) = o_{pr}(1)$;

(iii) $\|\hat{f}_n - f_n\|_2 = O_{pr}((n^{-1}p_n)^{1/2})$;

(iv) $|\hat{\theta}_n - \theta_n| = O_{pr}(n^{-1/2}p_n)$.

PROOF. According to the maximum likelihood equation for $\hat{\theta}_n$,

$$(\hat{\theta}_n-\theta_n)'D_n(\hat{\theta}_n-\theta_n) = (\hat{\theta}_n-\theta_n)'S_n.$$

Thus the first result follows from Lemmas 10 and 11. The second result now follows from (2) and Lemma 7 of Stone (1986). The third result follows from the first two results and Theorem 1(i). Since $(\hat{\theta}_n-\theta_n)'u = 0$, it follows from (5), the first result, and (12) of Stone (1986) that

$$|\hat{\theta}_n-\theta_n|^2 + p_n(C_n(\hat{\theta}_n)-C_n(\theta_n))^2 = |\hat{\theta}_n - \theta_n + (C_n(\hat{\theta}_n)-C_n(\theta_n)u|^2$$
$$= O(p_n\|\log \hat{f}_n - \log f_n\|_2^2) = O_{pr}(n^{-1}p_n^2)$$

and hence that the last result is valid.

The next result follows from (2), Lemma 4 and Lemma 12(i).

LEMMA 13. *There are positive constants* M_1 and M_2 *such that, except on an event whose probability tends to zero with* n,

$$M_1 \le f_n(\cdot;\ \theta_n+t(\hat{\theta}_n-\theta_n)) \le M_2 \qquad \text{for } 0 \le t \le 1.$$

Let $\nabla C_n(\theta)$ denote the gradient of $C_n(\theta)$; that is, the p_n-dimensional vector of elements

$$\frac{\partial C_n(\theta)}{\partial \theta_j}.$$

LEMMA 14. $C_n(\hat{\theta}_n) - C_n(\theta_n) = \nabla C_n(\theta_n)'(\hat{\theta}_n-\theta_n) + O_{pr}(n^{-1}p_n)$.

PROOF. Observe that

$$C_n(\hat{\theta}_n) - C_n(\theta_n) = \nabla C_n(\theta_n)'(\hat{\theta}_n-\theta_n) + (\hat{\theta}_n-\theta_n)'R_n(\hat{\theta}_n\ \theta_n),$$

where R_n is the $p_n \times p_n$ matrix defined by

$$R_n = \int_0^1 (1-t)H_n(\theta_n+t(\hat{\theta}_n-\theta_n))dt.$$

The desired result now follows from (1), Lemma 12(iv), Lemma 13 and (12) of Stone (1986).

LEMMA 15. *There is a positive constant* M *such that, except on an event whose probability tends to zero with* n,

$$\sum_j \left[\sum_k \sum_m \max_{0\le t\le 1}\left|\frac{\partial^3 C_n}{\partial\theta_j\partial\theta_k\partial\theta_m}(\theta_n+t(\hat{\theta}_n-\theta_n))\right||\tau_k|\right]^2 \le Mp_n^{-2}|\tau|^2 \qquad \textit{for} \quad \tau\in\Theta_n.$$

PROOF. It is easily seen that

$$\begin{aligned}\frac{\partial^3 C_n(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_m} = {} & \int B_{nj}B_{nk}B_{nm}f_n(\cdot;\,\theta) \\ & - \int B_{nj}B_{nk}f_n(\cdot;\,\theta)\int B_{nm}f_n(\cdot;\,\theta) \\ & - \int B_{nj}B_{nm}f_n(\cdot;\,\theta)\int B_{nk}f_n(\cdot;\,\theta) \\ & - \int B_{nk}B_{nm}f_n(\cdot;\,\theta)\int B_{nj}f_n(\cdot;\,\theta) \\ & + \int B_{nj}f_n(\cdot;\,\theta)\int B_{nk}f_n(\cdot;\,\theta)\int B_{nm}f_n(\cdot;\,\theta).\end{aligned}$$

The desired result now follows from Lemma 13 and the basic properties of B-splines.

It follows easily from Theorem 1(i) that

$$\max_{1\le j\le p_n}\left|\frac{\partial C_n}{\partial\theta_j}(\theta_n)\right| = O(p_n^{-1}) \tag{6}$$

and hence that

$$|G_n(y)| \sim 1 \quad \text{uniformly for} \quad y\in I. \tag{7}$$

The next result follows from (5) and Lemma 14.

LEMMA 16. $\|\log\hat{f}_n - \log f_n - G_n(\cdot)'(\hat{\theta}_n-\theta_n)\|_\infty = O_{pr}(n^{-1}p_n)$.

For $\theta\in\Theta_n$ let $\min f_n(\cdot;\,\theta)$ and $\max f_n(\cdot;\,\theta)$ respectively denote the minimum and maximum values of $f_n(y;\,\theta)$ as y ranges over I.

LEMMA 17. *For each* $\theta\in\Theta_n$, $H_n(\theta)$ *is a positive definite symmetric linear transformation on* Θ_{n0}. *There are positive constants* M_1 *and* M_2 *such that*

$$M_1p_n^{-1}|\tau|^2\min f_n(\cdot;\,\theta) \le \tau'H_n(\theta)\tau \le M_2p_n^{-1}|\tau|^2\max f_n(\cdot;\,\theta)$$

for $n\ge 1$, $\theta\in\Theta_n$ *and* $\tau\in\Theta_{n0}$.

PROOF. The symmetry of $H_n(\theta)$ follows directly from its definition. By (12) of Stone (1986) there are positive constants M_1 and M_2 such that

$$M_1 p_n^{-1}|\tau|^2 \le \|s_n(\cdot;\ \tau)\|_2^2 \le M_2 p_n^{-1}|\tau|^2 \qquad \text{for } \tau \in \Theta_n.$$

It follows from (1) that

$$\tau' H_n(\theta)\tau \le \|s_n(\cdot;\ \tau)\|_2^2 \max f_n(\cdot;\ \theta) \le M_2 p_n^{-1}|\tau|^2 \max f_n(\cdot;\ \theta)$$

for $\tau \in \Theta_n$. It also follows from (1) that if $\tau \in \Theta_{n0}$ then

$$\begin{aligned} \tau' H_n(\theta)\tau &\ge \|s_n(\cdot;\ \tau - au)\|_2^2 \min f_n(\cdot;\ \theta) \\ &\ge M_1 p_n^{-1}|\tau - au|^2 \min f_n(\cdot;\ \theta) \\ &\ge M_1 p_n^{-1}|\tau|^2 \min f_n(\cdot;\ \theta). \end{aligned}$$

Thus the conclusion of the lemma is valid.

LEMMA 18. *There are positive constants* M_1 *and* M_2 *such that*

$$M_1 n p_n^{-1}|\tau|^2 \le \tau' \mathrm{Cov}(S_n)\tau \le M_2 n p_n^{-1}|\tau|^2 \qquad \textit{for } n \ge 1 \textit{ and } \tau \in \Theta_{n0}.$$

PROOF. Since

$$\tau' \mathrm{Cov}(S_n)\tau = n\int (s_n(\cdot;\ \tau) - a)^2 f,$$

where $a = \int s_n(\cdot;\ \tau) f$, the result follows from the argument used to prove Lemma 17.

Consider the approximation $\hat{\varphi}_n \in \Theta_{n0}$ to $\hat{\theta}_n - \theta_n$ defined by $\mathcal{I}_n \hat{\varphi}_n = S_n$. Then $\hat{\varphi}_n = \mathcal{I}_n^{-1} S_n$ and hence $G_n(y)'\hat{\varphi}_n = G_n(y)'\mathcal{I}_n^{-1} S_n$. It follows easily from (7) and Lemma 17 that

(8) $$|\tau' \mathcal{I}_n^{-1} G_n(y)| = O(n^{-1} p_n |\tau|) \qquad \text{uniformly in } \tau \in \Theta_n \text{ and } y \in I.$$

LEMMA 19. $\max_{1 \le j \le p_n} |\hat{\varphi}_{nj}| = O_{pr}((n^{-1} p_n \log(p_n))^{1/2})$.

PROOF. Since $\mathcal{I}_n^{-1} \mathrm{Cov}(S_n) \mathcal{I}_n^{-1}$ is the covariance matrix of $\hat{\varphi}_n$, it follows from Lemmas 17 and 18 that $\max_j \mathrm{Var}(\hat{\varphi}_{nj}) = O(p_n/n)$. Observe that

$$\hat{\varphi}_{nj} = \Sigma_1^n (\mathcal{I}_n^{-1} G_n(Y_i))_j.$$

According to (8),

$$\sup_{y \in I} \max_{1 \le j \le p_n} |(\mathcal{I}_n^{-1} G_n(y))_j| = O(p_n/n).$$

The desired result now follows from Bernstein's inequality.

LEMMA 20. $|\hat{\theta}_n - \theta_n - \hat{\varphi}_n|^2 = O_{pr}(n^{-2}p_n^3\log(p_n))$.

PROOF. It follows from the maximum likelihood equation that

$$\hat{\theta}_n - \theta_n = \hat{\varphi}_n - \mathcal{I}_n^{-1}(D_n-\mathcal{I}_n)(\hat{\theta}_n-\theta_n).$$

According to Lemma 17,

$$|\mathcal{I}_n^{-1}(D_n-\mathcal{I}_n)(\hat{\theta}_n-\theta_n)|^2 = O_{pr}(1)(p_n/n)^2|(D_n-\mathcal{I}_n)(\hat{\theta}_n-\theta_n)|^2.$$

The (j, k)th element of $D_n-\mathcal{I}_n$ can be written as

$$n\sum_m A_{njkm}(\hat{\theta}_{nm}-\theta_{nm}),$$

where

$$A_{njkm} = \int_0^1 (1-t)\frac{\partial^3 C_n}{\partial\theta_j\partial\theta_k\partial\theta_m}(\theta_n+t(\hat{\theta}_n-\theta_n))dt.$$

Thus the jth element of $(D_n-\mathcal{I}_n)(\hat{\theta}_n-\theta_n)$ equals

$$n\sum_k\sum_m A_{njkm}(\hat{\theta}_{nk}-\theta_{nk})(\hat{\theta}_{nm}-\theta_{nm}).$$

Hence by Lemmas 12 and 15

$$|(D_n-\mathcal{I}_n)(\hat{\theta}_n-\theta_n)|^2 = O_{pr}(n\max_{1\le j\le p_n}(\hat{\theta}_{nj}-\theta_{nj})^2)$$

and therefore

$$|\hat{\theta}_n - \theta_n - \hat{\varphi}_n|^2 = O_{pr}(n^{-1}p_n^2\max_{1\le j\le p_n}(\hat{\theta}_{nj}-\theta_{nj})^2).$$

Consequently, by Lemma 19,

$$\max_{1\le j\le p_n}(\hat{\theta}_{nj}-\theta_{nj})^2 = O_{pr}(n^{-1}p_n\log(p_n) + n^{-1}p_n^2\max_{1\le j\le p_n}(\hat{\theta}_{nj}-\theta_{nj})^2).$$

Thus by (2)

$$\max_{1\le j\le p_n}(\hat{\theta}_{nj}-\theta_{nj})^2 = O_{pr}(n^{-1}p_n\log(p_n)),$$

which yields the desired result.

The first conclusion to Theorem 2 is contained in Lemma 8; the second and fourth conclusions are contained in Lemma 12; and the third conclusion follows from (2) and Lemmas 19 and 20. We will now verify the fifth conclusion. It follows from (5) and Lemma 14 that

(9) $$\|\log \hat{f}_n - \log f_n - G_n(\cdot)'(\hat{\theta}_n-\theta_n)\|_\infty = O_{pr}(n^{-1}p_n).$$

Since

$$G_n(y)'\hat{\varphi}_n = \sum_j B_{nj}(y)\hat{\varphi}_{nj} - \sum_j \frac{\partial C_n}{\partial \theta_j}(\theta_n)\hat{\varphi}_{nj},$$

it follows from (6) and Lemma 19 that

$$(10) \qquad \|G_n(\cdot)'\hat{\varphi}_n\|_\infty = O_{pr}((n^{-1}p_n\log(p_n))^{1/2}).$$

The fifth conclusion follows from (2), (7), (9), (10), Theorem 1(i), Lemma 12(ii), and Lemma 20.

We will now verify the sixth conclusion. It follows from (6) and Lemma 20 that

$$(11) \qquad \nabla C_n(\theta_n)'(\hat{\theta}_n - \theta_n - \hat{\varphi}_n) = O_{pr}(n^{-1}p_n\log^{1/2}(p_n)).$$

It follows from Lemma 20 (and, of course, the basic properties of B-splines) that

$$(12) \qquad \sum_j |\hat{\theta}_{nj} - \theta_{nj} - \hat{\varphi}_{nj}| \int B_{nj} = O_{pr}(n^{-1}p_n\log^{1/2}(p_n)).$$

By (11) and (12),

$$(13) \qquad \int |G_n(\cdot)'(\hat{\theta}_n - \theta_n - \hat{\varphi}_n)| = O_{pr}(n^{-1}p_n\log^{1/2}(p_n)).$$

The sixth conclusion is a consequence of (2), (7), (10), (13), Theorem 1(i), Lemma 20, and the following result; while the seventh conclusion follows from the sixth conclusion and Theorem 1.

LEMMA 21. $\max_{0\le x\le 1} |\int_0^x f_n(y)G_n(y)'\hat{\varphi}_n dy| = O_{pr}(n^{-1/2})$.

PROOF. Observe that

$$\mathrm{Var}(\nabla C_n(\theta_n)'\hat{\varphi}_n) = \mathrm{Var}(\nabla C_n(\theta_n)'\mathcal{I}_n^{-1}S_n) = (\mathcal{I}_n^{-1}\nabla C_n(\theta_n))'\mathrm{Cov}(S_n)\mathcal{I}_n^{-1}\nabla C_n(\theta_n).$$

Thus it follows from (6), Lemma 17 and Lemma 18 that

$$\mathrm{Var}(\nabla C_n(\theta_n)'\hat{\varphi}_n) = O(np_n^{-1}|\mathcal{I}_n^{-1}\nabla C_n(\theta_n)|^2) = O(n^{-1}p_n|\nabla C_n(\theta_n)|^2) = O(n^{-1})$$

and hence that

$$\nabla C_n(\theta_n)'\hat{\varphi}_n = O_{pr}(n^{-1/2}).$$

Consequently, to prove the desired result it suffices to verify that

$$(14) \qquad \max_{0\le x\le 1} |\sum_j \hat{\varphi}_{nj}\int_0^x f_n(y)B_{nj}(y)dy| = O_{pr}(n^{-1/2}).$$

For any particular value of x, all but a bounded number of terms $\int_0^x f_n(y)B_{nj}(y)dy$ are equal to $\int_0^1 f_n(y)B_{nj}(y)dy$ or to zero. By (2) and Lemma 19, the total contribution of the bounded number of exceptional terms is

$$O(p_n^{-1} \max_{1\le j\le p_n} |\hat{\varphi}_{nj}|) = O_{pr}(p_n^{-1}(n^{-1}p_n\log(p_n))^{1/2}) = o_{pr}(n^{-1/2}).$$

Thus, by the form of the support of the B-splines B_{nj}, $1 \le j \le p_n$ (see de Boor (1978)), to verify (14) it suffices to verify

$$\max_{1\le k\le p_n} |\Sigma_1^k \hat{\varphi}_{nj}\int f_n B_{nj}| = o_{pr}(n^{-1/2}). \tag{15}$$

Let $\mathcal{J}$ be a subset of consecutive integers in $\{1,\cdots,p_n\}$ and let J denote the number of integers in $\mathcal{J}$. Let τ denote the p_n-dimensional vector having elements $\int f_n B_{nj}$ for $j \in \mathcal{J}$ and zero otherwise. Then

$$|\tau|^2 = O(Jp_n^{-2}) \quad \text{uniformly in } \mathcal{J} \text{ and } n. \tag{16}$$

Since

$$\mathrm{Var}(\sum_{j\in\mathcal{J}} \hat{\varphi}_{nj}\int f_n B_{nj}) = (\mathcal{J}_n^{-1}\tau)'\mathrm{Cov}(S_n)(\mathcal{J}_n^{-1}\tau),$$

it follows from (16) and Lemmas 17 and 18 that

$$\mathrm{Var}(\sum_{j\in\mathcal{J}} \hat{\varphi}_{nj}\int f_n B_{nj}) = O(Jn^{-1}p_n^{-1}) \quad \text{uniformly in } \mathcal{J} \text{ and } n. \tag{17}$$

Observe next that

$$\sum_{j\in\mathcal{J}} \hat{\varphi}_{nj}\int f_n B_{nj} = \Sigma_1^n \int(\sum_{j\in\mathcal{J}} f_n B_{nj}(\mathcal{J}_n^{-1}G_n(Y_i))_j) = \Sigma_1^n X_{n\mathcal{J}i}. \tag{18}$$

Here $X_{n\mathcal{J}i}$, $1 \le i \le n$, are independent random variables having mean zero; and, by (7), Lemma 17, and the basic properties of B-splines,

$$|X_{n\mathcal{J}i}| \le b_n \quad \text{with } b_n = O(n^{-1}p_n^{1/2}). \tag{19}$$

It follows from (17)-(19) and Bernstein's inequality that there is a $\beta > 0$, which does not depend on n or $\mathcal{J}$, such that

$$\begin{aligned} &\Pr(|\sum_{j\in\mathcal{J}} \hat{\varphi}_{nj}\int f_n B_{nj}| \ge An^{-1/2}(J/p_n)^c) \\ &\le 2[\exp(-\beta An^{1/2}p_n^{-1/2}(J/p_n)^c) + \exp(-\beta A^2(p_n/J)^{1-2c})] \end{aligned} \tag{20}$$

for $A > 0$ and $0 < c < .5$.

Set $R_n = \min[r\colon 2^r \ge p_n]$. For $0 \le r \le R_n$, let $\mathcal{M}_{nr}$ denote the collection of all sets of integers of the form

$$\{(m-1)2^r+1,\cdots,m2^r\}, \qquad \text{where } 1 \le m \le p_n/2^r.$$

It follows from (2) and (20) that, for any $\epsilon > 0$, A can be chosen sufficiently large so that

(21) $$\Pr\Big(\Big|\sum_{j\in\mathcal{J}} \hat{\varphi}_{nj}\int f_n B_{nj}\Big| \ge A n^{-1/2}(J/p_n)^c \text{ for some } \mathcal{J} \in \cup_0^{R_n} \mathcal{M}_{nr}\Big)$$
$$\le \epsilon \quad \text{for all } n \ge 1.$$

For $1 \le k \le p_n$, $\{1,\cdots,k\}$ can be written as a disjoint union of sets $\mathcal{J} \in \mathcal{M}_{nr}$ such that for $0 \le r \le R_n$, there is at most one such $\mathcal{J} \in \mathcal{M}_{nr}$. Thus it follows from (21) that (15) holds, as desired.

4. Proof of Theorem 3. Now

$$\operatorname{Var} G_n(y)'\hat{\varphi}_n = G_n(y)'\mathcal{I}_n^{-1}\operatorname{Cov}(S_n)\mathcal{I}_n^{-1}G_n(y), \tag{22}$$

so it follows from (7), Theorem 1(i), and Lemmas 17 and 18 that

$$\operatorname{Var} G_n(y)'\hat{\varphi}_n \sim n^{-1}p_n. \tag{23}$$

LEMMA 22. *Uniformly in* $y \in I$,

$$\mathcal{L}\left[\frac{G_n(y)'\hat{\varphi}_n}{SD(G_n(y)'\hat{\varphi}_n)}\right] \to \mathcal{N}(0,1) \quad \textit{as} \quad n \to \infty.$$

PROOF. Observe that $G_n(y)'\hat{\varphi}_n = \Sigma_1^n Z_{ni}$, where $Z_{ni} = G_n(y)'\mathcal{I}_n^{-1}G_n(Y_i)$. For each n, the random variables $Z_{n1}, \cdots, Z_{nn}$ have mean zero and are independent and identically distributed. Moreover,

$$|G_n(y)'\mathcal{I}_n^{-1}G_n(Y_i)|^2 \le |G_n(y)'\mathcal{I}_n^{-1}G_n(y)|\ |G_n(Y_i)'\mathcal{I}_n^{-1}G_n(Y_i)| = O(n^{-2}p_n^2).$$

The desired result now follows from (2), (23) and the central limit theorem (see the corollary on page 201 of Chung, 1974).

LEMMA 23. *There is a positive constant* M *such that*

$$|\tau'\operatorname{Cov}(S_n)\tau - \tau'\mathcal{I}_n\tau| \le Mnp_n^{-1}\delta_n|\tau|^2 \qquad \textit{for} \quad n \ge 1 \quad \textit{and} \quad \tau \in \Theta_n.$$

PROOF. Set $a_n = \int s_n(\cdot;\ \tau)f$ and $a_n^* = \int s_n(\cdot;\ \tau)f_n$. Then

$$\tau'\operatorname{Cov}(S_n)\tau = n\int(s_n(\cdot;\ \tau)-a_n)^2 f$$

and

$$\tau'\mathcal{I}_n\tau = n\int(s_n(\cdot;\ \tau)-a_n^*))^2 f_n.$$

The desired result now follows easily from Theorem 1(i) and (12) of Stone (1986).

It follows from (7), Theorem 1(i), Lemma 17 and Lemma 23 that there is a positive constant M such that, for $n \ge 1$ and $y \in I$,

$$|G_n(y)'\mathcal{I}_n^{-1}\operatorname{Cov}(S_n)\mathcal{I}_n^{-1}G_n(y) - G_n(y)'\mathcal{I}_n^{-1}G_n(y)| \le Mn^{-1}p_n\delta_n. \tag{24}$$

It follows from (2), (7), (9) and Lemma 20 that

$$\|\log \hat{f}_n - \log f_n - G_n(\cdot)'\hat{\varphi}_n\|_\infty = o_{pr}((n^{-1}p_n)^{1/2}).$$

Thus by (23) and Lemma 22,

$\mathcal{L}((\log \hat{f}_n - \log f_n)/SD(G_n(y)'\hat{\varphi}_n)) \to \mathcal{N}(0, 1)$ uniformly in y as $n \to \infty$. It follows easily from this together with (2), (22)-(24) and Theorem 1(i) that the first and third conclusions of Theorem 3 are valid.

LEMMA 24. *Uniformly in* $\tau \in \Theta_n$,

$$|(\mathcal{I}_n(\hat{\theta}_n)-\mathcal{I}(\theta_n))\tau|^2 = O_{pr}(np_n^{-1}\log(p_n))|\tau|^2.$$

PROOF. Observe that

$$\frac{\partial^2 C_n}{\partial\theta_j\partial\theta_k}(\hat{\theta}_n) - \frac{\partial^2 C_n}{\partial\theta_j\partial\theta_k}(\theta_n) = \sum_m \int_0^1 \frac{\partial^3 C_n}{\partial\theta_j\partial\theta_k\partial\theta_m}(\theta_n+t(\hat{\theta}_n-\theta_n))(\hat{\theta}_{nm}-\theta_{nm})dt.$$

Thus the desired result follows from Lemmas 15, 19 and 20.

Since $\hat{\mathcal{I}}_n^{-1} - \mathcal{I}_n^{-1} = \mathcal{I}_n^{-1}(\mathcal{I}_n-\hat{\mathcal{I}}_n)\hat{\mathcal{I}}_n^{-1}$, the next result follows from Lemmas 13, 17 and 24.

LEMMA 25. *Uniformly in* $\tau \in \Theta_{n0}$,

$$|(\hat{\mathcal{I}}_n^{-1} - \mathcal{I}_n^{-1})\tau|^2 = O_{pr}(n^{-3}p_n^3\log(p_n))|\tau|^2.$$

LEMMA 26. $|\hat{G}_n(y) - G_n(y)|^2 = O_{pr}(1/n)$ *uniformly in* y.

PROOF. Observe that

$$\hat{G}_n(y) - G_n(y) = -(\nabla C_n(\hat{\theta}_n) - \nabla C_n(\theta_n)).$$

Since

$$\nabla C_n(\hat{\theta}_n) - \nabla C_n(\theta_n) = (\textstyle\int_0^1 H_n(\theta_n+t(\theta_n-\hat{\theta}_n))dt)(\hat{\theta}_n-\theta_n),$$

the desired result follows from Theorem 2(i), Lemma 13 and Lemma 17.

The next result follows from (7), (24), Theorem 1(i), and Lemmas 17, 25, and 26.

LEMMA 27. *Uniformly in* y,

$$\hat{G}_n(y)'\hat{\mathcal{I}}_n^{-1}\hat{G}_n(y) - G_n(y)'\mathcal{I}_n^{-1}\mathrm{Cov}(S_n)\mathcal{I}_n^{-1}G_n(y) = O_{pr}((n^{-3}p_n^3\log(p_n))^{1/2} + n^{-1}p_n\delta_n).$$

The second conclusion of Theorem 3 follows from (2), (22), (23), Theorem 1(i), Theorem 2(v), and Lemma 27.

5. Proof of Theorem 4. For a given value of $x \in [0, 1]$, set

$$g_n(\theta) = \int_0^x f_n(y; \theta)dy.$$

It is easily seen that

$$\nabla g_n(\theta_n) = \int_0^x G_n(y)f_n(y)dy. \tag{25}$$

The next result follows from (2), (9), (10), (13), (25) and Theorem 1(i).

LEMMA 28. $\hat{F}_n(x) - F_n(x) = \nabla g_n(\theta_n)'\hat{\tau}_n + o_{pr}(n^{-1/2})$ *uniformly in* x.

Observe that

$$\mathrm{Var}(\nabla g_n(\theta_n)'\hat{\tau}_n) = \nabla g_n(\theta_n)'\mathcal{I}_n^{-1}\mathrm{Cov}(S_n)\mathcal{I}_n^{-1}\nabla g_n(\theta_n). \tag{26}$$

By (6) and (25),

$$|\nabla g_n(\theta_n)|^2 = O(p_n^{-1}) \quad \text{uniformly in } x. \tag{27}$$

It follows from (26), (27), Theorem 1(i), Lemma 17, and Lemma 23 that

$$\mathrm{Var}(\nabla g_n(\theta_n)'\hat{\tau}_n) = \nabla g_n(\theta_n)'\mathcal{I}_n^{-1}\nabla g_n(\theta_n) + O(n^{-1}\delta_n) \quad \text{uniformly in } x. \tag{28}$$

According to the Cramer-Rao inequality

$$\nabla g_n(\theta_n)'\mathcal{I}_n^{-1}\nabla g_n(\theta_n) \le n^{-1}\mathrm{Var}_{\theta_n}(I_{[0,\ x]}(Y)) = n^{-1}F_n(x)(1-F_n(x)). \tag{29}$$

LEMMA 29. *Suppose that* $p_n \to \infty$ *as* $n \to \infty$. *Then*

$$\lim_n \frac{\nabla g_n(\theta_n)'\mathcal{I}_n^{-1}\nabla g_n(\theta_n)}{n^{-1}F(x)(1-F(x))} = 1 \quad \textit{uniformly for } x \textit{ in compact subsets of } \mathrm{int}(I).$$

PROOF. Choose $\tau_n \in \Theta_{n0}$. By Schwarz's inequality

$$\nabla g_n(\theta_n)'\mathcal{I}_n^{-1}\nabla g_n(\theta_n) \ge \frac{(\nabla g_n(\theta_n)'\tau_n)^2}{\tau_n'\mathcal{I}_n\tau_n}. \tag{30}$$

By (1)

$$\tau_n'\mathcal{I}_n\tau_n = n\ \mathrm{Var}_{\theta_n}(s_n(Y;\ \tau_n)). \tag{31}$$

It follows from (25) that

$$\nabla g_n(\theta_n)'\tau_n = E_{\theta_n}(I_{[0,\ x]}(Y)(s_n(Y;\ \tau_n)-E_{\theta_n}s_n(Y;\ \tau_n)). \tag{32}$$

The desired result follows from (29)-(32), Theorem 1(ii), and the construction of $s \in \mathcal{S}_n$ used in the proof of that result.

The proof of the next result is similar to that of Lemma 22.

LEMMA 30. *Uniformly for* x *in compact subsets of* int(I)

$$\mathcal{L}\left[\frac{\nabla g_n(\boldsymbol{\theta}_n)'\hat{\boldsymbol{\gamma}}_n}{\mathrm{SD}(\nabla g_n(\boldsymbol{\theta}_n)'\hat{\boldsymbol{\gamma}}_n)}\right] \to \mathcal{N}(0,\ 1) \qquad \textit{as} \quad n \to \infty.$$

The first conclusion of Theorem 4 follows from (28) and Lemmas 28-30. The second conclusion follows from the first conclusion and Theorem 1.

REFERENCES

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*, Wiley, New York.

de BOOR, C. (1968). On the uniform approximation by splines. *J. Approx Theory* 1 219-235.

de BOOR, C. (1976). A bound on the L_∞-norm of the L_2-approximation by splines in terms of a global mesh ratio. *Math. Comp.* 30 765-771.

de BOOR, C. (1978). *A Practical Guide to Splines.* Springer-Verlag, New York.

CRAIN, B. R. (1974). Estimation of distributions using orthogonal expansions. *Ann. Statist.* 2, 454-463.

CRAIN, B. R. (1976a). Exponential models, maximum likelihood estimation, and the Haar condition. *J. Amer. Statist. Assoc.* 71, 737-740.

CRAIN, B. R. (1976b). More on estimation of distributions using orthogonal expansions. *J. Amer. Statist. Assoc.* 71, 741-745.

CRAIN, B. R. (1977). An information theoretic approach to approximating a probability distribution. *SIAM J. App. Math.* 32, 339-346.

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

JOHANSEN, S. (1979). *Introduction to the Theory of Regular Exponential Families*, Lecture Notes, 3, The Institute of Mathematical Statistics, University of Copenhagen, Copenhagen.

LEHMANN, E. L. (1983). *Theory of Point Estimation*, Wiley, New York.

LEONARD, T. (1978). Density estimation, stochastic processes and prior information (with discussion). *J. Roy. Statist. Soc. Ser. B* 40 113-146.

NEYMAN, J. (1937). 'Smooth' test for goodness of fit. *Skand. Akt.* 20 149-199.

SCHUMAKER, L. L. (1981). *Spline Functions: Basic Theory.* Wiley, New York.

SILVERMAN, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist.* 10 795-810.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* 8 1348-1360.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* 10 1040-1053.

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

STONE, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* 14 590-606.

STONE, C. J. and KOO, C. Y. (1986). Logspline density estimation. *AMS Contemporary Mathematics Series*, AMS, Providence, to appear.

TITTERINGTON, D. M. (1985). Common structure of smoothing techniques in statistics. *Internat. Statist. Rev.* 53, 141-170.

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

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 estimates. 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. (December 1985, revised Dec. 1986). As others see us: A case study in path analysis. Prepared for the Journal of Educational Statistics.

56. LE CAM, L. and YANG, G. L. (January 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies.

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 polynomials, 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.

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.

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 November 1987). 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. (March 1987). A finite version of de Finetti's theorem for exponential families, with uniform asymptotic estimates.

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). Nonparametric quantile regression with censored data.

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

95. RIZZARDI, F. (Aug 1987). Two-Sample t-tests where one population SD is known.

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 cancer.

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.

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.

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.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987). 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. (September 1987). Large sample theory of estimation in biased sampling regression models I.

116. RITOV, Y. and BICKEL, P.J. (September 1987). Unachievable 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. (October 1987). 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. and MACGIBBON, B. (November 1987). 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). Optimal stopping of two-parameter processes on hyperfinite 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). Vapnik-Červonenkis classes.

130. STONE, C.J. (November 1987). Uniform error bounds involving logspline models.

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

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.

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.