Large-Sample Inference for Logspline Models

By Charles J. Stone
University of California, Berkeley

Technical Report No. 171
September 1988

This research was supported in part by National Science Foundation
Grants MCS83-10257 and DMS-8600409.

Department of Statistics
University of California
Berkeley, California

LARGE-SAMPLE INFERENCE FOR LOGSPLINE MODELS[1]

BY CHARLES J. STONE

University of California, Berkeley

Let f be a continuous and positive unknown density on a known compact interval $\mathcal{I}$. Let F denote the distribution function of f and let $Q = F^{-1}$ denote its quantile function. A finite-parameter exponential family model based on B-splines is constructed. Maximum-likelihood estimation of the parameters of the model based on a random sample of size n from f yields estimators $\hat{f}$, $\hat{F}$ and $\hat{Q}$ of f, F, and Q respectively. Under mild conditions, if the number of parameters tends to infinity in a suitable manner as $n \to \infty$, these estimators achieve the optimal rate of convergence. The asymptotic behavior of the corresponding confidence bounds is also investigated. In particular, it is shown that the standard errors of $\hat{F}$ and $\hat{Q}$ are asymptotically equal to 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 families, B-splines, maximum likelihood, function estimation, rates of convergence, confidence intervals.

1. Introduction. Let Y be a random variable that ranges over a subinterval $\mathcal{I}$ of $\mathbb{R}$ and has unknown density f on $\mathcal{I}$. Let $Y_1,\cdots,Y_n$ be a random sample of size n from the distribution of Y. This random sample can be used for inference about the density of Y and other aspects of the distribution of Y.

The classical parametric approach is to start by "assuming" a fixed parametric model for f involving a J-dimensional vector $\boldsymbol{\theta} \in \Theta$ of unknown parameters. We write the resulting model as $f(\cdot;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. The maximum-likelihood estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained by maximizing the log-likelihood function

$$l(\boldsymbol{\theta}) = \sum_i \log(f(Y_i;\boldsymbol{\theta})), \quad \boldsymbol{\theta} \in \Theta.$$

This estimator is especially attractive when the parametric model for f has the form of an exponential family in which $\boldsymbol{\theta}$ is a natural parameter.

To obtain such a form, choose a J-dimensional vector space $\mathcal{S}$ of functions on $\mathcal{I}$ such that (for identifiability) the zero function on $\mathcal{I}$ is the only function in $\mathcal{S}$ that equals a constant almost everywhere on $\mathcal{I}$. Let $B_1,\cdots,B_J$ be a basis of $\mathcal{S}$. Given the column vector $\boldsymbol{\theta} = (\theta_1,\cdots,\theta_J)^{\mathrm{T}} \in \mathbb{R}^J$, set

$$s(\cdot;\boldsymbol{\theta}) = \sum_1^J \theta_j B_j \quad \text{and} \quad C(\boldsymbol{\theta}) = \log\left[\int_{\mathcal{I}} \exp(s(y;\boldsymbol{\theta}))dy\right].$$

Then

$$\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^J : C(\boldsymbol{\theta}) < \infty\}$$

is a convex subset of $\mathbb{R}^J$, which is assumed to be nonempty and open. The corresponding J-parameter exponential family is given by

$$f(\cdot;\boldsymbol{\theta}) = \exp(s(\cdot;\boldsymbol{\theta}) - C(\boldsymbol{\theta})), \quad \boldsymbol{\theta} \in \Theta.$$

For $\boldsymbol{\theta} \in \Theta$, let $\mathbf{H}(\boldsymbol{\theta})$ denote the Hessian matrix of $C(\cdot)$ at $\boldsymbol{\theta}$, which is the $J \times J$ matrix having entry $\partial^2 C(\boldsymbol{\theta})/\partial\theta_j\partial\theta_k$ in row j and column k for $1 \le j,k \le J$. This matrix is positive definite. Thus $C(\cdot)$ is strictly convex on Θ. The log-likelihood function is given by

$$l(\boldsymbol{\theta}) = \sum_i [s(Y_i;\boldsymbol{\theta}) - C(\boldsymbol{\theta})], \quad \boldsymbol{\theta} \in \Theta.$$

The Hessian matrix of the log-likelihood function of $C(\cdot)$ at $\boldsymbol{\theta}$ is $-\mathbf{I}(\boldsymbol{\theta}) = -n\mathbf{H}(\boldsymbol{\theta})$, where the information matrix $\mathbf{I}(\boldsymbol{\theta})$ is positive definite and hence its inverse $[\mathbf{I}(\boldsymbol{\theta})]^{-1}$ is also

positive definite. Since the log-likelihood function is strictly concave, the maximum-likelihood estimator is unique if it exists.

Let τ be a real-valued parameter depending on f. Under the assumption that f belongs to the indicated exponential family, $\tau = g(\boldsymbol{\theta})$ for some function g on Θ. The maximum-likelihood estimator of τ is given by $\hat{\tau} = g(\hat{\boldsymbol{\theta}})$. Suppose that g is continuously differentiable on Θ. Let $\nabla g(\boldsymbol{\theta})$ denote the gradient of g at $\boldsymbol{\theta}$, which is the J-dimensional column vector whose jth entry is $\partial g(\boldsymbol{\theta})/\partial\theta_j$. The asymptotic standard deviation (ASD) and standard error (SE) of $\hat{\tau}$ are the nonnegative quantities defined by

$$\mathrm{ASD}(\hat{\tau}) = \left[(\nabla g(\boldsymbol{\theta}))^{\mathrm{T}}[\mathbf{I}(\boldsymbol{\theta})]^{-1}\nabla g(\boldsymbol{\theta})\right]^{1/2} \quad \text{and} \quad \mathrm{SE}(\hat{\tau}) = \left[(\nabla g(\hat{\boldsymbol{\theta}}))^{\mathrm{T}}[\mathbf{I}(\hat{\boldsymbol{\theta}})]^{-1}\nabla g(\hat{\boldsymbol{\theta}})\right]^{1/2}.$$

Suppose that $\nabla g(\boldsymbol{\theta}) \neq \mathbf{0}$, where $\boldsymbol{\theta}$ is the true parameter. Then

$$\mathrm{Dist}\left[\frac{\hat{\tau}-\tau}{\mathrm{ASD}(\hat{\tau})}\right] \approx N(0,1) \quad \text{and} \quad \mathrm{Dist}\left[\frac{\hat{\tau}-\tau}{\mathrm{SE}(\hat{\tau})}\right] \approx N(0,1), \quad n \gg 1.$$

Consequently, $\hat{\tau} \pm z_{1-.5\alpha}\mathrm{SE}(\hat{\tau})$ is an asymptotic $(1-\alpha)$-level confidence interval for $\hat{\tau}$. Here $\Phi(z_p) = p$, Φ being the standard normal distribution function.

In the nonparametric approach, we do not assume a fixed finite-parameter model for f, but we can still make use of finite-parameter models as approximations. In order for the corresponding maximum-likelihood estimators and asymptotic confidence bounds for parameters τ defined in terms of f to be reliable, we need the modelling error to tend to zero as $n \to \infty$; for this, it is necessary that the number of parameters tend to infinity as $n \to \infty$.

A thorough study of the use of such approximate models requires a combination of mathematically rigorous asymptotics and computer simulation. In the present paper we concentrate on the asymptotic approach, which is a blend of parametric and nonparametric theory. To carry it out, we require that $\mathcal{I}$ be compact and that f be continuous and positive on $\mathcal{I}$. Also, we restrict our attention to function spaces $\mathcal{S}$ consisting of splines of fixed order q (piecewise polynomials of degree less than q), to bases consisting of B-splines, and (for simplicity) to equally spaced knots. The constant function 1 is in $\mathcal{S}$, so we impose the constraint $\Sigma\theta_j = 0$ and thereby end up with a $(J-1)$-parameter identifiable exponential family. Since $\log(f(\cdot;\boldsymbol{\theta})) \in \mathcal{S}$, we refer to this family as

a *logspline model.*

The main results of this paper and their motivation and application will now be described in an informal manner.

Let $\| \ \|_2$ and $\| \ \|_\infty$ denote the usual L_2 and L_∞ norms of functions on $\mathcal{I}$. Set

$$\delta = \inf_{s \in \mathcal{S}} \|s - \log(f)\|_\infty .$$

As pointed out in Section 2, we know that $\delta = o(1)$ as $n \to \infty$ (under the supposition that $J \to \infty$ as $n \to \infty$). Under appropriate smoothness conditions on f, it is also known that $\delta = O(J^{-p})$, where the positive number p depends on the smoothness condition. In particular, if f is m-times continuously differentiable, where $1 \le m \le q$, then $\delta = O(J^{-m})$.

Set $f^* = f(\cdot;\theta^*)$, where θ^* maximizes the expected value of the log-likelihood function. It is shown in Stone (1989) that $\|f - f^*\|_\infty = O(\delta)$, the proof being based on a similar result of de Boor (1976) involving L_2 projections. Such results are surprising, since minimizing the L_2 error of approximation and maximizing the expected log-likelihood do not appear to be closely related to minimizing the L_∞ error of approximation.

The nonparametric viewpoint suggests looking at parameters that are values of the density function at individual points or of the corresponding distribution function or quantile function. We refer to $\hat{f} = f(\cdot;\hat{\theta})$ as the *logspline density estimator* of f. As a special case, if $\mathcal{S}$ is a collection of piecewise-constant functions on $\mathcal{I}$, then $\hat{f}$ is a histogram density estimator.

For technical reasons, we require that $J = o(n^{.5-\varepsilon})$ for some $\varepsilon > 0$. This is slightly stronger than the assumption $J = o(n^{.5})$ that arises in Portnoy (1986, 1988).

We will show that

$$\|\hat{f} - f^*\|_2 = O_{pr}((J/n)^{1/2}) \quad \text{and} \quad \|\hat{f} - f^*\|_\infty = O_{pr}([J \log(J)/n]^{1/2}).$$

(The result for $\|\hat{f} - f^*\|_2$ is plausible: there are about n/J trials per unknown parameter, so the asymptotic standard deviation of the estimator of each parameter should be proportional to $(J/n)^{1/2}$.) Suppose that $\delta = O(J^{-p})$, where $p > .5$. Set $\gamma = 1/(2p+1)$ and $r = p/(2p+1)$. By choosing $J \sim n^{\gamma}$, we get that $\|\hat{f} - f\|_2 = O_{pr}(n^{-r})$. (Here $a_n \sim b_n$ means that a_n/b_n is bounded away from zero and infinity.) By choosing $J \sim [n/\log(n)]^{\gamma}$, we get

that $\|\hat{f}-f\|_\infty = O_{pr}([\log(n)/n]^r)$. Under suitable specifications, these are the well-known optimal rates of convergence for nonparametric density estimation; see Stone (1980, 1982, 1983).

Let F, F^* and $\hat{F}$ denote the distribution functions of f, f^* and $\hat{f}$ respectively. We will show that

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

Thus, provided that $\delta/J = O(n^{-1/2})$, we get that $\|\hat{F}-F\|_\infty = O_{pr}(n^{-1/2})$, which is well known to be the optimal rate of convergence for estimation of an unknown distribution function in both parametric and nonparametric settings.

Let $\hat{F}^{emp}$ denote the empirical distribution function, defined as usual by

$$\hat{F}^{emp}(y) = \frac{1}{n}\#\{i: 1 \le i \le n \text{ and } Y_i \le y\}, \quad y \in \mathbb{R}.$$

Then

$$SD(\hat{F}^{emp}(y)) = \left[\frac{F(y)[1-F(y)]}{n}\right]^{1/2}, \quad y \in \mathcal{I}.$$

It is well known that, for $n \gg 1$,

$$\text{Dist}\left[\frac{\hat{F}^{emp}(y)-F(y)}{SD(\hat{F}^{emp}(y))}\right] \approx N(0, 1) \quad \text{uniformly on compact subsets of int}(\mathcal{I}),$$

where int($\mathcal{I}$) denotes the interior of $\mathcal{I}$. We will show that, for $n \gg 1$,

$$\text{Dist}\left[\frac{\hat{F}(y)-F^*(y)}{ASD(\hat{F}(y))}\right] \approx N(0,1) \quad \text{uniformly over compact subsets of int}(\mathcal{I})$$

and

$$\frac{ASD(\hat{F}(y))}{SD(\hat{F}^{emp}(y))} \approx 1 \quad \text{uniformly over compact subsets of int}(\mathcal{I}).$$

According to these results, the performance of the parametric estimator $\hat{F}(y)$ is similar to that of the nonparametric estimator $\hat{F}^{emp}(y)$ when $n \gg 1$ (recall that $J \to \infty$ as $n \to \infty$). The results can be used to obtain asymptotic confidence bounds for $F^*(y)$ or $F(y)$ based on the logspline model. They also lead us to conjecture that

$$\|\hat{F} - \hat{F}^{emp}\|_\infty = o_{pr}(n^{-1/2}) \quad \text{if } \delta/J = o(n^{-1/2}).$$

Let $Q = F^{-1}$, $Q^* = (F^*)^{-1}$ and $\hat{Q} = \hat{F}^{-1}$ denote the quantile functions corresponding to f, f^* and $\hat{f}$ respectively. We will show that

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

Thus, provided that $\delta/J = O(n^{-1/2})$, we get that $\|\hat{Q}-Q\|_\infty = O_{pr}(n^{-1/2})$, which again is the optimal rate of convergence.

Let $\hat{Q}^{emp}$ denote the empirical quantile function as usually defined (for example, $\hat{Q}^{emp}(.5)$ is the sample median). It is well known that, for $n \gg 1$,

$$\text{Dist}\left[\frac{\hat{Q}^{emp}(p)-Q(p)}{\text{ASD}(\hat{Q}^{emp}(p))}\right] \approx N(0,1) \quad \text{uniformly over compact subsets of } (0, 1),$$

where

$$\text{ASD}(\hat{Q}^{emp}(p)) = \left[\frac{p(1-p)}{nf^2(Q(p))}\right]^{1/2}.$$

We will show that, for $n \gg 1$,

$$\text{Dist}\left[\frac{\hat{Q}(p)-Q^*(p)}{\text{ASD}(\hat{Q}(p))}\right] \approx N(0,1) \quad \text{uniformly over compact subsets of } (0, 1)$$

and

$$\frac{\text{ASD}(\hat{Q}(p))}{\text{ASD}(\hat{Q}^{emp}(p))} \approx 1 \quad \text{uniformly over compact subsets of } (0, 1).$$

These results can be used to obtain asymptotic confidence bounds for $Q^*(p)$ or $Q(p)$ based on the logspline model.

In parallel with the analytic approach and in continuation of the work of Stone and Koo (1986a, 1986b), Charles Kooperberg and I are currently using computer simulation to determine the finite-sample performance of inference based on logspline models. An important advantage of the computational approach is that attractive but mathematically unwieldy modifications can be studied. In our investigation, we have focussed on positive random variables Y ($\mathcal{I} = (0, \infty)$). By using a suitable data-dependent transformation, which behaves like a power transformation at infinity and a logarithmic transformation at the origin, we first transform Y to a real-valued random variable ($\mathcal{I} = (-\infty, \infty)$). So far, we have used cubic splines. The first knot has been placed at the minimum value in the random sample $Y_1,\cdots,Y_n$, the last knot has been placed at the maximum value, and a number of intermediate knots have been placed at selected order statistics of the random sample. Also, the selection of the number of intermediate knots and the order statistics used in placing them has been made in a data-independent manner. To reduce the standard errors of logspline estimators of extreme quantiles, we

have imposed linear restrictions on the fitted splines to the left of the first knot and to the right of the last knot. Thus the estimated distribution of the transformed random variable has exponential tails.

We have focussed on confidence bounds for quantiles, especially on the 90% upper confidence bound (UCB) for $Q(p)$ with $p \approx 1$. As applied to the transformed random variables, such confidence bounds are of the form

$$\hat{Q}(p) + t\, \mathrm{SE}(\hat{Q}(p)),$$

where the indicated standard error is obtained by ignoring the data-dependence of the preliminary transformation and the knot selection. In order to achieve a satisfactory approximation to the desired coverage probability over a broad range of n, p, and underlying distributions, we have chosen t by adaption to the exponential distribution (using computer simulation as in the use of the bootstrap) instead of using the value $t = z_{.9} \approx 1.28$ suggested by normal approximation.

The results obtained to date confirm the necessity of having $J \to \infty$ as $n \to \infty$. The performance of the nominal 90% logspline UCB's for extreme quantiles ($1-p = 1/n$, $.1/n$, or $.01/n$ with $n = 50$ or 500) is satisfactory, although perhaps not quite as good as that of the best procedures that emerged in the fairly extensive computer simulation study by Breiman, Stone and Kooperberg (1988). As expected, if there is an order statistic that serves as a 90% UCB for $Q(p)$ (which, for $n \gg 1$, can only happen if $1-p > 2/n$), then the performance of the UCB based on the logspline model is very similar to that of the order statistic.

Some advantages of inference based on logspline models over that based on kernel density estimators are as follows: logspline density estimators are automatically positive and can be chosen to achieve the optimal rate of convergence n^{-r} for values of r that are arbitrarily close to .5, logspline estimators of extreme quantiles are sensible, the corresponding UCB's are reliable, and standard inferential tools for treating exponential families are applicable. A disadvantage of logspline models based on quadratic and higher-order splines is that the evaluation of $C(\theta)$ and its partial derivatives requires computationally intensive numerical integration.

The logspline approach could probably be extended to handle multidimensional distributions by using tensor-product B-splines or certain function spaces arising in connection with the finite-element method. In three or more dimensions, however, the approach may be impractical.

In Stone (1985), the goal was to achieve the optimal rate of convergence for nonparametric estimation of the best additive approximation to the regression function; it was realized by using the least-squares method to fit an additive spline. The setup of the follow-on paper, Stone (1986), included additive logistic regression and other nonparametric extensions of generalized linear models (see McCullagh and Nelder, 1983). The goal there was to achieve the optimal rate of convergence for nonparametric estimation of the best additive approximation to the response function. This goal was realized by using maximum likelihood to fit an additive spline. Some of the results in Stone (1985, 1986) and the present paper are summarized in Stone (1987).

It would be worthwhile to include the setup of Stone (1986) and that of the present paper in a common framework. Consider, in particular, a random variable Y whose density depends on a real-valued variable x. It seems reasonable to use a logspline model $f(\cdot;\boldsymbol{\theta}_x)$ based on linear splines to approximate the unknown density, where the dependence of each entry of $\boldsymbol{\theta}_x$ on x is approximated by a cubic spline. Tensor-product B-splines arise naturally in this manner. Stone (1989) contains a start at the mathematical analysis of this approach (namely, the extension of Theorem 1(i) in the next section).

Logspline models are flexible exponential families. Previously, Neyman (1937) and Crain (1974, 1976a, 1976b, 1977) considered other such families. Leonard (1978) and Silverman (1982) considered various nonparametric estimators of $\log(f)$, an attractive feature being that the corresponding estimator of f itself is automatically positive.

Precise results will be stated in the next section and proven thereafter. The proofs use refinements of the known analytic properties of splines that have been developed in Stone (1985, 1986, 1989). They also use refinements of standard techniques for determining the asymptotic behavior of maximum-likelihood estimators of the unknown parameters of an exponential family that are applicable when the number of parameters

tends to infinity as $n \to \infty$. Some of these refinements were developed in Stone (1986). Portnoy (1988) has also studied the asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. It might be interesting to compare Portnoy's approach to that of Stone (1986) and the present paper.

2. Statement of Results. Let $Y_1, Y_2 \cdots$ be independent and identically distributed random variables taking on values in a known compact interval $\mathcal{I}$ having positive length; without loss of generality, let $\mathcal{I} = [0, 1]$. These random variables are assumed to have a continuous and positive density f on $\mathcal{I}$. Let F denote the distribution function of f and let $Q = F^{-1}$ denote its quantile function. Given the positive integer n, we refer to $Y_1, \cdots, Y_n$ as the random sample of size n.

For simplicity in notation, we suppress the dependence on n of various quantities after these quantities are defined.

Let q denote a positive integer. Given $n \geq 1$, let $K = K_n$ denote a positive integer. Let $\mathcal{I}$ be partitioned into subintervals

$$\mathcal{I}_k = [(k-1)/K, k/K), \ 1 \leq k < K, \quad \text{and} \quad \mathcal{I}_K = [(K-1)/K, 1].$$

Let $\mathcal{S} = \mathcal{S}_n$ denote the collection of functions s on $\mathcal{I}$ satisfying the following two properties: s is a polynomial of order q (degree less than q) on each of the subintervals $\mathcal{I}_1, \cdots, \mathcal{I}_K$; if $q \geq 2$, s is $(q-2)$-times continuously differentiable on $\mathcal{I}$. Then $\mathcal{S}$ is a vector space of dimension $J = J_n = q+K-1$, which is a space of polynomial splines of order q with simple knots at k/K for $1 \leq k < K$. The functions in $\mathcal{S}$ are piecewise-constant, linear, quadratic or cubic splines according as $q = 1, 2, 3,$ or 4. Consider the usual B-spline basis $B_j = B_{nj}$, $1 \leq j \leq J$, of $\mathcal{S}$ (see de Boor, 1978). These functions are nonnegative and sum to 1 on $\mathcal{I}$. Also, there is a fixed positive integer J_0, depending on q but not on n, such that (i) the support of each B_j is contained in the convex hull of J_0 consecutive knots and (ii) if $|k-j| > J_0$, the supports of B_j and B_k are disjoint. (The support of B_j is the subset of $\mathcal{I}$ on which $B_j > 0$.)

Let $\Theta = \Theta_n$ denote the collection of all J-dimensional (column) vectors. Given $\theta \in \Theta$, set

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

$$s(\cdot;\theta) = s_n(\cdot;\theta) = \sum_j \theta_j B_j,$$

$$C(\theta) = C_n(\theta) = \log\left[\int \exp(s(\cdot;\theta))\right],$$

and

$$f(\cdot;\theta) = f_n(\cdot;\theta) = \exp(s(\cdot;\theta) - C(\theta)).$$

Then

$$\int f(\cdot;\theta) = 1, \quad \theta \in \Theta.$$

Let $F(\cdot;\theta) = F_n(\cdot;\theta)$ and $Q(\cdot;\theta) = Q_n(\cdot;\theta)$ denote the distribution function and quantile function corresponding to $f(\cdot;\theta)$. Set

$$\Lambda(\theta) = \Lambda_n(\theta) = E\left[\log(f(Y;\theta))\right] = \int \log(f(\cdot;\theta)) f = \int s(\cdot;\theta) f - C(\theta), \quad \theta \in \Theta,$$

where Y has density f.

It is assumed from now on that $J \geq 2$ for all n. The exponential family $f(\cdot;\theta)$, $\theta \in \Theta$, is not identifiable, for if we add a constant to each element of θ, we do not change $f(\cdot;\theta)$. Let $\Theta_0 = \Theta_{n0}$ denote the $(J-1)$-dimensional subspace of Θ consisting of those vectors $\theta \in \Theta$ whose entries add up to zero.

Let $H(\theta) = H_n(\theta)$ denote the Hessian matrix of $C(\cdot)$ at θ, which is the $J \times J$ matrix having entry $\partial^2 C(\theta)/\partial\theta_j \partial\theta_k$ in row j and column k for $1 \leq j,k \leq J$. It is an elementary and well known property of exponential families (see Lehmann, 1983) that if $\theta, \tau \in \Theta$, then

$$\tau^T H(\theta)\tau = \int [s(\cdot;\tau) - a]^2 f(\cdot;\theta), \quad \text{where } a = \int s(\cdot;\tau) f(\cdot;\theta). \tag{1}$$

Thus $\tau^T H(\theta)\tau > 0$ if τ is a nonzero element of Θ_0. Consequently, $C(\cdot)$ is strictly convex on Θ_0. Since $-H(\theta)$ is the Hessian matrix of $\Lambda(\cdot)$ at θ, $\Lambda(\cdot)$ is strictly concave on Θ_0. If $\theta \in \Theta_0$ and $\theta \neq 0$, then

$$\Lambda(t\theta) = t\int s(\cdot;\theta) f - \log\left[\int \exp(ts(\cdot;\theta))\right]$$

and $s(\cdot;\theta)$ is not almost everywhere equal to a constant on $\mathcal{I}$; thus $\Lambda(t\theta) \to -\infty$ as $|t| \to \infty$ (consider the maximum value of $s(\cdot;\theta)$ on $\mathcal{I}$). It follows that for each $n \geq 1$, there is a

unique $\boldsymbol{\theta}^* = \boldsymbol{\theta}_n^* \in \Theta_0$ that maximizes $\Lambda(\cdot)$ on Θ_0. Set $f^* = f_n^* = f(\cdot;\boldsymbol{\theta}^*)$, $F^* = F_n^* = F(\cdot;\boldsymbol{\theta}^*)$ and $Q^* = Q_n^* = Q(\cdot;\boldsymbol{\theta}^*)$.

It follows from the assumption on f that $\log(f)$ is continuous and hence bounded on $\mathcal{I}$. Set

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

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

THEOREM 1. (i) $\|f^* - f\|_\infty = O(\delta)$;

(ii) $\|F^* - F\|_\infty = O(\delta/J)$;

and

(iii) $\|Q^* - Q\|_\infty = O(\delta/J)$.

Let $l(\cdot) = l_n(\cdot)$ be the log-likelihood function based on the logspline model and the random sample of size n, which is given by

$$l(\boldsymbol{\theta}) = \sum_i \log\left[f(Y_i;\boldsymbol{\theta})\right] = \sum_i \left[s(Y_i;\boldsymbol{\theta}) - C(\boldsymbol{\theta})\right].$$

Then $l(\cdot)$ is a strictly concave function on Θ_0. Let $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_n \in \Theta_0$ denote the maximum-likelihood estimator of $\boldsymbol{\theta} \in \Theta_0$ based on the random sample of size n. Then $\hat{\boldsymbol{\theta}}$ is unique if it exists. (The log-likelihood function has a maximum if and only if there is no nonconstant function $s \in \mathcal{S}$ such that $Y_1, \cdots, Y_n$ all maximize s.) Set $\hat{f} = \hat{f}_n = f(\cdot;\hat{\boldsymbol{\theta}})$, $\hat{F} = \hat{F}_n = F(\cdot;\hat{\boldsymbol{\theta}})$ and $\hat{Q} = \hat{Q}_n = Q(\cdot;\hat{\boldsymbol{\theta}})$.

From now on it is assumed that

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

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

(ii) $|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*| = O_{pr}(J/n^{1/2})$;

(iii) $\max_{1\le j\le J} |\hat{\theta}_j-\theta_j^*| = O_{pr}([J \log(J)/n]^{1/2})$;

(iv) $\|\hat{f}-f^*\|_2 = O_{pr}([J/n]^{1/2})$;

(v) $\|\hat{f}-f^*\|_\infty = O_{pr}([J \log(J)/n]^{1/2})$;

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

and

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

Let $\mathbf{I}(\boldsymbol{\theta}) = \mathbf{I}_n(\boldsymbol{\theta}) = n\mathbf{H}(\boldsymbol{\theta})$ denote the information matrix based on the random sample of size n and let $\mathbf{I}^{-1}(\boldsymbol{\theta}) = \mathbf{I}_n^{-1}(\boldsymbol{\theta})$ denote the inverse to $\mathbf{I}(\boldsymbol{\theta})$ viewed as a linear transformation on Θ_0. Set $\mathbf{I}^* = \mathbf{I}_n^* = \mathbf{I}(\boldsymbol{\theta}^*)$, $(\mathbf{I}^*)^{-1} = (\mathbf{I}_n^*)^{-1} = \mathbf{I}^{-1}(\boldsymbol{\theta}^*)$, $\hat{\mathbf{I}} = \hat{\mathbf{I}}_n = \mathbf{I}(\hat{\boldsymbol{\theta}})$, and $\hat{\mathbf{I}}^{-1} = \hat{\mathbf{I}}_n^{-1} = \mathbf{I}^{-1}(\hat{\boldsymbol{\theta}})$. Given $y \in \mathcal{I}$, let $\mathbf{G}^*(y) = \mathbf{G}_n^*(y) \in \Theta_0$ and $\hat{\mathbf{G}}(y) = \hat{\mathbf{G}}_n(y) \in \Theta_0$ denote the J-dimensional vectors having elements

$$G_j^*(y) = B_j(y) - \frac{\partial C}{\partial \theta_j}(\boldsymbol{\theta}^*) \quad \text{and} \quad \hat{G}_j(y) = B_j(y) - \frac{\partial C}{\partial \theta_j}(\hat{\boldsymbol{\theta}}), \quad 1 \le j \le J.$$

respectively. The asymptotic standard deviation (ASD) and standard error (SE) of $\hat{f}(y)$ are defined by

$$\mathrm{ASD}(\hat{f}(y)) = f^*(y)\left[[\mathbf{G}^*(y)]^{\mathrm{T}}(\mathbf{I}^*)^{-1}\mathbf{G}^*(y)\right]^{1/2}$$

and

$$\mathrm{SE}(\hat{f}(y)) = \hat{f}(y)\left[[\hat{\mathbf{G}}(y)]^{\mathrm{T}}\hat{\mathbf{I}}^{-1}\hat{\mathbf{G}}(y)\right]^{1/2}.$$

THEOREM 3. *Suppose that* $J \to \infty$ *as* $n \to \infty$. *Then, uniformly in* $y \in \mathcal{I}$,

$$\mathrm{ASD}(\hat{f}(y)) \sim (J/n)^{1/2},$$

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

and

$$\mathrm{Dist}\left[\frac{\hat{f}(y)-f^*(y)}{\mathrm{ASD}(\hat{f}(y))}\right] \to N(0,1).$$

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

$$\frac{\mathrm{ASD}(\hat{F}(y))}{\mathrm{SD}(\hat{F}^{\mathrm{emp}}(y))} \to 1 \quad \textit{uniformly on compact subsets of } \mathrm{int}(\mathcal{I}),$$

$$\mathrm{Dist}\left[\frac{\hat{F}(y)-F^*(y)}{\mathrm{ASD}(\hat{F}(y))}\right] \to N(0,1) \quad \textit{uniformly on compact subsets of } \mathrm{int}(\mathcal{I}),$$

$$\frac{\mathrm{ASD}(\hat{Q}(p))}{\mathrm{ASD}(\hat{Q}^{\mathrm{emp}}(p))} \to 1 \quad \textit{uniformly on compact subsets of } (0, 1).$$

and

$$\mathrm{Dist}\left[\frac{\hat{Q}(t)-Q^*(t)}{\mathrm{ASD}(\hat{Q}(t))}\right] \to N(0,1) \quad \textit{uniformly on compact subsets of } (0, 1).$$

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

3. Proof of Theorem 1.

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

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

The proof of Theorem 1(i) is contained in Stone (1989). We now verify Theorem 1(ii), from which Theorem 1(iii) follows easily. It can be assumed that $J \to \infty$ as $n \to \infty$. It will be shown below that there is a positive constant M satisfying the following condition: for $n \geq 1$ and $x \in [2MJ^{-1}, 1-MJ^{-1}]$ there is an $s \in \mathcal{S}$ such that $-1 \leq s \leq 1$,

$s = 1$ on $[MJ^{-1}, x-MJ^{-1}]$, and $s = 0$ on $[x + MJ^{-1}, 1]$. According to Lemma 1,

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

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

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

$$s(y) = \int_0^y [c_1 B_1(z) - c_2 B_2(z)]dz, \quad 0 \le y \le 1,$$

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

4. Proof of Theorem 2.

LEMMA 2. *For any positive number M there is a $\delta > 0$ such that if f and g are positive densities on $\mathcal{I}$, $\|g\|_\infty \le M$ 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 $a_0 = \int[\log(f)-\log(g)]$. 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*

$$|C(\theta)| \le M\|\log(f(\cdot;\theta))\|_\infty, \quad \theta \in \Theta_0.$$

PROOF. Since

$$\log(f(\cdot;\boldsymbol{\theta})) = \sum_j [\theta_j - C(\boldsymbol{\theta})]B_j, \quad \boldsymbol{\theta} \in \Theta,$$

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 $n \geq 1$, $\boldsymbol{\theta} \in \Theta$ and $\|\log(f_n(\cdot;\boldsymbol{\theta})) - \log(f_n)\|_2 \leq MJ^{-1/2}$, then*

$$\|\log(f(\cdot;\boldsymbol{\theta}^* + t(\boldsymbol{\theta} - \boldsymbol{\theta}^*))\|_\infty \leq M_1, \quad 0 \leq t \leq 1,$$

and

$$\Lambda(\boldsymbol{\theta}) - \Lambda(\boldsymbol{\theta}^*) \leq -\delta \|\log(f(\cdot;\boldsymbol{\theta})) - \log(f^*)\|_2^2.$$

PROOF. Without loss of generality, it can be assumed that $\boldsymbol{\theta} \in \Theta_0$. 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 $\boldsymbol{\theta}$ are as in the statement of the present lemma, then $\|\log(f^*)\|_\infty \leq M_2$ and $\|\log(f(\cdot;\boldsymbol{\theta}))\|_\infty \leq M_2$. Thus by Lemma 3 there is a positive constant M_3 depending on M such that

$$\|s(\cdot;\boldsymbol{\theta}^* + t(\boldsymbol{\theta} - \boldsymbol{\theta}^*))\|_\infty \leq M_3, \quad 0 \leq t \leq 1,$$

which yields the first conclusion of the lemma. Observe that

$$\frac{d}{dt} \Lambda(\boldsymbol{\theta}^* + t(\boldsymbol{\theta} - \boldsymbol{\theta}^*))\big|_{t=0} = 0.$$

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

$$\Lambda(\boldsymbol{\theta}) - \Lambda(\boldsymbol{\theta}^*) = -\frac{1}{2}\int [s(\cdot;\boldsymbol{\theta} - \boldsymbol{\theta}^*) - a]^2 f(\cdot;\boldsymbol{\theta}^* + t(\boldsymbol{\theta} - \boldsymbol{\theta}^*))$$

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*

$$\Pr\left[\left|\frac{l(\boldsymbol{\theta}) - l(\boldsymbol{\theta}^*)}{n} - [\Lambda(\boldsymbol{\theta}) - \Lambda(\boldsymbol{\theta}^*)]\right| \geq c\|\log(f(\cdot;\boldsymbol{\theta})) - \log(f^*)\|_2\right] \leq 2e^{-\delta n c^2}$$

for $n \geq 1$, $\boldsymbol{\theta} \in \Theta$, and $0 < c \leq MJ^{-1/2}$.

PROOF. Write

$$l(\boldsymbol{\theta}) - l(\boldsymbol{\theta}^*) - n[\Lambda(\boldsymbol{\theta}) - \Lambda(\boldsymbol{\theta}^*)] = \sum_i Z_i,$$

where

$$Z_i = \log(f(Y_i;\boldsymbol{\theta})) - \log(f^*(Y_i)) - E[\log(f(Y_i;\boldsymbol{\theta})) - \log(f^*(Y_i))].$$

Set $a = \|\log(f(\cdot;\theta)) - \log(f^*)\|_2$. By Lemma 7 of Stone (1986) there is a positive number M_1 such that $\|\log(f(\cdot;\theta)) - \log(f^*)\|_\infty \le M_1 a J^{1/2}$. Thus there is a positive constant M_2 such that $|Z_i| \le M_2 a J^{1/2}$ and $\mathrm{Var}(Z_i) \le M_2 a^2$. 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* $\theta_1, \theta_2 \in \Theta$, *then*

$$\left| \frac{l(\theta_2) - l(\theta_1)}{n} - [\Lambda(\theta_2) - \Lambda(\theta_1)] \right| \le 2\|\log(f(\cdot;\theta_2)) - \log(f(\cdot;\theta_1))\|_\infty.$$

Set $\mathcal{F} = \mathcal{F}_n = \{f(\cdot;\theta): \theta \in \Theta\}$. It is convenient to define the "diameter" of a subset $\mathcal{A}$ of $\mathcal{F}$ as $\sup\{\|\log(f_2) - \log(f_1)\|_\infty : f_1, f_2 \in \mathcal{F}\}$. 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* $\varepsilon > 0$ *and* $\delta > 0$, *there is an* $M > 0$ *such that the following is valid*:

$$\left\{ f(\cdot;\theta): \theta \in \Theta \text{ and } \|\log(f(\cdot;\theta)) - \log(f^*)\|_2 \le n^{\varepsilon}(J/n)^{1/2} \right\}$$

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

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

$$\|\log(\hat{f}) - \log(f^*)\|_2 = O_{pr}(n^{\varepsilon}[J/n]^{1/2}), \quad \varepsilon > 0.$$

PROOF. Set $c = c_n = n^{\varepsilon}(J/n^{-1})^{1/2}$ and

$$\Theta_1 = \Theta_{n1} = \{\theta \in \Theta_0: \|\log(f(\cdot;\theta)) - \log(f^*)\|_2 \le c\}.$$

Then Θ_1 is a compact set whose boundary, relative to Θ_0, is contained in

$$\Theta_2 = \Theta_{n2} = \{\theta \in \Theta_0: \|\log(f(\cdot;\theta)) - \log(f^*)\|_2 = c\}.$$

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

$$\Lambda(\theta) - \Lambda(\theta^*) \le -\delta c^2, \quad \theta \in \Theta_2.$$

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

$$l(\theta) < l(\theta^*), \quad \theta \in \Theta_2,$$

and hence $l(\cdot)$ has a local maximum in the relative interior of Θ_1. The desired conclusions now follow from the strict concavity of $l(\theta)$ on Θ_0. (The first conclusion of

the lemma can also be obtained from the necessary and sufficient condition for the log-likelihood function to have a maximum that was mentioned in Section 2.)

Let $\mathbf{S}(\boldsymbol{\theta}) = \mathbf{S}_n(\boldsymbol{\theta}) \in \Theta_0$ denote the gradient of $l(\cdot)$ at $\boldsymbol{\theta}$; that is, the J–dimensional vector whose jth entry is

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_j} = \sum_i \left[B_j(Y_i) - \frac{\partial C}{\partial \theta_j}(\boldsymbol{\theta}) \right].$$

Set $\mathbf{S}^* = \mathbf{S}_n^* = \mathbf{S}(\boldsymbol{\theta}^*)$. Then $E\mathbf{S}^* = 0$ and

$$E|\mathbf{S}^*|^2 = n\sum_j \mathrm{Var}\left[B_j(Y)\right] \le n\sum_j EB_j^2(Y) = nE\sum_j B_j^2(Y) = O(n).$$

Consequently, the following result is valid.

LEMMA 9. $|\mathbf{S}^*|^2 = O_{pr}(n)$.

The maximum-likelihood equation for $\hat{\boldsymbol{\theta}}$ is $\mathbf{S}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, which can be rewritten as $\mathbf{D}(\hat{\boldsymbol{\theta}}–\boldsymbol{\theta}^*) = \mathbf{S}^*$, where $\mathbf{D} = \mathbf{D}_n$ is the $J \times J$ matrix defined by

$$\mathbf{D} = n\int_0^1 \mathbf{H}(\boldsymbol{\theta}^*+t(\hat{\boldsymbol{\theta}}–\boldsymbol{\theta}^*))dt.$$

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

$$(\hat{\boldsymbol{\theta}}–\boldsymbol{\theta}^*)^T\mathbf{D}(\hat{\boldsymbol{\theta}}–\boldsymbol{\theta}^*) \ge \delta n\|\log(\hat{f})– \log(f^*)\|_2^2.$$

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

$$\max_{0\le t\le 1} \|\log(f(\cdot;\boldsymbol{\theta}^*+t(\hat{\boldsymbol{\theta}}–\boldsymbol{\theta}^*)))\|_\infty = O_{pr}(1).$$

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

LEMMA 11. $(\hat{\boldsymbol{\theta}}–\boldsymbol{\theta}^*)^T\mathbf{S}^* = O_{pr}\left[(nJ)^{1/2}\right]\|\log(\hat{f})–\log(f^*)\|_2$.

PROOF. Let 1 denote the J-dimensional vector each of whose coordinates is 1. Then $1^T\mathbf{S}^* = 0$ since $\mathbf{S}^* \in \Theta_0$. Now

$$\log(\hat{f}) – \log(f^*) = \sum_j [\hat{\theta}_j–\theta_j^*–C(\hat{\boldsymbol{\theta}})+C(\boldsymbol{\theta}^*)]B_j, \tag{3}$$

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

$$\begin{aligned} |(\hat{\boldsymbol{\theta}}–\boldsymbol{\theta}^*)^T\mathbf{S}^*|^2 &= |[\hat{\boldsymbol{\theta}}–\boldsymbol{\theta}^*–(C(\hat{\boldsymbol{\theta}})–C(\boldsymbol{\theta}^*))1]^T\mathbf{S}^*|^2 \\ &\le |(\hat{\boldsymbol{\theta}}–\boldsymbol{\theta}^*)–(C(\hat{\boldsymbol{\theta}})–C(\boldsymbol{\theta}^*))1|^2|\mathbf{S}^*|^2 \\ &= O_{pr}(nJ)\|\log(\hat{f})–\log(f^*)\|_2^2 \end{aligned}$$

as desired.

LEMMA 12. (i) $\|\log(\hat{f})-\log(f^*)\|_2 = O_{pr}([J/n]^{1/2})$;

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

(iii) $\|\hat{f}-f^*\|_2 = O_{pr}([J/n]^{1/2})$;

and

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

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

$$(\hat{\theta}-\theta^*)^T D(\hat{\theta}-\theta^*) = (\hat{\theta}-\theta^*)^T S^*.$$

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 $(\hat{\theta}-\theta^*)^T 1 = 0$, it follows from (3), the first result, and (12) of Stone (1986) that

$$|\hat{\theta}-\theta^*|^2 + J[C(\hat{\theta})-C(\theta^*)]^2 = |\hat{\theta}-\theta^*+[C(\hat{\theta})-C(\theta_n^*)]1|^2$$

$$= O\left[J\|\log(\hat{f})-\log(f^*)\|_2^2\right] = O_{pr}(J^2/n)$$

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,*

$$M_1 \le f(\cdot;\theta^*+t(\hat{\theta}-\theta^*)) \le M_2, \quad 0 \le t \le 1.$$

Let $\nabla C(\theta)$ denote the gradient of $C(\cdot)$ at θ; that is, the J-dimensional vector having entries $\partial C(\theta)/\partial\theta_j$, $1 \le j \le J$.

LEMMA 14. $C(\hat{\theta}) - C(\theta^*) = [\nabla C(\theta^*)]^T(\hat{\theta}-\theta^*) + O_{pr}(J/n)$.

PROOF. Observe that

$$C(\hat{\theta}) - C(\theta^*) = [\nabla C(\theta^*)]^T(\hat{\theta}-\theta^*) + (\hat{\theta}-\theta^*)^T R(\hat{\theta}-\theta^*),$$

where $R = R_n$ is the $J \times J$ matrix defined by

$$R = \int_0^1 (1-t)H(\theta^*+t(\hat{\theta}-\theta^*))dt.$$

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,*

$$\sum_j \left[\sum_k \sum_m \max_{0\le t\le 1} \left|\frac{\partial^3 C}{\partial\theta_j\partial\theta_k\partial\theta_m}\left[\boldsymbol{\theta}^*+t(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)\right]\right| |\tau_k|\right]^2 \le MJ^{-2}|\boldsymbol{\tau}|^2, \quad \boldsymbol{\tau}\in\Theta.$$

PROOF. It is easily seen that

$$\begin{aligned}\frac{\partial^3 C(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k\partial\theta_m} &= \int B_jB_kB_m f(\cdot;\boldsymbol{\theta}) \\ &\quad - \int B_jB_k f(\cdot;\boldsymbol{\theta}) \int B_m f(\cdot;\boldsymbol{\theta}) \\ &\quad - \int B_jB_m f(\cdot;\boldsymbol{\theta}) \int B_k f(\cdot;\boldsymbol{\theta}) \\ &\quad - \int B_kB_m f(\cdot;\boldsymbol{\theta}) \int B_j f(\cdot;\boldsymbol{\theta}) \\ &\quad + \int B_j f(\cdot;\boldsymbol{\theta}) \int B_k f(\cdot;\boldsymbol{\theta}) \int B_m f(\cdot;\boldsymbol{\theta}).\end{aligned}$$

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

It follows easily from Theorem 1(i) that

$$\max_{1\le j\le J} \left|\frac{\partial C}{\partial\theta_j}(\boldsymbol{\theta}^*)\right| = O(J^{-1}) \tag{4}$$

and hence that

$$|\mathbf{G}^*(y)| \sim 1 \text{ uniformly for } y\in I. \tag{5}$$

The next result follows from (3) and Lemma 14.

LEMMA 16. $\left\|\log(\hat{f})-\log(f^*)-[\mathbf{G}^*(\cdot)]^{\mathrm{T}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)\right\|_\infty = O_{\mathrm{pr}}(J/n)$.

For $\boldsymbol{\theta}\in\Theta$ let $\min f(\cdot;\boldsymbol{\theta})$ and $\max f(\cdot;\boldsymbol{\theta})$ respectively denote the minimum and maximum values of $f(\cdot;\boldsymbol{\theta})$.

LEMMA 17. *For each $\boldsymbol{\theta}\in\Theta$, $\mathbf{H}(\boldsymbol{\theta})$ is a positive definite symmetric linear transformation on Θ_0. There are positive constants M_1 and M_2 such that*

$$M_1J^{-1}|\boldsymbol{\tau}|^2 \min f(\cdot;\boldsymbol{\theta}) \le \boldsymbol{\tau}^{\mathrm{T}}\mathbf{H}(\boldsymbol{\theta})\boldsymbol{\tau} \le M_2J^{-1}|\boldsymbol{\tau}|^2 \max f(\cdot;\boldsymbol{\theta})$$

for $n\ge 1$, $\boldsymbol{\theta}\in\Theta$ and $\boldsymbol{\tau}\in\Theta_0$.

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

$$M_1J^{-1}|\boldsymbol{\tau}|^2 \le \|s(\cdot;\boldsymbol{\tau})\|_2^2 \le M_2J^{-1}|\boldsymbol{\tau}|^2, \quad \boldsymbol{\tau}\in\Theta.$$

It follows from (1) that

$$\boldsymbol{\tau}^{\mathrm{T}}\mathbf{H}(\boldsymbol{\theta})\boldsymbol{\tau} \le \|s(\cdot;\boldsymbol{\tau})\|_2^2 \max f(\cdot;\boldsymbol{\theta}) \le M_2 J^{-1}|\boldsymbol{\tau}|^2 \max f(\cdot;\boldsymbol{\theta})$$

for $\boldsymbol{\tau} \in \Theta$. It also follows from (1) that if $\boldsymbol{\tau} \in \Theta_0$, then

$$\begin{aligned}\boldsymbol{\tau}^{\mathrm{T}}\mathbf{H}(\boldsymbol{\theta})\boldsymbol{\tau} &\ge \|s(\cdot;\boldsymbol{\tau}-a\mathbf{1})\|_2^2 \min f(\cdot;\boldsymbol{\theta}) \\ &\ge M_1 J^{-1}|\boldsymbol{\tau}-a\mathbf{1}|^2 \min f(\cdot;\boldsymbol{\theta}) \\ &\ge M_1 J^{-1}|\boldsymbol{\tau}|^2 \min f(\cdot;\boldsymbol{\theta}).\end{aligned}$$

Thus the conclusion of the lemma is valid.

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

$$M_1 n J^{-1}|\boldsymbol{\tau}|^2 \le \boldsymbol{\tau}^{\mathrm{T}}\mathrm{Cov}(\mathbf{S}^*)\boldsymbol{\tau} \le M_2 n J^{-1}|\boldsymbol{\tau}|^2, \quad n \ge 1 \text{ and } \boldsymbol{\tau} \in \Theta_0.$$

PROOF. Since

$$\boldsymbol{\tau}^{\mathrm{T}}\mathrm{Cov}(\mathbf{S}^*)\boldsymbol{\tau} = n\int[s(\cdot;\boldsymbol{\tau})-a]^2 f,$$

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

Consider the approximation $\hat{\boldsymbol{\varphi}} = \hat{\boldsymbol{\varphi}}_n \in \Theta_0$ to $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*$ defined by $\mathbf{I}^*\hat{\boldsymbol{\varphi}} = \mathbf{S}^*$. Then $\hat{\boldsymbol{\varphi}} = (\mathbf{I}^*)^{-1}\mathbf{S}^*$ and hence $[\mathbf{G}^*(y)]^{\mathrm{T}}\hat{\boldsymbol{\varphi}} = [\mathbf{G}^*(y)]^{\mathrm{T}}(\mathbf{I}^*)^{-1}\mathbf{S}^*$. It follows easily from (5) and Lemma 17 that

$$(6) \qquad |\boldsymbol{\tau}^{\mathrm{T}}(\mathbf{I}^*)^{-1}\mathbf{G}^*(y)| = O(n^{-1}J|\boldsymbol{\tau}|) \quad \text{uniformly in } n \ge 1, \boldsymbol{\tau} \in \Theta \text{ and } y \in \mathcal{I}.$$

LEMMA 19. $\max_{1\le j\le J} |\hat{\varphi}_j| = O_{\mathrm{pr}}([n^{-1}J\log(J)]^{1/2})$.

PROOF. Since $(\mathbf{I}^*)^{-1}\mathrm{Cov}(\mathbf{S}^*)(\mathbf{I}^*)^{-1}$ is the covariance matrix of $\hat{\boldsymbol{\varphi}}$, it follows from Lemmas 17 and 18 that $\max_j \mathrm{Var}(\hat{\varphi}_j) = O(J/n)$. Observe that

$$\hat{\varphi}_j = \sum_i ((\mathbf{I}^*)^{-1}\mathbf{G}^*(Y_i))_j.$$

According to (6),

$$\sup_{y\in\mathcal{I}} \max_{1\le j\le J} |((\mathbf{I}^*)^{-1}\mathbf{G}^*(y))_j| = O(J/n).$$

The desired result now follows from Bernstein's inequality.

LEMMA 20. $|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*-\hat{\boldsymbol{\varphi}}|^2 = O_{\mathrm{pr}}\left[n^{-2}J^3\log(J)\right]$.

PROOF. It follows from the maximum-likelihood equation that

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = \hat{\boldsymbol{\varphi}} - (\mathbf{I}^*)^{-1}(\mathbf{D}-\mathbf{I}^*)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*).$$

According to Lemma 17,

$$|(\mathbf{I}^*)^{-1}(\mathbf{D}-\mathbf{I}^*)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)|^2 = O_{pr}\Big[[J/n]^2\,|(\mathbf{D}-\mathbf{I}^*)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)|\Big].$$

The entry in row j and column k of $\mathbf{D}-\mathbf{I}^*$ can be written as

$$n\sum_m A_{jkm}(\hat{\theta}_m-\theta_m^*),$$

where

$$A_{jkm} = A_{njkm} = \int_0^1 (1-t)\frac{\partial^3 C}{\partial\theta_j\partial\theta_k\partial\theta_m}\Big[\boldsymbol{\theta}^*+t(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)\Big]\,dt.$$

Thus the jth entry of $(\mathbf{D}-\mathbf{I}^*)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)$ is

$$n\sum_k\sum_m A_{jkm}(\hat{\theta}_k-\theta_k^*)(\hat{\theta}_m-\theta_m^*).$$

Hence by Lemmas 12 and 15

$$|(\mathbf{D}-\mathbf{I}^*)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)|^2 = O_{pr}\Big[n\max_{1\le j\le J}(\hat{\theta}_j-\theta_j^*)^2\Big]$$

and therefore

$$|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*-\hat{\boldsymbol{\varphi}}|^2 = O_{pr}\Big[n^{-1}J^2\max_{1\le j\le J}(\hat{\theta}_j-\theta_j^*)^2\Big].$$

Consequently, by Lemma 19,

$$\max_{1\le j\le J}(\hat{\theta}_j-\theta_j^*)^2 = O_{pr}\Big[n^{-1}J\log(J) + n^{-1}J^2\max_{1\le j\le J}(\hat{\theta}_j-\theta_j^*)^2\Big].$$

Thus by (2)

$$\max_{1\le j\le J}(\hat{\theta}_j-\theta_j^*)^2 = O_{pr}(n^{-1}J\log(J)),$$

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 (3) and Lemma 14 that

$$\|\log(\hat{f})-\log(f^*)-[\mathbf{G}^*(\cdot)]^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)\|_\infty = O_{pr}(J/n). \tag{7}$$

Since

$$[\mathbf{G}^*(y)]^T\hat{\boldsymbol{\varphi}} = \sum_j B_j(y)\hat{\varphi}_j - \sum_j\frac{\partial C}{\partial\theta_j}(\boldsymbol{\theta}^*)\hat{\varphi}_j,$$

it follows from (4) and Lemma 19 that

$$\|[\mathbf{G}^*(\cdot)]^T\hat{\boldsymbol{\varphi}}\|_\infty = O_{pr}([n^{-1}J\log(J)]^{1/2}). \tag{8}$$

The fifth conclusion follows from (2), (5), (7), (8), Theorem 1(i), Lemma 12(ii), and Lemma 20.

We will now verify the sixth conclusion. It follows from (4) and Lemma 20 that

$$[\nabla C(\boldsymbol{\theta}^*)]^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*-\hat{\boldsymbol{\varphi}}) = O_{pr}(n^{-1}J[\log(J)]^{1/2}). \tag{9}$$

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

$$\sum_j |\hat{\theta}_j-\theta_j-\hat{\varphi}_j| \int B_j = O_{pr}(n^{-1}J[\log(J)]^{1/2}). \tag{10}$$

By (9) and (10),

$$\int |[\mathbf{G}^*(\cdot)]^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*-\hat{\boldsymbol{\varphi}})| = O_{pr}(n^{-1}J[\log(J)]^{1/2}). \tag{11}$$

The sixth conclusion is a consequence of (2), (5), (7), (8), (11), Theorem 1(i), Lemma 20, and the following result; the seventh conclusion follows from the sixth conclusion and Theorem 1.

LEMMA 21. $\max_{0\le x\le 1} \left| \int_0^x f^*(y)[\mathbf{G}^*(y)]^T\hat{\boldsymbol{\varphi}}dy \right| = O_{pr}(n^{-1/2}).$

PROOF. Observe that

$$\mathrm{Var}([\nabla C(\boldsymbol{\theta}^*)]^T\hat{\boldsymbol{\varphi}}) = \mathrm{Var}([\nabla C(\boldsymbol{\theta}^*)]^T(\mathbf{I}^*)^{-1}\mathbf{S}^*) = [(\mathbf{I}^*)^{-1}\nabla C(\boldsymbol{\theta}^*)]^T\mathrm{Cov}(\mathbf{S}^*)(\mathbf{I}^*)^{-1}\nabla C(\boldsymbol{\theta}^*).$$

Thus it follows from (4), Lemma 17 and Lemma 18 that

$$\mathrm{Var}([\nabla C(\boldsymbol{\theta}^*)]^T\hat{\boldsymbol{\varphi}}) = O(nJ^{-1}|(\mathbf{I}^*)^{-1}\nabla C(\boldsymbol{\theta}^*)|^2) = O(n^{-1}J|\nabla C(\boldsymbol{\theta}^*)|^2) = O(n^{-1})$$

and hence that

$$[\nabla C(\boldsymbol{\theta}^*)]^T\hat{\boldsymbol{\varphi}} = O_{pr}(n^{-1/2}).$$

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

$$\max_{0\le x\le 1} \left| \sum_j \hat{\varphi}_j \int_0^x f^*(y)B_j(y)dy \right| = O_{pr}(n^{-1/2}). \tag{12}$$

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

$$O\left[J^{-1}\max_{1\le j\le J}|\hat{\varphi}_j|\right] = O_{pr}(J^{-1}[n^{-1}J\log(J)]^{1/2}) = o_{pr}(n^{-1/2}).$$

Thus, by the form of the support of the B-splines B_j, $1 \le j \le J$ (see de Boor (1978)), to verify (12) it suffices to verify

$$\max_{1\le j\le J} \left| \sum_{j=1}^k \hat{\varphi}_j \int f^*B_j \right| = O_{pr}(n^{-1/2}). \tag{13}$$

Let $\mathcal{J}$ be a subset of consecutive integers in $\{1,\cdots,J\}$ and let J' denote the number of integers in $\mathcal{J}$. Let $\boldsymbol{\tau}$ denote the J-dimensional vector having elements $\int f^* B_j$ for $j \in \mathcal{J}$ and zero otherwise. Then

$$|\boldsymbol{\tau}|^2 = O(J'/J^2) \text{ uniformly in } \mathcal{J} \text{ and } n. \tag{14}$$

Since

$$\mathrm{Var}\left[\sum_{j\in\mathcal{J}} \hat{\varphi}_j \int f^* B_j\right] = [(\mathbf{I}^*)^{-1}\boldsymbol{\tau}]^{\mathrm{T}}\mathrm{Cov}(\mathbf{S}^*)(\mathbf{I}^*)^{-1}\boldsymbol{\tau},$$

it follows from (14) and Lemmas 17 and 18 that

$$\mathrm{Var}\left[\sum_{j\in\mathcal{J}} \hat{\varphi}_j \int f^* B_j\right] = O\left[\frac{J'}{nJ}\right] \quad \text{uniformly in } \mathcal{J} \text{ and } n. \tag{15}$$

Observe next that

$$\sum_{j\in\mathcal{J}} \hat{\varphi}_j \int f^* B_j = \sum_i \left[\left[\sum_{j\in\mathcal{J}} f^* B_j ((\mathbf{I}^*)^{-1}\mathbf{G}^*(Y_i))_j\right] = \sum_i X_{\mathcal{J}i}\right. \tag{16}$$

Here $X_{\mathcal{J}i} = X_{n\mathcal{J}i}$, $1 \le i \le n$, are independent random variables having mean zero; by (5), Lemma 17, and the basic properties of B-splines,

$$|X_{\mathcal{J}i}| \le b \quad \text{with } b = b_n = O(n^{-1}J^{1/2}). \tag{17}$$

It follows from (15)–(17) and Bernstein's inequality that there is a $\beta > 0$, which does not depend on n or $\mathcal{J}$, such that

$$\begin{aligned}\Pr\left[\left|\sum_{j\in\mathcal{J}} \hat{\varphi}_j \int f^* B_j\right| \ge An^{-1/2}(J'/J)^c\right] \\ \le 2[\exp(-\beta A(n/J)^{1/2}(J'/J)^c) + \exp(-\beta A^2(J/J')^{1-2c})]\end{aligned} \tag{18}$$

for $A > 0$ and $0 < c < .5$. Set $R = R_n = \min[r: 2^r \ge J]$. For $0 \le r \le R$, let $\mathcal{M}_r = \mathcal{M}_{nr}$ denote the collection of all sets of integers of the form $\{(m-1)2^r+1,\cdots,m2^r\}$, where $1 \le m \le J/2^r$. It follows from (2) and (18) that, for any $\varepsilon > 0$, A can be chosen sufficiently large so that

$$\Pr\left[\left|\sum_{j\in\mathcal{J}} \hat{\varphi}_j \int f^* B_j\right| \ge An^{-1/2}(J'/J)^c \text{ for some } \mathcal{J} \in \bigcup_0^R \mathcal{M}_r\right] \le \varepsilon, \quad n \ge 1. \tag{19}$$

For $1 \le k \le J$, $\{1,\cdots,k\}$ can be written as a disjoint union of sets $\mathcal{J} \in \mathcal{M}_r$ such that for $0 \le r \le R$, there is at most one such $\mathcal{J} \in \mathcal{M}_r$. Thus it follows from (19) that (13) holds, as desired.

5. Proof of Theorem 3. Now

$$\text{Var}([\mathbf{G}^*(y)]^T\hat{\boldsymbol{\varphi}}) = [\mathbf{G}^*(y)]^T(\mathbf{I}^*)^{-1}\text{Cov}(\mathbf{S}^*)(\mathbf{I}^*)^{-1}\mathbf{G}^*(y), \tag{20}$$

so it follows from (5), Theorem 1(i), and Lemmas 17 and 18 that

$$\text{Var}([\mathbf{G}^*(y)]^T\hat{\boldsymbol{\varphi}}) \sim J/n. \tag{21}$$

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

$$\text{Dist}\left[\frac{[\mathbf{G}^*(y)]^T\hat{\boldsymbol{\varphi}}}{\text{SD}([\mathbf{G}^*(y)]^T\hat{\boldsymbol{\varphi}})}\right] \to N(0,1) \text{ as } n \to \infty.$$

PROOF. Observe that $[\mathbf{G}^*(y)]^T\hat{\boldsymbol{\varphi}} = \sum Z_i$, where $Z_i = Z_{ni} = [\mathbf{G}^*(y)]^T\mathbf{I}^{-1}\mathbf{G}^*(Y_i)$ for $1 \le i \le n$. For each n, the random variables $Z_1, \cdots, Z_n$ have mean zero and are independent and identically distributed. Moreover,

$$|[\mathbf{G}^*(y)]^T(\mathbf{I}^*)^{-1}\mathbf{G}^*(Y_i)|^2 \le |[\mathbf{G}^*(y)]^T(\mathbf{I}^*)^{-1}\mathbf{G}^*(y)|\ |[\mathbf{G}^*(Y_i)]^T(\mathbf{I}^*)^{-1}\mathbf{G}^*(Y_i)| = O(J^2/n^2).$$

The desired result now follows from (2), (21) and the central limit theorem (see the corollary on page 201 of Chung, 1974).

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

$$|\boldsymbol{\tau}^T\text{Cov}(\mathbf{S}^*)\boldsymbol{\tau} - \boldsymbol{\tau}^T\mathbf{I}^*\boldsymbol{\tau}| \le MnJ^{-1}\delta|\boldsymbol{\tau}|^2, \quad n \ge 1 \text{ and } \boldsymbol{\tau} \in \Theta.$$

PROOF. Set $a = a_n = \int s(\cdot;\boldsymbol{\tau})f$ and $a^* = a_n^* = \int s(\cdot;\boldsymbol{\tau})f^*$. Then

$$\boldsymbol{\tau}^T\text{Cov}(\mathbf{S}^*)\boldsymbol{\tau} = n\int[s(\cdot;\boldsymbol{\tau})-a]^2 f$$

and

$$\boldsymbol{\tau}^T\mathbf{I}^*\boldsymbol{\tau} = n\int[s(\cdot;\boldsymbol{\tau})-a^*)]^2 f^*.$$

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

It follows from (5), Theorem 1(i), Lemma 17 and Lemma 23 that there is a positive constant M such that, for $n \ge 1$ and $y \in \mathcal{I}$,

$$|[\mathbf{G}^*(y)]^T(\mathbf{I}^*)^{-1}\text{Cov}(\mathbf{S}^*)(\mathbf{I}^*)^{-1}\mathbf{G}^*(y) - [\mathbf{G}^*(y)]^T(\mathbf{I}^*)^{-1}\mathbf{G}^*(y)| \le Mn^{-1}J\delta. \tag{22}$$

It follows from (2), (5), (7) and Lemma 20 that

$$|\log(\hat{f})-\log(f^*)-[\mathbf{G}^*(\cdot)]^T\hat{\boldsymbol{\varphi}}|_\infty = o_{pr}([J/n]^{1/2}).$$

Thus by (21) and Lemma 22,

$$\text{Dist}\left[\frac{\log(\hat{f})-\log(f^*)}{\text{SD}([\mathbf{G}^*(y)]^T\hat{\boldsymbol{\varphi}})}\right] \to N(0,1) \quad \text{uniformly in } y \text{ as } n \to \infty.$$

It follows easily from this together with (2), (20)–(22) and Theorem 1(i) that the first and third conclusions of Theorem 3 are valid.

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

$$|(\hat{\mathbf{I}}-\mathbf{I}^*)\tau|^2 = O_{pr}\left[nJ^{-1}\log(J)\right]|\tau|^2.$$

PROOF. Observe that

$$\frac{\partial^2 C}{\partial\theta_j\partial\theta_k}(\hat{\boldsymbol{\theta}}) - \frac{\partial^2 C}{\partial\theta_j\partial\theta_k}(\boldsymbol{\theta}^*) = \sum_m \int_0^1 \frac{\partial^3 C}{\partial\theta_j\partial\theta_k\partial\theta_m}(\boldsymbol{\theta}^*+t(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*))(\hat{\theta}_m-\theta_m^*)dt.$$

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

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

LEMMA 25. *Uniformly in* $\tau \in \Theta_0$,

$$|[\hat{\mathbf{I}}^{-1}-(\mathbf{I}^*)^{-1}]\tau|^2 = O_{pr}(n^{-3}J^3\log(J))|\tau|^2.$$

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

PROOF. Observe that

$$\hat{\mathbf{G}}(y) - \mathbf{G}^*(y) = -[\nabla C(\hat{\boldsymbol{\theta}}) - \nabla C(\boldsymbol{\theta}^*)].$$

Since

$$\nabla C(\hat{\boldsymbol{\theta}}) - \nabla C(\boldsymbol{\theta}^*) = \left[\int_0^1 \mathbf{H}(\boldsymbol{\theta}^*+t(\boldsymbol{\theta}^*-\hat{\boldsymbol{\theta}}))dt\right](\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*),$$

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

The next result follows from (5), (22), Theorem 1(i), and Lemmas 17, 25, and 26.

LEMMA 27. *Uniformly in* y,

$$[\hat{\mathbf{G}}(y)]^T\hat{\mathbf{I}}^{-1}\hat{\mathbf{G}}(y) - [\mathbf{G}^*(y)]^T(\mathbf{I}^*)^{-1}\mathrm{Cov}(\mathbf{S}^*)(\mathbf{I}^*)^{-1}\mathbf{G}^*(y)$$
$$= O_{pr}\left[[n^{-3}J^3\log(J)]^{1/2} + n^{-1}J\delta\right].$$

The second conclusion of Theorem 3 follows from (2), (20), (21), Theorem 1(i), Theorem 2(v), and Lemma 27.

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

$$g(\theta) = g_n(\theta) = \int_0^x f(y;\theta)dy, \quad \theta \in \Theta.$$

It is easily seen that

$$\nabla g(\theta^*) = \int_0^x \mathbf{G}^*(y)f^*(y)dy. \tag{23}$$

The next result follows from (2), (7), (8), (11), (23) and Theorem 1(i).

LEMMA 28. $\hat{F}(x) - F^*(x) = [\nabla g(\theta^*)]^T\hat{\varphi} + o_{pr}(n^{-1/2})$ *uniformly in* x.

Observe that

$$\mathrm{Var}([\nabla g(\theta^*)]^T\hat{\varphi}) = [\nabla g(\theta^*)]^T(\mathbf{I}^*)^{-1}\mathrm{Cov}(\mathbf{S}^*)(\mathbf{I}^*)^{-1}\nabla g(\theta^*). \tag{24}$$

By (4) and (23),

$$|\nabla g(\theta^*)|^2 = O(J^{-1}) \quad \text{uniformly in } x. \tag{25}$$

It follows from (24), (25), Theorem 1(i), Lemma 17, and Lemma 23 that

$$\mathrm{Var}([\nabla g(\theta^*)]^T\hat{\varphi}) = [\nabla g(\theta^*)]^T(\mathbf{I}^*)^{-1}\nabla g(\theta^*) + O(\delta/n) \quad \text{uniformly in } x. \tag{26}$$

Let E_θ and Var_θ correspond to the assumption that Y has density $f(\cdot;\theta)$. According to the Cramer-Rao inequality,

$$[\nabla g(\theta^*)]^T(\mathbf{I}^*)^{-1}\nabla g(\theta^*) \le n^{-1}\mathrm{Var}_{\theta^*}\left[\mathrm{ind}_{[0,x]}(Y)\right] = n^{-1}F_n(x)[1-F_n(x)]. \tag{27}$$

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

$$\lim_n \frac{[\nabla g(\theta^*)]^T(\mathbf{I}^*)^{-1}\nabla g(\theta^*)}{n^{-1}F(x)[1-F(x)]} = 1 \quad \textit{uniformly for } x \textit{ in compact subsets of } \mathrm{int}(\mathcal{I}).$$

PROOF. Choose $\varphi = \varphi_n \in \Theta_0$. By Schwarz's inequality

$$[\nabla g(\theta^*)]^T(\mathbf{I}^*)^{-1}\nabla g(\theta^*) \ge \frac{([\nabla g(\theta^*)]^T\varphi)^2}{\varphi^T\mathbf{I}^*\varphi}. \tag{28}$$

By (1)

$$\varphi^T\mathbf{I}^*\varphi = n\,\mathrm{Var}_{\theta^*}(s(Y;\varphi)). \tag{29}$$

It follows from (23) that

$$[\nabla g(\theta^*)]^T\varphi = E_{\theta^*}\left[\mathrm{ind}_{[0,x]}(Y)[s(Y;\varphi)-E_{\theta^*}s(Y;\varphi)]\right]. \tag{30}$$

The desired result follows from (27)–(30), Theorem 1(ii), the construction of $s \in \mathcal{S}_n$ used in the proof of that result, and the corresponding choice of φ_n.

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

LEMMA 30. *Uniformly for x in compact subsets of* int($\mathcal{I}$),

$$\text{Dist}\left[\frac{[\nabla g(\theta^*)]^T\hat{\varphi}}{\text{SD}([\nabla g(\theta^*)]^T\hat{\varphi})}\right] \to N(0, 1) \quad \text{as } n \to \infty.$$

The first conclusion of Theorem 4 follows from (26) and Lemmas 28–30. The second conclusion follows from the first conclusion and Theorem 1.

REFERENCES

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.

BREIMAN, L., STONE, C. J. and KOOPERBERG, C. (1988). Confidence bounds for extreme quantiles. Technical Report Number 167, Department of Statistics, University of California, Berkeley.

CHUNG, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic Press, 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.

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.

MCCULLAGH, P. and NELDER, J. A. (1983). *Generalized Linear Models*. Chapman and Hall, London.

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

PORTNOY, S. (1986). On the central limit in R^p when $p \to \infty$. *Probab. Theory Related Fields* 73 571–583.

PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* 16 356–366.

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. In *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday* (M. H. Rezvi, J. S. Rustagi, and D. Siegmund, eds.) 393–406. Academic Press, New York.

STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705.

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

STONE, C. J. (1987). A nonparametric framework for statistical modelling. In *Proceedings of the International Congress of Mathematicans, Berkeley, California, 1986* 1052–1056. American Mathematical Society, Providence, RI.

STONE, C. J. (1989). Uniform error bounds involving logspline models. To appear in *Karlin Festschrift.* Academic Press, New York.

STONE, C. J. and KOO, C.-Y. (1986a). Additive splines in statistics. In *1985 Statistical Computing Section Proceedings of the American Statistical Association* 45–48. American Statistical Association, Washington.

STONE, C. J. and KOO, C.-Y. (1986b). Logspline density estimation. In *AMS Contemporary Mathematics Series* **29** 1–15. American Mathematical Society, Providence, RI.

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. Statistical Science, 1988, Vol.3 No. 2 239-271.

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, 579-611.

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. Scandinavian J. Statist., 1988, 15, 1-23.

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. Biometrika, 1987, 74, 799-808.

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. (Nov 1986, revised Aug 1988). 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. J. Amer. Statist. Assoc., 1988, 83, 744-749.

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. (Aug 1987, revised Oct 1988). 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. (Oct. 1987, revised Mar. 1988, Oct. 1988). Geometrizing rates of convergence, II.

121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. Statistics a Guide to the Unknown, pp. 249-260 (Eds. J.M. Tanur et al.) Wadsworth, Pacific Grove.

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. (Jan. 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. (Feb. 1988, revised Nov. 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. (Revised by Tech Report No. 180).
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. (Aug.1988, revised Feb. 1989). Submodel selection and evaluation in regression I. The X-fixed 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.

177. MILLAR, P.W. (October 1988). Gamma-funnels in the domain of a probability, with statistical implications.

178. DONOHO, D.L. and LIU, R.C. (October 1988). Hardest one-dimensional subfamilies.

179. DONOHO, D.L. and STARK, P.B. (October 1988). Recovery of sparse signals from data missing low frequencies.

180. FREEDMAN, D.A. and PITMAN, J.A. (Nov. 1988). A measure which is singular and uniformly locally uniform. (Revision of Tech Report No. 163).

181. DOKSUM, K.A. and HOYLAND, ARNLJOT (Nov. 1988, revised Jan. 1989). A model for step-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution.

182. DALANG, R.C., MORTON, A. and WILLINGER, W. (November 1988). Equivalent martingale measures and no-arbitrage in stochastic securities market models.

183. BERAN, R. (November 1988). Calibrating prediction regions.

184. BARLOW, M.T., PITMAN, J. and YOR, M. (Feb. 1989). On Walsh's Brownian Motions.

185. DALANG, R.C. and WALSH, J.B. (Dec. 1988). Almost-equivalence of the germ-field Markov property and the sharp Markov property of the Brownian sheet.

186. HESSE, C.H. (Dec. 1988). Level-Crossing of integrated Ornstein-Uhlenbeck processes

187. NEVEU, J. and PITMAN, J.W. (Feb. 1989). Renewal property of the extrema and tree property of the excursion of a one-dimensional brownian motion.

188. NEVEU, J. and PITMAN, J.W. (Feb. 1989). The branching process in a brownian excursion.

189. PITMAN, J.W. and YOR, M. (Mar. 1989). Some extensions of the arcsine law.

190. STARK, P.B. (Dec. 1988). Duality and discretization in linear inverse problems.

191. LEHMANN, E.L. and SCHOLZ, F.W. (Jan. 1989). Ancillarity.

192. PEMANTLE, R. (Feb. 1989). A time-dependent version of Pólya's urn.

193. PEMANTLE, R. (Feb. 1989). Nonconvergence to unstable points in urn models and stochastic approximations.

194. PEMANTLE, R. (Feb. 1989, revised May 1989). When are touchpoints limits for generalized Pólya urns.

195. PEMANTLE, R. (Feb. 1989). Random walk in a random environment and first-passage percolation on trees.

196. BARLOW, M., PITMAN, J. and YOR, M. (Feb. 1989). Une extension multidimensionnelle de la loi de l'arc sinus.

197. BREIMAN, L. and SPECTOR, P. (Mar. 1989). Submodel selection and evaluation in regression — the X-random case.

198. BREIMAN, L., TSUR, Y. and ZEMEL, A. (Mar. 1989). A simple estimation procedure for censored regression models with known error distribution.

199. BRILLINGER, D.R. (Mar. 1989). Two papers on bilinear systems: a) A study of second- and third-order spectral procedures and maximum likelihood identification of a bilinear system. b) Some statistical aspects of NMR spectroscopy, *Actas del 2° congreso lantinoamericano de probabilidad y estadistica matematica*, Caracas, 1985.

200. BRILLINGER, D.R. (Mar. 1989). Two papers on higher-order spectra: a) Parameter estimation for nonGaussian processes via second and third order spectra with an application to some endocrine data. b) Some history of the study of higher-order moments and spectra.

201. DE LA PENA, V. and KLASS, M.J. (April 1989). L bounds for quadratic forms of independent random variables.

202. FREEDMAN, D.A. and NAVIDI, W.C. (April 1989). Testing the independence of competing risks.

203. TERDIK, G. (May 1989). Bilinear state space realization for polynomial stochastic systems.

204. DONOHO, D.L. and JOHNSTONE, I.M. (May 1989). Minimax risk over l_p-Balls.

205. PEMANTLE, R., PROPP, J. and ULLMAN, D. (May 1989). On tensor powers of integer programs.

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.