

**A Decomposition for the Likelihood Ratio Statistic
and the Bartlett Correction — a Bayesian Argument**

By

**Peter J. Bickel
University of California, Berkeley⁽¹⁾⁽²⁾**

and

**J.K. Ghosh⁽²⁾
Indian Statistical Institute, Calcutta**

**Technical Report No. 111
August 1987
(revised June 1988)**

⁽¹⁾This paper was completed while the author was visiting AT&T Bell Telephone Labs,
the Courant Institute and the University of Chicago.

⁽²⁾This research was partially supported by ONR Contract N00014-80-C-0163

**Department of Statistics
University of California
Berkeley, California**

**A Decomposition for the Likelihood Ratio Statistic
and the Bartlett Correction — a Bayesian Argument**

by

Peter J. Bickel

University of California, Berkeley⁽¹⁾⁽²⁾

and

J. K. Ghosh⁽²⁾

Indian Statistical Institute, Calcutta

ABSTRACT

Let $l(\theta) = n^{-1} \log p(x, \theta)$ be the log-likelihood of an n -dimensional X under a p -dimensional θ . Let $\hat{\theta}_j$ be the mle under $H_j: \theta^1 = \theta_0^1, \dots, \theta^j = \theta_0^j$ and $\hat{\theta}_0$ be the unrestricted mle. Define T_j as

$$\left[2n \{l(\hat{\theta}_{j-1}) - l(\hat{\theta}_j)\} \right]^{1/2} \text{ sign } (\hat{\theta}_{j-1}^j - \theta_0^j).$$

Let $T = (T_1, \dots, T_p)$. Then under regularity conditions, the following theorem is proved: Under $\theta = \theta_0$, T is asymptotically $N(n^{-1/2} a_o + n^{-1} a, J + n^{-1} \Sigma) + O(n^{-3/2})$ where J is the identity matrix. The result is proved by first establishing an analogous result when θ is random and then making the prior converge to a degenerate distribution. The existence of the Bartlett correction to order $n^{-3/2}$ follows from the theorem. We show that an Edgeworth expansion with error $O(n^{-2})$ for T involves only polynomials of degree ≤ 3 and hence verify rigorously Lawley's (1956) result giving the order of the error in the Bartlett correction as $O(n^{-2})$.

(1) This paper was completed while the author was visiting AT&T Bell Telephone Labs, the Courant Institute and the University of Chicago.

(2) This research was partially supported by ONR Contract N00014-80-C-0163.

**A Decomposition for the Likelihood Ratio Statistic
and the Bartlett Correction — a Bayesian Argument**

by

Peter J. Bickel

University of California, Berkeley⁽¹⁾⁽²⁾

and

J. K. Ghosh⁽²⁾

Indian Statistical Institute, Calcutta

1. Introduction

Let $X = (X_1, \dots, X_n)$ be a vector of observations with joint density $p(x, \theta)$, $\theta \in \Theta$ open $\subset R^p$ where we do not assume a priori any particular structure on $p(x, \theta)$. Consider the hypothesis $H : \theta^1 = \theta_0^1, \dots, \theta^k = \theta_0^k$. Suppose that maximum likelihood estimates $\hat{\theta}$ and $\hat{\theta}_H$ for $\theta \in \Theta$ and $\theta \in H$ respectively are well defined. Then let

$$(1.1) \quad l(\theta) = n^{-1} \log p(X, \theta)$$

$$(1.2) \quad l(\hat{\theta}) = \max_{\Theta} l(\theta)$$

$$(1.3) \quad l(\hat{\theta}_H) = \max_H l(\theta),$$

and

$$(1.4) \quad \Lambda = 2n(l(\hat{\theta}) - l(\hat{\theta}_H))$$

the usual likelihood ratio test statistic. All these quantities, of course, depend on n but we suppress this dependence to ease the notation. There is a common approximation to the distribution of Λ which has the status of a folk theorem:

$$L_{\theta}(\Lambda) \approx \chi_k^2$$

for $\theta \in H$. Theoretically this can be interpreted as, for $\theta \in H$,

(1) This paper was completed while the author was visiting AT&T Bell Telephone Labs, the Courant Institute and the University of Chicago.

(2) This research was partially supported by ONR Contract N00014-80-C-0163.

$$(1.5) \quad P_{\theta}[\Lambda \leq t] = \chi_k^2(t) + o(1),$$

as $n \rightarrow \infty$. This result was proved by Wilks (1938) and extended by Wald (1943) in the i.i.d. case, extended to the Markov case by Billingsley (1961) and subsequently extended to many other dependent and nonstationary situations. Bartlett (1937) noted, in the particular case of the hypothesis of the equality of variances for $k+1$ normal populations, that the χ_k^2 distribution was a far better fit to the distribution of $k\Lambda/E_{\theta}\Lambda$ than to Λ itself. Following work by Box (1949) and Bartlett (1954), Lawley (1956) by ingenious and difficult cumulant calculations "established" the folk theorem that quite generally

$$(1.6) \quad P_{\theta}\left[\frac{k\Lambda}{\hat{E}} \leq t\right] = \chi_k^2(t) + O(n^{-2})$$

where

$$\hat{E} = k + \frac{\hat{b}}{n} = E_{\theta}(\Lambda) + O_p(n^{-3/2}),$$

and \hat{b} is a suitable estimate for the coefficient b of n^{-1} in the expansion of $E_{\theta}(\Lambda)$. Departing from an asymptotic formula for the conditional density of X given an ancillary due to Barndorff-Nielsen (1983), Barndorff-Nielsen and Cox (1984) showed that (1.6) can be expected to hold quite generally and derived formulas for estimating b in one important class of models. Efron (1985) established (for an important special case) a related result. Let,

$$T = \Lambda^{1/2} \operatorname{sgn}(\hat{\theta}^1 - \theta^1).$$

Then

$$(1.7) \quad P_{\theta}[T \leq t] = \Phi\left[\frac{t - \mu(\theta)}{\sigma(\theta)}\right] + O(n^{-3/2})$$

where

$$\mu(\theta) = \frac{a_0(\theta)}{\sqrt{n}} + \frac{a_1(\theta)}{n} + O(n^{-3/2})$$

$$\sigma^2(\theta) = 1 + \frac{c(\theta)}{n} + O(n^{-3/2}),$$

where a_0 , a_1 and c are suitable functions of θ , not depending on n . As P. McCullagh pointed out to us, this result implicitly already appears in Lawley (1956) and, in fact, $a_1 = 0$. It is easy to see that, for $k = 1$, (1.7) finally implies (1.6) (with $O(n^{-2})$ replaced by $O(n^{-3/2})$) with $\hat{\theta}$ estimating $a_0^2(\theta) + c(\theta)$.

Our aim in this paper is,

- a) To give a generalization of Efron's result to vector parameters. A closely related result appears in Barndorff-Nielsen (1986) and is again foreshadowed by Lawley (1956).
- b) To apply this extension to establish the validity of Bartlett's correction for the p variate joint distribution of the Λ statistics (deviances) arising from testing the nested hypotheses $H_k : \theta^j = \theta_o^j$, $j = 1, \dots, k$, within H_{k-1} for $k = 1, \dots, p$. That is, to show that, when the deviances are standardized by their asymptotic expectations to order $\frac{1}{n}$, their joint distribution, under θ_o , differs from that of p independent identically distributed χ^2_1 variables by an error of order n^{-2} . This result is also implicit in Lawley (1956) although the calculations are purely formal. For the case of a single statistic Λ , this can be obtained in a

rigorous fashion under appropriate regularity conditions from Chandra and Ghosh (1979).

c) To give Bayesian analogues of both of these results which we believe provide a key to understanding the Bartlett phenomenon. The Bayesian analogue is interesting in its own right, is fairly easy to establish and is the basic step in our arguments for a) and b).

Here is a discussion of the motivation and the structure of our Bayesian argument when we restrict to the familiar case of i.i.d. observations from a smooth parametric family. It has been proved in Chandra and Ghosh (1979) that the distributions of the likelihood ratio, as well as Wald's and Rao's score statistic, have asymptotic expansions in powers of n^{-1} , which are valid in the sense of Bickel (1974). These types of expansions have been around for a long time, see Box (1949). When viewed as formal expansions for the density $p_n(\chi^2)$ of one of these statistics, they are of the form $ce^{-\chi^2/2}(\chi^2)^{k/2-1} \{1 + \psi_1(\chi^2)n^{-1} + \dots\}$, where the coefficients ψ 's are polynomials in χ^2 . It is easy to check that adjustment of such a statistic through multiplication or division by a constant of the form $(1 + bn^{-1})$ will knock off the coefficient of n^{-1} in the expansion for the adjusted statistic, iff ψ_1 is linear. By examining various examples one can convince oneself that ψ_1 is not linear for Wald's or Rao's statistic. Moreover it is far from clear why ψ_1 is linear for the likelihood ratio statistic. This paper is addressed to clearing up mysteries of this kind as well as to exploring the duality between the Bayesian and the frequentist set up which, to first order, was studied extensively by Le Cam under the rubric of the Bernstein-von Mises theorem.

Our Bayesian route could be followed to produce a relatively transparent proof of linearity of ψ_1 . However, since we want to do more, namely, derive the asymptotic expansion for the joint distribution of the p deviances statistics up to $O(n^{-2})$, we first note, in a similar vein, that here also the question boils down to the structure of the polynomials that appear as coefficients of powers of n^{-1} in the expansion. The relevant results for this purpose are the Lemmas A2 through A4 in the Appendix. These lemmas need to be applied to the vector $T(\theta, X)$ of the signed square roots of the likelihood ratio statistics, defined in section 2. That the distribution of these statistics has a valid Edgeworth expansion can be shown using Theorem 2 of Bhattacharya and Ghosh (1978). In the frequentist set-up the sort of structure one needs for the polynomials is specified in the conclusion of Theorem 3. It turns out that one needs the polynomials corresponding to $n^{-1/2}$ and n^{-1} to be of degree at most one and two respectively. To prove this one first obtains a similar result in the Bayesian set-up, namely, Theorem 1, which provides an expansion for the posterior distribution of $T(\theta, X)$ given X . The likelihood factor in the posterior $\exp\{nl(\theta) - nl(\hat{\theta})\}$ is exactly the sum of squares of the components of T and so no expansion is needed. The coefficient polynomials in the asymptotic expansion arise only from the Taylor expansions of the prior density $\pi(\theta)$ around $\hat{\theta}$ and a stochastic expansion of the Jacobian of the transformation of $(\theta - \hat{\theta})$ to $T(\theta, X)$ viewed as a function of random θ . For reasons that are not hard to see, in these latter expansions the degree of the coefficient polynomial matches the power of n^{-1} , vide Lemmas 3.1 and 3.2. These facts are at the heart of the proof of Theorem 1. Theorem 1 would fail for Wald's or

Rao's statistic because the likelihood factor $\exp\{nl(\theta) - nl(\hat{\theta})\}$ can't be written as the square of either of them exactly and so an expansion of this term is called for too. Finally, Theorem 3 follows because Theorem 1 is true for a set of priors which is dense in the weak topology.

Our expansions may be used to set up Bayesian or frequentist confidence intervals — see the discussion following Corollary 1.

We propose to carry out our program without relying on the i.i.d. sampling assumption, under conditions such as those of Bickel, Götze, and van Zwet (1985) which emphasize that we are, as with the original Wilks result, dealing with a phenomenon which depends only on the asymptotic stability of l and its derivatives, moderate deviation properties of $\hat{\theta}$ and related estimates and the existence of Edgeworth expansions for the distribution of T . Simple conditions implying those we give may be specified in the case of Markov and independent nonidentically distributed observations in the same way as is done in Bickel, Götze and van Zwet (1985).

A feature of our approach is that calculations are kept to a minimum so that, we believe, the phenomena are transparent. The disadvantage here is that unlike our predecessors we do not arrive at formulae for the (estimated) coefficient \hat{b} needed in the correction. It is, however, worth pointing out that, in situations which are like simple random sampling and where computing power is readily available, we can obtain \hat{b} without knowing its form by applying the jackknife for bias reduction - see Efron (1982) for example. That is, we calculate Λ_{-i} , the

Λ statistic for the data $X_j, j \neq i$ and put

$$\hat{b} = \sum_{i=1}^n (\Lambda_{-i}) - nk.$$

The paper is organized as follows. Section 2 contains the statements of the main theorems plus the necessary assumptions and notations. Section 3 contains the proofs of our results. Four simple technical lemmas are in the appendix.

2. The Main Results

Since we intend to use tensor notation for arrays we subsequently identify vector components by superscripts, for example, $\theta = (\theta^1, \dots, \theta^p)$. For given $\theta \in \Theta$ define $\hat{\theta}_j$ as the maximum likelihood estimate of θ when $\theta^1, \dots, \theta^j$ are fixed, i.e.,

$$(2.1) \quad l(\hat{\theta}_j) = \max \left\{ l(\tau) : \tau^1 = \theta^1, \dots, \tau^j = \theta^j \right\}.$$

We shall in the sequel assume that these quantities exist and are unique but will sketch at the end of the section how this requirement can be weakened. Define, $T = (T^1, \dots, T^p)$ where

$$(2.2) \quad T^j \equiv n^{1/2} \left[2(l(\hat{\theta}_{j-1}) - l(\hat{\theta}_j)) \right]^{1/2} \operatorname{sgn}(\hat{\theta}_{j-1}^j - \theta^j).$$

Note that T is a function of θ and \mathbf{X} .

Let π be a prior density on Θ . Let P denote the joint distribution of (θ, \mathbf{X}) and $P(\cdot | \mathbf{X})$ the conditional (posterior) probability distribution of (θ, \mathbf{X}) given \mathbf{X} . Let $r = n^{-1/2}$ and consider the posterior density of $r^{-1}(\theta - \hat{\theta})$ given by

$$\pi(h | \mathbf{x}) \equiv \exp \{ l(\hat{\theta} + rh) - l(\hat{\theta}) \} \pi(\hat{\theta} + rh) / N(\mathbf{x})$$

where

$$(2.3) \quad N(\mathbf{X}) = \int \exp\{l(\hat{\theta} + rh) - l(\hat{\theta})\} \pi(\hat{\theta} + rh) dh.$$

Let

$$\phi(t) = (2\pi)^{-p/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^p (t^i)^2\right\}$$

the standard p variate normal density. Let $\pi_T(t | \mathbf{X})$ denote the posterior density of T , (which exists under our assumptions with probability $1 - O(r^{m+1})$).

Notation: We postulate $m+3$ continuous derivatives for $l(\theta)$, $\pi(\theta)$ and write $l_{i_1 \dots i_k}$ for

$\frac{\partial^k l}{\partial \theta^{i_1} \dots \partial \theta^{i_k}}$, etc. Following tensor notation, we indicate arrays by their elements. Thus l^i

is a vector, l_{ij} a matrix etc. We also follow the Einstein convention of summing over a subscript which is repeated in a superscript, e.g. $l_{ij} l^i = \sum_i l_{ij} l^i$. Occasionally we denote a vector array by symbols like v_i , so that $v_i t^i$ stands for $\sum_i v_i t^i$

Here are the main results stated under regularity conditions which appear at the end of the section.

Theorem 1. If B_m holds then

$$(2.4) \quad E_P \int |\pi_T(t | \mathbf{X}) - \pi_m(t, \mathbf{X})| dt = O(r^{m+1})$$

where

$$\pi_m(t, \mathbf{X}) = \phi(t) \left[1 + P_m(r, \mathbf{X}, \pi) + Q_m(rt, \mathbf{X}, \pi) \right] 1(\mathbf{X} \in S),$$

P_m is a polynomial in r of degree m , Q_m is a polynomial in rt of degree m , both without

constant terms and with coefficients which are rational functions of $l_{b_1 \dots b_k}(\hat{\theta})$ and $\pi_{b_1 \dots b_k}(\hat{\theta})/\pi(\hat{\theta})$ for $1 \leq k \leq n+2$, and $P[X \notin S] = O(r^{m+1})$ where S is given in Section 3.

Write

$$P_m(r, \mathbf{X}, \pi) = \sum_{k=1}^m P_{mk}(\mathbf{X}, \pi) r^k$$

$$Q_m(u, \mathbf{X}, \pi) = \sum_{k=1}^m Q_{mb_1 \dots b_k}(\mathbf{X}, \pi) u^{b_1} \dots u^{b_k}.$$

and note that P_m , Q_m and S depend on n .

Notes. 1) It is necessary to keep the indicator of S in π_m since the coefficients P_{mk} , $Q_{mb_1 \dots b_k}$ need not be bounded outside S .

The proof of Theorem 1 actually also yields that if $X \in S$, i.e. with probability $1 - O(r^{m+1})$, the random quantity

$$\int |\pi_T(t | \mathbf{X}) - \pi_m(t | \mathbf{X})| dt$$

is $O(r^{m+1})$.

2) Since P_{mk} and $Q_{mb_1 \dots b_k}$ depend on r they are not uniquely defined. Since

$$(2.4a) \quad E \left| \int \pi_m(t, \mathbf{X}) dt - 1 \right| = O(r^{m+1})$$

it is easy to see that we can always take $P_{mo} = Q_{mo} = 0$ and suppose all P_{mk} for k odd to be zero. For example, suppose we are given a set of $P_{mk}^{(1)}$ and associated Q_m . Note that

$$P_m^{(1)} r + \int Q_{m1} r t \phi(t) dt = O(r^2) \text{ if } m \geq 1$$

Therefore, $P_m^{(1)} = O(r)$. Hence we can define the following set $P_{mk}^{(2)}$ satisfying (2.4).

$$P_{m0}^{(2)} = 0, P_{m2}^{(2)} = \frac{P_m^{(1)}}{r} + P_{m2}^{(1)} + P_{m3}^{(1)} r, P_{mk}^{(2)} = 0 \text{ for } k \text{ odd and } P_{mk}^{(2)} = P_{mk}^{(1)} + r P_{m(k+1)}^{(1)} \text{ for } k \text{ even.}$$

even and ≥ 4 .

3) Note that (2.4a) for $m = 2, 3$ implies $E |\int \pi_2(t, X) dt - 1| = E |P_{22} r^2 - Q_{2ij} \delta^{ij} r^2| = O(r^3)$. In view of Notes 1,2 and the above relation we deduce, putting $m = 1, 2$ in (2.4), that with probability $1 - O(r^2)$ and $1 - O(r^3)$ respectively the posterior distribution of T is $N_p(rQ_{1i}, J)$ with error $O(r^2)$ and $N_p(rQ_{2i}, J + r^2(2Q_{2ij} - Q_{2i}Q_{2j}))$ with error $O(r^3)$, where $N_p(\mu, \Sigma)$ is the p variate normal distribution with mean μ and dispersion matrix Σ , and J is the $p \times p$ identity matrix. These are the multivariate Bayesian analogues of Efron's (1985) result.

4) The relation (2.4a) for $m = 3$ implies as above that

$$E |P_{32} r^2 - Q_{3ij} \delta^{ij} r^2| = O(r^4)$$

and hence that π_3 may be written as

$$rQ_{3i} t^i + r^2 Q_{3ij} (t^i t^j - \delta^{ij}) + r^3 Q_{3ijk} t^i t^j t^k + O(r^4)$$

which has the structure of $g(t)$ of Lemma A2 up to $O(r^4)$. This fact will be used in the proof of Theorem 2.

Let $c_k(\cdot)$ denote the χ_k^2 density,

$$D^j \equiv (T^j)^2 = 2n(l(\hat{\theta}_{j-1}) - l(\hat{\theta}_j))$$

the deviance, and

$$\tilde{D}^j = D^j / (1 + 2r^2 Q_{2jj})$$

the standardized (Bartlett corrected) deviance. If $\pi_D, \pi_{\tilde{D}}$ are the corresponding posterior densities of these vectors $D = (D^1, \dots, D^p)$ and $\tilde{D} = (\tilde{D}^1, \dots, \tilde{D}^p)$, then one has the following

result.

Theorem 2. Under B_1

$$(2.5) \quad E_P \left\{ \int \left| \pi_D(u | \mathbf{X}) - \prod_{j=1}^p c_1(u^j) \right| dt \mathbf{1}(\mathbf{X} \in S) \right\} = O(n^{-1})$$

while under B_3 ,

$$(2.6) \quad E_P \left\{ \int \left| \pi_{\tilde{D}}(u | \mathbf{X}) - \prod_{j=1}^p c_1(u^j) \right| du \mathbf{1}(\mathbf{X} \in S) \right\} = O(n^{-2}).$$

In fact (vide Note 1), with probability $1 - O(n^{-1})$ and error $O(n^{-1})$ the posterior distribution of D is that of p independent χ_1^2 , while for \tilde{D} the same claim holds with probability $1 - O(n^{-2})$ and error $O(n^{-2})$.

From this we deduce

Corollary 1. a) Under B_1 , if π_{Λ} is the posterior distribution of Λ given by (1.4)

$$(2.7) \quad E_P \left\{ \int \left| \pi_{\Lambda}(u | \mathbf{X}) - c_k(u) \right| du \mathbf{1}(\mathbf{X} \in S) \right\} = O(n^{-1}).$$

b) Let $\tilde{\Lambda} = \Lambda / \left[1 + 2r^2 k^{-1} \sum_{j=1}^k Q_{2jj} \right]$

Then under B_3

$$(2.8) \quad E_P \left\{ \int \left| \pi_{\tilde{\Lambda}}(u | \mathbf{X}) - c_k(u) \right| du \mathbf{1}(\mathbf{X} \in S) \right\} = O(n^{-2}).$$

So (2.7) says that the posterior distribution of Λ is χ_k^2 with error $O(n^{-1})$ while (2.8) is the Bayesian analogue of the Bartlett phenomenon. The posterior distribution of the ‘‘Bartlett standardized’’ statistic, $\tilde{\Lambda}$, is χ_k^2 with error $O(n^{-2})$.

These results can in principle be used to set Bayesian posterior confidence regions for θ to order n^{-1} , n^{-2} in a variety of ways. For instance, $\{\theta : \Lambda \leq \chi_p(1-\alpha)\}$ where χ_p is the $1-\alpha$ percentile of χ_p^2 and $\Lambda = 2(l(\hat{\theta}) - l(\theta))$ has posterior probability $1-\alpha$ with error $O(n^{-1})$ while $\{\theta : \tilde{\Lambda} \leq \chi_p(1-\alpha)\}$ has posterior probability $1-\alpha$ with error $O(n^{-2})$. Of course regions could be based on other functions of D_j , \tilde{D}_j , for instance on $\max_j D_j$ or $\max_j \tilde{D}_j$. They could also be used in investigating the old question of what choices of model and prior lead to posterior probability regions which are also frequentist regions with error $O(n^{-2})$, see, for example Stein (1985), Welch and Peers (1963). However, more detailed computation of the Q_j than we provide seems necessary for this endeavor.

We use these results only in establishing the corresponding result in the frequentist case.

Theorem 3. Suppose that F_m holds and the density of T , $p_T(t|\theta)$ admits an Edgeworth expansion such that, if $i^2 = -1$,

$$(2.9) \quad \left| \int e^{i\mathbf{v}_j t^j} [p_T(t|\theta) - \phi(t) \{1 + \sum_{k=1}^m r^k R_k(t, \theta)\}] dt \right| = O(r^{m+1})$$

uniformly in compact sets of θ and \mathbf{v} , where the $R_k(\cdot, \theta)$ are continuous in θ and polynomials in t , independent of r . Then, the R_k are of at most degree k in t .

As in notes (2) and (3) after Theorem 1, it is clear that (2.9) implies, on taking $\mathbf{v} = 0$, that $R_1(t, \theta) = R_{1j} t^j$, $R_2(t, \theta) = R_{2ij} (t^i t^j - \delta^{ij}) + R_{2i} t^i$ where δ^{ij} is the Kronecker delta. In the following we shall need a condition analogous to (2.9), namely,

$$(2.9a) \quad \left| \int e^{i\mathbf{v}_j (t^j)^2} [p_T(t|\theta) - \phi(t) \{1 + \sum_{k=1}^m r^k R_k(t, \theta)\}] dt \right| = O(r^{m+1})$$

uniformly in compact sets of θ and all v . We deduce our generalization of Efron's result.

Corollary 2: If $m = 1$, the characteristic function of p_T differs from that of $N(rR_{1j}, J)$ by $O(r^2)$ and if $m = 2$, from $N(rR_{ij}, J + r^2(2R_{2j} - R_{1i}R_{1j}))$ by $O(r^3)$.

Theorem 4. If the assumptions of Theorem 3 and (2.9a) hold for $m = 1$, then, uniformly in v ,

$$(2.10) \quad \int e^{ivu} [p_D(u|\theta) - \prod_{j=1}^p c_1(u^j)] du = O(n^{-1})$$

i.e. the approximation $\prod_{j=1}^p c_1(u^j)$ is good to order n^{-1} .

(b) Further let

$$\tilde{D}^j = D^j / (1 + 2r^2 R_{2jj}).$$

If (2.9), (2.9a) and F_m hold for $m = 3$, then uniformly in v ,

$$(2.11) \quad \int e^{iv_j u^j} [p_{\tilde{D}}(u|\theta) - \prod_{j=1}^p c_1(u^j)] du = O(n^{-2}).$$

Corollary 3. Under the conditions of Theorem 4, uniformly in v ,

$$(2.12) \quad \int e^{ivu} [p_{\Lambda}(u|\theta) - c_k(u)] du = O(n^{-1}).$$

$$(2.13) \quad \int e^{ivu} [p_{\tilde{\Lambda}}(u|\theta) - c_k(u)] du = O(n^{-2}).$$

It turns out that $T^i = r^{-1}(\hat{\eta}^i - \eta^i) + O(r)$ (see (3.6) and (3.19)), and $r^{-1}(\hat{\eta}^i - \eta^i)$ is up to $O(r)$ a linear function of the first derivatives of the log likelihood evaluated at θ . In fact it is possible to stochastically expand T in terms of the derivatives of the log likelihood evaluated at θ , with a leading linear term. In the i.i.d. case if enough moments are finite, we can talk of a formal Edgeworth expansion for the density or distribution function of T , and under the

same assumptions the rigorous expansion of the characteristic function of T that we require is valid, vide the introduction in Bhattacharya and Ghosh (1978). This is all that one needs to justify the Bartlett correction and the related results as given in Theorem 4. If one wants these results to be valid for the distribution function in the sense of Bickel (1974), it is enough to assume that the Edgeworth expansion for the density of T is valid in the L_1 -sense. This assumption may be verified via Theorem 2(a) of Bhattacharya and Ghosh (1978), if the derivatives of the log-likelihood appearing in the stochastic expansion for T up to $o_p(n^{-3/2})$ have an absolutely continuous joint distribution. Actually, instead of absolute continuity, it is enough to assume Cramer's condition, vide condition C of Bhattacharya and Ghosh (1978), and apply their Theorem 2(b) instead of Theorem 2(a).

We note again that a form of Theorem 4 appeared in Barndorff-Nielsen (1986) (with error $O(n^{-3/2})$). Barndorff-Nielsen's results focus on conditional inference given asymptotic ancillary statistics. His work implicitly requires conditions for the validity of saddlepoint expansions for the conditional density. These in turn imply but are not necessary for the validity of Edgeworth expansions for the conditional density. The Edgeworth expansions may be used in conjunction with our "Bayesian" result to derive the appropriate analogues of Theorem 4. We believe our Bayesian route makes matters easier and more transparent. The assumptions below may appear rather strong but, as indicated in the remarks, they hold quite generally. Moreover, they are quite natural if one is to develop a rigorous, rather than a formal, argument.

Suppose we estimate the correction factor and adjust the likelihood ratio statistic in (1.6).

If in Corollary 3 we replace $\tilde{\Lambda}$ by $k \Lambda / (k + \frac{\hat{b}}{n})$ then the conclusion of Corollary 3 holds

under suitable regularity conditions. This fact was first noted by Barndorff - Nielsen and Hall (1987). The most brutal condition is to suppose that,

$$(2.14) \quad \hat{b} = b(\theta) + r c_i t^i + \Delta(\theta)$$

where

$$E_\theta |\Delta(\theta)| = O(r^2).$$

Of course (2.14) is motivated by a stochastic expansion such as,

$$(2.15) \quad \hat{b} \equiv b(\hat{\theta}) = b(\theta) + d_i (\hat{\theta}^i - \theta^i) + O_p(r^2)$$

and the expansion

$$\hat{\theta}^i - \theta^i = r \hat{D}_{ij} T^i + O_p(r^2)$$

for a suitable \hat{D}_{ij} , see Lemma 3.2. To show that (2.14) and the assumptions of Corollary 3

are enough for this result we need only note that the difference between the Fourier transforms of $\tilde{\Lambda}$ and $k \Lambda / (k + \hat{b} / n)$ at v can be written as (with an appropriate constant $M(\theta)$) as,

$$M(\theta) \int \exp \left[\left(-\frac{1}{2} \sum_{i=1}^p [t^i]^2 \right) + i v \sum [t^i]^2 \left[\sum [t^i]^2 (c_i t^i) \right] r^3 dt + O(r^4) \right]$$

uniformly on compact v subsets. The integral vanishes by symmetry.

Condition (2.14) is too brutal but can readily be replaced by the possibility of further expansion of (2.15) and large deviation estimates for $\hat{\theta} - \theta$. Alternatively, we can simply suppose that the Edgeworth expansion of $k \Lambda (k + \hat{b} / r^2)^{-1}$ agrees with that of

$\tilde{\Lambda}(1 - (k + b(\theta)r^2)^{-1}r^2 c_i T^i)$ with error of order r^2 . This kind of replacement can be proved in a standard fashion under the usual protocols for asymptotic expansions of maximum likelihood estimates. See Pfanzagl (1974), for example.

We postulate nonrandom arrays λ_i, λ_{ij} etc. and write,

$$l_{i_1 \dots i_k}(\theta) = \lambda_{i_1 \dots i_k}(\theta) + \Delta_{i_1 \dots i_k}(\theta).$$

Here are our conditions. Let $|\cdot|$ denote the l_1 norm on R^p . For all $0 < M < \infty$ and some $0 < \delta < 1, \varepsilon_n \downarrow 0$.

$$B_m: \quad (i) \quad P[|\hat{\theta} - \theta| \geq Mr^{1-\delta}] = O(r^{m+1})$$

$$(ii) \quad P[|\hat{\theta} - \theta| \leq Mr^{m+2}] = O(r^{m+1})$$

Let

$$A = \{x: \text{for all } j, \{\theta: |\hat{\theta}(x) - \theta| \leq M_1 r^{1-\delta}\} \subset \{\theta: |\hat{\theta}_j(x, \theta) - \hat{\theta}(x)| \leq M_2 r^{1-\delta}\}\}.$$

For all $0 < M_1 < \infty$ there exists $0 < M_2 < \infty$ such that

$$(iii) \quad P[X \notin A] = O(r^{m+1})$$

$$(iv) \quad P[\sup\{|\Delta_{i_1 \dots i_k}(\hat{\theta} + rv)|: |v| \leq Mr^{1-\delta}\} \geq \varepsilon_n] = O(r^{m+1}), \quad 1 \leq k \leq m+3$$

(v) The maps $\theta \rightarrow \lambda_{i_1 \dots i_k}(\theta)$ are continuous, $1 \leq k \leq m$

(vi) The matrix $\|-\lambda_{ij}(\theta)\|$ is positive definite for all θ

(vii) a) π vanishes off a compact $K \subset \Theta$

$$b) \quad P[\sup\{|\pi_{i_1 \dots i_{m+2}}(\hat{\theta} + rv)|/\pi(\hat{\theta}): |v| \leq Mr^{-\delta}\} \geq r^{-\delta}] = O(r^{m+1})$$

F_m : Uniformly on compacts in θ ,

(i) $P_\theta[|\hat{\theta} - \theta| \geq Mr^{1-\delta}] = O(r^{m+1})$

(ii) $P_\theta[|\hat{\theta} - \theta| \leq Mr^{m+2}] = O(r^{m+1})$

(iii) $P_\theta[X \notin A] = O(r^{m+1})$, for A defined in B_m

(iv) $P_\theta[\sup\{|\Delta_{i_1 \dots i_k}(\theta + rv)| : |v| \leq Mr^{1-\delta}\} \geq \varepsilon_n] = O(r^{m+1})$, $1 \leq k \leq m+3$

(v) Condition (v) of B_m

(vi) Condition (vi) of B_m

Remarks. a) We give a qualitative discussion of the “Bayesian” conditions B_m . The frequentist conditions F_m can be viewed in an analogous fashion.

(i). Variations of the M.L.E. $\hat{\theta}$ from θ of order $n^{-1/2(1-\delta)}$ occur with very small probability.

Thus we can safely think about Taylor expanding $l(\theta)$ and $l(\hat{\theta}_j(\theta))$ around $\hat{\theta}$.

(ii). This condition says that $r^{-1}(\hat{\theta} - \theta)$ has approximately a bounded density near 0. It is needed to ensure that the map $\theta - \hat{\theta} \rightarrow T(\theta, x)$ is 1-1 and otherwise well behaved with high probability.

(iii). This condition assumes that both $\hat{\theta}$ and $\hat{\theta}_j$ are close to θ and each other simultaneously.

It is needed for expansions of $l(\hat{\theta}_j(\theta))$.

(iv). The coefficients of the Taylor expansion differ little from constants, or more specifically, $l(\theta)$ and its derivatives behave like averages of i.i.d. variables.

(v). Smoothness conditions needed to permit replacement of quantities such as

$\lambda_{i_1 \dots i_k}(\hat{\theta}_j(\theta))$ appearing as approximations to coefficients in the Taylor expansion of $l(\hat{\theta}_j(\theta))$ by $\lambda_{i_1 \dots i_k}(\hat{\theta})$.

(vi). Nonsingularity of the information matrix is necessary even for the statement of the Bernstein-von Mises theorem.

(vii). We need to expand $\log \pi(\theta)$ around $\hat{\theta}$. Condition a) is useful for technical reasons. while b) is needed to control $\log \pi$ and its derivatives near the boundary of K where $\log \pi \rightarrow -\infty$.

b) The validity of F_m , and B_m other than (ii), (iii) has been checked for independent nonidentically distributed and Markov dependent observations in Bickel, Götze, and van Zwet (1985). In particular these conditions hold for exponential families in the i.i.d. case. They also hold in many examples for such families in the independent non-identically distributed case, e.g., in regression and GLIM models. Another example is the class of aperiodic irreducible finite state Markov chains with stationary completely unknown transition matrix.

c) Condition B_m (ii) in fact follows from the other B_m conditions since they guarantee an Edgeworth expansion for $\pi(h | \mathbf{X})$.

An Edgeworth expansion uniform on θ compacts for the distribution of $r^{-1}(\hat{\theta} - \theta)$ implies F_m (i) and (ii).

Condition F_m or B_m (iii) holds if the log likelihood is convex.

d) The conditions on existence of the estimate $\hat{\theta}_j$ can be replaced by requiring the existence of a preliminary estimate $\tilde{\theta}$ with appropriate moderate deviation properties and then redefining the $\hat{\theta}_j$ as the result of $m+1$ iterations of the Newton-Raphson method applied to the appropriate likelihood equations. See Theorem 4 of Bickel *et al.* (1985).

e) In the situation of d) suppose that F_m (iv)-(vi) hold and that, uniformly on θ compacts, for all $0 < M < \infty$,

$$(2.16) \quad P_{\theta}[|\tilde{\theta} - \theta| \geq Mr^{1-\delta}] = O(r^{m+1})$$

$$P_{\theta}[|\tilde{\theta} - \theta| \leq Mr^{m+2}] = O(r^{m+1}).$$

Let $A^* = \{x: \text{for all } j, \{\theta: |\tilde{\theta} - \theta| < M_1 r^{1-\delta}\} \subset \{\theta: |\hat{\theta}_j - \tilde{\theta}| < M_2 r^{1-\delta}\}\}$. Then uniformly on θ compacts

$$P_{\theta}[X \in A^*] = O(r^{m+1}).$$

If we redefine the set B of Section 3 so that B (ii) is replaced by

$$|\hat{\theta}^b - \theta^b| > M^* r^{m+2}, \quad |\tilde{\theta}^b - \theta^b| < r^{1-\delta}$$

then the proof of Theorems 4 and 5 goes through.

3. Proofs

We need to analyze $\pi_T(t | X)$ where we assume that X belongs to a set S on which the map $h \rightarrow T(\hat{\theta} + rh, X)$, $|h| < M r^{-\delta}$ is invertible with non-vanishing Jacobian and the matrix $\| -l_{ij}(\hat{\theta}) \| = \hat{C}$ is positive definite. We explain the transformation in more detail and give S

below. Let \hat{D} be the unique lower triangular matrix with positive diagonal such that

$$(3.1) \quad \hat{D} \hat{D}^T = \hat{C}$$

and

$$(3.2) \quad L(\eta) = l(\hat{D}^{-1}\eta).$$

If $\|l_{ij}(\hat{\theta})\|$ is the Hessian of l at $\hat{\theta}$ and $\hat{\eta} = \hat{D} \hat{\theta}$ then in the usual notation,

$$(3.3) \quad -L_{ij}(\hat{\eta}) = J,$$

the $p \times p$ identity. This in the Bayesian domain corresponds to standardizing the Fisher information at θ to be J as is done in the corresponding frequentist calculations. Further define $\hat{\eta}_j$

by:

$$(3.4) \quad L(\hat{\eta}_j) = \max \{L(\gamma) : \gamma^1 = \eta^1, \dots, \gamma^j = \eta^j\}$$

and

$$(3.5) \quad \tilde{T}^i(\eta) = r \left[2(L(\hat{\eta}_{i-1}) - L(\hat{\eta}_i)) \right]^{1/2} \operatorname{sgn}(\hat{\eta}_{i-1}^i - \eta^i).$$

It is easy to verify that

$$(3.6) \quad T(\hat{\theta} + rh) = \tilde{T}(\hat{\eta} + r \hat{D}h).$$

Now $\hat{D} r^{-1}(\theta - \hat{\theta})$ has posterior density

$$(3.7) \quad \pi(\hat{D}^{-1}h | \mathbf{X}) \propto \det(\hat{D})^{-1}$$

and hence,

$$(3.8) \quad \pi_T(t | \mathbf{X}) = \exp\left(-\frac{1}{2} \sum_{i=1}^p (t^i)^2 \pi(\hat{D}^{-1}(\hat{\eta} + rh(t))) \det\|h_j^i(t)\| / M(\mathbf{X})\right)$$

where $h(t)$ is defined by

$$(3.9) \quad \tilde{T}(\hat{\eta} + rh(t)) = t$$

and

$$h_j^i(t) = \frac{\partial h^i}{\partial t_j}(t)$$

and

$$M(\mathbf{X}) = \int \exp(-1/2(\sum_{i=1}^p (t^i)^2) \pi(\hat{D}^{-1}(\hat{\theta} + rh(t))) \det(\|h_j^i(t)\|).$$

For fixed \mathbf{X} , let R_X be the image of $\{h : |h| < Mr^{-\delta}\}$ under the map $h \rightarrow T(\hat{\theta} + rh, \mathbf{X})$.

From (3.8) it's clear that our task in proving Theorem 1 is to exhibit the set S such that, for

$t \in R_X$, h is uniquely defined by (3.9) and such that

$$(3.10) \quad h(t) = t + r P(t, \mathbf{X}) + O(r^{m+1})$$

$$(3.11) \quad h_j^i(t) = \delta_{ij} + r P_{ij}(t, \mathbf{X}) + O(r^{m+1})$$

where P and P_{ij} are polynomials in t and to identify the order of the polynomials. Here

$O(r^{m+1})$ means that the remainder is bounded on S by Mr^{m+1} for a generic constant M

independent of n .

We define B as the set where

$$(i) \quad \sup \left\{ \left| \pi_{i_1 \dots i_{m+2}}(\hat{\theta} + rv) \right| / \pi(\hat{\theta}) : |v| \leq Mr^{-\delta} \right\} \leq r^{-\delta}$$

$$(ii) \quad M^* r^{m+2} < |\hat{\theta}^b - \theta^b| < r^{1-\delta}, \quad 1 \leq b \leq p$$

$$(iii) \quad \sup \left\{ |\Delta_{i_1 \dots i_k}(\hat{\theta} + rv)| : |v| \leq Mr^{-\delta} \right\} \leq \varepsilon_n.$$

Note that, by B_m ,

a) $P \left[(r^{-1}(\hat{\theta} - \theta), \mathbf{X}) \in B^c \right] = O(r^{m+1})$

b) The \mathbf{x} sections of B intersect each quadrant in an open convex set since $|\cdot|$ is the l_1 -norm.

There exists a generic constant $C > 0$ such that on B ,

c) $\sup \left\{ |l_{i_1} \dots i_k}(\hat{\theta} + rh)| : |h| \leq Mr^{-\delta} \right\} \leq C$

d) $C^{-1} \leq \lambda \leq \bar{\lambda} \leq C$ where $\lambda, \bar{\lambda}$ are the minimal and maximal eigenvalues of $\| -l_{ij}(\hat{\theta}) \|$

e) $|\hat{\theta}_j - \hat{\theta}_{j-1}| \leq M_2 r^{1-\delta}, |\hat{\theta}_j - \hat{\theta}| \leq M_1 r^{1-\delta}$.

We let \tilde{S} be the image of B under the map $(h, x) \rightarrow (T(\theta(x) + rh, x), x)$ and S be just the projection of \tilde{S} on the \mathbf{x} axis, i.e., the set of all \mathbf{x} satisfying (i) and (iii) above.

Convention: Expressions such as $\hat{\eta}_i(\eta)$ are calculated at $\eta = \hat{\eta} + rh$.

Lemma 3.1. On B , for $j \geq i+1$

$$(3.12) \quad \hat{\eta}_i^j = \hat{\eta}^j + \sum_{k=2}^{m+1} N_{b_1 \dots b_k}^{ij} r^k h^{b_1} \dots h^{b_k} + O(r^{m+1})$$

where $N_{i_1 \dots i_k}$ are polynomials in the derivatives $L_{i_1} \dots i_t$ of L (evaluated at $\hat{\eta}$), with $t \leq k$

and $h = r^{-1}(\eta - \hat{\eta})$ with no constant term. Let $d = \hat{\eta}_{i-1}^i - \eta^i$. Then

$$(3.13) \quad \hat{\eta}_{i-1}^i - \hat{\eta}_i^j = \sum_{k=1}^{m+1} M_k^{ij} d^k + O(|d|^{m+2})$$

where M_k^{ij} are polynomials in $L_{i_1} \dots i_k$ and rh which vanish at $h = 0$.

Proof: Write L_{ab} etc. for derivatives of L evaluated at $\hat{\eta}$. For $j \geq i+1$

$$0 = L_j(\hat{\eta}_i) - L_j(\hat{\eta}) = L_{jb} (\hat{\eta}_i^b - \hat{\eta}^b) + \dots$$

$$(3.14) \quad + \frac{1}{(m+1)!} L_{j, b_1 \dots b_{m+1}} \prod_{k=1}^{m+1} (\hat{\eta}_i^{b_k} - \hat{\eta}^{b_k}) + O(r^{m+1}).$$

To see this note first that $\hat{\eta}_i = \hat{D} \hat{\theta}_i$ and hence in view of e) $|\hat{\eta}_i - \hat{\eta}| \leq M_3 r^{1-\delta}$. Therefore

applying c) and d) again the relevant derivatives of order up to $m + 2$ of L at $\hat{\eta}$ are bounded and (3.14) follows. Note that by (3.3) $L_{ab} = -\delta_{ab}$ and that

$$\hat{\eta}_i^b - \hat{\eta}^b = -rh^b \quad \text{for } b \leq i.$$

So we can rewrite (3.14) in the form

$$(3.15) \quad \delta_{jb} u^b = P_j(u, rh) + O(r^{m+1}), \quad j \geq i+1$$

where $u^b = \hat{\eta}_{i-1}^b - \hat{\eta}_i^b$ and P_j is a polynomial of degree $(m + 1)$ in u and rh with no term of combined degree less than 2 and bounded coefficients which are polynomials in the $L_{i_1 \dots i_j}$.

Claim (3.12) follows from a standard Lagrange inversion argument. For (3.13) write for

$$j \geq i + 1$$

$$(3.16) \quad 0 = L_j(\hat{\eta}_i) - L_j(\hat{\eta}_{i-1}) = -L_{jb}(\hat{\eta}^*) e^b$$

where $\hat{\eta}^*$ is an intermediate value and $e^b = \hat{\eta}_{i-1}^b - \hat{\eta}_i^b$.

Note that

$$(3.17) \quad e^b = 0, \quad b \leq i-1, \quad e^i = d$$

and

$$L_{jb}(\hat{\eta}^*) = -\delta_{jb} + O(r)$$

so that (3.16) yields, for $j \geq i + 1$,

$$(3.18) \quad |\hat{\eta}_{i-1}^j - \hat{\eta}_i^j| = O(r)|d|.$$

Expand further to get

$$(3.19) \quad L_{jb}(\eta_{i-1}) e^b + \cdots + \frac{1}{(m+1)!} L_{jb_1 \cdots b_{m+1}}(\eta_{i-1}) e^{b_1} \cdots e^{b_{m+1}} + O(|d|^{m+2}) = 0.$$

Rewrite (3.19) in the form

$$A_{jb} e^b + A_{jb_1 b_2} e^{b_1} e^{b_2} + A_{jb_1 \cdots b_{m+1}} e^{b_1} \cdots e^{b_{m+1}} = a_1 d + \cdots + a_{m+1} d^{m+1} + O(d^{m+1})$$

where the indices b, b_1, \dots, b_m range from $i + 1$ to p

$$A_{jb_1 \cdots b_k} = \frac{L_{jb_1 \cdots b_k}}{k!}(\eta_{i-1})$$

and the a_i are polynomials in the $L_{jb_1 \cdots b_k}(\eta_{i-1})$ and the e^b . Expand $A_{jb_1 \cdots b_k}$ around η to

$m+1 - k$ terms and use (3.12) to conclude that with remainder $O(r^{m+1})$ all the $A_{jb_1 \cdots b_k}$ are

polynomials in $L_{jb_1 \cdots b_t}$ and rh . Finally note that for $b \geq i + 1$

$$e^b = \eta_{i-1}^b - \eta_i^b = (\eta_{i-1}^b - \eta^b) - (\eta_i^b - \eta^b)$$

can by (3.12) itself be written as a polynomial of rh and $L_{jb_1 \cdots b_t}$ so that the a_j are also, up to order $m + 1$ polynomials in rh and $L_{jb_1 \cdots b_t}$, for $t \leq m + 1$. The lemma follows. \square

Lemma 3.2. On B

$$(3.19) \quad \tilde{T}^i(\eta + rh) = h^i + r^{-1} Q^i(rh) + O(r^{m+1})$$

where Q is a polynomial of degree $m + 1$ in rh with no constant or linear term and coefficients which are polynomials in $L_{b_1 \cdots b_k}$, $k \leq m + 2$.

Proof: By definition

$$(3.20) \quad \begin{aligned} \tilde{T}^i(\eta + rh) &= r^{-1} \left[- \sum_{k=1}^{m+2} \frac{2}{k!} L_{b_1 \cdots b_k}(\eta_{i-1}) \prod_{t=1}^k (\eta_{i-1}^{b_t} - \eta_i^{b_t}) \right. \\ &\quad \left. + O(|\eta_{i-1} - \eta_i|^{m+3}) \right]^{1/2} \operatorname{sgn}(\eta_{i-1}^i - \eta^i). \end{aligned}$$

Note that $L_b(\mathbf{n}_{i-1}) = 0$ for $b \geq i$ and $\mathbf{n}_{i-1}^b = \mathbf{n}_i^b$, $b \leq i-1$ so that the first term in the sum vanishes. Expand the coefficients around \mathbf{n} and use (3.18) and (3.13) to get

$$(3.21) \quad \tilde{T}^i(\mathbf{n} + rh) = r^{-1}(d + \sum_{k=2}^{m+2} c_k d^k) + O(r^{-1}|d|^{m+2})$$

where the c_k are polynomials in rh . Now substitute for d from (3.12)

$$(3.22) \quad d = rh^i + \sum_{k=2}^{m+1} N_{b_1 \dots b_k}^{i-1, i} r^k h^{b_1} \dots h^{b_k} + O(r^{m+1})$$

and the lemma follows. \square

Lemma 3.3.

(i) If O_i , $i = 1, \dots, 2^p$ are the quadrants of R^p then $\tilde{T}(\mathbf{n} + rh)$ maps $O_i \cap B_x$ into O_i for all i .

(ii) \tilde{T} is continuously differentiable on $O_k \cap B_x$ for $1 \leq k \leq 2^p$. Let

$$\tilde{T}_j^i = \frac{\partial \tilde{T}^i}{\partial h_j}$$

Then \tilde{T}_j^i is lower triangular and

$$(3.23) \quad \tilde{T}_j^i = 1 + P^i(rh) + O(r^{m+1})$$

where P^i is a polynomial of degree $m+1$ with no constant term and coefficients in $L_{b_1 \dots b_k}$, $k \leq m+2$.

(iii) \tilde{T} is 1-1.

Proof:

(i) We need to show that on B ,

$$(3.24) \quad \operatorname{sgn}(\hat{\eta}_{i-1}^i - \eta^i) = \operatorname{sgn} h, \quad i = 1, \dots, p.$$

By (3.12) on B ,

$$\hat{\eta}_{i-1}^i - \eta^i = rh^i (1 + rM_1(h)) + r^{m+2}M_2(h)$$

where M_1 is a polynomial in h with bounded coefficients and $|M_2(h)|$ is bounded by M_2 for all $(x, h) \in B$. But $(x, h) \in B \Rightarrow aM^* r^{m+1} < |h^i| < a^{-1}r^{-\delta}$ where a is positive constant depending only on the constant C of d).

Choose M^* so that

$$(3.25) \quad aM^* > M_2.$$

The relation (3.24) follows from

$$(3.26) \quad \hat{\eta}_{i-1}^i(\hat{\eta} + aM^* r^{m+2}) - \eta^i > (aM^* - M_2)r^{m+2} + O(r^{m+3}) > 0$$

and

$$\frac{d}{dh^i} \{h^i(1 + rM_1(h))\} = 1 + O(r).$$

(ii) It is easy to see that $\tilde{T}(\hat{\eta} + rh)$ is continuously differentiable on B with derivatives

$$\tilde{T}_j^i = |\tilde{T}^i|^{-1} \left[L_k(\hat{\eta}_{i-1}) \frac{\partial \hat{\eta}_{i-1}^k}{\partial h^j} - L_k(\hat{\eta}_i) \frac{\partial \hat{\eta}_i^k}{\partial h^j} \right].$$

Note that,

$$\frac{\partial \hat{\eta}_{i-1}^a}{\partial \eta^b} = 0, \quad a, b \geq i$$

$$= \delta_{ab}, \quad a \leq i-1$$

and $L_k(\hat{\eta}_{i-1}) = 0, k \geq i$. So $i < j \Rightarrow \tilde{T}_j^i = 0$ while

$$(3.27) \quad \tilde{T}_i^i = -r^{-1} |\tilde{T}^i|^{-1} L_i(\mathfrak{N}_i)$$

Now write

$$\begin{aligned}
 L_i(\mathfrak{N}_i) &= L_{ib}(\mathfrak{N}_{i-1})(\mathfrak{N}_i^b - \mathfrak{N}_{i-1}^b) + \sum_{k=1}^{m+1} \frac{L_{ib_1 \dots b_k}(\mathfrak{N}_{i-1})}{k!} \prod_{j=1}^k (\mathfrak{N}_i^j - \mathfrak{N}_{i-1}^j) \\
 &\quad + O(|\mathfrak{N}_{i-1}^i - \mathfrak{N}_i^i|^{m+2}) \\
 (3.28) \quad &= \sum_{k=1}^{m+1} P_k(rh) d^k + O(d^{m+2})
 \end{aligned}$$

by (3.13), where $d = \mathfrak{N}_{i-1}^i - \mathfrak{N}_i^i$ and P_k are polynomials in rh such that $P_1(0) = 1$. Now apply (3.21) and (3.28) to (3.27) and then substitute (3.22) for d and (ii) follows.

(iii) Follows from Lemma 1 of the Appendix. \square

Proof of Theorem 1. By Lemma 3.3 formula (3.8) is valid for $(x, t) \in \tilde{S}$. Moreover, from

Lemma 3.2

$$(3.29) \quad h^i(t) = t^i + r^{-1} P^i(rt) + O(r^{m+1})$$

where P^i is a polynomial of degree $m+1$ in rt with no constant or linear term and coefficients which are polynomials in $L_{b_1 \dots b_k}$, $k \leq m+2$. From (3.23) and (3.29)

$$\begin{aligned}
 \det \|h_j^i(t)\| &= \det \|\tilde{T}_j^i(\mathfrak{N} + rh(t))\|^{-1} = \prod_{i=1}^p \tilde{T}_i^i(\mathfrak{N} + rh(t))^{-1} \\
 &= \prod_{i=1}^p (1 + P^i(rh(t)))^{-1} + O(r^{m+1}) \\
 (3.30) \quad &= 1 + V(rt) + O(r^{m+1})
 \end{aligned}$$

where V is a polynomial of degree $m + 1$ in rt with no constant term and coefficients which are polynomials in $L_{b_1} \dots b_k$, $k \leq m + 2$.

Moreover, from (3.29) and B_m (i),

$$(3.31) \quad \pi(\hat{\theta} + r\hat{D}^{-1}h(t)) = \pi(\hat{\theta}) \left[1 + \frac{\pi_b(\hat{\theta})}{\pi(\hat{\theta})} U^b(rt) + \dots + \frac{\pi_{b_1 \dots b_{m+2}}(\hat{\theta})}{\pi(\hat{\theta})} U^{b_1}(rt) \dots U^{b_{m+2}}(rt) \right] \\ + O(r^{m+1} \pi(\theta))$$

where the U^b are polynomials of degree $\leq m + 1$ with no constant term. Substituting back (3.30) and (3.31) in (3.8) provides an approximation to the numerator in (3.8) and integrating this we get an approximation to the denominator in (3.8). Together these approximations ensure that

$$E_P \int |\pi_T(t | \mathbf{X}) - \phi(t)(1 + Q_m^*(rt, x, \pi)) \mathbf{1}[(t, \mathbf{X}) \in \tilde{S}]| dt = O(r^{m+1})$$

for a suitable Q_m^* . We get Q_m by dropping all terms of degree $m + 1$ in Q_m^* . The coefficients are evidently polynomials in $L_{b_1} \dots b_k(\hat{\theta})$ and $\frac{\pi_{b_1 \dots b_k}(\hat{\theta})}{\pi(\hat{\theta})}$, $1 \leq k \leq m + 1$. But the former are polynomials in the elements of \hat{D}^{-1} which are rational functions of $L_{ij}(\hat{\theta})$.

Now,

$$(3.32) \quad E_P \int \phi(t)[Q_m(rt, x, \pi) - Q_m^*(rt, x, \pi)] \mathbf{1}[(t, \mathbf{X}) \in \tilde{S}] dt = O(r^{m+1})$$

since for $\mathbf{x} \in S$ all coefficients in both functions are bounded. Further,

$$(3.33) \quad E_P \int \pi_T(t | \mathbf{X}) \mathbf{1}((t, \mathbf{X}) \notin \tilde{S}) dt = P[(T, \mathbf{X}) \notin \tilde{S}] = O(r^{m+1})$$

by B_m . Finally,

$$E_P \int \phi(t) Q_m(rt, x, \pi) 1(X \in S, |t| \leq M^* r^{m+1} \text{ or } |t| \geq r^{-\delta}) dt = O(r^{m+1})$$

and the theorem follows. \square

Proof of Theorem 2 and Corollary 1: Evidently since D and \tilde{D} are simple transforms of T

we need merely check that the approximation to the density of D , (\tilde{D} respectively) obtained by

applying the usual transformation formula to $\pi_m(\cdot, X)$ agree with $\prod_{k=1}^p c_1(u^k)$ with error

$O(r^{m+1})$ for $m = 1, 3$ respectively. This follows readily from Lemmas A2 and A3 in the

appendix if we identify π_m with $g(t)$ for $m = 2, 3$, and note that $R_{jj} = O(n^{-1})$. Relation (2.6)

follows from Lemma A2 and A3. Corollary 1(a) follows immediately from (2.5), while 1(b)

follows from (2.6) and Lemma A4. \square

Proof of Theorem 3: Evidently $F_m \Rightarrow B_m$ for π satisfying (vii). It is shown in Ghosh, Sinha

and Joshi (1982) and Bickel *et al.* (1985) that the set of all such π is dense in the set of all pri-

ors under weak convergence. Now (2.9) implies that for any π concentrating on a compact,

the characteristic function of T satisfies the approximation

$$\begin{aligned}
 (3.34) \quad \int e^{i v_j t^j} p_T(t) dt &= \iint e^{i v_j t^j} p_T(t | \theta) \pi(\theta) d\theta dt \\
 &= e^{i v_j t^j} \phi(t) \left[1 + \sum_{k=1}^m r^k \int R_k(t, \theta) \pi(\theta) d\theta \right] dt + O(r^{m+1}) \\
 &= \exp\left\{-\frac{1}{2} \sum_{j=1}^p (v^j)^2\right\} \left[1 + \sum_{k=1}^m r^k \int P_k(v, \theta) \pi(\theta) d\theta \right] + O(r^{m+1})
 \end{aligned}$$

where $\exp\left\{-\frac{1}{2} \sum_{j=1}^p (v^j)^2\right\} P_k(v, \theta)$ is the Fourier transform of $\phi(t) R_k(t, \theta)$, so that the P_k 's are

also polynomials in v . On the other hand, Theorem 1 yields

$$\begin{aligned}
 \int \exp\left\{\sum_{j=1}^p (v^j)^2\right\} p_T(t) dt &= E_P \left[\int \exp\left\{\sum_{j=1}^p (v^j)^2\right\} \pi_m(t, \mathbf{X}) 1(\mathbf{X} \in S) dt \right] + O(r^{m+1}) \\
 &= \exp\left\{-\frac{1}{2} \sum_{j=1}^p (v^j)^2\right\} \left[1 + \sum_{k=1}^m r^k t^{b_1} \cdots t^{b_k} E Q_{mb_1 \cdots b_k}(\mathbf{X}, \pi) 1(\mathbf{X} \in S) \right] \\
 (3.35) \quad &\quad + O(r^{m+1})
 \end{aligned}$$

Therefore, multiplying by $\exp\left\{\frac{1}{2} \sum_{j=1}^p (v^j)^2\right\}$ we get

$$\begin{aligned}
 1 + \sum_{k=1}^m r^k \int P_k(v, \theta) \pi(\theta) d\theta &= 1 + \sum_{k=1}^m r^k c_{b_1 \cdots b_k}(\pi) v^{b_1} \cdots v^{b_k} \\
 (3.36) \quad &\quad + O(r^{m+1})
 \end{aligned}$$

where O is now uniform for $|v| \leq M$ by the hypothesis of Theorem 3

Define, as usual

$$\Delta_{b_1 \cdots b_p} f(t^1, \dots, t^p) = \begin{bmatrix} \Delta_1^{b_1} & \cdots & \Delta_p^{b_p} \end{bmatrix} f(t^1, \dots, t^p)$$

where the $b_j = 0, \dots, p$, $\sum_{j=1}^p b_j = l$ and

$$\Delta_k f = f(t^1, \dots, t^{k-1}, t^k + \varepsilon, t^{k+1}, \dots, t^p) - f(t^1, \dots, t^p)$$

and Δ_k^p represents an operator product. Apply $\Delta_{b_1 \cdots b_p}$ to both sides of (3.36) considered as

functions of v . If $l > m$ we obtain

$$(3.37) \quad \sum_{j=1}^m r^j \varepsilon^{-l} \int \Delta_{b_1 \cdots b_p} P_j(\varepsilon, \theta) \pi(\theta) d\theta = O(r^{m+1} \varepsilon^{-l}).$$

Let $\varepsilon \downarrow 0$ more slowly than $r^{\frac{1}{l}}$. Then (3.37) yields

$$\int \frac{\partial^p P_k}{\partial^{b_1} u_1 \cdots \partial^{b_p} u_p}(v, \theta) \pi(\theta) d\theta = 0 \quad \text{for all } v, \text{ for all } k \leq m.$$

But by assumption the integrand is continuous in θ . Since π ranges over a dense set we conclude that the integrand vanishes identically in θ . So P_k is a polynomial of degree $\leq k$ and hence so is R_k . \square

Theorem 4 and Corollary 3 follow from Theorem 3 in the same fashion as Theorem 2 and Corollary 1 follow from Theorem 1. \square

Acknowledgement

We thank Ole Barndorff-Nielsen, Ib Skovgaard and Peter McCullagh for some crucial references.

APPENDIX

Lemma A1. Suppose $f : C^\circ \rightarrow R^p$ where C° is an open convex set in R^p . Suppose f is differentiable with Hessian \dot{f} and,

$$(A.1) \quad |\dot{f} - J| < 1$$

where J is the identity and $|M|$ is the operator norm on matrices. Then \dot{f} is nonsingular and f is 1-1.

Proof: By (A1), \dot{f} is nonsingular,

$$\dot{f}^{-1} = J - (\dot{f} - J) + (\dot{f} - J)^2 \dots$$

If $f(a) = f(b)$ then

$$0 = \int_0^1 \dot{f}(a + \lambda(b - a)) d\lambda(b - a)$$

or

$$(b-a) = -\int_0^1 (f(a + \lambda(b-a)) - J) d\lambda(b-a).$$

Then, by (A.1)

$$|b-a| \leq \max_{\lambda} |\dot{f}(a + \lambda(b-a)) - J| |b-a| < |b-a|$$

unless $b=a$.

Lemma A2. Let

$$g(t) = \phi(t) \left[1 + R_i t^i + R_{ij} (t^i t^j - \delta^{ij}) + R_{ijk} t^i t^j t^k \right]$$

be the density of a finite measure μ on \mathbb{R}^p where δ^{ij} is Kronecker delta, and let

$$g_o(t) = \phi(t) \left[1 + R_{jj} ((t^j)^2 - 1) \right]$$

similarly correspond to μ_o . Let $h(t) = (t^i)^2$. Then

$$\mu h^{-1} = \mu_o h^{-1}.$$

Proof: The densities of μh^{-1} and $\mu_o h^{-1}$ at $(|u^1|, \dots, |u^p|)$ differ by the term

$$2^p E \left[R_i \varepsilon_i |u^i|^{1/2} + R_{ij}^* \varepsilon_i \varepsilon_j |u^i|^{1/2} |u^j|^{1/2} + \varepsilon_i \varepsilon_j \varepsilon_k R_{ijk} |u^i|^{1/2} |u^o|^{1/2} |u^k|^{1/2} \right] = 0$$

where $R_{ij}^* = R_{ij}(1 - \delta_{ij})$ and ε_i are independent ± 1 with probability $1/2$. \square

Lemma A3. Suppose $\sum_j |R_{jj}| = o(1)$. Then

$$\int g_o(t) - \prod_{j=1}^p (1 + 2R_{jj})^{-1/2} \phi_1(t^j(1 + 2R_{jj})^{-1/2}) dt = O(\sum_j R_{jj}^2)$$

where ϕ_1 is the standard normal density.

Proof: Taylor expand. \square

Lemma A4. Suppose $\sum_1^k |c_{jj} - 1| = o(1)$ and $Z_j^2 = c_j Y_j^2$ where Y_j 's are i.i.d. $N(0, 1)$. Let

$U = \sum_1^k Y_j^2$ and $V = (\sum Z_i^2)(1 + \sum(c_j - 1)/k)^{-1}$. Then U and V have the same characteristic function up to $O(\sum(c_j - 1)^2)$.

Proof. Compute the characteristic function of V , take logarithms and expand. \square

REFERENCES

- [1] Barndorff-Nielsen, O. E. (1983). On a formula for the conditional distribution of the maximum likelihood estimator. *Biometrika* **70**, 343-65.
- [2] Barndorff-Nielsen, O. E. and Cox, D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. R. Statist. Soc. B* **46**, 483-95.
- [3] Barndorff-Nielsen, O. E. and Cox, D. R. (1986). Inference on full or partial parameters based on the standardized log likelihood ratio statistic. *Biometrika* **73**, 307-322.
- [4] Barndorff-Nielsen, O.E. and Hall, P. (1988). On the level error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* **75**, 374-378.
- [5] Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. R. Soc. A* **160**, 268-282.
- [6] Bhattacharya, R.N. and Ghosh, J.K. (1978). Validity of formal Edgeworth expansion. *Ann. Statist.* **6**, 434-451.
- [7] Bickel, P.J. (1974). Edgeworth expansions in nonparametric statistics. *Ann. Statist.* **2**, 1-20.
- [8] Bickel, P. J., Götze, F., and van Zwet, W. R. (1985). A simple analysis of third order efficiency of estimates. *Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack*

Kiefer, L. LeCam and R. A. Olshen, eds. Wadsworth Inc., Belmont.

- [9] Billingsley, P. (1961). *Statistical Inference for Markov Processes*. The University of Chicago Press, Chicago.
- [10] Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* **36**, 317-346.
- [11] Chandra, T. and Ghosh, J. K. (1979). Valid asymptotic expansion for the likelihood ratio statistic and other perturbed χ^2 variables. *Sankhyā A* **41**, 22-47.
- [12] Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. SIAM, Philadelphia.
- [13] Efron, B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika* **72**, 45-58.
- [14] Ghosh, J. K., Sinha, B. K., and Joshi, S. M. (1982). Expansions for posterior probability and integrated Bayes risk. *Proc. III Purdue Symp. on Decision Theory and Related Topics*, vol. II, J. Berger and S. Gupta, eds. Academic Press, New York.
- [15] Götze, F. and Hipp, C. (1978). Asymptotic expansions under moment conditions. *Z. Warsch* **42**, 67-87.
- [16] Lawley, D. N. (1956). A general method for approximating to the distribution of the likelihood ratio criteria. *Biometrika* **43**, 295-303.

[17] Pfanzagl, J. (1974). Asymptotically optimum estimation and test procedures. In: *Proceedings of Prague Symposium on Asymptotic Statistics v.1* (J. Hájek, Ed.), 201-272. Charles University, Prague.

[18] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Mat. Soc.* **54**, 426-482.

[19] Welch, B. N. and Peers, B. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *JRSS B* **35**, 318-329.

[20] Wilks, S. S. (1938). The large sample distribution of the likelihood ratio statistic for testing composite hypotheses. *Ann. Math. Statist.* **9**, 60-62.

[21] Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. *Sequential Methods in Statistics: Banach Center Publication* **16**, 485-514.

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 $E \max_{1 \leq k \leq n} S_k^+ / ES_n^+$: 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. (Dec. 1987, revised June 1989). 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. CANCELLED
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. (Mar. 1988, revised May 1989). Asymptotics for multivariate trimming.
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 subproblems.
179. DONOHO, D.L. and STARK, P.B. (October 1988). Recovery of sparse signal when the low frequency information is missing.
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, Aug. '89). Models for variable 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 2º 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.

206. MILASEVIC, P. and NOLAN, D. (May 1989). Estimation on the sphere: A geometric approach.
207. SPEED, T.P. and YU, B. (July 1989). Stochastic complexity and model selection: normal regression.
208. DUBINS, L.E. (June 1989). A group decision device: Its pareto-like optimality.
209. BREIMAN, L. (July 1989). Fitting additive models to regression data.
210. PEMANTLE, R. (July 1989). Vertex-reinforced random walk.
211. LE CAM, L. (August 1989). On measurability and convergence in distribution.
212. FELDMAN, R.E. (July 1989). Autoregressive processes and first-hit probabilities for randomized random walks.
213. DONOHO, D.L., JOHNSTONE, I.M., HOCH, J.C. and STERN, A.S. (August 1989). Maximum entropy and the nearly black object.
214. DONOHO, D.L. (August, 1989). Statistical estimation and optimal recovery.

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.