On the Bernstein-von Mises theorem

by

Lucien Le Cam

University of California, Berkeley

Technical Report No. 57
February 1986

*Research Partially Supported by
National Science Foundation Grant DMS85-40314

Department of Statistics
University of California
Berkeley, California

On the Bernstein-von Mises theorem *

by Lucien Le Cam
University of California, Berkeley

1. Introduction

The present paper is essentially a collection of known, but perhaps unpublished, results on the approximability of posterior distributions by Gaussian ones. The initial result is due to Laplace (1809) but the most common appellation is "Bernstein-von Mises theorem". The paper came about as an answer to two questions. The first, posed by Albert Lo, is whether the differentiability in quadratic mean conditions that, in the i.i.d case, insure the validity of the LAN conditions also yield the Bernstein-von Mises phenomenon. The second, posed by David Blackwell was complex but related to the query: "What are necessary and sufficient conditions for the Bernstein- von Mises theorem?"

The fact that non-trivially tautological conditions do not seem to be known prompted the writing of the paper. In it, after stating the problem in Section 2, we start by an example of sufficient conditions in the i.i.d case. This is for two reasons: to answer Albert Lo's question and to prepare the reader for the remainder by pointing out the various features that have to be checked. Indeed, the proof involves three steps: 1) a global step where one proves a consistency type of a result, 2) a semi local step where one proves that the posterior distributions concentrate on sets that have diameter of the order of $1/\sqrt{n}$ and finally, 3) a strictly local study on these sets. The details of the proofs are not given. They can be found in the author's forthcoming monograph (Le Cam, 1986).

This constitute Section 3. Section 4 gives necessary and sufficient conditions for the validity of what corresponds to the strictly local arguments. However, it does so under some restrictions that are not too terrible but not actually necessary for the validity of Gaussian approximations.

Section 5 gives some sufficient conditions for the case of independent observations that, individually, contain a negligible amount of information. Here we can carry out the first two steps described in Section 3 at one stroke, but at the price of fairly severe assumptions.

In all sections we refer for many proofs or parts of proofs to (Le Cam 1986).

* partially supported by National Science Foundation, DMS 85-40314

It will be seen that the third, strictly local part of the argument can be carried out under the LAN or LAMN conditions. The latter are often satisfied when observations are taken on stochastic processes instead of independent sequences. However we have refrained from saying anything about such dependent cases because it is difficult to write out nice, simple and general conditions that permit to carry out the global and semi-local steps of the proofs. There do exist such conditions in the literature. See for instance I. Basawa and B. L. S. Prakasa Rao (1980).

2. Notations and statement of the problem.

Let $\mathbf{X}$ and Θ be two sets. Let $\mathbf{A}$ be a σ-field of subsets of $\mathbf{X}$ and let $\mathbf{B}$ be a σ-field of subsets of Θ. Let S be a positive measure on $\mathbf{A}\times\mathbf{B}$. Assuming some regularity conditions one can disintegrate S, writing it in the form

$$S(dx,d\theta) = S'(dx)F_x(d\theta) = P_\theta(dx)\mu(d\theta)$$

where each F_x and each P_θ is a probability measure and where S' and μ are the corresponding marginal measures. The general question one could ask: "Under what conditions on the family $\{P_\theta: \theta \in \Theta\}$ and the measure μ can one approximate the posterior distributions F_x by Gaussian ones?"

For this to make sense the space $(\Theta,\mathbf{B})$ must be suitable for the definition of Gaussian distributions. This can be done for all linear spaces and for certain groups.

Here we shall concern ourselves only with the cases where Θ is a finite dimensional vector space except that we shall say a few words about subsets of such spaces.

To make the question precise, one must specify the type of approximation sought. We shall concern ourselves only with approximation for the total variation norm, also called L_1-norm. Thus if G_x is the presumed approximation, the distance will be measured by

$$||F_x - G_x|| = \sup_\phi |\int \phi(\theta)[F_x(d\theta) - G_x(d\theta)]|$$

taken over all $\mathbf{B}$-measurable ϕ such that $|\phi| \le 1$.

There is also the matter of the fact that in F_x the variable x moves through $\mathbf{X}$. We shall consider the convergence to zero of integrals such as $\int ||F_x - G_x||S'(dx)$ or $\int ||F_x - G_x||P_{\theta_0}(dx)$ for a selected point θ_0. This means, in particular, that we shall not bother with statements such as: $||F_x - G_x||$ tends to zero almost surely for P_{θ_0}. These may be nice, but for the situations described in Sections 4 and 5, they do not make sense. Thus, to simplify life, we have omitted them.

As mentioned above, we need some regularity conditions to insure the existence of disintegrations of the measure S. It usually comes in the form $S(dx,d\theta) = P_\theta(dx)\mu(d\theta)$. So there is no problem on that part. For the other side, $S'(dx)F_x(d\theta)$, we shall always assume that Θ is a Borel set in a Polish space and that S' is σ-finite.

3. Sufficient conditions for the i.i.d case

In the standard i.i.d. case, one considers probability measures $\{p_\theta: \theta \in \Theta\}$ on some σ-field, say $\mathbf{F}$, and, for each integer n, one takes n replicates of the system $\{\mathbf{F},\{p_\theta: \theta \in \Theta\}\}$. The measures under study are the product measures $P_{\theta,n}$ corresponding to n independent identically distributed observations. Note that the system $\{p_\theta: \theta \in \Theta\}$ remains fixed, independent of n. The case where one takes n replicates of systems $\{p_{\theta,n}: \theta \in \Theta\}$, (a $\theta \in \Theta_n$) that change as n changes and goes to infinity is different. It will be treated under other assumptions in Section 5.

A possible set of assumptions is as follows.

(A1) $\theta_0 = 0$ and Θ is the intersection of a closed set with an open set. (The $\theta_0 = 0$ is to avoid excess of indices.)

(A2) The equality $p_s = p_t$ implies $s = t$.

(A3) If $t_\nu \to t$ for the Euclidean topology of Θ then $\int \phi dp_{t_\nu} \to \int \phi dp_t$ for every bounded measurable ϕ.

(A4) There is a compact $K \subset \Theta$, an $\epsilon < 1/2$, and an integer n such that there is a test function ω_n, $0 \leq \omega_n \leq 1$, based on n observations with $\int (1-\omega_n) dP_{0,n} < \epsilon$ and $\int \omega_n dP_{\theta,n} < \epsilon$ for all $\theta \in K^c$.

(A5) A process ξ with covariance kernel $E\xi(s)\xi(t) = \int \sqrt{dp_s dp_t}$ admits at $\theta = 0$ a derivative in quadratic mean that has a non singular covariance matrix.

(A6) If p_t'' is the part of p_t that is p_0 singular, then $\frac{1}{|t|^2}\|p_t''\|$ tends to zero as $t \to 0$.

(A7) The prior measure μ is independent of n. It is a probability measure that has a density f with respect to the Lebesgue measure λ.

(A8) Let $B(\epsilon)$ be the ball of radius ϵ centered at zero. There is some number $a \in (0,\infty)$ such that

$$\lim_{\epsilon \to 0} \frac{1}{\lambda[B(\epsilon)]} \int I_{B(\epsilon)}^{(t)} |f(t) - a| \lambda(dt) = 0.$$

Furthermore

$$\liminf_{\epsilon \to 0} \frac{\lambda[B(\epsilon) \cap \Theta]}{\lambda[B(\epsilon)]} > 0.$$

Under these conditions, consider estimates T_n and Γ_n, defined on the first n observations, such that the values of T_n are in $\mathbb{R}^k$ and that those of Γ_n are positive definite matrices on $\mathbb{R}^k$. Let $H_{x,n}$ be the measure defined by

$$H_{x,n}(B) = \int_{B \cap \Theta} \exp\{-\frac{n}{2}(t - T_n)\Gamma_n(t - T_n)\} \lambda(dt),$$

and let

$$G_{x,n} = ||H_{x,n}||^{-1}H_{x,n}$$

Theorem 1. Let all the conditions (A1) to (A8) be satisfied. Let $F_{x,n}$ be the posterior distribution of θ given the value x of the first n observations. Then there exist estimates T_n and Γ_n such that

$$\int ||F_{x,n} - G_{x,n}|| P_{0,n}(dx)$$

tends to zero as $n \to \infty$.

Proof. See Le Cam (1986) Chapter 17, Section 7. The proof involves several steps. The first one is as follows. Let C be the complement of a neighborhood of zero. One first shows that $F_{x,n}(C) \to 0$ in $P_{0,n}$ probability. For this the first four assumptions are used as well as assumptions (A7) and (A8). These latter imply that μ is not too thin around zero.

The next step is to consider sets C_n of the type $\{\theta: \theta \in \Theta, |\theta| \geq \frac{b_n}{\sqrt{n}}\}$ where b_n tends to infinity as $n \to \infty$. One shows that $F_{x,n}(C_n)$ also tends to zero in probability. This allows one to reduce the considerations to sets $\Theta \cap V_n$ with $V_n = \left\{\theta: |\theta| \leq \frac{b}{\sqrt{n}}\right\}$ for fixed b. To these one applies an argument similar to the one described in Section 4. One shows that

$$\int ||F_{x,n} - G_{x,n}|| S_n'(dx) \to 0$$

for measures $S_n' = \frac{1}{\mu(V_n)} \int_{V_n} P_{\theta,n}\mu(d\theta)$. Then one replaces S_n' by $P_{0,n}$ by a contiguity argument. This local part uses restriction (A5) to obtain a linear + quadratic approximation to the logarithms of likelihood ratios.

Note that we have not assumed that zero is interior to Θ. Thus our $G_{x,n}$ are not necessarily Gaussian, but Gaussian truncated to Θ. This is a small refinement that is sometimes useful. We shall not bother with such refinements in the sequel. We shall, however, need to consider in Section 5 subsets of Euclidean spaces defined otherwise than the $V_n = \{\theta: |\theta| \leq \frac{b}{\sqrt{n}}\}$ and we shall have to <u>assume</u> that they look like the V_n's.

One could ask whether the conditions imposed here are necessary. The answer is, of course, certainly not. We have worked at the rate $\sqrt{n}$. One can obtain results for other rates. We have assumed the existence of uniformly consistent tests in (A4) when it would be enough to assume existence of consistent tests of $P_{0,n}$ against $\frac{1}{\mu(K^c)} \int_{K^c} P_{\theta,n}\mu(d\theta)$. We have kept μ independent of n. It could be made a measure of variable norm dependent of n, but tending to something that is, near zero, like the Lebesgue measure on Θ, as long as one imposes growth condition on its tails and so forth.

However, our aim here was not to look for utmost generality, but to point out the types of arguments involved. As we shall see, the purely local arguments can be generalized immensely. The semi-local and global arguments need a lot of care. Conditions under which they can be carried out are not simple to formulate in more complex cases.

4. The local part of the theory

In this section it will be assumed that the parameter space Θ is a Euclidean space, $\Theta = \mathbb{R}^k$. For each integer n, one considers an experiment $\{P_{\theta,n}: \theta \in \Theta\}$ on a σ-field $\mathbf{A}_n$ and a prior measure μ_n. We shall also assume given a function Γ_n that is $\mathbf{A}_n$-measurable and takes its values in the space of positive definite symmetric matrices on $\mathbb{R}^k$. As before, the conditional distributions of θ given $\mathbf{A}_n$ will be denoted $F_{x,n}$. The question is whether one can find $\mathbf{A}$-measurable functions T_n such that if $G_{x,n}$ is the Gaussian measure whose density with respect to the Lebesgue measure λ is proportional to

$$\exp\{-\frac{1}{2}(\theta - T_n)'\Gamma_n(\theta - T_n')\}$$

then the $F_{x,n}$ are approximable by the $G_{x,n}$. Note that, here, the exponent in the Gaussian density uses the matrices Γ_n themselves, not the n Γ_n of Section 3. This means that we assume that all necessary normalizations and transformations have already been carried out. We shall answer the question under the following restrictions.

Let $\mu_{n,s}$ be the prior measure μ_n shifted by s. That is, $\mu_{n,s}$ is defined by the equality

$$\int \gamma(\theta)\mu_{n,s}(d\theta) = \int \gamma(\theta + s)\mu_n(d\theta)$$

(B1) The sequence $||\mu_n||$ is bounded. For every fixed $c \in (0,\infty)$, the quantity $\sup_s\{||\mu_{n,s} - \mu_n||;\ |s| < c\}$ tends to zero as n tends to infinity.

Let S_n be the joint measure defined by $S_n(dx,d\theta) = P_{\theta,n}(dx)\mu_n(d\theta)$.

(B2) For each $\epsilon > 0$ there is an integer N and a number $b < \infty$ such that $n \geq N$ implies

$$S_n\{||\Gamma_n|| + ||\Gamma_n^{-1}|| > b\} < \epsilon.$$

To state the result we shall use the following notation. Let $S_{n,t}$ be S_n shifted by t. Explicitly $S_{n,t}$ is defined by the equality

$$\int\int u(x,\theta)S_{n,t}(dx,d\theta) = \int\int u(x,\theta + t)S_n(dx,d\theta).$$

Let $\Lambda_{n,t}$ be the logarithm of likelihood ratio $\Lambda_{n,t} = \log\frac{dS_{n,t}}{dS_n}$.

Let S_n' be the marginal distribution of x for S_n.

Theorem 2. Let condition (B1) and (B2) be satisfied. Then the existence of Gaussian kernels $G_{x,n}$ with matrices Γ_n such that

$$\int ||G_{x,n} - F_{x,n}||S_n'(dx)$$

tends to zero is equivalent to the combination of the following conditions:

a) For every sequence $\{s_n\}$ such that $s_n \to 0$ the norms $||S_{n,s_n} - S_n||$ tend to zero.

b) For every $t \in \mathbb{R}^k$ the sequences $\{S_{n,t}\}$ and $\{S_n\}$ are contiguous.

c) For every pair (s,t) of elements of $\mathbb{R}^k$ the difference

$$\Lambda_{n,s+t} - \Lambda_{n,s} - \Lambda_{n,t} + s'\Gamma_n t$$

tends to zero in S_n-probability.

d) There are $\mathbf{A}_n$-measurable functions $\hat{\theta}_n$ such that for every $\epsilon > 0$ there exist integers N and b such that $n \geq N$ implies $S_n\{|\hat{\theta}_n - \theta| > b\} < \epsilon$.

The proof is long and tortuous. It can be found in Le Cam (1986). We shall sketch some of its ingredients below, but first a few words about the assumptions and the statement (a)-(b). Assumption (B1) is not necessary for the validity of Gaussian approximations to posterior distributions. However, it helps. What it says is that, asymptotically, the prior distributions behave somewhat like the Lebesgue measure. Assumption (B2) says that, in an appropriate scale, the presumed Gaussian approximations do not stray to infinity and do not degenerate. One can readily conceive of situations where Gaussian approximations do exist but where the relevant matrices cannot be so controlled. The statements (a) and (b) of the theorem are due in large part to this control over the size of the matrices. Statement (c) in the theorem is an expression of the requirement that log likelihoods must be approximately quadratic on Θ. Finally, the requirement (d) says that if the posterior distributions are approximately Gaussian, with well behaved matrices, then a Bayesian statistician who believes in his prior can estimate θ successfully. Thus, the requirements (a) to (d) are not surprising. That they do imply the stated Gaussian approximability is not so easy to see. One must first find the appropriate centering variables T_n. This can be done by correcting the $\hat{\theta}_n$ in a manner that is very similar to the procedure used by the present author (Le Cam, 1960) under the LAN conditions and by R. B. Davies, (1985) and others under the LAMN conditions. See also Le Cam (1986), Chapter 11. The bound on the integrals of $||F_{x,n} - G_{x,n}||$ can be derived through the use of a simple inequality as follows.

Let S_1 and S_2 be two finite positive measures on $\mathbf{A}_n \times \mathbf{B}$ where $\mathbf{B}$ is the Borel field of Θ. Let S_0 be another such measure that dominates both S_1 and S_2. Each S_i admits a disintegration $S_i(dx,d\theta) = S_i'(dx)F_{x,i}(d\theta)$. A possible disintegration of S_i, for $i = 1,2$, is given by

$$\phi_i'(x)F_{x,i}(d\theta) = \phi_i(x,\theta)F_{x,0}(d\theta)$$

where $\phi_i = \dfrac{dS_i}{dS_0}$ and $\phi_i'(x) = \int \phi_i(x,\theta)F_{x,0}(d\theta)$.

Now let r be defined by

$$r(x,\theta) = \frac{\phi_2(x,\theta)}{\phi_1(x,\theta)} \frac{\phi_1'(x)}{\phi_2'(x)}$$

on the set $A_1 \cap A_2$ with $A_i = \{x: 0 < \phi_i'(x) < \infty\}$ this is also equal to

$$\int \frac{\phi_2(x,\theta)}{\phi_1(x,\theta)} \frac{\phi_1(x,\xi)}{\phi_2(x,\xi)} F_{x,2}(d\xi).$$

Then

$$||F_{x,1} - F_{x,2}|| = 2\int \{1 - \min[1,r(x,\theta)]\} F_{x,1}(d\theta)$$

on $A_1 \cap A_2$. On the complement of A_i one can define $F_{x,i}$ arbitrarily and therefore make $F_{x,1}$ and $F_{x,2}$ coincide outside of $A_1 \cap A_2$. Thus to prove that the integrals $\int ||F_{x,1} - F_{x,2}|| S_{n,1}'(dx)$ are small, it is sufficient to show that the ratio r_n is close to unity in $S_{n,1}$ probability.

Once the appropriate centerings T_n have been obtained it is possible, with some effort, to look at the appropriate ratios $\dfrac{\phi_{2,n}(x,\theta)}{\phi_{1,n}(x,\theta)}$ and, even though effort is needed, the argument is not terribly difficult.

5. The case of independent observations

In this section we consider a situation where one has a finite or infinite sequence of experiments $E_j = \{p_{\theta,j}: \theta \in \Theta\}$, $j \in J$, and where, for a given θ, the observations X_j; $j \in J$ are independent with X_j distributed according to $p_{\theta,j}$. To obtain an asymptotic theory, the entire system will be made to depend on an integer n, so that Θ becomes Θ_n and $p_{\theta,j}$ becomes $p_{\theta,j,n}$, for instance. Also there will be a prior measure μ_n onΘ_n. Our aim is to give certain conditions under which the posterior distributions $F_{x,n}$ can be shown to concentrate on small neighborhoods of the true value θ_0, thus carrying out at the same time the global and semi-local steps described in Section 3. Then we pass to the task of describing conditions leading to asymptotic normality of the $F_{x,n}$.

For the first part, we shall use definitions as follows omitting the index n for simplicity. Take two elements s and t of Θ. Let $h_j^2(s,t)$ be the square Hellinger distance

$$h_j^2(s,t) = \frac{1}{2}\int(\sqrt{dp_{s,j}} - \sqrt{dp_{t,j}})^2.$$

Let $H_n^2 = \sum_j h_j^2$. We shall use H as a "distance" on Θ and assume that it is in fact a distance, that is, that $H(s,t) = 0$ implies s=t and that $0 \leq H(s,t) < \infty$. This last condition is always fulfilled if J is finite.

Let P_θ be the product measure $P_\theta = \prod_j p_{\theta,j}$. There are cases where the experiment $\{P_\theta; \theta \in \Theta\}$ is infinitely divisible. In such a case instead of H^2 one can use the expression $K^2(s,t) = -\log\int\sqrt{dP_s dP_t}$. The theorems given below for H are also applicable to K.

Consider the metric space (Θ,H) and a number $\tau \geq 0$. The dimension of Θ for H at the level τ is the smallest number $D(\tau)$ such that every set of diameter $2a \geq 2\tau$ can be covered by $2^{D(\tau)}$ or fewer sets of diameter a. We shall assume here that the following condition is satisfied.

(C1) The function H is a metric for Θ. For each $\tau > 0$ the dimension $D(\tau)$ of Θ for H at the level τ is finite. Furthermore Θ is complete for H.

This makes (Θ,H) a locally compact space. The completeness condition is actually irrelevant. One can always complete the space. However together with (C2) it will insure the existence of the desired disintegration.

(C2) The measure μ gives finite mass to bounded subsets of Θ and the marginal $S' = \int P_\theta\mu(d\theta)$ is σ-finite.

For any measurable set $V \subset \Theta$, let $M_V = \int_V P_\theta\mu(d\theta)$ and $P_V = \frac{1}{\mu(V)}M_V$.

The following result is proved in Le Cam (1986). It is a consequence of a result of L. Birgé ((1981) unpublished) about test between balls of the type $\{\theta: H(\theta,\theta_i) \le b\}$, $i = 1,2$.

Take a number $z > 0$ and an integer k. Let $B_k = \{\theta: H^2(\theta,\theta_0) < z + k\}$ and let $A_k = B_{k+1} \setminus B_k$.

Theorem 3. Let V be the ball $V = \{\theta: H(\theta,\theta_0) \le \alpha\}$ and let D_k be the dimension of Θ for the level $\tau_k = \frac{1}{18}(z+k)^{1/2}$. Assume that $(18\alpha)^2 \le z + n$. Then

$$\int F_x(A_k)P_V(dx) \le \{\{(72)[\frac{z+k+1}{z+k}]^{1/2}\}^{D_k} + \frac{\mu(A_k)}{\mu(V)}\}\times\exp[-\frac{1}{2}(z+k)].$$

Actually the result stated and proved in Le Cam (1986) chapter 16, section 6, uses the measure P_{θ_0} instead of P_V. The substitution introduces another term, bounded by 2α, that would be a nuisance here.

Note that D_k decreases as k increases. Thus the term in $[\frac{z+k+1}{z+k}]^{D_k/2}$ will be of little importance for k or z large. Simple addition and the remark that D_k decreases as k increases yields the following result

Proposition 1. Assume $(18\alpha)^2 \le z$ and let $C(z) = \{\theta; H^2(\theta,\theta_0) \ge z\}$. Then

$$\int F_x[C(z)]P_V(dx) \le \frac{\sqrt{e}}{\sqrt{e}-1}[(72)(\frac{z+1}{z})^{1/2}]^{D[\tau(z)]}e^{-\frac{1}{2}z}$$
$$+ e^{-1/2z}\sum_{k=0}^{\infty}\{\frac{\mu[A_k(z)]}{\mu(V)}e^{-\frac{1}{2}k}\}$$

where $\tau(z) = \frac{1}{18}\sqrt{z}$ and where we have written $A_k(z)$ instead of A_k to remind the reader that it depends on z. Now let us pass to a situation where the experiments and the prior measure depend on an integer n. Then every one of the objects mentioned above acquires an extra subscript n. Since this make the notation more cumbersome we shall not write out the n. Instead, we shall say that the objects that do not depend on n are fixed. With this agreement consider the following conditions.

(C3) There is a fixed z and a fixed α such that $\sum_{k=0}^{\infty}\frac{\mu[A_k(z)]}{\mu(V)}e^{-\frac{1}{2}k}$ remains bounded (by a fixed bound).

(C4) For a fixed α, the sequences $\{P_{\theta_0}\}$, $\{P_\theta\}$ are contiguous whenever $H(\theta,\theta_0) \le \alpha$.

(C5) There is a fixed τ such that $D(\tau)$ remains bounded. We can then state the following.

Proposition 2. Let the conditions (C1) to (C5) be satisfied. Then for every fixed $\epsilon > 0$, there is a fixed b such that

$$\int F_x[C(b)]P_{\theta_0}(dx) < \epsilon.$$

The proof is easy. One applies Proposition 2 and use the contiguity of $\{P_{\theta_0}\}$ and $\{P_V\}$.

This result means that the posterior distributions concentrate on the balls of the type $\{\theta: H(\theta,\theta_0) \leq b\}$. It is therefore analogous to the result mentioned in Section 3 where the balls were of the type $\left\{\theta: |\theta| \leq \frac{b}{\sqrt{n}}\right\}$. If for instance the $p_{\theta,j}$ do not depend on j and if J_n has n elements, the balls $\{\theta: H(\theta,\theta_0) \leq b\}$ can also be written in the form $\{\theta: \sqrt{n}\, h(\theta,\theta_0) \leq b\}$, thus exhibiting the $\sqrt{n}$ type of convergence.

Condition (C3) calls for some comments. Suppose for instance that Θ and μ are fixed. Restore the subscript n and assume, as is often the case, that as $n \to \infty$, $H_n(\theta,\theta_0) \to \infty$ for every fixed pair (θ,θ_0). Then the neighborhood $V_n = \{\theta: H_n(\theta,\theta_0) \leq \alpha\}$ shrinks as $n \to \infty$. The condition that $\sum_{k=0}^{\infty} \frac{\mu(A_{k,n})}{\mu(V_n)} e^{-\frac{1}{2}k}$ remains bounded imposes a serious restriction on the rate of decrease of $\mu(V_n)$. In the argument of Section 3 this restriction was easily met because, for the Lebesgue measure, balls have a measure proportional to a fixed power of their diameter.

There are many situations where for points such that $H_n(\theta_n,\theta_0)$ remains bounded each individual observation yield an asymptotically negligible amount of information in the sense that $\sup_j h_n(\theta_n,\theta_0)$ tends to zero. Then, at least for such pairs, the metric $H_n(\theta_0,\theta_0)$ is an approximation to the - $\log \int \sqrt{dP_{\theta_0,n} dP_{\theta,n}}$ mentioned at the beginning of this section. At least in such cases, the result of Proposition 2 is reasonably satisfactory.

In such cases, and in many other (dependent) situations, the logarithms of likelihood ratios can often be shown to admit linear-quadratic approximations. Then a Bernstein-von Mises result can be obtained by the combination of Proposition 2 with the local result of Section 4. For this to make sense, one needs to have appropriate parametrization through a Euclidean space. A possibility is as follows, dropping the subscript n again for simplicity.

Let $\mathbf{M}_0(\Theta)$ be the space of finite signed measures m that have finite support on Θ and are such that $m(\Theta) = 0$. Let $\Lambda(s,\theta_0) = \log \frac{dP_s}{dP_{\theta_0}}$. Let us say that $\{P_\theta: \theta \in \Theta\}$ is under quadratic control near θ_0 if the following requirement is satisfied:

(C6) There exist quadratic forms Q, possibly random (depending on the observations) that satisfy the following conditions:

1) They are defined on $\mathbf{M}_0$ and positive semi definite there.

2) If $\{m\}$ is a sequence of elements $m \in \mathbf{M}_0(\Theta)$ such that the cardinality of the support of m remains bounded (independently of n) and the support remains at bounded distance of θ_0 then

$$\int [\Lambda(s,\theta_0) + \frac{1}{2}Q(\delta_s - \delta_{\theta_0})]m(ds)$$

tends to zero in P_{θ_0} probability whenever Q(m) does.

The possibility of Euclidean parametrization then corresponds to the following requirement.

(C7) There are measurable maps ξ from Θ to a fixed Euclidean space $(\mathbb{R}^k, |\cdot|)$ such that $\xi(\theta_0) = 0$ and such that 1) $|\xi(\theta)|$ remains bounded if and only if $H(\theta,\theta_0)$ does, and, 2) the quadratic forms Q of condition (C6) may be taken of the form

$$Q(m) = |M\int \xi(s)m(ds)|^2$$

for some, possibly random, matrix M. Note that (C7) does not assume that the maps ξ are injective or surjective. There may be subsets of very large dimension in Θ that are mapped at a single point. An example will be given below.

However, if one wants to have Gaussian distributions, one needs to fill out the Euclidean space in some way. One possibility is as follows.

(C8) Let $B = \{v \in \mathbb{R}^k, |v| \leq b\}$ for a fixed b. Let W(c) be the image by ξ of the ball $\{\theta: H(\theta,\theta_0) \leq c\}$. Then, for each fixed b there is a fixed c such that the Hausdorff distance between B and $B \cap W(c)$ tends to zero.

Finally, if we want to use the results of Section 4, we need to impose adequate restrictions on the measures μ. Let ν be the image of μ by ξ. Let ν_q be its translate by a vector q, and let $\nu_{q,B}$ be the restriction of ν_q to B.

(C9) For each fixed ball $B = \{v \in \mathbb{R}^k, |v| \leq b\}$ the norms $||\nu_{q,B} - \nu_{0,B}||$ tend to zero as long as $|q|$ remains bounded. Also $||\nu_{0,B}||$ remains bounded.

One can then state the following.

Theorem 4. Let the conditions (C1) to (C9) be satisfied. Let $\tilde{F}_{x,n}$ be the posterior distribution of $\xi(\theta)$. Then there are Gaussian $G_{x,n}$ such that $\int ||\tilde{F}_{x,n} - G_{x,n}|| P_{\theta_0,n}(dx)$ tends to zero.

Proof. According to Proposition 2, it is sufficient to look at what happens on balls of the type $B(b) = \{\theta: H(\theta,\theta_0) \leq b\}$. The experiments $\mathbf{E}_b = \{P_\theta; \theta \in B(b)\}$ indexed by such a ball could be reparametrized by subsets of the Euclidean space

$\mathbb{R}^k$ at a cost that is asymptotically negligible. Indeed consider two points s and t of B(b) (variable with n). Note that the contiguity imposed in (C4) implies that $Q(\delta_s - \delta_{\theta_0})$ will remain bounded in probability. Thus the same will be true of $|M\xi(s)|$ by (C6) and (C7). Now suppose in addition that $|\xi(s) - \xi(t)| \to 0$. Then $M\xi(s) - M\xi(t)$ will tend to zero in probability. It follows then from (C6) that $\Lambda(t,\theta_0) - \Lambda(s,\theta_0)$ will tend to zero in probability. Thus $||P_s - P_t||$ will tend to zero. Thus replacing a P_s by a P_t such that $\xi(s) = \xi(t)$ will not change the experiments **E**(b) except by an amount that is asymptotically negligible. One can select representatives in a measurable way, or at least in a way that is universally measurable, by anyone of a number of procedures. Thus, eliminating parts of Θ whose distance to θ_0 tends to infinity, we can proceed as if the experiment was indexed by $\mathbb{R}^k$ or a part of it.

Now refer back to Theorem 2, Section 4. If we consider only the modified (as decided above) experiments **E**(b), the conditions (a) (b) (c) (d) of the theorem are obviously satisfied. Now, it can be proved that these conditions already imply the validity of condition (B2) of Section 4. (See Le Cam (1986) Chapter 12 Section 4, Proposition 5). The condition (C9) is a modest variant of the condition (B1) of Section 4. It can serve the same purpose and the proof can proceed in the same way as that of Theorem 2.

Now a few comments about the conditions (C1) to (C9). The local ones, (C6) to (C9) are very much like the conditions of Section 4 and subject to the same comments. For instance, one cannot claim that (C9) is necessary, only that it helps. The conditions (C6) (C7) may seem peculiar. The wording of (C7) allows some inverse images $\xi^{-1}(v)$ of points $v \in \mathbb{R}^k$ to have high cardinality. This is because there are many "non parametric" situations where that could happen. A natural example is obtainable from mixtures of Poisson distributions. There one takes i.i.d observations $X_1, \cdots, X_n$, where $P_r[X_j = r] = \int e^{-\lambda}\frac{\lambda^r}{r!}P_n(d\lambda)$ for some probability measure P_n on $[0,\infty)$. Assume that there is a λ_0 such that $n\int(\lambda - \lambda_0)^2 P_n(d\lambda)$ remains bounded and such that, for some integer q, the moments $n\int|\lambda - \lambda_0|^q P_n(d\lambda)$ tend to zero. Here the parameter set Θ_n consists of a class of probability measures. However it can be shown that locally, that is as P_n concentrates around λ_0, the only relevant parameters are the first two moments $\sqrt{n}\int(\lambda - \lambda_0)P_n(d\lambda)$ and $n\int(\lambda - \lambda_0)^2 P_n(d\lambda)$. This allows construction of a map ξ from Θ_n to $\mathbb{R}^2$ with the properties described in (C7). For a detailed description, see Le Cam and Traxler (1978). There are very many examples of that type. Note, however, that if one lets P_n vary through all possible distributions, the metric dimension condition (C1) is far from being satisfied.

That condition (C1) and the associated (C5) are not as simple and natural as one might think. That balls $\{\theta\colon H_n(\theta,\theta_0) \le b\}$ for b fixed would be required to satisfy such a requirement is all right. In fact the pair (C6) (C7) requires something that is essentially much more restrictive. Some difficulties occur for "large" values of the distances $H_n(\theta,\theta_0)$. To show what is involved, take the case of i.i.d. observations $X_1, \cdots, X_n$ that are normally distributed, $N(\theta,1)$, on the line and let $\Theta = \mathbb{R}$. Then $H_n(\theta,\theta_0) \leq n$. However the condition (C1) cannot be satisfied since it would imply that $\mathbb{R}$ can be covered by a finite number of intervals of length one.

The finite dimension restriction (C1) is satisfied here if instead of H_n one uses the distance K defined by $K^2(s,t) = -\log\int\sqrt{dP_s dP_t}$ since $\sqrt{8}K(s,t) = |s-t|$. Condition (C1) is also satisfied for H_n if instead of taking for Θ the entire line $\mathbb{R}$, one restricts Θ to be a finite interval $\Theta = [a,b]$.

One could modify the conditions, requiring only that (C1) and (C5) hold for suitable balls $B_n = \{\theta\colon H_n(\theta,\theta_0) \leq b_n\}$ provided that one can show otherwise that $F_{x,n}(B_n^c)$ tends to zero in $P_{\theta_0,n}$ probability. This is actually what condition (A4), Section 3, is about. One could mimic its effect here, at least if the μ_n are finite, by assuming that for some $V_n = \{\theta\colon H_n(\theta,\theta_0) \leq b\}$, b fixed, there exist test functions ω_n such that

$$\int(1-\omega_n)dP_{V_n} \quad\text{and}\quad \frac{\mu_n(B_n^c)}{\mu_n(V_n)}\int\omega_n dP_{B_n^c}$$

both tend to zero. This can often be verified directly in applications, but it is not a very aesthetic requirement. Finally, note that we allowed the quadratic forms Q of (C6) or the matrices M of (C7) to be random. This is as it should be, since one needs to cover at least the cases where the LAMN conditions are satisfied. These were discussed by many authors, but mostly as a phenomenon that occurs naturally in the analysis of stochastic processes. Here we were interested in the case of independent observations. That the phenomenon can occur there as well is easy to see: Consider an experiment $\mathbf{E} = \{P_\theta\colon \theta \in \Theta\}$ where $\Theta = \mathbb{R}^k$ and where the log likelihoods are $\log\frac{dP_\theta}{dP_0} = t'S - \frac{1}{2}t'\Gamma t$ for some random vector S and some random positive definite matrix Γ. Assume that the P_θ are mutually absolutely continuous and that, conditionally given Γ, the distribution of S is Gaussian $N(0,\Gamma)$ if the distributions are induced by P_0. This is the standard mixed normal case.

Now suppose in addition that the distribution of Γ is infinitely divisible. Then, for any given integer n, one can represent the experiment $\mathbf{E}$ as a direct product of experiments $\mathbf{E}_n = \{p_{\theta,j,n};\ j = 1,2,\ldots n\}$. To do this one just takes i.i.d matrices $\Gamma_{j,n}$ whose distribution is the n^{th} root of that of Γ and one reconstitutes a version of S as a sum $S = \sum_j S_{j,n}$ where $S_{j,n}$ is conditionally $N(0,\Gamma_{j,n})$. We presume that even in the standard i.i.d case where the $p_{\theta,j,n}$ would be independent of j and n, one could construct examples that satisfy the LAMN conditions, but not the more restrictive LAN. However, we do not have a natural example of such a case.

References

[1] Basawa, I. V. and Prakasa Rao, B. L. S. (1980). Statistical inference for stochastic process. Academic Press

[2] Bernstein, S. (1917). Theory of Probability.

[3] Birgé, L. (1981). Private communication.

[4] Laplace, P. S. (1809). Mémoire sur les intégrales définies et leur application aux probabilités, et spécialement à la recherche du milieu qu'il faut choisir entre les résultats des observations. Mémoires presentés à l'Académie des Sciences, Paris.

[5] Le Cam, L. (1960). Locally asymptotically normal families of distributions. University of California Pub. Statistics. Vol 3, No 2, pp 37-98.

[6] Le Cam, L. (1986). Asymptotic methods in statistical decision theory. Springer Verlag.

[7] Le Cam, L. and Traxler, R. (1978). On the asymptotic behavior of mixtures of Poisson distributions. Z. F. Wahrscheinlichkeitstheorie u.v.G. Vol 44, pp 1-45.

[8] von Mises, R. (1931). Wahrscheinlichkeitsrechnung. Springer Verlag, Berlin.

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. <u>Statistics a Guide to the Unknown</u>, 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 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.

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.

208. DUBINS, L.E. (June 1989). A group decision device: Its pareto-like optimality.

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.