Some recent results in the asymptotic theory of statistical estimation

By L. Le Cam[1]

University of California, Berkeley

1. **Introduction**.

One of the simplest results in asymptotic theory of estimation is the Hájek-Le Cam asymptotic minimax theorem. Besides being simple, it has many applications. We review the theorem and give brief indications on some applications.

The theorem is called Hájek-Le Cam because it was proved by Hájek (1972) for the asymptotically normal (more precisely LAN) case. There was a previous theorem by Le Cam (1953). Hájek's result was substantially extended in Le Cam (1979).

Section 2 below gives a summary of definitions and notation. Section 3 reviews the asymptotic minimax theorem. Section 4 indicates how the theorem can be applied to problems recently studied by Donoho and Liu (1990), by M. Low (1989) and by Golubev and Nussbaum (1990). For further

[1]Research supported by NSF grant DMS-8701426

applications of the asymptotic minimax theorem, see Millar (1983).

2. **Definitions and notation**.

We shall use the definitions of Le Cam (1986) with indication of conditions under which these definitions reduce to the more usual ones.

An *experiment* $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ will be given by a σ-field $\mathcal{A}$ carried by a set $\mathcal{X}$ and a family $\{P_\theta; \theta \in \Theta\}$ of probability measures on $\mathcal{A}$. The set Θ is usually called the parameter space. The L-space $L(\mathcal{E})$ of an experiment $\mathcal{E}$ is the set of all finite signed measures defined on $\mathcal{A}$ and dominated by some convergent sum $\sum_\theta c_\theta P_\theta$, $c_\theta \geq 0$, $\sum_\theta c_\theta < \infty$. Let $\mathcal{E}$ and $\mathcal{F}$ be two experiments, with $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ on a σ-field $\mathcal{A}$ and $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ on some other σ-field $\mathcal{B}$. A *transition* T from $L(\mathcal{E})$ to $L(\mathcal{F})$ is a positive linear map from $L(\mathcal{E})$ to $L(\mathcal{F})$ such that $\|T\mu\| = \|\mu\|$ if $\mu \geq 0$. Here $\|\mu\|$ is the L_1-norm $\|\mu\| = \sup_f\{|\int f d\mu|; |f| \leq 1, f \text{ measurable}\}$. The deficiency $\delta(\mathcal{E}, \mathcal{F})$ is the number $\delta(\mathcal{E}, \mathcal{F}) = \inf_T \sup_\theta \|Q_\theta - TP_\theta\|$ where the inf is over all transitions. The distance $\Delta(\mathcal{E}, \mathcal{F})$ is $\max\{\delta(\mathcal{E}, \mathcal{F}), \delta(\mathcal{F}, \mathcal{E})\}$. Two experiments $\mathcal{E}$ and $\mathcal{F}$ are equivalent if $\Delta(\mathcal{E}, \mathcal{F}) = 0$.

The reader who would prefer working only with transitions given by Markov kernels can satisfy himself or herself that all transitions from $L(\mathcal{E})$ to $L(\mathcal{F})$ are given by Markov kernels if 1) The family $\{P_\theta\}$ is dominated and 2) the Q_θ are Borel measures on a Borel subset of a complete separable metric space.

An estimation problem consists of an experiment $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ together with a set Z and a loss function W defined on $\Theta \times Z$ to $(-\infty, +\infty]$

such that $\inf_z W_\theta(z) > -\infty$. The set Z will also be assumed to carry a vector lattice Γ of bounded numerical functions, complete for the sup norm and such that $1 \in \Gamma$.

A decision procedure ρ is then a transition ρ from $L(\mathcal{E})$ to the dual Γ' of Γ (for the sup norm). Such a transition has a value $\gamma\rho P$ for $\gamma \in \Gamma$ and $P \in L(\mathcal{E})$. (This is a contraction of $\int[\int \gamma(z)K(dz,x)]P(dx)$.) The risk of ρ at θ is $R(\theta,\rho) = W_\theta \rho P_\theta = \sup\{\gamma\rho P_\theta; \gamma \in \Gamma, \gamma \leq W_\theta\}$.

Here again the reader who prefers to work with Markov kernels (K, as above) can assume that 1) the $\{P_\theta\}$ are dominated 2) Z is compact, $\Gamma = C(Z)$ and each W_θ is lower semicontinuous.

An estimation problem given by an experiment $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ and a loss function W has a set $\mathcal{R}(\mathcal{E}, W)$ of possible risk functions, the set of functions f from Θ to $(-\infty, +\infty]$ such that there is a decision procedure ρ for which $W_\theta \rho P_\theta \leq f(\theta)$ for all $\theta \in \Theta$.

Often we shall need to work with subsets $F \subset \Theta$. Then $\mathcal{E}_F$ will be $\mathcal{E}_F = \{P_\theta; \theta \in F\}$.

3. **The asymptotic minimax theorem**.

The distance defined in Section 2 gives a topology on the set of (equivalence classes) of experiments indexed by a set Θ. Another topology is the weak topology: A directed set $\{\mathcal{E}_\nu\}$; $\mathcal{E}_\nu = \{P_{\theta,\nu} : \theta \in \Theta\}$ converges weakly to $\mathcal{F}$ if for every *finite* set $F \subset \Theta$, the distances $\Delta(\mathcal{E}_{\nu,F}, \mathcal{F}_F)$ tend to zero. This is equivalent to convergence in distribution of the vector of likelihood ratios $\{\frac{dP_{t,\nu}}{dP_{s,\nu}}, t \in F\}$ for all $s \in F$.

To state the theorem call a loss function V *special* if $V_\theta \in \Gamma$ for each $\theta \in \Theta$.

Theorem 1 *Let f be a function that does not belong to $\mathcal{R}(\mathcal{F}, W)$. Then there is a special $V \leq W$, a number $\alpha > 0$, a finite set F and an $\epsilon > 0$ such that if $\Delta(\mathcal{E}_F, \mathcal{F}_F) < \epsilon$ then $f + \alpha$ restricted to F does not belong to $\mathcal{R}(\mathcal{E}_F, V)$.*

For a proof, see Le Cam (1979) or Le Cam (1986) pages 109-110.

Remark 1. There is a weaker version of the theorem that might be easier to visualize. Let $\{\mathcal{E}_\nu\}$ be a directed family of experiments $\mathcal{E}_\nu = \{P_{\theta,\nu}; \theta \in \Theta\}$. Assume that the $\mathcal{E}_\nu$ converge weakly to $\mathcal{F}$ and that for each ν the function f_ν belongs to $\mathcal{R}(\mathcal{E}_\nu, W)$. If f_ν converges pointwise to f then $f \in \mathcal{R}(\mathcal{F}, W)$. This easy version is not sufficient for applications where one wants to truncate W. The fact that the finite set F the special V and the ϵ of Theorem 1 depends only on the triplet $(\mathcal{F}, W, f)$ is also lost in the weaker version.

Remark 2. Theorem 1 has been stated in the general framework of Section 2 with procedures that are "transitions". If one wants to restrict oneself to transitions representable by Markov kernels it is sufficient to put restrictions on the limit $\mathcal{F}$ and the loss W. Call $\mathcal{R}(\mathcal{F}, W, \text{Markov})$ the set of functions defined as in Section 2 for $\mathcal{R}(\mathcal{F}, W)$ but for transitions that are Markov kernels. It is enough to assume that $\mathcal{R}(\mathcal{F}, W) = \mathcal{R}(\mathcal{F}, W, \text{Markov})$ for the limit experiment $\mathcal{F}$. Assumptions that insure this are given in Section 2 and in Le Cam (1986) pages 11-14. No assumptions need to be placed on the experiments $\mathcal{E}$ such that $\Delta(\mathcal{E}, \mathcal{F}) < \epsilon$.

Theorem 1 uses only weak convergence to $\mathcal{F}$ of the experiments $\mathcal{E}$. There is another mode of convergence that is usually available at very little cost. It is as follows.

Take a fixed $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ and call a set $S \subset \Theta$ compact if the set $\{Q_\theta; \theta \in S\}$ is compact in $L(\mathcal{F})$ for the L_1-norm. Let $\{\mathcal{E}_\nu\}$ be a directed family of experiments, $\mathcal{E}_\nu = \{P_{\theta,\nu}; \theta \in \Theta\}$. It is said to converge to $\mathcal{F}$ on compacts if for each compact set S the restrictions $\mathcal{E}_{\nu,S}$ are such that $\Delta(\mathcal{E}_{\nu,S}, \mathcal{F}_S)$ tends to zero.

The standard LAN conditions of Le Cam (1960) imply convergence on compacts. (Hájek's 1972 do not). According to Lindae (1972) convergence on compacts follows from pointwise convergence plus some tail equicontinuity of differences $\|P_{s,\nu} - P_{t,\nu}\|$, $s, t \in S$ compact. In many cases one would wish to consider convergence on *precompact* sets instead of compacts. The precompact convergence can be reduced to the compact one by completing $\mathcal{F}$. This can be achieved without any difficulty.

Now if $\mathcal{E}_\nu$ converges on compacts to $\mathcal{F}$, Theorem 1 is certainly applicable, but can one say more? In the direction of lower bounds for the risk, perhaps very little can be said. However here are two results, that are of some interest.

Theorem 2 *Assume that, for compacts defined as above, $\mathcal{E}_\nu$ converges to $\mathcal{F}$ on compacts and that W is bounded (that is $\sup\{|W_\theta(z)|;\ \theta \in \Theta,\ z \in Z\} < \infty$.) Then if $f \in \mathcal{R}(\mathcal{F}, W)$ there is for each ν an $f_\nu \in \mathcal{R}(\mathcal{E}_\nu, W)$ such that $f_\nu \to f$ uniformly on the compact subsets of Θ.*

This is easy to see. It tends to indicate that some results that can be achieved on the limit $\mathcal{F}$ can also be achieved asymptotically on the directed set $\{\mathcal{E}_\nu\}$.

Another result extends the lower bound of Theorem 1. To state it, let $W_\theta^c = c \wedge W_\theta$ for $c \geq 0$. For risk functions $W_\theta \sigma_\nu P_{\theta,\nu}$ that might not be measurable, let $\int_* W_\theta \sigma_\nu P_{\theta,\nu}\ \mu(d\theta)$ be the lower integral, supremum of integrals of measurable functions not exceeding $W_\theta \sigma_\nu P_{\theta,\nu}$. Consider an experiment $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ and loss functions satisfying the following assumption:

(A) If Θ is pseudometrized by the distance $d(s,t) = \|Q_s - Q_t\|$ then the risk functions $W_\theta^c \rho Q_\theta$ are Borel measurable in θ for all c and all procedures ρ available in $\mathcal{F}$.

We shall state our next theorem assuming that $d(s,t) = \|Q_s - Q_t\|$ is in fact a metric on Θ. Modifications for a more general case are easy.

Theorem 3 *Suppose that condition* (A) *above is satisfied for the experiment* $\mathcal{F}$ *and that* $d(s,t)$ *defined above is a metric.*

Let μ *be a finite Radon measure on* Θ *(metrized by* d*). Assume that* $W \geq 0$*, and let* $A = \inf_\rho \int W_\theta \rho Q_\theta \mu(d\theta)$ *be the Bayes risk for* μ *and* $\mathcal{F}$*.*

Then for each $b < A$ *there is a* $c < \infty$*, a compact* $K \subset \Theta$ *and an* $\alpha > 0$ *such that if* $\Delta(\mathcal{E}_K, \mathcal{F}_K) < \alpha$ *then* $\inf_\sigma \int_* I_K(\theta) W_\theta^c \sigma P_\theta \mu(d\theta) \geq b$*, the infimum being over all procedures* σ *available for* $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$

Proof. Let ρ_0 be a Bayes procedure for $\mathcal{F}$, W and μ. Let $\mathcal{V}$ the class of special loss functions V with $V \leq W$. Then, by definition, $W_\theta \rho_0 Q_\theta = \sup_{V \in \mathcal{V}} V_\theta \rho_0 Q_\theta = \sup_c \sup_{\mathcal{V}} V_\theta^c \rho_0 Q_\theta = \sup_c W_\theta^c \rho_0 Q_\theta$. Thus if the $W_\theta^c \rho_0 Q_\theta$ are measurable $\int W_\theta \rho_0 Q_\theta \mu(d\theta) = \sup_c \int W_\theta^c \rho_0 Q_\theta \mu(d\theta)$. Since $W \geq 0$, this implies that for any number b', $b < b' < A$ there is a finite c and a compact $K \subset \Theta$ such that $\int_K W_\theta^c \rho_0 Q_\theta \mu(d\theta) > b'$. Let $\alpha > 0$ be such that $b + \|\mu\| c\alpha <$

b'. If $\Delta(\mathcal{E}_K, \mathcal{F}_K) < \alpha/2$ there is a transition T from $L(\mathcal{F}_K)$ to $L(\mathcal{E}_K)$ such that $\|P_\theta - TQ_\theta\| < \alpha$ for all $\theta \in K$. This T extends to a transition from $L(\mathcal{F})$ to $L(\mathcal{E})$. Thus, if σ is any procedure on $\mathcal{E}$, the procedure $\rho = \sigma T$ defined for $\mathcal{F}$ is such that $|W_\theta^c(\sigma T)Q_\theta - W_\theta^c \sigma P_\theta| < c\alpha$ for all $\theta \in K$. This implies

$$\int_* (W_\theta \sigma P_\theta)\mu d(\theta) \geq \int_* I_K(\theta) W_\theta \sigma P_\theta \mu(d\theta)$$

$$\geq \int_* I_K(\theta) W_\theta^c \sigma P_\theta \mu(d\theta)$$

$$\geq \int_K W_\theta^c(\sigma T) Q_\theta \mu(d\theta) - \|\mu\| c\alpha \geq b.$$

Hence the result. □

Remark 1. It should be noted that the measurability requirement (A) is imposed only on the limit experiment $\mathcal{F}$, not on the approximating experiments $\mathcal{E}$. In the cases considered in the literature the functions $\theta \rightsquigarrow W_\theta^c \rho Q_\theta$ are in fact continuous. Thus measurability is not a serious problem. However it seems to be needed for the validity of Theorem 3.

Remark 2. Let $\mathcal{M}$ be a class of Radon probability measures on Θ. The conclusion of the theorem can be replaced by: Let b denote any number strictly inferior to $\sup_\mu \inf_\rho \int W_\theta \rho Q_\theta \mu(d\theta)$. Then there is a compact $K \subset \Theta$ and numbers $\alpha > 0$ and $c < \infty$ such that if $\Delta(\mathcal{E}_K, \mathcal{F}_K) < \alpha$ one has

$$\sup_\mu \inf_\sigma \int_* W_\theta^c \sigma P_\theta \mu(d\theta) \geq b.$$

This can be seen as in Theorem 3 taking a Bayes procedure ρ_0 for a μ that almost achieves the $\sup_\mu$ for procedures on $\mathcal{F}$.

Remark 3. One might ask whether the conclusion of Theorem 3 would remain valid under only weak convergence of the experiments instead of compact convergence. This is perhaps not so. The difficulty arises from the fact that pointwise convergence of a bounded directed set of functions does not imply convergence of their integrals.

4. **Some applications**.

A) Let us start by an example of M. Low (1989) since it is very simple. Consider, on the line $\mathbb{R}$, a fixed probability density f_0 (with respect to Lebesgue measure) such that $f_0(0) > 0$, $\sup_x f(x) < \infty$ and such that f_0 be continuous at zero. Let $\{\alpha_n\}$ and $\{\beta_n\}$ be nondecreasing sequences of positive numbers such that $\alpha_n \to \infty$ and $(\alpha_n^2 \beta_n)(f_0(0)n)^{-1} \to 1$. Consider the class H of functions from $\mathbb{R}$ to $\mathbb{R}$ such that $\int h^2 < \infty$, $\int |h| < \infty$ and $\sup_x |h(x)| < \infty$. Let h_n be the number $h_n = \int \alpha_n^{-1} h(\beta_n x) f_0(x) dx$. Define $f_n(h,x) = [1 + \alpha_n^{-1} h(\beta_n x) - h_n] f_0$ if $1 + \alpha_n^{-1} h(\beta_n x) - h_n \geq 0$. Let $f_n(h,x) = f_0(x)$ otherwise. The standard Gaussian shift experiment $\mathcal{G}$ of H is one where one takes under $\theta = 0$ the distribution G_0 of a Gaussian linear process Z indexed by H and such that $E\langle Z, h\rangle = 0$ and $E|\langle Z, h\rangle|^2 = \|h\|^2 = \int h^2(x) dx$. For another value $h \in H$ one takes for G_h the measure $dG_h = \exp\{\langle Z, h\rangle - \frac{1}{2}\|h\|^2\} dG_0$.

Now let $\mathcal{E}_n = \{P_h^n; h \in H\}$ be defined by taking for P_h^n the joint distribution of n independent observations from the density $f_n(h,x)$. Low shows that $\mathcal{E}_n$ converges weakly to the Gaussian $\mathcal{G}$ as $n \to \infty$.

By restricting oneself to subsets of H one can obtain a variety of results from Theorem 1 (or 2). For instance Low considers a set of densities subject to a condition $\sup_x |f^k(x)| \leq M$ and estimates of $f(0)$. By selecting $\alpha_n = c_1 n^{k(2k+1)^{-1}}$ he shows that the appropriate rate of convergence of the estimate is in $n^{k(2k+1)^{-1}}$. This was known otherwise but Low obtains the exact limit of the risk for several loss functions.

The technique of rescaling through coefficients α_n and β_n had been previously used by Has'minskii (1979) to study estimation of a mode. For $\beta_n \equiv 1$, it has been used extensively.

B) A more complicated example appears in a paper by Golubev and Nussbaum (1990). They consider the problem of estimating a signal $t \rightsquigarrow f(t)$, $t \in [0,1]$ when the observations are of the form $Y_i = f(x_{i,n}) + \xi_i$, $i = 1, \ldots, n$ with, for instance $x_{i,n} = i/n$ and with noise ξ where the ξ_i are independent, mean zero, fixed variance σ^2 and fourth moment $E\xi_i^4$ less than a fixed constant c. The problem has been studied by many authors. A major breakthrough is due to Pinsker (1980) who considered the case where the ξ_i are Gaussian. Pinsker and subsequent authors consider the Sobolev class $W_2^m = \{f \in L_2; D^m f \in L_2\}$ where L_2 is the Hilbert space of the Lebesgue measure on $[0,1]$. For the subset $W_2^m(B) = \{f \in W_2^m; \|D^m f\|^2 \leq B\}$ let $\Delta = \lim_n \inf_{\hat{f}} \sup_f n^{(2m)(2m+1)^{-1}} E_{f,n}\|\hat{f} - f\|^2$ where the sup is on $f \in W_2^m(B)$ and the inf is over all estimators depending on n observations. The papers of Pinsker (1980) and Nussbaum (1985) give the result

$$\Delta = \gamma(m) B^r \sigma^{4mr} \quad \text{for} \quad r = (2m+1)^{-1}$$

and $\gamma(m) = (2m+1)^r [m/\pi(m+1)]^{2mr}$. The fact that the ξ_i were Gaussian was essential in the proofs. Golubev and Nussbaum use only the restrictions

$E\xi_i = 0$, $E\xi_i = \sigma^2$, $E\xi_i^4 \leq c$ and obtain a similar result.

The proof is full of ingenious devices. The relation with Theorem 1, 2 and 3 is obtainable through a series of arguments that go about as follows. Consider a particular f_0, for instance $f_0 \equiv 0$ and deviations from it. Let $W_2^{m,0}$ on $[0, 1]$ be that part of W_2^m formed by functions whose derivatives of order $0, 1, \ldots, m$ vanish at 0 and 1. For $f \in W_2^{m,0}$ one can obtain an orthogonal expansion $f = \sum_j c_j \varphi_j$ with $\|\varphi_j\| = 1$ and $\|D^m \varphi_j\|^2 = \lambda_j$ increasing in j. Now take an integer q and for $k = 1, 2, \ldots, n$ let $I_{k,q} = ((k-1)/q, k/q]$. Transport $W_2^{m,0}$ to $I_{k,q}$, by proper scaling. Look at deviations of the type $\sum_k \sum_{j=1}^s \varphi_{j,k,q}(x) f_{j,k}$ where $\varphi_{j,k,q}$ is φ_j transported to $I_{k,q}$ and put equal to zero outside $I_{k,q}$. Take only deviations that remain in $W_2^m(B)$. This allows to separate the observations by classes, the k-th class yielding a model $y_i = \sum_{j=1}^s \varphi_{j,k,q}(x_{i,n}) f_{j,k} + \xi_i$ for those $x_{i,n}$ that fall in $I_{k,q}$.

Golubev and Nussbaum let q depend on n, so it becomes $q(n)$ of the order of n^r. They then proceed to show that the part of the regression model restricted to one of the intervals $I_{k,q}$ converges to a Gaussian shift one.

Selecting the parameters $f_{j,k}$ independently according to some measure ν one can try to find a lower bound on the Bayes risk.

The bound in the limit is given by Theorem 1 or 3 for each of the subintervls $I_{k,q}$; $k = 1, 2, \ldots, q$. Since the Bayes risk for the entire problem is $q(n)$ times the risk on each $I_{k,q}$ the global lower bound can be computed for each fixed s. Then one will let s tend to infinity. Of course this is only a brief sketch of the method of proof. There are many other difficult steps on the way. One of them is to make sure that the product measure ν^{sq} on $\mathbb{R}^{sq}$ concentrates on the Sobolev ball $W_2^m(B)$. This was also crucial in Pinsker

(1980).

In Low (1989) or Golubev and Nussbaum (1990) Theorems such as Theorems 1, 2 and 3 are used to reduce a complex problem to one in which the distributions are Gaussian and where one can often get more precise information.

C) The estimation problem treated by Donoho and Liu (1990) differs considerably from the one described in (B) above. Yet the two are closely connected. Let $\mathcal{F}$ be a class of probability densities with respect to Lebesgue measure λ on an interval $[-a, +a]$ of the line. Assume that $\mathcal{F}$ is convex, closed and bounded for the L_2-norm, $\|f\|^2 = \int f^2 d\lambda$. Donoho and Liu study the problem of estimating the value $T(f)$ of a real valued *linear* function T defined on $\mathcal{F}$ when one takes n independent observations $X_1, \ldots, X_n$ from some $f \in \mathcal{F}$. For example one may want to estimate the value at zero of the k-th derivative of f subject to a local constraint on the m-th derivative, with $k \leq m$.

Let ν_n be the empirical measure of the first n observations. One can either limit oneself to estimates $\hat{T}$ that are *linear affine* in ν_n (with risk R_A indicated by a suffix A) or use any arbitrary measurable function $\hat{T}$ of ν_n (with risk R_M, indicated by a suffix M). A first remark is that, for *affine* estimates and square loss the problem of estimation of T is not more difficult than the estimation problem for a certain Gaussian shift experiment where one observes $Y = f + \sigma_n W$, $f \in \mathcal{F}$, W a white noise or a Gaussian process defined on subsets of $[-\alpha, +\alpha]$, with expectations zero and a given covariance function. This is quite analogous to (B) above, but now we need to estimate only the value of $T(f)$ instead of the whole f as in (B).

Let $\mathcal{G}_n$ be the Gaussian experiment with observations $Y = f + \sigma_n W$, $f \in \mathcal{F}$. Donoho and Liu proceed as follows

1) $\mathcal{F}$ being as described, there is a worst pair $(f_{0,n}, f_{1,n})$ of elements of $\mathcal{F}$ such that the minimax risk for *affine* estimates and for the one dimensional system $S_n = \{f_{\theta,n} = (1-\theta)f_{0,n} + \theta f_{1,n};\ \theta \in [0,1]\}$ is the same as the minimax risk for affine estimates for the entire $\mathcal{G}_n$. Furthermore the estimate for the worst pair is minimax for $\mathcal{G}_n$ among affine estimates. It is given by an explicit formula.

2) Consider the problem of estimating θ for the segment S_n described above and observations

$$\int u(t)Y(dt) \quad \text{where} \quad u = (f_{1,n} - f_{o,n})\|f_{1,n} - f_{o,n}\|^{-1}.$$

By sufficiency, this is equivalent to the problem where all of Y would be observed.

For the problem the risk R_A for affine estimates is certain function $\sigma \rightsquigarrow R_A(\sigma)$ of the standard deviation σ of $\int u(t)Y(dt)$. Similarly for the minimax risk $R_M(\sigma)$ for all measurable estimates. From Ibragimov-Has'minskii (1984) one knows that $\sup_\sigma R_A(\sigma)/R_M(\sigma)$ is bounded by a constant μ^*. From Donoho, Liu and McGibbon (1989) one knows that $\mu^* \leq 5/4$. This essentially solves the problem for the Gaussian case, at least if one considers a 25% margin acceptable.

The method "almost" solves the initial problem of estimation of T defined on $\mathcal{F}$ for the independent observations $X_1, \ldots, X_n$ at least if one selects σ_n and the white noise W properly, since for *affine* estimates the two problems are essentially asymptotically equivalent. (Asymptotically only because to get exact equivalence one has to select the Gaussian set function W with a covariance that depends on the true f_0). However that is for *affine* estimates. Would there be a possibility of doing much better for estimation of $T(f)$ by general measurable functions of the $X_1, \ldots, X_n$?

Donoho and Liu resolve the difficulty, at least for usual cases, by an appeal to a theorem similar to Theorem 2, Section 3 above.

Let $P_{\theta,n}$ be the joint distribution of $X_1, \ldots, X_n$ for the densities $f_{\theta,n} = (1-\theta)f_{0,n} + \theta f_{1,n}$, $\theta \in [0,1]$. Consider the experiments $\mathcal{E}_n = \{P_{\theta,n}; \theta \in [0,1]\}$. Consider also a Gaussian experiment

$$\mathcal{F}_n = \{Q_{\theta,n}; \theta \in [0,1]\}$$

where $Q_{\theta,n}$ is $\mathcal{N}(\theta, \sigma_n^2)$ on the line. One can prove the following

Proposition 1 *Assume that the Lévy distance between the distribution under $P_{0,n}$ of $\sum_{j=1}^n \left[\frac{f_{1,n}(X_j)}{f_{0,n}(X_j)} - 1\right]$ and a normal distribution $\mathcal{N}(0, \tau_n^2)$ tends to zero as $n \to \infty$. Assume that τ_n stays bounded. Then if $\tau_n^2 \sigma_n^2 \to 1$ the distance $\Delta(\mathcal{E}_n, \mathcal{F}_n)$ between the experiments $\mathcal{E}_n = \{P_{\theta,n}; \theta \in [0,1]\}$ and the Gaussian $\mathcal{F}_n$ tends to zero.*

This is easy to see. It follows then that the difference between the minimax risk $R_M(\mathcal{E}_n)$ for $\mathcal{E}_n$ and $R_M(\mathcal{F}_n)$ for $\mathcal{F}_n$ tends to zero.

Of course, the bulk of the argumentation of Donoho and Liu takes place on the Gaussian experiment. Donoho and Nussbaum have now extended these arguments to the estimation of certain *quadratic* functionals of the density f instead of linear ones. That the problem can be very different can be seen from an article of Bickel and Ritov (1990). The subject is still progressing.

References

[1] Bickel, P.J. and Ritov, Y. (1990). "Achieving information bounds in non and semiparametric models." *Ann. Statist.* vol.18, No.2, 925-938

[2] Donoho D.L. and Liu, R.C. (1990). "Geometrizing rates of convergence, III." Tech. Report #138, Department of Statistics, Berkeley.

[3] Donoho, D.L., Liu, R.C., and MacGibbon, B. (1988). "Minimax risk for hyperrectangles." Tech. Report #123, Department of Statistics, Berkeley.

[4] Donoho, D.L. and Nussbaum, M. (1990). "Minimax quadratic estimation of a quadratic functional." Tech. Report #236, Department of Statistics, Berkeley.

[5] Golubev, G.K. and Nussbaum, M. (1990). "A risk bound in Sobolev class regression." *Ann. Statist.*, vol.18, #2, 758-778.

[6] Hájek, J. (1972). "Local asymptotic minimax and admissibility in estimation". *Proc. 6th Berkeley Symp. Math. Stat. Proba.*, vol.1, 175-194.

[7] Has'minskii, R.Z. (1979). "Lower bound for the risk of nonparametric estimates of the mode." *Contribution to Statistics: J. Hájek memorial volume* (Jurečková, J. ed.) Akademia Praha, 90-97.

[8] Ibragimov, I.A. and Has'minskii, R.Z. (1984). "On nonparametric estimation of the value of a functional in Gaussian white noise." *Theory Probab. Applications* **24**, 18-32.

[9] Le Cam, L. (1953). "On some asymptotic properties of maximum likelihood estimates and related Bayes estimates." *Univ. California Pub. Statist.* **1**, 125-142.

[10] Le Cam, L. (1960). "Locally asymptotically normal families of distributions." *Univ. California Pub. Statist.* **3**, 37-98.

[11] Le Cam, L. (1979). "On a theorem of J. Hájek." *Contributions to Statistics: J. Hájek memorial volume* (Jurečková, J. ed.) Akademia Praha, 119-137.

[12] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

[13] Lindae, D. (1972). "Distributions of likelihood ratios and convergence of experiments." Unpublished Ph.D Thesis, Univ. California, Berkeley.

[14] Low, M.G. (1989). "Local convergence of nonparametric density estimation problems to Gaussian shift experiments on a Hilbert space." Tech. Report #225, Department of Statistics, Berkeley.

[15] Millar, P.W. (1983). "The minimax principle in asymptotic statistical theory." *Ecole d'Eté de Probabilités, Saint Flour*, 1981, Lecture Notes in Mathematics, Springer-Verlag, Berlin, Heidelberg, New York.

[16] Nussbaum, M. (1985). "Spline smoothing in regression models and asymptotic efficiency in L_2". *Ann. Statist.* **13**, 984-997.

[17] Pinsker, M.S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* **16**, 120-133.