Comments on Le Cam lecture
JSM San Francisco 2003

David Pollard
Yale University

http://www.stat.yale.edu/~pollard/

ISSN (0090-5364)

THE ANNALS
of
STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Memorial Articles

Asymptotic Equivalence and Deficiency Distance

Inverse Problems

Prediction from Extrema

Imputation

Correction

Vol. 30, No. 3—June 2002

An idea and a hope

- Statistical model: $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$

- Partition Θ into disjoint regions $\Theta_1, \ldots, \Theta_k$, with points $\theta_i \in \Theta_i$, such that it is "difficult to distinguish statistically between" models $\{\mathbb{P}_\theta : \theta \in \Theta_i\}$.

- Hope: Statistical decision problems for $\mathcal{P}$ are "comparable in difficulty" to the analogous problems for the submodel $\{\mathbb{P}_{\theta_i} : i = 1, \ldots, k\}$.

The Annals of Statistics
1973, Vol. 1, No. 1, 38-53

CONVERGENCE OF ESTIMATES UNDER DIMENSIONALITY RESTRICTIONS[1]

BY L. LECAM

University of California, Berkeley

Consider independent identically distributed observations whose distribution depends on a parameter θ. Measure the distance between two parameter points θ_1, θ_2 by the Hellinger distance $h(\theta_1, \theta_2)$.

Suppose that for n observations there is a good but not perfect test of θ_0 against θ_n. Then $n^{\frac{1}{2}}h(\theta_0, \theta_n)$ stays away from zero and infinity. The usual parametric examples, regular or irregular, also have the property that there are estimates $\hat{\theta}_n$ such that $n^{\frac{1}{2}}h(\hat{\theta}_n, \theta_0)$ stays bounded in probability, so that rates of separation for tests and estimates are essentially the same.

The present paper shows that need not be true in general but is correct under certain metric dimensionality assumptions on the parameter set. It is then shown that these assumptions imply convergence at the required rate of the Bayes estimates or maximum probability estimates.

1. Introduction. Let $\mathscr{X}$ be a set carrying a σ-field $\mathscr{A}$ and a family of probability measures $\{p_\theta; \theta \in \Theta\}$. Let $\mathscr{A}^n$ be the product of n copies of $\mathscr{A}$ and let P_θ^n be the product measure which corresponds to the distribution of n independent observations from p_θ.

It is a familiar phenomenon that, when Θ is the real line, a number of well worn regularity restrictions imply the existence of estimates $\hat{\theta}_n$ such that $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$ stays bounded in P_θ^n probability. Another familiar phenomenon occurs if p_θ is the uniform distribution of $(0, \theta)$. There, the usual estimates are such that $n(\hat{\theta}_n - \theta)$ stays bounded in P_θ^n probability.

In both examples the factors $n^{\frac{1}{2}}$ or n correspond to a certain natural rate of separation of the measures P_θ^n which can be described in terms of the Hellinger distance of the measures. If P and Q are two probability measures on the same σ-field, their Hellinger distance $H(P, Q)$ will be defined by

$$\begin{aligned} H^2(P, Q) &= \tfrac{1}{2} \int |(dP)^{\frac{1}{2}} - (dQ)^{\frac{1}{2}}|^2 \\ &= 1 - \rho(P, Q) , \end{aligned}$$

where $\rho(P, Q)$ is the affinity $\rho(P, Q) \doteq \int (dP\, dQ)^{\frac{1}{2}}$.

Letting $h(s, t) = H(p_s, p_t)$ the two factors $n^{\frac{1}{2}}$ and n correspond now to the same rate. In both cases the statement is that $n^{\frac{1}{2}}h(\hat{\theta}_n, \theta)$ stays bounded in probability.

For any two sequences $\{s_n\}$, $\{t_n\}$ inequalities of the type $0 < a \leqq n^{\frac{1}{2}}h(s_n, t_n) \leqq b < \infty$ correspond to the fact that the best test between $p_{s_n}^n$ and $p_{t_n}^n$ has probabilities of error which do not tend to zero or unity. Thus the two examples

Received September 14, 1971.

[1] This paper was prepared with the partial support of the U.S. Army Research Office (Durham) Grant DA-ARO-D-31-124-G1135.

Key words and phrases. Bayes estimates, maximum probability estimates, rate of convergence.

Kraft (1955, p133): " ... theorem, which, together with the proof, is due to L. Le Cam."

It has been shown in [2] that when A and B are dominated families of measures the number $\pi(A, B)$ is precisely equal to

$$\pi(A, B) = 1 - \inf\{D(P, Q); P \in \tilde{A}, Q \in \tilde{B}\}$$

where the sets $\tilde{A}$ and $\tilde{B}$ are the convex hulls of A and B respectively.

Consider now direct products $\{\mathscr{X}^n, \mathscr{A}^n\}$ and the corresponding product measures P^n for measures P defined on $\mathscr{A}$. Define numbers $\pi_n(A, B; \varphi)$ and $\pi_n(A, B)$ as follows. For a test function φ defined on $\{\mathscr{X}^n, \mathscr{A}^n\}$, let

$$\pi_n(A, B; \varphi) = \sup\{\textstyle\int (1-\varphi)\, dP^n + \int \varphi\, dQ^n; P \in A, Q \in B\}\,.$$

Let $\pi_n(A, B)$ be the infimum of this over all $\mathscr{A}^n$ measurable test functions φ.

To compute $\pi_n(A, B)$ would amount to the computation of the L_1 distance between the convex hulls of sets such as $A^n = \{P^n; P \in A\}$. This is usually difficult but bounds may occasionally be obtained through use of the Hellinger distances. In fact, if A is reduced to the one element P and B is reduced to the single element Q, then $\pi = \pi(P, Q)$ is the L_1-norm of the infimum $P \wedge Q$. This is related to the affinity $\rho = \rho(P, Q)$ by the inequalities

$$\pi^2 \leqq \rho^2 \leqq 1 - (1-\pi)^2 = \pi(2-\pi)\,.$$

We shall need repeatedly the following easy lemmas.

LEMMA 1. *Let P and Q be two probability measures on $\{\mathscr{X}, \mathscr{A}\}$. If $n^{\frac{1}{2}}H(P, Q) \leqq y \leqq 1$ then*

$$D(P^n, Q^n) \leqq y(2 - y^2)^{\frac{1}{2}}\,.$$

Similarly, if $nH^2(P, Q) \geqq \beta \geqq 0$ then

$$D(P^n, Q^n) \geqq 1 - e^{-\beta}\,.$$

PROOF. For the first inequality note that $H^2(P, Q) \leqq y^2/n$ is equivalent to $\rho(P, Q) \geqq 1 - y^2/n$. This gives $\rho^n(P, Q) \geqq (1 - y^2/n)^n \geqq 1 - y^2$. Hence

$$D^2(P^n, Q^n) \leqq (1 - \rho^{2n}) \leqq 1 - (1 - y^2)^2 = y^2(2 - y^2)\,.$$

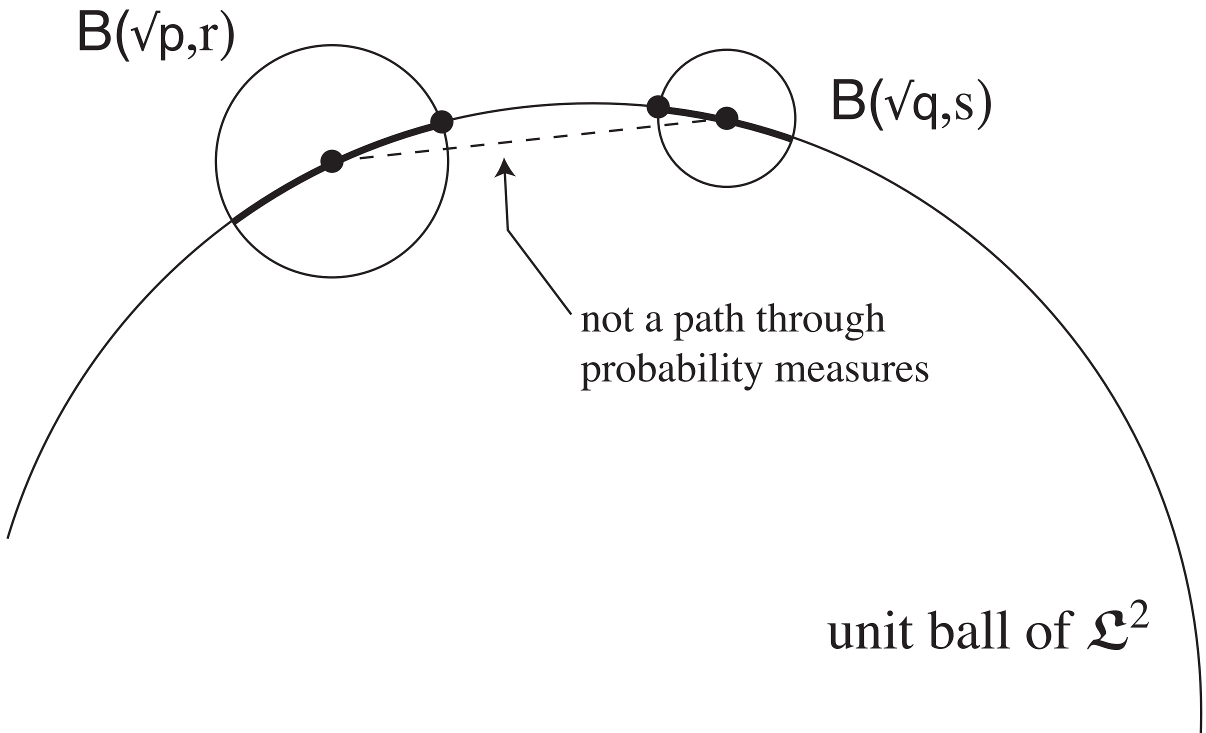
For the second inequality one can write $\rho(P, Q) \leqq (1 - \beta/n)$. Thus $\rho^n(P, Q) \leqq (1 - \beta/n)^n \leqq e^{-\beta}$. Since $D(P^n, Q^n) \geqq H^2(P^n, Q^n) = 1 - \rho^n(P, Q)$, the result follows.

A rather immediate consequence of these inequalities is that estimates cannot converge faster than the usual $n^{\frac{1}{2}}$ rate where the distance used is the Hellinger distance. Since this may be needed to place the results in perspective, we shall state it formally.

Let $\{p_\theta; \theta \in \Theta\}$ be a family of probability measures on $\{\mathscr{X}, \mathscr{A}\}$. Let $h(s, t) = H(p_s, p_t)$.

PROPOSITION 1. *Let $\{\theta_{n,i}\}$; $i = 1, 2$ be two sequences of elements of Θ. For each n, let T_n be a map from $\mathscr{X}^n$ to Θ. Assume that for both values of i the quantities $n^{\frac{1}{2}}h(T_n, \theta_{n,i})$ converge to zero in $P^n_{\theta_{n,i}}$ probability. Then the possible cluster points of the sequence $n^{\frac{1}{2}}h(\theta_{n,1}, \theta_{n,2})$ are only zero and infinity.*

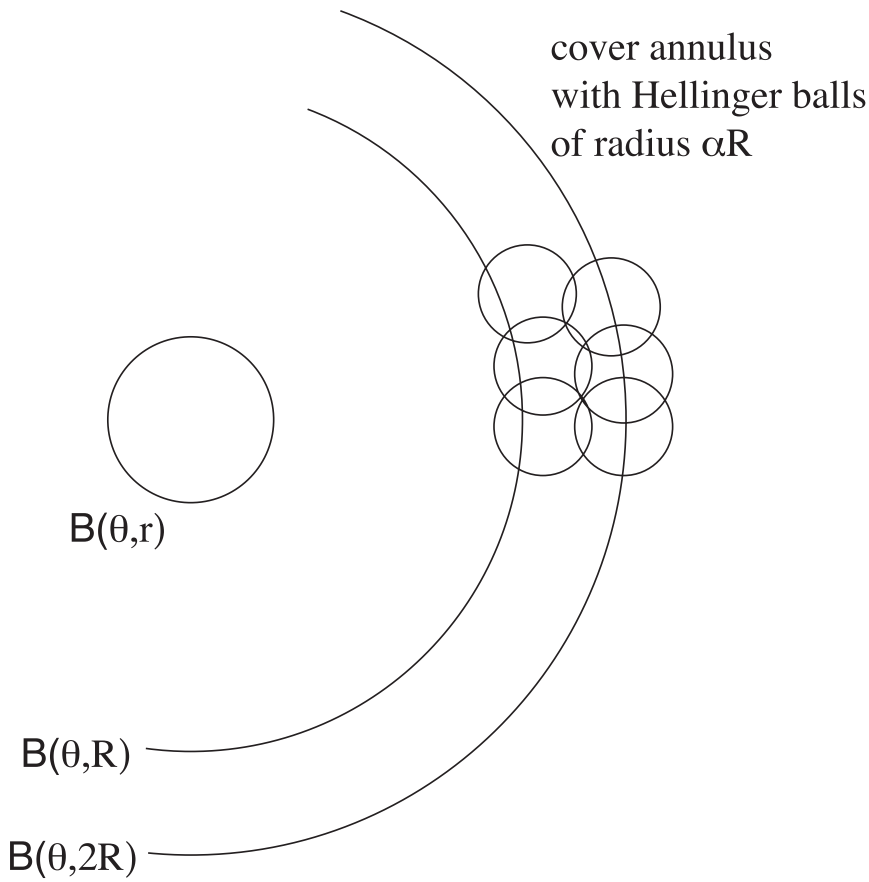
Tests between Hellinger balls



- Hellinger distance $\leftrightarrow$ total variation distance $\leftrightarrow$ existence of tests between convex sets of probability measures (minimax theorem)

- Use likelihood ratio test between centers or between closest points?
- See Le Cam & Yang (2000, p224), Le Cam(1986, §16.4) and Birgé (198*, 2003).

Le Cam (1973)



- Test $B(\theta, r)$ against each of the balls in the covering of the annulus.

- How many balls needed to cover? How good a test if $r \ll R$?

- Be pessimistic for $B(\theta, r)$.