Statistics, Probability and Game Theory
IMS Lecture Notes – Monograph Series (1996) Volume 30

COMPARISON OF EXPERIMENTS - A SHORT REVIEW

L. LE CAM
University of California, Berkeley

Abstract

In its present form, the subject of Comparison of Experiments was introduced into Statistics by D. Blackwell and C. Stein in 1951. We trace its development up to the publication of E. N. Torgersen's monumental treatise in 1991. The story leads us through the representation theorems of V. Strassen, convolution theorems of C. Boll and the use of a distance between experiments.

1. Introduction. Following Blackwell (1951) we shall call "experiment" a mathematical structure composed of the following pieces:

1) A set Θ called the parameter set, or the set of states of nature.

2) For each $\theta \in \Theta$ a probability measure P_θ on a σ-field $\mathcal{A}$ of subsets of a set X.

The idea is that, somewhere, there is a "true state of nature", that the statistician observes a random variable with values in $(X, \mathcal{A})$ and that he models the effect of the state of nature on that observation by the probability measure P_θ.

The definition covers the case of sequential experiments where the stopping rule has previously been chosen. Discussion of the choice of stopping rule would need additional mathematical objects.

Now consider the plight of a statistician who could carry out an experiment $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ on $(X, \mathcal{A})$ or a different experiment $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ on (Y, B), but not *both* $\mathcal{E}$ and $\mathcal{F}$. Assuming that the costs of observation are not taken into account, should the statistician prefer to carry out $\mathcal{E}$ or $\mathcal{F}$? The answer to such a question is complex or impossible depending on the statistician's goals. We shall deal here only with a statistician who wants to minimize risks, that is expected losses, and with definitions in which $\mathcal{E}$ is claimed to be better than $\mathcal{F}$ if that happens no matter what the loss functions are.

The comparison procedure was introduced, following a suggestion of von Neumann, in an unpublished RAND memorandum entitled "Reconnaissance in game theory" by Bohnenblust, Shapley and Sherman, (1949). Blackwell immediately noticed the links with statistics and produced another RAND memorandum (#241, 1949).

Formal definitions of the comparison criteria are given in Section 2 below. Section 3 recalls some of the main results, such as the Blackwell-Sherman-Stein theorem. Section 4 indicates that the problem was of interest in some other branches of mathematics and recalls theorems of Cartier, Fell and Meyer (1964) as well as work of V. Strassen (1965).

Section 5 is about invariance and an unpublished Thesis of C. Boll. Section 6 discusses some of the ramifications of the theory when one uses only "approximate" comparisons with introduction of "distances" between experiments. For a more complete account of the theory, the reader should consult the book by E.N. Torgersen, (1991).

2. Criteria for comparison. Suppose that, besides our two experiments $\mathcal{E}$ and $\mathcal{F}$, one is given a set Z of possible decisions and a loss function $(\theta, z) \rightsquigarrow W_\theta(z) \in [0, \infty)$.

A decision procedure for $\mathcal{E}$ is a Markov kernel, that is a map $x \rightsquigarrow \rho_x$ from X to probability measures on $(Z, \mathcal{C})$ where $\mathcal{C}$ is a σ-field such that the W_θ are measurable. The function $x \rightsquigarrow \rho_x$ is assumed to be measurable in the sense that, for $C \in \mathcal{C}$, the real valued function $x \rightsquigarrow \rho_x(C)$ is $\mathcal{A}$-measurable. (This measurability condition will be discussed further below). One can then define the risk function $\theta \rightsquigarrow R(\theta, \rho) = \int\int W_\theta(z)\rho_x(dz)P_\theta(dx)$.

Let $R(\mathcal{E}, W)$ be the set of all functions of θ that are either actual risk functions or larger than such risk functions. Proceed similarly for $\mathcal{F}$ getting $R(\mathcal{F}, W)$.

If W was given it would make good sense to say that $\mathcal{E}$ is better than $\mathcal{F}$ for W if $R(\mathcal{F}, W) \subset R(\mathcal{E}, W)$. This leads to a first criterion:

(C1) $\mathcal{E}$ is statistically better than $\mathcal{F}$ if $R(\mathcal{F}, W)$ is a subset of $R(\mathcal{E}, W)$ for every loss function W such that $0 \leq W \leq 1$.

(The restriction $0 \leq W \leq 1$ is irrelevant here. It is meant for use in Section 6.)

A second possible criterion is as follows:

(C2) $\mathcal{E}$ is better than $\mathcal{F}$ if one can find on $(X \times Y, \mathcal{A} \times \mathcal{B})$ a probability measure M_θ, $\theta \in \Theta$ such that:

(i) The marginal of M_θ on $\mathcal{A}$ is P_θ and the marginal on $\mathcal{B}$ is Q_θ.

(ii) For $\{M_\theta; \theta \in \Theta\}$ on $\mathcal{A} \times \mathcal{B}$ the σ-field generated by $\mathcal{A}$ is sufficient.

This criterion can be strengthened to read:

(C3) The experiment $\mathcal{F}$ is reproducible from $\mathcal{E}$ if there is a Markov kernel $(x, B) \rightsquigarrow K_x(B)$, $x \in X$, $B \in B$ such that $Q_\theta(B) = \int K_x(B)P_\theta(dx)$, all $\theta \in \Theta$, $B \in B$.

Finally, we shall also consider another criterion expressible in terms of linear operators. To do this, let $L(\mathcal{E})$ be the Banach lattice generated by $\mathcal{E}$. That is the space of finite signed measures carried by $\mathcal{A}$ and dominated by some convergent series $\sum_\theta a_\theta P_\theta$ where $a_\theta \geq 0$ and $a_\theta = 0$ except for a countable subset of Θ.

Define $L(\mathcal{F})$ similarly. Call a map T from $L(\mathcal{E})$ to $L(\mathcal{F})$ a *transition* if it is positive, linear and preserves the mass of positive elements.

(C4) There is a transition T from $L(\mathcal{E})$ to $L(\mathcal{F})$ such that $Q_\theta = TP_\theta$ for all $\theta \in \Theta$.

In criterion (C3) and in the definition of decision procedures, one might allow a bit more freedom to the Markov kernels. For instance in (C3) it is not necessary for the validity of the definition that $x \rightsquigarrow K_x(B)$ be actually $\mathcal{A}$-measurable. It might only be integrable for each P_θ, $\theta \in \Theta$. Similarly for the definition of the map $x \rightsquigarrow \rho_x$. The criterion (C2) could be modified accordingly replacing the conditional expectation given the σ-field generated by $\mathcal{A}$ by projections that yield bounded functions integrable for each P_θ. This makes the situation a bit complex. The complexity almost disappears if the family $\{P_\theta; \theta \in \Theta\}$ is dominated.

Generally one sees easily that (C3) implies (C2) and that (C2) implies (C1). Whether (C2) implies (C3) depends on how well behaved the triplet $(\mathcal{F}, \mathcal{Y}, \mathcal{B})$ is. Most authors consider only the case where $(\mathcal{Y}, \mathcal{B})$ is "Polish", that is where $\mathcal{Y}$ is Borel isomorphic to a Borel subset of a complete separable metric space and where $\mathcal{B}$ is the family of Borel subsets of $\mathcal{Y}$. If in addition $\{P_\theta; \theta \in \Theta\}$ is dominated then all criteria (C2), (C3) and (C4) are equivalent. There are other possibilities as described by Doob (1952), pages 28-29 and 623. The domination condition can be relaxed to some extent. For this see Torgersen (1991), pages 10-13.

All these possibilities do not change much the conceptual aspects of the problem. They are just a reflection of the complexity of a measure theoretic formulation. For instance (C4) always implies a criterion very similar to (C1), namely that $\mathcal{R}(\mathcal{F}, W) \subset \bar{\mathcal{R}}(\mathcal{E}, W)$ where the bar means closure for pointwise convergence.

The interesting problem is whether (C1) implies any of the other criteria. We shall see that (C1) always implies (C4), but we shall first look at the initial steps of the theory and at results obtained for very different mathematical purposes.

Let us note in passing that we have not used the standard terminology that if $\mathcal{E}$ and $\mathcal{F}$ satisfy (C3) then $\mathcal{E}$ is *sufficient* for $\mathcal{F}$. This is because, barring difficulties that occur in quantum theory, or any field such as psychology where the mere act of measurement modifies the "state of nature", if there is truly a "state of nature" it presumably generates a measure, say S_θ, on $\mathcal{A} \times \mathcal{B}$. For such measures $\{S_\theta; \theta \in \Theta\}$ the σ-field generated by $\mathcal{A}$ may not be sufficient even if (C2) or (C3) are satisfied.

Indeed look at a pair (X, Y) of independent random variables where X and Y are both $\mathcal{N}(\theta, 1)$ and where $\mathcal{E}$ consists of observing X only while $\mathcal{F}$ consist of observing Y only.

This distinction should be kept in mind when we discuss "deficiency" versus "insufficiency".

3. The Blackwell-Sherman-Stein theorem. Consider the situation where the parameter set Θ is a finite set. Let $S = \sum_\theta P_\theta$. Then the Radon-Nikodym densities $f_\theta = dP_\theta/dS$ are well defined almost every where S. One can assume that they are non negative and add up to unity. The image of S by the vector map $x \rightsquigarrow \{f_\theta(x); \theta \in \Theta\}$ is a measure M on the unit simplex U of $\mathbb{R}^\Theta$. It is called the Blackwell canonical measure of the experiment $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$. The image of P_θ by the same map is the measure $u_\theta \cdot M$ that has density u_θ with respect to M, the function u_θ being the θ-th coordinate of $u \in U$.

By sufficiency, one sees that nothing is lost when passing from $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ on $(X, \mathcal{A})$ to $\mathcal{E}^* = \{U_\theta \cdot M; \theta \in \Theta\}$. Indeed the pair $(\mathcal{E}, \mathcal{E}^*)$ satisfies (C3) and the pair $(\mathcal{E}, \mathcal{E}^*)$ satisfies (C2) and also (C3) upon, if needed, slight completion of X. Thus to compare experiments when Θ is finite, it is enough to compare their canonical versions. (This was already observed by Bohnenblust, Shapley and Sherman (1949)).

To state the results available in this case, we need the concept of "dilation". Suppose that C is a compact convex subset of a vector space. A dilation D on C is a Markov kernel $x \rightsquigarrow D_x$ such that the barycenter of D_x is x itself. That is $\int yD(x, dy) = x$.

A form of the Blackwell-Sherman-Stein theorem states the following:

Theorem 1. *Let $\mathcal{E}$ and $\mathcal{F}$ be two experiments indexed by the same finite parameter set Θ. Assume that they are put in their Blackwell canonical form on the unit simplex U of $\mathbb{R}^\Theta$. Let $M_\mathcal{E}$ and $M_\mathcal{F}$ be their respective canonical measures. Then the following conditions are all equivalent:*

i) The pair $(\mathcal{E}, \mathcal{F})$ satisfies (C1).

ii) The pair $(\mathcal{E}, \mathcal{F})$ satisfies (C2), (C3) and (C4).

iii) For every continuous convex function φ on U one has $\int \varphi dM_\mathcal{E} \geq \int \varphi dM_\mathcal{F}$

iv) The measure $M_\mathcal{E}$ is a dilation of $M_\mathcal{F}$

For a proof in the case where both X and Y are finite sets, see Blackwell and Girshick (1954), page 328. These authors give also a number of other properties equivalent to the statements in Theorem 1. Results similar to those of Theorem 1 were first obtained by Blackwell, (1949), (1951), in the case where Θ has only two elements. For that case some results had previously been obtained by R.F. Muirhead, (1903), as reported in Hardy, Littlewood and Pólya, (1934), Theorem 45. The connection between this and the comparison of experiments was noted by Sherman, (1951). This author and C. Stein (1951) extended Blackwell's result to the case where Θ, X and Y are all finite sets. The extension to the case where X and Y are compact subsets of Euclidean spaces, Θ being still finite, was given by Blackwell, 1953. Boll (1955) replaced the finiteness condition on Θ by a domination condition on $\{P_\theta; \theta \in \Theta\}$.

The equivalence of (iii) and (iv) of Theorem 1, for arbitrary convex compact subsets of locally convex linear spaces was of interest in the Choquet theory of representation of points of the convex set of barycenters of measures carried by their extreme points. To avoid measurability difficulties one has to assume that the set of extreme points is at least "universally measurable". This is the case for compact metrisable sets where the extreme points form a G_δ. Choquet's representation theorem gives then a much more informative result than the Krein-Milman theorem.

A paper by P. Cartier, J. Fell and P. Meyer (1964) proves the equivalence of three ordering criteria for positive measures carried by a compact convex metrisable set X. The criteria are as follows:

(a) For every continuous convex function f on X one has $\int f d\mu \leq \int f d\nu$.

(b) There exist a dilation D such that

$$\nu = \int D_x \mu(dx).$$

(c) For every decomposition $\mu = \mu_1 + \ldots + \mu_n$ of μ one can find a decomposition $\nu = \nu_1 + \ldots + \nu_n$ such that v_i has the same total mass and barycenter as μ_i.

(The last criterion was proposed by Loomis, 1962). At about the same time V. Strassen (1965) linked the problem to a problem of representation of linear functionals. Strassen considered a separable Banach space X with dual Z and what he calls "support functions" h satisfying

$$h(x+y) \leq h(x) + h(y) \text{ and } h(ax) = ah(x)$$

for $a \geq 0$.

Take a probability space $(\Omega, \mathcal{B}, \mu)$ and a measurable map $\omega \rightsquigarrow h_\omega$ to continuous support functions such that $\int \|h_\omega\| \mu(d\omega) < \infty$. Let z be a linear functional $z \in Z$ such that $z(x) \leq \int h_\omega(x)\mu(d\omega)$ for all $x \in X$. Then Strassen shows that z is an integral $z = \int z_\omega \mu(d\omega)$ where $\omega \rightsquigarrow z_\omega$ is measurable and such that $z(x) \leq h_\omega(x)$ for all ω and x.

From this it is easy to derive the Cartier-Fell-Meyer result and the Blackwell-Sherman-Stein theorem.

Many results of this nature can be found in Torgersen (1991) but with the additional twist that one does not require straight inequalities but approximate ones. See Section 6.

The arguments involving Blackwell's canonical measures on a simplex use the finiteness of the set Θ. It was noted later, see for instance Le Cam (1986), that for infinite Θ a suitable substitute for Blackwell's *canonical* measure is a Choquet *conical* measure of resultant unity. Of course, Θ as such does not appear in the Cartier-Fell-Meyer paper.

4. Comparison of experiments and invariance. We have already noted that C. Boll (1955) extended the Blackwell-Stein results to the case where Θ is infinite (but $\{P_\theta; \theta \in \Theta\}$ dominated). This was done under Polish type assumptions on the spaces $(X, \mathcal{A})$ and $(\mathcal{Y}, B)$.

The main result of Boll's unpublished thesis was relative to a case where Θ, X and Y are all the same, equal to a locally compact group. To express the result we shall use the following notation. If μ is a finite Radon measure on the locally compact group X and if $\alpha \in X$ let $S^\alpha\mu$ be "μ shifted by α on the left". This is defined by $\int f(\alpha x)\mu(dx) = \int f(x)(S^\alpha\mu)dx)$ for f continuous with compact support. If λ and μ are two such measures let $\lambda * \mu$ be the convolution defined by $\int f(z)(\lambda * \mu)(dz) = \int\int f(xy)\lambda(dx)\mu(dy)$.

Consider a pair of experiments $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ and $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ both given by Radon measures on $X = Y = \Theta$. Let us say that the pair $(\mathcal{E}, \mathcal{F})$ is shift invariant if for α and θ in Θ there is a θ' such that $S^\alpha P_\theta = Q_\theta$ and $S^\alpha Q_{\theta'} = Q_\theta$.

Boll's results imply the following

Theorem 2. *Let $(\mathcal{E}, \mathcal{F})$ be a shift invariant pair such that $\mathcal{E}$ is better than $\mathcal{F}$. Assume that the group admits almost invariant means on the left and that the P_θ are dominated by the Haar measure of X. Then there is a probability measure M such that $Q_\theta = P_\theta * M$.*

This is just one of Boll's results. To state his other results too much notation would be needed. A somewhat simplified form occurs in Le Cam (1986), page 120, Definition 1.

Boll's proof was rather complex. It was noted later (Le Cam, (1964) (1972)) that one can decompose it in two parts. First $\mathcal{E}$ better than $\mathcal{F}$ implies the existence of a transition T such that $Q_\theta = TP_\theta$. The Markov-Kakutani theorem implies then the existence of a transition that commutes with shifts: $S^\alpha T = TS^\alpha$. (See Eberlein 1949).

In a second part one shows that, since the P_θ are dominated by the Haar measures, a T that commutes with shifts is a convolution. This second part can be obtained through a slight modification of a proof of J. Wendel (1952). Wendel shows that a continuous linear map T on the set of measures dominated by Haar measure to itself satisfying $T(\lambda * \mu) = (T\lambda)\mu$ must be of the form $T\lambda = \lambda * M$ for some finite signed measure M. Wendel's theorem does not seem to have been known to Statisticians till decades later.

The subject was further investigated by Torgersen (1972), who derived numerous consequences, extended the convolution criterion to the case where the deficiency $\delta(\mathcal{E}, \mathcal{F})$ is $\leq \epsilon$, (see Section 6) and applied it to numerous computations of deficiencies.

There does not seem to be much prospect to a removal of the "amenability" restriction on the groups. See Paterson (1983). Nor is there much prospect to a removal of the condition of domination by Haar measures.

The theorem has some further consequences in asymptotic theory. It can be used in a simple proof of the Hájek convolution theorem under the LAN conditions. Far reaching results for the case where Θ is a Banach space (hence not locally compact) and the P_θ are Gaussian cylinder measures have been obtained by Moussatat (1976), Millar (1985) and van der Vaart (1991). This result has been taken as a basic definition of asymptotic optimality in Bickel-Klaassen-Ritov-Wellner (1993).

5. Distances between experiments. Let $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ and $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ be two experiments indexed by the same parameter set Θ. Typically, $\mathcal{E}$ and $\mathcal{F}$ are not comparable according to the criteria of Section 2, but one can always introduce a "distance" between $\mathcal{E}$ and $\mathcal{F}$. This was done in Le Cam (1964). One can describe it as follows. It involves the spaces $L(\mathcal{E})$ and $L(\mathcal{F})$ of criterion (C4), Section 2, metrized by their L_1-norm $\|M\| = \sup_\varphi\{|\int \varphi dM|; \varphi$ measurable; $|\varphi| \leq 1\}$. Define a "deficiency" $\delta(\mathcal{E}, \mathcal{F})$ by

$$\delta(\mathcal{E}, \mathcal{F}) = \inf_T \sup_\theta \frac{1}{2}\|Q_\theta - TP_\theta\|$$

where T runs over all transitions from $L(\mathcal{E})$ to $L(\mathcal{F})$. Define $\delta(\mathcal{F}, \mathcal{E})$ similarly and let $\Delta(\mathcal{E}, \mathcal{F}) = \max[\delta(\mathcal{E}, \mathcal{F}), \delta(\mathcal{F}, \mathcal{E})]$. This Δ is a pseudo metric on the class of experiments indexed by Θ. Two experiments such that $\Delta(\mathcal{E}, \mathcal{F}) = 0$ will be called equivalent.

Since (C1) and (C4) are nearly equivalent one can use also another definition. Take loss functions W such that $0 \leq W \leq 1$ and consider the spaces $\mathcal{R}(\mathcal{E}, W)$ and $\mathcal{R}(\mathcal{F}, W)$ of (C1). Let $\bar{\mathcal{R}}$ be the pointwise closure of $\mathcal{R}$. Then $\delta(\mathcal{E}, \mathcal{F}) \leq \epsilon$ if and only if for every such W, every $g \in \mathcal{R}(\mathcal{F}, W)$ there is an $f \in \bar{\mathcal{R}}(\mathcal{E}, W)$ such that $f(\theta) \leq g(\theta) + \epsilon$, all $\theta \in \Theta$

(Torgersen (1991) does not use a single number ϵ but a function $\theta \rightsquigarrow \epsilon_\theta \geq 0$. This gives more flexibility.).

The use of the L_1-norm $\|\cdot\|$ is suggested by statistical considerations: $1 - \frac{1}{2}\|P - Q\|$ is the optimal sum of errors for a test between P and Q.

Although Le Cam (1964) was written independently, we later (1990) became acquainted with an unpublished paper of C. Stein given to us by E. Lehmann. The paper was written when Stein was in Chicago, perhaps in 1951. Stein does not restrict the loss functions to be bounded above by unity. He requires instead that the convex hull of $\{W_\theta(z); z \in Z\}$ contain a function bounded by unity. Stein applies his definition to the case where Θ has only two points and where one considers sequences $(\mathcal{E}^n, \mathcal{F}^n)$ where for instance $\mathcal{E}^n$ is obtained by replicating $\mathcal{E}$ n-times independently. In the case where $\mathcal{E}$ and $\mathcal{F}$ are not trivial, Stein shows that $\mathcal{E}^n$ and $\mathcal{F}^n$ are asymptotically equivalent if and only if $\mathcal{E}$ and $\mathcal{F}$ have the same Kullback-Leibler numbers. On the contrary, with Le Cam's definition of distance, if $\mathcal{E}$ and $\mathcal{F}$ are not trivial then $\Delta(\mathcal{E}^n, \mathcal{F}^n) \to 0$. This discrepancy is not as serious as it may appear. One can vary $\mathcal{E}$ and $\mathcal{F}$ as n varies, getting $\mathcal{E}_n$ and $\mathcal{F}_n$. Then one looks at distances $\Delta(\mathcal{E}_n^n, \mathcal{F}_n^n)$.

Stein's paper also answers a question of D. Blackwell. It had been proved by Blackwell that if $\mathcal{E}$ is better than $\mathcal{F}$ then $\mathcal{E}^n$ is better than $\mathcal{F}^n$. Blackwell asked whether a converse could be true. Stein gives an example where $\mathcal{E}$ and $\mathcal{F}$ are not comparable but $\mathcal{E}^2$ is better than $\mathcal{F}^2$. One can wonder whether this example was ever brought to Blackwell's attention. In the mid-sixties he started Torgersen working on comparison of experiments by a similar inquiry. Torgersen produced counterexamples constructed especially for the purpose but in (1973) Hansen and Torgersen, looking at standard Gaussian linear models for a different purpose, found that they can exhibit the phenomenon when variances are unknown.

Returning to Le Cam's distance Δ, let us note that, typically, it is not easy to evaluate, although Torgersen gave precise computable formulas for it in many examples. However one can often get usable bounds for it. In the case of finite Θ, Le Cam (1969) and Torgersen (1970) show that it is bounded by and therefore uniformly equivalent to the dual Lipschitz distance of the canonical measures. More generally, for arbitrary Θ, one can often find a coupling procedure whereby $\mathcal{E}$ and $\mathcal{F}$ are realized on the same observation space. Suppose then that they are both dominated by a probability measure μ. Let $X(\theta) = dP_\theta/d\mu$ and $Y(\theta) = dQ_\theta/d\mu$. If the coupling is such that $\sup_\theta \int |X(\theta) - Y(\theta)| d\mu \leq 2\epsilon$ then $\Delta(\mathcal{E}, \mathcal{F}) \leq \epsilon$.

For this see Le Cam and Yang (1990), page 16, the paper by M. Nussbaum on "Asymptotic equivalence of density estimation and Gaussian white noise", and the forthcoming book by A.N. Shiryaev and V.G. Spokoiny.

The introduction of the distance Δ made it possible to look at the convergence of a sequence $\{\mathcal{E}_n\}$ of experiments to a limit $\mathcal{F}$, but it also makes it possible to consider pairs $(\mathcal{E}_n, \mathcal{F}_n)$ and ask whether $\Delta(\mathcal{E}_n, \mathcal{F}_n) \to 0$. Here the parameter space Θ need not be fixed. It can be Θ_n, depending on n. A most remarkable example of such a convergence occurs in the paper of Nussbaum cited above. He considers an experiment $\mathcal{E}$ given by densities f on $[0,1]$ with respect to Lebesgue measure. The densities are subject to a Hölder restriction: $|f(x) - f(y)| \leq C|x-y|^\alpha$ with $\alpha > 1/2$ and a positivity restriction: $f(x) \geq \epsilon > 0$. Let $\mathcal{E}^n$ be $\mathcal{E}$ repeated n-times independently. Let $\mathcal{F}_n$ be the experiment in which one observes a stochastic process $\{Y(t); t \in [0,1]\}$ such that $dY(t) = \sqrt{f(t)}dt + \frac{1}{2\sqrt{n}}dW(t)$ where W is standard Brownian motion.

Nussbaum shows that $\Delta(\mathcal{E}^n, \mathcal{F}_n) \to 0$ as $n \to \infty$.

This is done by first showing that the result holds for certain subsets $\Theta_n(f_0)$ of the set Θ of densities described above. Then Nussbaum shows that one can estimate the f_0 rapidly enough to fit the various pieces together. A quilt-patch technique of this type had been used by Le Cam (1986), Chapter

11, but only in cases where the densities corresponding to the $\Theta_n(f_0)$ stay contiguous and subject to restrictions of a finite dimensional character.

Nussbaum has recently extended his results to the case of estimation of regression curves.

One of the techniques used in the paper on densities is a Poissonization technique: Instead of carrying out $\mathcal{E}^n$ one observes a Poisson variable N with $EN = n$. One then carries out $\mathcal{E}^N$. Call the Poissonized experiment $\mathcal{P}_n$. Under Nussbaum's assumption $\Delta(\mathcal{E}^n, \mathcal{P}_n) \to 0$. This has now been proved by Low and Nussbaum under much weaker assumptions but there are cases where $\Delta(\mathcal{E}^n, \mathcal{E}^{n+1}) \geq 1/2$ while $\Delta(\mathcal{P}_n, \mathcal{P}_{n+r})$ is always inferior to $r/\sqrt{2n}$. See Le Cam (1986), pages 170-171.

Of course if one has a distance such as Δ one can also define some other distances. One of the most used is one in which the parameter sets are rescaled and one takes a distance such as Δ but computed only on compact subsets of the rescaled Θ_n. See Bickel-Klaassen-Ritov-Wellner, (1993). Another possibility is to take distances only on subsets of bounded cardinality. According to the results available in the finite parameter set case this is then equivalent to the ordinary convergence of the distributions of likelihood ratios. Finally, for Θ fixed, one can consider convergence on fixed finite subsets. This is called weak convergence and is already enough to yield the Hájek-Le Cam asymptotic minimax theorem, and, with appropriate group structure, the Hájek-Le Cam convolution theorem. All of this uses Le Cam's (1964) definitions of deficiencies and distances. It has been observed by Le Cam (1974) or Le Cam (1986), page 67, that under certain circumstances one can define another measure of loss of information. This is the case where one considers an experiment $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ given by measures on a σ-field $\mathcal{B}$ and where the experiment $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ is obtained by using a sub-σ-field $\mathcal{A} \subset \mathcal{B}$ and the restriction P_θ of Q_θ to $\mathcal{A}$. There $\delta(\mathcal{E}, \mathcal{F}) = 0$ means that $\mathcal{A}$ is sufficient or at least sufficient for all dominated subfamilies of $\mathcal{F}$. However, instead of measuring the loss of information incurred by the restriction to $\mathcal{A}$ by $\delta(\mathcal{E}, \mathcal{F})$ one can ask how much should one modify the Q_θ to make $\mathcal{A}$ sufficient? That is, one can measure the loss of information by an $\eta(\mathcal{E}, \mathcal{F})$ defined as $\inf_S \sup_\theta \frac{1}{2}\|Q_\theta - S_\theta\|$ where the infimum is taken over all families $\{S_\theta\}$ for which $\mathcal{A}$ is sufficient. There are various other numbers that could be defined in similar ways. It is now known (see Espen Norberg, preprint (1995)) that all the numbers introduced in Le Cam (1986), pp. 67-71 are the same. It is clear that one has always $\delta(\mathcal{E}, \mathcal{F}) \leq \eta(\mathcal{E}, \mathcal{F})$. However, it may happen that $\delta(\mathcal{E}, \mathcal{F})$ is very small compared to $\eta(\mathcal{E}, \mathcal{F})$. For finite Θ the "insufficiency" η is always bounded by a constant multiple of the deficiency. Another case where deficiencies and insufficiencies have been computed exactly is the case of Gaussian shift experiments. Following Hansen and Torgersen (1973), consider an experiment $\mathcal{E}$ where the distributions for θ are $\mathcal{N}(\theta, I)$ on $\mathbb{R}^k$. Let

$\mathcal{E}^m$ be the product of m replicas of $\mathcal{E}$ and let $\mathcal{E}^{m+r}$ be obtained by using $\mathcal{E}^m$ and r additional observations. Here $\mathcal{A}$ is the σ-field generated by the first m observations among the $n = m + r$ total.

If one lets m tend to infinity, with r fixed or much smaller than m then the deficiencies $\delta(\mathcal{E}^m, \mathcal{E}^{m+r})$ behave like $C_1 \frac{1}{\sqrt{k}}(\frac{r}{m})$ while the insufficiencies behave like $C_2 \frac{1}{\sqrt{k}}(\sqrt{\frac{r}{m}})$, C_1 and C_2 being appropriate constants.

This reflects the fact that, in our criteria (C2) or (C3) one *chooses* a joint distribution for the observations of $\mathcal{E}^m$ and $\mathcal{E}^{m+r}$ while here the joint distribution has already been prescribed.

Another possible description is that in $\delta(\mathcal{E}, \mathcal{F})$ one tries to reproduce the Q_θ from the P_θ with small "bias" while in the computation of insufficiencies one tries to reproduce the $Q_\theta = P_\theta^{m+r}$ from P_θ^m with small expected loss, the loss being always measured by L_1-norm distances. This latter effort can also be considered as an attempt to estimate the conditional distributions of the extra r variables given the first m, or in general, the conditional distributions on $\mathcal{B}$ given $\mathcal{A}$.

The use of "insufficiencies" is the basis of assertions of asymptotic sufficiency in Le Cam (1960), Le Cam (1986) or Le Cam and Yang (1990).

6. Further developments. In all the preceding sections, the families of measures such as $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ were families of *probability* measures. In particular they were positive and so were the corresponding Blackwell canonical measures.

According to Torgersen (1991) one can consider general families of measures $\{M_\theta \theta \in \Theta\}$ where the M_θ are finite *signed* measures.

The main results of the theory can be applied to such a case, with some precautions. See for instance Torgersen (1991) page 510 where the loss functions are real valued, not positive as in our (C1). See also the implication for the algebraic total mass of the measures in equation 9.21, page 511.

In the case of finite Θ, what would correspond to the Blackwell canonical measure would be the image by the vector map $\{\frac{dM_\theta}{dS}; \theta \in \Theta\}$ of the measure S defined by $dS = \sum_\theta |dP_\theta|$.

The extension has enabled Torgersen to proceed to "local comparison of experiments" by comparing families of measures such as $P_s - P_t$ and $Q_s - Q_t$ for pairs s and t in a neighborhood of a point θ_0.

Also the theory so extended is closely linked to what Marshall and Olkin (1979) call majorization of measures, to what is called Schur convexity. It is also related to results of Karlin and Rinott (1983).

To describe all of this would take much space. We can only refer the reader to Torgersen (1991). See especially Chapter 9.

REFERENCES

BICKEL, P. J., KLAASSEN, C., RITOV, Y. AND WELLNER, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press.

BLACKWELL, D. (1949). Comparison of reconnaissances. Unpublished Rand Memorandum RM. 241.

BLACKWELL, D. (1951). Comparison of experiments. *Proc. 2nd Berkeley Symp. on Math. Stat. and Probab.* 93-102.

BLACKWELL, D. (1953). Equivalent comparison of experiments. *Ann. Math. Stat.* **24** 269-272.

BLACKWELL, D. AND GIRSHICK, A. (1954). *Theory of Games and Statistical Decisions.* J. Wiley and Sons.

BOHNENBLUST, H. F., SHAPLEY, L. S. AND SHERMAN, S. (1949). Reconnaissance in Game Theory. Unpublished Rand Memorandum RM. 208.

BOLL, C. (1955). *Comparison of experiments in the infinite case and the use of invariance in establishing sufficiency.* Unpublished Ph.D. Thesis, Stanford.

CARTIER, P., FELL, J. M. G. AND MEYER, P. A. (1964). Comparaison des mesures portées par un ensemble convexe compact. *Bull. Soc. Math. France* **92** 435-445.

DOOB, J. L. (1952). *Stochastic Processes.* J. Wiley and Sons.

EBERLEIN, W. F. (1949). Abstract ergodic theorems and weak almost periodic functions. *Trans. Amer. Math. Soc.* **67** 2717-240.

HANSEN, O. H. AND TORGERSEN, E. N. (1974). Comparison of linear normal experiments. *Ann. Statist.* **2**, 367-373.

HARDY, G. H., LITTLEWOOD, J. E. AND PÓLYA, G. (1934). *Inequalities.* Cambridge Univ. Press.

KARLIN, S. AND RINOTT, Y. (1983). Comparison of measures, multivariate majorization and applications to Statistics. *Studies in econometric time series and multivariate statistics.* Academic Press.

LE CAM, L. (1964). Sufficiency and Approximate sufficiency. *Ann. Math. Stat.* **35**, 1419-1455.

LE CAM, L. (1969). *Théorie asymptotique de la décision statistique.* Presses de l'Université de Montréal.

LE CAM, L. (1974). On the information contained in additional observations. *Ann. Statist.* **2** 630-649.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer Verlag.

LE CAM, L. AND YANG, G. L. (1990). *Asymptotics in Statistics. Some basic concepts.* Springer Verlag.

MARSHALL, A. W. AND OLKIN, I. (1979). *Inequalities: Theory of majorization and its applications.* Academic Press.

MILLAR, P. W. (1985). Nonparametric applications of an infinite dimensional convolution theorem. *Z. Wahrsch. ver. Geb.* **68** 545-556.

MOUSSATAT, W. (1976). *On the asymptotic theory of statistical experiments and some of its applications.* Unpublished Thesis, Univ. of Cal. Berkeley.

MUIRHEAD, R. F. (1903). Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. *Proc. Edinburgh Math. Soc.* **21** 144-157.

NORBERG, E. (1995). On insufficiency. Preprint, Thesis Cand. Scient.

NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian whitenoise. To appear in *Ann. Statist.* Preprint (1994).

PATERSON, A. L. T. (1983). Amenability and translation experiments. *Canadian J. Math* **25** 49-58.

SHERMAN, S. (1951). On a theorem of Hardy, Littlewood, Pólya and Blackwell. *Proc. Nat. Acad. Sciences* **37** 826-831.

SHIRYAEV, A. N. AND SPOKOINY, V. G. (1996). *Statistical experiments and Decisions: Asymptotic Theory.* To be published.

STEIN, C. (1951). *Notes on a Seminar on Theoretical Statistics. I. Comparison of experiments.* Dittoed, University of Chicago.

STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Stat.* **36** 423-439.

TORGERSEN, E. N. (1970). Comparison of experiments when the parameter space is finite. *Z. Warsch. verw. Geb.* **16** 219-249.

TORGERSEN, E. N. (1972). Comparison of translation experiments. *Ann. Math. Stat.* **43** 1383-1399.

TORGERSEN, E. N. (1991). *Comparison of Statistical Experiments.* Cambridge Univ. Press.

VAN DER VAART, A. W. (1991). An asymptotic representation theorem. *Internat. Statist. Review* **58** 97-121.

WENDEL, J. G. (1952). Left centralizers and isomorphisms of group algebras. *Pacific J. Math.* **2** 251-261.

L. LE CAM
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CA 94720-3860
lecam@stat.berkeley.edu