LIMITS OF EXPERIMENTS

L. LE CAM
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

In a recent paper J. Hájek [4] proved a remarkably simple result on the limiting distributions of estimates of a vector parameter θ. It turns out that this result, as well as many of the usual statements about asymptotic behavior of tests or estimates, can be obtained by a general procedure which consists roughly in passing to the limit first and then arguing the case for the limiting problem. This passage to the limit relies on some general facts which are perhaps not entirely elementary. They depend heavily on the techniques of L. LeCam [8]. However, these general facts are of interest by themselves. If they are taken for granted the basic result of Hájek [4] and many results of A. Wald [13] become available immediately.

The present paper is organized as follows. Section 2 recalls a number of definitions and theorems which are variations on those given by the author in [8]. We have used here again a simplified definition of "experiments" barely different from the one given in [8]. There is no essential difficulty in returning to the more usual description, at least under appropriate restrictions. However the simplified (or "more abstract" as is claimed by some) description avoids measure theoretic technicalities and makes the arguments more transparent.

Section 3 gives further theorems concerning experiments with a fixed set of indices. It uses the metric introduced in [8] to define a weak topology on the space of experiments indexed by a given set Θ. Although the compactness statements proved in this section are not absolutely essential to the remainder of the paper, they do produce a number of simplifications.

The metric of [8] was intended, in part, to insure a certain continuity of risk functions, at least if loss functions stay bounded. The purpose of Section 4 is to show that a similar type of lower semicontinuity still exists for the weak topology of Section 3, even if the loss functions are only bounded from below.

In Section 5 we consider two types of limits: (1) limits of experiments in the weak topology of Section 3, and (2) experiments formed by taking limiting distributions of certain statistics. The main result is that the experiments of second type are always weaker than those of the first type. Statistics systems for which the two coincide are characterized. Another result of Section 5 is the existence of transitions which are convolutions in the case of shift invariant experiments. For this see also E. Torgersen [12] and H. Heyer [6].

Section 6 elaborates a few examples indicating some of the results implied by the previous propositions. It reproduces partially some results of Hájek, [4] and

[5]. Also it improves some results of LeCam [9] on the misbehavior of super-efficient estimates.

In the normal situation a simplified direct proof of Hájek's convolution result was communicated to us by P. J. Bickel.

Any resemblance between our results and those of Hájek is not entirely accidental, since the present paper was greatly modified after Hájek's presentation during the Symposium.

2. Definitions relative to experiments

Let Θ be a set. An experiment indexed by Θ is usually represented by a σ-field $\mathscr{A}$ carried by a set $\mathscr{X}$ and a family $\{P_\theta : \theta \in \Theta\}$ of probability measures on $\mathscr{A}$. We shall use instead a description which is obtainable from the usual one by ignoring the set $\mathscr{X}$.

Recall that an L-space is a Banach lattice whose norm satisfies the relation $\|\mu + \nu\| = \|\mu\| + \|\nu\|$ if $\mu \geqq 0$ and $\nu \geqq 0$. The dual of an L-space L is another Banach lattice M whose norm satisfies the relation $\|f \vee g\| = \|f\| \vee \|g\|$ for $f \geqq 0$ and $g \geqq 0$.

DEFINITION 1. *Let Θ be a set. An experiment $\mathscr{E}$ indexed by Θ is a function $\theta \rightsquigarrow P_\theta$ from Θ to an L-space L. This function is subject to the restriction that $P_\theta \geqq 0$ and $\|P_\theta\| = 1$.*

Let S be a subset of the L-space L. The band generated by S is the smallest linear subspace L_0 of L such that: (1) $S \subset L_0$, (2) if $\mu = \Sigma_{j=1}^n c_j |\mu_j|$ with $\mu_j \in S$ and $|\nu| \leqq \mu$ then $\nu \in L_0$, and (3) L_0 is complete for the norm.

If the experiment $\mathscr{E}: \theta \rightsquigarrow P_\theta$ is such that $S = \{P_\theta ; \theta \in \Theta\}$ has for band the whole range space L, we shall say that $\mathscr{E}$ generates L.

Since the above definition ignores the set of $\mathscr{X}$ of the more usual description it is necessary to translate to the present language the definition of "statistics." This is done by the "transitions" described below.

DEFINITION 2. *Let L_1 and L_2 be two L-spaces A transition from L_1 to L_2 is a positive linear map A from L_1 to L_2 such that $\|A\mu^+\| = \|\mu^+\|$ for all $\mu \in L_1$.*

Suppose that Z is a completely regular topological space. Let $C^b(Z)$ be the Banach space of bounded continuous numerical functions defined on Z. Let $C^*(Z)$ be the dual of $C^b(Z)$. This space $C^*(Z)$ is an L-space for the natural order. Thus if $\mathscr{E}$ is an experiment generating an L-space L, one can consider "transitions" from L to $C^*(Z)$. It will be convenient to call these transitions *statistics with values in Z*. It is clear that they correspond very exactly to the ordinary idea of *randomized* Z-valued statistics except for the circumstance that the randomization distributions need not be countably additive.

Consider a given L-space L with dual M. The formula $I\mu = \|\mu^+\| - \|\mu^-\|$ defines an element of M. The unit ball B of M is the set $B = \{u \in M; |u| \leqq I\}$. Statistical tests correspond to the positive part of this unit ball. The following result is the fundamental tool in [8].

THEOREM 1. *Let L be an L-space with dual M. Let H be a sublattice of M whose unit ball $\{u; u \in H, |u| \leq I\}$ is $\sigma(M, L)$ dense in the unit ball of M. Assume also $I \in H$. Let Z be a completely regular space. Let $\mathscr{M}$ be the set of all transitions from L to $C^*(Z)$. Let $\mathscr{M}_0$ be the subset of $\mathscr{M}$ consisting of transitions T such that*

(1) *$\gamma T \in H$ for each $\gamma \in C^b(Z)$,*

(2) *there is a finite set $F \subset Z$ such that for every $\mu \in L$ the image T_μ of μ by T is carried by F.*

In $C^b(Z) \times L$ let $\mathscr{K}$ be the class of finite unions of rectangles $K_1 \times K_2$ such that either (1) K_1 is $\sigma[C^b(Z), C^*(Z)]$ compact and K_2 is norm compact in L or, (2) K_1 is norm compact and K_2 is $\sigma(L, M)$ compact.

Then $\mathscr{M}_0$ is dense in $\mathscr{M}$ for the topology of uniform convergence on the sets $S \in \mathscr{K}$.

With a slight change in notation this is Theorem 1 in [8].

Some other definitions of [8] which will be needed below are as follows.

Let Θ be a given set. Let $\mathscr{E} : \theta \rightsquigarrow P_\theta$ be an experiment generating an L-space $L(\mathscr{E})$. Let $\mathscr{F} : \theta \rightsquigarrow Q_\theta$ be another experiment indexed by the same set Θ. It generates a space $L(\mathscr{F})$.

DEFINITION 3. *The deficiency of $\mathscr{E}$ relative to $\mathscr{F}$ is the number*

$$\delta(\mathscr{E}, \mathscr{F}) = \inf_A \sup_\theta \|A P_\theta - Q_\theta\|, \tag{2.1}$$

where the infimum is taken over all transitions from $L(\mathscr{E})$ to $L(\mathscr{F})$. The "distance" between $\mathscr{E}$ and $\mathscr{F}$ is the number

$$\Delta(\mathscr{E}, \mathscr{F}) = \max\,[\delta(\mathscr{E}, \mathscr{F}), \delta(\mathscr{F}, \mathscr{E})]. \tag{2.2}$$

This "distance" is only a pseudometric. It becomes an actual distance if two experiments whose distance is zero are considered *equivalent.*

DEFINITION 4. *For a given set Θ the equivalence class of an experiment $\mathscr{E}$ will be called the* type *of $\mathscr{E}$ and denoted $\dot{\mathscr{E}}$.*

Let $\mathscr{E}$ be an experiment $\mathscr{E} : \theta \rightsquigarrow P_\theta$ indexed by a set Θ. Let S be a subset of Θ. The experiment $\mathscr{E}$ restricted to S, that is the function $\theta \rightsquigarrow P_\theta$ defined on S only, will be denoted $\mathscr{E}_S$. If $\mathscr{E}$ and $\mathscr{F}$ are two experiments indexed by Θ the deficiency $\delta(\mathscr{E}_S, \mathscr{F}_S)$ will also be called the deficiency of $\mathscr{E}$ relative to $\mathscr{F}$ on the set S.

According to Wald a statistical decision problem is given by a triplet $\{\mathscr{E}, D, W\}$ where $\mathscr{E}$ is an experiment indexed by a set Θ and W is a function from $\Theta \times D$ to $[-\infty, +\infty]$.

In all the situations encountered below D will be a completely regular space. We shall identify the decision functions, also called decision procedures to the transitions from $L(\mathscr{E})$ to $C^*(D)$. If for each $\theta \in \Theta$ the loss function $z \rightsquigarrow W_\theta(z)$ defined on D by W is an element of $C^*(D)$, and if T is a decision procedure, the *risk* at θ of the procedure T is defined by the value $W_\theta T P_\theta$.

In order to state the result which was the main object of [8], we need a very particular class of decision spaces (D, W). A more general situation will be described in Section 4.

DEFINITION 5. *Let $\mathcal{S}$ be the class of decision spaces (D, W) formed by pairs where*

(1) *D is a compact, convex subset of the space $\mathcal{F}\{\Theta, [-1, +1]\}$ of functions from Θ to $[-1, +1]$,*

(2) *the set D has finite* linear *dimension,*

(3) *the value $W_\theta(z)$ of the loss at (θ, z), $\theta \in \Theta$, $z \in D$ is the value of the element z of $\mathcal{F}\{\Theta, [-1, +1]\}$ at the point θ.*

Note that for each θ, the function $z \rightsquigarrow W_\theta(z)$ is linear in z. Also $\|W\| = \sup\{|W_\theta(z)|, \theta \in \Theta, z \in D\}$ is always finite since $|W_\theta(z)| < 1$.

The main result of [8] can then be phrased as follows.

THEOREM 2. *Let $\mathcal{E} : \theta \rightsquigarrow P_\theta$ and $\mathcal{F} : \theta \rightsquigarrow Q_\theta$ be two experiments indexed by Θ. For any given $\varepsilon \in [0, 2]$ the following statements are all equivalent.*

(1) *There is a transition K from $L(\mathcal{E})$ to $L(\mathcal{F})$ such that* $\sup_\theta \|KP_\theta - Q_\theta\| \leqq \varepsilon$.

(2) *If T is a decision procedure relative to the experiment $\mathcal{F}$ and a decision space $(D, W) \in \mathcal{S}$ there is a procedure S of $\mathcal{E}$ to the same space (D, W) such that*

$$W_\theta S P_\theta \leqq W_\theta T Q_\theta + \varepsilon \|W\| \tag{2.3}$$

for every $\theta \in \Theta$.

(3) *If $\varepsilon' > \varepsilon$, if μ is a probability measure with finite support on Θ, and T, D, W are as in* (2), *there is an S from $\mathcal{E}$ to (D, W) such that*

$$\int (W_\theta S P_\theta)\mu(d\theta) \leqq \int (W_\theta T Q_\theta)\mu(d\theta) + \varepsilon' \|W\|. \tag{2.4}$$

That (1) $\Rightarrow$ (2) $\Rightarrow$ (3) is obvious. The implication (3) $\Rightarrow$ (1) is the subject of [8] where it is shown also that in (3) one can assume that the procedure T is "special restricted" in the sense described by the sublattice range and finite support of Theorem 1.

An immediate corollary of Theorem 2 is as follows.

COROLLARY. *The deficiency $\delta(\mathcal{E}, \mathcal{F})$ is equal to $\sup_S \delta(\mathcal{E}_S, \mathcal{F}_S)$ where the supremum extends to all* finite *subsets $S \subset \Theta$.*

Furthermore, there is a transition K which achieves the infimum in $\inf_A \sup_\theta \|AP_\theta - Q_\theta\|$.

It is perhaps appropriate to include here a few words of warning on the difference between the setup just described and the more usual one. In the latter an experiment indexed by Θ is a system $\{\mathcal{X}, \mathcal{A}, P_\theta; \theta \in \Theta\}$ consisting of a σ-field $\mathcal{A}$ carried by a set $\mathcal{X}$ and a family $\{P_\theta; \theta \in \Theta\}$ of probability measures on $\mathcal{A}$. What we have called "transitions" are replaced by Markov kernels. If $\{\mathcal{X}, \mathcal{A}\}$ and $\{\mathcal{Y}, \mathcal{B}\}$ are two measurable spaces, a Markov kernel from $\mathcal{X}$ to $\mathcal{Y}$ is a function $K(B, x)$ defined on $\mathcal{B} \times \mathcal{X}$ and such that it is a probability measure as function of B and measurable as function of x. It is well known that if $\mathcal{B}$ is the Borel field of an analytic set $\mathcal{Y}$ in a Polish space and if the P_θ are dominated by a finite measure every one of our transitions from $\mathcal{L}(\mathcal{E})$ to measures on $\mathcal{Y}$ can be represented by a Markov kernel. The general situation is quite different.

In fact J. Denny [3] has pointed out to us the following result.

Let $\mathscr{E} = \{P_\theta; \theta \in \Theta\}$ be an experiment given by probability measures P_θ on the Borel sets of the real line. Let $\mathscr{E}^n$ be the direct product of n copies of $\mathscr{E}$. That is, $\mathscr{E}^n$ corresponds to n independent identically distributed observations, each distributed according to some P_θ. Assume that P_θ is nonatomic.

Then there is an experiment $\mathscr{F} = \{Q_\theta; \theta \in \Theta\}$ given by probability measures P_θ on a certain σ-field of subsets of the real line which has the following properties:

(1) for every integer n the experiments $\mathscr{E}^n$ and $\mathscr{F}^n$ are equivalent (in our sense),

(2) for every n, the sum of the observations is a sufficient statistic for $\mathscr{F}^n$.

To obtain $\mathscr{F}$ one just restricts the P_θ to a Hamel base which has elements in common with every perfect set.

The apparent teratology does not come from our definition of experiment but from the fact that we have elected to work with a category in which the morphisms are "transitions" which may or may not be Markov kernels. If we had restricted ourselves to Markov kernels we would not even be able to cover the situations described by the usual (Halmos-Savage) definition of sufficiency. As shown below, the category used here is just the appropriate one to reflect those properties which can be expressed in terms of joint distributions of *finite* (or countable) sets of likelihood ratios. One can also put it differently by stating that the system described here relies on the feeling that Boolean algebras of events are a more primitive notion that the points of the families of sets by which one represents the algebras.

3. A weak topology in experiments indexed by a given set

In this section Θ will be a fixed set. We are interested in the class of all experiments which can be indexed by Θ. Since the range L-space is unspecified this "class" does not ordinarily qualify as a "set." We shall see however that the corresponding equivalence classes called experiment types form a set $E(\Theta)$. It is this set which will be topologized.

When Θ is *finite*, D. Blackwell [2] used certain canonical measures to characterize experiment types as follows. Let U be the unit simplex of the product space R^Θ. Specifically U is the set

$$U = \{x; x \in R^\Theta, x_\theta \geqq 0, \Sigma_\theta x_\theta = 1\}. \tag{3.1}$$

A *canonical measure* on U is a positive measure μ such that $\int x_\theta \, d\mu = 1$ for each $\theta \in \Theta$. Each canonical measure μ on U defines an experiment $\theta \rightsquigarrow p_\theta$ where p_θ is the probability measure defined on U by $dp_\theta = x_\theta \, d\mu$.

Conversely, let $\mathscr{E}; \theta \rightsquigarrow P_\theta$ be an experiment indexed by the finite set Θ. Let $m = \Sigma_\theta P_\theta$. Consider the Radon-Nikodym densities u_θ defined by $dP_\theta = u_\theta \, dm$. (Here each u_θ is an element of the dual M of $L(\mathscr{E})$ and $0 \leqq u_\theta \leqq I$ with $\Sigma_\theta u_\theta = I$.) Let u be the vector $u = \{u_\theta; \theta \in \Theta\}$. Let γ be an element of $C^b(U)$. The lattice algebraic properties of M allow the definition of $\gamma(u)$ as element of M. Thus, there is a well defined transition S from $L(\mathscr{E})$ to $C^*(U)$ whose value at $(\gamma, \lambda) \in$

$C^b \times L$ is $\gamma S\lambda = \langle \gamma(u), \lambda \rangle$. In another notation if λ is a probability measure $\lambda \in L(\mathscr{E})$ the value $S\lambda$ is the joint distribution of the likelihood ratios $\{u_\theta; \theta \in \Theta\}$ for an initial distribution λ. Consider then the image $\mu = Sm$ of $m = \Sigma P_\theta$ by S. This is obviously a canonical measure on the simplex U.

DEFINITION 6. *Provide the space R^Θ with the maximum coordinate norm, $|x| = \sup_\theta |x_\theta|$. Let μ be a signed measure on U. The Dudley norm of μ is the number*

$$\|\mu\|_D = \sup \left\{\left|\int f d\mu\right| : f \in \Lambda\right\}, \tag{3.2}$$

where Λ is the set of functions f defined on U and such that $|f| \leqq 1$ and $|f(x) - f(x')| \leqq |x - x'|$ for all x and x' in U.

One basic simple fact concerning such canonical measures is as follows.

PROPOSITION 1. *Let $\mathscr{E}$ and $\mathscr{F}$ be two experiments indexed by the finite set Θ. Let μ and v be the corresponding canonical measures on the simplex $U \subset R^\Theta$. Then $\Delta(\mathscr{E}, \mathscr{F}) \leqq \|\mu - v\|_D$. Also $\Delta(\mathscr{E}, \mathscr{F}) = 0$ if and only if $\mu = v$.*

COROLLARY. *Let $E(\Theta)$ be the set of experiment types indexed by the finite set Θ. Metrize $E = E(\Theta)$ by the experiment distance Δ. Metrize the set K of canonical measures on $U \subset R^\Theta$ by the Dudley norm. Then both E and K are compact metric spaces and the canonical one to one correspondence between them preserves the corresponding topologies and uniformities.*

PROOF. The statement that $\Delta(\mathscr{E}, \mathscr{F}) = 0$ if and only if $\mu = v$ was already proved by Blackwell in [2]. The inequality $\Delta(\mathscr{E}, \mathscr{F}) \leqq \|\mu - v\|_D$ can easily be obtained by applying the implication (3) $\Rightarrow$ (1) of Theorem 2. This is done for instance in LeCam [10]. For further results see Torgersen [11]. The corollary is immediate since U is a compact set.

Although the above proposition is not actually indispensible for the results of Sections 4 and 5, it does afford a certain convenience and allows statements of theorems in which existence of certain limits can be assumed without actual loss of generality. For the same reason it will be appropriate to state a related proposition for the case where Θ is an infinite set. For this purpose let us recall that if α is a subset of a set Θ the experiment $\mathscr{E} : \theta \rightsquigarrow P_\theta$ restricted to α is denoted $\mathscr{E}_\alpha$. If α is finite we shall topologize the set of experiment types $E(\alpha)$ by the distance Δ, equivalent to the Dudley distance of canonical measures.

DEFINITION 7. *Let Θ be an arbitrary set. Let E be a set of experiment types indexed by Θ. By the weak topology of E will be meant the weakest topology which makes the map $\mathscr{E} \rightsquigarrow \mathscr{E}_\alpha$ continuous for all the finite subsets α of Θ.*

To study the class $E(\Theta)$ of experiment types indexed by Θ it is convenient to investigate first the relations between $E(\alpha)$ and $E(\beta)$ for two *finite* subsets $\alpha \subset \beta \subset \Theta$.

It is obvious from the definitions that the restriction map from $E(\beta)$ to $E(\alpha)$ is a continuous map of $E(\beta)$ onto $E(\alpha)$.

For each finite set $\alpha \subset \Theta$ let E'_α be an experiment type $E'_\alpha \in E(\alpha)$. Such a family $\{E'_\alpha\}$ will be called *compatible* if whenever $\alpha \subset \beta$ the experiment type E'_α is the restriction to α of the experiment type E'_β. Let $E'(\Theta)$ be the set of all compatible families of experiment types. This is in an obvious way the projective limit of the sets $E(\alpha)$. Since each $E(\alpha)$ is a compact Hausdorff space, the usual projective limit theorems insure that $E'(\Theta)$ is a compact Hausdorff space whose natural map into $E(\alpha)$ is in fact *onto*. This leads to the following statement.

PROPOSITION 2. *Let Θ be an arbitrary set. Then the class $E(\Theta)$ of experiment types indexed by Θ is in one to one correspondence with the space of compatible families of experiment types $E'(\Theta)$. For its weak topology $E(\Theta)$ is a compact Hausdorff space. For the metric Δ the space $E(\Theta)$ is a complete metric space.*

PROOF. If $E(\Theta)$ and the compatible families $E'(\Theta)$ are the same set then $E(\Theta)$ is compact for the weak topology because $E'(\Theta)$ is compact. The completeness statement for the distance Δ follows because the uniform structure induced by Δ is stronger than the weak structure and because the set of pairs $(\mathscr{E}, \mathscr{F})$ such that $\Delta(\mathscr{E}, \mathscr{F}) \leqq \varepsilon$ is closed for the weak topology. Furthermore if a compatible family $\{E'_\alpha\} \in E'(\Theta)$ derives from an experiment $\mathscr{E}$ indexed by Θ, the type of $\mathscr{E}$ is well determined as can be seen from the corollary of Theorem 2. Thus the proposition will be proved if we show that any compatible family $\{E'_\alpha\}$ can be obtained from an experiment in the sense of Definition 1.

For this purpose consider two finite sets $\alpha \subset \beta \subset \Theta$. Let R^α be the Cartesian product corresponding to α. Let U_α be the unit simplex of R^α. Let $C_\alpha = C(U_\alpha)$ be the Banach space of continuous functions on U_α. Let C^*_α be the space of Radon measures on U_α. Finally let K_α be the subset of C^*_α formed by canonical measures on U_α. For $\alpha \subset \beta$ and $y \in U_\beta$ let $s_{\alpha,\beta}(y) = \Sigma_{\theta\in\alpha} y_\theta$. Finally, for all $y \in U_\beta$ such that $s_{\alpha,\beta}(y) > 0$ let $\Pi_{\alpha,\beta}(y) \in U_\alpha$ be defined by $[\Pi_{\alpha,\beta} y]_\theta = [s_{\alpha,\beta}(y)]^{-1} y_\theta$.

Let $\mathscr{E}_\beta : \theta \rightsquigarrow P_\theta$ be an experiment indexed by β. Let μ_β be the corresponding canonical measure image of $m_\beta = \Sigma_{\theta\in\beta} P_\theta$. Let $m_\alpha = \Sigma_{\theta\in\alpha} P_\theta$ and let μ_α be the corresponding canonical measure for the experiment $\mathscr{E}_\alpha = \{P_\theta; \theta \in \alpha\}$. The canonical form of $\mathscr{E}_\beta$ is given by the measures $p_{\theta,\beta}$ defined by $dp_{\theta,\beta} = y_\theta d\mu_\beta$. Similarly, the canonical form of $\mathscr{E}_\alpha$ is given by measures $dp_{\theta,\alpha} = x_\theta d\mu_\alpha$. For $\theta \in \alpha$ the above relation can be read $x_\theta = [\Pi_{\alpha,\beta} y]_\theta = [s_{\alpha,\beta}(y)]^{-1} y_\theta$. Starting from the transformation $\Pi_{\alpha,\beta}$ one can define a transformation $A_{\alpha,\beta}$ from C^*_β to C^*_α and its transpose $A'_{\alpha,\beta}$ from C_α to C_β according to the formula $[\phi A'_{\alpha,\beta}](y) = [s_{\alpha,\beta}(y)]\phi[\Pi_{\alpha,\beta}(y)]$ and $< \phi, A_{\alpha,\beta}\mu > = < \phi A'_{\alpha,\beta}, \mu >$ for $\phi \in C_\alpha$ and $\mu \in C^*_\beta$. The transformation $A'_{\alpha,\beta}$ is well defined if $[\phi A'_{\alpha,\beta}](y)$ is put equal to zero whenever $s_{\alpha,\beta}(y) = 0$. One verifies easily that it is a transformation from C_α to C_β. Furthermore its transpose $A_{\alpha,\beta}$ is a transformation of C^*_β into C^*_α which maps the canonical measures K_β *onto* K_α. Finally, by virtue of its construction, the transformation $A_{\alpha,\beta}$ is such that, with the notation used above, $A_{\alpha,\beta} p_{\theta,\beta} = p_{\theta,\alpha}$ for each $\theta \in \alpha$. This can be stated in a different way as follows. Let $\{E'_\alpha\}$, $\alpha \subset \Theta$ be a compatible family of experiment types. For each E'_α, let $\{p_{\theta,\alpha}; \theta \in \alpha\}$ be its canonical representative. If $\alpha \subset \beta$, the restriction of E'_β to E'_α has for canonical representative the linear operation $A_{\alpha,\beta}$. To associate to the compatible family $\{E'_\alpha\}$ an experiment in the sense of Definition 1 one can then proceed as follows.

For each $\alpha \subset \Theta$ take a $\phi \in C^{\alpha}$ and all its images $\phi A'_{\alpha,\gamma}$ for $\gamma \supset \alpha$. If $\phi \in C_{\alpha}$ and $\psi \in C_{\beta}$ are such that $\phi A'_{\alpha,\gamma} = \psi A'_{\alpha,\gamma}$ for $\gamma = \alpha \cup \beta$ call them equivalent. Call such an equivalence class positive if the generating ϕ is positive. Also if $\phi \in C_{\alpha}$, $\psi \in C_{\beta}$ define the sum of their classes as the class of $\phi A'_{\alpha,\gamma} + \psi A'_{\beta,\gamma}$ with $\gamma = \alpha \cup \beta$. It is readily verified that these operations on classes are well defined. The set H of classes so obtained is a vector space and in fact a vector lattice. A compatible family $\{E'_{\alpha}\}$ defines, for each $\theta \in \Theta$, a linear functional P_{θ} on H as follows: if $\phi \in C_{\alpha}$ and $\theta \in \alpha$ define $< \phi, P_{\theta} > = \int \phi dp_{\theta,\alpha}$. To obtain an experiment in the sense of Definition 1, it is sufficient to take, in the dual of H, the band generated by these functionals. This completes the proof of the proposition.

Let us mention now a corollary of Propositions 1 and 2 which is often useful in specific computations.

Let Θ be an arbitrary set. Let $\tilde{\Theta}$ be the set of probability measures with *finite* support on Θ. Each element $\tau \in \tilde{\Theta}$ can be given by a finite set $\alpha \in \Theta$ and numbers τ_{θ}. $\theta \in \alpha$ such that $\tau_{\theta} \geqq 0$ and $\Sigma_{\theta\in\alpha} \tau_{\theta} = 1$. Let $\mathscr{E} : \theta \rightsquigarrow P_{\theta}$ be an experiment indexed by Θ. For each $\tau \in \tilde{\Theta}$ define another element H_{τ} of $L(\mathscr{E})$ by $dH_{\tau} = \Pi_{\theta}(dP_{\theta})^{\tau_{\theta}}$. This is a well defined Hellinger product. The map $\tau \rightsquigarrow \|H_{\tau}\|$ from $\tilde{\Theta}$ $[0, 1]$ will be called the Hellinger transform of the experiment $\mathscr{E}$. It is easily seen that if μ is the canonical measure of $\mathscr{E}_{\alpha}$, restriction of $\mathscr{E}$ to the support α of τ, then $H_{\tau} = \int(\Pi_{\theta\in\alpha}x_{\theta}^{\tau_{\theta}})\, d\mu$.

PROPOSITION 3. *Let $E(\Theta)$ be the set of experiment types indexed by Θ. There is a one to one correspondence between experiment types and their Hellinger transforms. The weak topology of $E(\Theta)$ is the same as the topology of pointwise convergence of the Hellinger transforms.*

PROOF. That the Hellinger transforms determines the type of an experiment results from the uniqueness of the Laplace transform. The standard arguments show that the uniform structure of $E(\Theta)$ can be described as follows. Take a particular finite set $\alpha \subset \Theta$ and an $\varepsilon > 0$. Consider the set $S(\alpha, \varepsilon)$ of $\tau \in \tilde{\Theta}$ such that τ is carried by α and $\min\{\tau_{\theta}; \theta \in \alpha\} \geqq \varepsilon$. Call two Hellinger transforms H and H' close of order $(\alpha, \varepsilon, \delta)$ if $|H_{\tau} - H'_{\tau}| < \delta$ for all $\tau \in S(\alpha, \varepsilon)$. This defines a uniformity on the space of Hellinger transforms. It is precisely the uniformity induced by the topology of $E(\Theta)$. This results from the fact that the functions $x \rightsquigarrow \Pi_{\theta}x_{\theta}^{\tau_{\theta}}$ form, for $\tau \in S(\alpha, \varepsilon)$, an equicontinuous set of functions of U_{α}.

4. Lower semicontinuity of risk functions

The present section indicates some continuity relations for risk functions when the set of experiments is topologized by the weak topology. To state a decision problem we need, in addition to the experiment $\mathscr{E}$, a set D of possible decisions and a loss function W. We shall always assume below that the following condition is satisfied.

The decision space D is completely regular. For each $\theta \in \Theta$, the loss function W_{θ} is the pointwise supremum of a nonempty subset of $C^{b}(D)$. The risk at θ of a decision function ρ is defined by the equality

$$R(\theta, \rho) = \sup_{\gamma} \{\gamma\rho P_\theta;\, \gamma \in C^b(D),\, \gamma \leqq W_\theta\}. \tag{4.1}$$

This condition will be labeled the lower semicontinuity assumption, or for short, the l.s.c. assumption.

For a given experiment $\mathscr{E}$ and a given pair (D, W), let $\mathscr{D} = \mathscr{D}(\mathscr{E}, D, W)$ be the set of all available decision functions, that is, by definition, the set of all transitions from $L(\mathscr{E})$ to $C^*(D)$. This set $\mathscr{D}(\mathscr{E})$ will be given the topology of pointwise convergence on $C^b(D) \times L(\mathscr{E})$.

Let $K = K(\Theta)$ be the set of all positive finite nonnull measures which have finite support on Θ. For each $\mu \in K(\Theta)$, let

$$\chi(\mu) = \inf\left\{\int R(\theta, \rho)\mu(d\theta) : \rho \in \mathscr{D}(\mathscr{E})\right\}. \tag{4.2}$$

This will be called the envelope of the risk functions. Finally, let $\mathscr{R}(\mathscr{E})$ be the set of functions f from Θ to $(-\infty, +\infty]$ which are such that $R(\theta, \rho) \leqq f(\theta)$ for all $\theta \in \Theta$ and some $\rho \in \mathscr{D}(\mathscr{E})$.

In all these definitions it may become necessary to indicate which experiment and which pair (D, W) is involved. This will be done by writing $\chi(\mu; \mathscr{E})$ or $\chi(\mu; \mathscr{E}, W)$ or $\chi(\mu; W)$ and similar expressions for R and $\mathscr{R}$ according to which sets need to be specified.

One of the basic results from which many of the usual general statements of decision theory can be derived is as follows.

PROPOSITION 4. *Assume that the l.s.c. condition is satisfied. Then $\mathscr{D}(\mathscr{E})$ is a compact Hausdorff space and the function $\rho \rightsquigarrow \mathscr{R}(\theta, \rho)$ is lower semicontinuous on $\mathscr{D}(\mathscr{E})$. Furthermore this function is also convex in ρ. Finally a function f from Θ to $(-\infty, +\infty]$ belongs to $\mathscr{R}$ if and only if*

$$\chi(\mu) \leqq \int f(\theta)\mu(d\theta) \tag{4.3}$$

for every $\mu \in K(\Theta)$.

PROOF. The compactness and lower semicontinuity assertions are almost immediate consequences of the definitions. For the last assertion note that it is enough to consider the subset Θ_f of Θ where f is finite. Consider then a finite set $A \subset \Theta_f$ and an $\varepsilon > 0$. Let S be the set of measures $\mu \in K(\Theta)$ which are carried by A and such that $\inf\{\mu[\{\theta\}], \theta \in A\} \geqq \varepsilon$. On such a set S, integrals of risk functions are either everywhere finite or everywhere infinite. The usual theorems on separation of convex sets show then that a function f form S to $(-\infty, +\infty]$ agrees on S with an element of $\mathscr{R}$ if and only if $\int f d\mu \geqq \chi(\mu)$ for all $\mu \in S$. A passage to the limit as $\varepsilon \to 0$ and another passage to the limit letting A increase to Θ_f gives the result.

PROPOSITION 5. *Let Θ and the pair (D, W) be fixed. Let $E(\Theta)$ be the set of all experiment types indexed by Θ. Suppose that the l.s.c. assumption is satisfied. For each $\mathscr{E} \in E(\Theta)$, let $\mu \to \chi(\mu, \mathscr{E})$ be the corresponding envelope function. Let S be an arbitrary subset of $K(\Theta)$. Then $\mathscr{E} \rightsquigarrow \sup\{\chi(\mu, \mathscr{E});\, \mu \in S\}$ is a lower semicontinuous function of $\mathscr{E}$ for the weak topology of $E(\Theta)$.*

PROOF. It is obviously sufficient to prove the result for a set S reduced to a single probability measure μ.

Since μ has finite support, it is also sufficient to prove the result for Θ finite. Assuming this, note that $\chi(\mu, \mathscr{E}, W)$ is the supremum $\sup_V \chi(\mu, \mathscr{E}, V)$ where V runs through loss functions V such that $V_\theta \in C^b(D)$ and $V_\theta \leqq W_\theta$ for all θ. Indeed, there is a decision function σ such that $\chi(\mu, \mathscr{E}, W) = \int R(\theta, \sigma)\mu(d\theta)$. Suppose that $\mathscr{E} = \{P_\theta; \theta \in \Theta\}$. For this fixed σ and for any $b(\theta) < W_\theta \sigma P_\theta$ one can find a $V_\theta \leqq W_\theta$ such that $V_\theta \sigma P_\theta \geqq b(\theta)$. Let $\|V\| = \sup\{|V_\theta(t)|; \theta \in \Theta, t \in D\}$. This is finite since $V_\theta \in C^b(D)$ and since Θ is finite. However we have also

$$|\chi(\mu, \mathscr{E}, V) - \chi(\mu, \mathscr{F}, V)| \leqq \|\mu\| \, \|V\| \, \Delta(\mathscr{E}, \mathscr{F}). \tag{4.4}$$

Thus $\mathscr{E} \rightsquigarrow \chi(\mu, \mathscr{E}, V)$ is continuous and the supremum $\chi(\mu, \mathscr{E}, W)$ is lower semicontinuous.

COROLLARY 1. *Let f be a function from Θ to $(-\infty, +\infty]$. Suppose that f does not belong to $\mathscr{R}(\mathscr{E})$. Then there is an $\alpha > 0$ and a weak neighborhood G of $\mathscr{E}$ in $E(\Theta)$ such that $f + \alpha$ does not belong to any $\mathscr{R}(\mathscr{F})$, $\mathscr{F} \in G$.*

PROOF. The relation $f \notin \mathscr{R}(\mathscr{E})$ implies $\int f d\mu < \chi(\mu, \mathscr{E})$ for some $\mu \in K(\Theta)$. Thus, there is an $\alpha > 0$ such that if $g = f + \alpha$ then $\int g d\mu < \chi(\mu, \mathscr{E})$. The neighborhood $G = \{\mathscr{F}; \chi(\mu, \mathscr{F}) > \int g d\mu\}$ satisfies the required conditions.

Some particular cases of this corollary will be of interest in the following section. We shall restate two of them in a slightly different language. Recall that $\mathscr{E}_S$ means the experiment $\mathscr{E}: \theta \rightsquigarrow P_\theta$ with θ restricted to the set $S \subset \Theta$.

COROLLARY 2. *Suppose that r is an admissible element of $\mathscr{R}(\mathscr{E})$. Let θ_0 be a given element of Θ and let b be a number $b < r(\theta_0)$. There is an $\varepsilon > 0$, a finite set $S \subset \Theta$ and an $\alpha > 0$ such that if $f \in \mathscr{R}(\mathscr{F})$ for some $\mathscr{F}$ such that $\Delta(\mathscr{F}_S, \mathscr{E}_S) < \varepsilon$ satisfies the inequality $f(\theta_0) \leqq b$ then $f(\theta) > r(\theta) + \alpha$ for some $\theta \in S$.*

COROLLARY 3. *Let $a = \sup\{\chi(\mu, \mathscr{E}); \mu \in K(\Theta), \|\mu\| = 1\}$. Let b be a number $b < a$. There is a finite set S and an $\varepsilon > 0$ such that if $\Delta(\mathscr{F}_S, \mathscr{E}_S) < \varepsilon$ then $\sup\{f(\theta): \theta \in S\} > b$ for every $f \in \mathscr{R}(\mathscr{F})$.*

Note that even when W is bounded, in which case $\mathscr{E} \rightsquigarrow \chi(\mu, \mathscr{E}, W)$ is continuous in $\mathscr{E}$ for each fixed μ, we cannot conclude that the minimax risk $\sup\{\chi(\mu, \mathscr{E}); \mu \in K(\Theta), \|\mu\| = 1\}$ is a continuous function of $\mathscr{E}$. This would be true if instead of the weak topology we used the metric Δ on $E(\Theta)$.

5. Limits of experiments; distinguished statistics

Let Z be a completely regular space with its Banach space of bounded continuous numerical functions $C^b(Z)$. Let $C^*(Z)$ be the dual of $C^b(Z)$. We shall often call the elements of $C^*(Z)$ "measures" on Z or integrals even though this is an abuse of language. The space C^* can be topologized by the weak topology $\sigma[C^*(Z), C^b(Z)]$. To distinguish it from other weak topologies we shall call that one the *vague* topology of $C^*(Z)$. Recall that the positive elements of norm unity of $C^*(Z)$ (abusively called probability measures or distributions here) form a vaguely compact set.

PROPOSITION 6. *Let N be a directed set and let Θ be another set. For each $n \in N$ let $\mathscr{E}_n: \theta \rightsquigarrow F_{\theta,n}$ be an experiment such that each $F_{\theta,n}$ belongs to $C^*(Z)$. Assume that for each $\theta \in \Theta$ the $F_{\theta,n}$ converge vaguely to a limit F_θ. Let $\mathscr{F}$ be the experiment $\mathscr{F} = 0 \rightsquigarrow F_\theta$. On the space of experiment types $E(\Theta)$ let $\mathscr{E}$ be a cluster point of the directed set of types $\dot{\mathscr{E}}_n$ for the weak topology of $E(\Theta)$.*

Then $\mathscr{F}$ is weaker than $\mathscr{E}$ or more precisely $\delta(\mathscr{E}, \mathscr{F}) = 0$.

PROOF. According to the corollary of Theorem 2 it is sufficient to prove the result assuming that Θ is *finite*. This will be assumed henceforth. According to Proposition 1 one can also assume without loss of generality that $\Delta(\mathscr{E}_n, \mathscr{E}) \to 0$.

Let K be a compact convex subset of some Euclidean space and let W be a loss function defined on $\Theta \times K$. Assume that $|W| \leqq 1$ and that for each $\theta \in \Theta$ the map $t \rightsquigarrow W_\theta(t)$ is continuous. For any decision procedure σ provided by $\mathscr{F}$ let $R(\theta, \sigma; \mathscr{F})$ be the risk of σ at θ. Take an $\varepsilon > 0$ and suppose $n(\varepsilon)$ so large that $n \geqq n(\varepsilon)$ implies $\Delta(\mathscr{E}_n, \mathscr{E}) < \varepsilon$.

According to Theorem 1 for a fixed σ there is a decision procedure ρ such that $\sup_\theta |R(\theta, \sigma; \mathscr{F}) - \mathscr{R}(\theta, \rho; \mathscr{F})| < \varepsilon$ and such that ρ is *continuous* in the sense that its transpose maps $C(K)$ into $C^b(Z)$. This procedure ρ may also be applied to $\mathscr{E}_n$ giving a risk $R_n(\theta, \rho) = W_\theta \rho F_{\theta,n}$. Since $W_\theta \rho \in C^b(Z)$ there is an $n(\varepsilon, \sigma)$ such that $n \geqq n(\varepsilon\sigma)$ implies $|R_n(\theta, \rho) - R(\theta, \rho)| < \varepsilon$. This gives $R_n(\theta, \rho) \leqq R(\theta, \sigma; \mathscr{F}) + 2\varepsilon$. However since $\Delta(\mathscr{E}_n, \mathscr{E}) < \varepsilon$ for $n \geqq \max[n(\varepsilon), n(\varepsilon, \sigma)]$ there is a procedure of $\mathscr{E}$ such that

$$R(\theta, \sigma'; \mathscr{E}) \leqq R_n(\theta, \rho) + \varepsilon \leqq R(\theta, \sigma; \mathscr{F}) + 3\varepsilon. \tag{5.1}$$

This is true for every K, every W and every σ. Thus $\sigma(\mathscr{E}, \mathscr{F}) \leqq 3\varepsilon$ according to Theorem 2. This proves the desired result.

It is easy to construct examples where $\delta(\mathscr{F}, \mathscr{E}) = 2$. In other words, it may happen that $\mathscr{F}$ is trivial but $\mathscr{E}$ is perfect. However our next proposition shows that under special circumstances one can obtain the equivalence of $\mathscr{E}$ and $\mathscr{F}$.

Suppose again that N is a directed set and that Θ and Z are given. *Assume that Θ is finite* and that Z is completely regular. For each $n \in N$ let $\mathscr{E}_n: \theta \rightsquigarrow P_{\theta,n}$ be an experiment indexed by Θ. Let U be the unit simplex of R^Θ and let S_n be the canonical transition from $L(\mathscr{E}_n)$ to $C^*(U)$. This is the usual likelihood ratio vector described in Section 3. The experiment $\theta \rightsquigarrow S_n P_{\theta,n}$ is therefore the canonical representative of $\mathscr{E}_n$.

Let T_n be any statistic from $\mathscr{E}_n$ to the completely regular space Z. (See Definition 2 and the remarks following it.) Let us recall that the distribution of T_n given θ, usually written $\mathscr{L}[T_n|\theta]$ is written $T_n P_{\theta,n}$ in the present notation.

PROPOSITION 7. *With the notation just described, assume that for each $\theta \in \Theta$ the distributions $T_n P_{\theta,n}$ converge vaguely to a limit F_θ. Assume also that the experiments $\mathscr{E}_n$ converge to a limit $\mathscr{E}$. Let $\mathscr{F}$ be the experiment $\mathscr{F}$; $\theta \rightsquigarrow F_\theta$. The following conditions are equivalent:*

(a) $\Delta(\mathscr{E}, \mathscr{F}) = 0$,

(b) *for each $\varepsilon > 0$ there is a transition Γ_ε such that*

$$\limsup_n \sup_\theta \|S_n P_{\theta,n} - \Gamma_\varepsilon T_n P_{\theta,n}\|_D \leqq \varepsilon \tag{5.2}$$

for the Dudley norm of measures on U. Furthermore Γ_ε is such that its transpose maps $C^b(U)$ into $C^b(Z)$.

If these conditions are satisfied then $\Delta(\mathscr{E}_n, \mathscr{F}_n) \to 0$ for $\mathscr{F}_n = \{T_n P_{\theta,n};\, \theta \in \Theta\}$.

PROOF. Let $G_{\theta,n} = S_n P_{\theta,n}$ and let $F_{\theta,n} = T_n P_{\theta,n}$. Since $\mathscr{E}'_n = \theta \rightsquigarrow G_{\theta,n}$ is the canonical form of $\mathscr{E}_n$, the convergence of $\mathscr{E}_n$ to $\mathscr{E}$ implies that $\|G_{\theta,n} - G_\theta\|_D \to 0$ for some limit measure G_θ. Suppose now that (b) is satisfied. Then $\Gamma_\varepsilon F_{\theta,n}$ converges to $\Gamma_\varepsilon F_\theta$. Thus the inequality in (b) may be replaced by the relation

$$\limsup_n \sup_\theta \|G_\theta - \Gamma_\varepsilon F_\theta\| \leqq \varepsilon. \tag{5.3}$$

In particular $G_\theta = \lim_{\varepsilon \to 0} \Gamma_\varepsilon F_\theta$.

Let $\mathscr{F}_\varepsilon$ be the experiment $\theta \rightsquigarrow \Gamma_\varepsilon F_\theta$. This is obviously weaker than $\mathscr{F}$. As $\varepsilon \to 0$ the family $\mathscr{F}_\varepsilon$ has some cluster point say $\mathscr{F}^+$. Since each $\delta(\mathscr{F}, \mathscr{F}_\varepsilon) = 0$ one has also $\delta(\mathscr{F}, \mathscr{F}^+) = 0$. However by Proposition 6, the experiment $\theta \rightsquigarrow G_\theta$ is weaker than $\mathscr{F}^+$, hence also weaker than $\mathscr{F}$. Since $\mathscr{E} = \theta \rightsquigarrow G_\theta$, this gives $\delta(\mathscr{F}, \mathscr{E}) = 0$. Thus $\Delta(\mathscr{E}, \mathscr{F}) = 0$, according to Proposition 6.

Conversely, assume $\Delta(\mathscr{E}, \mathscr{F}) = 0$. Let Λ be the set of Lipschitz functions used to define the Dudley norm on $C^*(U)$. The equality $\Delta(\mathscr{E}, \mathscr{F}) = 0$ implies the existence of a transition Γ such that $G_\theta = \Gamma F_\theta$. Since Λ is a compact subset of $C(U)$, Theorem 1 implies the existence, for each $\varepsilon > 0$, of a transition Γ_ε satisfying the continuity requirement of (b) and such that $|f\Gamma F_\theta - f\Gamma_\varepsilon F_\theta| \leqq \varepsilon$ for all $f \in \Lambda$ and all θ. This can be written $\|G_\theta - \Gamma_\varepsilon F_\theta\|_D \leqq \varepsilon$ and implies (b). The last statement is a consequence of the triangle inequality.

Let us note that the relation $\Delta(\mathscr{E}_n, \mathscr{F}_n) \to 0$ can be interpreted to mean that the statistics T_n are asymptotically sufficient. However $\Delta(\mathscr{E}_n, \mathscr{F}_n) = 0$ would not necessarily imply $\Delta(\mathscr{E}, \mathscr{F}) = 0$. For this reason we shall introduce a definition.

DEFINITION 8. *Let $\{T_n, n \in N\}$ be a net of statistics as described before the statement of Proposition* 5. *If Θ is finite and condition* (b) *of Proposition* (5) *is satisfied the net $\{T_n;\, n \in N\}$ will be called distinguished. If Θ is finite, "distinguished" will mean that condition* (b) *is satisfied for every* finite *subset of Θ.*

Consider again a fixed arbitrary set Θ and a directed set N. For each $n \in N$ let $\mathscr{E}_n\colon \theta \rightsquigarrow P_{\theta,n}$ be an experiment indexed by Θ. Let Z and Z' be two completely regular spaces. For each n let T_n be a statistic from $\mathscr{E}_n$ to Z and let T'_n be a statistic from $\mathscr{E}_n$ to Z'.

PROPOSITION 8. *Assume that the distributions $T_n P_{\theta,n}$ converge vaguely to a limit F_θ and that the distributions $T'_n P_{\theta,n}$ converge vaguely to a limit F'_θ. If $\{T_n;\, n \in N\}$ is distinguished, there is a transition M such that $F'_\theta = M F_\theta$ for every $\theta \in \Theta$.*

PROOF. When Θ is finite, this is an immediate consequence of the definitions. If Θ is infinite it is still true, according to Theorem 2, that $\theta \rightsquigarrow F'_\theta$ is a weaker experiment than $\theta \rightsquigarrow F_\theta$. Hence the result.

The theorem given by Hájek in [4] appears similar to the above except for the fact that where we obtain the existence of a general transition, Hájek obtains a transition representable by convolution. This more precise statement can be derived from the above Proposition 8 under suitable assumptions, as we shall now show.

Let $\mathscr{F} = \{F_\theta; \theta \in \Theta\}$ and let $\mathscr{F}' = \{F'_\theta; \theta \in \Theta\}$.

Consider a transition A from $L(\mathscr{F})$ to itself and a transition A' from $L(\mathscr{F}')$ to itself.

DEFINITION 9. *The pair (A, A') leaves the system $(\mathscr{F}, \mathscr{F}')$ invariant if*:

(1) *A restricted to $\{F_\theta; \theta \in \Theta\}$ is a permutation,*

(2) *A' restricted to $\{F'_\theta; \theta \in \Theta\}$ is a permutation,*

(3) *If $AF_\theta = F_\xi$ then $A'F'_\xi = F'_\theta$.*

NOTE. If the maps $\theta \rightsquigarrow F_\theta$ and $\theta \rightsquigarrow F'_\theta$ are one to one, each one of A and A' induces a permutation of the set Θ. Condition 3 in Definition 9 says then that the permutation induced by A' is the *inverse* of that induced by A. It is more customary to write the definition in such a way that A and A' induce the *same* permutation. However the present formulation is more convenient here.

Suppose that $(\mathscr{F}, \mathscr{F}')$ is invariant by (A, A') and suppose that K is a transition from $\mathscr{L}(\mathscr{F})$ to $\mathscr{L}(\mathscr{F}')$ such that $KF_\theta = F'_\theta$. Then $A'KA$ is also such a transition.

Let $\mathscr{M}$ be the set of all transitions K from $\mathscr{L}(\mathscr{F})$ to $L(\mathscr{F}')$ such that $KF_\theta = F'_\theta$ for all $\theta \in \Theta$. For the topology of pointwise convergence on $M(\mathscr{F}') \times \mathscr{L}(\mathscr{F})$, this is a compact convex set which is transformed into itself by the continuous linear transformation $\mathscr{L} \rightsquigarrow A'KA$. An application of the Markov-Kakutani fixed point theorem gives immediately the following result.

PROPOSITION 9. *Assume that the experiments $\mathscr{F} = \{F_\theta; \theta \in \Theta\}$ and $\mathscr{F}' = \{F'_\theta; \theta \in \Theta\}$ are such that*:

(1) *$\mathscr{F}'$ is weaker than $\mathscr{F}$,*

(2) *there is a family (A_g, A'_g), $g \in G$, of transition pairs leaving the system $(\mathscr{F}, \mathscr{F}')$ invariant such that the induced family of transformations on $\mathscr{M}$ is either abelian or a solvable group or more generally a semigroup which admits almost invariant means.*

Then there is a transition K from $L(\mathscr{F})$ to $L(\mathscr{F}')$ such that $F'_\theta = KF_\theta$ for all θ and $A'_gKA_g = K$ for all $g \in G$.

In asymptotic theory one often encounters the following situation which is an important special case of the one just described.

Suppose that $\mathscr{F}$ and $\mathscr{F}'$ are as in Proposition 8 but that the two spaces Z and Z' are one and the same. Suppose also that $Z = Z'$ is a locally compact group and that the measures F_θ and F'_θ are Radon measures on Z. For any finite Radon measures μ on Z and any element $\alpha \in Z$ define the measure $\alpha\mu$, called μ shifted by α, by the equality $\int f(z)[\alpha\mu](dz) = \int f(\alpha z)\mu(dz)$.

One can say that the pair $(\mathscr{F}, \mathscr{F}')$ is invariant by the (left) group shifts if for each $\alpha \in Z$ the operations $F_\theta \rightsquigarrow \alpha F_\theta$ and $F'_\theta \rightsquigarrow \alpha F'_\theta$ are permutations and if $\alpha F_\theta = F_\xi$ implies $\alpha F'_\theta = F'_\xi$. A transition K which is "invariant" can then be described as a transition which commutes with the group shifts. That is for every $\alpha \in Z$ and $\mu \in L(\mathscr{F})$ one has $\alpha K\mu = K\alpha\mu$. It is to be expected that such transitions will in fact turn out to be convolutions by a fixed probability measure. However we have been able to prove this only under special assumptions.

PROPOSITION 10. *Let Z be a locally compact group. Let $\mathscr{F} = \{F_\theta; \theta \in \Theta\}$ be an experiment defined by Radon measures on Z. Let K be a transition from $L(\mathscr{F})$*

to finite Radon measures on Z. Assume that K commutes with the group shifts. Then there is a probability measure Q such that $K\mu$ is the convolution $K\mu = \mu * Q$ if and only if the transpose of K transforms $C^b(Z)$ into $C^b(Z)$. This happens in particular if the F_θ are either all discrete or all absolutely continuous with respect to one of the Haar measures of Z.

PROOF. Consider the case where all the F_θ are absolutely continuous with respect to a Haar measure. Let f be a bounded measurable function defined on Z. Then $\int |f| d\mu = 0$ for all $\mu \in L(\mathscr{F})$ if and only if f is locally equivalent to zero for the Haar measure on Z. Indeed, suppose that $f \geqq 0$ has compact support. By Fubini's theorem

$$\int \left\{ \int f(\alpha z)\mu(dz) \quad \lambda(d\alpha) = \int \mu(dz) \int f(\alpha z)\lambda(d\alpha) \right. \tag{5.4}$$
$$= \|\mu\| \int f(\alpha)\lambda(d\alpha)$$

for the right Haar measure λ on Z.

Let H_0 be the space of equivalence classes bounded measurable functions on Z, for the local equivalence relation defined by the Haar measure. According to [7] the space H_0 admits a lifting which commutes with the shifts. Let H be the range of the lifting. That is H consists of the functions which have been selected as representatives of classes in H_0. Note that if $g \in C^b(Z)$ the representative of the class of g is g itself.

Since K is a transition from $L(\mathscr{F})$ to finite Radon measures on Z, its transpose K^t maps the dual of the space of finite Radon measures into H_0. However, composing K^t with the lifting, we can instead define K^t as a map into H. With this agreement consider $K^t g$ for some $g \in C^b(Z)$. The equality $\alpha K\mu = K\alpha\mu$, $\mu \in L(\mathscr{F})$, $\alpha \in Z$, implies $< \alpha^t K^t g, \mu > \; = \; < K^t \alpha^t g, \mu >$ for $\mu \in L(\mathscr{F})$ and $\alpha \in Z$. However $\alpha^t K^t g$ is also in the lifting and therefore the *almost* everywhere equality just indicated implies that $\alpha^t K^t g = K^t \alpha^t g$ *everywhere*. The same conclusion would be available if we had assumed that the F_θ are discrete or that $K^t g$ can be taken equal to an element of $C^b(Z)$.

Let then $\mathscr{K}$ be the space of continuous functions with compact support on Z. Fix a $z \in Z$ and evaluate $K^t g$ at z for each $g \in \mathscr{K}$. This gives a positive linear functional Q_z on $\mathscr{K}$. The equality $\alpha^t K^t g = K^t \alpha^t g$ yields $Q_{\alpha z} = \alpha Q_z$, that is $Q_z = zQ$ with $Q = Q_0$. Thus

$$(K^t g)(z) = \int g(t) Q_z(dt) = \int g(zt) Q(dt). \tag{5.5}$$

The result follows.

6. Applications to standard examples

In many of the customary applications of large sample theory one is interested in the asymptotic behavior of certain test or estimate in the vicinity of a given value of the parameter. This can be formalized as follows. One is given a directed set N, for instance the set of integers. For each $n \in N$, let $\hat{\mathscr{E}}_n = \{Q_{\xi,n}; \xi \in \Xi_n\}$ be an experiment indexed by a parameter space Ξ_n. One is given in addition, for each $n \in N$, a subset Θ_n of the k-dimensional Euclidean space Θ and a function $\theta \rightsquigarrow \xi_n(\theta)$ from Θ_n to Ξ_n. The small vicinity of interest is the set $\xi_n(\Theta_n)$. For instance Ξ_n may be the k-dimensional Euclidean space itself and the functions ξ_n may be of the type $\xi_n(\theta) = \xi_n(0) + \delta_n\theta$ with δ_n tending to zero.

In brief the experiments of interest are not really the $\hat{\mathscr{E}}_n$ but the experiments $\mathscr{E}_n = \{P_{\theta,n}; \theta \in \Theta_n\}$ with $P_{\theta,n} = Q_{\xi_n(\theta),n}$.

To avoid extra complications we shall consider only cases where the following conditions are satisfied.

(A1) *If m and n are elements of N such that $m < n$ then $\Theta_m \subset \Theta_n$. Furthermore $\Theta = \cup_n \Theta_n$.*

(A2) *For every finite subset $S \subset \Theta$ the experiments $\mathscr{E}_{n,S} = \{P_{\theta,n}; \theta \in S\}$ have a limit in the sense of Section 2 and 3.*

One can extend arbitrarily the map $\theta \rightsquigarrow P_{\theta,n}$ to the whole of Θ. The experiment $\{P_{\theta,n}; \theta \in \Theta\}$ will still be denoted $\mathscr{E}_n$.

According to Section 3 these experiments $\mathscr{E}_n$ converge weakly to a certain limit $\mathscr{E}$, which is, of course, independent of the manner in which the $P_{\theta,n}$ are defined outside Θ_n.

The following examples are taken from the standard statistical example list.

EXAMPLE 1. Let $r_n(t)$ be the likelihood ratio $r_n(t) = dP_{t,n}/dP_{0,n}$. One assumes that there are k-dimensional random vectors Y_n and a positive definite matrix Γ such that $r_n(t) - \exp\{tY_n - \frac{1}{2}t\Gamma t'\}$ converges in $P_{0,n}$ probability to zero for each $t \in \Theta$. Furthermore one assumes that the distribution $\mathscr{L}[Y_n|P_{0,n}]$ converges to a Normal distribution with mean zero and covariance matrix Γ.

The approximation to $r_n(t)$ can also be rewritten $\exp\{\frac{1}{2}X_n\Gamma X_n'\} \exp\{-\frac{1}{2}(X_n - t)\Gamma(X_n - t)'\}$ with $\Gamma X_n = Y_n$. The limit experiment $\mathscr{E}$ is the experiment $\{P_\theta; \theta \in \Theta\}$ with P_θ normal with mean θ and covariance matrix Γ^{-1}.

EXAMPLE 2. Let N be the set of integers. Consider n independent, identically distributed variables $U_1, U_2, \cdots, U_n$, which are uniformly distributed on the interval $[0, 1 + \delta_n\theta]$ with θ such that $1 + \delta_n\theta > 0$ and $\delta_n = 1/n$. Let Y_n be the maximum of the observations $U_1, \cdots, U_n$. Then $\mathscr{L}\{n[(1 + \delta_n\theta) - Y_n]|1 + \delta_n\theta\}$ converges to the exponential distribution which has density e^{-x} on the positive part of the line. The measures P_θ of the limit experiment $\mathscr{E} = \{P_\theta; \theta \in \Theta\}$ have densities $e^{-(x-\theta)}$, $x \geqq \theta$.

EXAMPLE 3. Let $U_{n,j}$, $j = 1, 2, \cdots, n$, be independent identically distributed with individual distribution uniform on the interval $[\theta/n, (\theta/n) + 1]$. Let $Z_n' = n \min_j U_{n,j}$ and let $Z_n'' = n \max_j [U_{n,j} - 1]$. The pair (Z_n', Z_n'') has a limiting distribution equal to that of a pair (S, T) with $S = \theta + X$ and $T = \theta - Y$ where X and Y are independent variables such that $P[X \geqq x] = P[Y \geqq x] = e^{-x}$ for $x > 0$. The limit experiment $\mathscr{E}$ is the experiment where $P_\theta = \mathscr{L}\{(\theta + X, \theta - Y)\}$.

In all three examples we have stated a description of the limit experiment $\mathscr{E}$. Of course one needs to prove that $\mathscr{E}_n$ tend to the limit. For normal Example 1 note that the statistics $\{X_n\}$ are "distinguished" in the sense of Proposition 7. For the other two examples it is enough to remark that the statistics $X_n = n(Y_n - 1)$ in Example 2 and (Z'_n, Z''_n) in Example 3 are sufficient statistics whose *densities* converge. Thus they are automatically distinguished.

For any one of these examples let $\{T'_n\}$ be some other family of statistics. Suppose for instance that for each $\theta \in \Theta$ the distributions $\mathscr{L}(T'_n|\theta)$ converge to a limit F'_θ and that F'_θ is F'_0 shifted by the amount θ. That is $\mathscr{L}[T'_n - \theta|\theta]$ tends to a limit F'_0.

In Examples 1 and 2, Proposition 9 and 10 insure the existence of some probability measure Q such that $F'_0 = Q * F_0$ for the distribution $F_0 = \lim \mathscr{L}[X_n|\theta = 0]$.

In Example 3 things are more complicated since the shift group does not operate transitively on the plane. Introducing new variables $\xi = \frac{1}{2}(S + T)$ and $\eta = S - T$, let $H_\theta = \mathscr{L}[(\xi, \eta)|\theta]$. Clearly $\mathscr{L}[(\xi, \eta)|\theta] = \mathscr{L}[(\xi + \theta, \eta)|0] = H$ say. Simple computation shows then that the limiting distribution $G = F'_0$ must be obtainable by the formula

$$\int f(z)G(dz) = \iiint f(z + x)\mu(dz|y)C(dx|y)M(dy), \tag{6.1}$$

where C and M are the conditional and marginal distributions such that $H(dx, dy) = C(dx|y)M(dy)$ and where for each y the symbol $\mu(dz|y)$ represents a probability measure in z.

In other words, for each y one convolutes the conditional distribution $C(dx|y)$ with some probability measure μ which depends on y. Then one averages the result over all values of y according to the marginal distribution M.

Some other statements can be obtained by application of the results of Section 4. We shall state a few for the case of the normal Example 1. Take as loss function the quadratic $(\theta - t)\Gamma(\theta - t')$. Then the normal vector X of the limit experiment is a minimax estimate with risk identically equal to the dimension k of Θ. Suppose then that $\{T'_n\}$ is any sequence of estimates. Let $R_n(\theta)$ be the risk of T'_n at θ. Let $\varepsilon > 0$ be given. By Proposition 5, Corollary 3, there is some *finite* set $S \subset \Theta$, some n_0 such that $\sup\{R_n(\theta); \theta \in S\} > k - \varepsilon$ for all $n \geqq n_0$. Furthermore S and n_0 do not depend on the choice of T'_n. This gives part of a result of Hájek in [5].

If $k \leqq 2$, the estimate X is also admissible. Corollary 2 of Proposition 5 says then that if $a = \sup_{n>m} R_n(\theta_0) < k$ for a given θ_0 there is a finite set S an n_0 and an $\alpha > 0$ such that $\sup\{R_n(\theta); \theta \in S\} \geqq k + \alpha$ for $n \geqq n_0$. Here again α, n_0 and S depend only on θ_0 and $a < k$. This strengthens considerably a result of [9] according to which super efficiency at one point must imply misbehavior nearby. The result of [9] was proved only for $k = 1$. Of course the result does not extend to $k \geqq 3$ since X is no longer admissible.

It was proved in [13] that the χ^2 test which rejects $\theta = 0$ if $X'\Gamma X$ is larger than a given c_0 has best minimum power over the surface $\theta\Gamma\theta' = c_1$. Suppose then that $\alpha_0 = P_0\{X'\Gamma X \geqq c_0\}$, $\beta_0 = P_\theta[X'\Gamma X \geqq c_0]$ for $\theta\Gamma\theta' = c_1$. Let $\alpha_1 \leqq \alpha_0$

and $\beta_1 \geqq \beta_0$ be two numbers. Suppose for instance $\beta_1 > \beta_0$ and let ϕ_n be tests such that $\int \phi_n dP_{0,n} \leqq \alpha_1$. Then there is some finite subset S, $S \subset \{\theta; \theta\Gamma\theta' = c_1\}$, and some n_0 such that if $n \geqq n_0$ we have $\int \phi_n dP_{\theta,n} < \beta_1$ for at least some $\theta \in S$. This is another application of Proposition 5, Corollary 3.

For such tests one can even say more. A. Birnbaum in [1] proved that every convex subset of the Euclidean space is admissible as a test of $\theta = 0$ against $\theta \neq 0$. Thus suppose that $\alpha_1 \leqq \alpha_0$, $\beta_1 \geqq \beta_0$ with at least one of the inequalities holding strictly. Let t be a particular point such that $t\Gamma t' = c_1$. There is then an n_0, and $\varepsilon > 0$ and a finite set $S \subset \Theta$ such that if $n \geqq n_0$ the inequalities $\int \phi_n dP_{0,n} \leqq \alpha_1$ and $\int \phi_n dP_{t,n} \geqq \beta_1$ imply $\varepsilon + \int \phi_n dP_{\theta,n} \leqq P_\theta[X'\Gamma X \geqq c_1]$ for some $\theta \in S$.

Analogous statements can be made for the cases of Examples 2 and 3. We shall leave them to the care of the reader.

REFERENCES

[1] A. BIRNBAUM, "Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests," *Ann. Math. Statist.*, Vol. 26 (1955), pp. 21–36.

[2] D. BLACKWELL, "Comparison of experiments," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 93–102.

[3] J. L. DENNY, telephone conversation, 1968.

[4] J. HÁJEK, "A characterization of limiting distributions of regular estimates," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, Vol. 14 (1970), pp. 323–330.

[5] ———, "Local asymptotic minimax and admissibility in estimation," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 197x, Vol. 1, pp. xxx–xxx.

[6] H. HEYER, "Erschöpfheit und Invarianz beim Vergleich von Experimenten," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, Vol. 12 (1969), pp. 21–25.

[7] A. IONESCU TULCEA and C. IONESCU TULCEA, "On the existence of a lifting commuting with the left translations of an arbitrary locally compact group," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1967, Vol. 2, Part 1, pp. 63–97.

[8] L. LECAM, "Sufficiency and asymptotic sufficiency," *Ann. Math. Statist.*, Vol. 35 (1964), pp. 1419–1455.

[9] ———, "On some asymptotic properties of maximum likelihood estimates and related Bayes estimates," *Univ. California Publ. Statist.*, Vol. 1, (1953), pp. 277–330.

[10] ———, *Théorie Asymptotique de la Décision Statistique*, Montreal, University of Montreal Press, (1969).

[11] E. N. TORGERSEN, "Comparison of experiments when the parameter space is finite," Ph.D. thesis, University of California Berkeley, 1968.

[12] ———, "Comparison of translation experiments," unpublished manuscript, May 1969.

[13] A. WALD, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, Vol. 54, (1943), pp. 426–482.