LIMITS OF EXPERIMENTS AND A THEOREM OF J. HÁJEK

By

L. LeCam*

University of California, Berkeley

I. Introduction. In a recent paper [1] J. Hájek proved a remarkably simple result on the limiting distributions of estimates of a vector parameter θ .

The present paper gives an alternate proof of the same result. The proof given here cannot be called simple or elementary. It relies on some general facts where are themselves consequences of results in [2]. The reader less inclined to abstraction may find another approach in [3]. However, if these general facts are granted the proof of the result itself becomes immediate. Since these general facts appear to be of interest by themselves and not widely known we have emphasized them.

A direct simple proof of Hájek's result has been given by P. Bickel in [4].

The present paper is organized as follows: Section 2 recalls a number of definitions and theorems which are variations

---

*This paper was written with partial support of U.S. Army Research Office, Contract DA-ARO-D-31-124-G1135.

on those given by the present author in [2]. The third section considers more specifically the behavior of experiments under passages to the limit. The main result there is that when distributions converge the experiment formed by their limits is weaker than the limit of the experiments the distributions defined on their way to the limit. This is intuitively obvious but requires proof.

An application of this to the case where one statistic has limiting distributions whose experiment coincides with the limit of the experiments give a Markov kernel statement. Shift invariance of the limit experiments allow this Markov kernel to be a convolution kernel. This yields Hajek's result.

II. Experiments indexed by a given set. Recall that an abstract L-space (hereinafter called L-space) is a Banach lattice whose norm satisfies the relation $||\mu + \nu|| = ||\mu|| + ||\nu||$ if $\mu \geq 0$, $\nu \geq 0$. Let Θ be a set.

Definition 1. An experiment $\mathcal{E}$ indexed by Θ is a function $\theta \rightsquigarrow P_\theta$ form Θ to an L-space L such that $P_\theta \geq 0$ and $||P_\theta|| = 1$.

We shall say that $\mathcal{E}$ generates L if L is the smallest L-space containing the family $\{P_\theta; \theta \in \Theta\}$.

Definition 2. Let L_1 and L_2 be two abstract L-spaces.

A transition A from L_1 to L_2 is a positive linear map from L_1 to L_2 such that $||A\mu^+|| = ||\mu^+||$ for all $\mu \in L_1$.

<u>Definition</u> 3. Let Θ be a given set. Let $\mathcal{E} : \theta \rightsquigarrow P_\theta$ be an experiment generating a space L_1 and let $\mathfrak{F} : \theta \rightsquigarrow Q_\theta$ be an experiment generating a space L_2.

The deficiency $\delta(\mathcal{E},\mathfrak{F})$ of $\mathcal{E}$ relative to $\mathfrak{F}$ is the number

$$\delta(\mathcal{E},\mathfrak{F}) = \inf_A \sup_\theta ||AP_\theta - Q_\theta||$$

where the infimum is taken over all transitions from L_1 to L_2.

<u>Definition</u> 4. The distance between the experiments $\mathcal{E}$ and $\mathfrak{F}$ is the number

$$\Delta(\mathcal{E},\mathfrak{F}) = \max[\delta(\mathcal{E},\mathfrak{F}), \delta(\mathfrak{F},\mathcal{E})] \quad .$$

<u>Definition</u> 5. Two experiments $\mathcal{E}$ and $\mathfrak{F}$, both indexed by Θ are called equivalent if $\Delta(\mathcal{E},\mathfrak{F}) = 0$. If $\delta(\mathcal{E},\mathfrak{F}) = 0$ then $\mathcal{E}$ is called better than $\mathfrak{F}$ and $\mathfrak{F}$ is called weaker than $\mathcal{E}$.

Let $\mathcal{E} : \Theta \rightsquigarrow P_\theta$ be an experiment generating an L-space L. This Banach space L has a dual M which is an abstract M-space in the usual sense $(||f^+ \vee g^+|| = ||f^+|| \vee ||g^+||.)$ Furthermore M has an "identity" I defined by $\langle I,\mu\rangle = ||\mu^+|| - ||\mu^-||$.

It is possible to define on M a multiplication $(f,g) \to fg$. This is the unique bilinear function from $M \times M$ to M such that $If = fI = f$ and $f^+g^+ \geq 0$. For this, M becomes a real Banach algebra isomorphic to the space $C(Z)$ of continuous functions on a certain compact Hausdorff space Z. This space is the Kakutani space of M. The isomorphism is an isomorphism for the vector lattice, algebra and Banach space structures of M and $C(Z)$. Under the isomorphism the space L becomes a band in the space of Radon measures on Z.

The above formulation differs some from the usual one of [3] and from the formulation used in [2]. In this latter paper an experiment indexed by Θ is defined as:

1) a vector lattice E with a unit I
2) a family $\{P_\theta ; \theta \in \Theta\}$ of positive normalized linear functionals on E
3) The vector lattice E is a vector lattice (for the pointwise operations) of bounded functions on some set $\mathcal{X}$ and I is the function identically unity. Furthermore E is complete for uniform convergence.

It can easily be shown that condition (3) does not have any essential bearing on anything proved in [2]. It was thrown in there for convenience. With condition (3) removed the present definition agrees with that of [2] except for the fact that no particular sublattice $\dot{E}$ of M need to be singled out.

The more customary definition of an experiment is that it is a family $\{P_\theta : \theta \in \Theta\}$ of probability measures on a set $\mathcal{X}$ carrying a σ-field $\mathcal{A}$.

This is rather inconvenient for many purposes. However an "abstract" experiment $\mathcal{E}$ does admit at least one and usually many representations of this type.

Among the results established in [2] figure the following. Define the deficiency $\delta_S(\mathcal{E},\mathfrak{F})$ of $\mathcal{E}$ with respect to $\mathfrak{F}$ on the set $S \subset \Theta$ as

$$\delta_S(\mathcal{E},\mathfrak{F}) = \inf_T \sup_{\theta\in S} ||TP_\theta - Q_\theta||$$

where T is a transition from the L-space of $\mathcal{E}_S = \{P_\theta : \theta\in S\}$ to the L-space of $\mathfrak{F}$ (or F_S).

<u>Theorem</u> 1 (See theorem 3 of [2]).

<u>For any arbitrary</u> Θ <u>the deficiency</u> $\delta(\mathcal{E},\mathfrak{F})$ <u>is the supremum of</u> $\delta_S(\mathcal{E},\mathfrak{F})$ <u>as</u> S <u>runs through the finite subsets of</u> Θ.

Let $\mathcal{E}$ be an experiment indexed by Θ and generating a space L. Let H be a $W(M,L)$ closed subalgebra of M such that $I \in H$. The restrictions of elements of L to H form another L-space, say L_H, quotient of L. One calls H "sufficient" if the experiment $\mathcal{F}\{\theta \rightsquigarrow P'_\theta\}$ with $P'_\theta = [P_\theta | H]$ equal to P_θ restricted to H is equivalent to $\mathcal{E}$.

<u>Theorem</u> 2 (Proposition 10 of [2]). <u>For each experiment</u> $\mathcal{E}$ <u>there exists a unique minimal sufficient subalgebra</u> H.

The function $\theta \rightsquigarrow P'_\theta = P_\theta$ restricted to the minimal H is called the minimal form of $\mathcal{E}$.

<u>Theorem</u> 3 <u>Let</u> $\mathcal{E}$ <u>and</u> $\mathcal{F}$ <u>be two experiments,</u> $\mathcal{E} = \{P_\theta : \theta \epsilon \Theta\}$ <u>indexed by</u> Θ. <u>The following are equivalent</u>:

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

b) <u>For the minimal forms</u> $\{H(\mathcal{E}), L(\mathcal{E}_H)\}$ <u>and</u> $[H(\mathcal{F}), L_H(\mathcal{F})]$ <u>there is a positive isometry between</u> $L_H(\mathcal{E})$ <u>and</u> $L_H(\mathcal{F})$ <u>which extends the correspondence</u>

$$[P_\theta | H(\mathcal{E})] \longleftrightarrow [Q_\theta | H(\mathcal{F})] \quad .$$

c) <u>There is a positive isometry between the linear spaces spanned by</u> $\{P_\theta : \theta\epsilon\Theta\}$ <u>and</u> $\{\mathcal{E}_\theta : \theta\epsilon\Theta\}$ <u>which extends the correspondence</u> $P_\theta \longleftrightarrow Q_\theta$.

<u>Proof</u> (See [2] proposition 12 and corollary.)

The usual definition, with a space $\mathcal{X}$ carrying a σ-field $\mathcal{A}$, allows the definition of "statistics" with appropriate randomizations when needed. The following definition gives an appropriate replacement for this concept.

Definition 6 Let Z be a completely regular space with space of bounded continuous functions $C^b(Z)$. Let $\mathcal{E}$ be an experiment generating an L-space $L(\mathcal{E})$.

A statistic (randomized!) defined on $\mathcal{E}$ is a transition from $L(\mathcal{E})$ to some other L-space L-space L' .

A statistic T with values in Z is a transition T from $L(\mathcal{E})$ to the space of bounded linear functionals on $C^b(Z)$. The law (or distribution) of T if θ is true is the image of P_θ by T .

When Θ is finite a particular representation used by D. Blackwell [5] is of interest. Let $\mathcal{E} = \{P_\theta ; \theta \in \Theta\}$ generate L . Let $\mu = \sum_\theta P_\theta$. There are elements u_θ in the dual M of L such that $\langle vu_\theta, \mu\rangle = \langle v, P_\theta\rangle$ for every $v \in M$. These elements are such that $u_\theta \geq 0$ and $\sum_\theta u_\theta = I$. Let Z be the Kakutani space of M . The map $z \rightsquigarrow \{u_\theta(z) ; \theta \in \Theta\}$ sends Z into the finite dimensional space R^Θ . In fact the range of this map is in the unit simplex $U = \left\{\{x_\theta\} ; \theta \in \Theta\} , x_\theta \geq 0 , \Sigma x_\theta = 1\right\}$ of R^Θ . The image ν of μ by this map is called the canonical measure of $\mathcal{E}$. It is such that $\int x_\theta d\nu = 1$ for each $\theta \in \Theta$. Conversely any positive measure ν on the Borel sets of U such that $\int x_\theta d\nu = 1$ for all θ defines an experiment $\mathcal{E}_\nu = \{Q_\theta ; \theta \in \Theta\}$, with $dQ_\theta = x_\theta d\nu$.

It is convenient to make R^Θ into a Banach space by the maximum coordinate norm. If $x \in R^\Theta$ and $x = \{x_\theta\}$; $\theta \in \Theta$ then $|x| = \max|x_\theta|$.

Let Λ be the space of functions f defined on the simplex U and such that $|f| \leq 1$ and $|f(x) - f(x')| \leq |x-x'|$. For any signed measure μ of U define its Dudley norm $||\mu||_D$ by

$$||\mu||_D = \sup\{|\int f d\mu| \; ; \; f \in \Lambda\} \; .$$

The following proposition is proved in [6]. See also [7].

<u>Proposition</u> 1 <u>Let</u> Θ <u>be a finite set</u>. <u>Let</u> $\mathcal{E}$ <u>and</u> $\mathfrak{F}$ <u>be two experiments indexed by</u> Θ . <u>Let</u> ν_1 <u>and</u> ν_2 <u>be the corresponding canonical measures on the simplex</u> U <u>of</u> R^Θ . <u>Then</u>

$$\Delta(\mathcal{E}, \mathfrak{F}) \leq ||\nu_1 - \nu_2||_D \; .$$

<u>In addition</u> $\Delta(\mathcal{E}, \mathfrak{F}) = 0$ <u>if and only if</u> $\nu_1 = \nu_2$. <u>The distance</u> Δ <u>and the Dudley norm of the difference of canonical measures are uniformly equivalent</u>.

Let us remark that the simplex U is compact (for its norm). Therefore the set of all canonical measures is compact for the Dudley norm. Let us remark also that for a given experiment $\mathcal{E}$ the associated canonical measure can be obtained from any arbitrary equivalent representation in the form of an "experiment" as defined in [2] or any equivalent representation in the form of an "experiment" of the customary "set, σ-field σ-additive measures" limit. The "canonical measure" obtained from these representations is always the same if "measures on U" are defined to be Radon extensions of linear functionals on $C(U)$.

Definition 7 Let Θ be a set. Let $\mathcal{E}$ be an experiment indexed by Θ . The class of all experiments which are indexed by Θ and equivalent to $\mathcal{E}$ will be called the type of $\mathcal{E}$.

This type will often be denoted $\dot{\mathcal{E}}$

Consider an arbitrary set Θ and an experiment $\mathcal{E} : \theta \rightsquigarrow P_\theta$ indexed by Θ . For each finite subset $F \subset \Theta$, let $\mathcal{E}_F$ be the experiment $\{\theta \rightsquigarrow P_\theta \,;\, \theta \in F\}$, let R^F be the corresponding finite dimensional space with unit simplex U_F . Each $\mathcal{E}_F$ defines a canonical measure v_F . These measures are "compatible" in the following sense. If $F_1 \subset F_2$ are two finite subsets of Θ , let s be the function defined on U_{F_2} by the sum of coordinates $s(y) = \sum_{\theta \in F_1} y_\theta$. Let Π be the transformation $y \rightsquigarrow \Pi_y$ from part of U_{F_2} to U_{F_1} defined by

$$x_\theta = \frac{y_\theta}{s(y)}$$

if $\Theta \in F_1$ and $s(y) > 0$.

If ν_{F_2} and ν_{F_1} arise from the same experiment then ν_{F_1} is the image of ν_{F_2} by the following transformation. First multiply ν_{F_2} by s getting a measure $s \cdot \nu_{F_2}$. Second take the image of this by the transformation Π . More generally let us call two signed measures μ_1 and μ_2 carried respectively by U_{F_1} and U_{F_2} "compatible" if $\mu_1 = \Pi[s \cdot \mu_2]$.

<u>Proposition</u> 2 <u>Let</u> Θ <u>be an arbitrary set</u>. <u>For each finite subset</u> F <u>of</u> Θ <u>let</u> ν_F <u>be a measure on the corresponding simplex</u> U_F . <u>Assume that each</u> ν_F <u>is</u> "canonical", <u>that is</u> $\nu_F \geq 0$ <u>and</u> $\int x_\theta \nu_F(dx) = 1$ <u>for each coordinate</u> $\theta \in F$. <u>Assume that the measures are "compatible" in the above described sense</u>.

<u>Then they are the measures induced by an experiment</u> $\mathcal{E}$ <u>indexed by</u> Θ . <u>Furthermore the type of</u> $\mathcal{E}$ <u>is well determined</u>.

<u>Proof</u> Order the finite subsets of Θ by inclusion. For each finite set $\alpha \subset \Theta$ consider the corresponding unit simplex U_α . Let C_α be the space $C(U_\alpha)$ of continuous functions on U_α . If $\alpha \subset \beta$ are two finite subsets define a function $y \rightsquigarrow s_{\alpha,\beta}(y)$ on U_β by the sum of coordinates $s_{\alpha,\beta}(y) = \Sigma\{y_\theta \ ; \ \theta \in \alpha\}$. Define a map $\Pi_{\alpha,\beta}$ by

$$(\Pi_{\alpha,\beta} y)_\theta = [s_{\alpha,\beta}(y)]^{-1} y_\theta$$

if $\Theta \in \alpha$ and $s_{\alpha,\beta}(y) > 0$.

If $\varphi \in C_\alpha$ let $\varphi A_{\alpha,\beta}$ be the function defined on U_β by

$$(\varphi A_{\alpha,\beta})(y) = s_{\alpha,\beta}(y)\ \varphi[\Pi_{\alpha,\beta}(y)] \quad .$$

One can verify that the operation so described is a positive linear map of C_α into C_β . Furthermore $(\varphi A_{\alpha,\beta})A_{\beta,\gamma} = \varphi A_{\alpha,\beta}$ if $\alpha \subset \beta \subset \gamma$.

Define a set E as follows. An element of E_1 consists of a continuous function φ belonging to some C_α and all its transforms $\varphi A_{\alpha,\beta}$ for $\alpha \subset \beta$. Identify two such elements, say φ coming from C_β and ψ coming from C_γ if $\varphi A_{\beta,\zeta} = \psi A_{\gamma,\zeta}$ for $\zeta = \beta \cup \gamma$ and therefore for all $\zeta \supset \beta \cup \gamma$. These classes form a set E . On E we can define a order structure: $f \in E$ is ≥ 0 if a function φ giving it birth is non-negative. Also one can define a vector structure.

If $\varphi \in C_\alpha$, $\psi \in C_\beta$ and $\gamma \supset \alpha \cup \beta$ one can define $\varphi + \psi$ to mean $\varphi A_{\alpha,\beta} + \psi A_{\beta,\gamma}$. Upon checking that the class so defined does not depend on the choice of γ one can verify that E is indeed a vector space and in fact a vector lattice. For each $\alpha \subset \Theta$ let $\mathcal{R}_\alpha$ be the space of (signed) Radon measures on the compact U_α . If $\alpha \subset \beta$ the may $A_{\alpha,\beta}$ has an adjoint $A'_{\alpha,\beta}$ defined by $\langle \varphi A_{\alpha,\beta}, \mu \rangle = \langle \varphi, A'_{\alpha,\beta}\mu \rangle$ which map $\mathcal{R}_\beta$ into $\mathcal{R}_\alpha$.

It is clear that the transformation $f \rightsquigarrow A'_{\alpha,\beta}\mu$ is precisely the transformation described previously between canonical measures. Note that the coordinate functions are compatible by the transformations $A_{\alpha,\beta}$. That is if $\Theta \in \alpha$ and φ is the function x_Θ on U_α then $\varphi(\Pi_{\alpha,\beta}y)s_{\alpha,\beta}(y)$ is the coordinate y_Θ in U_β.

This implies the following relation: For α and β such that $\alpha \subset \beta$ let μ_α and μ_β be canonical measures such that $\mu_\alpha = A'_{\alpha,\beta}\mu_\beta$. Let $\Theta \in \alpha$ be a particular point and let $x_\alpha(\Theta)$ (resp $x_\beta(\Theta)$) be the corresponding coordinate functions. Let P_α be the measure $dP_\alpha = x_\alpha(\Theta)d\mu\alpha$ and let $dP_\beta = x_\beta(\Theta)d\mu_\beta$. Then $P_\alpha = A'_{\alpha,\beta}P_\beta$.

One can define a class of linear functionals E' on the vector space E by taking all families $\mu = \{\mu_\alpha\}$ with $\alpha \subset \Theta$ running through all finite sets larger than some α_μ and with $\mu_\alpha \in \mathcal{R}_\alpha$ such that $\mu_\alpha = A'_{\alpha,\beta}\mu_\beta$ if $\alpha \subset \beta$ and $||\mu|| = \sup_\alpha ||\mu_\alpha|| < \infty$.

Any compatible system $\{v_\alpha\}$ of canonical measures gives for each $\Theta \in \Theta$ such a family of measures $P_\Theta = \{P_{\Theta,\alpha}\}$ which are positive and such that $||P_\Theta|| = 1$.

With the obvious definition of positivity and with the above definition of norm the space E' can be completed to an abstract L-space L .

In summary each compatible family of canonical measures can be used to define a map $\theta \rightsquigarrow P_\theta$ into a suitable L-space. The experiment so defined has the appropriate type for each finite subset of Θ . Therefore its type is well defined. This concludes the proof of the proposition.

Proposition 3 Let Θ be an arbitrary set. Let $\overline{\mathcal{E}} = \overline{\mathcal{E}}(\Theta)$ be the class of experiment types indexed by Θ . Then $\overline{\mathcal{E}}$ is a set. With the metric Δ the space $(\overline{\mathcal{E}},\Delta)$ is a complete metric space. If $\mathcal{E}$ is an experiment type $\mathcal{E} \in \overline{\mathcal{E}}(\Theta)$ and F is a finite set $F \subset \Theta$ define $\mathcal{E}_F$ by restricting the set of indices to F . Let W be the weakest topology on $\overline{\mathcal{E}}(\Theta)$ for which all the maps $\mathcal{E} \rightsquigarrow \mathcal{E}_F$ are continuous (for Δ on $\overline{\mathcal{E}}(F)$) . Then $\overline{\mathcal{E}}(\Theta)$ is a compact Hausdorff space for W .

Proof That $\overline{\mathcal{E}}(\Theta)$ is a set follows from the construction of proposition 2. Similarly $[\overline{\mathcal{E}}(\Theta),W]$ is compact because each of the $[\overline{\mathcal{E}}(F),W]$, $F \subset \Theta$ and F finite is compact. Finally $[\overline{\mathcal{E}}(\Theta),\Delta]$ is complete because the structure induced by Δ is stronger than that of W and because the set of pairs $\mathcal{E},\mathfrak{F}$ such that $\Delta(\mathcal{E},\mathfrak{F}) \leq \varepsilon$ is closed for W .

Remark 1 Although the above propositions 2 and 3 do not have direct bearing on the remainder of this paper we have included them to indicate that the present definition of experiment is as general as need be for most purposes. It would be unfortunate if a Cauchy sequence of experiments did not have an experiment for limit.

Remark 2 Consider two experiments $\theta \rightsquigarrow P_\theta$ and $\theta \rightsquigarrow Q_\theta$ with P_θ in an L-space F and Q_θ is an L-space G. In the usual circumstances where the P_θ are a dominated family of measures and where the Q_θ are Borel measures on a Polish space a transition T as defined in definition 2 is representable by a Markov kernel transition. However this need not be the case in general.

In fact Jack Denny has pointed out to us the following theorem. Let $\mathcal{E} = \{P_\theta ; \theta\in\Theta\}$ be an experiment given by probability measures P_θ on the Borel sets of the real line. Let $\mathcal{E}^n$ be the direct product of $\mathcal{E}$ by itself n-times (that is $\mathcal{E}^n$ corresponds to the usual product measures). Assume that P_θ non atomic. Then there is an experiment $\mathfrak{F} = \{Q_\theta ; \theta\in\Theta\}$, where the Q_θ are σ-additive measures on a suitable σ-field of subsets of the real line, such that

1) For each integer n the experiments $\mathcal{E}^n$ and $\mathfrak{F}^n$ are equivalent.
2) For $\mathfrak{F}^n$ the sum of the observations is a sufficient statistic.

To obtain the Q_θ one just restrict P_θ to a Hamel base which has points in each perfect set.

III. Limits of experiments.

Let Θ be a set and let Z be a fixed completely regular space. Define integrals or measures on Z as linear functionals on $C^b(Z)$. Call a filter $\{\mu_\alpha\}$ of measures vaguely convergent to μ_0 if for each $f \in C^b(Z)$ the integrals $\int f d\mu_\alpha$ converge to $\int f d\mu_0$.

Proposition 4 For each integer n let $\mathcal{E}_n$ be an experiment indexed by Θ and defined by measures $F_{\theta,n}$ on the space Z. Assume that for each θ the $F_{\theta,n}$ converge vaguely to a limit F_θ. On the space $\overline{\mathcal{E}}(\Theta)$ of experiment types let $\mathcal{E}$ be a cluster point of the sequence of types $\dot{\mathcal{E}}_n$ for the weak topology of $\overline{\mathcal{E}}(\Theta)$. Let $\mathfrak{F}$ be the experiment $\mathfrak{F} = \{\theta \rightsquigarrow F_\theta\}$.

Then $\mathfrak{F}$ is weaker than $\mathcal{E}$. That is $\delta(\mathcal{E}, \mathfrak{F}) = 0$.

Proof The proposition remains true if filters are considered instead of sequences. In either case there are cluster points because of proposition 3. According to theorem 1 it is enough to prove the result assuming that Θ is finite. This will be assumed henceforth.

In this case one can also assume that $\Delta(\mathcal{E}_n, \mathcal{E}) \to 0$ since this is certainly true for some suitably chosen subsequence.

Let C be a compact convex subset in some Euclidean space and let W be a loss function defined on $\Theta \times C$. Assume that $|W| \leq 1$ and that for each θ the map $t \rightsquigarrow W_\theta(t)$ is continuous. For any procedure σ let $R(\Theta,\sigma) = R(\Theta,\sigma;\mathfrak{F})$ be the risk at Θ in the experiment $\mathfrak{F}$.

Take an $\varepsilon > 0$ and suppose N so large that $n \geq N$ implies $\Delta(\mathcal{E}_n,\mathcal{E}) < \varepsilon$.

According to theorem 1 of [2], for any fixed σ there is a decision procedure ρ which is <u>continuous</u> and such that $\sup_{\Theta}|R(\Theta,\sigma) - R(\Theta,\rho)| < \varepsilon$. This procedure ρ may also be applied to $\mathcal{E}_n$ giving a risk $R_n(\Theta,\rho) = W_\Theta \rho F_{\Theta,n}$. Since $W_\Theta \rho$ is an element of $C^b(Z)$ there is an $N(\varepsilon,\sigma)$ such that $n \geq N(\varepsilon,\sigma)$ implies $|R_n(\Theta,\rho) - R(\Theta,\rho)| < \varepsilon$. This gives $R_n(\Theta,\rho) < R(\Theta,\sigma) + 2\varepsilon$. However since $\Delta(\mathcal{E}_n,\mathcal{E}) < \varepsilon$ for $n \geq \max[N,N(\varepsilon,\sigma)]$, this implies the existence of a procedure σ' of $\mathcal{E}$ such that

$$R(\Theta,\sigma'; \mathcal{E}) \leq R(\Theta,\sigma; \mathfrak{F}) + 3\varepsilon \quad .$$

Since this is true for every W and every σ one concludes that $\delta(\mathcal{E},\mathfrak{F}) \leq 3\varepsilon$ by application of theorem 3 of [2]. This proves the desired result.

<u>Remark</u> It is easy to construct examples where $\delta(\mathfrak{F},\mathcal{E}) = 2$. In other words it may happen that $\mathfrak{F}$ is trivial but $\mathcal{E}$ is perfect. However our next proposition shows that under special circumstances one can obtain equivalence between $\mathcal{E}$ and $\mathfrak{F}$.

Suppose Θ finite. For each integer n let

$\mathcal{E}_n = \{P_{\theta,n};\ \theta\in\Theta\}$ be an experiment. Taking a representation of the $P_{\theta,n}$ by probability measures on a set χ_n carrying a σ-field $\mathcal{A}_n$ one can define Radon Nikodym densities $\mu_{\theta,n} = \frac{dP_{\theta,n}}{d\mu_n}$ and a corresponding map S_n to the unit simplex of R^{Θ}. More abstractly this S_n is a transition from the L-space of $\mathcal{E}_n$ to the L-space dual of $C^b(R^{\Theta})$. The experiment $\{S_n P_{\theta,n};\ \theta\in\Theta\}$ is the standard representation of $\mathcal{E}_n$ on the unit simplex.

Let T_n be any statistic from $\mathcal{E}_n$ to a completely regular space Z. (See definition 6).

Proposition 5 With the notation just described assume that for each $\theta \in \Theta$ the distributions $\mathcal{L}(T_n|\theta) = T_n P_{\theta,n}$ converge vaguely to a limit F_θ. Assume also that the experiment $\mathcal{E}_n$ converge to a limit $\mathcal{E}$. Let $\mathfrak{F}$ be the experiment $\mathfrak{F} = \{F_\theta;\ \theta\in\Theta\}$. The following conditions are equivalent

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

b) For each $\varepsilon > 0$ there is a transition Γ_ε whose transpose maps $C^b(R^{\Theta})$ into $C^b(Z)$ and for which $\limsup_n \sup_\theta ||S_n P_{\theta,n} - \Gamma_\varepsilon T_n P_{\theta,n}||_D \le \varepsilon$ for the Dudley norm $||\mu||_D$.

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

Proof Let $G_{\theta,n} = S_n P_{\theta,n}$ and let $F_{\theta,n} = T_n P_{\theta,n}$. Since $\mathcal{E}_n' = \{G_{\theta,n};\ \theta\in\Theta\}$ is the canonical form of $\mathcal{E}_n$ on the simplex of R^{Θ} , convergence of $\mathcal{E}_n$ to a limit $\mathcal{E}$ implies convergence in Dudley norm of $G_{\Theta,n}$ to a limit G_{Θ} . Also $\Gamma_{\varepsilon}F_{\Theta,n} \to \Gamma_{\varepsilon}F_{\Theta}$ since the transpose of Γ_{ε} maps continuous functions into continuous functions. Thus (b) may be replaced by

$$\lim_n \sup \sup_{\Theta} ||G_{\Theta}-\Gamma_{\varepsilon}F_{\Theta}||_D \leq \varepsilon \quad .$$

In particular $G_{\Theta} = \lim_{\varepsilon\to 0} \Gamma_{\varepsilon}F_{\Theta}$. Therefore by proposition 4 the experiment $\mathcal{E} = \{G_{\Theta};\ \Theta\epsilon\Theta\}$ is weaker than $\{\Gamma_{\varepsilon}F_{\Theta}\varepsilon\ \Theta\epsilon\Theta\}$. This is in turn weaker than $\mathfrak{F}$. Thus $\delta(\mathfrak{F},\mathcal{E}) = 0$. Since $\delta(\mathfrak{F},\mathcal{E}) = 0$ by proposition 4 again we have obtained $\Delta(\mathcal{E},\mathfrak{F}) = 0$.

Conversely, assume $\Delta(\mathcal{E},\mathfrak{F}) = 0$). There is then a transition Γ such that $G_{\Theta} = \Gamma F_{\Theta}$ for all Θ . The set Λ of Lipschitz functions bounded by unity on the simplex U is a compact set for the uniform norm. Thus by theorem 1 of [2] there is a "special" transition Γ_{ε} which sends $C^b(U)$ into $C^b(Z)$ and is such that $|fG_{\Theta} - f\Gamma_{\varepsilon}F_{\Theta}| < \varepsilon$ for all $f \in \Lambda$ and all Θ . This gives (b).The last statement is a consequence of the fact that $\Delta(\mathcal{E}_n,\mathcal{E}) \to 0$ and $\Delta(\mathfrak{F}_n,\mathfrak{F}) \to 0$ implies $\Delta(\mathcal{E}_n,\mathfrak{F}_n) \to 0$ if $\Delta(\mathcal{E},\mathfrak{F}) = 0$.

Consider now an arbitrary set Θ and two fixed completely regular spaces Z_1 and Z_2 with then spaces $C^b(Z_i)$. Let T_n be a statistic from $\mathcal{E}_n$ to Z_1 and let V_n be a statistic from $\mathcal{E}_n$ to Z_2 .

For any subset $A \subset \Theta$, let $\mathcal{E}_{n,A}$ be the experiment $\{P_{\theta,n}; \theta \epsilon A\}$ so that $\mathcal{E}_n = \mathcal{E}_{n,\Theta}$.

Proposition 6 Assume that for each finite set $A \subset \Theta$, the experiment types $\dot{\mathcal{E}}_{n,A}$ have a limit. Assume also that for $\theta \in \Theta$ the distributions $\mathcal{L}(T_n|\theta)$ converge vaguely to a limit F_θ and the distributions $\mathcal{L}(V_n|\theta)$ converge vaguely to a limit G_θ. Finally, assume that for each finite subset $A \subset \Theta$ the statistics T_n satisfy condition (b) of proposition 5. Then there is a transition M such that $MF_\theta = G_\theta$ for all $\theta \in \Theta$.

This is an immediate consequence of propositions 4 and 5 and theorem 1.

To obtain Hájek's result from this proposition one can proceed to an argument by invariance as follow. Let $\mathfrak{F} = \{F_\theta; \theta\epsilon\Theta\}$ and $\mathcal{G} = \{G_\theta; \theta\epsilon\Theta\}$. Let $L(\mathfrak{F})$ be the L_space of $\mathfrak{F}$ and let $M(\mathfrak{F})$ to itself. Let A_1 be a positive linear map of $M(\mathfrak{F})$ into itself such that $1A_1 = 1$. One says that the pair (A, A_1) leaves $\mathfrak{F}$ invariant if

1) A restricted to $\{F_\theta; \theta\epsilon\Theta\}$ is a permutation.

2) For each $\mu \in M(F)$ and each θ one has $(\mu A_1)(AF_\theta) = uF_\theta$

Suppose that (B, B_1) is a pair of transformations analogous to (A, A_1) which leaves $\mathcal{G}$ invariant. We shall say that $[(A, A_1), (B, B_1)]$ leaves the pair $\mathfrak{F}$, invariant if in addition

3) $AF_{\theta} = F_{\theta'}$ implies $BG_{\theta'} = G_{\theta}$,

Let K be a transition from $L(\mathfrak{F})$ to $L(\mathcal{G})$. Then BKA is also such a transition. Furthermore if $KF_{\theta} = G_{\theta}$ then $(BKA)F_{\theta} = G_{\theta}$ by condition (3) above. Let $\mathcal{M}$ be the set of all transitions K such that $KF_{\theta} = G_{\theta}$ for all θ . The system $\mathcal{S} = [A,A_1)$, $(B,B_1)]$ maps $\mathcal{M}$ into itself by the operation $K \rightsquigarrow BKA$.

Consider not only one system $\mathcal{S}$ but a family $\{\mathcal{S}_{\alpha}\}$ of such systems and the induced family of transformations of $\mathcal{M}$. It has been shown in [2] that, if the induced family of transformations is either abelian or a solvable group, there is an element M of $\mathcal{M}$ which is invariant in the sense that $BMA = M$ for all the transformations of the family.

Thus we can add to proposition 6 the following corollary:

Proposition 7 Assume that all the conditions of proposition 6 are satisfied. Assume in addition that the pair of experiments $(\mathfrak{F}, \mathcal{G})$, $\mathfrak{F} = \{F_{\theta}; \theta\epsilon\Theta\}$, $\mathcal{G} = \{G_{\theta}; \theta\epsilon\Theta\}$ is invariant by a solvable or compact group of transformations. Then there is an invariant transition M such that $MF_{\theta} = G_{\theta}$.

In several problems one encounters an even more restricted situation as follows. Consider the situation described by proposition 6 but assume in addition that the two spaces Z_1 and Z_2 in which T_n and V_n take their values are the same space Z . Assume also that Z is a locally compact group and that the experiments $\mathfrak{F} = F_\Theta; \Theta\epsilon\Theta\}$ and $\mathcal{H} = \{G_\Theta; \Theta\epsilon\Theta\}$ are invariant by the shift operations of the group. More precisely, if μ is a finite measure on Z and $\alpha \in Z$ let $\alpha\mu$ be the measure defined by

$$\int f(z)(\alpha\mu)(dz) = \int g(\alpha z)\mu(dz) \quad .$$

One can say that $\mathfrak{F}$ is "invariant" if the operation $F_\Theta \rightsquigarrow \alpha F_\Theta$ is for each $\alpha \in Z$ a permutation of $\{F_\Theta; \Theta\epsilon\Theta\}$. Let us assume that this is the case for $\mathfrak{F}$ and $\mathcal{H}$ both, and that in addition $\alpha F_\Theta = F_{\Theta'}$ implies $\alpha G_\Theta = G_{\Theta'}$. A transition which is invariant by the operations of the group can then be described as follows.

Let Z be a topological group. Let $\mathfrak{F}$ and $\mathcal{H}$ be two experiments which are invariant by the group shifts. Let $L(\mathfrak{F})$ and $L(\mathcal{H})$ be the corresponding L-spaces. A transition M from $L(\mathfrak{F})$ to $L(\mathcal{H})$ is "invariant" if it commutes with the shift operations. That is for every $\alpha \in Z$ and $\mu \in L(\mathfrak{F})$ one has $(\alpha M)\mu = (M\alpha)\mu$. Note that if the pair $(\mathfrak{F}, \mathcal{H})$ is shift invariant then $\mu \in L(\mathfrak{F})$ implies $\alpha\mu \in L(\mathfrak{F})$ for all $\alpha \in Z$.

It is to be expected that such transitions will in fact turn out to be convolutions by a fixed measure. However, we have been able to prove this only in special cases.

Assuming that Z is locally compact and that the F_Θ and G_Θ are Radon measures one can show that a transition from $L(\mathfrak{F})$ to $L(\mathcal{H})$ which commutes with shifts is the convolution operation for some suitable probability measure at least in the following cases.

a) All the F_θ are discrete.

b) The transpose of M transforms $C^b(Z)$ into $C^b(Z)$.

c) All the F_θ are absolutely continuous with respect to the left Haar measure of the group.

Case (a) is rather trivial. For case (b) let $g \in C^b(Z)$. Consider the transpose operations. By assumption $\langle\alpha^t(M^t g), \nu\rangle = \langle M^t(\alpha^t g), \nu\rangle$ for all $\nu \in L(\mathcal{V})$. Since the transformed functions are still in $C^b(Z)$ this equality implies that $[\alpha^t M^t g](z) = [M^t \alpha^t g](z)$ for every $z \in Z$ and every $\alpha \in Z$.

Apply M^t to the space $\mathcal{K}$ of functions with compact support on Z. For each $g \in \mathcal{K}$ and $z \in Z$ evaluate $M^t g$ at z. As function of g this gives a positive linear functional P_z on $\mathcal{K}$. The equation $\alpha^t M^t g = (M^t \alpha^t) g$ everywhere can then be written $\alpha^{-1} P_{\alpha z} = P_z$. This implies the desired result.

For case (c) note that according to [8] the equivalence classes of the bounded measurable functions admit a lifting which commutes with the group operations. Considering again the equality $\langle\alpha^t(M^t g), \nu\rangle = \langle M^t(\alpha^t g), \nu\rangle$ valid for $\nu \in L(\mathcal{V})$, take a representative f of $M^t g$ in the lifting. Then $\alpha^t f = M^t(\alpha^t g)$ everywhere since $\alpha^t f$ is also in the lifting. Thus we are in the same situation as in case (b).

A particular situation, with additional countability assumptions, of case (c) appears in [3]. We have included case (b) since the condition given there is also necessary.

IV Some applications.

An application of the preceding arguments is a proof of the theorem given by J. Hájek in [1].

One considers there a sequence of experiments indexed by some open set U of a Euclidean space and a sequence $\{\delta_n\}$ of numbers $\delta_n > 0$ which tend to zero. The experiments which have limits are experiments of the type $\mathcal{E}_n = \{P_{\theta_0+\delta_n\theta,n}; |\theta| \leq b\}$ for a fixed $\theta_0 \in U$. In fact Hájek considers a slightly more general situation involving matrices instead of the sequence δ_n but this does not change the essential features of the argument.

Under the assumptions given there, one sequence of statistics T_n is such that for each θ $\mathcal{L}\{\delta_n^{-1}(T_n-\theta_0-\delta_n\theta)|\theta_0+\delta_n\theta\}$ converges to a given non-degenerate Normal distribution F the other sequence V_n is such that $\mathcal{L}\{\delta_n^{-1}(V_n-\theta_0-\delta_n\theta)|\theta_0+\}_n\theta\}$ converges to some distribution G. The particular relation assumed between T_n and the likelihood ratios shows immediately that the conditions of proposition 5 are satisfied. It follows then from proposition 4 and 5 and the invariance argument that G is convolution of F by some measure Q.

To show that the present argument is not limited to the

Normal case consider independent identically distributed observations $X_1, X_2, \cdots, X_n$ which are uniformly distributed on the interval $[0, 1+\delta_n\Theta]$ with Θ such that $1+\delta_n\Theta > 0$ and $\delta_n = \frac{1}{n}$. Let Y_n be the maximum of the observations. Then

$$\mathcal{L}\{n[(1+\delta_n\Theta) - Y_n] \mid 1+\delta_n\Theta\}$$

converges to the exponential distribution which has density e^{-x} on the line.

Let T_n be the statistic $T_n = n(1-Y_n)$. For fixed Θ this has a limiting distribution F_Θ which has density $h_\Theta(x) = e^{-(x+\Theta)}$ for $x > -\Theta$ and zero otherwise. Let $\{V_n\}$ be another sequence of statistics such that $\mathcal{L}[n(1+\delta_n\Theta) - V_n \mid 1+\delta_n\Theta]$ has a limiting distribution G for each fixed Θ. We are again in a shift invariant situation. Hence G will be the convolution $G = F*Q$ of F with some probability measure Q provided we can show that T_n satisfied the conditions of proposition 5.

For this purpose, let $\Theta_1, \Theta_2, \cdots, \Theta_\varepsilon$ be k values of Θ written so that $\Theta_1 < \Theta_2 < \cdots < \Theta_k$. Let f_Θ be the uniform density from 0 to $1+\delta_n\Theta$. Consider the integral $\varphi_n(\alpha) = \int(\prod_j f_{\Theta_j}^{\alpha_j})dx$ for values of α_j such that $\Sigma\,\alpha_j = 1$, $\alpha_j \geq 0$. The corresponding Hellinger integral for the experiment $\mathfrak{F}_n$ is simply $[\varphi_n(\alpha)]^n$. It is almost immediate that this tends to $\varphi(\alpha) = \int (\prod h_{\Theta_j}^{\alpha_j})dx$. Hence the result.

References

[1] J. Hájek - "A characterization of limiting distributions of regular estimates". Zeitschrift für Wahrscheinlichkeitstheorie u.v.G. Band 14, (1970) pp. 323-330

[2] L. LeCam - "Sufficiency and approximate sufficiency". Ann. Math. Stat. Vol. 35, No. 4 (1964) pp. 1419-1455

[3] H. Heyer - "Erschöpftheit and Invarianz beim Vergleich von Experimenten". Zeitschrift fur Wahrscheinlichkeitstheorie u.v.G Band 12, (1969) pp. 21-25

[4] P. J. Bickel - (Private communication)

[5] D. Blackwell - "Comparison of experiments" Second Berkeley Symposium on Mathematical Statistics and Probability, Univ. of Calif. Press-Berkeley 1951 pp. 93-102

[6] L. LeCam - Théorie asymptotique de la décision statistique Univ. of Montreal Press-(1969)

[7] E. Torgersen - "Comparison of experiments when the parameter space is finite". Ph.D. thesis University of California, Berkeley (1968)

[8] A. Ionescu Tulcea and C. Ionescu Tulcea - "On the existence of a lifting commuting with the left translations of an arbitrary locally compact group" Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol II part I, Univ. of Calif. Press-(1967 pp. 63-97.