An Infinite Dimensional Convolution Theorem

By L. Le Cam
U.C. Berkeley

(Research supported by NSF Grant DMS 9001710)

1. Introduction.

In 1970, J. Hájek published a remarkable result on the limiting distribution of "regular" sequences of estimates. This was done under the so called LAN conditions that involve a Euclidean space $\mathbb{R}^k$. For Hájek's convolution result, the abelian, locally compact additive group structure of $\mathbb{R}^k$ plays an important role. Inagaki [1970] obtained a similar result but under considerably more restrictive assumptions. A simplified proof was soon given by P.J. Bickel. It was not published separately but appears in the book of G.G. Roussas [1972].

Le Cam [1972] offered a different proof of Hájek's result. It was based on properties of limits of experiments together with an application of a theorem of C.H. Boll [1955]. Le Cam's result applies to locally compact groups that admit almost invariant means. The Gaussian character of the special limit in the LAN case is noticeably absent. It is replaced by a domination assumption.

Unfortunately the local compactness condition on the group does not allow direct extension to the infinite dimensional set-up of non-parameteric or semi-parametric statistics. Extensions for that situation were given by Moussatat [1976] and by Millar [1985]. These authors retained the Gaussian assumption on the special limit. H. Luschgy [1987] does not use a Gaussian assumption but conditions on the comparability of finite dimensional projections.

Recently, several authors have proposed alternate proofs of the convolution theorem. Among them one should mention particularly by D. Pollard [1990] and A. van der Vaart [1989] and [1991]. The tract by van der Vaart

[1989] contains several extensions of Hájek's result for the Gaussian case. As already mentioned the paper by van der Vaart [1991] contains infinite dimensional extensions for a particular case that includes the Gaussian situation. However, van der Vaart [1991], end of page 104, says that the general case of non Gaussian situations appears to be unsolved. The purpose of the present paper is to state a result that applies to experiments obtained by shifting general cylinder measures, as long as their finite dimensional projections are dominated. That some sort of domination condition is necessary can be shown by examples on the line, or on the plane. See Sections 5 and 9 below.

The problem itself is described in Section 2, which introduces the necessary terminology and notation. In the remaining sections we have retained the method of proof used in Le Cam [1972]. The argument is split into three parts. The first involves connections between passages to the limit for experiments and passages to the limit for distributions. A second part of the proof, given in Section 4 is an application of the Markov-Kakutani fixed point theorem in a form taken from Eberlein [1949]. This is where "almost invariant means" are involved. At this point one is led to consider positive linear operators that commute with the group shifts. The combination of parts 2 and 3 of the argument contains a slight extension of a theorem of C.H. Boll [1955]. This is stated in Section 5.

Section 6 introduce the relevant notation and terminology for cylinder measures. The convolution theorem itself is stated in Section 7, with a brief sketch of its proof. Section 8 retrieves from the general convolution theorem a result of Millar [1985]. Section 9 contains various comments on the problem considered here.

We have not given complete proofs because of lack of space. In particular proofs of the results in Section 7 are only sketched. A complete proof will be contained in a forthcoming technical report of the Department of Statistics at Berkeley. This is a revision of Tech. Rep. No. 269, which unfortunately contains a gap.

At the time of this writing, we believe that the sketch of proof given here can be made into a formal proof, but cannot entirely exclude the possibility that gaps may remain. That the situation may be somewhat delicate can be seen from the examples of Section 9.

2. Notation and terminology.

Blackwell [1953] has called "experiment" a system that consists of a set Θ and of a map $\theta \rightsquigarrow P_\theta$ to probability measures on a σ-field $\mathcal{A}$ of subsets of a set $\mathbf{X}$. For technical reasons, it is more convenient to use a slightly different definition. One considers an abstract L-space L in the sense of Kakutani, that is a Banach lattice in which the norm satisfies $\|\mu + \nu\| = \|\mu\| + \|\nu\|$ for positive elements μ and ν of L. An experiment $\mathcal{E}$ indexed by Θ is then a family $\mathcal{E} = \{P_\theta : \theta \in \Theta\}$ where each P_θ is a positive element of norm unity in L.

This definition covers in particular the case of cylinder measures that will be encountered in Sections 6 and 7. Any experiment $\mathcal{E}$ defines a subspace $L(\mathcal{E})$ of L. It is the smallest closed subspace of L that contains all the P_θ, $\theta \in \Theta$ and all the elements of L dominated by convex combinations of the P_θ. See Le Cam [1964] or [1986].

Let $\mathcal{E}$ and $\mathcal{F}$ be two experiments indexed by the same set Θ. A transition from $L(\mathcal{E})$ to $L(\mathcal{F})$ is a positive linear map that preserves the norm of positive elements. If $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ and $\mathcal{F} = \{Q_\theta : \theta \in \Theta\}$, the deficiency of $\delta(\mathcal{E}, \mathcal{F})$ of $\mathcal{E}$ with respect to $\mathcal{F}$ is the number

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

where the inf is taken over all transitions T from $L(\mathcal{E})$ to $L(\mathcal{F})$. One says that it is stronger than $\mathcal{F}$ or $\mathcal{F}$ weaker than $\mathcal{E}$ if $\delta(\mathcal{E}, \mathcal{F}) = 0$.

(Note the position of the letters. The T is placed to the right of P_θ. This is in keeping with the notation $P_\theta f$ for the expectation of f under P_θ used by de Finetti or the notation $P_\theta k$ for Markov processes when P_θ is the initial distribution and $(P_\theta k)(B) = \int P_\theta(dx) k(x, B)$ for a Markov kernel k. The notation will be used consistently in the sequel, so that, for instance, we shall write $\theta\tau$ for the image of θ by a map τ).

The "distance" $\Delta(\mathcal{E}, \mathcal{F})$ between $\mathcal{E}$ and $\mathcal{F}$ is $\Delta(\mathcal{E}, \mathcal{F}) = \delta(\mathcal{E}, \mathcal{F}) \vee \delta(\mathcal{F}, \mathcal{E})$. One says that $\mathcal{E}$ and $\mathcal{F}$ are *equivalent*, or of the *same type* if $\Delta(\mathcal{E}, \mathcal{F}) = 0$.

The "distance" Δ defines a topology on the class of experiment types indexed by Θ. It also defines a weak topology: A net or filter $\mathcal{E}_\nu$ converges weakly to $\mathcal{F}$ if for any finite subset $S \subset \Theta$ the distances $\Delta(\mathcal{E}_{\nu,S}, \mathcal{F}_S)$ tend to zero, $\mathcal{E}_{\nu,S}$ being $\mathcal{E}_\nu$ with the set of indices restricted to $S \subset \Theta$.

Let Z be a set carrying a vector lattice Γ of bounded numerical functions. Assume that the constant functions belong to Γ. A *statistic with values in*

Z and available on $\mathcal{E}$ will be a transition S from $L(\mathcal{E})$ to the dual of Γ (for its uniform norm).

Let $\{S_\nu\}$ be a sequence, or net, or filter of statistics with values in Z but with S_ν defined on some experiment $\mathcal{E}_\nu = \{P_{\theta,\nu} : \theta \in \Theta\}$. The net $\{S_\nu\}$ is said to converge in distribution for $P_{\theta,\nu}$ to a limit F_θ if for each $\gamma \in \Gamma$ the evaluations $P_{\theta,\nu}S_\nu\gamma$ (of the images $P_{\theta,\nu}S_\nu$ of the $P_{\theta,\nu}$ in the dual of Γ) converge to $F_\theta\gamma$.

3. Convergence of experiments and convergence in distribution

We shall apply the following theorem, for which see Le Cam [1986], Proposition 1, page 100.

A set Z with a vector lattice Γ of bounded numerical functions such that all the constant functions belong to Γ will be called *a decision space*.

Theorem 1. *Let $\{\mathcal{E}_\nu = P_{\theta,\nu} : \theta \in \Theta\}$ be a net or sequence of experiments all indexed by the same set Θ. Let (Z, Γ) be a particular decision space. Let S_ν be a statistic available on $\mathcal{E}_\nu$ and taking values in Z. Assume that*

i) The $\mathcal{E}_\nu$ converge weakly to a limit $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$

ii) The distributions $P_{\theta,\nu}S_\nu$ converge to a limit F_θ for each $\theta \in \Theta$.

Then the experiment $\mathcal{E}$ is stronger than $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$

Proof. See Le Cam [1986], page 100. However, since that proof is complicated, here is a simple proof communicated by D. Pollard [1990].

Consider first the case where Θ is a finite set. The weak convergence of $\mathcal{E}_\nu$ to $\mathcal{E}$ implies the existence of transitions T_ν such that $\|P_\theta T_\nu - P_{\theta,\nu}\|$ tends to zero for each θ. The transition T_ν takes $L(\mathcal{E})$ to $L(\mathcal{E}_\nu)$. Combine it with the S_ν obtaining that $\|P_\theta T_\nu S_\nu - P_{\theta,\nu}S_\nu\| \leq \|P_\theta T_\nu - P_{\theta,\nu}\|$ tends to zero. Transitions from $L(\mathcal{E})$ to the dual Γ' of Γ form a compact set for the topology of pointwise convergence on $L(\mathcal{E}) \times \Gamma$. Take a cluster point B of the $\{T_\nu S_\nu\}$ for that topology. It will be such that $P_\theta B\gamma = F_\theta\gamma$ for all $\gamma \in \Gamma$. Hence $P_\theta B = F_\theta$.

To obtain the theorem for general Θ one can apply the foregoing to the experiments $\mathcal{E}_S = \{P_\theta; \theta \in S\}$ and $\mathcal{F}_S = \{F_\theta : \theta \in S\}$ for each finite subset

$S \subset \Theta$. This gives transitions B_S such that $P_\theta B_S = F_\theta$ for $\theta \in S$. These transitions are defined only in $L(\mathcal{E}_S)$ but they can be extended to transitions, say A_S, defined on the entire $L(\mathcal{E})$. A cluster point A of the A_S as S increases in Θ will be such that $P_\theta A = F_\theta$ for all $\theta \in \Theta$. Hence the result.

This result has been converted to "an asymptotic representation theorem", Theorem 3.1 of van der Vaart [1991]. It was an essential tool in the proof of the Hájek convolution theorem given by Le Cam [1972]. Another tool was an application of the Markov-Kakutani fixed point theorem, described below.

4. An application of the Markov-Kakutani theorem

Let $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ and $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ be two experiments indexed by the same set Θ. Let $\mathcal{S}$ be a set of pairs where S is a transition from $L(\mathcal{E})$ to $L(\mathcal{E})$ and S' is a transition from $L(\mathcal{F})$ to $L(\mathcal{F})$. Let us say that $\mathcal{S}$ leaves the pair $(\mathcal{E}, \mathcal{F})$ invariant if the following conditions hold for each pair $(S, S') \in \mathcal{S}$:

i) S restricted to the *set* $\mathcal{E}$ is a permutation,

ii) S' restricted to the *set* $\mathcal{F}$ is a permutation,

iii) if $P_{\theta_1} S = P_{\theta_2}$ then $Q_{\theta_2} S' = Q_{\theta_1}$.

We have emphasized the word "set" in (i) and (ii) to indicate that here we mean the *range* of the functions $\theta \rightsquigarrow P_\theta$ and $\theta \rightsquigarrow Q_\theta$ not the functions themselves. For condition (iii) the functions are involved and for a pair (S, S') the S' is a pseudo-inverse of S on the indices. Note that if T is a transition from $L(\mathcal{E})$ to $L(\mathcal{F})$ then STS' is also such a transition

Theorem 1. *Let $\mathcal{S}$ be a system of pairs (S, S') as described. Assume that $\mathcal{S}$ leaves the pair $(\mathcal{E}, \mathcal{F})$ invariant. Consider the transformations $T \rightsquigarrow STS'$ on transitions from $L(\mathcal{E})$ to $L(\mathcal{F})$. Assume that $\mathcal{E}$ is stronger than $\mathcal{F}$ and that the set of transformations $T \rightsquigarrow STS'$ admit almost invariant means (acting on their left).*

Then there is a transition T_0 from $L(\mathcal{E})$ to $L(\mathcal{F})$ such that $ST_0S' = T_0$ for all $(S, S') \in \mathcal{S}$ and such that $P_\theta T_0 = Q_\theta$ for all $\theta \in \Theta$.

For a proof see Le Cam 1986, Chapter 8, Section 2, Theorem 2. This result will be applied in Section 7 in a case where there is no need to worry about the existence of almost invariant means since they automatically exist in abelian cases. It will also be applied below in connection with Boll's theorem.

5. Boll's convolution theorem

C. Boll [1955] considers a situation describable as follows. One has a locally compact group $\mathbf{X}$ and two Radon probabilities P and Q on $\mathbf{X}$. Let PS^α be the measure P shifted by α so that if $P = L(X)$ then $PS^\alpha = L(X\alpha)$ (or $L(X+\alpha)$ if the group is abelian noted additively). Similarly, let QS^α be Q shifted by α. Let $\mathcal{E} = \{PS^\theta; \theta \in \mathbf{X}\}$ and $\mathcal{F} = \{QS^\theta : \theta \in \mathbf{X}\}$.

Lemma 1. *Suppose that $\mathcal{E}$ is better than $\mathcal{F}$. Let S^θ be the shift by θ and let S'^θ be the shift by the inverse of θ. Assume that the transformations $T \rightsquigarrow S^\theta T S'^\theta$, $\theta \in \mathbf{X}$ admit almost invariant means (acting on their left). Then there is a T_0 such that for all θ one has $T_0 = S^\theta T_0 S'^\theta$ and $PS^\theta T_0 = QS^\theta$.*

This is a direct consequence of Theorem 1, Section 4. It leads to the consideration of positive linear operations T such that $TS^\theta = S^\theta T$, that is transitions that commute with the shifts. Those are the object of the following theorem.

Theorem 1. *Let P be dominated by the Haar measure of $\mathbf{X}$. Let $\mathcal{E} = \{PS^\theta; \theta \in \mathbf{X}\}$ and let T be a transition from $L(\mathcal{E})$ to Radon measures on $\mathbf{X}$. If T commutes with shifts, that is if $TS^\theta = S^\theta T$, then T is obtainable from convolution with a Radon probability m so that for $\mu \in L(\mathcal{E})$ one has $\mu T = m * \mu$.*

The combination of Lemma 1 and Theorem 1 gives C. Boll's theorem. Actually Boll had assumed some countability restrictions. A proof without such restrictions is in Le Cam [1986]. There the proof makes use of the lifting theorem of A. and C. Ionescu Tulcea [1967]. One can bypass this theorem as follows. Let $M(\mathcal{E})$ be the dual of $L(\mathcal{E})$. Call a $\gamma \in M(\mathcal{E})$ continuous (under shift) if $\|S^\theta\gamma - \gamma\|$ tends to zero as $\theta \to 0$. One can show that such continuous elements of $M(\mathcal{E})$ are equivalence classes of a well determined uniformly continuous function on $\mathbf{X}$. Then lifting is automatic.

For many applications of Boll's theorem to Statistics, see E.N. Torgersen [1972], Hansen and Torgersen [1974], Bondar and Milnes [1981] and Torgersen [1991].

Theorems similar to Theorem 1 above have been proved long ago in the mathematical literature see Wendel [1952]. Bochner and Chandrasekharan [1949] give results of the same type for operators on Hilbert space. They also give a result analogous to our Theorem 1 for the line and Lebesgue measure (see Remark, page 215). (For this reference, we are indebted to David Brillinger). See also the paper by Brainerd and Edwards [1966] and Paterson [1983]..

In Theorem 1, the Haar measure plays a special role. It turns out that $L(\mathcal{E})$ is *stable* in the sense that if $\mu \in L(\mathcal{E})$ then $\mu S^\theta \in L(\mathcal{E})$ for all $\theta \in \mathbf{X}$. It is also *irreducible*, that is it does not contain any stable sub-band different from $L(\mathcal{E})$ or $\{0\}$.

Even on the line, we do not know whether there exist stable irreducible band other than the space of finite signed measures dominated by the Lebesgue measure or the band of finite signed purely atomic measures.

This last band arises by considering $\mathbb{R}$ as a discrete group, with Haar measure giving mass 1 to each point. It is covered by Theorem 1. If one takes a P that is constituted by a purely atomic part and a part dominated by the Lebesgue measure, then there are transitions satisfying the assumptions of Theorem 1, but not the conclusion. One obtains such a T by convoluting the absolutely continuous part with some measure m_1 and the discrete part with another measure m_2.

6. Cylinder measures

Let $\mathbf{X}$ be a locally convex vector space. Cylinder measures on $\mathbf{X}$ can be defined as follows. (See L. Schwartz [1973].)

Let $\mathcal{C}$ be the class of all closed linear subspaces of $\mathbf{X}$ that have finite codimension. If $F \in \mathcal{C}$ the quotient $\mathbf{X}/F$ is finite dimensional and there is a canonical projection Π_F of $\mathbf{X}$ into $\mathbf{X}/F$. If $G \in \mathcal{C}$ and $G \subset F$, there is a canonical projection $\Pi_{G,F}$ of $\mathbf{X}/G$ onto $\mathbf{X}/F$ and $\Pi_F = \Pi_G \Pi_{G,F}$.

Definition 1. A cylinder measure μ on $\mathbf{X}$ is a collection $\{\mu_F; F \in \mathcal{C}\}$ of ordinary (σ-additive) finite signed measures with μ_F on $\mathbf{X}/F$ and satisfying

$\mu_F = \mu_G \Pi_{G,F}$ for all G, F in $\mathcal{C}$ with $G \subset F$. A cylinder probability is a cylinder measure where all the μ_F are probability measures.

Cylinder probabilities can also be defined differently as follows. Let g be a real valued function defined on $\mathbf{X}$.

Call g an F-invariant function if $g(x_1) = g(x_2)$ whenever $x_1 - x_2 \in F$. Clearly such a function can be obtained by a composition $x \rightsquigarrow (x\Pi_F)\gamma$ where γ is defined on $\mathbf{X}/F$. Let B_F be the space of all functions defined in this way for γ a bounded Borel function on $\mathbf{X}/F$. Let $B = U\{B_F; F \in \mathcal{C}\}$.

Definition 2. A cylinder probability μ on $\mathbf{X}$ is positive linear functional μ defined on B, such that $\langle \mu, 1 \rangle = 1$ and such that, if $F \in \mathcal{C}$, the restriction of μ to B_F is σ-smooth.

Definition 1 and 2 are obviously equivalent. Let $\mathbf{X}'$ be the topological dual of $\mathbf{X}$ and let $\mathbf{X}'^*$ be the *algebraic* dual of $\mathbf{X}'$. It is a theorem of Bochner [1947] that every cylinder probability on $\mathbf{X}$ arises from a σ-additive probability on $\mathbf{X}'^*$ equipped with the smallest σ-field that makes the elements of $\mathbf{X}'$ measurable. Conversely a σ-additive probability on $\mathbf{X}'^*$ for that σ-field yields a cylinder probability on $\mathbf{X}$. Note that our projections Π_F of $\mathbf{X}$ onto $\mathbf{X}/F$ admit uniquely defined extensions, projections of $\mathbf{X}'^*$ onto $\mathbf{X}/F$. In that sense, an experiment given by cylinder measures on $\mathbf{X}$ is equivalent to one given by σ-additive measures on $\mathbf{X}'^*$.

7. A convolution theorem for cylinder measures

Consider two locally convex spaces $\mathbf{X}$ and $\mathcal{Y}$ and a continuous map τ from $\mathbf{X}$ into $\mathcal{Y}$.

On $\mathbf{X}$ let P be a cylinder probability. Let Q be a cylinder probability on $\mathcal{Y}$. Let $\mathcal{E}$ be the experiment $\mathcal{E} = \{PS^\theta; \theta \in \mathbf{X}\}$ and let $\mathcal{F} = \{QS^{\theta\tau}; \theta \in \mathbf{X}\}$.

Let us say that $\mathcal{E}$ is projection dominated if for each $F \in \mathcal{C}$ the image $P\Pi_F$ on $\mathbf{X}/F$ is dominated by the Lebesgue measure of $\mathbf{X}/F$.

We aim to prove the following theorem.

Theorem 1. *Let $\mathcal{E}$ and $\mathcal{F}$ be as described. Assume that $\mathcal{E}$ is projection dominated and stronger than $\mathcal{F}$. Then there is a cylinder measure m on $\mathcal{Y}$ such that*

$$Q = (P\tau) * m$$

The proof is long and somewhat complex. We shall only sketch what it involves. Possible steps are as follows

a) Consider $L(\mathcal{E})$ with $\mathcal{E}$ projection dominated. Then, for $F \in \mathcal{C}$, the projection of $L(\mathcal{E})$ by Π_F is the entire space $L(\mathbf{X}/F)$ of finite signed measures dominated by the Lebesgue measure of $\mathbf{X}/F$.

b) Since $\mathcal{E}$ is stronger than $\mathcal{F}$ there exists a transition K from $L(\mathcal{E})$ to $L(\mathcal{F})$ such that $PS^{\theta}K = QS^{\theta\tau}$ for all $\theta \in \mathbf{X}$ and such that $S^{\theta}KS^{-\theta\tau} = K$ for all $\theta \in \mathbf{X}$.

   (See Theorem 1, Section 4).

   This K may not be uniquely defined, even if it is a convolution. We shall assume from now on that a particular K has been selected.

c) Let $\mathcal{C}_2$ be the family of closed subspaces of $\mathcal{Y}$ that have finite codimension in $\mathcal{Y}$. Let Π_H be the canonical projection of $\mathcal{Y}$ onto $\mathcal{Y}/H$. Let K be as in (b). Suppose that for every $H \in \mathcal{C}_2$ and $\mu \in L(\mathcal{E})$ the images $\mu K \Pi_H$ have the form $\mu K \Pi_H = (\mu\tau\Pi_H) * m_H$ where m_H is some probability measure on $\mathcal{Y}/H$. Then $\mu K = (\mu\tau) * m$ for a cylinder probability m on $\mathcal{Y}$.

d) Assertion (c) shows that it will be enough to prove the theorem for the case where $\mathcal{Y}$ is finite dimensional and the map τ is onto. By convoluting PK with a Gaussian $\mathcal{N}(0, \sigma^2 I)$ on $\mathcal{Y}$ (finite dimensional), one sees that it is possible without loss of generality to assume that Q is dominated by the Lebesgue measure of $\mathcal{Y}$.

   From now on we assume that $\mathcal{Y}$ is finite dimensional, that τ is onto and that $F = \{x : x \in \mathbf{X}, x\tau = 0\}$. Furthermore Q has been convoluted with a small Gaussian measure.

   The space $\mathbf{X}$ can be written as a direct sum, or cartesian product of F with a subspace of $\mathbf{X}$ isomorphic to $\mathbf{X}/F$. We shall just chose such a subspace and still call it $\mathbf{X}/F$. (This is to avoid an excess of notation).

e) Let V be the subspace of $M(\mathcal{E})$ formed by elements $\varphi \in M(\mathcal{E})$ such that $\varphi = S^{\theta}\varphi$ for all $\theta \in F$ and such that $\|S^{\beta}\varphi - \varphi\|$ tends to zero if β tends to zero in $\mathbf{X}/F$. It is a vector lattice. Consider the experiment

$\mathcal{D} = \{P^* S^\beta; \beta \in \mathbf{X}/F\}$ where P^* is the restriction of P to V. It would be enough to show that $\mathcal{D}$ is equivalent to $\{P\Pi_F S^\beta; \beta \in \mathbf{X}/F\}$.

Remark that if $G \in \mathcal{G}$, $G \subset F$ decreases the experiments $\{P\Pi_G S^\beta; \beta \in \mathbf{X}/F\} = \mathcal{E}(G)$ converge to $\{PS^\beta; \beta \in \mathbf{X}/F\}$. Thus, for any compact $C \subset \mathbf{X}/F$ and any $\epsilon > 0$ there is a $G \in \mathcal{G}$, $G \subset F$ such that for parameter sets restricted to C

$$\delta(\mathcal{E}(G)_C, \mathcal{D}_C) < \epsilon.$$

f) Let F be as specified and let $G \in \mathcal{C}$ be such that $G \subset F$. Call G-F-shuffle an operation T carried out as follows.

One take a finite partition of unity $\{u_j; j \in J\}$ formed by elements $u_j \in B_G$ and, $[\mu \circ u_j]$ denoting the measure that has density u_j with respect to μ, one write $\mu T = \sum (\mu \circ u_j) S^{\theta_j}$ for points $\theta_j \in F$. For a given partition of unity and for given θ_j's, this is a positive linear operation on cylinder measures on $\mathbf{X}$.

One can show that if μ_i are cylinder probabilities on $\mathbf{X}$ such that the $\mu_i \Pi_G$ are dominated by the Lebesgue measure of $\mathbf{X}/G$, then for every $\epsilon > 0$ there is a G-F-shuffle T such that

$$\|\mu_1 \Pi_G - (\mu_2 T)\Pi_G\| < \epsilon.$$

A G-F-shuffle leaves all the elements of V invariant. Thus it does not change the projections on $\mathbf{X}/F$.

g) Given $G \in \mathcal{C}$, $G \subset F$, using a compactness argument and parts (a) and (f) one can show that if $\mu \in L(\mathcal{E})$, there is a $\nu \in L(\mathcal{E})$ such that μ and ν agree on V and such that $\nu \Pi_G$ is the product of $\mu \Pi_F$ by a certain probability measure on the complement of $\mathbf{X}/F$ in $\mathbf{X}/G$.

h) The experiment $\mathcal{D}$ is unchanged if one replaces the restriction of P to V by the restriction to a lattice that contains both V and a space B_G, $G \in \mathcal{G}$, $G \subset F$.

i) The combination of (e) and (h) implies that $\mathcal{D}$ is equivalent to $\{P\Pi_F S^\beta; \beta \in \mathbf{X}/F\}$.

It should be possible to obtain a simpler proof of Theorem 1 of this section by "lifting" the elements of V that are continuous under shift as in Section 5. However, the necessary argument keeps escaping us.

8. A theorem of P.W. Millar

Millar [1985] uses the situation described in Section 7. However, he uses linear maps ψ_i, $i = 1, 2$ where ψ_1 maps $\mathbf{X}$ into a space $\hat{\mathbf{X}}$ and ψ_2 maps $\mathcal{Y}$ into a space $\hat{\mathcal{Y}}$ in such a way that $P\psi_1$ and $Q\psi_2$ become Radon measures on $\hat{\mathbf{X}}$ and $\hat{\mathcal{Y}}$ respectively.

That his result can be obtained from Theorem 1, Section 7 is implied by the following lemma in which R is the image by τ of P on $\mathcal{Y}$ or $\hat{\mathcal{Y}}$ as the case may be.

Lemma 1. *Let R and M be cylinder measures on $\mathcal{Y}$ (resp $\hat{\mathcal{Y}}$), suppose that $R * M$ is a Radon measure on $\mathcal{Y}$ (resp $\hat{\mathcal{Y}}$). Then there is a z in the algebraic dual $\mathcal{Y}^{'*}$ of the dual $\mathcal{Y}'$ of $\mathcal{Y}$ (resp: algebraic dual of the dual of $\hat{\mathcal{Y}}$) such that RS^z and MS^{-z} are both Radon measures on $\mathcal{Y}$ (resp $\hat{\mathcal{Y}}$).*

This lemma is classical. It is an expression of Paul Lévy's principle: Convolution decreases concentration.

One cannot avoid the $z \in \mathcal{Y}^{'*}$ because $(RS^z) * (MS^{-z}) = R * M$. However one can take $z = 0$ if, for instance, R is symmetric around zero, or Radon on $\mathcal{Y}$.

9. Final remarks.

Remark 1. Theorem 1, Section 7 uses the condition that all projections $P\Pi_F$, $F \in \mathcal{C}$ are dominated by the Lebesgue measure of $\mathbf{X}/F$. It should be noted that this is a condition that depends on the topology used on $\mathbf{X}$. For instance take for $\mathbf{X}$ the space of sequences $x = \{x_n\}$ where x_n tends to a limit as $n \to \infty$. It can be given the norm $\|x\| = \sup_x |x_n|$, or it can be given the topology of coordinatewise convergence: $x_\nu = \{x_{\nu,n}\}$ converges to $x_0 = \{x_{0,n}\}$ if $x_{\nu,n} \to x_{0,n}$ for each n.

Let p be a probability measure on the line. Suppose that it is absolutely continuous with respect to Lebesgue measure, has compact support and expectation zero. If $Y_1, Y_2, \ldots, Y_n, \ldots$ i.i.d. with distribution p, let $X_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$. Let P be the distribution on $\mathbf{X}$ where the coordinates of X have the distribution of $X_1, X_2, \ldots$. Then P is a cylinder probability on $\mathbf{X}$ for either topology. However P satisfies the projection domination of Theorem 1 for the coordinatewise convergence of $\mathbf{X}$, but not for the topology induced by the norm. To see this, note that if P is shifted by $x \in \mathbf{X}$ then almost all sample path converge to $\lim x_n$. That is continuous for the norm.

It follows from this that the conditions of Theorem 1, Section 7 need further investigation.

Remark 2. The projection domination of Theorem 1, Section 7, cannot be entirely omitted. An example is as follows. Let $\mathbf{X}$ be $\mathbb{R}^2$ with coordinates denoted x and y. Let C be the circle $C = \{(x, y); x^2 + y^2 = 1\}$. Take on C a probability measure P dominated by the measure arc length on C. Take it so that it projects on the x-axis on the measure that has density $[1 - |x|]^+$ with respect to the Lebesgue measure on the line. To do this it is enough to take a probability measure that has density proportional to $[1-|x|]^+\sqrt{1 - x^2}$ with respect to the arc length measure.

Now the shifts $\{PS^{(x,y)}, (x, y) \in \mathbb{R}^2\}$ give an experiment $\mathcal{E}$ that is perfect in the sense that if $(x_1, y_1) \neq (x_2, y_2)$ then $PS^{(x_1,y_1)}$ and $PS^{(x_2,y_2)}$ are disjoint. The projections on the x-axis give an experiment $\mathcal{F}_1 = \{Q_1 S^x; x \in \mathbb{R}\}$ where Q_1 has density $[1 - |x|]^+$ with respect to the Lebesgue measure.

Now let $\mathcal{F} = \{QS^x; x \in \mathbb{R}\}$ where Q is uniform on $[-1/2, +1/2]$. (To be consistent with previous notation, we should write $QS^{(x,y)\tau}$ with $(x, y)\tau = x$). This is certainly weaker than $\mathcal{E}$, since $\mathcal{E}$ is perfect. However, there is no way Q could be obtained as $Q = Q_1 * M$ for some probability measure M, since the variances and the ranges do not allow this. On the contrary, with the particular choice of P, we have $Q_1 = Q * Q$.

It was not obvious at first sight that such misbehavior could not arise under the conditions of Theorem 1, Section 7.

Remark 3. The sketch of proof used in Section 7 shows that Theorem 1 of that section is not really very far from Theorem 5.2 of van der Vaart [1991].

Remark 4. Theorem 1 of Section 7 has a remarkable consequence as follows. Let $\mathbf{X}$, $\mathcal{Y}$, τ, P and Q be as in Section 7, Theorem 1. Consider two other locally convex spaces $\mathbf{X}_1$ and $\mathcal{Y}_1$. Let φ be a continuous linear map from $\mathbf{X}$ to $\mathbf{X}_1$. Let ψ be a continuous linear map from $\mathcal{Y}$ to $\mathcal{Y}_1$. Finally let ω be a continuous linear map from $\mathbf{X}_1$ to $\mathcal{Y}_1$. Assume that $\varphi\omega = \tau\psi$.

In this situation, one obtains two experiments. One of them is $\mathcal{E}_1 = \{P_1 S^{\theta\varphi}; \theta \in \mathbf{X}\}$ where P_1 is the image $P\varphi$ on $\mathbf{X}_1$. The other is $\mathcal{F}_1 = \{Q_1 S^{\theta\tau\psi}; \theta \in \mathbf{X}\}$.

It follows from Theorem Section 7 that if $\mathcal{E}$ is projection dominated and stronger than $\mathcal{F}$, then $\mathcal{E}_1$ is projection dominated and stronger than $\mathcal{F}_1$. If one could prove this directly without appealing to Theorem 1, Section 7, then the theorem itself would be immediate.

Remark 5. There are many situations in which one obtains limiting experiments that are not Gaussian, but where the "natural" shifts do not satisfy the conditions of Theorem 1, Section 7. See for instance Prakasa Rao [1968] and Ibragimov and Has'minskii [1981]. In the LAMN cases of Jeganathan [1981] one obtain transitions that are *conditionally* representable by convolution.

For some other cases where the convolution result may be applicable, see Le Cam [1975] or Chapter 8 Section 5, Proposition 7 of Le Cam [1986]. The experiment considered there arises by looking at extreme values of a sequence of i.i.d. variables. A multitude of other cases can be obtained from the stable processes studied by C. Hesse [1991].

Acknowledgement

I am indebted to a referee for pointing out the works of Paterson [1983] and Wendel [1952].

References

[1] Blackwell, D. [1953]. "Equivalent comparisons of experiments." *Ann. Math. Stat.* **24**, 265-272.

[2] Bochner, S. [1947]. "Stochastic processes." *Ann. of Math.* **48**, 1014-1061.

[3] Bochner, S. and Chandrasekharan, K. [1949]. *Fourier Transforms.* Annals of Mathematics Studies. Princeton University Press.

[4] Boll, C. [1955]. "Comparison of experiments in the infinite case and the use of invariance in establishing sufficiency." Unpublished Ph.D. Thesis, Stanford.

[5] Brainerd, B. and Edwards, R.E. [1966]. "Linear operators which commute with translations." I and II. *J. Austr. Math. Soc.* VI 289-350.

[6] Eberlein, W.F. [1949]. "Abstract ergodic theorems and weak almost periodic functions." *Trans. Amer. Math. Soc.* **67**, 217-240.

[7] Hájek, J. [1970]. "A characterization of limiting distributions of regular estimates." *Z. Wahrsch. Verw. Gebiete* **14**, 323-320.

[8] Hansen, O.H. and Torgersen, E.N. [1974]. "Comparison of linear normal experiments." *Ann. Statist.* **2**, 367-373.

[9] Hesse, C. [1991]. "On a class of linear models in economics." Tech. Report, U.C. Berkeley, Department of Statistics.

[10] Ibragimov, I. and Has'minskii, R.Z. [1981]. *Statistical Estimation: Asymptotic Theory.* Springer Verlag. New York, Heidelberg, Berlin.

[11] Inagaki, N. [1970]. "On the limiting property of a sequence of estimators with uniformity property." *Ann. Inst. Statist. Math.* **22**, 1-13.

[12] Ionescu Tulcea, A. and C. [1967]. "On the existence of a lifting commuting with left translations of an arbitrary locally compact group." *Proc. 5th Berkeley Symp. Math. Statist. Probab.* vol II, 63-97.

[13] Jeganathan, P. [1981]. "On a decomposition of the limit distribution of a sequence of estimators." *Sankhyā* **Vol. 43**, Series A, pt. 1, 26-36.

[14] Le Cam, L. [1972]. "Limits of experiments." *Proc. 6th Berkeley Symp. Math. Statist. Probab.* **Vol. I**, 249-261.

[15] Le Cam, L. [1975]. "Construction of asymptotically sufficient estimates in some non-Gaussian situations." *Proc. IASPS Satellite Symposium on Asymptotic Methods in Statistics.* J. Hájek ed. Prague.

[16] Le Cam, L. [1986]. *Asymptotic methods in statistical decision theory.* Springer Verlag. New York, Heidelberg, Berlin.

[17] Luschgy, H. [1987]. "Comparison of shift experiments on a Banach space." *Mathematical Statistics and Probability.* M.L. Puri et al (eds). Vol. A, 217-230. Reidel Pub. Co.

[18] Millar, P.W. [1985]. "Nonparametric applications of an infinite dimensional convolution theorem." *Z. Wahrsch. verw. Gebiete* **68**, 545-556.

[19] Moussatat, W. [1976]. "On the asymptotic theory of statistical experiments and some of its applications." Unpublished Ph.D. Thesis. University of California, Berkeley.

[20] Paterson, A.L.T. [1983]. "Amenability and translation experiments." *Canadian J. Math.* **25**, 49-58.

[21] Pollard, D. [1990]. Private communications, Summer 1990.

[22] Prakasa Rao, B.L.S. [1968]. "Estimation of the location of the cusp of a continuous density." *Ann. Math. Statist.* **39**, 76-87.

[23] Roussas, G.G. [1972]. *Contiguous Probability measures: Some applications in Statistics.* Cambridge Univ. Press.

[24] Schwartz, L. *Radon measures an arbitrary Topological spaces and cylindrical measures.* Oxford University Press. (See pages 74 and 174).

[25] Torgersen, E.N. [1972]. "Comparison of translation experiments." *Ann. Math. Stat.* **43**, 1383-1399.

[26] Torgersen, E.N. [1991]. *Comparison of Statistical Experiments.* Cambridge Univ. Press. xx + 675 pages.

[27] van der Vaart, A. [1989]. Statistical estimation in large parameter spaces CWI Tract #44. Centre for Mathematics and Computer Science. Amsterdam.

[28] van der Vaart, A. [1991]. "An asymptotic representation theorem." *Internat. Statist. Review* **58**, 97-121.

[29] Wendel, J.G. [1952]. "Left centralizers and isomorphisms of group algebras." *Pacific J. Math.* **2**, #3, 251-261.