An Infinite Dimensional Convolution Theorem

By L. Le Cam[1]

U.C. Berkeley

1. Introduction

In 1970, J. Hájek published a remarkable result on the limiting distribution of estimates. This was done under the so called LAN conditions which involve a Euclidean space $\mathbb{R}^k$. For Hájek's convolution result, the abelian locally compact group structure of $\mathbb{R}^k$ is important. It should also be mentioned that, at about the same time, Inagaki (1970) obtained a similar result but under considerably more restrictive assumptions. Hájek's proof was complicated. 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].

After the publication of Hájek's paper, Le Cam [1972] offered a different proof based on properties of limits of experiments and on a convolution result previously given by C.H. Boll [1955]. Le Cam's 1972 result applies to certain locally compact groups. The Gaussian character of the special limit in the LAN case is noticeably absent. It is replaced by a domination assumption. However the locally compact nature of the group did not allow direct extensions to the infinite dimensional set up used in non-parametric statistics. An extension for that situation was given by Moussatat [1976] and by Millar [1985]. These authors retained the Gaussian assumption on the special limit. More recently several authors have proposed alternate proofs of the convolution theorem. Among them one should mention more particularly D. Pollard (1990) and A. van der Vaart [1989] and [1991]. In his [1991] paper, van der Vaart says: "It appears to be unknown whether the Euclidean space $\mathbb{R}^m$ in Theorem 5.1, which plays the role of both sample space and parameter

[1]Research supported by NSF Grant DMS-9001710

space, can be replaced by a more general, infinite dimensional space. In fact, only results for the situation where V has a Gaussian distribution have been obtained so far ..."

One of the main purposes of the present paper is to prove a convolution theorem valid for arbitrary cylinder measures under the sole restriction that their finite dimensional images satisfy a domination condition.

One should ask whether the domination condition is necessary. That something is necessary should have been obvious from the statement of Proposition 10 and example 3 page 269 of Le Cam [1972]. Unfortunately, it took 20 years for this author to notice the obvious. A description of the situation is given in Section 5 below.

The problem itself is described in Section 2, which introduces the necessary terminology and notation. For the rest we have followed the pattern of proof suggested in Le Cam [1972]. It splits the argument into three separate parts. The first one involves passages to the limit for experiments and for distributions. It is explained in Section 3. The second part, given in Section 4, is an application of the Markov-Kakutani fixed point theorem. It involves "almost invariant means" and is responsible for one of the main restrictions on the groups or semigroups considered here. Part three in Section 5 uses arbitrary locally compact groups, which may or may not have almost invariant means. It shows that an equivariant transition from measures dominated by Haar measures to finite signed measures on the group is given by a convolution. This completes the theorem for the case of locally compact amenable groups. The combination of parts two and three contains an extension of a result obtained by C.H. Boll in 1955.

Section 6 gives definitions relative to cylinder measures. Section 7 gives a proof of a convolution theorem for that situation. Section 8 retrieves from that theorem the result of Millar [1985] relative to actual countably additive measures.

We are aware that the pattern of proof used here is not the simplest available for the Euclidean case. One can obtain simpler proofs as shown by D. Pollard [1990] and by van der Vaart [1991]. However, our decomposition in three steps has the merit of showing where the various assumptions, such as for instance amenability, enter into the picture.

The classical Hájek-Le Cam convolution theorem has many applications. So does the extension by Millar [1985] even though it is restricted to the Gaussian case. Applications of theorems similar to that of Boll [1955] have been described by E.N. Torgersen [1972] and by Hansen and Torgersen [1974]. See also Bondar and Milnes [1981]. Situations involving cylinder measures that are not Gaussian occur naturally various contexts. It is hoped that our theorem will be found to be applicable in some of these situations.

2. Notation and terminology

The situation considered by Hájek [1970] is one in which one has a sequence (or net) $\{\mathcal{E}_n\}$ of experiments $\mathcal{E}_n = \{P_{\theta,n}; \theta \in \Theta\}$ indexed by a fixed space Θ. The $P_{\theta,n}$ are probability measures on spaces $(\mathbf{X}_n, \mathcal{A}_n)$. For such sequences, Le Cam [1972] introduced a concept of weak convergence. It can be shown that if $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ is another experiment on a space $(\mathcal{Y}, \mathcal{B})$ the $\mathcal{E}_n$ converge weakly to $\mathcal{F}$ if and only if for every finite subset $J \subset \Theta$ the joint distribution under $P_{i,n}$ $i \in J$ of likelihood ratios $dP_{j,n}/dP_{i,n}$, $j \in J$, converge in the usual sense to corresponding distributions for $\mathcal{F}$. According to Le Cam [1964], an experiment $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ defines an L-space $L(\mathcal{E})$. It is the smallest vector space that contains all the P_θ; $\theta \in \Theta$ and has the following properties: a) if μ and ν belong to $L(\mathcal{E})$ so do their maximum $\mu \vee \nu$ and their minimum $\mu \wedge \nu$; b) the space is closed for the L_1-norm (= total variation); c) if $\mu \geq 0$ belongs to $L(\mathcal{E})$ so does every finite measure dominated by μ. One says that $L(\mathcal{E})$ is the *band* generated by $\mathcal{E}$.

A *transition* T from a space such as $L(\mathcal{E})$ to another such space, say $L(\mathcal{F})$, is a positive linear map from $L(\mathcal{E})$ into $L(\mathcal{F})$ such that if $\mu \in L(\mathcal{E})$ is positive then its image μT has the same norm as μ, that is $\|\mu T\| = \|\mu\|$. Such transitions are often, but not always, representable by Markov kernels operations so that $(\mu T)(B) = \int \mu(dx) K(x, B)$.

If Z is a completely regular topological space and if Γ is the space of bounded continuous numerical functions on Z we shall call "statistic" available on $\mathcal{E}_n = \{P_{\theta,n}; \theta \in \Theta\}$ any transition T_n from $L(\mathcal{E}_n)$ to the dual Γ' of Γ. This is a generalization of the concept of randomized transformation from $(\mathbf{X}_n, \mathcal{A}_n)$ to Z. The image $P_{\theta,n}T_n$ in Γ' called the distribution of T_n under $P_{\theta,n}$. As element of Γ' it can be evaluated at each $\gamma \in \Gamma$. The evaluation of γ will be denoted $P_{\theta,n}T_n\gamma$. Here one can interpret this symbol as the value at γ of $P_{\theta,n}T_n$ or as the value at $T_n\gamma$ (element of the dual $M(\mathcal{E}_n)$) of the element $P_{\theta,n}$ of $L(\mathcal{E}_n)$.

One says that the $P_{\theta,n}T_n$ converge in distribution if $P_{\theta,n}T_n\gamma$ converges to a limit for all $\gamma \in \Gamma$.

In the situation considered by Hájek [1970] and Le Cam [1972], the space Θ itself is a topological group. There is a particular sequence $\{T_n^*\}$ of "statistics" with values in Θ called *distinguished* in Le Cam [1972]. It is assumed that the distributions $P_{\theta,n}T_n^*$ converge to limits F_θ^*.

Consider then another sequence $\{T_n\}$ of Θ valued "statistics" and suppose that the distributions $P_{\theta,n}T_n$ also tend to limits, say F_θ.

The theorem says that under suitable assumptions on the group Θ, on the family $\{F_\theta^*; \theta \in \Theta\}$ and under a "regularity" assumption on the T_n, the limit F_θ is obtained from F_θ^* by convoluting it with a probability measure Q. That is $F_\theta = Q * F_\theta^*$ for all θ.

Our task is to elucidate under what conditions the result might be valid.

3. Limits of experiments and limits of distributions

Suppose given a sequence (or net) $\{\mathcal{E}_n\}$ of experiments $\mathcal{E}_n = \{P_{\theta,n}; \theta \in \Theta\}$ and corresponding "statistics" T_n with values in the completely regular space Z. Suppose that $\mathcal{E}_n$ tends to $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ in the weak sense of experiments. Suppose also that the distributions $P_{\theta,n}T_n$ converge to limits F_θ on Z. Then the family $\{F_\theta : \theta \in \Theta\}$ itself is an experiment, say $\mathcal{F}'$, indexed by Θ. According to an observation of Le Cam [1972], the experiment $\mathcal{F}'$ is always weaker than $\mathcal{F}$. Specifically, the following is true.

Proposition 1. *Let $\mathcal{F}$ and $\mathcal{F}'$ be as described. Then there is a transition* A *from $L(\mathcal{F})$ to $L(\mathcal{F}')$ such that $Q_\theta A = F_\theta$ for all θ.*

Proof. This is contained in Le Cam [1972] but the proof given there is not entirely convincing. The proof in Le Cam [1986] is complicated. So we shall use an argument communicated to us by D. Pollard [1990]. Consider a finite subset $J \subset \Theta$ and the restriction $\mathcal{F}_J = \{Q_j; j \in J\}$ of $\mathcal{F}$ to J. The weak convergence of $\mathcal{E}_n$ to $\mathcal{F}$ implies the existence of transitions $A_{J,n}$ from $L(\mathcal{F}_J)$ to $L(\mathcal{E}_{n,J})$ such that $\|Q_jA_{J,n} - P_{j,n}\|$ tends to zero for each $j \in J$. Combine this with the transitions T_n obtaining that $\|Q_jA_{J,n}T_n - P_{j,n}T_n\| \leq \|Q_jA_{J,n} - P_{j,n}\|$ tends to zero. Let $B_{J,n} = A_{J,n}T_n$. This is a transition from $L(\mathcal{F}_J)$ to the dual space Γ' of our space of continuous functions Such transitions from a compact set for the weak topology that makes the evaluations $PB\gamma$, $P \in L(\mathcal{F}_j)$ $\gamma \in \Gamma$ continuous. (See for instance Le Cam [1986], page 8). Take a cluster point

B_J the sequence $\{B_{J,n}, n = 1, 2...\}$. Since

$$Q_j B_{J,n}\gamma - P_{j,n} T_n \gamma \text{ and } P_{j,n} T_n \gamma - F_j \gamma$$

all tend to zero, one concludes that $Q_j B_J = F_j$ for $j \in J$. Now B_J is defined only on $L(\mathcal{F}_J)$ but it can be extended to $L(\mathcal{F})$ itself. Let A_J be a transition defined on $L(\mathcal{F})$ that extends B_J and consider $\{A_J\}$ as a net indexed by finite subsets of Θ. It has at least one weak cluster point, say A and A is such that $Q_j A = F_j$ for $j \in J$ and all finite J. Therefore $Q_\theta A = F_\theta$ for all θ. This concludes the proof of the proposition. □

Note that it is just claimed that $\mathcal{F}'$ is weaker than $\mathcal{F}$. It is easy to give examples where $\mathcal{F}'$ is strictly weaker than $\mathcal{F}$.

The case when $\mathcal{F}$ and $\mathcal{F}'$ are equivalent is a special one. In Le Cam [1972] or [1986], a sequence $\{T_n\}$ such that the two limits $\mathcal{F}$ and $\mathcal{F}'$ are equivalent is called *distinguished*. To be distinguished a sequence $\{T_n\}$ must have some asymptotic sufficiency properties, but that is not enough. The situation is described in some detail in Le Cam [1986], Chapter 7, Section 3.

4. An application of the Markov-Kakutani fixed point theorem

Let $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ be an experiment formed by measures Q_θ on a space $(\mathbf{X}, \mathcal{A})$. If α is a permutation of the set Θ let $\theta\alpha$ be the image of θ by α. It happens fairly often that one has a set, say G, of such permutations and that $\mathcal{F}$ is invariant under the action of each $\alpha \in G$. By this is meant that the experiment $\{Q_{\theta\alpha} : \theta \in \Theta\}$ is equivalent to $\mathcal{F}$ itself. Then there are transitions, say S^α, of $L(\mathcal{F})$ to $L(\mathcal{F})$ such that $Q_\theta S^\alpha = Q_{\theta\alpha}$. There are also transitions S'^α such that $Q_{\theta\alpha} S'^\alpha = Q_\theta$.

These transitions are not necessarily uniquely defined, but they exist. Consider also another experiment $\mathcal{F}' = \{F_\theta : \theta \in \Theta\}$ given by measures on a space $(\mathbf{X}', \mathcal{A}')$. Suppose that $\mathcal{F}'$ is also invariant by the action of each $\alpha \in G$. Then there are transitions R'^α and R^α such that $F_\theta R'^\alpha = F_{\theta\alpha}$ and $F_{\theta\alpha} R^\alpha = F_\theta$.

Let us suppose in addition that $\mathcal{F}'$ is weaker than $\mathcal{F}$, as happened in Section 3. Then there are transitions B from $L(\mathcal{F})$ to $L(\mathcal{F}')$ such that $Q_\theta B = F_\theta$ for all θ and, as a consequence $Q_{\theta\alpha} B = F_{\theta\alpha}$. This implies $Q_\theta S^\alpha B = F_\theta R'^\alpha$ and $Q_\theta S^\alpha B R^\alpha = F_\theta$. Here B maps $L(\mathcal{F})$ into $L(\mathcal{F}')$ and $S^\alpha B R^\alpha$ has the same property. Note that $S^\alpha B R^\alpha$ may be different from B.

Assuming selected particular transitions S^α and R^α for each $\alpha \in G$ we shall say that B is invariant under the pairs (S^α, R^α), $\alpha \in G$, if $S^\alpha B R^\alpha = B$ for all $\alpha \in G$. Now let $\mathcal{B}$ denote the set of all transitions from $L(\mathcal{F})$ to $L(\mathcal{F}')$ such that $Q_\theta B = F_\theta$ for all θ. It can be made into a topological space by giving it the weakest topology for which the maps $B \rightsquigarrow \mu B f$, $\mu \in L(\mathcal{F})$, f in the dual of $L(\mathcal{F}')$, are all continuous. A simple observation recorded in Le Cam [1986], Chapter 8, Section 2, is that for this topology the set $\mathcal{B}$ is a compact convex Hausdorff space. Each map $B \rightsquigarrow S^\alpha B R^\alpha$ is a continuous map of $\mathcal{B}$ into itself.

This is a classic situation for possible application of the Markov-Kakutani fixed point theorem, however that theorem assumes the existence of *almost invariant means*. If those almost invariant means are written to the right of their argument, their definition can be described as follows: For every neighborhood V of the origin in the space of continuous linear maps of $L(\mathcal{F})$ to $L(\mathcal{F}')$, for every finite subset $J \subset G$ and every B_0 in $\mathcal{B}$ there is a linear map M whose restriction to $\mathcal{B}$ is a continuous mapping of $\mathcal{B}$ into itself such that $S^\alpha B_0 M R^\alpha - B_0 M \in V$ for every $\alpha \in J$. This leads to the following statement.

Proposition 2. *Let $\mathcal{F}$ and $\mathcal{F}'$ and $\mathcal{B}$ be as described. Assume that for each $\alpha \in G$ a pair (S^α, R^α) has been selected. Assume also that, acting on $\mathcal{B}$, this system admits right almost invariant means. Then $\mathcal{B}$ contains a fixed point. That is, there is an $A \in \mathcal{B}$ such that $S^\alpha A R^\alpha = A$ for all $\alpha \in G$.*

This follows readily from an argument of Eberlein [1949]. The argument is reproduced in Le Cam [1986], Chapter 8, Section 2, but Eberlein's paper contains many other results.

Note that once the pairs (S^α, R^α) have been selected one can generate from them a semi-group. The pair (S^α, R^α) followed by (S^β, R^β) gives the pair $(S^\beta S^\alpha, R^\alpha R^\beta)$. It is not necessary to assume that the (S^α, R^α) give a representation (restricted to G) of the group of permutations of Θ, although that will often be the case.

The system (S^α, R^α) will admit almost invariant means whenever the semigroup so formed is commutative, or, if it is a group, it is a solvable group. Another case where not only almost invariant means, but actually invariant ones exist is when the system (S^α, R^α), $\alpha \in G$ is a compact group.

Proposition 2 brings us a step closer to the Hájek-Le Cam convolution theorem. However note that the invariant A is a map from $L(\mathcal{F})$ to $L(\mathcal{F}')$.

These two experiments live in different spaces $(\mathbf{X}, \mathcal{A})$ and $(\mathbf{X}', \mathcal{A}')$. Even if the space $(\mathbf{X}, \mathcal{A})$ and $(\mathbf{X}', \mathcal{A}')$ are the same, A need not be a "convolution" since convolution is not defined in that degree of generality.

We have mentioned above that the transitions S^α may not be uniquely defined. To see this on a simple example look at n independent variables $X_1, X_2, \ldots, X_n$ each distributed as $\mathcal{N}(\theta, 1)$ with $\theta \in I\!R$. Let α be the shift by the amount α so that, here "θ_α" means $\theta + \alpha$. A transformation S^α could be obtained by shifting each coordinate so that x_j goes to $x_j + \alpha$. It could also be obtained by shifting by α the average $\bar{X}_n$ of the X_j and then reconstituting the conditional distribution of $X_1 - \bar{X}_n$, $X_2 - \bar{X}_n, \ldots, X_n - \bar{X}_n$ through randomization. Note that while the coordinate shift has an inverse operation, the randomization just described does not. Thus what we have called S'^α is not necessarily an inverse of S^α. There is however a case where the S^α are necessarily isometries of $L(\mathcal{F})$ onto itself. For this see Le Cam [1986] Chapter 8, Section 2, Lemma 2.

5. A convolution theorem on groups

In this section we consider an experiment $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ as before. However it will be assumed that Θ itself is a locally compact group and that the measures Q_θ are probability measures on the group Θ. If $\alpha \in \Theta$ we shall let S^α be the right shift by the amount α so that if Q_θ is the distribution of a variable X, the measure $Q_\theta S^\alpha$ is the distribution of $X\alpha$ (or $X + \alpha$ if the group is abelian and noted additively). Under these circumstances one can prove the following theorem

Theorem 1. *Assume that for each θ and α in Θ the image $Q_\theta S^\alpha$ belongs to the L space $L(\mathcal{F})$. Let A be a transition from $L(\mathcal{F})$ to Radon measures on Θ. Assume that A commutes with the shifts so that $\mu A S^\alpha = \mu S^\alpha A$ for every $\mu \in L(\mathcal{F})$.*

Assume in addition that all the Q_θ are absolutely continuous with respect to the Haar measure of Θ. Then A is a convolution by a probability measure m so that $\mu A = m * \mu$.

This is proved in Le Cam [1986], Chapter 8, Section 3, Proposition 1. In conjunction with Proposition 2, Section 4 above it contains a result of C. Boll [1955]. Together with the results of Section 3 and 4, it gives a version of the Hájek-Le Cam convolution theorem for sequences of statistics $\{T_n\}$ and

$\{T_n^*\}$ where $\{T_n^*\}$ is distinguished and both sequences are "regular" in the sense that the limiting distributions F_θ and F_θ^* of $\{T_n\}$ and T_n^* are such that $F_{\theta\alpha} = F_\theta S^\alpha$ and $F_{\theta\alpha}^* = F_\theta^* S^\alpha$. Our result differs from Boll's mostly by the fact that we have no separability restriction. This is due to the fact that we can now use a lifting result of A. and C. Ionescu Tulcea [1967]. Of course Theorem 1 does not require the existence of almost invariant means, but that is because we already assume that A commutes with the shifts.

Theorems similar to Theorem 1 above must have been proved long ago in the mathematical literature, however we have not found an exact analogue. Bochner and Chandrasekharan [1949] give similar results for operators on Hilbert space. They also give the theorem for the line and Lebesgue measure in a remark page 215. (For these references we are indebted to David Brillinger).

The theorem is often stated by assuming that the family $\{Q_\theta\}$ is dominated by a σ-finite measure, instead of the Haar measure. It is easily seen that a family $\mathcal{F} = \{Q_\theta : \theta \in \Theta\}$ such that $Q_\theta S^\alpha \in L(\mathcal{F})$ for all $\alpha \in \Theta$ is dominated by the Haar measure whenever it is dominated by a σ-finite measure. See for instance Torgersen [1972].

By contrast domination by the Haar measure does not imply domination by a σ-finite measure. For an example consider the additive group of the real line with the discrete topology. Then Haar measure is the measure that gives mass unity to each individual point. It is not dominated by a σ-finite measure and a family $\{Q_\theta\}$ obtained by shifting one particular totally atomic measure is not dominated by a σ-finite measure.

It was observed in Le Cam [1972] and [1986] that the conclusion of Theorem 1 remains valid if one assumes that all the Q_θ are discrete, totally atomic, instead of assuming domination by the Haar measure. This might suggest that the conclusion of Theorem 1 remains valid if each Q_θ consist of a purely atomic part and a part dominated by the Haar measure. Unfortunately, that is not the case, as we shall now demonstrate.

Take for Θ the real line $\mathbb{R}$ with its ordinary topology and its Lebesgue measure λ. Let $Q_0 = \frac{1}{2}(H_0 + K_0)$ where H_0 is dominated by the Lebesgue measure and K_0 is purely atomic. Let $Q_\theta = Q_0 S^\theta$ and let $\mathcal{F} = \{Q_\theta, \theta \in \Theta\}$. The space $L(\mathcal{F})$ is a direct sum of two bands, say L_0 and L_1, where L_0 consists of the purely atomic finite measures and L_1 consists of measures dominated by λ. One can construct an operation A that commutes with the shifts and is not a convolution operation as follows: Take two distinct probability measures, say m_0 and m_1. Convolute the discrete part of $\mu \in L(\mathcal{F})$ with m_0 and convolute the absolutely continuous part with m_1. This gives an operation A from $L(\mathcal{F})$ to measures. It does commute with the shifts but is not representable by a single convolution. The dual of the space $L(\mathcal{F})$ splits into two parts. It can be represented as follows. Take two separate copies of the line, say $\mathbb{R}_0$ and $\mathbb{R}_1$ and make a direct sum $\mathbb{R}_0 \cup \mathbb{R}_1$. An element of the dual of $L(\mathcal{F})$ can be represented by a pair of functions (f_0, f_1). The function f_0 is bounded, otherwise arbitrary. It lives on $\mathbb{R}_0$ and gets integrated with respect to the discrete measures. The function f_1 lives on $\mathbb{R}_1$. It is bounded, Lebesgue measurable and get integrated with respect to the absolutely continuous measures. If one applies the operation A described above to, say, a bounded uniformly continuous g on the line the image Ag consists of a pair such as the above (f_0, f_1).

In the present case one can show that any transition A that commutes with the shifts is made up by convolution with a pair (m_0, m_1) of measures as just described. However one can ask what is the situation more generally? Here we do not have complete answer. A start toward an answer is as follows.

Let $\mathcal{M}$ be the space of all finite Radon measures on the line $\mathbb{R}$. Let $\mathcal{V}$ be a band in $\mathcal{M}$. Call it *stable* under shift if $\mu \in \mathcal{V}$ implies $\mu S^\alpha \in \mathcal{V}$ for all $\alpha \in \mathbb{R}$. If $\mathcal{V}_1$ and $\mathcal{V}_2$ are two stable bands, their intersection is also stable. If $\mathcal{V}$ is a stable band, the band formed by measures disjoint from $\mathcal{V}$ is also stable.

Let $\mathcal{M} = \oplus_j \mathcal{V}_j$ be a decomposition of $\mathcal{M}$ into disjoint bands. For each j, let m_j be a probability measure. Define an operation A by $\mu A = \sum_j m_j * \mu_j$ where μ_j is the component of μ in the band $\mathcal{V}_j$. This gives a positive linear map that commutes with shifts and is not a convolution. It is not known whether all transitions that commute with shifts can be represented in such a manner, but that appears doubtful. The band of measures dominated by the Lebesgue measure, L_1 and the band L_0 of discrete measures have a special property. They are *irreducible*. That is they do not contain any other stable bands except themselves and the trivial band formed by the zero measure. There are probably other irreducible bands but they have not been studied. Following a suggestion of Steve Evans it appear possible to obtain an irreducible band from the Hausdorff measure that gives mass one to the Cantor set, or to any perfect set that has self similarity properties. A stable band $\mathcal{V}$ is irreducible if and only if for every positive non zero element $\mu \in \mathcal{V}$ the smallest band containing the set $\{\mu S^\theta; \theta \in \Theta\}$ is $\mathcal{V}$ itself.

It is conceivable that transitions commuting with shifts and defined on an irreducible band would be given by convolutions. It is conceivable, but it appears to be unknown whether such transition are given by Markov kernels.

At any rate it is clear that any band that is not irreducible admits many transitions that commute with shifts but are not convolutions. Such transitions do not seem to possess one of the most interesting properties of convolutions, namely that convolution decreases concentration.

There is a large literature on linear operations that commute with shifts. See for instance L. Schwartz [1952], Helson [1954] and Brainerd and Edwards [1966]. However, it does not seem relevant to linear operations defined on bands that consist of singular measures. Also there is a large literature on linear maps T such that $(\mu \times \nu)T$: $(\mu T) * (\nu T)$ for the convolution operation $*$. However if $\mathcal{V}$ is a stable band formed by singular measures, it need not be closed under convolution. See for instance H. Rubin [1967]. So the problems are related, but different.

6. Cylinder measures

Let $\mathbf{X}$ be a real locally convex space and let $\mathcal{C}$ be the family of subspaces of $\mathbf{X}$ that are closed and have finite codimension. For any $F \in \mathcal{C}$ one can form the quotient $\mathbf{X}/F$ by calling two points x_1 and x_2 equivalent if $x_1 - x_2 \in F$. This quotient $\mathbf{X}/F$ is a finite dimensional space. The canonical projection of $\mathbf{X}$ onto $\mathbf{X}/F$ will be denoted $\Pi(\mathbf{X}, \mathbf{X}/F)$. In keeping with the notation of previous sections, we shall let such projections operate on points written to their left, so that $x\Pi$ is the image of the point x by the operation Π. A similar notational convention will be used for measures. If G is another element of $\mathcal{C}$ such that $G \subset F$ there is a canonical map $\Pi(\mathbf{X}/G, \mathbf{X}/F)$ of $\mathbf{X}/G$ onto $\mathbf{X}/F$ and

$$\Pi(\mathbf{X}, \mathbf{X}/F) = \Pi(\mathbf{X}, \mathbf{X}/G)\Pi(\mathbf{X}/G, \mathbf{X}/F).$$

Now consider a collection $\{\mu_F, F \in \mathcal{C}\}$ of ordinary finite signed measures such that μ_F lives on the finite dimensional quotient $\mathbf{X}/F$.

Such a collection is called a *cylindrical measure* on $\mathbf{X}$ if for any pair (G, F) of elements of $\mathcal{C}$ such that $G \subset F$ one has $\mu_G \Pi(\mathbf{X}/G, \mathbf{X}/F) = \mu_F$. It is called a cylindrical probability if the μ_F are probability measures.

Cylindrical probabilities arise naturally as "distributions" of linear stochastic processes. If $\mathcal{Y}$ is the dual of $\mathbf{X}$ a linear stochastic process is a map

$y \rightsquigarrow \langle y, X\rangle$ from $\mathcal{Y}$ to random variables $\langle y, X\rangle$ on some probability space $(\Omega, \mathcal{A}, P)$ such that

$$\langle \alpha_1 y_1 + \alpha_2 y_2, X\rangle = \alpha_1\langle y_1, X\rangle + \alpha_2\langle y_2, X\rangle$$

for pairs (y_1, y_2) of elements of $\mathcal{Y}$ and pairs (α_1, α_2) of scalars. Any finite collection $\{y_j; j \in J\}$ of elements of $\mathcal{Y}$ gives rise to a finite dimensional random vector $\{\langle y_j, X\rangle; j \in J\}$. If $F \subset \mathbf{X}$ is the space on which all the $\{y_j; j \in J\}$ vanish, the vectors $\{\langle y_j, x\rangle; j \in J\}$ can be considered as taking values in $\mathbf{X}/F$ and having there a distribution μ_F. The collection of these μ_F is a cylindrical probability. It is a theorem of Bochner [1947] that every cylindrical probability on $\mathbf{X}$ can be realized as a countably additive probability measure on the algebraic dual $\mathcal{Z}$ of $\mathcal{Y}$. It can also be realized as a finitely additive measure on the space $\mathbf{X}$ itself. For our purposes it will be convenient to consider a cyclindrical probability as a linear functional on a space D defined as follows. A real valued function γ defined on $\mathbf{X}$ will be called F-unchanged if $\gamma(x_1) = \gamma(x_2)$ whenever $x_1 - x_2 \in F$. Let D_F be the space of all bounded uniformly continuous functions that are F-unchanged. Let D be the union $D = \cup_F\{D_F; F \in \mathcal{C}\}$. Let $\mu = \{\mu_F; F \in \mathcal{C}\}$ be a cylindrical probability. If $\gamma \in D_F$ and $G \subset F$ is another element of $\mathcal{C}$ then γ also belongs to D_G and the expectations $\mu_F\gamma$ and $\mu_G\gamma$ are the same. Thus μ defines a positive linear functional on D. Any positive linear functional μ on D such that $\mu 1 = 1$ defines a cylinder measure if its restrictions to the spaces D_F, $F \in \mathcal{C}$ are σ-smooth. Indeed a positive σ-smooth linear functional μ defined on D_F and such that $\mu 1 = 1$ extends uniquely to a probability measure on $\mathbf{X}/F$. More exactly μ has a unique σ-smooth extension to bounded measurable functions on $\mathbf{X}/F$ (Daniell [1918]). The restriction to uniformly continuous functions will be convenient in the arguments carried out below. Note that the elements of D extend by continuity to bounded uniformly continuous functions on the algebraic dual $\mathcal{Z}$ of the dual $\mathcal{Y}$ of $\mathbf{X}$. Thus there is no need to make a distinction between cylindrical probabilities on $\mathbf{X}$ and cylindrical probabilities on $\mathcal{Z}$.

If μ and ν are two cylindrical measures on $\mathbf{X}$ their convolution $\mu * \nu$ is well defined as the collection $\mu * \nu = \{\mu_F * \nu_F, F \in \mathcal{C}\}$. The convolution operation is commutative.

All vector spaces used below will be assumed to be locally convex, even if this is not said explicitly.

7. A convolution theorem for cylindrical measures

Let P and Q be two cylindrical probabilities on the locally convex space $\mathbf{X}$. Let $\Theta = \mathbf{X}$ and let S^θ be the shift by θ so that $xS^\theta = x + \theta$ and $PS^\theta = \mathcal{L}(X + \theta)$ if $P = \mathcal{L}(X)$. Let $\mathcal{E}$ be the experiment $\mathcal{E} = \{PS^\theta; \theta \in \Theta\}$ and let $\mathcal{F} = \{QS^\theta; \theta \in \Theta\}$. We shall say that $\mathcal{E}$ is *projection dominated* if for each $F \in \mathcal{C}$ the images $PS^\theta \Pi(\mathbf{X}, \mathbf{X}/F)$ on $\mathbf{X}/F$ are dominated by a finite or σ-finite measure.

Theorem 2. *Let P and Q be two cylindrical probabilities on $\mathbf{X}$. Assume that the experiment $\mathcal{E} = \{PS^\theta; \theta \in \Theta\}$ is projection dominated. Assume also that the experiment $\mathcal{E}$ is stronger than $\mathcal{F} = \{QS^\theta; \theta \in \Theta\}$. Then there is a cylindrical probability M such that Q is the convolution $Q = P * M$.*

Note. That $\mathcal{E}$ is stronger than $\mathcal{F}$, or more informative than $\mathcal{F}$ can be taken to mean that there is a positive linear map K from the L-space $L(\mathcal{E})$ of $\mathcal{E}$ to the L-space $L(\mathcal{F})$ of $\mathcal{F}$ such that $\|\mu^+ K\| = \|\mu^+\|$ and such that $PS^\theta K = QS^\theta$ for all θ. For other definitions of "more informative" see Blackwell [1953].

Actually, according to a referee's suggestion, one needs to amend Theorem 2 to cover some of the classical results obtained when $\mathcal{E}$ is a Gaussian shift experiment. The formulation involves then two linear spaces $\mathcal{X}$ and $\mathcal{Y}$ and a continuous linear map A from $\mathcal{X}$ to $\mathcal{Y}$. One has an experiment $\mathcal{E}$ given by cylinder measures P_θ on $\mathcal{X}$ and another experiment $\mathcal{F}$ given by cylinder measures Q_θ on $\mathcal{Y}$.

Theorem 3. *Let $\mathcal{E}$ and $\mathcal{F}$ be as described. Assume that $P_\theta = P^{S_\theta}$ and $Q_\theta = QS^{\theta A}$ for all $\theta \in \Theta = \mathcal{X}$. Assume also that $\mathcal{E}$ is projection dominated and stronger than $\mathcal{F}$. Then there is a cylinder measure M on $\mathcal{Y}$ such that $Q = (PA) * M$.*

The proofs of Theorem 2 and Theorem 3 given below depend on the use of operations called *F-shuffles.* To define an F-shuffle let f_i, $i \in J$ be a finite partition of unity by elements f_i of $M(\mathcal{E})$. That is, $f_i \in M(\mathcal{E})$, $f_i \geq 0$ and $\sum_{i \in J} f_i = 1$. If $\mu \in L(\mathcal{E})$ let $(\mu \circ f_i)$ be the cylinder measure that has density f_i with respect to μ. (That is if $v \in M(\mathcal{E})$ and if $f_i v$ denote the usual product of f_i by v in $M(\mathcal{E})$, one has $\langle \mu \circ f_i, v \rangle = \langle \mu, f_i v \rangle$). For each $i \in J$ let β_i be an element of F. The operation that transforms μ into $\mu T = \sum_i (\mu \circ f_i) S^i$ is a positive linear operation that preserves the mass of positive elements. It will be called an F-shuffle.

One of the essential properties of F-shuffles is that the *image μT and μ*

have exactly the same marginals on $\mathcal{X}/F$. To describe another feature let us consider two elements F and G of $\mathcal{C}$ with $G \subset F$ and the quotients $\mathcal{X}/F$ and $\mathcal{X}/G$. One can identify $\mathcal{X}/F$ to a subspace, say Y of $\mathcal{X}/G$. This space Y has in $\mathcal{X}/G$ a complement, say Z, namely the subspace of $\mathcal{X}/G$ that is mapped to zero by the projection $M(\mathcal{X}/G, \mathcal{X}/F)$. Here we shall use (locally and temporarily) the notation W for $\mathcal{X}/G$ with subspaces Y and Z as described and call Π the projection of W on Y that cancels Z.

Lemma 1. *Let W, Y and Z be as described. Let μ_1 and μ_2 be two probability measures on W. Assume that both μ_i are dominated by the Lebesgue measure of W. Assume also that their marginals on Y by M are the same. Then for every $\epsilon > 0$ there is a Z-shuffle T such that $\|\mu_1 - \mu_2 T\| < \epsilon$.*

Proof. If the dimension of Z is m, let $u_1, u_2, \ldots, u_m$ be a basis of Z. If $z \in Z$ can be written $z = \sum z_j u_j$ let $|z| = \max_j |z_i|$.

For every $y \in Y$ let $\mu_{i,y}$ on Z be the conditional distribution of z on Z for μ_i. Let v be the common marginal of the μ_i on Y. The $\mu_{i,y}$ can also be taken so they are dominated by the Lebesgue measure of Z. Furthermore, for every $\epsilon > 0$, $\epsilon < 1/2$, there is a $b < \infty$ such that if $\mathcal{B} = \{z; z \in Z, |z| \leq b\}$ then $\int \mu_{i,y}(\mathcal{B}^c) v(dy) \leq \frac{\epsilon^2}{64}$ for $i = 1, 2$. Thus, except for a set $A \subset Y$ such that $v(A) \leq \frac{\epsilon}{8}$ one has $\mu_{i,y}(\mathcal{B}^c) \leq \frac{\epsilon}{8}$.

For $y \in A$, replace both $\mu_{i,y}$ by a probability measure multiple of the Lebesgue measure restricted to $\mathcal{B}$. For $y \in A^c$ replace the $\mu_{i,y}$ by their restrictions to $\mathcal{B}$ renormalized to be probability measures. This gives new conditional distributions say $\mu_{i,y}$, and $\int \|\mu_{i,y} - \mu'_{i,y}\| v(dy) \leq \frac{\epsilon}{8}$.

Now take on Z a probability measure ξ carried by a ball $\{z : |z| \leq \delta\}$ and having with respect to the Lebesgue measure a density that satisfies a Lipschitz condition. One can select δ so that if $\mu''_{i,y} = \mu'_{i,y} * \xi$ then $\int \|\mu'_{i,y} - \mu''_{i,y}\| v(dy) \leq \frac{\epsilon}{8}$.

The $\mu''_{i,y}$ are all carried by the same compact of Z. They have there Lebesgue densities that satisfy the same Lipschitz condition. So they can be approximated within $\frac{\epsilon}{8}$ by a finite number of them, say π_r, $r = 1, 2, \ldots, k_0$. Let A_{r_1,r_2} be the set in Y where r_1 is the first index such that $\|\mu''_{1,y} - \pi_{r_1}\| \leq \frac{\epsilon}{8}$ and r_2 is the first index such that $\|\mu''_{2,y} - \pi_{r_2}\| \leq \frac{\epsilon}{8}$. On A_{r_1,r_2} replace the $\mu_{1,y}$ by π_{r_1} and the $\mu_{2,y}$ by π_{r_2}.

For $y \in A_{r_1,r_2}$ consider the distributions π_{r_1} and π_{r_2} on Z. Take an integer $k > 2$ and make a partition of Z by cubes whose sides are of the form

$(\frac{j_s}{2^k}, \frac{j_s+1}{2^k}]$, where $s = 1, 2, \ldots, w$ indicates the coordinate that is restricted. For such a cube, say C_h, where $h = (j_1, \ldots, j_m)$ replace the measure π_{r_i} restricted to C_h by a measure proportional to the Lebesgue measure on C_h having the same total norm. This gives new measures $\bar{\pi}_{r_i}$. Selecting k large enough one can insure that $\|\bar{\pi}_{r_i} - \pi_{r_i}\| < \frac{\epsilon}{8}$. Now it is clear that one can obtain $\bar{\pi}_{r_1}$ from $\bar{\pi}_{r_2}$ by shifting appropriate slices of the type $\alpha(k_1, k_2)\bar{\pi}_{r_2} I(C_{k_2})$ where $\alpha(k_1, k_2) \in [0, 1]$. This gives a shuffle T^* on Z such that $\|\bar{\pi}_{r_1} - \bar{\pi}_{r_2} T^*\| = 0$ and therefore $\|\mu_{1,y} - \mu_{2,y} T^*\| \leq \epsilon$ for all $y \in Y$. The result follows by forming the corresponding Z-shuffle T on the measures carried by W. □

Now let us return to the experiment $\mathcal{E}$ with its spaces $L(\mathcal{E})$ and $M(\mathcal{F})$. The Z-shuffle of Lemma 1 uses bounded measurable functions defined on $W = \mathcal{X}/G$. They can be identified to elements of $M(\mathcal{E})$ that are G-unchanged. The shifts from Z, kernel of the map $M(\mathcal{X}/G, \mathcal{X}/F)$ can also be identified with shifts S^β for $\beta \in F$. Thus the result of Lemma 1 can be restated as follows:

Let μ_1 and μ_2 be two elements of $L(\mathcal{E})$ that have the same marginals on $\mathcal{X}/F$. Then for every $\epsilon > 0$, every $G \in \mathcal{C}$, $G \subset F$, there is an F-shuffle T such that the projections $= \mu_{i,G}$ on $\mathcal{X}/G$ satisfy $\|\mu_{1,G} - (\mu_2 T)_G\| < \epsilon$.

From this one can deduce the following corollary.

Lemma 2. *Let $\mathcal{S}$ be the closure pointwise on $L(\mathcal{E}) \times M(\mathcal{E})$ of the set of F-shuffles considered as a subset of the set of transitions from $L(\mathcal{E})$ to the dual of $M(\mathcal{E})$. Let μ_1 and μ_2 be two elements of $L(\mathcal{E})$ whose marginals on $\mathcal{X}/F$ are identical. Then there is a $T \in \mathcal{S}$ such that $\|\mu_1 - \mu_2 T\| = 0$.*

Proof. The set $\mathcal{S}$ is compact for the topology of pointwise convergence on $(L(\mathcal{E}) \times M(\mathcal{E})$.

The set of pairs (G, ϵ) where $G \in \mathcal{G}$ and where $\epsilon > 0$ is directed decreasingly by the natural inclusion order and the order of the line. For each (G, ϵ) Lemma 1 and its reinterpretation gives an F-shuffle $T_{G,\epsilon}$ such that $\|\mu_{1,G} - (\mu_2 T_{G,\epsilon})_G\| < \epsilon$.

Take a limit T of these $T_{G,\epsilon}$ along an ultrafilter finer than the tails of our directed set of (G, ϵ). This limit will satisfy $\|\mu_{1,G} - \mu_2 T)_G\| \leq \epsilon$ for every $G \in \mathcal{G}$ such that $G \subset F$ and every $\epsilon > 0$. Thus $\mu_1 = \mu_2 T$ as claimed □

Note. The limit T of Lemma 2 still maps $M(\mathcal{E})$ into $M(\mathcal{E})$ if it acts on its right.

We are now ready for the proof of Theorem 3 of which Theorem 2 is a special case.

Proof. The Markov-Kakutani fixed point theorem implies the existence of transition K from $L(\mathcal{E})$ to $L(\mathcal{F})$ such that $S^{\theta}KS^{-\theta A} = K$ and $P_\theta K = Q_\theta$ for all θ. Select such a K and keep it fixed throughout the rest of the proof.

In the space $\mathcal{Y}$, let $\mathcal{H}$ be the family of all closed subspaces of finite co-dimension. For $H \in \mathcal{H}$ let Π_H be the canonical map of $\mathcal{Y}$ onto $\mathcal{Y}/H$. One obtains a map $x \rightsquigarrow xA\Pi_H$ of $\mathcal{X}$ into $\mathcal{Y}/H$ by applying A first and then Π_H. Let $F(H) = \{x : xA\Pi_H = 0\}$ be the kernel of $A\Pi_H$ in $\mathcal{X}$.

Since $\mathcal{Y}/H$ is finite dimensional and since A is continuous, $F(H)$ has finite co-dimension in $\mathcal{X}$ and it belongs to the class $\mathcal{C}$.

Let D'_H be the space of numerical functions defined on $\mathcal{Y}$ by expressions of the type $y \rightsquigarrow (y\Pi_H)\varphi$ where φ is bounded, uniformly continuous on $\mathcal{Y}/H$. These functions define a subspace of $M(\mathcal{F})$. As such they are mapped by K (acting on its right) into elements of $M(\mathcal{E})$. Let γ be a positive element of D'_H and let $\eta = K\gamma$. If $\beta \in F(H)$ then $S^{-\beta A}\gamma = \gamma$. Since $S^{\beta}KS^{-\beta A} = K$ this also gives $S^{\beta}KS^{-\beta A}\gamma = K\gamma$, hence $S^{\beta}\eta = \eta$ for all $\beta \in F(H)$. Now let T be an $F(H)$-shuffle on $\mathcal{X}$. Then $T\eta = \eta$. Indeed let the shuffle be given by a partition of unity $\{f_j; j \in J\}$ and shifts $\beta_j \in F(H)$. If $\mu \in L(\mathcal{E})$, write ν_j for the measure that has density f_j with respect to μ. Then $\mu = \sum \nu_j$ and $\sum_j \nu_j S^{\beta_j} = \mu T$. Also $\langle \nu_j S^{\beta_j}, \eta\rangle = \langle \nu_j, S^{\beta_j}\eta\rangle = \langle \nu_j, \eta\rangle$. So $\langle \mu, \eta\rangle = \langle \mu, T\eta\rangle$ for all $\mu \in L(\mathcal{E})$.

The same invariance property holds for any T that is in the set $\mathcal{S}$ of limits of $F(H)$-Shuffles in Lemma 2, provided that one considers these limits as maps from $M(\mathcal{E})$ to $M(\mathcal{E})$.

We claim that this implies that for any $\mu \in L(\mathcal{E})$ the value $\langle \mu, \eta\rangle$ depends only on the marginal $\mu_{F(H)}$ of μ on $\mathcal{X}/F(H)$. Indeed let μ_1 and μ_2 be two cylinder probabilities $\mu_i \in L(\mathcal{E})$ that have the same marginals on $\mathcal{X}/F(H)$. Then, according to Lemma 2, there is a $T \in \mathcal{S}$ such that $\mu_1 = \mu_2 T$ and therefore $\langle \mu_1, \eta\rangle = \langle \mu_2, T\eta\rangle = \langle \mu_2, T\eta\rangle = \langle \mu_2, \eta\rangle$. Since $\eta = K\gamma$ with $\gamma \geq 0$ the element η is also positive. The projection of $L(\mathcal{E})$ on $\mathcal{X}/F(H)$ is the entire space L_1 of measures dominated by the Lebesgue measure of $\mathcal{X}/F(H)$ and one can identify η to a positive element of the dual M_1 of L_1. This element of M_1 will still be denoted η to avoid an excess of notational complication.

According to a theorem of A_1 and C_1 Ionescu Tulcea thee is a positive linear map ℓ of M_1 to bounded measurable functions on $\mathcal{X}/F(H)$ such that if $\varphi \in M_1$ then the equivalence class of ℓ_φ is φ itself. One can select this lifting ℓ so that it commutes with shifts on $\mathcal{X}/F(H)$.

The functions elements of the space D'_H originated from the space, say $\bar{D}'_H$ of bounded uniformly continuous functions on $\mathcal{Y}/H$. This $\bar{D}'_H$ contains the space $\mathcal{K}_H$ of continuous functions with compact support on $\mathcal{Y}/H$. For $\gamma \in \mathcal{K}_H$ let $\langle \nu_H, \gamma \rangle$ be the evaluation of $\ell_K(\Pi_H \nu)$ at zero. In keeping with the notation where $\gamma \in \bar{D}'_H$ acts on its left, let $[0, \ell_K(\Pi_H \gamma)]$ be this evaluation also called $\langle \nu_H, \gamma \rangle$. This gives a positive linear functional on $\mathcal{K}_H$. It admits an integral representation $\langle \nu_H, \gamma \rangle = \int [z, \gamma] \nu_H(dz)$ where ν_H is a Radon measure on $\mathcal{Y}/H$.

One can identify $\mathcal{X}/F(H)$ to a subspace of $\mathcal{X}$, written as direct sum of $F(H)$ and $\mathcal{X}/F(H)$. With this identification, one can apply the shifts S^θ to measures and functions on $\mathcal{X}/F(H)$.

Of course, for $\theta \in F(H)$ the shift S^θ on $\mathcal{X}/F(H)$ is the identity map. With this notation rewrite the relation $S^\theta K S^{-\theta A} = K$ as $S^\theta K = K S^{\theta A}$ and evaluate $\ell K(\Pi_H \gamma)$ at $x \in \mathcal{X}/F(H)$. One obtains $[x, \ell K(\Pi_H \gamma)] = [0, S^x \ell K(\Pi_H \gamma)] = [0, \ell K S^{xA}(\Pi_H \gamma)] = \int [x A \Pi_H + z, \gamma] \nu_H(dz)$. Then, if $\mu \in L(\mathcal{E})$ has marginal $\mu_{F(H)}$ on $\mathcal{X}/F(H)$ and if $\gamma \in \mathcal{K}_H$, one can write $\mu K(\Pi_H \gamma) = \int \{ \int [x A \Pi_H + z, \gamma] \nu_H(dz) \} \mu_{F(H)}(dx)$. In other words, for $\gamma \in \mathcal{K}_H$ the value $\mu K(\Pi_H \gamma)$ is obtained by integrating γ with respect to the convolution of ν_H by the image of μ by the map $\Pi_{F(H)} A \Pi_H$. This is true for $\gamma \in \mathcal{K}_H$ but it remains true for instance for $\gamma \in \bar{D}'_H$. Indeed K considered as map from $M(\mathcal{F})$ to $M(\mathcal{E})$ is continuous for the weak topologies $w[M(\mathcal{F}), L(\mathcal{F})]$ and $w[M(\mathcal{E}), L(\mathcal{E})]$ and $\mathcal{K}_H$ is dense in $\bar{D}'_H$.

To proceed it is convenient to look at the convolution operation so obtained as a map sending D'_H into $D_{F(H)}$. Now take two spaces H_i, $i = 1, 2$, $H_i \in \mathcal{H}$ and $H_1 \subset H_2$. We get two convolutions maps, by measures ν_{H_i}. However, since K is fixed and since $D'_{H_2} \subset D'_{H_1}$ the measure ν_{H_2} must be the restriction of ν_{H_1} to D'_{H_2}. Hence the ν_H, $H \in \mathcal{H}$ define a cylinder measure M on $\mathcal{Y}$ and the convolution relations can be summarized in the formula $\mu K \varphi = \langle (\mu A) * M, \varphi \rangle$. This completes the proof of Theorem 3 and therefore also of Theorem 2. □

Remark 1. We have used the lifting theorem of A. and C. Ionescu Tulcea [1967]. Its full force is not needed. It would have been sufficient to use what is called a *linear lifting* that commutes with shifts. For the finite dimensional spaces $\mathcal{X}/F(H)$ such linear liftings are easy to obtain. A proof was given by Dieudonné [1951]. It is reproduced in LeCam [1986] page 126. For the situation considered here there are many other methods used to convert "almost invariance" to "invariance". See Boll [1955] or Berk and Bickel [1968].

Remark 2. One of the immediate consequences of Theorem 3 is that the experiment $\mathcal{F} = \{Q_\theta : \theta \in \Theta\}$, weaker than $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ is also weaker than the experiment $\mathcal{E}A = \{P_\theta A; \theta \in \Theta\}$ obtained by mapping the P_θ to $\mathcal{Y}$ by the linear map A.

One can state this as follows.

Corollary. *Lt $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ be projection dominated and given by cylinder measures on $\mathcal{X} = \Theta$. Let B_1 be a continuous linear map to a vector space (locally convex) $\mathcal{Z}$ and let B_2 be a continuous linear map from $\mathcal{Z}$ to $\mathcal{Y}$. Let $A = B_1 B_2$ be the resulting map from $\mathcal{X}$ to $\mathcal{Y}$. On $\mathcal{Y}$, let $\mathcal{F} = \{QS^{\theta A}; \theta \in \Theta\}$.*

Suppose that $P_\theta = PS^\theta$ on $\mathcal{X}$ and that $\mathcal{E}$ is stronger than $\mathcal{F}$. Then the experiment $\mathcal{G} = \{P_\theta B_1; \theta \in \Theta\}$ is also stronger than $\mathcal{F}$ and there is a cylinder measure M on $\mathcal{Y}$ such that $Q = [(PB_1)B_2] * M$.

In particular continuous linear maps preserve the order of experiments that are obtained by shifting one cylinder measure if the stronger one is projection dominated.

Remark 3. The cylinder measure M obtained in the proof of Theorem 3 is well defined only because we selected a particular K such that $S^\theta K S^{-\theta A} = K$. Generally there may be several M such that $Q = (PA) * M$. This is so even on the line: Suppose that P on $I\!R$ has characteristic function $\tilde{P}(t) = [1 - |t|]^+$. Then $P * M_1 = P * M_2$ if $\tilde{M}_1(t) = \tilde{M}_2(t)$ for $t \in [-1, +1]$. There are many such pairs (M_1, M_2) according to Polyá's theorem on characteristic functions that are convex in t for $t \geq 0$. The possibility $P * M_1 = P * M_2$ for $M_1 \neq M_2$ cannot arise if P is Gaussian.

Remark 4. The case where P is a Gaussian cylinder measure can be treated without recourse to the shuffle operations of lemmas 1 and 2. To see this note that if $\{P_\theta; \theta \in \Theta\}$ is Gaussian shift the function $(\theta_1, \theta_2) \rightsquigarrow -8 \log \int \sqrt{dP_{\theta_1} dP_{\theta_1}}$ (if finite valued as in Millar [1985]), is the square of a Hilbertian or prehilbertian norm on $\Theta = \mathcal{X}$. Then one can write $\mathcal{X}$ as a direct sum $\mathcal{X} = F(H) \oplus V$ where V is isomorphic to $\mathcal{X}/F(H)$ but orthogonal to $F(H)$ for the prehilbertian norm. For shifts S^θ where $\theta \in V$ the experiment $\{PS^\theta; \theta \in V\}$ is equivalent to the experiment $\{(PM)S^\theta; \theta \in V\}$ where M is the canonical projection on $V = \mathcal{X}/F(H)$, by a sufficiency argument. Thus $\{(PM)S^\theta; \theta \in V\}$ is also stronger than $\{QS^{\theta A}; \theta \in V\}$. One can then apply the finite dimensional convolution theorem. The general result follows.

Remark 5. Theorem 3 says that, under its conditions, the experiment $\mathcal{F} = \{Q_\theta; \theta \in \Theta\}$ is also weaker than $\{P_\theta A; \theta \in \Theta\}$. That this need not be the case if the projection domination is removed can be seen on the following example. Take $\mathcal{X}$ to be the plane $\mathbb{R}^2$ with first coordinate called x and second called y. Let C be the circle $C = \{(x,y) : x^2 + y^2 = 1\}$. Place on C a measure P that has a density with respect to the Lebesgue measure of C. Take $\Theta = \mathbb{R}^2$. Let A be the projection on the second axis. One can choose P so that its projection on the second axis $\mathcal{Y}$ has density $[1 - |y|]^+$ with respect to the Lebesgue measure of $\mathcal{Y}$.

Let Q be the measure carried by $\mathcal{Y}$ that has density unity with respect to the Lebesgue measure on $[-1/2, +1/2]$. Let $Q_\theta = QS^{\theta A}$. The experiment $\mathcal{E} = \{PS^\theta; \theta \in \Theta\}$ is certainly stronger than $\{QS^{\theta A}; \theta \in \Theta\}$ since $\mathcal{E}$ is "perfect" in the sense that any two different PS^{θ_i} are disjoint. However Q is not obtainable as $(PA) * M$, since, for instance, the support of Q is $[-1/2, +1/2]$ while that of PA is $[-1, +1]$. Here, for good measure, we have taken P so that $(PA) = Q * Q$.

One could attempt to spread P by convolution, for instance with a probability measure m carried by the ball (centered at zero) of radius ϵ in $\mathbb{R}^2$. Take m to be proportional to Lebesgue measure on that ball. Carry a similar operation on Q convoluting it with a probability measure m' proportional to the Lebesgue measure on $[-\epsilon, +\epsilon]$.

Then, for ϵ small enough, the shift experiments based on $P*m$ and $Q*m'$ cannot be comparable. Indeed if one takes a $\theta \in \mathbb{R}^2$ with norm $|\theta| \leq 1/100$ and an ϵ that is much smaller, say $\epsilon = 10^{-4}$, the affinity between $Q * m'$ and $(Q * m')S^{\theta A}$ is close to unity but the affinity between $(P * m)$ and $(P * m)S^\theta$ is small. Thus, if the experiments were comparable $Q * m'$ would be obtained as a convolution of the marginal of $P * m$ with some probability measure (by Theorem 3). This cannot be as can be seen from the supports of the measures.

8. A theorem of P.W. Millar

It has been customary to state theorems such as Theorem 2 and especially

Theorem 3 for the case where Q is a Radon measure on the space $\mathcal{Y}$. For instance Millar [1985] imbeds all spaces in suitable completions so that the measures P and Q become Radon measures. This complicate matters some since if one completes $\mathcal{X}$ to say, $\tilde{\mathcal{X}}$, the shifts used are only the S^{θ}, $\theta \in \Theta = \mathcal{X} \subset \tilde{\mathcal{X}}$. Actually there is no need to perform such embeddings as will follow from Proposition 1 below. To avoid excessive notation we have stated Proposition 1 in a simplified form using only one linear space $\mathcal{X}$. In the spirit of Theorem 3 it can be used in the space $\mathcal{Y}$ of Theorem 3 replacing the P in the Proposition by PA.

Proposition 1. *Let P and M be cylinder probabilities on $\mathcal{X}$. Assume that $Q = P * M$ extends to a Radon measure on $\mathcal{X}$. Then there is a shift S^z in the algebraic dual $\mathcal{Z}$ of the dual $\mathcal{Y}$ of $\mathcal{X}$ such that PS^z and MS^{-z} are extendable to Radon measures on $\mathcal{X}$. If one of P or M is already extendable to a Radon measure on $\mathcal{X}$, so is the other.*

Note. The algebraic dual $\mathcal{Z}$ of the dual $\mathcal{Y}$ of $\mathcal{X}$ can be identified to the completion of $\mathcal{X}$ for the weak topology $w^(\mathcal{X}, \mathcal{Y})$. All the functions $\gamma \in D$ considered previously are $w(\mathcal{X}, \mathcal{Y})$ uniformly continuous. They extend continuously to $\mathcal{Z}$. A cylinder measure on $\mathcal{X}$ is also a cylinder measure on $\mathcal{Z}$.

Proof. According to a theorem of E. Mourier improved by Prohorov [1966] and then Schwartz [1973], a cylinder probability Q on $\mathcal{X}$ extends to a Radon measure on $\mathcal{X}$ if and only if it has the following property. Take an $\epsilon \in (0, 1/5)$. Then there is a sequence $\{C_n\}$ of compact symmetric subsets of $\mathcal{X}$ such that for each $F \in \mathcal{C}$ the image of C_n by $\Pi(\mathcal{X}, \mathcal{X}/F)$ has measure at least $1 - \epsilon^{2n}$ for the projection measure $Q\Pi(\mathcal{X}, \mathcal{X}/F)$.

Now let us look at a fixed $F \in \mathcal{C}$ and the projected measures p_F, m_F and q_F so that $p_F * m_F = q_F$. The sets C_n project onto compacts $c_{n,F}$.

By assumption on Q one has $q_F(c_{n,F}) \geq 1 - \epsilon^{2n}$. Thus by the convolution formula on $\mathcal{X}/f$ there are elements y of $\mathcal{X}/f$ such that $p_F(c_{n,F} - y) \geq 1 - \epsilon^{2n}$. These form a certain set $A_{n,F}$ that is a compact set and the measure $m(A_{n,F})$ is at least $1 - \epsilon^n$. It follows that $B_F = \cap_n A_{n,F}$ is not empty and has measure at least $(1 - \frac{\epsilon}{1-\epsilon}) \geq 1/4$, since $\epsilon \leq 1/5$. Thus if P is shifted by any element y of B_F the shifted measure $p_F S^y$ projection of PS^y will give mass at least $1 - \epsilon^{2n}$ to $c_{n,F}$ for all n.

Now look at $G \subset F$, $G \in \mathcal{C}$. If $p_G(c_{n,G} - y) \geq 1 - \epsilon^{2n}$ then one will have $p_F((c_{n,G} - y)\Pi) \geq 1 - \epsilon^{2n}$ for the projection Π of $\mathcal{X}/G$ onto $\mathcal{X}/F$. This shows that B_G projects into a subset of B_F. All these sets are compact. Thus, as $G \subset F$ runs through $\mathcal{C}$, the projections of B_G into B_F have a non empty intersection. In other words, there are elements $y_F \in B_F$, $F \in \mathcal{C}$ such that $y_F = y_G \Pi(\mathcal{X}/G, \mathcal{X}/F)$ and such that, shifted by S^{y_F} the image of P gives mass at least $1 - \epsilon^{2n}$ to the projection of C_n. Such a system $\{y_F, f \in \mathcal{C}\}$ may or may not arise from an element in $\mathcal{X}$, but it obviously defines an element z of $\mathcal{Z}$.

Thus we have $Q = (PS^z) * (MS^{-z})$ and PS^z is extendable to a Radon measure. By the same argument there is some other shift $t \in \mathcal{Z}$ such that MS^{-t} extends to a Radon measure. Then $(PS^z) * (MS^{-t}) = (P * M)S^{z-t}$ is also extendable to a Radon measure on $\mathcal{X}$. Since $Q = P * M$ is also extendable, $z - t$, must belong to $\mathcal{X}$ and $MS^{-z} = (MS^{-t}S^{(t-z)}$ is also extendable to a Radon measure. Hence the result. □

Note that we cannot claim that the shift z is always in $\mathcal{X}$. That is because any shift S^z for $z \in \mathcal{Z}$ still yields cylinder measures PS^z and MS^{-z} with convolution

$$(PS^z) * (MS^{-z}) = P * M.$$

Proposition 1, above, is just another expression of Paul Lévy's principle: A convolution $P * M$ is less concentrated than each of P and M.

As mentioned earlier, Millar's theorem was only for the case where the experiment $\mathcal{E} = \{PS^\theta; \theta \in \Theta\}$ is obtained by shifting a Gaussian measure. In such a case one can use special arguments as sketched in Remark 4, Section 7. Another theorem, similar to Theorem 3, Section 7, has been given by van der Vaart [1991] for a non-Gaussian $\mathcal{E}$ formed by shifting product measures. By application of Proposition 1 these results become consequences of Theorem 3, Section 7.

9. Final remarks

It was already observed in LeCam [1972] that there are cases other than the LAN ones in which one can use the combination of the results of Sections 3 to 5. This happens for instance when one shifts densities f that have certain singularities. However, in many cases the limit experiments obtained in this fashion are not of the type $\mathcal{E} = \{PS^\theta; \theta \in \Theta\}$ where P is a cylinder measure on the space $\mathcal{X} = \Theta$ itself. For instance limits of triangular arrays of product measures, as described in LeCam [1986], are infinitely divisible and therefore products of a Gaussian experiment by a Poisson experiment. Poisson experiments are obtained by using a ring $\mathcal{A}$ of subsets of a suitable space and observing for each $A \in \mathcal{A}$ an ordinary Poisson variable $N(A)$ with expectation $\lambda(A)$. For disjoint systems $A_1, A_2, \ldots, A_m$ with $A_j \in \mathcal{A}$, the Poisson variables $N(A_j)$ are mutually independent. The experiment is parametrized by the intensity measure λ allowed to vary in a set Θ. Such experiments are not invariant under the "natural shifts" which would consist of adding another measure to the intensity λ. To see what happens one can look at the examples given by Prakasa Rao [1968] and by Ibragimov and Has'minskii [1981].

However, in many such cases, as in the LAMN cases of Jeganathan [1981], one can still obtain transitions that are *conditionally* representable by convolutions. See LeCam [1972].

The situation is then reminiscent of the one described in Section 5 for bands that are stable but not irreducible.

There are however many non-Gaussian situations where one obtains experiments $\mathcal{E}$ of the form $\mathcal{E} = \{PS^\theta; \theta \in \Theta\}$ with P a cylinder measure on $\mathcal{X} = \Theta$ as required for Theorem 3, Section 7, and where in addition one has distinguished sequences of statistics as in Section 3. One can, for instance, consider experiments $\mathcal{E}_n$ that tend to a limit $\mathcal{E}$ given by an exponential family that is shift invariant. See for instance LeCam [1986], Chapter 8, Section 5 and in particular Proposition 7 and the work by LeCam [1975]. The experiment of Proposition 7, Chapter 8 Section 5 of LeCam [1986] is a limit experiment obtainable by looking at extreme values of a sequence of i.i.d. random variables.

A multitude of other cases can be obtained from the stable processes studied by C. Hesse [1991]. There the distinguished statistics may have to be infinite dimensional.

Note that Theorem 2 and 3 apply to the abelian shift group of a linear space $\mathcal{X}$. It seems possible to extend the proof to other cases where the space $\mathcal{X}$ is a projective limit of solvable locally compact groups and for "cylinder measures" defined appropriately. However, one should beware of the fact that convolution theorems apply only to special situations. Even though they give at once information about a variety of loss functions, they need a particular structure. Such structure is noticeably absent in the asymptotic minimax theorem (see LeCam [1986], page 109) which can often be used directly to obtain a variety of results.

Acknowledgement. The author wishes to extend his heartfelt thanks to the anonymous referee who suggested that Theorem 2 be rewritten to become Theorem 3, Section 7.

References

[1] Berk, R. and Bickel, P. [1968]. "On invariance and almost invariance." *Ann. Math. Stat.* **39**, 1573-1576.

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

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

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

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

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

[7] Daniell, P.J. [1918]. "Integrals in an infinite number of dimensions." *Ann. of Math.* **20**, 281-288.

[8] Dieudonné, J. [1951]. "Sur le théorème de Lebesgue-Nikodym, IV: *J. Indian Math. Soc. NS* **15**, 77-86.

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

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

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

[12] Helson, H. [1954]. "Isomorphisms of abelian group algebras." *Arkiv för Matematik* **2**, 475-487.

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

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

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

[16] 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.

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

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

[19] 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.

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

[21] Lehmann, E.L. [1988]. "Comparing location experiments." *Ann. Statist.* **16**, 521-533.

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

[23] Mourier, E. [1953]. "Eléments aléatoires dans un espace de Banach." *Ann. Inst. H. Poincaré* **13**, 161-244.

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

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

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

[27] Prohorov, Yu. V. [1961]. "The method of characteristic functionals." *Proc. 4th Berkeley Symp. Math. Statist. Probab.* I, 403-419.

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

[29] Rubin, H. [1967]. "Supports of convolutions of identical distributions". *Proc. 5th Berkeley Symp. Math. Statist. Probab.* II, 415-422.

[30] Schwartz, L. [1951]. *Théorie des distributions* vol. II. Hermann, Paris.

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

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

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

[31] van der Vaart, A. [1991]. "An asymptotic representation theorem."

Internat. Statist. Review **58**, 97-121.