SOME THEOREMS ON THE DISTRIBUTION OF SUMS OF INDEPENDENT LINEAR STOCHASTIC PROCESSES *

By

L. Le Cam

University of California, Berkeley, California

1. Introduction. The present paper originated in an attempt to organize and simplify some of the known results relative to the asymptotic behavior of empirical cumulative distribution functions. Such empirical cumulative distributions can conveniently be regarded as sums of independent random variables whose values lie in suitable infinite dimensional spaces. It is therefore natural to attempt to apply to this situation some of the methods which have proved so successful in handling similar problems for finite dimensional variables.

A particularly important result for the finite dimensional case is Kolmogorov's inequality on dispersion of sums. Unfortunately this inequality does not extend to the infinite dimensional case, although we shall give here a weaker result which is still valid in general vector spaces.

After describing what will be meant by a linear stochastic process in section 2, we give in section 3 a proof of the inequality on dispersion of sums referred to above. Section 4 is devoted to possible applications.

2. Linear processes. Let $\mathcal{Y}$ be a real vector space and let $(\Omega, \mathcal{A}, P)$ be a probability space. Let $\mathcal{V}$ be the space of all real-valued measurable functions on $(\Omega, \mathcal{A}, P)$. We shall be concerned here with entities which can conveniently be described as linear maps from the space $\mathcal{Y}$ to a space such as $\mathcal{V}$. Another possible description is that the processes considered here, and called thereafter "linear processes," are families $\{\langle X, y\rangle ; y \in \mathcal{Y}\}$ of real-valued random variables subject to the restriction that for every pair (y_1, y_2) of elements of $\mathcal{Y}$ and every pair (α_1, α_2) of real numbers

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

Since, however, we shall be concerned mostly with distribution problems, it will be more convenient to ignore the basic space $\{\Omega, \mathcal{A}, P\}$ and reconstruct special probability spaces whenever necessary.

All computations of probabilities will be made from the joint distributions of finite sets $\{\langle X, y_j\rangle; y_j \in \mathcal{Y}\}$ as is usual in the theory of stochastic processes. In this respect the first theorem needed is a theorem of Bochner which can be formulated as follows.

Let Ω be the space of all linear functionals on the vector space $\mathcal{Y}$. Let $\mathcal{A}$ be the smallest σ-field of subsets of Ω with respect to which all the functions $\omega \longrightarrow \langle \omega, y\rangle$, $y \in \mathcal{Y}$ are measurable.

Theorem 1 (Bochner). Let φ be a complex-valued function defined on $\mathcal{Y}$. In order that there exist a probability measure P on $\{\Omega, \mathcal{A}\}$ such that

$$\varphi(y) = \int e^{i\langle \omega, y\rangle} P(d\omega)$$

for every $y \in \mathcal{Y}$, it is necessary and sufficient that

1) $\varphi(0) = 1.$

2) $\alpha \longrightarrow \varphi(\alpha y)$ <u>is for each</u> y <u>continuous in the real variable</u> α.

3) <u>If</u> c <u>is a complex-valued function defined on</u> $\mathcal{Y}$ <u>and differing from zero only on a finite subset of</u> $\mathcal{Y}$ <u>then</u>

$$\sum_{y} \sum_{y'} c(y)\overline{c(y')}\ \varphi(y-y') \geq 0.$$

<u>If these conditions are satisfied the corresponding</u> P <u>is uniquely determined on</u> $\mathcal{A}$ <u>by the function</u> φ.

The family $\{\langle\omega, y\rangle;\ y \in \mathcal{Y}\}$ described in the theorem is a linear process. Further, according to the theorem, every linear process admits of such a representation. If X is a linear process on the space $\mathcal{Y}$, the characteristic function $\varphi(y) = E\ e^{i\langle X,y\rangle}$ defines a measure P on the space $[\Omega, \mathcal{A}]$ of theorem 1. This measure P will be called the distribution of the linear process X.

Consider now the space Φ_0 of complex functions defined on Ω by expressions of the type

$$f(\omega) = \sum_{y} \alpha(y)e^{i\langle\omega,y\rangle}$$

where α is a complex function vanishing outside of a finite subset of $\mathcal{Y}$. Let $\overline{\Phi}_0$ be the space of functions which are uniform limits of functions of Φ_0. Further let Φ be the subspace of $\overline{\Phi}_0$ formed of <u>real</u> functions. Clearly Φ is a uniformly complete algebra containing the constant functions. Hence Φ is also a lattice for the usual point by point operations.

The distribution P of a linear process X induces on Φ a positive linear functional E according to the formula

$$Ef = Ef(X) = \int f(\omega)P(d\omega).$$

The linear functional E will be called the expectation attached to X.

Since the measure P is defined only on a very small σ-field it is usually necessary or convenient to extend its domain of definition. Also the space Ω is often much too large for convenient handling. For these reasons we shall restrict our attention to processes X which satisfy special properties of tightness as defined below. In this definition the letters Ω, E and Φ need not have the meaning associated to them above.

Definition. Let E be a linear functional on a vector lattice Φ of bounded numerical functions on a set Ω. Let $\mathcal{F}$ be a family of subsets of Ω. Suppose that for every $\varepsilon > 0$ there is a $\delta > 0$ and an $F \in \mathcal{F}$ such that the inequalities

1) $\sup\{|f(x)|;\ x \in \Omega\} \leq 1$,

2) $\sup\{|f(x)|;\ x \in F\} \leq \delta$,

imply $|Ef| \leq \varepsilon$. Then E will be called $\mathcal{F}$-tight on Φ.

For obvious reasons we shall limit our considerations to directed families $\mathcal{F}$ such that $F_i \in \mathcal{F}$; $i=1,2$, implies the existence of an $F_3 \in \mathcal{F}$ for which $F_1 \cup F_2 \subset F_3$. It is clear that the notion of $\mathcal{F}$-tightness is not altered if to $\mathcal{F}$ one adds all the subsets of its elements, so that one may assume that $G \subset F$ and $F \in \mathcal{F}$ implies $G \in \mathcal{F}$. A family $\mathcal{F}$ satisfying these supplementary restrictions will be called a cofilter.

The property of $\mathcal{F}$-tightness is evidently a property of continuity. In fact, let T_0 on Φ be the topology of uniform convergence on the elements of $\mathcal{F}$. Let T_1 be the strongest locally convex topology which coincides with T_0 on the uniformly bounded subsets of Φ. To say that E is $\mathcal{F}$-tight is to say that E is T_1 continuous on Φ.

The space Ω involved in the definition of $\mathcal{F}$-tightness plays a relatively inessential role. For instance let $\hat{\Omega}$ be the completion of Ω for the uniform structure induced by Φ on Ω. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two directed families of subsets of Ω. Let $\mathcal{F}_i' = \{\overline{F}; F \in \mathcal{F}_i\}$ where $\overline{F}$ is the closure of F in $\hat{\Omega}$ and let $\mathcal{F}_i''$ be the cofilter generated by $\mathcal{F}_i'$. If $\mathcal{F}_1'' = \mathcal{F}_2''$ the notions of tightness for $\mathcal{F}_1$ and $\mathcal{F}_2$ are equivalent and equivalent to the notions of $\mathcal{F}_i'$ or $\mathcal{F}_i''$-tightness.

Consider now a vector space $\mathcal{Y}$ with algebraic dual Ω and let $\mathcal{X}$ be a vector subspace of Ω. Let X be a linear process over $\mathcal{Y}$. If there is no $y \in \mathcal{Y}$ such that $y \neq 0$ but $\langle x,y\rangle = 0$ for every $x \in \mathcal{X}$ we shall say that $(\mathcal{X},\mathcal{Y})$ is a dual system. For such a dual system we shall denote by $w(\mathcal{X},\mathcal{Y})$ the weak topology induced by $\mathcal{Y}$ on $\mathcal{X}$ and by $\tau(\mathcal{X},\mathcal{Y})$ the strongest locally convex topology for which the dual of $\mathcal{X}$ is $\mathcal{Y}$.

Even if $\mathcal{X}$ does not separate $\mathcal{Y}$ there are always locally convex topologies $\mathcal{C}$ on $\mathcal{Y}$ for which the dual of $(\mathcal{Y},\mathcal{C})$ is $\mathcal{X}$ but these topologies are not necessarily separated.

Consider then a subspace $\mathcal{X}$ of $\mathcal{Y}$ and let $\mathcal{K}$ be the family of all $w(\mathcal{X},\mathcal{Y})$ compact convex symmetric subsets of $\mathcal{X}$. <u>We shall restrict our considerations to processes</u> X <u>whose expectations are</u> $\mathcal{K}$-<u>tight on the family</u> Ψ <u>of restrictions to</u> $\mathcal{X}$ <u>of the elements of the space</u> Φ <u>defined above</u>.

Such an expectation E can be extended to a larger space J as follows. Let B be a subset of Ψ which is <u>bounded</u> for the uniform norm on $\mathcal{X}$. Let $\mathcal{N}$ be the space of all numerical functions on $\mathcal{X}$. Let $\bar{B}$ be the closure of B in $\mathcal{N}$ for the structure of uniform convergence on the elements of $\mathcal{K}$. Let $\bar{\Psi}$ be the union of all the sets $\bar{B}$ obtained in this fashion. Clearly E possesses an extension by continuity to the whole of $\bar{\Psi}$ and this extension is still $\mathcal{K}$-tight on $\bar{\Psi}$. The extension $\bar{E}$ so obtained can now be extended further by the Mac-Shane-Bourbaki procedure.

We shall denote by the letter J the space of all bounded numerical functions which are integrable in the Mac-Shane-Bourbaki sense for every $\mathcal{K}$-tight expectation on $\bar{\Psi}$. The extension of E to J will be denoted $\tilde{E}$ and called the Radon extension of E.

If F is a $w(\mathcal{X}, \mathcal{Y})$ closed subset of $\mathcal{X}$ then its indicator I_F belongs to J. The expectation $\tilde{E}(I_F)$ will be called the probability of F and denoted $\tilde{P}(F)$. These probabilities always satisfy the following "separability" requirement.

<u>If</u> K <u>is a</u> $w(\mathcal{X}, \mathcal{Y})$ <u>closed symmetric convex subset of</u> $\mathcal{X}$ <u>and</u> K^0 <u>is the polar</u> $K^0 = \{y: y \in \mathcal{Y};\ \sup[|\langle x,y\rangle|;\ x \in K] \leq 1\}$, <u>then there is a countable subset</u> D <u>of</u> K^0 <u>such that</u>

$$\tilde{P}(K) = P\left\{\sup_{y \in D} |\langle X,\ y\rangle| \leq 1\right\}.$$

To recognize tight expectations, or by abuse of language tight linear processes, one may occasionally use the following criterion.

<u>Let</u> $\mathcal{T}$ <u>be a locally convex topology on</u> $\mathcal{Y}$ <u>for which the dual of</u> $\mathcal{Y}$ <u>is</u> $\mathcal{X}$. <u>Let</u> $\mathcal{F}$ <u>be the family of equicontinuous subsets of</u> $\mathcal{X}$. <u>A linear process</u> X <u>over</u> $\mathcal{Y}$ <u>is</u> $\mathcal{F}$<u>-tight on</u> Ψ <u>if and only if for every</u> $\varepsilon > 0$ <u>there is a</u> $\mathcal{T}$<u>-neighborhood</u> V <u>of the origin of</u> $\mathcal{Y}$ <u>such that</u>

$$P\left\{\sup_{y\in F}|\langle X,y\rangle| \leq 1\right\} \geq 1 - \varepsilon$$

<u>for every finite subset</u> F <u>of</u> V.

When the topology $\mathcal{C}$ of $\mathcal{Y}$ is metrisable, the statement that X is $\mathcal{F}$-tight is equivalent to the statement that E is σ-smooth on $\{\Psi, \mathfrak{X}\}$. Also it is equivalent to the statement that the P outer measure of $\mathcal{X}$ in Ω is unity.

Note that $\mathcal{K}$-tightness is nothing but $\mathcal{F}$-tightness for the family $\mathcal{F}$ of sets which are $\tau\{\mathcal{Y},\mathcal{X}\}$ equicontinuous.

Suppose now that $\tilde{P}$ is $\mathcal{K}$-tight on Ψ, and let H be the intersection of all the $w[\mathcal{X},\mathcal{Y}]$ closed convex symmetric sets $C \subset \mathcal{X}$ such that $\tilde{P}(C) = 1$. Then $\tilde{P}(H) = 1$. A similar result holds for $w(\mathcal{X},\mathcal{Y})$ closed linear subspaces of $\mathcal{X}$. In this case H is simply the polar of the set of elements $y \in \mathcal{Y}$ such that $\int e^{i\alpha\langle x,y\rangle} P(dx) = \varphi(\alpha y) = 1$ for every real number α. This last set is also the polar of H.

Probability measures which are $\mathcal{K}$-tight on $\mathcal{X}$ are already restricted enough to behave in a tractable manner in many problems. In fact, in spite of the great apparent generality inherent in the arbitrariness of dual systems $\{\mathcal{X}, \mathcal{Y}\}$, the study of $\mathcal{K}$-tight measures can for most purposes be carried out as if $\mathcal{Y}$ was a Frechet space of dual $\mathcal{X}$.

This can be seen as follows. If P is $\mathcal{K}$-tight on $\mathcal{X}$ there is an increasing sequence $\{K_n\}$ of $w(\mathcal{X},\mathcal{Y})$ compact convex symmetric subsets of $\mathcal{X}$ such that $\tilde{P}\{[\bigcup_n K_n]^c\} = 0$. If H is the smallest closed linear subspace of $\mathcal{X}$ for which $\tilde{P}(H) = 1$ one may replace K_n by $K_n' = K_n \cap H$ without essential change. Consider then the space $\mathcal{Z}$ generated by the sequence $\{K_n \cap H\}$. Taking a quotient of $\mathcal{Y}$ if necessary one may assume that $\mathcal{Z}$ separates the points of $\mathcal{Y}$. Topologize $\mathcal{Y}$ by the topology of uniform convergence on the sets K_n' and complete the space so obtained. This gives a Frechet space $\bar{\mathcal{Y}}$. The space $\bar{\mathcal{Y}}$ is precisely the space of linear functionals defined on $\mathcal{Z}$ whose restrictions to the compacts K_n' are continuous. Further, a linear functional which is continuous on $\bar{\mathcal{Y}}$ is an element of $\mathcal{Z}$ so that $\mathcal{Z}$ is the dual of the Frechet space $\bar{\mathcal{Y}}$.

One can even go somewhat further and show that each one of the compacts $K_n' \in \mathcal{Z}$ can be taken to be the closure of the union $A_n = \bigcup_k A_{n,k}$ of a countable family $\{A_{n,k};\ k=1,2,\cdots\}$ of $\tau(\mathcal{Z},\bar{\mathcal{Y}})$ compact convex symmetric subsets of $\mathcal{Z}$.

To show this consider for each integer n the measure μ_n truncation of $\tilde{P}$ to K_n' defined by $\mu_n(S) = \tilde{P}[S \cap K_n']$. There is a smallest $w[\mathcal{Z},\mathcal{Y}]$ compact convex symmetric subset B_n of K_n' such that $\mu_n(B_n) = \|\mu_n\| = \tilde{P}(K_n')$. We shall now show that B_n is the closure of a union $\bigcup_k A_{n,k}$ of $\tau(\mathcal{Z},\bar{\mathcal{Y}})$ compact sets. Let M_n be the set of y's $y \in \bar{\mathcal{Y}}$ which vanish on B_n. Let $\mathcal{Z}_n$ be the subspace of $\mathcal{Z}$ spanned by B_n. The topology $\tau[\mathcal{Z}_n, \bar{\mathcal{Y}}/M_n]$ induced on $\mathcal{Z}_n$ by the quotient space $\bar{\mathcal{Y}}/M_n$ is stronger than $\tau[\mathcal{Z},\bar{\mathcal{Y}}]$.

It is therefore sufficient to consider the case where $\bar{\mathcal{Y}}$ is a Banach space of dual $\mathcal{Z}$ with unit ball B such that B is the smallest $w(\mathcal{Z},\bar{\mathcal{Y}})$ compact symmetric convex set for which $\tilde{P}(B) = 1$. Let L be the space of $\tilde{P}$ integrable functions for the Radon measure $\tilde{P}$ on the compact B. Let S be the unit ball of L for the norm

$$\|f\| = \int |f(x)|\tilde{P}(dx)$$

and let S_n be the subset of S consisting of functions f such that $|f(x)| \leqq n$ for each $x \in B$. To each $f \in L$ associate a "center of gravity" $G(f)$ in $\mathcal{Z}$ by the formula

$$G(f) = \int f(x) x \tilde{P}(dx).$$

The map $f \longrightarrow G(f)$ is a continuous map from the Banach space L to $\mathcal{Z}$ topologized by $\tau[\mathcal{Z}, \tilde{\mathcal{Y}}]$ and the set S is mapped into B. Therefore, according to the Grothendieck form of a theorem of Dunford and Pettis [1] the image $G(S_n)$ of S_n by G is a $\tau[\mathcal{Z}, \tilde{\mathcal{Y}}]$ compact subset of B. Let A_n be the closure of $G(S_n)$ in $\mathcal{Z}$. Then A_n is a symmetric convex $\tau[\mathcal{Z}, \tilde{\mathcal{Y}}]$ compact subset of $\mathcal{Z}$. Further, $A_n \subset A_{n+k} \subset \frac{n+k}{n} A_n$. Let $A = \bigcup_n A_n$ and let $\bar{A}$ be the closure of the convex symmetric set A.

If C is a convex subset of $\mathcal{Z}$ whose closure $\bar{C}$ does not intersect A_n, we must have $\tilde{P}(\bar{C}) < \frac{1}{n}$. If not there is an $f \geqq 0$, $f \in S$ such that $f(x) = 0$ for $x \in A_n$ and $f(x) = [\tilde{P}(\bar{C})]^{-1}$ for $x \in \bar{C}$. For this f we must have $G(f) \in B \cap \bar{C}$ and $G(f) \in A_n$. This is impossible so that $\tilde{P}(\bar{C}) < 1/n$. Similarly, if $\tilde{P}(\bar{A}) < 1$ there is a $y \in \tilde{\mathcal{Y}}$ such that $|\langle z,y \rangle| \leq 1$ for $z \in \bar{A}$ and such that $\alpha = \tilde{P}\{z: \langle z, y \rangle > 1\} > 0$. Taking a function f equal to α^{-1} on $\{z: \langle z, y \rangle > 1\}$ and zero otherwise leads to a contradiction. Since B was by assumption the smallest $w(\mathcal{Z}, \tilde{\mathcal{Y}})$ compact convex symmetric subset of $\mathcal{Z}$ such that $\tilde{P}(B) = 1$ we must have $B \subset \bar{A}$ hence $B = \bar{A}$.

One cannot conclude from this that $P(A) = 1$ as can be seen from the following example. Let $\mathcal{Y}$ be the space of continuous functions on the interval $T = [0, 1]$. The space $\mathcal{Y}$ is a Banach space for the uniform norm and its dual $\mathcal{Z}$ is the space of Radon integrals on T.

For each $t \in T$ let δ_t be the probability measure giving mass unity to the point t. Let X be the linear process defined by $\langle X, y\rangle = \langle \delta_t, y\rangle = y(t)$ where t is taken at random according to the Lebesgue measure λ on T. The positive part B^+ of the unit ball B of $\mathcal{Z}$ is a $w(\mathcal{Z}, \mathcal{Y})$ compact convex set such that $\tilde{P}(B^+) = 1$. Further, let A_n be the set of signed measures μ on T which are such that there is a bounded measurable function f, $|f| \leqq n$ for which $\langle \mu, y\rangle = \langle \lambda, fy\rangle$ for every $y \in \mathcal{Y}$. Then B is the closure (for $\tau(\mathcal{Z}, \mathcal{Y})$) of $\bigcup_n A_n$; however every $\tau(\mathcal{Z}, \mathcal{Y})$ convex compact subset of B has measure zero for P.

Such a situation cannot occur if the topology $\tau(\mathcal{Z}, \mathcal{Y})$ is metrisable. This can be easily derived from the above results or proved directly as follows. Suppose that $\mathcal{Z}$ possesses a metrisable locally convex topology for which its dual is $\mathcal{Y}$ (then this topology is $\tau(\mathcal{Z}, \mathcal{Y})$). Consider the characteristic function φ defined by $\varphi(y) = E\, e^{i\langle X, y\rangle}$. An immediate application of Šmulian's theorem shows that φ is $w(\mathcal{Y}, \mathcal{Z})$ continuous on the $w(\mathcal{Y}, \mathcal{Z})$ compacts of $\mathcal{Y}$.

Thus φ is also continuous for the topology of uniform convergence on the $\tau(\mathcal{Z}, \mathcal{Y})$ precompact subsets of $\mathcal{Z}$. Since these precompact sets have dense countable subsets it follows that $\tilde{P}$ is carried by a separable subset of $\mathcal{Z}$ hence is $\mathcal{F}$-tight for the family $\mathcal{F}$ of precompact subsets of $\mathcal{Z}$.

In particular, let $\tilde{P}$ be a $\mathcal{K}$-tight probability measure on a dual system $\{\mathcal{X}, \mathcal{Y}\}$. Let T be a linear map from $\mathcal{X}$ to a Frechet space $\mathcal{Z}$. Assume that T is continuous for the $\tau(\mathcal{X}, \mathcal{Y})$ topology of $\mathcal{X}$ and the metric topology of $\mathcal{Z}$. Then the image of $\tilde{P}$ by A is $\mathcal{F}$-tight on $\mathcal{Z}$ for the family $\mathcal{F}$ of strongly compact subsets of $\mathcal{Z}$.

The converse proposition is also true when $\mathcal{X}$ is the dual of a Frechet space $\mathcal{Y}$. Explicitly, assume that $\mathcal{Y}$ is a Frechet space of dual $\mathcal{X}$ and let X be a linear process on $\mathcal{Y}$ such that for every $\tau(\mathcal{X}, \mathcal{Y})$ continuous linear map T of $\mathcal{X}$ into a Frechet space $\mathcal{Z}$ of dual $\mathcal{Z}'$ the process TX is $\mathcal{F}$-tight for the family $\mathcal{F}$ of $w(\mathcal{Z}, \mathcal{Z}')$ compact of $\mathcal{Z}$ then X itself is $\mathcal{K}$-tight on $\mathcal{X}$.

In this statement TX is supposed to be defined by the equation $\langle TX, z' \rangle = \langle X, T'z' \rangle$ where T' is the transpose of T. Since T is $\tau(\mathcal{X}, \mathcal{Y})$ continuous T' maps $\mathcal{Z}'$ into $\mathcal{Y}$.

In many problems one is led to consider the following type of situation. For a given linear space $\mathcal{Y}$ one selects a locally convex topology $\mathcal{T}$ on $\mathcal{Y}$ for which the dual of $\{\mathcal{Y}, \mathcal{T}\}$ is a space $\mathcal{X}$. The topology $\mathcal{T}$ need not be separated. If X is a linear process over $\mathcal{Y}$ it may happen that X is not only $\mathcal{K}$-tight but that X is also $\mathcal{F}$-tight on the space Ψ for the family $\mathcal{F}$ of $\mathcal{T}$-equicontinuous subsets of $\mathcal{X}$. Since $\mathcal{K}$-tightness is equivalent to $\mathcal{F}$-tightness for the family $\mathcal{F}$ of sets which are $\tau[\mathcal{Y}, \mathcal{X}]$ equicontinuous a restriction of $\mathcal{F}$-tightness for a system $\{\mathcal{X}, \mathcal{Y}, \mathcal{T}\}$ is a stronger restriction than $\mathcal{K}$-tightness. The theorems stated above for $\mathcal{K}$-tight processes apply also to $\mathcal{F}$-tight processes in an obvious way.

In the sequel we shall almost always work with processes which are $\mathcal{F}$-tight on Ψ for the family $\mathcal{F}$ of equicontinuous subsets of $\mathcal{X}$ for a topology $\mathcal{C}$.

Let then X_1 and X_2 be two $\mathcal{F}$-tight linear processes. Their sum $X_1 + X_2$ is also $\mathcal{F}$-tight.

Further, call the processes X_1 and X_2 independent if for every finite set $F \in \mathcal{Y}$ the sets $\{\langle X_1, y\rangle; y \in F\}$ and $\{\langle X_2, y\rangle; y \in F\}$ are independent. It is equivalent to say that for $\varphi_j = E\, e^{i\langle X_j, y\rangle}$ we have

$$E\, e^{i\alpha\langle X_1, y\rangle + i\beta\langle X_2, y\rangle} = \varphi_1(\alpha y)\varphi_2(\beta y)$$

identically in $y \in \mathcal{Y}$ and the real variables α and β.

If X_j, $j=1,2$ are $\mathcal{F}$-tight processes and $X_3 = X_1 + X_2$ the Radon extensions $\tilde{P}_j$ of the distributions of the processes X_j satisfy the convolution theorem in the sense that

$$\int f(z)\tilde{P}_3(dz) = \int \tilde{P}_1(dx)\int f(x+y)\tilde{P}_2(dy)$$

$$= \int \tilde{P}_2(dy)\int f(x+y)\tilde{P}_1(dx)$$

for every f in the space J defined previously.

This representation of $\tilde{P}_3$ will be used frequently in the sequel.

3. <u>The concentration of sums</u>. Consider a system $\{\mathcal{X}, \mathcal{Y}, \mathcal{C}\}$ formed by a linear space $\mathcal{Y}$ carrying a locally convex topology $\mathcal{C}$ for which the dual of $\mathcal{Y}$ is $\mathcal{X}$. Let $\{x_j; j=1,2,\cdots,n\}$ be a finite sequence of elements of $\mathcal{X}$. Let K be a $w(\mathcal{X}, \mathcal{Y})$ closed

convex symmetric subset of $\mathcal{X}$. An inequality on the concentration of sums of independent random variables can be obtained if we can first derive inequalities for the number of sums of the type $\Sigma(\pm x_j)$ which belong to K. For this purpose we shall recall the following lemma, due to Sperner and Erdős [2].

Let S be a finite set and let $\mathcal{S}$ be a family of subsets of S. For every set F let $\kappa(F)$ be its cardinality. For a given integer $r \geq 1$ let us say that $\mathcal{S}$ has property π_r if for every pair F_i; $i=1,2$ of elements of $\mathcal{S}$ the inclusion $F_1 \subset F_2$ implies that $\kappa[F_2 \cap F_1^c] < r$.

Lemma (Sperner-Erdős). *Let $\mathcal{S}$ be a family of distinct subsets of a set S. Suppose that $\mathcal{S}$ has property π_r. Then*

$$\kappa(\mathcal{S}) \leq \Sigma \binom{n}{j}; \quad \frac{n-r}{2} \leq j < \frac{n+r}{2}$$

with $n = \kappa(S)$.

In particular if the elements of $\mathcal{S}$ are not comparable then

$$\kappa(\mathcal{S}) \leq \binom{n}{j} \text{ with } \frac{n-1}{2} \leq j < \frac{n+1}{2}$$

hence

$$\kappa(\mathcal{S}) \leq \frac{1}{\sqrt{n+1}} 2^n.$$

Returning to the n points $\{x_j; j=1,2,\cdots,n\}$ in the space $\mathcal{X}$, let K be a closed convex symmetric subset of $\mathcal{X}$. Consider sequences $\varepsilon = \{\varepsilon_j; j=1,2,\cdots,n\}$ such that $\varepsilon_j = \pm 1$. Let $S(\varepsilon) = \Sigma_j \varepsilon_j x_j$. For a given $x \in \mathcal{X}$ let E_x be the set of sequences ε such that

$$S(\varepsilon) - x \in K.$$

If $\varepsilon \in E_x$ and k is an integer, $k=1,2,\cdots,n$ let $\varepsilon(k)$ be the sequence obtained from ε by reversing the sign of ε_k. Explicitly, if

$$\varepsilon = \{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_{k-1},\varepsilon_k,\varepsilon_{k+1},\cdots,\varepsilon_n\}$$

then

$$\varepsilon(k) = \{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_{k-1},-\varepsilon_k,\varepsilon_{k+1},\cdots,\varepsilon_n\}.$$

This gives $S(\varepsilon) - S[\varepsilon(k)] = 2\varepsilon_k x_k$. Therefore $S[\varepsilon(k)]$ will be an element of $x + K$ only if $x_k \in K$.

For a particular $\varepsilon' \in E_x$ and a particular integer k one obtains a sequence $y = \varepsilon'(k)$. This y may be obtainable from other sequences, say ε^r, $r=2,3,\cdots,\nu$ with $\varepsilon^r \in E_x$. Note that $\varepsilon^r(j) = \varepsilon^s(j)$ implies $\varepsilon^r = \varepsilon^s$ so that, if $y = \varepsilon^r(j_r)$ the correspondence between r and j_r is necessarily one to one.

For $\varepsilon \in E_x$, let $\nu(x;\ \varepsilon,k)$ be the number of sequences $\varepsilon' \in E_x$ such that $\varepsilon(k) = \varepsilon'(j)$ for some j. Let $\nu = \nu(x) = \max\{\nu(x;\ \varepsilon,k)\}$ the maximum being taken over all $\varepsilon \in E_x$ and all integers k.

<u>Since each</u> $\varepsilon \in E_x$ <u>provides</u> n <u>different sequences</u> $\varepsilon(k)$; $k=1,2,\cdots,n$ <u>and since each such sequence may originate from at most</u> ν <u>elements of</u> E_x <u>the number of distinct sequences formed from</u> E_x <u>is at least equal to</u> $(n/\nu)\,\kappa(E_x)$.

<u>Therefore if</u> $x_k \notin K$ <u>and if</u> $x + K$ <u>contains</u> m <u>points</u> $S(\varepsilon)$ <u>the complement of</u> $x + K$ <u>contains at least</u> nm/ν <u>points</u> $S(\varepsilon)$.

<u>Lemma</u> 3. <u>Let</u> $\{x_j;\ j=1,2,\cdots,n\}$ <u>be</u> n <u>points of</u> $\mathfrak{X}$. <u>Assume</u> $x_j \notin K$ <u>and let</u> $\xi_j;\ j=1,2,\cdots,n$ <u>be a sequence of independent random variables such that</u> $P_r[\xi_j = 1] = P_r[\xi_j = -1] = 1/2$. <u>Let</u>

$$S = \sum_{j=1}^{n} \xi_j x_j,$$

<u>then, for every</u> $x \in \mathcal{X}$

$$P_r[S \in x + K] \leq \frac{1}{(n+1)^{1/3}}.$$

<u>Proof</u>. Consider a particular $x \in \mathcal{X}$.

Let $\nu(x) = \sup\{\nu[x; \varepsilon,k]; \varepsilon \in E_x, k=1,2,\cdots,n\}$ be defined as above and let $\nu = \sup[\nu(x); x \in \mathcal{X}]$. Since the number of elements $S(\varepsilon)$ situated outside of $x + K$ is at least (n/ν) times the number m situated inside, we can state that

$$P[S \in x + K] \leq \frac{m}{m + \frac{n}{\nu} m} = \frac{\nu}{n+\nu}.$$

Consider now a particular point $b \in \mathcal{X}$ such that $\nu(b) = \nu = \sup[\nu(x); x \in \mathcal{X}]$ and let $E = E_b$. There is a particular sequence ε' and an integer j_1 such that $\nu[b; \varepsilon', j_1] = \nu$. Hence there are sequences ε^r and integers j_r; $r=2,3,\cdots,\nu$, such that $S(\varepsilon^r) \in b + K$ and $\varepsilon^r(j_r) = \varepsilon'(j_1)$ for $r=2,3,\cdots,\nu$. Letting $y = \varepsilon^r(j_r)$ one can write

$$S(\varepsilon^r) - S(y) = S(\varepsilon^r) - S[\varepsilon^r(j_r)] = 2\, \varepsilon^r_{j_r} x_{j_r}.$$

To simplify the notation one can assume, without changing anything essential that $j_r = r$ and $\varepsilon^r_{j_r} x_{j_r} = x_r$. Let $a = b - S(y)$ and let $\Delta(\varepsilon) = S(\varepsilon) - S(y)$. We are then in a situation where for $r=1,2,\cdots,\nu$ we have

$$\Delta(\varepsilon^r) = 2\, x_r \in a + K,$$

$$x_r \notin K.$$

These conditions imply in particular that for every $s=1,2,\cdots,\nu$ and every $r=1,2,\cdots,\nu$ we have $x_r - x_s \in K$. Thus for any fixed $s=1,2,\cdots,\nu$, the set $x_s + K$ is a closed convex set which does not contain the origin but contains all the points x_r; $r=1,2,\cdots,\nu$.

The sum $S = \Sigma\xi_j x_j$ can also be written in the form $S = T + Z$ with

$$T = \sum_{j=1}^{\nu} \xi_j x_j$$

$$Z = \sum_{j=\nu+1}^{n} \xi_j x_j .$$

Since T and Z are independent we have

$$P\,[S \in x + K] \leqq \sup_z P_r[T \in z + K].$$

It is therefore sufficient to consider the behavior of T alone. Now $T = \Sigma_{j=1}^{\nu} \xi_j x_j$ with $x_j - x_s \in K$ and $x_s \notin K$.

Let $\varepsilon = \{\varepsilon_j;\ j=1,2,\cdots,\nu\}$ be a sequence such that $\varepsilon_j = \pm 1$ and let $T(\varepsilon) = \Sigma_{j=1}^{\nu} \varepsilon_j x_j$. To each such sequence ε associate the set $F(\varepsilon)$ of integers $j=1,2,\cdots,\nu$ such that $\varepsilon_j = +1$. If ε' and ε'' are different and such that $F(\varepsilon') \subset F(\varepsilon'')$ then

$$T(\varepsilon'') - T(\varepsilon') = 2 \sum_{j \in G} x_j$$

where G is a certain nonempty subset of the set of integers $\{1,2,\cdots,\nu\}$. Let s be an element of G so that $x_s + K$ contains all the elements x_j with $j \in G$. Since $x_s + K$ does not include the origin and is convex and closed, there is a $y \in \mathcal{Y}$ and a $\delta > 0$ such that $\langle z, y\rangle \geqq \delta$ for every $z \in x_s + K$. Letting $\alpha = \langle x_s, y\rangle$ this implies

$$\langle x_s + u,\ y\rangle = \alpha + \langle u,\ y\rangle \geqq \delta$$

for every $u \in K$. However, since K is symmetric we also have $\alpha - \langle u, y\rangle \geqq \delta$ for every $u \in K$. In particular

$$\sup\{|\langle u,\ y\rangle|;\ u \in K\} \leqq \alpha - \delta.$$

Consider now the sum $w = 2\ \Sigma_{j\in G}\ x_j$. For this sum we have

$$\langle w,\ y\rangle = 2\langle x_s,\ y\rangle + 2\ \Sigma\{\langle x_r,\ y\rangle;\ r \in G,\ r \neq s\}$$

$$\geqq 2\alpha + 2[\kappa(G) - 1]\delta.$$

Suppose now that λ is a positive number such that $w \in \lambda K$. Then, according to the above inequality

$$\lambda \geqq \frac{2\alpha + 2[\kappa(G) - 1]\delta}{\alpha - \delta} > 2.$$

In particular $w \notin 2K$. Hence, if $T(\varepsilon') \in (x+K)$ and $T(\varepsilon'') = T(\varepsilon') + w$ then $T(\varepsilon'') \notin x + K$ this implies that the sequences ε for which $T(\varepsilon) \in x + K$ give rise to sets $F(\varepsilon)$ which are not comparable.

An obvious application of the Sperner-Erdös lemma gives then

$$P\,[T \in x + K] \leqq \frac{1}{\sqrt{\nu + 1}}.$$

Summarizing, there is some integer $\nu \geqq 1$ such that both inequalities

$$P\,[S \in x + K] \leqq \frac{\nu}{n+\nu} \leqq \frac{\nu}{n+1}$$

and

$$P\,[S \in x + K] \leqq \frac{1}{\sqrt{\nu+1}}$$

are valid. It follows that

$$P\ [S \in x + K] \leq \frac{1}{(n+1)^{1/3}}$$

as claimed.

The preceding lemma 3 will allow us to derive an inequality on the concentration of sums of symmetric linear processes as follows.

Consider a system $\{\mathcal{X}, \mathcal{Y}, \mathcal{T}\}$ formed by the linear space $\mathcal{Y}$ with a locally convex topology $\mathcal{T}$ for which the dual of $\mathcal{Y}$ is $\mathcal{X}$. Consider only linear processes which are $\mathcal{F}$-tight for the family $\mathcal{F}$ of equicontinuous subsets of $\mathcal{X}$ and take probabilities for the Radon extensions of the distributions.

Theorem 2. Let $\{Z_j;\ j=1,2,\cdots,n\}$ be n independent $\mathcal{F}$-tight linear processes over $\mathcal{Y}$. Let Q_j be the Radon extension of the distribution of Z_j and let M be the measure $M = \Sigma_j\ Q_j$.

Suppose that each Z_j has a symmetric distribution and let K be a $w(\mathcal{X}, \mathcal{Y})$ closed convex symmetric subset of $\mathcal{X}$.

Then, for every $x \in \mathcal{X}$ and for $W = \Sigma_j Z_j$ we have

$$\frac{2}{3}\ P[W \in x + K] \leq \frac{1}{(s+1)^{1/3}}$$

with

$$s = M(K^c).$$

Proof. Let y_0 be an arbitrary nonzero element of $\mathcal{Y}$. If $x \in \mathcal{X}$ let $\check{x} = x$ whenever $\langle x, y_0\rangle \geq 0$ and let $\check{x} = -x$ if $\langle x, y_0\rangle \geq 0$. One may assume $P\ \{\langle Z_j;\ y_0\rangle = 0\} = 0$ if it is so desired. The operation $x \longrightarrow \check{x}$ applied to the process Z_j gives a process $\check{Z}_j$. The distribution of Z_j is the same as that of $\xi_j \check{Z}_j$

for real-valued random variables ξ_j which take values (-1) and (+1) with equal probabilities, equal to one-half. Therefore we may argue as if W was equal to $\Sigma\xi_j\check{Z}_j$. Let θ be the set of values $\theta = \{\check{Z}_j;\ j=1,2,\cdots,n\}$. If $N(\theta)$ is the number of $\check{Z}_j$ such that $\check{Z}_j \notin K$, lemma 3 gives

$$p(\theta) = \Pr[W \in x + K|\theta] \leqq \frac{1}{[N(\theta)+1]^{1/3}}\,.$$

Furthermore, $EN = s$ and variance $N \leqq s$.

For any $t \in [0, s]$ one can write

$$\frac{1}{(t+1)^{1/3}} \leqq \frac{1}{(s+1)^{1/3}} + [t-s]^{-}\left[1 - \frac{1}{(s+1)^{1/3}}\right]\frac{1}{s}.$$

Therefore

$$P[W \in x + K] \leqq \frac{1}{(s+1)^{1/3}} + \left[1 - \frac{1}{(s+1)^{1/3}}\right] E\left[\frac{N}{s} - 1\right]^{-}.$$

Since $E[N-s]^{-} = \frac{1}{2}E|N-s| \leqq \frac{1}{2}\sqrt{s}$ this implies

$$P[W \in x + K] \leqq \frac{1}{(s+1)^{1/3}} + \frac{1}{2}\,\frac{1}{(s+1)^{1/3}}\,\frac{(s+1)^{1/3} - 1}{\sqrt{s}}\,.$$

Hence

$$P[W \in x + K] \leqq \frac{3}{2}\,\frac{1}{(s+1)^{1/3}}$$

as claimed.

For small values of s it may be interesting to use a different bound. In fact when $s < 1$ the expectation $E[(N/s) - 1]^{-}$ is precisely equal to the probability that N be equal to zero, hence inferior to e^{-s}. In this case one would obtain

$$P[W \in x + K] \leq \frac{1}{(s+1)^{1/3}} + e^{-s}\left[1 - \frac{1}{(s+1)^{1/3}}\right]$$

$$= 1 - [1 - e^{-s}]\left[1 - \frac{1}{(s+1)^{1/3}}\right].$$

In particular $P[W \in x + K]$ can be equal to unity only if $s = 0$.

On the real line, a procedure due to Kolmogorov makes it possible to extend the validity of an improved version of theorem 2 to variables which are not symmetrically distributed in such a way that the order of magnitude of the bounds be not substantially altered. We have been unable to achieve this in the infinite dimensional situation. However, the procedure used earlier by Paul Lévy in a similar connection is still applicable. For this purpose we shall need two lemmas, which are essentially due to Paul Lévy.

__Lemma__ 4. __Let__ X __and__ Y __be two independent identically distributed linear processes over the space__ $\mathcal{Y}$. __Let__ K __be a closed convex symmetric subset of__ $\mathcal{X}$. __Then__

$$\{P\,[X - x \in K]\}^2 \leq P_r\{X - Y \in 2K\}.$$

The other lemma is the lemma on the increase of dispersion which has been already used in the proof of lemma 3; namely, if X and Y are independent and $P_r[X + Y \in x_0 + K] \geq \beta$ there is an $x \in \mathcal{X}$ such that $P_r[X \in x + K] \geq \beta$.

__Theorem__ 3. __Let__ $\{X_j;\ j=1,2,\cdots,n\}$ __be__ n __independent linear processes over__ $\mathcal{Y}$. __Assume that the processes__ X_j __are__ $\mathcal{T}$-__tight and take probability statements relative to the Radon extensions__.

Let $S = \Sigma_{j=1}^{n} X_j$ *and let* K *be a* $w(\mathcal{X}, \mathcal{Y})$ *closed convex symmetric subset of* $\mathcal{X}$. *Let*

$$C[2K; X_j] = \sup_{x \in \mathcal{X}} P[X_j \in x + 2K]$$

and

$$s = \sum_j \{1 - C[2K; X_j]\}.$$

Then, for every $x \in \mathcal{X}$ *we have*

$$P[S \in x + K] \leq \frac{1.3}{(s+1)^{1/3}}.$$

Proof. Let $\{Y_j; j=1,2,\cdots,n\}$ be n independent linear processes which are independent of $\{X_j; j=1,2,\cdots,n\}$. Suppose that the distributions of X_j and Y_j are identical. Let $T = \Sigma Y_j$ and let $W = S - T = \Sigma Z_j$, with $Z_j = X_j - Y_j$.

According to theorem 2

$$P[W \in 2K] \leq \frac{3}{2(t+1)^{1/3}}$$

with $t = \Sigma P_r[Z_j \notin 2K]$. By Paul Lévy's lemma on the increase in dispersion

$$P[Z_j \in 2K] \leq \sup_x P[X_j \in x + 2K] = C[2K; X_j].$$

This implies $t \geq s$. According to lemma 4

$$P[S \in x + K] \leq \{P_r[W \in 2K]\}^{1/2},$$

hence the result.

Remark. The quantities $C[2K; X_j]$ which occur in the statement of the theorem may occasionally be difficult to evaluate. In such cases it may be helpful to use the following lemma adapted from a result of Loève [3], page 247.

Lemma 5. Let X and Y be two independent identically distributed $\mathcal{T}$-tight linear processes. Let A be an arbitrary subset of $\mathcal{Y}$. For each $y \in A$ let $\mu(y)$ be a median of the numerical random variable $\langle X, y\rangle$ and let $a(y)$ be an arbitrary number. Then, for every $\alpha > 0$

$$\frac{1}{2}\tilde{P}\left\{\sup_{y\in A}|\langle X,y\rangle - \mu(y)| > \varepsilon\right\} \leq \tilde{P}\left\{\sup_{y\in A}|\langle X-Y, y\rangle > \varepsilon\right\}$$

$$\leq 2\,\tilde{P}\left\{\sup_{y\in A}|\langle X,y\rangle - a(y)| > \frac{\varepsilon}{2}\right\}.$$

To the preceding results one must add the rather obvious but important remark that the concentration of a sum $S = \Sigma X_j$ can be bounded by functions of the concentration of the real variables $\langle S, y\rangle$.

Let V_j be the real-valued variable $V_j = \langle X_j, y\rangle$ and for every positive number τ let

$$\gamma_j(y,\tau) = \sup_a P[a < V_j < a + \tau].$$

Kolmogorov's argument gives

$$P\,[b \leq \Sigma V_j \leq b + \lambda] \leq 2\ \text{Int}\left[1 + \frac{\lambda}{\tau}\right]\left\{\sum_j\,[1 - \gamma_j(y,\tau)]\right\}^{-1/2}.$$

From this inequality one can obtain in an obvious manner the following result. Let K be the set

$$K = \{x\colon \sup|\langle x,y\rangle| \leq 1;\ y \in F\}$$

where F is a finite subset of $\mathcal{Y}$. For a positive number τ let

$$\beta_j(\tau) = \sup_a P\left\{\sup_{y\in F}|\langle X_j, y\rangle - a(y)| \leq \tau\right\}.$$

Let K be the closed convex set

$$K = \left\{x: \sup_{y\in F}|\langle x, y\rangle| \leq 1\right\}.$$

Then, for every $\lambda > 0$ and every τ

$$P\,[S \in x + \lambda K] \leq 2[\kappa(F)]^{1/2}\ \mathrm{Int}\left[1 + \frac{\lambda}{\tau}\right]\left\{\Sigma\,[1 - \beta_j(\tau)]\right\}^{-1/2}.$$

Further inequalities can be obtained through the use of the Normal approximation theorem.

Some of the inequalities obtainable on the real line cannot be extended to linear spaces or Banach spaces in general as can be seen from the following examples.

Let $\mathcal{Y}$ be the space of summable sequences of real numbers. If $y \in \mathcal{Y}$ is such that $y = \{y_j;\ j=1,2,\cdots,\}$ let $\|y\| = \Sigma_j|y_j|$. Let $\mathcal{X}$ be the space of bounded sequences $x = \{x_j;\ j=1,\cdots\}$ with $\|x\| = \sup_j\ |x_j|$.

The Banach space $\mathcal{X}$ is the dual of the Banach space $\mathcal{Y}$. Let z_j be the sequence whose entries are all identically zero except the jth one which is unity. Let $Z_j = z_j$ with probability one-half and let $Z_j = -z_j$ with probability one-half. For each integer n the sum $S_n = \Sigma_{j=1}^n Z_j$ takes values in the unit ball of $\mathcal{X}$. Thus if K is the ball $K = \{x: \|x\| \leq 1 - \varepsilon\}$, $0 < \varepsilon < 1$ we have $P\,[Z_j \in K^c] = 1$ and $P_r\{S_n \in (1/1-\varepsilon)K = 1$. This shows that unless further assumptions are added one cannot hope to obtain bounds on the probability of falling in a convex set from similar bounds on the summands involving smaller convex sets.

In the preceding example, the variables Z_j are not identically distributed. The following construction indicates that an assumption of identity of distributions is not sufficient to allow a definite improvement on the situation. Consider again the sequences z_k defined above. For a given integer n, let $\{Z_j'; j=1,2,\cdots,n\}$ be a sequence of independent identically distributed variables such that

$$P[Z_j' = z_k] = \frac{1}{2n^3}, \qquad P[Z_j' = -z_k] = \frac{1}{2n^3}$$

for $k=1,2,\cdots,n^3$. Let $T_n = \Sigma_{j=1}^n Z_j'$. The probability that the values taken by the Z_j be all disjoint is larger than $[1-(n/n^3)]^n$. Therefore, except for cases having a total probability not in excess of $1 - (1-(1/n^2))^n \leqq 1/n$, the sum T_n has a norm $\|T_n\|$ exactly equal to unity although $P\{\|Z_j'\| < 1\} = 0$.

As n tends to infinity the two sums S_n and T_n just constructed behave in essentially different ways. However, we have $P[\|Z_j\| = 1] = 1$ and $P\{\|S_n\| = 1\}$ and also $P[\|Z_j'\| = 1] = 1$ and $P\{\|T_n\| = 1\} > 1 - 1/n$.

The linear process $S_\infty = \Sigma_{j=1}^\infty Z_j$, sum of an infinite sequence of independent random elements Z_j such that $P[Z_j = z_j] = P[Z_j = -z_j] = 1/2$, can also be used to illustrate the meaning of the definitions given in section 2. When the topology $\mathcal{C}$ of $\mathcal{Y}$ is the topology defined by the norm, the system Ψ of functions introduced in section 2 is precisely the space of all bounded numerical functions whose restrictions to the balls of $\mathcal{X}$ are $w(\mathcal{X}, \mathcal{Y})$ continuous.

The family $\mathcal{F}$ of equicontinuous subsets of $\mathcal{X}$ can be replaced by the family of balls in $\mathcal{X}$. The Radon extension $\tilde{P}$ of the distribution of S_∞ is defined for all closed balls in $\mathcal{X}$. Therefore the domain of $\tilde{P}$ includes all the closed balls and all the open balls in $\mathcal{X}$. However, in a system of axioms of set theory in which the continuum is not a weakly inaccessible cardinal it can be proved that $\tilde{P}$ cannot be extended to a σ-additive measure whose domain includes all the Borel subsets of the Banach space $\mathcal{X}$. In such a situation, there will be some bounded strongly continuous numerical functions defined on $\mathcal{X}$ which are not integrable for $\tilde{P}$. This type of circumstances cannot occur in a Hilbert space or a reflexive Banach space even if the space in question is not separable.

4. An application. Consider a linear space $\mathcal{Y}$ with a locally convex topology $\mathcal{C}$. Let $\mathcal{X}$ be the dual of $\{\mathcal{Y}, \mathcal{C}\}$.

If X is K-tight and K is a subset of $\mathcal{X}$ let

$$C[K; X] = \sup_x P[X \in x + K]$$

be the concentration of X at K.

For each integer n let $\{X_{n,j};\ j=1,2,\cdots,\}$ be a sequence of identically distributed K-tight linear processes over $\mathcal{Y}$. Let m_n and k_n be two integers and let

$$S_n = \sum_{j=1}^{k_n} X_{n,j},$$

$$T_n = \sum_{j=1}^{m_n} X_{n,j}.$$

Further, let ν_n be the integer part of (k_n/m_n).

<u>Lemma</u> 6. <u>If</u> K <u>is a closed convex symmetric subset of</u> $\mathcal{X}$ <u>and</u>

$$\varepsilon_n = \frac{5}{\nu_n\{C[K;\ S_n]\}^6}$$

then for each n there is an $x_n \in \mathcal{X}$ such that

$$P[T_n \in x_n + 2K] \geq 1 - \varepsilon_n.$$

Proof. This follows from theorem 3 by noting that S_n has the same distribution as a sum

$$R_n + \sum_{j=1}^{\nu_n} T_{n,j}$$

of independent variables such that $T_{n,j}$ has the same distribution as T_n.

Lemma 7. Assume that $\nu_n \longrightarrow \infty$ as $n \longrightarrow \infty$ and that there is a constant b independent of n such that

$$\liminf_{n \to \infty} P[S_n \in bK] > 0$$

then the inequality

$$\liminf_{n \to \infty} C[K; S_n] > 0$$

implies

$$\lim_{n \to \infty} P[T_n \in 6K] = 1.$$

Proof. Let $\varepsilon_n = 5/\nu_n C[K; S_n]^6$ as in lemma 6 and consider a sequence $x_n \in \mathcal{X}$ such that

$$P\{T_n \in x_n + 2K\} \geq 1 - \varepsilon_n.$$

Suppose that $x_n \notin 4K$. Then the closed convex sets $2K$ and $x_n + 2K$ are disjoint and the difference $x_n + 2K - 2K = x_n + 4K$ is a closed convex set which does not contain the origin of $\mathcal{X}$.

Therefore, there is a $y_n \in \mathcal{Y}$ such that

1) $|\langle x, y_n\rangle| \leq$ for $x \in 2K$,

2) $\langle x, y_n\rangle \geq 1$ for $x \in x_n + 2K$.

Let $X'_{n,j}$, S'_n, T'_n be the real variables defined by $X'_{n,j} = \langle X_{n,j}, y_n\rangle$; $S'_n = \langle S_n, y_n\rangle$; $T'_n = \langle T_n, y_n\rangle$. Kolmogorov's theorem, applied to these real variables gives inequalities of the following type. If ξ is a real random variable let

$$\gamma(\tau,\xi) = \sup_a P[a \leq \xi \leq a + \tau],$$

then for λ and τ positive we have

$$\gamma[\lambda;\ S'_n] \leq 2\ \mathrm{Int}\left[1 + \frac{\lambda}{\tau}\right]\left\{k_n\left[1 - \gamma(\tau, X'_{n,j})\right]\right\}^{-1/2}.$$

Also by definition of y_n

$$C[K;\ S_n] \leq \gamma[2;\ S'_n].$$

Hence, for every positive number τ there are numbers β_n such that

$$k_n\ P\left\{|X'_{n,j} - \beta_n| > \tau\right\} \leq \frac{4}{C[K;\ S_n]^2}\left[1 + \frac{2}{\tau}\right]^2.$$

This implies

$$m_n\ P\left\{|X'_{n,j} - \beta_n| > \tau\right\} \leq 4\left(\frac{\epsilon_n}{\nu_n^2}\right)^{1/3}\left[1 + \frac{2}{\tau}\right]^2.$$

The inequality $\gamma[2;\ S'_n] \geq C[K;\ S_n]$ implies also that for a suitable choice of β'_n we have

$$k_n\ E|X'_{n,j} - \beta'_n|^2\ I\left\{|X'_{n,j} - \beta'_n| < 1\right\} \leq A < \infty .$$

Under these circumstances there are constants c_n (truncated expectations) such that

$$S_n' - k_n c_n = \xi_n$$

and

$$T_n' - m_n c_n = \varphi_n$$

are random variables with relatively compact sequences of distributions on the line. Furthermore, φ_n tends in probability to zero. However, by definition of y_n we have

$$P\left[T_n' \geqq 1\right] \geqq P\left[T_n \in x_n + 2K\right] \geqq 1 - \varepsilon_n,$$

hence for n sufficiently large, $m_n c_n > 1/2$. This implies in turn

$$\langle S_n, y_n \rangle = S_n' \geqq \frac{k_n}{m_n} m_n c_n + \xi_n \geqq \frac{1}{2} \nu_n + \xi_n.$$

It follows that for every $0 < b < \infty$ we have

$$\lim_{n \to \infty} P\left[S_n \in (b+1)K\right] = 0$$

contrary to the assumption made. Hence, for n sufficiently large we shall have $x_n \in 4K$, hence

$$P\left[T_n \in 2K + 4K = 6K\right] \geqq 1 - \varepsilon_n.$$

This completes the proof of the lemma.

As an application of theorems 2 and 3 one can derive some results concerning Central limit theorems for linear processes. For independent real variables $\{X_j;\ j=1,2,\cdots,k\}$ such that the distribution $\mathcal{L}(X_j)$ of X_j be P_j, it is known that if the variables are suitably centered the convolution product

$$\mathcal{L}(\Sigma X_j) = \prod_{j=1}^{k} P_j$$

differs little from the Poisson exponential $\exp\{\Sigma[P_j - I]\}$, provided only that the dispersions of the variables involved be somewhat similar [4].

Also it appears difficult to extend this result to linear processes in general; some partial results can be obtained as follows.

For each integer n let k_n be an integer and let $\{X_{n,j};\ j=1,2,\cdots\}$ be a sequence of independent linear processes. We shall assume that $k_n \longrightarrow \infty$ as $n \longrightarrow \infty$ and that each $X_{n,j}$ has been written in a split form

$$X_{n,j} = [1 - \xi_{n,j}]U_{n,j} + \xi_{n,j}V_{n,j}$$

where the variables $\xi_{n,j}$, $U_{n,j}$, $V_{n,j}$ are all independent and $P[\xi_{n,j} = 1] = 1 - P[\xi_{n,j} = 0] = \alpha_{n,j} \leq \alpha_n$.

A sum $S_n = \Sigma_{j=1}^{k_n} X_{n,j}$ can then be written in the form $S_n = U_n + V_n + W_n$ with

$$U_n = \sum_{j=1}^{k_n} (1 - y_{n,j})U_{n,j}$$

$$V_n = \sum_{j=1}^{k_n} \xi_{n,j}V_{n,j}$$

$$W_n = \sum_{j=1}^{k_n} (y_{n,j} - \xi_{n,j})U_{n,j}.$$

The variables $y_{n,j}$ introduced in these formulas have the same distribution as the $\xi_{n,j}$ but are independent of all the other variables.

Let $Y_{n,j} = (1 - y_{n,j})U_{n,j}$ and let $A_{n,j} = \mathcal{L}(Y_{n,j})$ and $B_{n,j} = \mathcal{L}[\xi_{n,j} \; V_{n,j}]$. For the distribution of V_n the usual argument gives the inequality

$$\left\| \mathcal{L}(V_n) - \exp\left\{ \sum_{j=1}^{k_n} \left[B_{n,j} - I\right]\right\} \right\| \leq 2 \sum_{j=1}^{k_n} \alpha_{n,j}^2 .$$

Also, <u>if the variables</u> $Z_{n,j}$ <u>are identically distributed</u> a theorem of Prohorov implies that

$$\left\| \mathcal{L}(V_n) - \exp\left\{ \sum_{j=1}^{k_n} \left[B_{n,j} - I\right]\right\} \right\| \leq 4 \, \alpha_n .$$

Similar inequalities are applicable to the distribution of W_n.

Under favorable circumstances one may expect that the distribution P_n of S_n may be approximated by the exponential

$$Q_n = \exp\left\{ \sum_{j=1}^{k_n} (P_{n,j} - I) \right\} = \exp\left\{ \sum_{j=1}^{k_n} \left[(A_{n,j} - I) + (B_{n,j} - I) \right] \right\}.$$

Let $A'_{n,j}$ be the symmetric of $A_{n,j}$ and let $2\bar{A}_{n,j} = A_{n,j} + A'_{n,j}$. Let K be a closed convex symmetric subset of $\mathcal{X}$ such that

$$\liminf_{n\to\infty} \sup_x Q_n\left[x + \tfrac{1}{2} K\right] = \varepsilon > 0.$$

Then, if F_n is the measure

$$F_n = \exp\left\{ 2 \sum_{j=1}^{k_n} \left[\bar{A}_{n,j} - I\right] \right\}$$

we also have

$$\liminf_{n\to\infty} F_n(K) \geq \varepsilon^2 > 0.$$

Consequently, according to theorem 2, if $\{\beta_n\}$ is any sequence of numbers tending to zero and

$$G_n = \exp\left\{\beta_n \sum_{j=1}^{k_n} \left[\bar{A}_{n,j} - I\right]\right\}$$

we have

$$\lim_{n\to\infty} G_n(K) = 1.$$

Consider now the distribution $M_{n,j}$ of $Z_{n,j}$. This distribution may be written

$$M_{n,j} = \beta_{n,j} \bar{A}_{n,j}$$

with

$$\beta_{n,j} = \frac{2\alpha_{n,j}}{1-\alpha_{n,j}} \le \frac{2\alpha_n}{1-\alpha_n}.$$

An application of theorem 2 gives the following result.

Lemma 8. *Let* $\{X_{n,j};\ j=1,2,\cdots\}$ *be independent* K-*tight linear processes written in a split form as explained above.*

If K_n *is a closed convex subset of* $\mathcal{X}$ *such that*

a) $$\liminf_{n\to\infty} \sup_x Q_n\left[x + \tfrac{1}{2} K_n\right] > 0$$

and if in addition either

b) $$\sum_{j=1}^{k_n} \alpha_{n,j}^2 \longrightarrow 0$$

or

c) $$\alpha_n \longrightarrow 0$$

and the variables $Z_{n,j}$ *are identically distributed for each* n *then*

$$\lim_{n\to\infty} P[W_n \in K_n] = 1.$$

Note that when the split form $X_{n,j} = (1-\xi_{n,j})U_{n,j} + \xi_{n,j} V_{n,j}$ is chosen such that $V_{n,j}$ does not take any values in K the condition (b) that $\Sigma_{j=1}^{k_n} \alpha_{n,j}^2 \longrightarrow 0$ is a consequence of (a) and the condition that $\alpha_n \longrightarrow 0$. In fact, condition (a) implies then that $\Sigma_{j=1}^{k_n} \alpha_{n,j}$ stays bounded. A slight refinement of the preceding argument can be used to show that in any event the condition (b) can be replaced by the weaker condition $\Sigma_{j=1}^{k_n} \alpha_{n,j}^3 \longrightarrow 0$.

Consider now the case where for each n the linear processes $\{X_{n,j};\ j=1,2,\cdots\}$ are identically distributed and suppose that the split form $X_{n,j} = (1-\xi_{n,j})U_{n,j} + \xi_{n,j} V_{n,j}$ is also selected so that the distributions of $\xi_{n,j}$, $U_{n,j}$ and $V_{n,j}$ do not depend on j.

Let $Y_{n,j} = (1-y_{n,j})U_{n,j}$ as above. The distribution

$$Q_n' = \exp\left\{\sum_{j=1}^{k_n} \left[A_{n,j} - I\right]\right\}$$

can be written as the distribution of

$$U_n' = \sum_{j=1}^{N_n} Y_{n,j}$$

where N_n is independent of the $X_{n,j}$, $\xi_{n,j}$, etc. and has a Poisson distribution with expectation k_n. In this case let Δ_n be the difference $\Delta_n = U_n' - U_n$. This difference is a sum of a random number of terms $Y_{n,j}$ or $(-Y_{n,j})$ but the number of terms in the sum has a probability tending to zero of exceeding $k_n^{3/4}$. Similarly the number of terms in U_n' exceeds $k_n/2$ with probability tending to unity. Thus lemma 7 gives the following result.

Lemma 9. _If the_ $Y_{n,j}$ _are identically distributed for each_ n _and either_

$$\liminf_{n\to\infty} C[K;\ U_n] > 0$$

_or_

$$\liminf_{n\to\infty} C[K;\, U_n'] > 0$$

and for some $b > 0$

$$\liminf_{n\to\infty} P[U_n \in bK] + P[U_n' \in bK] > 0$$

then

$$\lim_{n\to\infty} P[W_n \in 6K] = 1$$

and

$$\lim_{n\to\infty} P[U_n' - U_n \in 6K] = 1.$$

The preceding lemmas lead to a theorem which can be used as a stepping stone toward the proof of Central limit theorems for linear processes. We shall formulate it as follows.

For each n let $\{X_{n,j};\ j=1,2,\cdots\}$ be a sequence of independent linear processes. A split form $X_{n,j} = (1-\xi_{n,j})U_{n,j} + \xi_{n,j} V_{n,j}$ will be called favorable to the closed symmetric convex set K if the following conditions are satisfied:

1) $U_{n,j}$, $\xi_{n,j}$, $V_{n,j}$ have distributions independent of j.

2) $\lim_{n\to\infty} P[\xi_{n,j} = 1] = 0$.

3) If $U_n = \Sigma_{j=1}^{k_n} (1-\xi_{n,j})U_{n,j}$ and U_n' is the corresponding Poisson sum

$$U_n' = \sum_{j=1}^{N_n} (1-\xi_{n,j})U_{n,j}$$

then

$$\liminf_{n\to\infty}\left\{C[K;\, U_n] + C[K;\, U_n']\right\} > 0.$$

4) For some $b < \infty$

$$\liminf_{n\to\infty}\left\{P\left[U_n \in bK\right] + P\left[U_n' \in bK\right]\right\} > 0.$$

Theorem 4. *For each integer* n *let* k_n *be another integer such that* $k_n \longrightarrow \infty$. *Let* $\{X_{n,j};\ j=1,2,\cdots\}$ *be a sequence of independent identically distributed* K-*tight linear processes. Let* $\mathcal{B}$ *be the family of closed convex symmetric subsets* K *of* $\mathcal{X}$ *such that for each* $\beta > 0$ *there is a split form of the* $\{X_{n,j}\}$ *favorable to* βK. *Let* $\mathcal{U}$ *be the uniform structure defined on* $\mathcal{X}$ *by the* vicinities $\{(x_1,x_2);\ (x_1-x_2) \in K\}$ *for* $K \in \mathcal{B}$.

Let

$$P_n = \prod_{j=1}^{k_n} P_{n,j} = \mathcal{L}\left[\sum_{j=1}^{k_n} X_{n,j}\right]$$

and let Q_n *be the exponential*

$$Q_n = \exp\left\{\sum_{j=1}^{k_n}\left[P_{n,j} - I\right]\right\}.$$

Then for every bounded $\mathcal{U}$-*uniformly continuous numerical function* f *defined on* $\mathcal{X}$ *we have*

$$\lim_{n\to\infty}\left|\int f(x)P_n(dx) - \int f(x)Q_n(dx)\right| = 0.$$

This is an immediate consequence of our previous lemmas.

5. *Empirical distribution functions*. Let T be an arbitrary set with a σ-field $\mathcal{S}$. For each integer n let p_n be a probability measure on $\mathcal{S}$ and let $\{\tau_{n,j};\ j=1,2,\cdots\}$ be a sequence of independent random elements having individually the distribution p_n on $\mathcal{S}$.

For a numerical function y defined on T let

$$s_n^2(y) = \int |y(t)|^2 p_n(dt).$$

Let $\mathcal{Y}_n$ be the space of square integrable functions on $\{T, \mathcal{O}, p_n\}$. This space is a Hilbert space for the norm s_n.

Each $\tau_{n,j}$ defines a linear process $\xi_{n,j}$ on $\mathcal{Y}_n$ according to the prescription

$$\langle \xi_{n,j}, y \rangle = y(\tau_{n,j}).$$

The normalized version of the empirical distribution function most frequently encountered in the statistical literature is the linear process

$$F_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\xi_{n,j} - p_n) = \sum_{j=1}^{n} X_{n,j}$$

with

$$X_{n,j} = \frac{1}{\sqrt{n}} (\xi_{n,j} - p_n).$$

Several well-known theorems on the asymptotic behavior of F_n can be expressed roughly as follows. First, the distribution of F_n "differs little" from that of $F_n^* = \Sigma_{j=1}^{N_n} X_{n,j}$ where N_n is a Poisson variable independent of the $X_{n,j}$ having expectation $E\, N_n = n$.

Second, and in a more restricted sense, the distribution of F_n "differs little" from that of the normal linear process G_n having expectation zero and the same covariance as F_n (or F_n^*).

A related and somewhat simpler process is the process

$$H_n = F_n^* + \frac{N_n - n}{\sqrt{n}} p_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{N_n} \xi_{n,j} - \sqrt{n}\, p_n.$$

This process H_n is decomposable in the sense that if $\{y_1, y_2, \cdots, y_k\}$ is any finite subset of two by two disjoint elements of $\mathcal{Y}_n$, then the variables $\{\langle H_n, \mathcal{Y}_r \rangle;\ r=1,2,\cdots,k\}$ are independent.

Note that F_n, F_n^* and H_n have expectations equal to zero and

$$E|\langle F_n, y\rangle|^2 = E|\langle F_n^*, y\rangle|^2 = s_n^2(y) - |\langle p_n, y\rangle|^2 \leq s_n^2(y).$$

In addition to the above processes it is convenient to introduce processes $F[n, \nu_n]$, $F^*[n, \nu_n]$ and $H(n, \nu_n)$ defined as follows. If M_n is a Poisson variable independent of the $X_{n,j}$ with $E\, M_n = \nu_n$ then

$$F[n, \nu_n] = \sum_{j=1}^{\nu_n} X_{n,j},$$

$$F^*[n,\nu_n] = \sum_{j=1}^{M_\nu} X_{n,j},$$

$$H(n, \nu_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^{M_n} \xi_{n,j} - \frac{\nu_n}{n} p_n = F^*(n,\nu_n) + \frac{M_n - \nu_n}{\sqrt{n}} p_n.$$

Finally we shall also use the symmetrized processes $\overline{H}(n,\nu_n) = H(n,\nu_n) - H'(n,\nu_n)$ where $H'(n,\nu_n)$ is independent of $H(n,\nu_n)$ and has the same distribution as $H(n,\nu_n)$. By the symbol $\overline{H}_n$ will be meant $\overline{H}(n,n)$.

Let B_n be a convex symmetric subset of $\mathcal{Y}_n$. Denote $\mathcal{Y}(B_n)$ the linear subspace of $\mathcal{Y}_n$ spanned by B_n and consider on $\mathcal{Y}(B_n)$ the norm

$$\|y\|_n = \inf\{\lambda;\ y \in \lambda B_n\}.$$

Let $\mathcal{X}(B_n)$ be the dual of the normed space $\mathcal{Y}(B_n)$. This space $\mathcal{X}(B_n)$ is a Banach space for the norm ρ_n defined by

$$\rho_n(x) = \sup\{|\langle x,y\rangle| ; \|y\|_n \leq 1\}.$$

A set $B_n \subset \mathcal{Y}_n$ will be called stochastically bounded if for every $\varepsilon > 0$ there is a subset $S \subset T$ and a number b such that (a) $p_n(T \setminus S) \leq \varepsilon$, (b) if $t \in S$ and $y \in B_n$ then $|y(t)| \leq b$.

In particular lattically bounded subsets of $\mathcal{Y}_n$ are stochastically bounded.

Lemma 10. In order that $\xi_{n,j}$ considered as a linear process over $\mathcal{Y}(B_n)$ be tight for the $w[\mathcal{X}(B_n), \mathcal{Y}(B_n)]$ compacts of $\mathcal{X}(B_n)$, it is necessary and sufficient that B_n be stochastically bounded.

If B_n is stochastically bounded the processes $X_{n,j}$, F_n, F_n^*, H_n and $\bar{H}_n$ are all tight for the weak compacts of $\mathcal{X}(B_n)$. Therefore, probability statements may be made for the Radon extensions of their distributions. This will be the meaning attached to the symbols $\tilde{P}$ used below.

Let C be the concentration function defined in the preceding sections

$$C(K; F) = \sup_x \tilde{P}[F \in x + K].$$

For any closed convex symmetric subset K of $\mathcal{X}(B_n)$ define $\pi_n[\nu_n, K]$ by

$$3\pi_n[\nu_n, K] = \tilde{P}[\bar{H}(n,\nu_n) \in K] + C[K; F(n,\nu_n)] + C[K; F^*(n,\nu_n)].$$

Lemma 11. Let $\{B_n\}$ be a sequence of convex symmetric subsets $\mathcal{Y}_n$ satisfying the conditions

(C_1) Each B_n is stochastically bounded.

(C_2) $\sup\{s_n^2(y);\ y \in B_n;\ n=1,2,\cdots\} < \infty$.

(C_3) If K_n is the polar of B_n in $\mathcal{X}(B_n)$ then for every $\alpha > 0$ there is a sequence $\{\nu_n\}$ such that

$$\lim_n \left(\frac{\nu_n^2}{n}\right) = \infty \quad \text{and} \quad \liminf_{n\to\infty} \pi_n[\nu_n, \alpha K_n] > 0.$$

Then $\rho_n(F_n - F_n^*)$ tends to zero in probability.

Proof. If condition C_3 is satisfied then for each $\alpha > 0$ there is a sequence $\{\nu_n\}$ such that $\nu_n^2/n \longrightarrow \infty$ and $\nu_n/n \longrightarrow 0$ and also $\pi_n[\nu_n, \alpha K_n] \longrightarrow 1$ as follows immediately from theorems 2 and 3.

Further, $y_n \in B_n$ implies

$$E\langle F(n,\nu_n),\ y_n\rangle = E\langle F^*(n,\nu_n),\ y_n\rangle = E\langle \bar{H}(n,\nu_n),\ y_n\rangle = 0$$

and

$$E|\langle F(n,\nu_n),y_n\rangle|^2 = E|\langle F^*(n,\nu_n),y_n\rangle|^2 \leq \frac{1}{2}\, E|\langle \bar{H}(n,\nu_n)y_n\rangle|^2$$
$$= \frac{\nu_n}{n}\, s_n^2(y_n).$$

One concludes easily that

$$\tilde{P}[F_n - F_n^* \in 3\alpha K_n] \longrightarrow 1.$$

Hence the result.

According to the preceding lemma, to find limiting distributions of functions of F_n one may in many cases replace F_n by F_n^*. Furthermore, it will often be possible to argue instead on H_n which differs from F_n^* only by the addition of a one-dimensional random variable.

The conditions C_1 and C_2 of lemma 11 are simple and easily verifiable. Unfortunately similar statements cannot be made for condition C_3. We shall give below examples of sequences $\{B_n\}$ which satisfy another condition, essentially much more restrictive than C_3 as follows.

Let B_n be a stochastically bounded subset of $\mathcal{Y}_n$ and let K_n be the polar of B_n in $\mathcal{X}(B_n)$. Let $S_n(\varepsilon, \alpha)$ be the smallest cardinal k such that there exists a linear subspace L of $\mathcal{X}(B_n)$ of dimension at most equal to k such that

$$\tilde{P}\{\bar{H}_n \in \overline{(L + \alpha K_n)}\} \geqq \varepsilon.$$

The cardinal $S_n(\varepsilon, \alpha)$ will be called the dimensional spread of $\bar{H}_n$ at the level (ε, α).

A sequence $\{B_n\}$ of stochastically bounded subsets of $\mathcal{Y}_n$ satisfies condition C_4 if for each ε and α there is a finite b such that the dimensional spread $S_n(\varepsilon, \alpha)$ stays smaller than b.

<u>A sequence</u> $\{B_n\}$ <u>which satisfies</u> C_1, C_2 <u>and</u> C_4 <u>necessarily satisfies</u> C_3 <u>in the sense that</u>

$$\liminf_{n\to\infty} \tilde{P}\left[\bar{H}_n \in \alpha K_n\right] > 0.$$

To give examples of situations where C_4 is satisfied one can use the following simple lemmas. The first of these will be recognized as a classical result of P. Lévy. The second is a particular form of results which can also be credited to P. Lévy. The third is the Hajek-Renyi form of Kolmogorov's inequality.

<u>Lemma</u> 12. <u>Let</u> $X_1, X_2, \cdots, X_m$ <u>be real random variables</u>. <u>Let</u> $S_k = \Sigma_{j=1}^{k} X_j$. <u>Assume that for each</u> k <u>the conditional distribution</u>

of $S_m - S_k$ *given* $S_1, S_2, \cdots, S_k$ *has median zero. Then*

$$P\left\{\max_k |S_k| > t\right\} \leq 2\, P\left\{|S_m| > t\right\}.$$

Lemma 13. *Let* $X_1, X_2, \cdots, X_m$ *be independent random variables having symmetric distributions. Let* $\sigma_k^2 = E\, X_k^2$ *and* $\sigma^2 = \Sigma_{j=1}^m \sigma_j^2$. *Let* $S = \Sigma_{j=1}^m X_j$.

If

$$E\, e^{iS} \leq \exp\left[-\frac{\sigma^2}{2} + \varepsilon\right]$$

with $\varepsilon \geq 0$, *then*

$$\sum_j P\left\{|X_j| > t\right\} \leq \frac{1}{g(t)}\left\{\varepsilon + \frac{1}{8}\left(1 - \frac{\delta^2}{2}\right)^{-1}\left(\sum_j \sigma_j^4\right)\right\}$$

with $g(t) = \cos t - 1 + t^2/2$ and $\delta^2 = \max_j \sigma_j^2$.

Lemma 14. *Let* $X_1, X_2, \cdots, X_n$ *be independent random variables with expectations zero and variances* $\sigma_j^2 = E\, X_j^2$. *Let* $S_k = \Sigma_{j=1}^k X_j$ *and let* $\{c_j;\ j=1,2,\cdots,n+1\}$ *be a nonincreasing sequence of numbers such that* $c_{n+1} = 0$. *Then*

$$P\left\{\sup_k |c_k\, S_k| > t\right\} \leq \frac{1}{t^2} \sum_{j=1}^n c_j^2 \sigma_j^2.$$

For applications to the present situation note that the characteristic function of the process $\overline{H}[n, \nu_n]$ is given by the expression

$$\log E \exp[i\langle \overline{H}(n, \nu_n), y\rangle]$$

$$= -\frac{\nu_n}{n} s_n^2(y) + 2\nu_n \int g\left[\frac{y(t)}{\sqrt{n}}\right] p_n(dt).$$

Lemmas 13 and 14 give immediately inequalities applicable to the processes H_n or $\overline{H}_n$ as follows.

Lemma 15. For each positive or negative integer k let u_k be an element of $\mathcal{Y}_n$ and let c_k be a nonnegative number. Assume that the u_k are two by two disjoint and that $c_k \geq c_{k+1}$. Furthermore, assume that $\beta = \Sigma_j c_j u_j \in \mathcal{Y}_n$. Let $y_k = \Sigma_{j=-\infty}^{k} u_j$ and $\delta^2 = \sup_k c_k^2 s_n^2(u_k)$. Then

$$g(t)\ P\left\{\sup_k |c_k \langle \bar{H}_n, u_k \rangle| > t\right\} \leq 2n \int g\left[\frac{\beta(\tau)}{\sqrt{n}}\right] p_n(d\tau)$$
$$+ \frac{1}{2}(1-\delta^2)^{-1}\ \delta^2 s_n^2(\beta),$$

and

$$P\left\{\sup_k |c_k \langle \bar{H}_n, y_k \rangle| > t\right\} \leq \frac{1}{t^2}\ s_n^2(\beta).$$

This lemma together with lemma 12 can be used to derive a number of results relative to empirical cumulative distribution functions.

To obtain theorems which can be used to derive the well-known result of Donsker and results of a similar nature relative to other norms or to multivariate cumulative distributions, we shall study symmetric convex sets B_n which are defined by means of sets $\{J_n;\ j \longrightarrow y_j;\ z_n;\ w_n\}$ as follows.

a) J_n is a totally ordered set,

b) the map $j \longrightarrow y_j$ associates to each $j \in J_n$ an indicator $y_j \in \mathcal{Y}_n$ in such a way that $j_1 \leq j_2$ implies $y_{j_1} \leq y_{j_2}$, almost everywhere,

c) z_n is an element of $\mathcal{Y}_n$,

d) w_n is a positive nonincreasing numerical function defined on J_n,

e) B_n is the smallest convex symmetric subset of $\mathcal{Y}_n$ which contains all the elements $w_n(j) y_j z_n$ for $j \in J_n$.

Let $S = \{j_0, j_1, \cdots, j_m\}$ be a finite subset of J_n such that $j_i \leq j_{i+1}$. To the set S one can associate an orthogonal projection Π_S of the Hilbert space $\mathcal{Y}_n$ into itself by the formula

$$\Pi_S \, x = \sum_{k=1}^{m} c_k u_k z_k$$

with $u_k = y_{j_k} - y_{j_{k-1}}$ and

$$c_k = \frac{\int x \; u_k z_n dp_n}{s_n^2(u_k \; z_n)} \, .$$

If $\kappa(S)$ is the cardinality of S, the rank of Π_S is at most $\kappa(S) - 1$. Furthermore, for each $j_k \in S$ let $r_k = \sup\{[W_n(j)/W_n(j_k)]; j_{k-1} < j \leq j_k\}$. Then

$$\Pi_S \, B_n \subset (\sup_k r_k) B_n .$$

It is enough to prove the corresponding membership relation for elements of the form $w_n(j) y_j z_n$. If $j \leq j_0$ or if $j \geq j_m$ or if j is one of the elements j_k; $k=0,1,2,\cdots,m$ of S the result is immediate. If $j_{k-1} < j \leq j_k$ then $\Pi_S \, y_j z_n$ is a linear combination of $y_{k-1} z_n$ and $y_k z_n$. The result follows easily.

To the system $\pi = \{J_n; \; j \longrightarrow y_j; \; S; \; z_n; \; w_n\}$ we shall associate numbers as follows.

a) $\alpha_{n,\pi}^2 = \sup\left\{s_n^2\left[(y_{j_0} + y_j - y_{j_m})\beta_n\right]; \;\; j \in J_n, \; y_j \geq y_{j_m}\right\}$,

b) $\sigma_{n,\pi}^2 = 2 \; s_n^2\left[(y_{j_m} - y_0)\beta_n\right]$,

c) $\delta_{n,\pi}^2 = \sup_{1 \leq k \leq m} s_n^2\left[(y_{j_k} - y_{j_{k-1}})\beta_n\right]$,

d) $\eta_{n,\pi} = 2n \int g\left[\frac{1}{\sqrt{n}} \beta_n\right] (y_{j_m} - y_{j_0}) \, dp_n$

with $g(\alpha) = \cos\alpha - 1 + \alpha^2/2$ and with β_n equal to a measurable function which is such that its equivalence class $\dot{\beta}_n$ is the supremum of the equivalence classes $4w_n(j)\dot{y}_j|\dot{z}_n|$ for $j \in J_n$, that is,

e) $\beta_n \in \dot{\beta}_n = 4\sup\{w_n(j)\dot{y}_j|\dot{z}_n|;\ j \in J_n\}$.

Proposition 1. With the above notation, let D be an arbitrary countable subset of B_n. Assume that the system π is such that $w_n(j) \leqq 4\, w_n(j_k)$ if $j \in J_n$ is such that $\dot{y}_{j_{k-1}} < \dot{y}_j \leqq \dot{y}_{j_k}$. Then, for every $\varepsilon > 0$

$$P\left\{\sup\left[|\langle H_n,\ (I-\Pi_S)y\rangle|;\ y \in D\right] > 3\varepsilon\right\}$$
$$\leqq 2\,\frac{\alpha^2_{n,\pi}}{\varepsilon^2} + \frac{2}{g(\varepsilon)}\left\{\eta_{n,\pi} + \frac{1}{4}\ \frac{\delta^2_{n,\pi}\sigma^2_{n,\pi}}{1-\delta^2_{n,\pi}}\right\}.$$

In addition, if $\overline{G}_n$ is the normal linear process having expectation zero and variances $E|\langle\overline{G}_n, y\rangle|^2 = 2\, s_n^2(y)$ then

$$P\left\{\sup\left[|\langle\overline{G}_n,\ (I-\Pi_S)y\rangle|;\ y \in D\right] > 3\varepsilon\right\}$$
$$\leqq \frac{\alpha^2_n}{\varepsilon^2} + \frac{1}{2g(\varepsilon)}\ \frac{\delta^2_{n,\pi}\sigma^2_{n,\pi}}{1-\delta_{n,\pi}}\ .$$

Proof. To make the following proof more readable we shall omit the subscripts n whenever possible. Thus J_n becomes J and w_n becomes w, etc.

Enlarging the family D and the set J if necessary one can immediately reduce the problem to the case where each element of D has the form $w(j)y_j z$ for $j \in J$.

Further, we may assume that D contains the elements $w(j_i)y_{j_i}z$

corresponding to S and elements $w(j_k)y_k z$ for $k=-1,-2,\cdots$ and $k=m+1,m+2,\cdots$ selected in such a way that $w(j_k) < 4w(j_{k+1})$ for all values of k. If the range $\{w(j);\ j \in J\}$ was not connected one could enlarge J to render it connected so that the preceding type of inequality would become possible.

Hajek form of Kolmogorov's inequality implies

$$P\left\{\sup\left[|\langle\overline{H}_n, w(j)y_j z\rangle|;\ y_j \leq y_{j_0}\right] > \varepsilon\right\} \leq \frac{2\ s_n^2[y_{j_0}\beta]}{\varepsilon^2}$$

and similar inequalities for $\langle\overline{H}_n, w(j)\ (y_j - y_{j_m})z\rangle$ for $y_j > y_m$.

To complete the proof let $u_k = y_{j_k} - y_{j_{k-1}}$ and $c_k = \sup\{w(j);\ j_{k-1} < j \leq j_k\}$. According to lemma 13

$$g(\varepsilon) \sum_{k=1}^{m} P\left\{c_k|\langle\overline{H}_n,\ y_k\ z\rangle| > \varepsilon\right\} \leq \sum_{k=1}^{m}\int g\left(\frac{c_k}{\sqrt{n}}\, u_k\ z\right)dP_n$$

$$+ \frac{1}{8}\left(1 - \frac{\delta^2}{2}\right)^{-1} \delta^2 \sum_{k=1}^{m} c_k^2\ s_n^2(u_k\ z)$$

with $\delta^2 = \sup_k c_k^2\ s_n^2(u_k\ z)$. The result follows from the inequality

$$c_k\ \dot{u}_k|\dot{z}| \leq 4\ w(j_k)\dot{u}_k|\dot{z}| \leq \dot{\beta}\ \dot{u}_k$$

and an application of lemma 12. The argument for $\overline{G}_n$ is exactly analogous.

<u>Corollary</u>. <u>Let</u> B_n <u>and the system</u> π <u>be as in the preceding proposition</u>. <u>Assume that</u> $\beta_n \in \mathcal{Y}_n$ <u>and let</u> K_n <u>be the polar of</u> B_n <u>in</u> $\mathfrak{X}(B_n)$. <u>Then both</u> $\overline{H}_n$ <u>and</u> $\overline{G}_n$ <u>are tight for the</u> $w\{\mathfrak{X}(B_n), \mathcal{Y}(B_n)\}$ <u>compacts of</u> $\mathfrak{X}(B_n)$. <u>Furthermore, if a process</u> $\overline{H}_n(I - \Pi_S)$ <u>is defined by</u> $\langle\overline{H}_n(I-\Pi_S), y\rangle = \langle\overline{H}_n, (I-\Pi_S)\ y\rangle$ <u>then</u>

$$\tilde{P}\left\{\bar{H}_n(I-\Pi_S) \notin 3\varepsilon K_n\right\} \leqq 2\,\frac{\alpha^2_{n,\pi}}{\varepsilon^2} + \frac{2}{g(\varepsilon)}\left[\eta_{n,\pi} + \frac{1}{4}\,\frac{\delta^2_{n,\pi}\sigma^2_{n,\pi}}{1-\delta^2_{n,\pi}}\right]$$

<u>for the Radon extension</u> $\tilde{P}$ <u>of the measure associated to this process</u>. <u>The corresponding inequalities hold for</u> $\bar{G}_n(1-\Pi_S)$ <u>similarly defined</u>.

For applications to convergence theorems, consider a sequence $\{B_n\}$ where each B_n is the symmetric convex set associated to a system $\{J_n;\ j \longrightarrow y_j;\ w_n;\ z_n\}$. Assume that the essential supremum β_n of the $w_n(j)y_j\ |z_n|$ satisfies the following uniform integrability condition.

(C_5) For every $\varepsilon > 0$ there is a number b such that

a) $$\int_{|\beta_n(\tau)|>b} \beta^2_n(\tau)p_n(d\tau) < \varepsilon.$$

b) If $w_n(j) > b$, then

$$\int y_j(\tau)\beta^2_n(\tau)p_n(d\tau) < \varepsilon.$$

c) If $w_n(j) < b^{-1}$ and $j_1 > j$, then

$$\int \beta^2_n(\tau)(y_{j_1}-y_j)(\tau)p_n(d\tau) < \varepsilon.$$

<u>Proposition</u> 2. <u>Let</u> $\{B_n\}$ <u>be a sequence of convex symmetric subsets</u>, $B_n \subset \mathcal{Y}_n$. <u>Assume that</u> B_n <u>is generated by a system</u> $\{J_n;\ j \longrightarrow y_j;\ z_n;\ w_n\}$ <u>satisfying the condition</u> C_5.

<u>For each</u> n <u>let</u> f_n <u>be a numerical function defined on</u> $\mathcal{X}(B_n)$. <u>Let</u> K_n <u>be the polar of</u> B_n <u>in</u> $\mathcal{X}(B_n)$. <u>Assume that</u> $\{f_n\}$ <u>satisfies the conditions</u>

(C_6) $\sup\{|f_n(x)|;\ x \in \mathcal{X}(B_n);\ n=1,2,\cdots\} \leqq A < \infty$.

(C_7) For every $\varepsilon > 0$ there is an integer N and a $\delta > 0$ such that $n \geqq N$ and $x_i \in \mathcal{X}(B_n)$ and $x_1 - x_2 \in \delta K_n$ implies

$$|f_n(x_1) - f_n(x_2)| < \varepsilon.$$

Let F_n be the normalized empirical distribution

$$F_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\xi_{n,j} - p_n).$$

Let G_n be the normal linear process having the same mean and covariance as F_n. Then

$$\lim_{n \to \infty} E^* f_n(F_n) - E^{**} f_n(G_n) = 0$$

for all positive linear functionals E^* and E^{**} which are extensions of the Radon expectations associated to F_n and G_n respectively.

Proof. If the proposition is valid for a sequence $\{B_n\}$ it is also valid for a sequence $\{B_n'\}$ such that $B_n' \subset B_n$. Thus it is permissible to assume that the measures p_n are nonatomic and that the ranges $\{w_n(j); j \in J_n\}$ and $\{s_n^2(y_j z_n); j \in J_n\}$ are connected. This can always be achieved by enlarging J_n and introducing supplementary indicators y_j.

The conditions C_5 imply that

$$\sup\{w_n^2(j) s_n^2(y_j z_n); j \in J_n, n=1,2,\cdots\} < \infty.$$

Therefore,

$$\sup\{w_n(j) \int |y_j z_n| dp_n; j \in J_n, n=1,2,\cdots\} < \infty.$$

Let Π_n be a projection corresponding to a finite set $S_n \subset J_n$

satisfying the conditions of proposition 1. From the inequalities $w_p(j) \leqq 4w_n(j_k)$ for $j_{k-1} < j \leqq j_k$ used there it follows that $\Pi_n B_n \subset 4 B_n$. <u>If in addition the Π_n have bounded rank</u> the usual central limit theorem implies that the difference between the distributions of $F_n \Pi_n$ and $G_n \Pi_n$ tend to zero.

Because of condition C_5 one can choose sets S_n such that $\alpha^2_{n,\pi} \leqq \varepsilon^4$. Also $\sigma^2_{n,\pi}$ stays bounded and, using the connectedness of the range one can choose the set S_n in such a way that $\delta^2_{n,\pi} \leqq \varepsilon g(\varepsilon)$.

The number of elements $\kappa(S_n)$ necessary for this stays bounded. In this case the $\eta_{n,\pi}$ of proposition 1 tends to zero. Therefore, for every ε' and δ', $\varepsilon' > 0$, $\delta' > 0$ there is a sequence $\{\Pi_n\}$ of projections of bounded rank such that

$$\limsup \tilde{P}\{\bar{H}_n(1 - \Pi_n) \notin \delta' K_n\} < \varepsilon,$$

$$\limsup \tilde{P}\{\bar{G}_n(1 - \Pi_n) \notin \delta' K_n\} < \varepsilon.$$

Since H_n has a concentration function at least equal to that of $\bar{H}_n$ and since $w_n^2(j)\int|y_j z_n|^2 dp_n$ stays bounded, a similar property holds for H_n. It follows that the same property holds for F_n^* and G_n hence also for F_n.

Summarizing, for every $\varepsilon > 0$ and $\delta > 0$ there is a number $k(\varepsilon) < \infty$ and a sequence $\{\Pi_n\}$ of projections having rank less than $k(\varepsilon)$ and such that $\Pi_n B_n \subset 4 B_n$ for which

$$\limsup \tilde{P}\{F_n(1 - \Pi_n) \notin \delta K_n\} < \varepsilon,$$

$$\limsup \tilde{P}\{G_n(1 - \Pi_n) \notin \delta K_n\} < \varepsilon.$$

The assumption made on f_n implies that for n large enough

$$|f_n(F_n) - f_n(F_n \prod_n)| < \varepsilon$$

except for cases having probability at most 2ε.

For any positive extensions of the Radon expectations this implies

$$\limsup_{n\to\infty} |E\, f_n(F_n) - E\, f_n(G_n)| \leq 2A\varepsilon + 2\varepsilon$$

$$+ \limsup_{n\to\infty} |E\, f_n(F_n \prod_n) - Ef_n(G_n \prod_n)|.$$

The Radon expectations are certainly well defined for functions of the type $x \longrightarrow \varphi(x \prod_n)$ where φ is a bounded continuous function on the finite dimensional space $\mathcal{E}_k = \mathcal{X}(B_n)\prod_n$. For such bounded continuous functions $E\, f(F_n \prod_n) - Ef(G_n \prod_n)$ converges to zero. Although the f_n need not be continuous or even measurable the result follows by a standard argument. This completes the proof of the proposition.

To illustrate possible applications of proposition 2 let us mention the following examples.

<u>Example</u> 1. Assume that the measures p_n are all equal and equal to the Lebesgue measure p on the interval $[0, 1]$. Let V_n be the empirical <u>cumulative</u> distribution function corresponding to n independent observations from p. Let V be the cumulative distribution of p and let w be a function defined on $[0, 1]$ such that (1) w is nonnegative, (2) $w(t)$ decreases in t for $t \in [0, \alpha]$, (3) $w(t)$ increases in t for $t \in [\beta, 1]$ with $0 < \alpha < \beta < 1$, (4) $w(t)$ is bounded for $t \in (\alpha, \beta)$, (5) $\int w^2(t)dt < \infty$. For instance, the function $w(t) = [t(1-t)]^{\alpha}$, $-1/2 < \alpha \leq 0$ has

all these properties. Then $\sup_t w(t)\sqrt{n}\ |V_n(t)-V(t)|$ has the same limiting distribution as $\sup_t w(t)|u(t)|$ where u is the Normal process having mean zero and the covariance $E\ u(t)u(s) = \min(s,t) - st$.

This can be seen by applying proposition 2 to the intervals $[0, \alpha]$, (α, β), $[\beta, 1]$ separately.

Example 2. Assume that the measures p_n are probability measures all equal to a measure p on the Euclidean plane. (The k dimensional case can be handled similarly.) Let V_n be the empirical cumulative corresponding to n independent observations from p and let V be the cumulative of p. Let U be a Normal process, defined on the plane, having expectation zero and covariance identical to that of $\sqrt{n}\ (V_n - V)$. Then $\sup_{(x,y)} \sqrt{n}\ |V_n(x,y) - V(x,y)|$ has for limiting distribution the distribution of $\sup_{x,y}|U(x,y)|$. To prove this order the plane by the usual lexicographical order and take this as the totally ordered set J_n of proposition 2.

Example 3. A further example of possible application refers to the Chernoff-Savage statistics. For simplicity we shall use conditions which are much too strong for certain applications but indicate how proposition 2 can be applied.

The functions considered by Chernoff and Savage are of the type

$$T_{m,n} = \beta(m,n)\left\{\int \varphi_{m,n}\left[\alpha_{m,n} f_m(s) + \alpha'_{m,n} g_n(s)\right] df_m(s) - \int \varphi_{m,n}\left[\alpha_{m,n} p_m(s) + \alpha'_{m,n} q_n(s)\right]\right\} dp_m(s)$$

where the β's and α's are suitable numbers and the p_m's and q_n's are probability measures on the line. The function f_m is the empirical cumulative distribution for a sample of m independent observations from p_m. Similarly, g_n is the cumulative obtained from n independent observations from q_n. Both m and n are assumed to increase indefinitely.

Let $h_{m,n} = \alpha_{m,n} p_m + \alpha'_{m,n} q_n$ and let X_m and Y_n be the random functions

$$X_m = \sqrt{m}\ (f_m - p_m),$$

$$Y_n = \sqrt{n}\ (g_n - q_n).$$

Furthermore, let $\Delta_{m,n}$ be defined by

$$\Delta_{m,n} = \frac{1}{\sqrt{m}}\, \alpha_{m,n}\ X_m + \frac{1}{\sqrt{n}}\, \alpha'_{m,n}\ Y_n.$$

With this notation the function $T_{m,n}$ may be written $T_{m,n} = T^{(1)}_{m,n} + T^{(2)}_{m,n} + T^{(3)}_{m,n}$ with

$$T^{(1)}_{m,n} = \beta(m,n) \int \varphi_{m,n}\left[h_{m,n}(s)\right] d\left[f_n(s) - p_m(s)\right],$$

$$T^{(2)}_{m,n} = \beta(m,n) \int \left\{\varphi_{m,n}\left[h_{m,n}(s) + \Delta_{m,n}\right] - \varphi_{m,n}\left[h_{m,n}(s)\right]\right\} dp_m(s)$$

$$T^{(3)}_{m,n} = \beta(m,n) \int \left\{\varphi_{m,n}\left[h_{m,n}(s) + \Delta_{m,n}\right] - \varphi_{m,n}\left[h_{m,n}(s)\right]\right\} d\left[f_m(s) - p_m(s)\right].$$

The first term $T^{(1)}_{m,n}$ is, except for the coefficient $\beta(m,n)$, a sum of independent identically distributed real random variables subject to the usual limit theorems. No elaboration on its behavior is necessary here.

To obtain some indication on the behavior of $T^{(2)}_{m,n}$ and $T^{(3)}_{m,n}$ we shall make the following assumptions.

1) $|\alpha_{m,n}| + |\alpha'_{m,n}|$ stays bounded.

2) $\frac{1}{\sqrt{m}} \beta(m,n)$ and $\frac{1}{\sqrt{n}} \beta(m,n)$ stay bounded.

3) The $\varphi_{m,n}$ satisfy a Lipschitz condition

$$|\varphi_{m,n}(u+v) - \varphi_{m,n}(u)| \le b_0|v|:$$

These assumptions can be relaxed to a noticeable extent by using the full strength of proposition 2. Thus one could replace (3) by local Lipschitz conditions which would allow unbounded derivatives. Also one could assume that the α's are random and replace (1) and (2) by "boundedness in probability" and assume that (3) holds only in a limiting sense. The above conditions, although they are too stringent for some applications are sufficient to indicate the possibilities of the method used here.

First we shall show that under conditions (1), (2), (3), the terms $T^{(3)}_{m,n}$ tend to zero in probability as m and n tend to infinity.

For this purpose, let $\mathfrak{X}$ be the space of bounded measurable functions on the line. Consider $\mathfrak{X}$ as a Banach space for the uniform norm. According to proposition 1, for every $\varepsilon > 0$ there is a finite ν and projections Π_m and Π_n of rank at most equal to ν such that $\|X_m(I-\Pi_m)\| < \varepsilon$ and $\|Y_n(1-\Pi_n)\| < \varepsilon$ except in cases of small probability. The Lipschitz condition (3) implies that it is sufficient to prove that

$$\beta(m,n) \int \left\{\varphi_{m,n}\left[h_{m,n}(s) + \frac{\alpha_{m,n}}{\sqrt{m}}(X_m\Pi_m) + \frac{\alpha'_{m,n}}{\sqrt{n}} Y_n\Pi_n\right] - \varphi_{m,n}[h_{m,n}(s)]\right.$$

$$d[f_m(s) - dp_n(s)]$$

tends to zero in probability.

The term $\alpha_{m,n}(X_m \prod_m)$ may be written $\alpha_{m,n} \Sigma_{j=1}^{\nu} c_{m,j} u_{m,j}$ where the $u_{m,j}$ are elements of $\mathcal{H}$ and the $c_{m,j}$ are random coefficients. One can assume $\|u_{m,j}\| \leq 1$. Also one can assume that $\Sigma_j |c_{m,j}|$ is bounded in probability.

Finally, using again the Lipschitzian character of $\varphi_{m,n}$ it is easily shown that it is sufficient to prove that

$$R_{m,n} = \beta(m,n)\int\left\{\varphi_{m,n}\left[h_{m,n}(s) + \frac{1}{\sqrt{m}} \sum_{j=1}^{\nu} c_{m,j}u_{m,j} + \frac{1}{\sqrt{n}} \sum_{j=1}^{\nu} c'_{n,j}v_{n,j}\right] - \varphi_{m,n}[h_{m,n}(s)]\right\} d[f_m(s) - p_m(s)]$$

tends to zero for functions $u_{m,j}$ and $v_{n,j}$ such that $\|u_{m,j}\| \leq 1$ and $\|v_{n,j}\| \leq 1$ and for random variables $c_{m,j}$ and $c'_{n,j}$ such that not only $\Sigma\{|c_{m,j}| + |c'_{m,j}|\}$ stays bounded but also such that the $c_{m,j}$ take values of the form $k\delta$, k integer $|k| \leq b$ for some $\delta > 0$. One can then classify the values of the original observations $\{\xi_i\}$ $\{\eta_i\}$ into $(\nu b)^2$ sets. On each one of these sets $R_{m,n}$ has the form

$$\frac{1}{m} \sum_{i=1}^{m} \{\rho(\xi_i) - E\rho(\xi_i)\}$$

where ρ is a certain bounded function. Since $|R_{m,n}|$ is smaller than the maximum of these sums it converges to zero in probability.

Thus to study the limiting behavior of $T_{m,n}$ it would be sufficient to be able to describe the limiting behavior of $T_{m,n}^{(2)}$.

<u>Under conditions</u> (1), (2) <u>and</u> (3) <u>the limiting behavior of</u> $T_{m,n}^{(2)}$ <u>is the same as that of</u>

$$\beta(m,n)\int\left\{\varphi_{m,n}\left[h_{m,n}(s) + \frac{\alpha_{m,n}}{\sqrt{m}} u_m(s) + \frac{\alpha'_{m,n}}{\sqrt{n}} v_n(s)\right] - \varphi_{m,n}[h_{m,n}(s)]\right\} dp_m(s)$$

<u>where u_m and v_n are normal processes having expectation zero and the same covariances as X_m and Y_n respectively</u>.

This follows immediately from proposition 2. Condition 3 implies that $\varphi_{m,n}$ is absolutely continuous with respect to the Lebesgue measure and has almost everywhere a derivative $\psi_{m,n}$. Thus, it is tempting to rewrite the foregoing expression as $T^{(4)}_{m,n} + T^{(5)}_{m,n}$ with

$$T^{(4)}_{m,n} = \beta(m,n)\int\left[\frac{\alpha_{m,n}}{\sqrt{m}} u_m(s) + \frac{\alpha'_{m,n}}{\sqrt{n}} v_n(s)\right]\psi_{m,n}[h_{m,n}(s)]\, dp_m(s)$$

and

$$T^{(5)}_{m,n} = \beta(m,n)\int\left\{\varphi_{m,n}[h_{m,n}(s) + w_{m,n}(s)] - \varphi_{m,n}[h_{m,n}(s)] - w_{m,n}(s)\, \psi_{m,n}[h_{m,n}(s)]\right\} dp_m(s),$$

where

$$w_{m,n}(s) = \frac{\alpha_{m,n}}{\sqrt{m}} u_m(s) + \frac{\alpha'_{m,n}}{\sqrt{n}} v_n(s).$$

The term $T^{(4)}_{m,n}$ has a Normal distribution. Under suitable conditions the term $T^{(5)}_{m,n}$ would be expected to tend to zero in probability. Note that to prove that $T^{(5)}_{m,n}$ tends to zero in probability it would be sufficient to show that

$$R_{m,n} = \beta(m,n)\int \left\{ \varphi_{m,n}\left[h_{m,n}(s) + \frac{\alpha_{m,n}}{\sqrt{m}} u_m(s) + \frac{\alpha'_{m,n}}{\sqrt{n}} v_n(s)\right] - \varphi_{m,n}\left[h_{m,n}(s)\right] - \left[\frac{\alpha_{m,n}}{\sqrt{m}} u_m(s) + \frac{\alpha'_{m,n}}{\sqrt{n}} v_n(s)\right] \psi_{m,n}\left[h_{m,n}(s)\right] \right\} dp_m(s)$$

tends to zero if the u_m and v_n are bounded sure functions, $\|u_m\| + \|v_n\| \leq b < \infty$. This can be shown easily by the method used to show that $T^{(3)}_{m,n}$ tends to zero. It follows, for instance, that $T^{(5)}_{m,n}$ tends to zero whenever $\varphi_{m,n}$ is a fixed function φ having a continuous derivative ψ. In more general situations one may have to verify that the p_m measure of the set of points s such that $h_{m,n}(s)$ is close to a discontinuity of $\psi_{m,n}$ is not excessive.

REFERENCES

[1] A. GROTHENDIECK, Sur les applications lineaires faiblement compactes d'espaces du type C(K). Canadian J. Math, 5, 129-173, 1953.

[2] P. ERDÖS, On a lemma of Littlewood and Offord. Bull. Amer. Math. Soc. 51, No. 12, 898-902, 1945.

[3] M. LOÈVE, Probability Theory, 2nd Ed., Van Nostrand, Princeton, New Jersey, 1960.

[4] A. N. KOLMOGOROV, Deux theorèmes asymptotiques uniformes pour des sommes de variables aléatoires. Teoriia Veroiastnosteii lee prim., 1, Part 4, 426-436, 1956.

Additional references

[5] S. BOCHNER, Harmonic Analysis and the Theory of Probability, Univ. of California Press, Berkeley and Los Angeles, 1955, 176 pages.

[6] L. LE CAM, Remarques sur les variables aléatoires dans les espaces vectoriels non separables. Publications de l'Institut de Statistique de l'Universite de Paris, 7, Parts 1-2, 39-53, 1958.

[7] M. D. DONSKER, Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems, Ann. Math. Stat., 23, 277-281, 1952.

[8] H. CHERNOFF and I. R. SAVAGE, Asymptotic normality and efficiency of certain nonparametric statistics. Ann. Math. Stat., 29, 972-994, 1958.

FOOTNOTE

This paper was prepared with the partial support of the Office of Ordnance Research, grant (DA-ARO(D)-31-124-G83).