LAWS OF LARGE NUMBERS
IN REAL AND ABSTRACT SPACES

by

THEOPHILUS OLABODE OGUNYEMI

Thesis submitted in partial fufilment of the require-ments for the Master of Science in

Department of Mathematics,
Faculty of Science,
Ahmadu Bello University,
Z A R I A,

1979

A C K N O W L E D G E M E N T S

I wish to express my profound gratitude to my Supervisor, Dr. W. Timoszyk, for the directives, guidance and the co-ordination received from him.

My gratitude also goes to the authority of the Mathematics Department, in particular, Dr. I.B. Mohammed and Dr. A.R. Kokan, through whose broad-based policy on admission to postgraduate studies I had the opportunity and encouragement to pursue the programme.

I am also grateful to Mr. S.I. Babafemi of the Department for the dedication with which he carried out the typing of the thesis.

Finally, my thanks to to my friends and relations who showed concern over the work and the time of its completion.

T.O. OGUNYEMI.

ABSTRACT

For most people the idea of probability is closely related to that of relative frequency. Therefore, it is natural that an attempt was made to construct a mathematical theory of probability using this concept. The results of the investigations on the relationship between probability and relative frequency are called the laws of the large numbers.

At present, the development of laws of large numbers based on the initial attempt is going on with extreme intensity, the development has gone far beyond the study of random variables in the real space. Infact, the development has become established in some important abstract spaces.

Having described briefly the stage of development of laws of large numbers, I now state the main purpose of this thesis:

1. To give a systematic review of the laws of large numbers in the real space.
2. To approach the proofs, in particular, on the strong laws of large numbers by the application of an important stochastic process - martingales.
3. To study the recent work on the laws of large numbers in some "well-behaved" abstract spaces, Hilbert space and general Banach spaces with "desirable" properties.

It is, however, ragretable that lack of materials on topological and geometrical properties of the abstract spa ce-restricts my urge of going deeper on (3) where the metric properties of normed spaces become a sophisticated tool in proving the law of large numbers.

Finally, it should be remarked that references to books and papers given at the end of the thesis are indicated with [ ] in the course of the work.

OGUNYEMI T.O.

We hereby recommend that the thesis prepared by THEOPHILUS OLABODE OGUNYEMI entitled LAWS OF LARGE NUMBERS IN REAL AND ABSTRACT SPACES be accepted in partial fuflfilment of the requirements for the degree of M.Sc.

1979.

---------------------------- Internal Examiner

---------------------------- External Examiner

DEDICATED TO:

MY LATE DAD
JOHN ANIFOWOSE OGUNYEMI

AND

MY LATE MUM
MARIAN AINA IBILOLA OGUNYEMI

Chapter 1

Definitions and Generalities

1.1 DIFFERENT TYPES OF THE LAWS OF LARGE NUMBERS

If we study the convergence of $n^{-1}S_n$, where $S_n = X_1 + X_2 + \ldots + X_n$, making use of different modes of convergence, then we obtain different types of the laws of large numbers. Many different concepts of convergence are used in probability theory. We shall, however, concentrate on the following types.

Let $X_1, X_2, \ldots,$ be a sequence of random variables.

DEFINITION 1.1.1. X_n converges to X with probability one (or almost surely) if $X_n(\omega) \to X(\omega)$ except on a set of measure zero ω belongs to the probability space Ω. For this kind of convergence, the symbol $X_n \xrightarrow{a.s} X$ is used. In a compact way, $X_n \xrightarrow{a.s} X$ if $P(\lim X_n = X) = 1$.

DEFINITION 1.1.2. X_n converges to X in probability or stochstically if for all $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0.$$

It will be written $X_n \xrightarrow{p} X$.

DEFINITION 1.1.3. X_n converges weakly to X in quadratic mean if X and $X_1, X_2, \ldots\ldots$, are square integrable and if

$$\lim_{n\to\infty} E|(X_n - X)^2| = 0.$$

It is written $X_n \xrightarrow{q.m} X$.

Using the above definitions, we can now give the definitions of the different kinds of laws of large numbers.

DEFINITION 1.1.4. Let $\{X_n\}$ be a sequence of random variables and

$$n^{-1}S_n = \frac{X_1 + \ldots\ldots + X_n}{n}.$$

Then, $\{X_n\}$ (with the sequence of $\{a_n\}$ of real numbers and random variables α) obeys the

(i) weak law of large numbers if

$$n^{-1}S_n - a_n \xrightarrow{p} \alpha;$$

(ii) strong law of large numbers if

$$n^{-1}S_n - a_n \xrightarrow{a.s} \alpha;$$

(iii) mean law of large numbers if

$$n^{-1}S_n - a_n \xrightarrow{q.m} \alpha.$$

Here a_n's are taken as the centering constants, and α as the limit.

It is worthwhile to take a look at the statistical meanings of the modes of convergence mentioned above. We investigate the example of coin tossing.

A coin is tossed n times independently and the random variables X_i, $(i = 1,2,\ldots\ldots, n)$, are defined by

$$X_i = \begin{cases} 1 \text{ if head appears} \\ 0 \text{ if tail appears} \end{cases}$$

In this case,

$$n^{-1}S_n = \frac{X_1 + \ldots\ldots + X_n}{n}$$

is the relative frequency of the heads after the nth tossing. In a concrete sequence of coin tossings, the relative frequencies can be given in a diagram. After n tossings, 2^n different curves of relative frequencies can be obtained depending on the results of the tossings.

A theorem stating that $n^{-1}S_n \longrightarrow 1/2$ in probability essentially states that almost the whole of the 2^n curves are near to $1/2$ if n is large enough but it can happen that all curves strongly deviate from $1/2$ infinitely many times as n tends to infinity.

LEMMA 1.1.1 The following two conditions are equavalent

$$X_n \xrightarrow{a.s} X$$

and $M_N = \sup_{n \geq N} |X_n - X| \xrightarrow{p} 0$ whenever $N \to \infty$.

PROOF. From definition, the first convergence implies the second. The converse statement follows from the facts that M_N is monotonically decreasing in N and any monotone sequence is surely convergent.

REMARKS: In particular, for series of independent random variables, convergence with probability and convergence in probability are equivalent. See [6].

Thus by the above lemma, we can say that $n^{-1}S_n \xrightarrow{a.s} \frac{1}{2}$ essentially states that, for large enough n, the relative frequency curves (except $o(2^n)$) are near to $\frac{1}{2}$ and do not leave a neighbourhood of $\frac{1}{2}$.

1.2 DIFFERENCE BETWEEN THE WEAK LAW AND THE STRONG LAW

Let $X_1, X_2, \ldots\ldots,$ be mutually independent random variables with a common distribution F and $E(X_k) = 0$ for all k.

The weak law of large numbers in this case states that for every $\varepsilon > 0$,

(1.2.1) - - - - $P(|n^{-1}S_n| > \varepsilon) \to 0$.

The fact does not eliminate the possibility that $n^{-1}S_n$ may become arbitrarily large for infinitely many n. In other words, the fact does not imply that $n^{-1}|S_n|$ remains small for all large n; it can happen that the law of large numbers applies but that $n^{-1}|S_n|$ continues to fluctuate between finite and infinite limits.

For example, in a symmetric random walk the probability that the particle passes through the origin at the nth step tends to zero, and yet it is certain that infinitely many such passages will occur. In practice, one is rarely interested in the probability in (1.2.1) for any particular large value of n.

A more interesting question is whether $n^{-1}|S_n|$ will ultimately become and remain small as n becomes large, that is, whether $n^{-1}|S_n| < \varepsilon$ simultaneously for all $n \geq N$. Accordingly, we ask for the probability of the event that $n^{-1}|S_n|$ tends to zero. If this event has probability one, we say that $\{X_k\}$ obeys the strong law of large numbers. In general, the finiteness of the expectations is a necessary condition for the strong law of large numbers.

We can interpret $n^{-1}|S_n| \to 0$ with probability 1 (given $E(X_k) = 0$) roughly by saying that with an overwhelming probability $n^{-1}|S_n|$ remains small for all $n \geq N$.

Apart from the interpretations of the weak law and strong law discussed above, there is another fact that $n^{-1}|S_n| \xrightarrow{p} 0$ can hold also for certain sequences $\{X_k\}$ without expectations. The following example illustrates this fact.

EXAMPLE Let us consider the sequence $\{Z_n\}$, $(n = 1,2,\ldots)$, of independent random variables where

$$P(Z_n = 1) = \frac{1}{n}$$

(1.2.2)-----

$$P(Z_n = 0) = 1 - \frac{1}{n} \ .$$

The sequence converges to zero in probability since from the equality

$$P(|Z_n| > \varepsilon) = P(Z_n = 1),$$

which holds for every $0 < \varepsilon < 1$, we obtain

$$\lim P(|Z_n| > \varepsilon) = \lim \frac{1}{n} = 0.$$

However, the considered sequence $\{Z_n\}$ does not satisfy

(1.2.3) - - - - $P(\lim_{n\to\infty} Z_n = 0) = 1.$

For denoting by A_n the event $\{Z_n = 1\}$, it follows from (1.2.2) that

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

(This show that $\{Z_n\}$ has no finite expectation.)

From the independence of the random variables and from Borel-Cantelli Lemma, (See [6]), it follows that the probability that an infinite number of the A_n's will occur equals one; hence with probability one, there will exist a subsequence of the subsequence $\{Z_n\}$ which is not convergent to zero. This obviously contradicts the relation (1.2.3). Thus $\{Z_n\}$ does not obey the strong law of large numbers.

The next theorem whose proof can be got in [2] shows that in the absence of a finite expectation the sequence of averages $n^{-1}S_n$ is unbounded with probability 1. In other words, a finite expectation is a necessary condition for the strong law of large numbers to hold.

THEOREM 1.2.1 (Converse to the strong law of large numbers.)

Let $X_1, X_2, \ldots$ be independent random variables with a common distribution. If $E(|X_1|) = \infty$, then for any numerical sequence $\{C_n\}$, with probability 1,

$$\lim\sup|n^{-1}S_n - C_n| = \infty.$$

1.3 STABILITY

Definition 1..3.1. If for some constants $a_1, b_1, a_2, b_2, \ldots$

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{i=1}^{n} (X_i - a_i)$$

exists in some sense of convergence, the sequence of rand variables $X_1, X_2, \ldots$ is said to be subject to the law of large numbers (relative to the centering constants $a_1, a_2, \ldots$ and scaling constants $b_1, b_2, \ldots$)

This section is devoted to the brief discussion on sums $S_n = \sum_{i=1}^{n} X_i$ of independent random variables $X_1, X_2, \ldots$ and, especially, on their limit properties - convergence and stability; the case where b_n (scaling factor) is not necessarily equal to n but satisfying the condition $b_n \uparrow \infty$ as $n \to \infty$. Stability property is an important root in the development of probability theory. Moreover, since medians always exist we take the centering constants a_n to be medi So we briefly seek conditions for almost sure stability of sequence S_n.

<u>DEFINITION 1.3.2:</u> Given two numerical sequences a_n and $b_n \uparrow \infty$, the sequence S_n is said to be stable in probability or almost surely if $(S_n/b_n) - a_n \xrightarrow{p} 0$ or $(S_n/b_n) - a_n \xrightarrow{a.s.} 0$.

Before giving the criteria for almost sure stability of S_n, the results of centering at medians and symmetrization will be birefly discussed. For a full discussion, see [6].

<u>Centering at Medians and Symmetrization</u>

Let F be the distribution function ~~of a random variab~~l. X. There exists at least one finite number μX called a median of X, such that

$$P(X \geq \mu X) \geq \frac{1}{2} \leq P(X < \mu X)$$

or, equivalently,

$$F(\mu X) \leq \frac{1}{2}\ ,\ F(\mu X + 0) \geq \frac{1}{2}\ .$$

The symmetrization procedure consists of assigning to a random variable X with characteristic function f a corresponding symmetrized random variable $X^s = X - X'$, where X and X′ are independent and identically distributed, and $f^s = |f|^2$ is the characteristic function of X^s.

<u>Weak Symmetrization Inequalities with Centering at Medians:</u>

For every $\varepsilon > 0$, and every a,

(i) $\frac{1}{2} P[(X - \mu X) \geq \varepsilon] \leq P[X^s \geq \varepsilon]$,

and

(ii) $\frac{1}{2} P[|X - \mu X| \geq \varepsilon] \leq P[|X^s| > \varepsilon] \leq 2P[|X-a| \geq \frac{\varepsilon}{2}]$.

<u>COROLLARY.</u> If $(X_n - a_n) \xrightarrow{P} 0$, then $X_n^s \xrightarrow{P} 0$ and $(a_n - \mu X_n) \longrightarrow 0$, and conversely.

Symmetrization Inequalities

For every $\varepsilon > 0$ and every $a_j, j \leq n$,

(i) $\frac{1}{2} P[\sup_j (X_j - \mu X_j) \geq \varepsilon] \leq P[\sup_j X_j^s \geq \varepsilon]$,

and

(ii) $\frac{1}{2} P[\sup_j |X_j - \mu X_j| \geq \varepsilon] \leq P[\sup_j |X_j^s| \geq \varepsilon]$

$$\leq 2P[\sup_j |X_j - a_j| \geq \tfrac{\varepsilon}{2}].$$

COROLLARY. If $(X_n - a_n) \xrightarrow{a.s.} 0$, then $X_n^s \xrightarrow{a.s} 0$ and $(a_n - \mu X_n) \to 0$, and conversely.

Given sequences a_n and $b_n \uparrow \infty$, we next search for cond[illegible] for almost sure stability of sequence S_n. On account of th[illegible] corollary to the symmetrization inequalities given above, a first condition is that $a_n = \mu(\frac{S_n}{b_n}) + o(1)$.

Thus it suffices to take $a_n = \mu(\frac{S_n}{b_n})$ and investigate conditions under which $\left(\frac{S_n}{b_n} - \mu\left(\frac{S_n}{b_n}\right)\right) \xrightarrow{a.s.} 0$.

We have $b_n \uparrow \infty$ and, moreover, assume that there exists a subsequence b_{n_k} and finite numbers α, β such that, for all k sufficiently large, $1 < \alpha \leq b_{n_{k+1}} \big\backslash b_{n_k} \leq \beta < \infty$.

Roughly speaking, this assumption means that the sequence b_n does not increase too fast, and is always satisfied (wi[illegible] an arbitrary $\beta > 1$) when $b_{n+1}/b_n \to 1$.

Let $S_{n_0} = 0$ and $T_k = \frac{S_{n_k} - S_{n_{k-1}}}{b_{n_k}}$

A.S. STABILITY CRITERION:

(i) $\frac{S_n}{b_n} - \mu\left(\frac{S_n}{b_n}\right) \xrightarrow{a.s} 0$ if, and only if,

(ii) $T_k - \mu T_k \xrightarrow{a.s} 0$ as $k \to \infty$ or, equivalently,

(11)' for every $\varepsilon > 0$,

$$\sum P\left[\,|T_k - \mu T_k| \geq \varepsilon\,\right] < \infty.$$

COROLLARY. If $|X_n| < b_n$, then $\frac{S_n - ES_n}{b_n} \xrightarrow{a.s} 0$ if, and only if, $T_k - ET_k \xrightarrow{a.s} 0$ as $k \to \infty$ or, equivalently, for $\varepsilon > 0$,

$$\sum P(|\,T_k - ET_k| \geq \varepsilon) < \infty.$$

COROLLARY. If the X_n's are centred at expectations and $\sum \frac{\mathrm{var}(X_n)}{b_n^2} < \infty$, then $\frac{S_n}{b_n} \xrightarrow{a.s} 0$.

REMARKS:- See ([6] pp. 252 - 254) for the proofs.

Chapter 2

INDEPENDENT RANDOM VARIABLES

Introduction

In this chpater we shall find necessary and sufficient conditions under which the laws of large numbers hold. The chapter consists of a section on the weak law and the other on the strong law.

Here $X = (X_1, X_2, \ldots, X_n)$ denotes a sequence of indepen... random variables, and

$S_n = X_1 + X_2 + \ldots + X_n$.

A. Weak Law. Under this section, we shall discuss the necessary and sufficient condition for the following various kinds of random variables.

(i) arbitery random variables,

(ii) independent random variables with $E|X_n| < \infty$,

and (iii) independent and identically distributed random variables (i.i.d.).

B. Strong Law. Here we only discuss the sufficient conditi... for the validity of strong law for:

(i) independent random variables with $E|X_n| < \infty$

since a good necessary and sufficient condition is unknown Sufficient and necessary condition will be dealt with in the case of:

(ii) independent and identically distributed random variables.

Some results on the rates of convergence and the Law of Iterated Logarithm will be given.

2.1. WEAK LAW

THEOREM 2.1.1. In order for the sequence $X_1, X_2, \ldots\ldots$ of random variables (not necessarily independent) to satisfy the relation

$$(2.1.1)\text{---}\quad \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| < \varepsilon \right) = 1$$

for any $\varepsilon > 0$, it is necessary and sufficient that

$$(2.1.2)\text{---}\quad E\left(\frac{\sum_{k=1}^{n} (X_k - EX_k)^2}{n^2 + \left(\sum_{k=1}^{n} (X_k - EX_k)\right)^2} \right) \to 0 \quad \text{as } n \to \infty.$$

PROOF: Suppose first that (2.1.2) is satisfied. Let $F_n(x)$ denote the distribution of the variable

$$\alpha_n = \frac{1}{n} \sum_{k=1}^{n} (X_k - EX_k)$$

Thenwe have the following inequalities:

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}(X_k - EX_k)\right| \geq \epsilon\right)$$

$$= P(|\alpha_n| \geq \epsilon)$$

$$= \int_{|x|\geq\epsilon} dF_n(x)$$

$$\leq \frac{1+\epsilon^2}{\epsilon^2}\int_{|x|\geq\epsilon}\frac{x^2}{1+x^2}\,dF_n(x)$$

$$\leq \frac{1+\epsilon^2}{\epsilon^2}\int_{|x|\geq\epsilon}\frac{x^2}{1+x^2}\,dF_n(x)$$

$$= \frac{1+\epsilon^2}{\epsilon^2}\,E\left(\frac{\alpha_n^2}{1+\alpha_n^2}\right).$$

Condition (2.1.2) implies that $\frac{1+\epsilon^2}{\epsilon^2}E\left(\frac{\alpha_n^2}{1+\alpha_n^2}\right) \to 0.$

Thus we have that

$$P\left(\left|\frac{1}{n}\sum(X_k - EX_k)\right| \geq \epsilon\right) \to 0.$$

<u>Necessity.</u> Assume (2.1.1) to hold.

Now,

$$P(|\alpha_n| \geq \varepsilon) = \int_{|x|\geq\varepsilon} dF_n(x)$$

$$\geq \int_{|x|\geq\varepsilon} \frac{x^2}{1+x^2}\, dF_n(x)$$

$$= \int_{\mathbb{R}} \frac{x^2}{1+x^2}\, dF_n(x) - \int_{|x|<\varepsilon} \frac{x^2}{1+x^2}\, dF_n(x)$$

$$\geq \int_{\mathbb{R}} \frac{x^2}{1+x^2}\, dF_n(x) - \varepsilon^2$$

$$= E\left(\frac{\alpha_n^2}{1+\alpha_n^2}\right) - \varepsilon^2.$$

Thus

$$0 \leq E\left(\frac{\alpha_n^2}{1+\alpha_n^2}\right) \leq \varepsilon^2 + P(|\alpha_n| \geq \varepsilon).$$

By first choosing ε sufficiently small and then n sufficiently large, the right - hand side of the last inequality can be made arbitrarily small. The hypothesis that $P(|\alpha_n| \geq \varepsilon) \to 0$ ends the proof.

NOTE: We should note that in more complicated situations, such as when the variables X_k are not assumed to have finite variances, the theorem just proved is of very little use in actually verifying the applicability of the law of large numbers, because condition (2.1.2) applies not to the individual variables but to their sum.

However, whenever the random variables $X_1, X_2, \ldots.$ have finite variances $\sigma_1^2, \sigma_2^2, \ldots$ such that

$$\sum_{1}^{n} \frac{\sigma_i^2}{n^2} \to 0 \text{ as } n \to \infty ,$$

the following theorems are consequences of the above theorem.

THEOREM 2.1.2 Let $X_1, X_2, \ldots.$ be mutually independent random variables with characteristic functions $\Phi_1, \Phi_2, \ldots\ldots$. If $a_j = E(X_j)$ exists for every j, then

$$P\left(\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} (X_j - a_j)\right) = 0 \text{ if, and only if}$$

$$\lim_{n\to\infty} \prod_{1}^{n} |\Phi_j(t|b_n)| = 1$$

uniformly in every finite t - interval.

REMARK, See [1] for the proof.

<u>THEOREM.</u> Let $\Delta(t) = [1 - F(t) + F(-t)]t$,

and $\sigma(t) = \frac{1}{t}\int_{-t}^{t} x^2 \, d\, F(x)$

$$= -\Delta(t) + \frac{2}{t}\int_{0}^{t} x\, \Delta(x)\, dx ,$$

with $\Delta(t) \to 0$ if and only if $\sigma(t) \to 0$.

Let X_k be independent with a common distribution F. In order that there exist constants μ_n such that, for each $\varepsilon > 0$, the probability

(2.1.3)--- $P(|n^{-1}S_n - \mu_n| > \varepsilon) \to 0$, it is necessary and sufficient that $\Delta(t) \to 0$ as $t \to 0$. In this case, (2.1.3) holds with

$$\mu_n = \int_{-n}^{n} x\, F(dx).$$

<u>REMARK.</u> See [2] for the proof.

Before leaving this section on the weak law of large numbers, we look into the estimation of the rate of converge.

2.2 ESTIMATION OF THE RATE OF CONVERGENCE

Let X_k be random variables on some probability space an let $a_{N,k}$ be real numbers.

Define

$$S_N = \frac{1}{N}\sum_{k=1}^{N} X_k$$

and $A_N = \sum_k a_{N,k}X_k$ or $\sum_k a_{N,k}(X_k - EX_k)$

depending on whether $E|X_k|$ is infinite or finite, respectively. (When dealing with A_N we will always have made enough assumptions that the sum makes sense).

This section treats the rates at which $P\{|S_n - \mu| \geq \varepsilon\}$ and $P\{|A_N| \geq \varepsilon\}$ converge to zero. If the random variables X_k's are i.i.d. with $EX_k = \mu$, then the probabilities $P\{|S_N - \mu| \geq \varepsilon\}$ and the rate at which they converge to zero are of interest when estimating μ.

Intuitively, when one has convergence, he would expect its rate to improve with increased averaging.
There is more averaging in the sum

$\frac{X_1}{2} + \frac{X_2}{4} + \frac{X_3}{4}$ than in the sum $\frac{X_1}{2} + \frac{X_2}{2}$, but how does the averaging in the first sum compare with the averaging in the $\frac{X_1}{3} + \frac{X_2}{3} + \frac{X_3}{3}$?

Also intuitively again, success in getting good convergence rates $P\{|A_N| \geq \varepsilon\}$ would seem to depend on how well one can measure the averaing due to the co-efficients $a_{N,k}$ in the sum A_N.

Exponential Rates. A necessary and sufficient condition for exponential convergence rates is given in the work of BAUM, L.E., KATZ, M. and READ R.R. (1961).

THEOREM 2.2.1 Suppose the X_k's are i.i.d. and that $Ee^{\theta|X_k|} < \infty$ for some $\theta > 0$.

Then, the probability

$$P_N(\varepsilon) = P\{|S_N - \mu| \geq \varepsilon\}$$

converges exponentially to 0 for any $\varepsilon > 0$, that is, there exist an $A > 0$ and $0 < \rho < 1$ such that

$$P_N(\varepsilon) \leq A\rho^N$$

if and only if, for all $\varepsilon > 0$, there exists a constant $\theta_\varepsilon > 0$ such that

$$\prod_{k=1}^{N} E(e^{\theta X_k}) \leq E\, e^{|\theta|\varepsilon N}$$

whenever $-\theta_\varepsilon \leq \theta \leq \theta_\varepsilon$.

REMARKS. An optimal ρ may be obtained from $E(e^{\theta X_k})$.

The exponential bounds obtained above have been extended to cover the probabilities $P\{|A_N| \geq \varepsilon\}$. The current "best" theorem along these lines seems to be the following.

THEOREM 2.2.2 If $X_1, X_2, \ldots\ldots$ are independent random variables having finite means and satisfying

$$P\{|X_k - EX_k| \geq x\} < \int_x^{\infty} M e^{-\tau t^p} dt \quad \text{for Some}$$

$M > 0$, $\tau > 0$, $1 < p < 2$, and for all $x > 0$, then there exist constants c_1 and c_2 depending on M, τ, and p such that, for every $\varepsilon > 0$,

$$P\{|A_N| \geq \varepsilon\} < A \exp\{-\min \left[C_1\left(\frac{\varepsilon}{\|a_N\|^2}\right)^2, C_2\left(\frac{\varepsilon}{\|a_N\|_q}\right)^p\right]\}$$

where $\|a_N\|_q$ is the l_q - norm of the sequence $a_N = \{a_{N,k}\}$ and $\frac{1}{p} + \frac{1}{q} = 1$.

REMARKS. It may be noted that $\|a_N\|_2$ and $\|a_N\|_q$ can be thought of as measuring a combination of the overall co-efficient weight and of the amount of averaging which the co-efficients provide. Unfortunately, the bounds in the theorem are not sharp; no bounds with optimality properties have yet been obtained.

Rate related to Moments.

Let X_k be i.i.d. random variables. Results of KATZ, and of BAUM and KATZ provide some theorems giving convergence rates under conditions related to moments.

THEOREM 2.2.3. If $t > 1$ and $E(|X_k|^t) < \infty$, then the following are equivalent:

(a) $E|X_k|^t < \infty$ and $EX_k = \mu$.

(b) $\sum_{n=1}^{\infty} n^{t-2} P\{|S_n - \mu| > \varepsilon\} < \infty$ for all $\varepsilon > 0$.

(c) $\sum_{n=1}^{\infty} n^{t-2} P\{\sup_{k \geq n} |A_k - \mu| > \varepsilon\} < \infty$ for all $\varepsilon > 0$.

An estimation of the rate of convergence of $P\{|S_N - \mu| > \varepsilon\}$ in the case of non-identically distributed random variables was obtained by FRANCK, W.E. and HANSON, D.L. (1966).

But for the convergence of $P\{|A_N| > \varepsilon\}$, we state the following theorem:

THEOREM 2.2.4. Suppose X_k's are independent but identically distributed, and that

$$F(x) = \sup_k P\{|X_k| \geq x\},$$

$$F^*(x) = \sup_k P\{|X_k - EX_k| \geq x\},$$

$$\sum_k |a_{N,k}|^2 \leq \alpha_N \text{ or } C\, N^{-\alpha},$$

and $$\sum_k |a_{N,k}|^t \leq \rho_N \text{ or } C\, N^{-\rho}.$$

(i) If $1 < t < 2$ and $x^t F^*(x) < M < \infty$ for all $x > 0$, then $P\{|A_N| > \varepsilon\} = O(\rho_N)$ for every $\varepsilon > 0$.

(ii) If $1 \leq t < 2$ and $x^t F^*(x) \to 0$ as $x \to \infty$, then $P\{|A_N| > \varepsilon\} = 0\ (\rho_N)$ for every $\varepsilon > 0$.

REMARK. For the proof see [16].

2.3. STRONG LAW

The weak law of large numbers states that under certain conditions the means of a sequence of independent random variables with 0 means converge in probability to 0. The question arises as to whether we can make a probability statement that all means in the sequence of means from some point sufficiently far out in the sequence are arbitrarily close to zero. This is answered by the strong law of large numbers.

THEOREM 2.3.1 Let $X_1, X_2, \ldots\ldots,$ be a sequence of independent random variables having 0 means and variances $\sigma_1^2, \sigma_2^2, \ldots.$ such that $\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} < +\infty$.

If $\bar{X}_n = (X_1 + X_2 + \ldots + X_n)/n$, then $(\bar{X}_1, \bar{X}_2, \ldots\ldots,)$ converges almost certainly to 0.

That is, for arbitrary $\delta > 0$ and $\varepsilon > 0$, there is an $N_{\delta,\varepsilon}$ such that $P\{|\bar{X}_n| < \varepsilon,\ n = N, N+1, \ldots N+k\} \geq 1-\delta$ for all $N \geq N_{\delta,\varepsilon}$ and for every k.

PROOF Let $E_N, E_{N+1}, \ldots\ldots, E_{N+k}$ be the events

$\{|\bar{X}_n| < \varepsilon,\ n = N, N+1, \ldots, N+k\}$, respectively.

Then

$$P\{|\bar{X}_n| < \varepsilon,\ n = N, N+1, \ldots\ldots, N+k\}$$

$$= P\{E_N \cap E_{N+1} \cap \ldots\ldots \cap E_{N+k}\}. \qquad (2.3.1)$$

But this probability has the value

$$1 - P\{E'_N \cup E'_{N+1} \cup \ldots\ldots \cup E'_{N+k}\}. \qquad (2.3.2)$$

Thus, for $N > N_{\delta,\varepsilon}$ and k, we must show that

$$P\{E'_N \cup E'_{N+1} \cup \ldots \cup E'_{N+k}\} < \delta. \qquad (2.3.3)$$

Let us partition the positive integers into sets $I_1, I_2, \ldots$, where I_α is the set $\{2^{\alpha-1}+1,\ 2^{\alpha-2}+2, \ldots, 2^\alpha\}$. Let F_α be the event for which at least one of the inequalities $\{|\bar{X}_n| < \varepsilon,\ n \in I_\alpha\}$ fails. Then, for some α and β, we have

$$F_\alpha \cup F_{\alpha+1} \cup \ldots \cup F_{\alpha+\beta} \supset E'_N \cup E'_N \cup \ldots \cup E'_{N+k}. \qquad (2.3.4)$$

Also, we have

$$P\{F_\alpha \cup F_{\alpha+1} \cup \ldots\ldots \cup F_{\alpha+\beta}\} \leq P\{F_\alpha\} + \ldots + P\{F_{\alpha+\beta}\}. \qquad (2.3.5)$$

Thus it is enough for $\sum_{\alpha=1}^{\infty} P\{F_\alpha\}$ to Converge.

Now F_α is the event that for at least one n in I_α we have $|\bar{X}_n| \geq \epsilon$, which may be written $|X_1 + ... + X_n| \geq n\,\epsilon$ which implies that

$$|X_1+X_2+....+X_n| > \left(\frac{\epsilon \,.\, 2^{\alpha-1}}{C_{2^\alpha}}\right) C_{2^\alpha} \,.$$

where $C^2_{2^\alpha} = \sigma_1^2 + ... + \sigma^2_{2^\alpha}$.

But if follows from Kolmogorov's Inequality, See [10], that

$$P\{F_\alpha\} \leq \frac{4\, C^2_{2^\alpha}}{\epsilon^2 . 2^{2\alpha}} \,.$$

Reversing the order of summation with respect to i and α and observing that $\sum 2^{-2\alpha}$ over all positive integers α for which $2^\alpha \geq i$ does not exceed $2i^{-2}$, we obtain

$$\sum_{\alpha=1}^{\infty} P\{F_\alpha\} \leq \frac{8}{\epsilon^2} \sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} \,.$$

Hence $\sum_{\alpha=1}^{\infty} P\{F_\alpha\}$ converges if $\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2}$ does, in which case we can make

$P\{F_\alpha\} + ... + P\{F_{\alpha+\beta}\} < \delta$ for all values of β by choosing α sufficiently large. It then follows from (2.3.4) and (2.3.5) that for sufficiently large N (2.3.3) holds.

This implies that

$$P(|\bar{X}_n| < \varepsilon,\ n = N, N+1, \ldots, N+k) \geq 1-\delta \text{ for all}$$

$N > N_{\delta,\varepsilon}$ and for k, and this completes the proof.

REMARKS: It should be noted that if all components in $(X_1, X_2, \ldots)$ have equal (finite) variances the condition $\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} < +\infty$ is automatically satisfied. As a matter of fact, it will be seen that $(\bar{X}_1, \bar{X}_2, \ldots)$ converges almost certainly to 0 if $X_1, X_2, \ldots,$ are independent and identically distributed with 0 means. See [2].

2.4. THE LAW OF ITERATED LOGARITHM (LIL)

In connection with the resulting types of laws of large numbers we are investigating the rate of convergence in this chapter. In the case of the convergence in probability on which the weak law is defined the definition of the rate of convergence is clear.

In the case of convergence with probability the rate of convergence will be characterised by the "largest" function f(n) for which

$$P(f(n) \mid \frac{S_n}{n} - a_n \mid = 0) = 1.$$

More precisely, we investigate next the class of functions f(n) for which the formula above holds.

The classical result of KOLMOGOROV, A.N. (1929) will be given for bounded independent random variables; the theorem of HARIMAN - WINTNER (1941) for identically distributed random variables will be given. The weighted i.i.d. random variables will also be touched. The general case of unbounded independent random variables has proved extremely elusive but the main finding which furnishes sufficient conditions that are no more stringent than finite variance in the i.i.d. case was treated by HENRY TEICHER (1974) and the result will be stated.

<u>THEOREM 2.4.1</u> (FELLER (1943)) Let $X_1, X_2, \ldots,$ be a sequence of bounded independent random variables with $|X_n| = O(s_n \backslash (\log\log s_n^2)^{\frac{1}{2}})$ where $EX_n = 0$, $EX_n^2 = \sigma_n^2$, $s_n^2 = \sum_{i=1}^{n} \sigma_i^2 \longrightarrow \infty$. Then the Law of Iterated Logarithm holds for $\{X_n\}$, that is,

$$P\left\{\limsup_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{s_n(2\,\mathrm{loglog}\, s_n^2)^{\frac{1}{2}}} = 1\right\} = 1.$$

REMARKS. For proof, see [6].

MARCINKIEWICZ and ZYGMUND (1937) noted that if $\{X_n\}$ are independent with

$$P\{X_n = \pm \frac{\exp(\lambda n \log n)}{(\log n)^{\frac{1}{2}}}\} = \frac{1}{2},$$

then the L.I.L fails provided the positive parameter λ is not too small.

In the case of i.i.d. random variables, Hartman and Wintner proved that the existence of a second moment is a sufficient condition for L I L.

THEOREM 2.4.2. Let $X_1, X_2, \ldots\ldots,$ be a sequence of i.i.d. random variables such that $EX_n = 0$, $EX_n^2 = 1$, then

$$P\left\{\limsup_{n\to\infty} \frac{X_1 + \ldots\ldots + X_n}{(2n \log\log n)^{\frac{1}{2}}} = 1\right\} = 1$$

REMARKS. It has been shown by STRASSEN [14] that the existence of a second moment is a necessary condition in the sense that

$$P\left\{\limsup_{n\to\infty} \frac{\left|\sum_{i=1}^{n} X_j\right|}{(n \log \log n)^{\frac{1}{2}}} = \infty\right\} = 1$$

when the variance X_1 is infinite.

<u>Weighted i.i.d. case</u>

Let Q denote the class of sequence $\{\sigma_n Y_n, n \geq 1\}$ where $\sigma_n \neq 0$, $n \geq 1$, $s_n^2 = \sum_1^n \sigma_j^2 \to \infty$, and where $\{Y_n\}$ are i.i.d. random variables with mean zero, variance $\sigma_Y^2 < \infty$ and the distribution F. Then, in Q, the necessary condition for two-sided L I L (i.e., symmetric L I L holding for $\{X_n\}$ and $\{-X_n\}$)

(2.4.1) --- $\sum_{n=1}^{\infty} P\{|X_n| > \delta\, s_n (\log\log s_n^2)^{\frac{1}{2}}\} < \infty$, where

$\delta > 2(2)^{\frac{1}{2}}$, becomes

(2.4.2)--- $\sum P\{ Y_1^2 > \delta \dfrac{n \log \log s_n^2}{\alpha_n} \} < \infty,\ \delta > \infty,$

where $\alpha_n = \dfrac{n\sigma_n^2}{s_n^2}$.

When α_n increases more rapidly than $\log_2 s_n^2$, (2.4.2) asserts that something beyond a finite second moment is necessary for the two-sided L I L in Q.

The next theorem, extending Strassen's necessity to weighted i.i.d. random variables, stipulates that nothing less than finite variance will do even when $\alpha_n = O(1)$ or more generally when

$$\alpha_n = O\left(\frac{n}{\log \log s_n^2}\right).$$

THEOREM 2.4.3. Let $\{\sigma_n, n \geq 1\}$ be non-zero constants satisfying $s_n = \sum_1^n \sigma_i^2 \to \infty$,

$$\sigma_n^2 = O\left(\frac{s_n^2}{\log \log s_n^2}\right),$$

If $\{Y, Y_n, n \geq 1\}$ are i.i.d. with $EY = 0$, $EY^2 = \infty$, then

$$P\left\{\limsup \frac{\left|\sum_{j=1}^n \sigma_j Y_j\right|}{s_n(\log \log s_n^2)^{\frac{1}{2}}} = \infty\right\} = 1.$$

REMARKS. See [14] for the proof.

Unbounded Case.

Next is the result for unbounded independent random variables which is fully discussed in [14].

THEOREM 2.4.4. If $\{X_n, n \geq 1\}$ are independent random variables with $EX_n = 0$, $EX_n^2 = \sigma_n^2$, $s_n^2 = \sum_{i=1}^n \sigma_i^2 \to \infty$, satisfying for some

$\delta > 0$,

(i) $\sum_{n=1}^{\infty} P\{|X_n| > \delta\, s_n(\log \log s_n^2)^{\frac{1}{2}}\} < \infty$

(ii) $\frac{1}{s_n^2} \sum_{j=1}^n \int_A x^2\, dF_j(x) = o(1)$ for all $\varepsilon > 0$, where $A = \{|X| > \varepsilon\, s_j\, (\log\log s_j^2)^{-\frac{1}{2}}\}$

(iii) $\sum_{n=1}^{\infty} \frac{1}{s_n^2(\log\log s_n^2)} \int_B x^2\, dF_n(x) < \infty$ for all $\varepsilon > 0$,

where $B = \{\varepsilon\, s_n(\log s_n^2)^{-\frac{1}{2}} < |X| \leq \delta s_n(\log\log s_n^2)^{\frac{1}{2}}\}$, then

the L I L holds for X_n.

That is,

$$P\left(\limsup_{n \to \infty} \frac{\sum_{j=1}^{n} X_j}{s_n(2\log\log s_n^2)^{\frac{1}{2}}} = 1\right) = 1.$$

REMARK Since all the three requirements involve only $|X_n|$, they imply the L I L for $\{-X_n\}$ as well.

Chapter 3

ORTHOGONAL RANDOM VARIABLES

INTRODUCTION. In the practical and theoretical applications of probability theory it is very difficult to check whether our random are independent. It is much easier to check the orthogonality of random variables. Hence it is very important to find the analogues of the theorems of Chapter 2 for the case in which the random variables are not independent but orthogonal.

3.1 PRELIMINARIES

The random variables are taken to be complex - valued and two random variables equal almost everywhere will be considered identical.

Let X and Y be random variables with $E|X^2| < \infty$, $E|Y^2| < \infty$. Then, if $E(X\bar{Y}) = E(X)\,E(\bar{Y})$, where $\bar{Y}$ is the complex conjugate of Y, the random variables are said to be uncorrelated; if $E(X\bar{Y}) = 0$, they are said to be orthogonal. Further, if X and Y are uncorrelated, $(X - EX)$ and $(Y - EY)$ are orthogonal.

Since the independent random variables in $\mathcal{L}_2$ (space of square-integrable functions) are orthogonal when centred at expectations, we assume, for the sake of analogy, that all our random variables are centred at expectations, unless otherwise stated. Since our random variables have finite second, hence, first moments, they can always be so centred and the assumption made does not restrict the generality. $E|X|^2$ is the variance of X and $E(X\bar{Y})$ is the covariance of X and Y.

If $X_1, X_2, \ldots\ldots\ldots$ are orthogonal random variables, then

$$E\left|\sum_{j=1}^{n} X_j\right|^2 = \sum_{j=1}^{n} E|X_j|^2, \text{ where}$$

$n = 1,2,\ldots\ldots,$ and as $n \to \infty$, we have

$$E\left|\sum_{j=1}^{n} X_j\right|^2 = \sum_{j=1}^{\infty} E|X_j|^2.$$

A fundamental inequality of the theory of orthogonal series which plays a role similar to that of the Kolmogorov Inequality in the theory of independent random variables is the following:

<u>THEOREM 3.1.1.</u> (RADEMARCHER - MENSOV INEQUALITY)

If $S_n = \sum_{j=1}^{n} X_j$ are consecutive sums of orthogonal random variables,

then

$$E\left(\max_{1\geq n<h} |S_n|^2\right) < \frac{\log 4h}{\log 2}^2 \sum_{j=1}^{n} E|X_j|^2.$$

REMARK. See [6] for the proof.

3.2. RESULTS ON MUTUALLY ORTHOGONAL RANDOM VARIABLES

We show below how little the qualitative hypotheses of some of the theorems on the law of large numbers for mutually independent random variables need be strengthened if only mutual orthogonality is presupposed.

THEOREM 3.2.1. Let the random variables X_n be orthogonal with $EX_n = 0$.

(i) If $\sum(\log^2 n)\, E|X_n|^2 < \infty$, then the series $\sum X_n$ converges in quadratic mean and with probability one.

(ii) If $\sum\left(\frac{\log n}{b_n}\right)^2 E|X_n|^2 < \infty$, $b_n \to \infty$,

then $\frac{1}{b_n} \sum_{j=1}^{n} X_j \to 0$ almost surely.

PROOF

We first state some lemmas whose proofs can be got in [6].

LEMMA 3.2.1. Let the random variables X_n be orthogonal.

(i) Then the series $\sum X_n$ converges in quadratic mean if and only if $\sum E|X_n|^2 < \infty$, and then

$$E|\sum X_n|^2 = \sum E|X_n|^2.$$

(ii) If $\sum \frac{E|X_n|^2}{b_n^2} < \infty$, $b_n \to \infty$, we have that

$$\frac{1}{b_n}\sum_{j=1}^{n} X_j \to 0 \text{ in q.m.}$$

LEMMA 3.2.2. If $S_n = \sum_{j=1}^{n} X_j$ are consecutive sums of orthogonal random variables and $\sum b_n E|X_n|^2 < \infty$, $b_n \to \infty$, then $S_n \to S$ in quadratic mean, and there exists a subsequence $S_{n_j} \to S$ with probability one such that, for every integer j, b_{n_j} be the first $b_n \geq j$.

Now, let $S_n = \sum_{j=1}^{n} X_j$. Under hypothesis (i) of the theorem $S_n \to S$ in q.m. according to Lemma (3.2.2) with b_n $b_n = \log n/\log 2$ yields $S_{2^j} \to S$ almost surely, thus (i) of the theorem will follow if we prove that

$$T_j = \max_{2^j \leq n < 2^{j+1}} |S_n - S_{2^j}| \to 0 \quad \text{almost surely.}$$

But by the Rademacher - Mensov Inequality, with $n = 2^{j+1} - 2^{j}$ yields, by elementary computation,

$$\sum_{j=1}^{\infty} E|T_j|^2 < (3/\log 2)^2 \sum_{n=1}^{\infty} \log^2 n \, E|X_n|^2 < \infty,$$

and the assertion follows from Borel-Cantelli lemma and Chebyshev Inequality. This proves (i), and (ii) follows by Kronecker's lemma.

Chapter 4

MARTINGALES

INTRODUCTION: In this chapter we will treat the application of martingales to the strong law of large numbers for arbitrary random variables and i.i.d. random variables.

A perusal of the former proofs reveals that the assumed independence of the variables was used only to derive certain inequalities among expectations, and hence we want to carry the main results over to martingales and submartingales. Such generalizations are important for many applications and they throw new light on the nature of theorems. See Example 2 below.

The concept of martingales is due to P.Levy but it was J.L. DOOB (1949) who started developing the theory.

4.1 DEFINITIONS OF MARTINGALES AND SEMI-MARTINGALES

DEFINITION 4.1.1. A stochastic process is the mathematical abstraction of an empirical process whose development is governed by probabilistic laws.

DEFINITION 4.1.2. A stochastic process $\{X_t, t \in T\}$, where X_t is in practice the observation at time t, and T is the time range involved, is called a martingale if $E|X_t| < \infty$ for all $t \in T$ and if, whenever $n \geq 1$ and $t_1 < \ldots\ldots < t_{n+1}$,

(4.1.1)---- $E\{X_{t_{n+1}} | X_{t_1}, \ldots\ldots, X_{t_n}\} = X_{t_n}$ with probability 1.

This is a strict-sense definition and martingale will always mean martingale in this strict sense.

A stochastic process with variables $\{X_t\}$ is called a martingale in the wide sense if $E|X_t|^2 < \infty$ for all $t \in T$ and if, whenever $n \geq 1$ and $t_1 < \ldots\ldots < t_{n+1}$,

$$E\{X_{t_{n+1}} | X_{t_1}, \ldots\ldots, X_{t_n}\} = X_{t_n} \text{ with probability one.}$$

EXAMPLES.

(1) Let $\eta, \xi_1, \xi_2, \ldots\ldots$ be any random variables with $E|\eta| < \infty$. If X_n is defined by

$X_n = E\{\eta | \xi_1, \ldots\ldots\ldots, \xi_n\}$, then the process is a martingale.

Infact,

$$E\{X_{n+1} | \xi_1, \ldots\ldots\ldots, \xi_n\}$$
$$= E\{E[\eta | \xi_1, \ldots\ldots\ldots, \xi_{n+1}] | \xi_1, \ldots\ldots, \xi_n\}$$
$$= E\{\eta | \xi_1 \ldots\ldots\ldots\ldots, \xi_n\}$$
$$= X_n.$$

That is, $E\{X_{n+1}|\xi_1,\ldots\ldots,\xi_n\} = X_n$ with probability 1.

Hence, since $X_1,\ldots\ldots,X_n$ are random variables on the sample space of $\xi_1,\xi_2,\ldots\ldots,\xi_n$, we have

$$E\{X_{n+1}|X_1,\ldots\ldots, X_n, \xi_1,\ldots\ldots,\xi_n\}$$

$$= E\{X_{n+1}|\xi_1,\ldots\ldots,\xi_n\}$$

$$= X_n \text{ with probability 1.}$$

Taking the conditional expectation of both sides with $X_1,\ldots\ldots, X_n$ fixed gives

$$E\{X_{n+1}\ |X_1,\ldots\ldots, X_n\} = X_n.$$

(2) Let $S_n = \sum_{j=1}^{n} Y_j$, $n = 1,2,\ldots,$. If the random variables Y_j are independent with $E|Y_j| < \infty$. Then the S_n process is a martingale if and only if $E\{Y_j\} = 0$ for $j > 1$, a semi-martingale if and only if the Y_j's are real and $E\{Y_j\} \geq 0$ for $j > 1$. See [1]. Furthermore, if the variances exist, $\{S_n^2\}$ is a semi-martingale.

We are now going to give a more flexible definition by considering the definition of a martingale in terms of σ-field.

<u>DEFINITION 4.1.3.</u> Let $Y_1,Y_2,\ldots\ldots,$ be random variables with expectations. Let $\mathbb{B}_1$, $\mathbb{B}_2,\ldots\ldots$ be σ-algebras of event satisfying $\mathbb{B}_1 \subset \mathbb{B}_2 \subset \ldots\ldots\ldots\ldots\ldots\ldots$

The sequence $\{Y_n\}$ is a martingale with respect to $\mathbb{B}_n$ if and only if

(4.1.2)---- $E\{Y_{n+1} | \mathbb{B}_n\} = Y_n$.

REMARKS (i) In most cases $\mathbb{B}_n$ will be generated by Y $Y_1, \ldots, Y_n$ and additional random variables depending on the past. The idea is that any random variable depending on the past must be measurable with respect to $\mathbb{B}_n$, and in this sense $\mathbb{B}_n$ represents the information contained in the past history of the process. As this information grows richer with time we shall suppose that the $\mathbb{B}_n$ increases, that is,

$$\mathbb{B}_1 \subset \mathbb{B}_2 \subset \mathbb{B}_3 \subset \ldots\ldots\ldots\ldots\ldots$$

(ii) It is to be noted that (4.1.2) implies that Y_n is $\mathbb{B}_n$ - measurable, and this has two important consequences. Since $\mathbb{B}_n \supset \mathbb{B}_{n-1}$, then iterated expectations' identity gives

$$E(E(Y_{n+1} | \mathbb{B}_n) | \mathbb{B}_{n-1}) = E(Y_{n+1} | \mathbb{B}_{n-1}) = Y_{n-1}.$$

By induction, it follows that

$$E(Y_{n+1} | \mathbb{B}_j) = Y_j \text{ where } j = 1,2,\ldots, n.$$

It follows in particular that every subsequence $Y_{\nu_1}, Y_{\nu_2}, \ldots\ldots$ of a martingale is again a martingale.

Next we note that since $\mathcal{B}_n$ contains the σ-algebra purely generated only by the variables $Y_1, Y_2, \ldots,$ the iterat expextations' identity shows that (4.1.2.) implies) (4.1.1).

DEFINITION 4.1.4. The sequence $\{Y_n\}$ is a semi-martingale if it satisfies the martingale definition (4.1.2) or (4.1.3) with the equality sign replaced by $\geq$.

DEFINITION 4.1.5. (Martingale Equality).

With the definition of conditional expectations, and changing the notation, (4.1.1) is equivalent to

$$\int_{\Lambda} X_t \, dP = \int_{\Lambda} X_s \, dP, \; s < t, \text{ for}$$

every ω-set Λ determined by conditions on a finite number of X_r's with $r \leq s$. This equation is called the Martingale Equality.

Replacing the equality by $\geq$ gives the Semi-Martingale Inequality.

4.2. APPLICATION TO THE STRONG LAW OF LARGE NUMBERS

We will now consider the application of martingale to the proofs of laws of large numbers for a sequence of random variables and for i.i.d. random variables. Before going into the theorems involved some other important lemmas and theorems will first be dealt with.

LEMMA 42.1. If u is a convex function and $\{Y_n\}$ a martingale, the $\{u(Y_n)\}$ is a submartingale provided the expectation of $u(Y_n)$ exists. In particular, $\{|Y_n|\}$ is a semi-martingale.

PROOF. The proof is immediate from Jensen's Inequality which applies to conditional expectations as well as to ordinary ones. It states that

$$E(u(Y_{n+1}) \mid \mathcal{B}_n) \geq u(E(Y_{n+1} \mid \mathcal{B}_n)),$$

and the right side equals $u(Y_n)$.

REMARKS. The same proof shows that if $\{Y_n\}$ is a semi-martingale and u a convex non-decreasing function, the $\{u(Y_n)\}$ is again a semi martingale, provided $u(Y_n)$ has an expectation.

THEOREM 4.2.1 (Kolmogorov's Inequality for Martingales.) If $Y_1, \ldots\ldots, Y_n$ constitute a martingale, then for $t > 0$,

$$P\{\max_{k \geq n} |Y_k| > t\} \leq t^{-1} E|Y_n|.$$

Next we prove some convergence theorems of martingales (wide-sense).

THEOREM 4.2.2. Let Z be a random variable with a finite second moment and let $\ldots\ldots \subset M_1 \subset M_2 \subset \ldots\ldots$, be closed linear manifold of random variables.

Let $M_{-\infty} = \bigcap_n M_n$ and let M_∞ be the closed linear manifold generated by the random variables in $\bigcup_n M_n$.

Then the random variables $E\{Z|M_{-\infty}\}$, $E\{Z|M_{-\infty+1}\}$,......., $E\{Z|M_1\}$,......, $E\{Z|M_\infty\}$ constitute a martingale (wide-sense), and

$$\text{(i)} \quad \lim_{n\to-\infty} E\{Z|M_n\} = E\{Z|M_{-\infty}\}$$

(4.2.1)---

$$\text{(ii)} \quad \lim_{n\to\infty} E\{Z|M_n\} = E\{Z|M_\infty\} \quad \text{with probability one.}$$

PROOF. The proof follows from the following two lemmas whose proofs can be got from [1].

LEMMA 4.2.1. Let ....,X_{-2},X_{-1} be random variables constituting a martingale (wide-sense). Let M_n be the closed linear manifold generated by ...., X_{n-1},X_n, and let

$$M = \bigcap_{-\infty}^{-1} M_n .$$

Then

$$\lim_{n\to\infty} X_n = E\{X_{-1}|M\}$$

$= X_{-\infty}$, and the random variables

$X_{-\infty}$,......,X_{-2},X_{-1} constitute a martingale (wide-sense).

LEMMA 4.2.2. Let X_1,X_2,......, be random variables constituting a martingale (wide-sense).

Then,

$$E\{|X_1|^2\} \leq E\{|X_2|^2\} \leq \ldots\ldots\ldots, .$$

If $\lim_{n\to\infty} E\{|X_n|^2\} = 1 < \infty$, then $\lim_{n\to\infty} X_n = X_\infty$ exists and the random variables $X_1, X_2, \ldots\ldots, X_\infty$ constitute a martingale (wide-sense).

REMARKS. By Lemma (4.2.1) the first limit in (4.2.1) follows. To show that the limit X of X_n is the projection $X_{-\infty}$ we note that the projection $X_{-\infty}$ is characterised by two conditions: it is in $M_{-\infty}$ (as is X since $X_n \in M_n$), and $Z - X_\infty$ is orthogonal to $M_{-\infty}$ (as is Z - X since $Z - X_n$ is orthogonal to M_n and therefore to $M_{-\infty}$). This completes the proof (4.2.1. (i)).

To prove (4.2.1 (ii)) we note that ....
$\leq E|X_1|^2 \leq \ldots\ldots\ldots\ldots \leq E|X_\infty|^2$ since X_t is a martingale sequence in the wide sense, so that the mean limit in (ii) exists by Lemma (4.2.2).

COROLLARY 4.2.1. Let Z, $Y_1, Y_2, \ldots\ldots$, be any random variables with finite second moments. Then if M_n is the closed linear manifold generated by Y_j's with $j \geq n$,

$$\text{(i)} \quad \lim_{n\to\infty} E\{ Z|Y_n, Y_{n+1}, \ldots\} = E\{Z|\bigcap_1^\infty M_n\}$$

(4.2.2)----

$$\text{(ii)} \quad \lim_{n\to\infty} E\{Z|Y_1,\ldots\ldots,Y_n\} = E\{Z|Y_1,Y_2,\ldots\}$$

with probability one. In particular, if Z is a random variable in a closed linear manifold generated by the Y's, the second limit can be replaced by Z.

PROOF. It follows on reducing (4.2.2 (i)) to (4.2.1 (')) by identifying M_n with M_{-n} in (4.2.1), and (4.2.2 (ii)) reduced to (4.2.1 (ii)) by identifying with M_n in (4.2.1 (ii)) the closed linear manifold generated by $Y_1,\ldots\ldots\ldots, Y_n$.

We are now in a position to apply the convergence theorem of wide-sense martingale to the strong law of large numbers for a sequence of random variables under certain conditions.

THEOREM 4.2.2. Let $\{X_n\}$ be a sequence of random variables such that $E\{X_n|M_{n-1}\} = 0$ for all n, where M_n is generated by X_j's with $j \geq n$. If $b_1 < b_2 < \ldots\ldots < +\infty$ and

(4.2.3)---- $\sum b_j^{-2}E(X_j^2) < \infty$, then with probability one,

$$\text{(4.2.4)-----} \quad \frac{X_1 + X_2 + \ldots\ldots\ldots + X_n}{b_n} \to 0.$$

PROOF We have to apply the equations in (4.2.2) with the following identifications:

$$Z \equiv X_1$$

$$Y_n = \sum_{j=1}^{n} b_j^{-1} X_j.$$

It is easily seen that $\{Y_n\}$ is a martingale sequence and that $E(Y_n^2)$ is bounded by the series in (4.2.3).

The preceding corollary therefore guarantees the almost sure convergence of $\{Y_n\}$, and by Kronecker's lemma this implies the assertion (4.2.4).

Before stating and proving the theorem on the application of strict-sense martingales to the strong law of large numbers for i.i.d. random variables, we shall first state the strict-sense versions of the convergence theorem of wide-sense martingales, Theorem (4.2.2.), and its Corollary (4.2.1).

<u>THEOREM 4.2.3.</u> Let Z be a random variable with $E|Z| < \infty$, and let $\cdots\subset M_1 \subset M_2 \subset \cdots$ be Borel fields of ω-sets. Let

$M_{-\infty} = \bigcap_n M_n$, and let M_∞ be the smallest Borel field of sets with $M_\infty \supset \bigcup_n M_n$.

Then

$$\text{(i)} \quad \lim_{n\to-\infty} E\{Z|M_n\} = E\{Z|M_{-\infty}\}$$

(4.2.5)---

$$\text{(ii)} \quad \lim_{n\to\infty} E\{Z|M_n\} = E\{Z|M_\infty\}$$

with probability one.

PROOF. See [1]. But it must be remarked that, if M_n is defined for sufficiently large or sufficiently small n, then (ii) or (i) of Theorem (4.2.3), respectively, remains applicable.

The following corollary covers the most important case of the above theorem, and it is the strict-sense version of Corollary (4.2.1).

COROLLARY 4.2.2 Let Z be any random variable with $E|Z| < \infty$, and let $Y_1, Y_2, \ldots\ldots.$ be any random variables. Then if $\mathcal{G}_n$ is the Borel field of the ω-sets determined by conditions on the Y_j's with $j \geq n$,

$$\text{(i)} \quad \lim_{n\to\infty} E\{Z|Y_n, Y_{n+1}, \ldots\} = E\{Z| \bigcap_1^\infty \mathcal{G}_n\}$$

(4.2.6)----

$$\text{(ii)} \quad \lim_{n\to\infty} E\{Z|Y_1, Y_2, \ldots, Y_n\} = E\{Z|Y_1, Y_2, \ldots\}$$

with probability one. In particular, if Z is a random variable on the sample space of the Y_j's, the second limit can be replaced by Z.

PROOF. To reduce (4.2.6 (i)) to (4.2.5 (i)) identify $\mathcal{G}_n$ with M_{-n}. To reduce (4.2.6 (ii)) to (4.2. 5 (ii)) identify with M_n the Borel field of ω-sets determined by conditions on the Y_j's for $j \geq n$.

In particular, if Z is a random variable on the sample space of the Y_j's, the second limit is Z itself, with probability 1, by the definition of conditional expectation.

The next theorem is an immediate consequence of Theorem (4.2.3), but its form makes it more useful in studying certain problems.

THEOREM 4.2.4. (i) Suppose that the random variables $W, \ldots\ldots, X_{-2}, X_{-1}$ constitute a martingale relative to the respective fields $F_W, \ldots\ldots, F_{-2}, F_{-1}$, and define

$$F_{-\infty} = \bigcap_{-\infty}^{-1} F_n .$$

Then

$$\lim_{n\to-\infty} X_n = E\{Z|F_{-\infty}\}$$

with probability 1, and the random variables

$$W,\ E\{Z|\ F_{-\infty}\}, \ldots\ldots\ldots, X_{-2}, X_{-1}$$

constitute a martingale relative to the Borel fields

$$F_W,\ F_{-\infty}, \ldots\ldots, F_{-2},\ F_{-1} .$$

(ii) Suppose that the random variables $X_1, X_2, \ldots\ldots, Z$ constitute a martingale relative to the respective Borel fields $F_1, F_2, \ldots\ldots, F_Z$, and let F_∞ be the smallest Borel field of ω-sets with

$$F_\infty \supset \bigcup_1^\infty F_n .$$

Then

$$\lim_{n\to\infty} X_n = E\{Z|\ F_\infty\}$$

with probability 1, and the random variables

$$X_1, X_2, \ldots\ldots\ldots, E\{Z|F_\infty\}, Z$$

constitute a martingale relative to the Borel fields

$$F_1, F_2, \ldots\ldots\ldots, F_\infty, F_Z.$$

<u>THEOREM 4.2.5.</u> (Strong Law of Large Numbers for i.i.d.)

Let $X_1, X_2, \ldots.,$ be mutually independent random variables with a common distribution function, and suppose $E|X_1| < \infty$. If $b_1 < b_2 < \ldots\ldots \to \infty$, then

$$\lim_{n\to\infty} \frac{X_1 + \ldots\ldots\ldots + X_n}{b_n} = E(X_1)$$

with probability one. Here take $b_n = n$.

<u>PROOF</u> If $Y_n = \sum_{j=1}^{n} b_n^{-1} X_j$ and $Z = X_1$, then according to Corollary (4.2.2), Theorem (4.2.3),

$$\lim_{n\to\infty} E\{X_1|Y_n, Y_{n+1}, \ldots\ldots\ldots\} = Y_{-\infty}$$

exists with probability 1.

It is clear that

$$E\{X_1|\ Y_n, Y_{n+1}, \ldots\} = E\{X_1|Y_n, X_{n+1}, X_{n+1}, \ldots\},$$

and since X_{n+1}, $X_{n+2}, \ldots\ldots\ldots$, are independent of the pair X_1, Y_n,

$$E\{X_1 | Y_n, X_{n+1}, \ldots\ldots\} = E\{X_1 | Y_n\} \text{ with probability one.}$$

Thus

$$\lim_{n\to\infty} E\{X_1 | X_1 + \ldots\ldots\ldots + X_n\} = Y_{-\infty} \quad \text{with probability one.}$$

By symmetry,

$$\begin{aligned} E\{X_1 | X_1, \ldots\ldots\ldots, X_n\} &= E\{X_j | X_1 + \ldots\ldots\ldots + X_n\}, \; j \leq n \\ &= \frac{1}{b_n} \sum_{j=1}^{n} E\{X_j | X_1 + \ldots\ldots\ldots + X_n\} \\ &= E\{\frac{1}{b_n} \sum_{j=1}^{n} X_j | X_1 + \ldots\ldots + X_n\} \\ &= Y_n, \end{aligned}$$

so that the theorem is proved except for the identification of the limit. To identify the limit we note that, in the first place, $Y_{-\infty}$ is unaffected by changes in any finite number of Y_j's and therefore, by the Zero-One Law,

$$Y_{-\infty} = \text{constant} = E\{Y_{-\infty}\} \text{ with probability one.}$$

In the second place, according to Theorem (4.2.4), the sequence

$$Y_{-\infty}, \ldots\ldots\ldots, E\{X_1 | Y_1, Y_2, Y_3, \ldots\ldots\}, E\{X_1 | Y_1, Y_2, \ldots\}, X_1$$

is a martingale, and therefore,

$$E\{Y_{-\infty}\} = E\{X_1\}.$$

That is,

$$\lim_{n\to-\infty} E\{Y_n\} = E\{Y_{-\infty}\} = E\{X_1\}.$$

Thus

$$\lim_{n\to\infty} \frac{X_1 + \dots\dots\dots + X_n}{b_n} = E\{X_1\} \text{ with probability one.}$$

Chapter 5

INDEPENDENT RANDOM VARIABLES TAKING VALUES IN ABSTRACT SPACES

5.1 PRELIMINARIES

In the early chapters we considered random variables taking values in the real space $\mathbb{R}$. Similar results can b obtained for random variables taking values in a Banach space. The real difficulty lies in the treatment of random variables (independent but not identically distributed) taking values in a Banach space. This problem has been treated in various forms by FORTET, R. and MOURIER, E. in their paper (1955-56) and (1965), BECK (1961), WOYCZYNSKI, W.A. (1973) and CHUNG, K.L. (1947). Here we shall follow the Chung's version of the strong law of large number because Beck's and Woyczynski's results from Chung's.

The situation is not much more complicated if the values of the random variables are in a Hilbert space becau in this case the variance of the sum of independent random variables is equal to the sum of the variances.

The Law of Iterated Logarithm in H-space will be stated; analogue of Strassen's functional version of the loglog law using Berry-Esseen type estimates for random variables with values in a real separable H-space. Briefly, we shall touch the strong laws of large numbers for Banach space-valued random variables which are subject to the Banach space analog of orthogonality called weak orthogonality. This is dealt with by BECK and WARREN (1972).

<u>DEFINITION 5.1.1</u> Let $\{X, \mathfrak{F}, P\}$ be a probability space, where X is a sample space, and let E be a Banach space. A function $\xi(\omega)$, $\omega \in X$, is called a B-random variable, or a random variable taking values in Banach space E, if $\xi(\omega) \in E$ for each $\omega \in X$ and $\{\omega\colon \xi(\omega) \in U\} \in \mathfrak{F}$ for any open subset U of E ("open" is with respect to the norm (strong) topology) or, in other words, if $\xi(\omega)$ is measurable.

<u>DEFINITION 5.1.2.</u> If $\|\xi(\omega)\|$ is integrable (with respect to P), then there exists an element $E(\xi)$ of B-space E, called the expectation of ξ, for which

$$f(E(\xi)) = \int_X f(\xi)\, dP = E(f(\xi))$$

for any bounded linear functional f defined on E.

The variance of ξ is defined as

$$D^2(\xi) = \int_X \| \xi - E \|^2 dP.$$

DEFINITION 5.1.3 (Uniform Smoothness) A normed space, and its norm, are said to be uniformly smooth if, for each $\varepsilon > 0$, there is an $\eta(\varepsilon) > 0$ for which $\|x\| \geq 1$, $\|y\| \geq 1$ and $\|x-y\| \leq \varepsilon$ always implies

$$\|x + y\| \geq \|x\| + \|y\| - \varepsilon \|x-y\| .$$

A normed space E is uniformly smooth if and only if, for each $\varepsilon > 0$, there is a $\rho(\varepsilon) > 0$ for which $\|x\| = 1$, $\|y\| \leq \rho$ always implies that

$$\|x+y\| + \|x-y\| \leq 2 + \varepsilon\|y\| .$$

DEFINITION 5.1.4 (Uniform Convexity) Following CLARKSON (1964), a normed space E, and its closed unit ball, are said to be uniformly convex if for each ε with $0 < \varepsilon < 2$ there exists a $\delta(\varepsilon) > 0$ for which it always follows from $\|x\| \leq 1$, $\|y\| \leq 1$ and $\|x-y\| \geq \varepsilon$ that

$$\|\tfrac{1}{2}(x+y)\| \leq 1 - \delta(\varepsilon).$$

Such a function $\delta(\varepsilon)$ is called a module of convexity for E.

E is uniformly convex if and only if it always follows from $\|x_n\| \leq 1$, $\|y_n\| \leq 1$ and $\lim_{n\to\infty} \|\tfrac{1}{2}(x_n+y_n)\| = 1$ that $\lim_{n\to\infty} \|x_n - y_n\| = 0$.

If E is uniformly convex, so is its completion.

EXAMPLE The Hilbert space l^2 of sequences is uniformly convex. In l^2 we have $\|x\|^2 = (x,x)$, and from this follows identity

$$\|x+\eta\|^2 + \|x-\eta\|^2 = 2(\|x\|^2 + \|\eta\|^2).$$

Thus if $\|x\| \leq 1$, $\|\eta\| \leq 1$ and $\|x-\eta\| \geq \varepsilon$,

then

$$\|\tfrac{1}{2}(x+\eta)\|^2 \leq 1 - \varepsilon^2/4,$$ so that l^2 is uniformly convex.

The following results give uniform convexity in terms of uniform smoothness. If the B-space E is uniformly convex, then E* (its dual) is uniformly smooth. If the B-space E is uniformly smooth, then E* is uniformly convex.

REMARKS. See [4], [5], [8], and [9] for full discussions.

5.2. INDEPENDENT RANDOM VARIABLES TAKING VALUES IN A HILBERT SPACE

The first theorem will be an analogue of the Kolmogorov's Inequality of the real space.

THEOREM 5.2.1 Let $\xi_1, \xi_2, \ldots\ldots, \xi_n$ be a sequence of independent random variables taking values in a Hilbert space H for which $E(\xi_i) = 0$, $(i = 1,2,\ldots\ldots,n)$. Then, we have

$$P\left(\sup_{1<k<n} \| \sum_{i=1}^{k} \xi_i \| > \varepsilon\right) \leq \sum_{k=1}^{n} \frac{D^2(\xi_k)}{\varepsilon^2} \quad \text{for } \varepsilon > 0.$$

REMARKS For the proof, see [7].

But it must be noted that

$$\sum_{k=1}^{n} E(\xi_\kappa)^2 = E[(\sum_{k=1}^{n} \xi_\kappa)^2]$$

for $n = 1,2,\ldots\ldots$, and that

$$\int_A (\xi_\kappa , \xi_j) \, dP = 0$$

whenever $j < \kappa$ and $A \in \beta(\xi_1,\ldots\ldots,\xi_{k-1})$, the smallest σ-algebra with respect to which the random variables $\xi_1,\ldots\ldots,\xi_{k-1}$ are measurable.

These two simple facts follow from the independence of the random variables and induced orthogonality of independent random variables in a Hilbert space.

THEOREM 5.2.2. (Strong Law of Large Numbers)

Let $\xi_1, \xi_2, \ldots\ldots$, be a sequence of independent random variables taking values in a Hilbert space H for which

$E(\xi_i) = 0$ (0 is the zero element of H)

and
$$\sum_{i=1}^{\infty} \frac{D^2(\xi_i)}{i^2} < \infty.$$

Then
$$P\left(\| \frac{1}{n} \sum_{i=1}^{n} \xi_i \| \to 0\right) = 1,$$ and if $\sum_{i=1}^{\infty} D^2(\xi_i) < \infty$ and $E(\xi_i) = 0$, then $\sum_{i=1}^{\infty} \xi_i$ is convergent with probability 1.

That is, there exists a random variable η (taking values in H) such that
$$P\left(\| \sum_{i=1}^{n} \xi_i - \eta \| \to 0\right) = 1.$$

PROOF Only the second assertion of the theorem will be proved since, by Kronecker's lemma, the first assertion follows from the second.

Let $\eta_j = \xi_1 + \xi_2 + \dots\dots\dots + \xi_j$,
$$\alpha_m(\omega) = \sup_j \{\| \eta_{m+j} - \eta_m \|\},$$

and
$$\alpha(\omega) = \inf_m \{\alpha_m(\omega)\}.$$

To show the convergence of $\sum_{j=1}^{\infty} \xi_j$ it is enough to prove that
$$P\{\alpha(\omega) = 0\} = 1.$$

By the Theorem (5.2.1) above, for any m and n, we have

$$P(\sup_{1<j\leq n} \| \eta_{m+j} - \eta_m \| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{j=m+1}^{n} D^2(\xi_j) .$$

Hence

$$P(|\alpha_m| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{j=m+1}^{\infty} D^2(\xi_j)$$

for any integer m, where ε is an arbitrary positive number. Thus for sufficiently large ε, $\alpha(\omega) = 0$ almost everywhere. Hence, $\sum_{j=1}^{\infty} \xi_j$ is convergent with probability one. Applying the Kronecker's lemma, we have that

$$P(\| \frac{1}{n} \sum_{j=1}^{n} \xi_j \| \longrightarrow 0) = 1.$$

REMARKS. The two theorems hold even if the random variables are not independent. This can be explained by the method of Centering at Conditional Expectations.

THEOREM 5.2.3. (Iterated Law of Large Numbers)

Let $X_1, X_2, \ldots\ldots\ldots\ldots$, be independent H-valued random variables such that $E(X_n) = 0$, $(n=1,2,\ldots\ldots)$, $\sup_n E\| X_n \|^3 < \infty$, and each X_n has a common covariance operator T where

$$(Tx,y) = E\{(X_n,x)(X_n,y)\},\ x,y \in H \text{ and } n = 1,2,\ldots\ldots.$$

Let μ be the mean zero Gaussian measure on H with covariance operator T and let K denote the unit ball of the Hilbert space H_μ which generates μ on H.

Then

$$P\left(\lim_{n\to\infty} \left\| \frac{X_1 + \ldots\ldots + X_n}{(2_n \text{ loglog } n)^{\frac{1}{2}}} - K \right\| = 0 \right) = 1,$$

and infact, with probability one, the sequence $\{(X_1 + \ldots\ldots\ldots + X_n)/(2n \text{ loglog } n)^{\frac{1}{2}}\}$ accumulates at every point of K.

REMARKS. The proof is furnished in ([13], page 397).

It is known that K is a compact subset of H. Thus the conclusions of the theorem are equivalent to saying that, with probability one, the sequence of the points $\{(X_1 + \ldots\ldots + X_n)/(2n \text{ loglog } n)^{\frac{1}{2}} : n \geq 3\}$ is conditionally compact in H and that with probability one the set of limit points of each sequence is precisely K.

5.3. INDEPENDENT RANDOM VALUES TAKING VALUES IN BANACH SPACE

The i.i.d. random variables do not pose a serious problem as the independent random variables. The strong law of large numbers for Banach space-valued random variables, in its generic form states:

<u>THEOREM 5.3.1.</u> Let E be a specified Banach space. Let $\{X_j\}$, $j = 1,2,\ldots\ldots$, be a sequence of E-valued random variables subject to a certain set of conditions. Then we have that $n^{-1} \sum_{j=1}^{n} X_j$ converges to a constant function in the norm topology of E almost surely.

<u>DEFINITION 5.3.1</u> The sequence $\{\xi_n\}$ and $\{\xi_n^*\}$ are called equivalent in the sense of Khinchin if

$$\sum_{j=1}^{\infty} P(\xi_j \neq \xi_j^*) < \infty.$$

<u>THEOREM 5.3.2.</u> (For i.i.d. random variables).

Let $X_1, X_2, \ldots\ldots$, be a sequence of independent and identically distributed random variables taking values in the Banach space E for which $E(X_i) = 0$.

Then

$$P(\| \frac{1}{n} \sum_{i=1}^{n} X_i \| \to 0) = 1.$$

<u>PROOF</u>. When E is finite-dimensional the proof follows from the fact that expectation of a finite-dimensional random vector of the expectations of the co-ordinates of the random vector.

We state the following lemma whose proof can be seen in [7] for proving the theorem.

LEMMA 5.3.1. The sequence C_1X_1, C_2X_2, ... obeys the strong law of large numbers, if

$$E(X_j) = 0$$

$$E|X_j|^p < \infty \text{ for } 1 \leq p < 2, \text{ and}$$

$$j^{-2-p} \sum_{n=[j^p]}^{\infty} \frac{|C_n^2|}{n^2} \leq M$$

where M is a positive constant, and $X_1, X_2, \ldots\ldots$, are i.i.d random variables.

Now let ε be an arbitrary positive number and let $X_1, X_2, \ldots\ldots$, be a dense sequence in E.

Set

$$K_j = \{X, \; \|X - X_j\| < \varepsilon\} \text{ , and}$$

$$A_j = K_j - \bigcup_{i=1}^{j-1} K_i \; .$$

Define the Banach space-valued random variable $X_i^*(\omega) = X_j$ whenever $X_i(\omega) \in A_j$, $(i = 1, 2, \ldots\ldots, \; j = 1, 2, \ldots\ldots)$.

Then

$$\| X_i^*(\omega) - X_i(\omega) \| \leq \varepsilon, \text{ and}$$

$$\sum_{j=1}^{\infty} \| X_j \| \, P\{X_i \in A_j\} = E(\| X_i^* \|)$$

$$\leq E(\| X_i \|) + \varepsilon$$

$$< \infty.$$

Choose an integer N such that

$$\sum_{j=N}^{\infty} \| X_j \| P\{X_i \in A_j\} < \varepsilon, \quad (i = 1,2,\ldots\ldots\ldots\ldots\ldots).$$

Let

$$\xi_i^* = \begin{cases} X_j \text{ if } X_i \in A_j, \ j \geq N \\ 0 \quad \text{if } j < N \end{cases}$$

and $\eta_i = X_i^* - \xi_i^*$, $\xi_i = X_i - \eta_i$.

Then $\{\xi_i^*\}$ and $\{\eta_i\}$ are sequences of independent Banach random variables for which

$$\| E(\xi_i^*) \| \leq E(\| \xi_i^* \|)$$

$$= \sum_{j=N}^{\infty} \| X_j \| P\{X_i \in A_j\}$$

(for i = 1,2,..)

$$< \varepsilon,$$

and

$$\| E(\eta_i) \| \leq \| E(X_i^*) \| + \| E(\xi^*{}_i) \|$$

$$= \| E(X_i^* - X_i) + E(X_i) \| + \| E(\xi_i^*) \|$$

$$= \| E(X_i - X_i^*) \| + \| E(\xi_i^*) \|$$

$$< \varepsilon + \varepsilon$$

$$= 2\varepsilon .$$

Furthermore, $\xi_i = X_i - \eta_i$

$$= (X_i - X_i^*) + (X_i^* - \eta_i)$$

$$= (X_i - X_i^*) + \xi_i^*,$$

and hence,

$$E(\| \xi_i \|) \leq E(\| X_i - X_i^* \|) + E(\| \xi_i^* \|)$$

$$\leq 2\varepsilon .$$

Clearly, since $E(X_i) = 0$,

$$X_i = (\eta_i - E(\eta_i)) + (\xi_i - E(\xi_i)).$$

$\{\eta_i - E(\eta_i)\}$ is a sequence of independent identically distributed random variables, taking only (N-1) distinct values. Thus it obeys the strong law of large numbers. (Note that $X_i^* = \xi_i^*$ whenever $X_i(\omega) \in A_j$ and $j \geq N$.)

Since

$$\| \frac{1}{n} \sum_{i=1}^{n} (\xi_i - E(\xi_i)) \|$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \| \xi_i \| + \frac{1}{n} \sum_{i=1}^{n} \| E(\xi_i) \|$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \| \xi_i \| + 2\varepsilon$$

$< K$ (a constant), we have that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \| \xi_i - E(\xi_i) \| \leq K.$$

So, by Lemma (5.3.1), we have that $\{\xi_i - E(\xi_i)\}$ oebys the strong law of large numbers. Thus we have from

$X_i = (\eta_i - E(\eta_i) + (\xi_i - E(\xi_i))$ that X_i obeys the strong law of large numbers. So the proof is completed.

Let $X_1, X_2, \ldots\ldots\ldots,$ be independent random variables with values in a Banach space E. We shall follow the Chung's version of the strong law of large numbers which holds if and only if E is of type p. Beck's version (1961) using the property of uniform convexity and Woyczynski's (1973) using the idea of uniform smoothness can be deduced from the Chung's version. Fortet, M.R. and Mourier, E (1955 - 56) and (1965) also worked on the problem.

Some of these authors make use of metric properties of normed spaces in their proofs because positive results are difficult to get except if the Banach space is "good" in some sense.

So we shall show that if $\{X_n\}$ satisfies (for some $1 \leq p \leq 2$)

$\sum_{n=1}^{\infty} n^{-p} E\| X_n \|^p < \infty$, then the strong law of large numbers holds for all independent sequences $\{X_n\}$ is and only if E is of type p.

Let $\{\varepsilon_n\}$ be a Bernoulli sequence, that is, ε_n's are independent real random variables with $P(\varepsilon_n = \pm 1) = \frac{1}{2}$ for all n. We define

(5.3.1)--- $C(E) = \{ \{X_n\} \in E^\infty : \sum_{j=1}^{\infty} \varepsilon_j X_j$ converges in probability$\}$, where E^∞ denotes infinite -dimensional real Banach space. If $L^p(\Omega,E)$ denotes the vector space of (equivalence classes of) functions f measurable on Ω into Banach space E such that $\| f(.)\| \in L^p(\Omega)$. The space $L^p(\Omega,E)$ is a Banach space with respect to the norm

$$\| f, L^p(\Omega,E)\| = \begin{cases} \left[\int_\Omega \| f(t)\|_E^p \, dt \right]^{1/p} & \text{if } 1 \leq p < \infty \\ \operatorname{ess\,sup}_{t \in \Omega} \| f(t)\|_E & \text{if } p = \infty \end{cases}$$

We can substitute "convergence in probability" with convergence in $L^p(E)$ for all $p \in \mathbb{R}^+$. Let $\| . \|_p$ denote the usual metric in $L_p(E)$ for $p \in \mathbb{R}^+$, and for $X_n \neq 0$, put

(5.3.2)---- $|X|_p = \sup_{1 \leq n < \infty} \| \sum_{j=1}^{n} \varepsilon_j X_j \|_p$ for $X = \{X_j\} \in C(E)$.

<u>LEMMA</u> 5.3.2. $\{C(E), |.|_p\}$ is a Banach space (hence a Fretchet space) for $p \in \mathbb{R}^+$.

REMARKS. The number (5.3.2) is finite since the sequence $\{ \| \sum_{i=1}^{n} \varepsilon_i X_i \| \}$ is convergent. Since for $X_n \neq 0$ $(n = 1,2,\ldots\ldots$ (5.3.2) is a norm on the linear space $C(E)$. Then a standard argument shows that a Cauchy sequence $\{X_n^{(k)}\}$ $(k = 1,2,\ldots\ldots)$ in $C(E)$ converges to $\{X_n\} \in C(E)$.

Since all the $|.|_p$-topologies are stronger than the product topology on $C(E)$, we find from the closed-graph theorem, see ([5], page 167), that all the $|.|_p$- topologies ($p \in \mathbb{R}^+$) are equivalent, that is:

LEMMA 5.3.3. Let $\{\varepsilon_n\}$ be a Bernoulli sequence, and define the set L by

$$L = \sum_{j=1}^{\infty} \varepsilon_j X_j \ : \{X_j\} \in C(E)\},$$

then $L \subsetneq L^p(E)$ for all $p \in \mathbb{R}^+$, and all the L^p - topologies, for $p \in \mathbb{R}^+$, coincide on L.

DEFINITION 5.3.2. Let $p \in [1,2]$, then E is said to be of type p if we have $l^p(E) \subsetneq C(E)$.

Since the injection from $l^p(E)$ into $C(E)$ has a closed graph in case E is of type p, we have that E is of type p, if and only, if there exists $A \in \mathbb{R}_+$ such that

$$(5.3.3)\text{----}\quad E\| \sum_{j=1}^{\infty} \varepsilon_j X_j \|^p \leq A \sum_{j=1}^{\infty} \| X_j \|^p$$

for all $\{X_j\} \in C(E)$.

(note that continuous mappings are bounded).

Finally, we state a lemma which follows from its equivalent under the real space $\mathbb{R}$.

LEMMA 5.3.4. If $\{X_n\}$ are independent E-valued random variables so that the series $\sum X_n$ converges in probability, then it converges almost surely.

Now let $\{X_n\}$ be a sequence of E-valued random variables satisfying

(5.3.4)----- $EX_n = 0$ for all $n \geq 1$. Then $\{X_n\}$ is said to satisfy the strong law of large numbers if

$$\frac{1}{n} \sum_{j=1}^{n} X_j \to 0 \text{ almost surely.}$$

We shall mainly consider here Chung's condition

(5.3.5)----- $\sum n^{-p} E\| X_n\|^p < \infty$

where p is some fixed positive number.

THEOREM 5.3.3 (Strong Law of Large Numbers)

Let $1 \leq p \leq 2$, then the following four statements are equivalent:

(5.3.6)---- E is of type p.

(5.3.7)---- $\exists$ $c > 0$ such that $E\|\sum_{j=1}^{n} X_j\|^p \leq c \sum_{j=1}^{n} E\| X_j\|^p$ for all independent $X_1, \ldots\ldots, X_n$ with mean 0 and pth moment.

(5.3.8)----- The strong law of large numbers holds for all independent sequences $\{X_n\}$, satisfying (5.3.4) and (5.3.5).

(5.3.9)----- If $\sum_{j=1}^{\infty} j^{-p} \| X_j \|^p < \infty$ and $\{\varepsilon_j\}$ is a Bernoulli sequence, then $n^{-1} \sum_{j=1}^{n} \varepsilon_j X_j \to 0$ in probability.

PROOF. Suppose (5.3.6) holds. Then there exists a constant B so that (see (5.3.3))

$$E \| \sum_{j=1}^{n} \varepsilon_j X_j \|^p \leq B \sum_{j=1}^{n} \| X_j \|^p$$

for all $X_1, \ldots\ldots\ldots, X_n \in E$ and for all n. So by Fubini's theorem we find

(5.3.10)------- $E \| \sum_{j=1}^{n} \varepsilon_j X_j \|^p \leq B \sum_{j=1}^{n} E \| X_j \|^p$

where $\{\varepsilon_j\}$ is a Bernoulli sequence independent of $X_1, \ldots\ldots\ldots, X_n$, and $X_1, \ldots\ldots\ldots\ldots\ldots, X_n$ are independent random vectors with mean 0 and finite pth moment. The following result (from HOFFMAN - JORGENSEN 1964)) will be stated without proof:

(5.3.11)---------$E \| \sum_{j=1}^{n} X_j \|^p \leq 8^p E \| \sum_{j=1}^{n} \varepsilon_j X_j \|^p$.

Multiplying (5.3.10) and (5.3.11) we get

$$E\| \sum_{j=1}^{n} X_j \|^p \leq 8^p B \sum_{j=1}^{n} E\| X_j \|^p .$$

So (6.3.7) holds with $c = 8^p B$.

Suppose that (5.3.7) holds, and that $\{X_j\}$ satisfies (5.3.4) and (5.3.5). Then the series $\sum j^{-1} X_j$ converges in $L^p(E)$ by (5.3.7), so by Lemma (5.3.4) it converges almost surely. Now we notice Kronecker's lemma is valid in any Banach space (with the same proof as in the real xase), so applying Kronecker's lemma to $(X_j(\omega))$, for those ω for which the series $\sum j^{-1} X_j(\omega)$ converges, we find that $n^{-1}(X_1 + \ldots\ldots + X_n) \to 0$ almost surely. And so (5.3.8) holds. (5.3.8) clearly implies (5.3.9).

Now suppose (5.3.9) holds, and let $\{X_j\}$ be a sequence in E so that

$$\sum_{j=1}^{\infty} j\text{-}p \| X_j \|^p < \infty .$$

Then from (5.3.9) and Lemma (5.3.3) we have that

$$\frac{1}{n} \sum_{j=1}^{n} \varepsilon_j X_j \to 0 \text{ in } L^p(E).$$

Now let

$$F = \{ \{X_j\} \in E^\infty : \sup_n \frac{1}{n} (E\| \sum_{j=1}^{n} \varepsilon_j X_j \|^p)^{1/p} < \infty \}.$$

Then F is a linear space and we may define the norm

$$|X| = \sup_n \frac{1}{n} (E\|\sum_{j=1}^{n} \epsilon_j X_j\|^p)^{1/p}$$

for $X = \{X_j\} \in F$, and by a standard argument, it can be shown that $(F, |.|)$ is a Banach space.

Now let

$$F_0 = \{ \{X_j\} \in E^{\infty} : \sum_{j=1}^{\infty} j^{-p}\| X_j\|^p < \infty\}$$

and $|X|_0 = \{ \sum_{j=1}^{\infty} j^{-p}\| X_j\|^p \}^{1/p}$, then

$(F_0, |.|_0)$ is a Banach space, and by assumption, we have $F_0 \subseteq F$. The injection $F_0 \to F$ is clearly a linear operator with closed graph, and it is continuous. That is, there exists $C > 0$ such that

$$E\|\sum_{j=1}^{n} \epsilon_j X_j\|^p \leq Cn^p \sum_{j=1}^{n} j^{-p}\| X_j\|^p$$

for all $X_1, \ldots\ldots, X_n \in E$ and all $n > 1$. Now let $X_1, \ldots\ldots, X_n \in E$ and define

$$Y_j = \begin{cases} 0 \text{ for } 1 \leq j \leq N \\ X_{j-N} \text{ for } N < j \leq N+n \end{cases}$$

where N is some integer.

Then

$$E\|\sum_{j=1}^{n}\varepsilon_j X_j\|^p = E\|\sum_{j=1}^{N+n}\varepsilon_j Y_j\|^p$$

$$\leq C(N+n)^p \sum_{j=N+1}^{N+n} \frac{\|X_{j-N}\|^p}{j^p}$$

$$\leq C\left(\frac{N+n}{N+1}\right)^p \sum_{j=1}^{n}\|X_j\|^p$$

for all $N \geq 1$, and so we have

$$E\|\sum_{j=1}^{n}\varepsilon_j X_j\|^p \leq C\sum_{j=1}^{n}\|X_j\|^p \text{ for all}$$

$X_1,.......,X_n \varepsilon E$ and for all $n \geq 1$. Hence, E is of type p and the theorem is proved.

REMARKS. A Banach space B is called convex in the sense of Beck if there is an integer j and $\varepsilon > 0$ such that for every sequence $X_1, X_2, ..., X_j$ of the elements of B with $\|X_i\| \leq 1$, $(i = 1,2,.....,j)$, we have

$$\|\pm X_1 \pm X_2 \pm .... \pm X_j\| \geq j(1-\varepsilon)$$

for some choice of the signs + and - .

Beck had shown that if E is Beck - convex, then the strong law of large numbers holds for all independent sequence $\{X_n\}$ satisfying

$EX_n = 0$ for all $n \geq 1$ and

(5.3.12)----- $\sup_n E\|X_n\|^2 < \infty$.

But E is Beck-convex if and only if E is of type p for some $p > 1$. Hence the result of Beck is included in Theorem (5.3.3), since (5.3.12) obviously implies (5.3.5) for $p > 1$.

DEFINITION 5.3.3. A topological vector space E is called a G_α - space ($0 < \alpha \leq 1$) if E is uniformly $(1+\alpha)$-smooth with modulus of smoothness given by

$$\rho(t) = \sup \{\tfrac{1}{2}(\|X+Y\| + \|X-Y\| - 2) : \|X\| = 1, \|Y\| = t\}$$

and ρ satisfying $\rho(t) = O(t^{1+\alpha})$ as $t \to 0$.

W.A. Woyczynski showed that if E is a G_α-space with $\alpha = p-1$, then the strong law of large numbers holds for all independent sequences $\{X_j\}$, satisfying $EX_n = 0$ for all $n > 1$ and (5.3.5).

So by Theorem (5.3.3) we find that every G_α-space is of the type $(1+\alpha)$ because according to the theorem the SLLN holds a Banach space E of type p (see definition) if

$$1 \leq p \leq 2$$

ie. $\quad 0 \leq p-1 \leq 1$

ie. $\quad 0 < \alpha \leq 1$

which implies $1 < \alpha+1 \leq 2$.

So with this argument it could be seen that Woyczynski's result is included in the theorem for $p = \alpha+1$ where $0 < \alpha \leq 1$.

5.4 STRONG LAWS OF LARGE NUMBERS FOR WEAKLY ORTHOGONAL SEQUENCES OF BANACH SPACE VALUED RANDOM VARIABLES

INTRODUCTION: This section studies SLLN for Banach space-valued random variables which are subject to the Banach space analog of orthogonality. So we shall relax the requirement of independence and consider instead a form of orthogonality.

ANATOLE BECK and WARREN, P. (1972) studied the Banach space analog of orthogonality called weak orthogonality. In the case of independent sequence of B-valued random variables identical distribution is sufficient for the strong law to hold. This is not so in the case of weakly orthogonal sequence as it is shown in the proof of the following example; see ([11]. pg. 922).

EXAMPLE: Let $B = C_0$, the subspace of l_∞ consisting of sequences which converge to zero. Then there exists a sequence of B-valued random variables which are

(i) uniformly bounded in norm
(ii) identically distributed
(iii) weakly orthogonal, but which
(iv) do not satisfy the strong law of large numbers.

This same example even shows that a sequence of identically distributed, uniformly bounded B-valued random variables satisfying the strong law of large numbers in the weak linear topology of B does not necessarily satisfy the Strong Law in the norm topology of B. This is interesting because similar sequences of B-valued (for B separable) random variables which satisfy the weak law of large numbers (convergence in probability) in the weak linear topology also must satisfy the weak law in the norm topology (TAYLOR, R.L. (1972)).

However, two theorems which state conditions when identical distribution and weak orthogonality are sufficient for the strong law will be given.

PRELIMINARIES:

Some vital definitions will be given and two lemmas without proofs (their proofs are given in [11]) will be stated. B denotes a Banach space as usual with norm $\|\cdot\|_B$, B^* denotes the dual of B, and $(\Omega, \mathfrak{F}, P)$ denotes a probability space.

DEFINITION 5.4.1. A mapping $X: \Omega \to B$ is strongly measurable if for every Borel set $C \subset B$, $X^{-1}(C)$ is measurable.

Strongly measurable functions from Ω into B, in other words, are called B-valued random variables. $L_p(\Omega, \mathfrak{F}, P, B)$, $1 \leq p < \infty$, hereafter written simply $L_p(\Omega, B)$, denotes the space of all B-valued random variables on Ω for which the norm

$$|||X||| = \left(\int_\Omega \| X(\omega) \|_B^p P(d\omega) \right)^{1/p} < \infty.$$

DEFINITION 5.4.2. A finite collection of B-valued random variables $X_1, \ldots\ldots, X_m$ is independent if for every collection of m Borel sets $C_1, \ldots\ldots, C_m \subset B$, we have

$$P\{\omega: X_1(\omega) \in C_1, \ldots, X_m(\omega) \in C_m\} = \prod_{i=1}^{m} P\{\omega: X_i(\omega) \in C_i\}.$$

An infinite collection of B-valued random variables is independent if every finite sub-collection is independent.

DEFINITION 5.4.3 A collection of B-valued random variables $X_1, X_2, \ldots\ldots$, is identically distributed, if for every Borel set $C \subset B$, and for all positive integers i and j, we have

$$P\{\omega: X_i(\omega) \in C\} = P\{\omega: X_j(\omega) \in C\}.$$

DEFINITION 5.4.4. A B-valued random variable X is strongly (Bochner) integrable if

$$\int_\Omega \| X(\omega) \|_B \; P(d\omega) < \infty.$$

This is not the usual definition of strong integrability but is employed here to avoid a longer definition.

DEFINITION 5.4.5 A sequence of B-valued random variables is weakly orthogonal if, for all $X^* \varepsilon B^*$, we have

$$E(X^*X_i \cdot X^*X_j) = \int X^*X_i(\omega) \cdot X^*X_j(\omega)P(d\omega) = 0$$

Strong Laws in B-Spaces. We shall consider only B-valued random variables which have their expectations equal to the zero element of B; general case follows by the use of "Centering at expectation."

We state the following two lemmas (see [11] for their proofs) which are to be used later.

LEMMA 5.4.1. Let B be a Banach space and let $X_1, X_2, \ldots\ldots$ be a weakly orthogonal sequence of identically distributed B-valued random variables with $Var(X_1) < \infty$. Then, for each $X^* \varepsilon B^*$, we have $X^*(Y_n) \to 0$ almost surely, where

$$Y_n = N^{-1}(X_1 + \ldots\ldots\ldots\ldots\ldots\ldots + X_n).$$

LEMMA 5.4.2. Let B be a separable Banach space and suppose $\|Y_n - Y\| \to 0$ almost surely, where Y_n and Y are B-valued random variables. If $X^*(Y_n) \to 0$ almost surely for each [illegible] $Y = 0$ almost surely.

Now, in [illegible] with some other stringent conditions, we will give a pair of theorems which state conditions when identical distribution and weak orthogonality are sufficient for the strong law of large numbers to hold.

THEOREM 5.4.1. Let B^* be a separable Banach space, C a convex subset of B with compact closure, and let $X_1, X_2, \ldots\ldots\ldots$ be a weakly orthogonal sequence of identically distributed B-valued random variables with $\mathrm{var}(X_1) < \infty$. If the range of $X_j \subseteq C$ for each j, then the sequence satisfies the strong law of large numbers.

PROOF. Since C is convex, it is clear that the range of Y_n is contained in C for each n. Furthermore, Lemma (5.4.1) implies that, for each $X^* \in B^*$, $X^*(Y_n) \to 0$ except possibly on a set $D(X^*)$ of measure zero.

Let $X_1^*, X_2^*, \ldots\ldots\ldots\ldots\ldots$ be a countable dense subset of B^*.

Let $D = \bigcup_{j=1}^{\infty} D(X_j^*)$.

Then, D has a measure zero and, if $\omega_o \notin D$, we have

$X_j^*(Y_n(\omega_o)) \to 0$ as $n \to \infty$ for each j.

Let $A(\omega_o) = \{Y_1(\omega_o),\ Y_2(\omega_o), \ldots\ldots\}$.

For each ω_o, $A(\omega_o)$ has at least the limit point in the closure of C since C is compact. Now suppose that $\omega_o \notin \Delta$ and $A(\omega_o)$ has at least two distinct limit points, say y_1 and y_2.

Let $M = \max \{\| y_1 \|, \| y_2 \|\}$. Then, since B^* separates points, there exists some $X^* \in B^*$ such that

$$|M[\hat{X}^*(y_1/M) - \hat{X}^*(y_2/M)| = |\hat{X}^*(y_1) - \hat{X}^*(y_2)|$$

$$= \varepsilon > 0.$$

Since the X_j^*'s are dense in B^*, we can find j, say $j = \rho$, such that

$$\| \hat{X}^* - X_\rho \| < \varepsilon/4M.$$

This implies that

$$|\hat{X}^*(y_i|M) - X_\rho^*(y_i|M)| < \varepsilon/4M \text{ for } i = 1,2,\ldots\ldots\ldots.$$

It follows that $|X_\rho^*(y_1) - X_\rho^*(y_2)| < \varepsilon/2$.

This, however, contradicts the fact that $X_\rho^*(Y_n(\omega_o)) \to 0$. Indeed, $A(\omega_o)$ has a unique limit point. Let $\hat{Y}: \Omega \to B$ be defined as the function which equals the unique limit point of $A(\omega)$ when $\omega \notin B$, and otherwise, is arbitrary. Clearly, $\| Y_n - \hat{Y} \| \to 0$ almost surely. Since B is separable, Lemma (5.4.2) implies that $Y = 0$ with probability one which completes the proof of the theorem.

<u>THEOREM 5.4.2.</u> Let B be a finite-dimensional Banach Space. Let $X_1, X_2, \ldots\ldots\ldots$, be a weakly orthogonal sequence of identically distributed B-valued random variables with $Var(X_1) < \infty$.

Then the sequence satisfies the strong law of large numbers.

PROOF. Let $b_1, \ldots\ldots, b_m$ be a Hamel basis for B. Every $X \in B$ can be uniquely written as

$$X = \sum_{i=1}^{m} \alpha_i b_i \ , \quad \text{where the}$$

α_i's are suitable scalars. Define linear maps $b_i^* : B \to \Phi$ into the field of scalars by

$$b_i^*(X) = \alpha_i \text{ if } X = \sum_{i=1}^{m} \alpha_i b_i .$$

These maps are continuous since every linear operator on a finite-dimensional space is continuous. Thus $b_i^* \in B^*$ for $i = 1,2,\ldots\ldots\ldots,m$ and $X = \sum_{i=1}^{m} b_i^*(X) b_i$.

Clearly, if $X_1, X_2, \ldots\ldots$ is a sequence such that for each i, $b_i^*(X_j) \to 0$, then $X_j \to 0$ as $j \to \infty$.

Suppose B is m-dimensional. Then, by Lemma (5.4.1), for $i = 1,2,\ldots\ldots\ldots,m$, $b_i^*(Y_n) \to 0$ as $n \to \infty$ except possibly on a set D_i of measure zero.
Let $D = \bigcup_{i=1}^{m} D_i$ so that D has measure zero.

Thus, for $\omega \notin D$, $b_i^*(Y_n(\omega)) \to 0$ for each $i = 1,2,\ldots\ldots,m$. Hence, $Y_n(\omega) \to 0$ as $n \to \infty$ except on a set of measure zero which is what we needed to prove.

REFERENCES

BOOKS

[1] DOOB, J.L., "Stochastic Processes", John Wiley and Sons, Inc., New York. (1953)

[2] FELLER, W., "Introduction to Probability Tehory and Its Applications," John Wiley and Sons, Inc., New York. (3rd Edition), (1957).

[3] FISZ, M., "Probability Theory and Mathematical Statistics" John Wiley and Sons, Inc., New York. (3rd Edition) (1963).

[4] GNEDENKO, B.V., "The Theory of Probability," Chelsea Publishing Company, New York. (4th Edition) (1968).

[5] GOTTFRIED KOTHE, "Topological Vector Spaces," Springer-Verlag Berlin. Heidelberg.. New York. (1969).

[6] LOEVE MICHEL, "Probability Theory," D. Van Nostrand Company, Inc., Princeton, New Jersey. (3rd Edition) (1963).

[7] REVESZ PAL, "The Laws of Large Numbers," Academic Press. New York. (1968).

[8] SCHAEFER, H.H., "Topological Vector Spaces," Springer — Verlag. New York. Heidelberg. Berlin. (1971).

[9] SINGER, I., "Bases in Banach Spaces I," Springer - Verlag Berlin. Heidelberg. New York. (197[illegible]).

[10] WILKS, S.S., "Mathematical Statistics," John Wiley and Sons, Inc., New York. (1962).

PAPERS

[11] ANATOLE, B. AND WARREN, P., "Strong Law of Large Numbers for Weakly Orthogonal Sequences of Banach Space-Valued Random Variables." The Annals of Probability (1974), Vol. 2, No. 5, pp. 918 - 925.

[12] HOFFMANN-JORGENSEN, J. AND PISIER, G., "The Law of Large Numbers and the Central Limit Theorem in Banach Spaces." The Annals of Probability (1976), Vol. 4, No.4, pp. 587 - 599.

[13] KUELBS, J. AND KURTZ, T., "Berry - Esseen Estimates in Hilbert Space and An Application to the Law of the Iterated Logarithm." The Annals of Probability (1974), Vol. 2, No.3, pp. 387 - 407.

[14] TEICHER, H. "On the Law of the Iterated Logarithm." The Annals of Probability (1974), Vol.2, No.4, pp. 714 - 728.

[15] VAKHANIA, N.N. AND TARIELADZE, V.I. "Covariance Operat... of Probability Measures in Locally Conv... Spaces," Theory of Probability and its Applications, Vol. XXIII, No.1, 1978. pp. 1 - 20.

[16] "Contributions to Ergodic Theory and Probability," (1970).