THE CONVERGENCE OF CONDITIONAL EXPECTATIONS ON A σ-ALGEBRA

by

Dennis Seiho Kira

B.Sc., Simon Fraser University, 1969

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Mathematics

SIMON FRASER UNIVERSITY

April 1972

TABLE OF CONTENTS

APPROVAL

Name: Dennis Seiho Kira

Degree: Master of Science

Title of Thesis: The Convergence of Conditional Expectations on a σ-Algebra

Examining Committee:

(A. R. Freedman)
Chairman

(D. J. Mallory)
Senior Supervisor

(A. Das)
Examining Committee

(J. J. Sember)
Examining Committee

(C. W. Kim)
External Examiner

Date Approved: April 6 /72

ABSTRACT

Let (Ω,F,μ) be a probability space, $C \subset F$ be a sub-σ-algebra. A generalized version of the Lebesgue dominated convergence theorem (GLDCT) was given by J. Doob in his book [4]. More precisely, if $\{X_n, n=1,2,\ldots\}$ is a sequence of non negative random variables converging almost surely to a r.v. X and $\sup_n X_n \in L_1$, then for any sub-σ-algebra $C \subset F$, a sequence of conditional expectations (Rádon-Nikodym derivatives), $\{E[X_n|C], n = 1,2,\ldots\}$ converges almost surely to $E[X|C]$. D. Blackwell and L. Dubins [2], Theorem 1, have shown that the condition $\sup_n X_n \in L_1$ cannot be weakened. That is, in a "certain sense", the condition $\sup_n X_n \in L_1$ is not only a sufficient condition but is also a necessary condition.

The purpose of this paper is to establish a condition on (Ω,F,μ) under which $\sup_n X_n \in L_1$ is a necessary condition. We prove that if (Ω,F,μ) is atomic then the condition $\sup_n X_n \in L_1$ is <u>not</u> a necessary condition for the GLDCT except for the case $E[X_n] \not\to E[X]$. In our final theorem, we establish our main objective by stating that if (Ω,F,μ) is nonatomic, then $\sup_n X_n \in L_1$ is a necessary condition for the GLDCT. We devote chapter 3 to the proof of these theorems.

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Dr. D.J. Mallory, Mathematics Department, Simon Fraser University, for his continuous aid and encouragement.

I also would like to thank the financial assistance from Simon Fraser University.

Finally, I would like to thank Miss Sue Smithe for typing the thesis.

INTRODUCTION

For many years, probability theory was mainly concerned with the study of independent random variables but from the beginning of this century, dependent random variables were seriously examined, chiefly by Markov and Lévy. Such an investigation resulted in the consideration of an abstract notion of conditional probability.

In probability theory, the conditional probability of an event A, given that an event B has already occurred, is introduced as a ratio $P(A \cap B)/P(B)$, where $P(B)$ is assumed to have a positive probability. With the development of measure theory by H. Lebesgue, A.N. Kolmogorov formulated the axiomatic model of probability theory in 1933. In his development of probability theory, the conditional probability of an event A given a σ-algebra $\mathcal{C} \subset \mathcal{F}$, $\mu(A|\mathcal{C})$, (definition 2.1.4), and the conditional expectation of a r.v. X given a σ-algebra $\mathcal{C} \subset \mathcal{F}$, $E[X|\mathcal{C}]$, (definition 2.1.5), are characterized as the <u>Rádon-Nikodym derivatives</u> with respect to certain measures.

It is well known that if $\{X_n, n = 1,2...\}$ is a sequence of integrable functions on a measure space $(\Omega,\mathcal{F},\mu)$, $X_n \to X$ μ almost everywhere, $|X_n| \leq Y$ μ-a.e. for each n, $Y \in L_1(\Omega,\mathcal{F},\mu)$, then the <u>Lebesgue dominated convergence theorem</u> (LDCT) asserts that $\int X_n d\mu \to \int X d\mu$. For his characterization of certain conditional expectations (theorem 9, page 27, [4]), J. Doob has shown that the LDCT can be generalized to the convergence of a sequence of conditional expectations. Specifically, let (Ω,F,μ) be a probability space, $\{X_n, n = 1,2...\}$ a sequence of non-negative random variables in $L_1(\Omega,F,\mu)$ such that $X_n \to X \in L_1(\Omega,F,\mu)$ μ almost surely, then $\sup_n X_n \in L_1$ implies that for any sub-σ-algebra $C \subset F$, $E[X_n|C] \to E[X|C]$ μ almost surely, [4], page 23. We shall refer to this theorem as the generalized Lebesgue dominated convergence theorem (GLDCT). It is clear that the LDCT is a special case of Doob's result since $\int X_n d\mu = E[X_n|\{\phi,\Omega\}]$ and $\int X d\mu = E[X|\{\phi,\Omega\}]$.

With this result in mind, D. Blackwell and L. Dubins have shown that in "some sense" $\sup_n X_n \in L_1(\Omega,F,\mu)$ is not only a sufficient condition but also a necessary condition. More precisely, theorem 1, [2], if $X_n \geq 0$, $X_n \to X$ μ almost surely, $X_n, X \in L_1(\Omega,F,\mu)$, $\sup_n X_n \notin L_1(\Omega,F,\mu)$, there are, on a suitable probability space (Ω^*,F^*,μ^*), random variables $\{X_n^*, n = 1,2...\}$, X^* and a sub-σ-algebra $C^* \subset F^*$ such that $X^*, X_1^*,...$ have the same joint distribution as $X, X_1, X_2 ...$ and

$$\mu^*(\{\omega^*: E[X_n^*|C^*](\omega^*) \to E[X^*|C^*](\omega^*)\}) = 0 .$$

In view of such a result, it is interesting to investigate the properties required by a probability space (Ω,F,μ) to conclude an "exact converse" to J. Doob's result. Namely, how rich must the structure of p.s. (Ω,F,μ) be, to conclude that there exists a σ-algebra $C \subset F$ such that $E[X_n|C](\omega) \not\to E[X|C](\omega)$ with probability one, when $\sup_n X_n \notin L_1(\Omega,F,\mu)$. (i.e. the converse statement holds on the original space.)

The purpose of this thesis is to establish conditions under which $\sup_n X_n \in L_1(\Omega,F,\mu)$ is a necessary condition for the generalized LDCT. We will show that if (Ω,F,μ) is assumed to be a nonatomic probability space, then $\sup_n X_n \in L_1$ is a necessary condition, theorem 3.3.1. However, if (Ω,F,μ) is purely atomic, then $\sup_n X_n \in L_1(\Omega,F,\mu)$ is shown to be not a necessary condition. (theorem 3.2.2 and corollary 3.2.1.)

Chapters one and two are of an introductory nature. We begin by stating the basic definitions and theorems from general measure and probability theory following P.R. Halmos [7], and M. Loéve [9], respectively. A conditional probability is first introduced from an elementary point of view. If a σ-algebra C is generated by a partition of a sample space Ω then we encounter no difficulties in defining $E[X|C]$ or $\mu(A|F)$, $A \in F$, explicitly (definition 2.1.2). However, in general, we cannot give an explicit characterization of conditional probabilities and expectations. With the help of the Rádon-Nikodym theorem, a general characterization of conditional probabilities and expectations is given in chapter 3. We also discuss Rényi's model of conditional probability space and note that A.N. Kolmogorov's model of probability theory is a special case of Rényi's model.

In chapter 3, we give a proof of our main result. We first note that if $E[X_n] \not\to E[X]$ then we have $E[X_n|\{\phi,\Omega\}] \not\to E[X|\{\phi,\Omega\}]$ μ-almost surely, but if $E[X_n] \to E[X]$ and (Ω,F,μ) is purely atomic, then for any σ-algebra C between $\{\phi,\Omega\}$ and F, we have $E[X_n|C](\omega) \to E[X|C](\omega)$ μ-almost surely. A proof of our main result is similar to D. Blackwell and L. Dubins' with more restriction on construction of the sub-σ-algebra C since we work only in the original space. Thus we are able to show that the result of D. Blackwell and L. Dubins applies to the original probability space if it is a nonatomic space.

CHAPTER 1

RESULTS FROM MEASURE AND PROBABILITY THEORY

In this chapter we shall discuss some basic and important results from general measure and probability theory. We refer the readers to the basic texts by P.R. Halmos [7], H. Royden [14], and J. Doob [4] for proof and details.

§1 Notations and definitions

In a probability theory, by a set Ω (a sample space), we mean a collection of certain events called elementary events and denote each of these elements by the Greek letter ω. The members of the set Ω are considered to be outcomes from a certain experiment, that is, elementary events are minimal events which are disjoint and one of these events is bound to occur in the experiment. We denote a set of these outcomes by the symbols A, B,... and call them the events A, B,... . It is clear from such an identification of events with sets that the ϕ,Ω are special events, and we shall refer to them as the impossible event and the sure event respectively. With each event A we associate the complementary event, denoted by A^c, such that an event A^c occurs if and only if an event A does not occur.

Before we state the axiomatic definition of the probability theory formulated in its present form by A.N. Kolmogorov in 1933, we shall introduce a few definitions.

Definition 1.1.1 An algebra of sets or a Boolean algebra is a collection F of subsets of Ω which satisfies the following conditions:

i) $A \cup B \in F$ whenever $A, B \in F$.

ii) $A^c \in F$ whenever $A \in F$.

A σ-algebra (Borel field, σ-field) is an algebra F with the following additional condition.

iii) If $\{A_n\}_{n=1}^{\infty}$ is a sequence of members of F then $\bigcup_{n=1}^{\infty} A_n \in F$.

From the above conditions on F, we note that whenever conditions ii, iii are satisfied then $\bigcap_{n=1}^{\infty} A_n \in F$ from De Morgan's formula.

Let Ω be a space, F a σ-algebra of subsets of Ω. If $A \subset \Omega$ then we say σ-algebra $\{F \cap A: F \in F\}$ is a restriction of the σ-algebra F to A and we denote it by $F|_A$.

Definition 1.1.2 Let Ω be a space, F a σ-algebra of subsets of Ω. A set function μ on F is a function which associates an extended real number to each member of F. A set function μ is said to be a countably additive measure (or simply measure) if it satisfies the following axioms:

i) $0 \leq \mu(F) \leq \infty$ for all $F \in F$.

ii) $\mu(\phi) = 0$.

iii) $\mu(\bigcup_{n=1}^{\infty} F_n) = \sum_{n=1}^{\infty} \mu(F_n)$ for all pairwise disjoint sequence

$\{F_n , n = 1,2\ldots\}$ such that $F_n \in F$.

Definition 1.1.3 Let C be a sub-σ-algebra contained in an σ-algebra F , μ' be a measure on C , and μ be a measure on F . If μ' and μ take the same values on sets in C then we say μ' is a restriction of μ and denote it by $\mu|_C$.

Definitions 1.1.4 Let Ω be a space, F a σ-algebra of subsets of Ω . The order pair (Ω,F) is said to be a measurable space and (Ω,F,μ) is called a measure space if μ satisfies definition 1.1.2.

Consider a measurable space (Ω,F,μ) . A partition of Ω is a finite or infinite disjoint sequence $\{F_n , n = 1,2\ldots\}$ of sets such that $\bigcup_{n=1}^{\infty} F_n = \Omega$. A measure μ on F is called a finite measure if and only if $\mu(\Omega) < \infty$. A measure μ on F is said to be a σ-finite measure if there is a partition of Ω such that $\mu(F_n) < \infty$ for each n .

Definition 1.1.5 A measure space (Ω,F,μ) with a σ-finite measure μ (a finite measure μ) is said to be a σ-finite measure space (a finite measure space).

Now we are in a position to state the axiomatic definition of probability space formulated by A.N. Kolmogorov.

Definition 1.1.6 A probability space (p.s.) (Ω,F,μ) is a measure space with $\mu(\Omega) = 1$. The members of F are then called events.

§2 Random variables

Definition 1.2.1 Let (Ω,F) and (E,E) be measurable spaces. A mapping $X: \Omega \to E$ is said to be measurable, or a random variable (henceforth abbreviated to r.v.), if and only if $X^{-1}(A) \in F$ for all $A \in E$. Frequently, we will write this as $X: (\Omega,F) \to (E,E)$.

In this paper we shall only be concerned with a real valued r.v.

Definition 1.2.2 Let $\{X_i , i \in I\}$ be a family of r.v. The σ-algebra generated by $\{X_i^{-1}(A): A \in E_i , i \in I\}$ is called the σ-algebra generated by the random variables $\{X_i , i \in I\}$ and is denoted by $T(\{X_i , i \in I\})$.

The existence of such a σ-algebra is the consequence of the following theorem.

Theorem 1.2.3 If C is a class of subsets of Ω , then there is a minimal σ-algebra, denoted by $T(C)$, containing all the sets in C .

Proof P.R. Halmos [7], page 26.

We shall denote the σ-algebra generated by all the open intervals of R by B_1 . Hence our r.v. will be in $X: (\Omega,F) \to (R,B_1)$.

With these r.v., we associate a function called the probability distribution of a r.v.

<u>Definition 1.2.4</u> Let X be a r.v. on p.s. (Ω,F,μ) . The probability distribution of r.v. X is a set function, $\mu_X(B) = \mu(\{\omega: X(\omega) \in B\})$ for all $B \in \mathcal{B}_1$, defined on $\mathcal{B}_1$.

From our definition of μ_X , it is clear that μ_X is a probability measure on $(\Omega,\mathcal{B}_1)$.

<u>Definition 1.2.5</u> A <u>discrete r.v.</u> is a r.v. taking at most countably many different values $a_1,a_2,\ldots$ with $\mu(\{a_n\}) = P_n > 0$ and $\sum_n P_n = 1$. A <u>continuous r.v.</u> has $\mu(\{a\}) = 0$ for all $a \in R$.

We note that if a σ-algebra C is generated by a discrete random variable then it is equivalent to the σ-algebra generated by a partition of Ω .

<u>Definition 1.2.6</u> Let $X_1, X_2,\ldots,X_n$ be r.v. on a p.s. (Ω,F,μ) . Then the function $\mu_{X_1,\ldots,X_n}(B_1\cap\ldots\cap B_n) = \mu(\{\omega: X_1(\omega) \in B_1,\ldots,X_n(\omega) \in B_n\})$ for $B_1,\ldots,B_n \in \mathcal{B}_1$ is called the joint probability distribution (joint distribution) of $X_1,\ldots,X_n$. For the case of $(X_1,X_2,\ldots)$, joint probability distribution is defined similarly.

In our work we do not require the concept of joint distribution but we have stated the above definition to clarify the statement of D. Blackwell and L. Dubins' result [2], theorem 1.

§3 <u>Independence</u>

In this section we state the very famous lemma called the Borel-Cantelli lemma which we need for the proof of our main theorem. Since the statement of the Borel-Cantelli lemma involves the idea of sequence of independent events we shall state a few definitions concerning the notion of an independence.

<u>Definition 1.3.1</u> Let $X_1, X_2, \ldots, X_n$ be the r.v. on p.s. (Ω, F, μ), then they are said to be independent if for any $B_1, B_2, \ldots, B_n \in B_1$, $\mu(X_1^{-1}(B_1) \cap \ldots \cap X_n^{-1}(B_n)) = \prod_{i=1}^{n} \mu(X_i^{-1}(B_i))$.

<u>Definition 1.3.2</u> Let (Ω, F, μ) be a p.s. and $F_1, \ldots, F_n$ be sub-σ-algebras contained in F. Then they are said to be independent if for any sets $F_1 \in F_1, \ldots, F_n \in F_n$ $\mu(F_1 \cap \ldots \cap F_n) = \prod_{i=1}^{n} \mu(F_i)$.

If A, B are events in F, then A, B are said to be independent if and only if $\mu(A \cap B) = \mu(A) \cdot \mu(B)$.

From the above two definitions, it is clear that the independence of r.v. is equivalent to the independence of σ-algebras generated by the r.v. $X_1, \ldots, X_n$.

The r.v. of an <u>infinite family</u> are said to be independent if and only if those in every finite subfamily are.

We now state the Borel-Cantelli lemma.

<u>Theorem 1.3.3</u> (Borel-Cantelli lemma) If the events $\{F_n, n = 1, 2 \ldots\}$ $F_n \in F$ are independent, then $\mu(\overline{\lim} F_n) = 0$ or 1 according as $\sum_{n=1}^{\infty} \mu(F_n) < \infty$ or $= \infty$.

<u>Proof</u> L. Breiman [3], pages 41-42, J. Doob [4], page 104.

§4 Convergence

We shall review some of the terminology used to describe the various convergence concepts of a sequence of random variables.

Definition 1.4.1 If $\{X_n , n = 1,2\ldots\}$ is a sequence of r.v. then the sequence $\{X_n , n = 1,2\ldots\}$ converges pointwise to a r.v. X on Ω if and only if, for any $\omega \in \Omega$ and $\varepsilon > 0$, there exists an integer $n_0 \in N$ such that for any $n \geq n_0$, $|X_n(\omega) - X(\omega)| < \varepsilon$. If a sequence of r.v. X_n converges pointwise to a r.v. X for all $\omega \in \Omega$ except possibly for those ω belonging to a set of probability zero, then X_n is said to converge almost surely (a.s.) or almost everywhere (a.e.) to the r.v. X .

Another important concept in convergence of a sequence of r.v. X_n in probability theory is convergence in probability or in measure.

Definition 1.4.2 A sequence of r.v. X_n is said to converge in probability (in measure) to X if for every $\varepsilon > 0$, $\lim_n \mu(\{\omega: |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0$

We state the following theorem to show some relationship between the two concepts of convergence.

Theorem 1.4.3 Let $\{X_n , n = 1,2\ldots\}$ be a sequence of r.v. converging a.s. to a r.v. X , then $\{X_n , n = 1,2\ldots\}$ converges in probability to X .

Proof P.R. Halmos [7], page 92.

However, the converse does not hold. For example, see [8], page 175.

Let us denote by $L_1(\Omega,F,\mu) = L_1$ the space of all r.v. X such that $\int|X|d\mu < \infty$. A sequence of r.v. $X_n \in L_1$ is said to converge in mean if and only if $\int|X_n - X|d\mu \to 0$ as $n \to \infty$.

The following theorem shows some relationship between the convergence in probability and in mean.

Theorem 1.4.4 If $\{X_n, n = 1,2...\}$ converges in mean to X then it converges in probability to X .

Proof P.R. Halmos [7], pages 104, 110.

For the converse statement of the above theorem, we need the following definition.

Definition 1.4.5 The r.v. X is said to be dominated by a r.v. Y if $|X| \leq Y$ μ-a.e. and a sequence of r.v. $\{X_n , n = 1,2...\}$ is said to be dominated by a r.v. Y if for each n Y_n is dominated by Y .

Theorem 1.4.6 If $\{X_n , n = 1,2...\}$ converges in probability to X and is dominated by some integrable r.v. Y then $\{X_n , n = 1,2...\}$ converges in mean to X .

Proof P.R. Halmos [7], page 110.

We state the following theorem which we require for the proof of our later theorem.

Theorem 1.4.7 If $\{X_n , n = 1,2...\}$ converges a.e. to a r.v. X and $\int|X_n|d\mu \to \int|X|d\mu$. Then $\int|X_n - X| \to 0$ as $n \to \infty$.

Proof E. Hewitt and K. Stromberg [8], page 209.

We conclude this chapter by stating the following very important theorem in integration theory.

Theorem 1.4.8 (Lebesgue dominated convergence theorem) Let $\{X_n , n = 1,2...\}$ be a sequence of r.v. converging μ-a.e. to a r.v. X . If there exists an integrable r.v. Y such that $\{X_n , n = 1,2...\}$ is dominated by Y μ-a.s., then X is integrable and $\lim_n \int X_n d\mu = \int X d\mu$.

Proof P.R. Halmos [7], H. Royden [14], page 229.

CHAPTER 2

CONDITIONAL PROBABILITIES AND EXPECTATIONS

A conditional expectation is one of the fundamental notions in probability theory and is a most frequently used concept. Before we introduce the formal definition and its basic underlying ideas, we shall make some comments about its usefulness and importance in probability theory. The concept of conditional expectations is used widely in Martingale and Markov theory. (For definitions, see L. Brieman [3], J. Doob [4].) One can also find some applications of martingale theory in the field of continuous-parameter stochastic processes, J. Doob [4], pages 190-370.

A. Rényi considers that the proper notion of probability theory is conditional probability and he has developed a new model of probability theory based on conditional probabilities. Although the axiomatic foundation of probability theory developed by A.N. Kolmogorov and others was satisfactory from a purely mathematical point of view, there arose some problems where Kolmogorov's model did not apply. One of the common features of these problems was that unbounded measures were used. (A. Rényi [12], pages 38-53.)

Because such an unpleasant situation arose in Kolmogorov's model, A. Rényi has developed the new model called the conditional probability space by allowing the usage of unbounded measures. This new model has Kolmogorov's model as a special case of it (theorem 2.3.1 [12], page 50.)

§1 Definitions and basic ideas

In probability theory, the relationship between the occurrence of an event A given that an event B has occurred, or the distribution of one set of r.v. given information concerning the observed values of another set, is introduced as the conditional probability.

In this section we begin with a conditional probability from an elementary point of view then characterize it as a Rádon-Nikodym derivative with respect to certain measures.

The basic idea behind a "conditioning" is that if we have some knowledge of a sample space, i.e. occurrence of an event B, then we can get some idea of the occurrence of another event A.

We now state the following definition of conditional probability of an event A given an event B which involves an idea of taking a ratio of two numbers.

Definition 2.1.1 Let $(\Omega,\mathcal{F},\mu)$ be a p.s. then for sets $A,B \in \mathcal{F}$ such that $\mu(B) > 0$, the conditional probability of an event A given that B has already occurred is defined to be a ratio $\mu(A \cap B)/\mu(B)$ and is denoted by the symbol $\mu(A|B)$.

For fixed B, $\mu(B) > 0$, we note that $\mu(\cdot|B)$ is a set function on $\mathcal{F}$.

The above definition does agree with our discussions and one's intuition as to what such a probability should be, since if an event B is known to occur, then the probability space is reduced to $(\Omega,\mathcal{F},\mu(\cdot|B))$. It is easy to see that $(\cdot|B)$ defines a probability measure on $\mathcal{F}$. The integral of a r.v. $X \in L_1$ with respect to this probability measure is said to be the conditional expectation of X given B and is defined as a point function.

$$E[X|B](\omega) = \int X \, d\mu(\cdot|B) = \frac{1}{\mu(B)} \int_B X \, d\mu \quad \text{on} \quad B . \tag{2.1.1}$$

We note that conditional expectation of X given B^c can also be defined in the same manner.

Consider a r.v. $Y: \Omega \to \{0,1\}$. Then from the above discussions we can consider the conditional expectation of X given Y to be the point function $E[X|Y](\omega)$ such that

$$E[X|Y](\omega) = E[X|\{\omega: Y(\omega) = 0\}](\omega) = \frac{1}{\mu(\{\omega: Y(\omega) = 0\})} \int_{\{\omega: Y(\omega) = 0\}} X d\mu \quad \text{for}$$

each $\omega \in \{\omega: Y(\omega) = 0\}$,

$$E[X|Y](\omega) = E[X|\{\omega: Y(\omega) = 1\}](\omega) = \frac{1}{\mu(\{\omega: Y(\omega) = 1\})} \int_{\{\omega: Y(\omega) = 1\}} X d\mu \quad \text{for}$$

each $\omega \in \{\omega: Y(\omega) = 1\}$.

Then $\int_{\{\omega: Y(\omega) = i\}} E[X|Y](\omega) \, d\mu|_{T(Y)} = \int_{\{\omega: Y(\omega) = i\}} X d\mu \, , \; i = 0,1$.

We can extend the above notion of conditional expectation of X given Y to the case where Y has a countably many values $a_1,a_2,\ldots$. Specifically, for each $\omega \in \{\omega: Y(\omega) = a_n\}$ define

$$E[X|Y](\omega) = \frac{1}{\mu(\{\omega: Y(\omega) = a_n\})} \int_{\{\omega: Y(\omega) = a_n\}} X d\mu .$$

Then we have

$$\int_{\{\omega: Y(\omega) = a_n\}} E[X|Y](\omega) \, d\mu|_{\mathcal{T}(Y)} = \int_{\{\omega: Y(\omega) = a_n\}} X d\mu \quad \text{for } n = 1,2\ldots .$$

Since $\mathcal{T}(Y)$ in the above equation is equivalent with the σ-algebra generated by the partition $\{B_n: B_n = \{\omega: \{Y(\omega) = a_n\}, n = 1,2\ldots\}$, we have

$$\int_B E[X|Y]d\mu|_{F_Y} = \int_B X \, d\mu \quad \text{for all } B \in \mathcal{T}(Y) .$$

This means that the above integral equation does not depend on the value of Y, but depends rather on the information regarding the location of ω in Ω.

In view of this we can define the conditional expectation of X given $\mathcal{T}(Y)$, $E[X|\mathcal{T}(Y)]$, as a point function

$$E[X|\mathcal{T}(Y)](\omega) = \frac{1}{\mu(B_n)} \int_{B_n} X \, d\mu \quad \text{for each } \omega \in B_n , n = 1,2\ldots .$$

We note that if we wish to consider $E[X|\mathcal{T}(Y)]$ to be a function defined on Ω then all we need to consider is the following sum,

$$E[X|\mathcal{T}(Y)](\omega) = \sum_{n=1}^{\infty} \left(\frac{1}{\mu(B_n)} \int_{B_n} X \, d\mu \right) 1_{B_n}(\omega) .$$

Let $\mathcal{B}$ be a σ-algebra generated by a partition $\{B_1, B_2 \ldots\}$ of Ω, then we have the following definition.

Definition 2.1.2 Let X be a r.v. with $X \in L_1(\Omega, \mathcal{F}, \mu)$. The conditional expectation of X with respect to the σ-algebra $\mathcal{B}$ denoted by $E[X|\mathcal{B}]$ is defined as

$$E[X|\mathcal{B}](\omega) = \sum_{n=1}^{\infty} E[X|B_n]\, 1_{B_n}(\omega) .$$

Since many of the interesting r.v. are of a continuous type (information given by continuous functions), we wish to consider the case when conditional expectations have a continuous r.v. as its "conditional part". We can not define the conditional expectation conditioned by a continuous r.v. in the same manner as the conditional expectation conditioned by a discrete r.v. or equivalently conditioned by a σ-algebra generated by a partition of Ω, since $\mu(\{\omega\colon\ Y(\omega) = a\}) = 0$ for all $a \in R$.

However, let us take a closer look at our discrete case. Let Y be a discrete r.v. with values $a_1, a_2 \ldots$, $\mathcal{B} = \mathcal{T}(\{B_n\colon\ B_n = \{Y^{-1}(a_n)\},\ n = 1,2\ldots\})$. Then from our discussion on discrete case

$$E[X|\mathcal{B}](\omega) = \frac{1}{\mu(B_n)} \int_{B_n} X\, d\mu \quad \text{for each } \omega \in B_n\ ,\ n = 1,2\ldots .$$

Now let us define the measure ν on $\mathcal{B}$ by the formula

$$\nu(B_n) = \int_{B_n} X\, d\mu .$$

Then $E[X|\mathcal{B}]$ is a ratio of two measures $\frac{\nu(B_n)}{\mu(B_n)}$ for $\omega \in B_n$, $n = 1,2\ldots$. Let f be a point function on B_n such that $f(\omega) = \frac{\nu(B_n)}{\mu(B_n)}(\omega) = \frac{\nu(B_n)}{\mu(B_n)}$ for $n = 1,2\ldots$. Then f is $\mathcal{B}$-measurable and

$$\int_{B_n} f(\omega)d\mu|_{\mathcal{B}} = \int_{B_n} X\, d\mu = \nu(B_n) \ , \ n = 1,2\ldots \ .$$

We shall use these two properties of f to characterize a conditional expectation of $X \in L_1$ given Y or equivalently a conditional expectation of X given $\mathcal{T}(Y)$, where Y is a continuous or discrete r.v.

In order to do this, we require the notion of taking a "derivative" of one measure with respect to another measure.

Consider the measure ν in our discrete case. We note that for any $B \in \mathcal{B}$, if $\mu(B) = 0$ then $\nu(B) = 0$. We shall state the following definition to formalize this.

Definition 2.1.3 Let $(\Omega,\mathcal{F},\mu)$ be a σ-finite measure space and ν a signed measure on $\mathcal{F}$, then ν is said to be absolutely continuous with respect to the measure μ if $\nu(F) = 0$ whenever $\mu(F) = 0$. Symbolically, this is denoted by $\nu \ll \mu$.

Note that we have $\nu \ll \mu|_{\mathcal{B}}$ in our discrete case.

We state the following theorem which states that if measure ν is absolutely continuous with respect to μ then under certain conditions ν can always be defined as an indefinite integral.

Theorem 2.1.4 (Rádon-Nikodym) Let $(\Omega,\mathcal{F},\mu)$ be a σ-finite measure space, and let ν be a σ-finite signed measure on F which is absolutely continuous with respect to μ, then there exists a finite valued measurable function $\frac{d\nu}{d\mu}$ on Ω such that

$$\nu(F) = \int_F \frac{d\nu}{d\mu}\, d\mu \quad \text{for all} \quad F \in F \ .$$

$\frac{d\nu}{d\mu}$ is unique upto a set of measure zero.

<u>Proof</u> P.R. Halmos [7], page 128.

In view of this result, we see that the function $\frac{d\nu}{d\mu}$ has the properties of f stated earlier.

Let Y be a continuous r.v. on a p.s. (Ω, F, μ). If we define ν to be a finite signed measure on $T(Y)$ by

$$\nu(F_y) = \int_{F_Y} X\, d\mu \quad \text{for each} \quad F_Y \in T(Y) \ , \text{ then}$$

$\nu \ll \mu_{|T(Y)}$ and from the Rádon-Nikodym theorem applied to $(\Omega, T(Y), \mu_{|T(Y)})$ we have

$$\nu(F_Y) = \int_{F_Y} \frac{d\nu}{d\mu_{|T(Y)}}\, d\mu_{|T(Y)} = \int_{F_Y} X\, d\mu \quad \text{for all} \quad F_Y \in T(Y) \ .$$

We now are in a position to state the following general definition of conditional expectation of an integrable r.v. X conditioned by a sub-σ-algebra $C \subset F$.

<u>Definition 2.1.5</u> Let (Ω, F, μ) be a p.s., let C be a sub-σ-algebra of F and X an integrable r.v. then the <u>conditional expectation</u> (c.e.) of X given C, denoted by $E[X|C]$ is any C-measurable r.v. such that

$$(*) \qquad \int_c E[X|C](\omega)\, d\mu_{|C} = \int_c X(\omega)\, d\mu \quad \text{for all} \quad c \in C .$$

The existence of $E[X|C]$ is assured by the Rádon-Nikodym theorem, however, $E[X|C]$ is unique only upto sets of measure zero since any C-measurable function satisfying the equation * can be considered as a conditional expectation of X given C.

We can define the conditional probability of an event $A \in F$ given a sub-σ-algebra C in the same manner.

<u>Definition 2.1.6</u> The <u>conditional probability</u> (c.p.) of $A \in F$ given sub-σ-algebra $C \in F$ is a r.v. $\mu(A|C)$ on a p.s. (Ω, F, μ) such that $\int_c \mu(A|C)\, d\mu = \mu(A \cap c)$ for all $c \in C$.

By letting $X = 1_A$ in our definition of c.e., it is clear that the c.p. of A given C is a special case of c.e.. We also note that c.p. can be considered as a function on $F \times \Omega$. That is, for each fixed $\omega \in \Omega$ we can consider $\mu(\cdot|C)(\omega)$ as a set function on Ω from definition 2.1.6.

If for μ-almost all $\omega \in \Omega$, $\mu(\cdot|C)$ is a probability measure on F, then we say the conditional probability is a regular c.p. (definition, L. Brieman [3], page 77).

If c.p. is a regular conditional probability, then we can define the conditional expectation of $X \in L_1$ given σ-algebra $C \subset F$ by the following equation

$$E[X|C](\omega) = \int X \, d\mu(\cdot|C)(\omega) \text{ a.s.}$$

Proposition 4.28, L. Brieman [3], page 77.

§2 Properties of conditional expectations

From the definition of c.e., we observe that the c.e. and expectation considered as a constant function have analogous properties. We shall state such properties without proof and refer the readers to J. Doob [4], and L. Brieman [3].

Let (Ω,F,μ) be a probability space.

Theorem 2.2.1 Let X, $Y \in L_1(\Omega,F,\mu)$, $a,b,c \in R$ and $C \subset F$ be a sub-σ-algebra, then the following properties hold.

i) If $X \geq 0$ μ-a.s. then $E[X|C] \geq 0$ μ-a.s. and if $X = c$ then $E[X|C] = c$ μ-a.s.

ii) $E[aX + bY|C] = aE[X|C] + bE[Y|C]$ μ-a.s.

iii) If $X \leq Y$ μ-a.s. then $E[X|C] \leq E[Y|C]$ μ-a.s.

iv) $|E[X|C] \leq E[|X||C]$ μ-a.s.

v) Let $X_n \geq 0$ and X_n converges monotonically to a r.v. $X \in L_1$. Then $E[X_n|C]$ converges μ-a.s. to $E[X|C]$.

vi) Let Y be an integrable r.v. and X a C-measurable r.v. such that $XY \in L_1$, then

$E[XY|\mathcal{C}] = X\, E[Y|\mathcal{C}]$ μ-a.s.

vii) If $X_n \to X$ in p-mean, then $E[X_n|\mathcal{C}] \to E[X|\mathcal{C}]$ in p-mean for $p \geq 1$.

viii) If a r.v. X is integrable and $\mathcal{B} \subset \mathcal{C} \subset \mathcal{F}$ are the sub-σ-algebras of $\mathcal{F}$, then $E[E[X|\mathcal{C}]|\mathcal{B}] = E[X|\mathcal{B}] = E[E[X|\mathcal{B}]|\mathcal{C}]$ μ-a.s.

Theorem 2.2.2 If $\mathcal{T}(X)$ and a σ-algebra $\mathcal{C} \subset \mathcal{F}$ are independent then $E[X|\mathcal{C}] = E[X]$ μ-a.s.

Proof For every $c \in \mathcal{C}$, $\int_c E[X|\mathcal{C}]\, d\mu_{|\mathcal{C}} = \int_c X\, d\mu = E[X1_c] = E[X] \cdot \mu(c) = \int_c E[X]\, d\mu_{|\mathcal{C}}$. The equation $E[X1_c] = E[X] \cdot \mu(c)$ follows from the independence of $\mathcal{T}(X)$ and $\mathcal{C}$.

By considering the extreme cases of sub-σ-algebras $\{\phi,\Omega\}$ and $\mathcal{F}$, we have $E[X|\{\phi,\Omega\}] = E[X]$ and $E[X|\mathcal{F}] = X$ μ-a.s. Clearly, if we let $\mathcal{C} = \{\phi,\Omega\}$ in theorem 2.2.2, then $E[E[X|\mathcal{F}]|\{\phi,\Omega\}] = E[X|\{\phi,\Omega\}] = E[X]$.

In the next section we shall give some characterization of c.e. as a linear transformation by noting similar properties shared by c.e. and linear transformations of a certain type.

§3 A conditional expectation as a linear transformation

From our characterization of $E[X|\mathcal{C}]$ as a Rádon-Nikodym derivative in section one of this chapter, if we consider $E[\cdot|\mathcal{C}]$ to be a mapping from $L_1(\Omega,\mathcal{F},\mu)$ into $L_1(\Omega,\mathcal{F},\mu)$, then we note that $E[\cdot|\mathcal{C}]$ is a contractive mapping taking a $\mathcal{F}$-measurable r.v. X into a $\mathcal{C}$-measurable r.v. $E[X|\mathcal{C}]$.

Here, we shall briefly give an outline of such a development. One of the reasons why one considers such a characterization comes from the study of dynamics of turbulence, where averaging operators are being used, G. Birkhoff [1].

In her paper, S.C. Moy has given one such characterization of c.e. [10], page 61, theorem 2.2: Let T be a linear transformation from L_1 into L_1 such that

i) $\|T\| \leq 1$,

ii) if X is bounded, then TX is bounded,

iii) $T(XTY) = TX \cdot TY$ for all bounded X , $Y \in L_1$,

iv) $T1 = 1$,

then $TX = E[X|\mathcal{B}]$ for all $X \in L_1$ where $\mathcal{B} = \{A \in \mathcal{F}: T(1_A \cdot X) = 1_A TX$ for all bounded $X\}$.

The above result of Moy was generalized by G.C. Rota [13], page 58, theorem 1, M. Olson [11], Z. Sidak [15], and R.G. Douglas [5].

CHAPTER 3

THE ALMOST SURE CONVERGENCE OF CONDITIONAL EXPECTATIONS

In this chapter we shall investigate the almost sure convergence of a sequence $\{E[X_n|C] , n = 1,2...\}$ when a sequence $\{X_n , n = 1,2...\}$ is assumed to converge almost surely to a r.v. X .

§1 Introduction

In his text [4], J. Doob has shown that if a sequence of non-negative integrable r.v. X_n converges μ-a.s. to an integrable r.v. X and $\sup_n X_n \in L_1$, then for any sub-σ-algebra $C \subset F$, the sequence $\{E[X_n|C] , n = 1,2...\}$ converges $\mu|C$ -a.s. to a r.v. $E[X|C]$. We shall refer to this fact as the generalized Lebesgue dominated convergence theorem (GLDCT).

The above theorem was noted by Doob in order to give a characterization of conditional expectation of integrable random variables. More precisely, (theorem 9.1 [4], page 27) in our notation, if $X \in L_1$ and a conditional probability with respect to C forms a probability measure on F (this was discussed in chapter 2, page 21), then

$$E[X|C] = \int_\Omega X \, d\mu(\cdot|C) .$$

In view of GLDCT result, one may wonder whether the converse to the GLDCT holds true. With regard to such a question, D. Blackwell and L. Dubins [2] have shown that in "some sense" the condition $\sup_n X_n \in L_1$ is not only a sufficient condition for the convergence of a sequence $\{E[X_n|C] , n = 1,2\ldots\}$ to a r.v. $E[X|C]$ $\mu|_C$ -a.s. for all sub-σ-algebras $C \subset F$, but is also a necessary condition. More exactly, if $X_n \geq 0$, $X_n \to X$ μ-a.s., X_n , $X \in L_1$ and $\sup_n X_n \notin L_1$, there are, on a suitable p.s. (Ω^*,F^*,μ^*) , r.v. $\{X_n^* , n = 1,2\ldots\}$, X^* , and a σ-algebra $C^* \subset F^*$ such that $X^*,X_1^*,\ldots$ have the same joint distribution as $X,X_1,X_2,\ldots$ and

$$\mu(\{\omega^*: \; E[X_n^*|C^*](\omega^*) \to E[X^*|C^*](\omega^*)\}) = 0 \; .$$

In view of D. Blackwell and L. Dubins' result, it is interesting to investigate characteristics of probability space (Ω,F,μ) for which we can conclude that if $\sup_n X_n \notin L_1$, then there exists a sub-σ-algebra C such that

$$\mu(\{\omega: \; E[X_n|C](\omega) \to E[X|C](\omega)\}) = 0 \; .$$

That is, under what condition is it necessary to construct a <u>new</u> suitable p.s. in D. Blackwell and L. Dubins' result? In corollary 3.2.6, we show that if a p.s. (Ω,F,μ) is atomic and $E[X_n] \to E[X]$ then for <u>any</u> σ-algebra $C \subset F$, $E[X_n|C]$ converges μ-a.s. to $E[X|C]$ with probability one, i.e., if the probability space (Ω,F,μ) is atomic, then the construction of a new suitable probability space may be necessary in order to obtain the

σ-algebra C for which convergence does not hold. We also note the following trivial fact: if $E[X_n] \not\to E[X]$ then there always exists a sub-σ-algebra C such that $E[X_n|C] \not\to E[X|C]$ with probability one by taking $C = \{\phi, \Omega\}$. If the p.s. (Ω, F, μ) is assumed to be a nonatomic p.s. then theorem 3.3.1 asserts that there exists a σ-algebra $C \subset F$ such that

$$\mu(\{\omega: E[X_n|C](\omega) \to E[X|C](\omega)\}) = 0 .$$

§2 Atomic and Nonatomic spaces

In this section we shall establish a few useful facts about atomic and nonatomic spaces. In the following discussion, let μ be a probability measure on a σ-algebra F of subsets of a sample space Ω.

Definitions 3.2.1 An atom of the measure μ is a set $A \in F$ such that $\mu(A) > 0$ and if $F \subset A$, $F \in F$ then $\mu(F) = \mu(A)$ or $\mu(F) = 0$. The probability measure μ is called atomic if $\Omega = \bigcup_{i=1}^{\infty} A_i$, $A_i \cap A_j \neq 0$ $i \neq j$ where A_i are atoms of the measure μ. A p.s. (Ω, F, μ) is said to be purely atomic or atomic if μ is atomic measure.

Definition 3.2.2 The probability measure μ is called atomless if μ has no atoms in F. A p.s. (Ω, F, μ) is said to be nonatomic if there is no atoms of μ in F.

Theorem 3.2.3 (Decomposition theorem of a p.s. Ω) Let (Ω, F, μ) be a p.s., then $\Omega = \Omega_1 \cup \Omega_2$, where $\Omega_1 = \bigcup_{i=1}^{\infty} A_i$, where A_i are atoms of $F_{|\Omega_1}$ and $\mu_{|F_{|\Omega 2}}$ has no atoms in $F|\Omega$.

Proof Hahn and Rosenthal [6], page 51.

We state and prove the following lemma which will be used to prove theorem 3.2.5.

Lemma 3.2.4 If a probability space (Ω,F,μ) is atomic and $X_n \to X$ in probability, then $X_n \to X$ μ-a.s.

Proof Suppose $X_n \to X$ in probability, then from the definition of convergence in probability we have for any $\varepsilon > 0$, $\lim_n \mu(\{\omega\colon |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0$. Now if $X_n \not\to X$ μ-a.s. then there exists an atom A_{i_0} such that $X_n \not\to X$ pointwise anywhere on A_{i_0}. Then for some $\varepsilon_0 > 0$ $A_{i_0} \subset \{\omega\colon |X_n(\omega) - X(\omega)| \geq \varepsilon_0\}$ for infinitely many n, hence $\mu(\{\omega\colon |X_n(\omega) - X(\omega)| \geq \varepsilon_0\}) \geq \mu(A_{i_0}) > 0$. This contradicts the assumption.

From the above lemma, we have the following theorem which gives a condition on a p.s. (Ω,F,μ) for the almost sure convergence of a sequence $E[X_n|C]$, $n = 1,2\ldots$ to a r.v. $E[X|C]$ when $X_n \to X$ μ-a.s.

Theorem 3.2.5 If (Ω,F,μ) is a purely atomic probability space and $X_n \geq 0$, $X \geq 0$, X_n, $X \in L_1(\Omega,F,\mu)$, $X_n \to X$ μ-a.s., and $E[X_n]$ converges to $E[X]$, then $E[X_n|C] \to E[X|C]$ μ-a.s. for all sub-σ-algebras $C \subset F$.

Proof From theorem 1.4.7, if $E[X_n] \to E[X]$ then $X_n \to X$ in mean. If $X_n \to X$ in mean, then from theorem 2.2.1 part viii, $E[X_n|C] \to E[X|C]$ in mean for all sub-σ-algebra $C \subset F$. Then $E[X_n|C] \to E[X|C]$ in probability from theorem 1.4.1 which implies $E[X_n|C] \to E[X|C]$ μ-a.s. for all sub-σ-algebras $C \subset F$ by lemma 3.2.4.

Note that if $E[X_n] \to E[X]$, then $E[X_n|C] \to E[X|C]$ in probability on any probability space. Hence, if C is a σ-algebra generated by a partition then $E[X_n|C] \to E[X|C]$ μ-a.s.

There are sequences of integrable random variables X_n such that $\sup X_n \notin L_1(\Omega,F,\mu)$, yet $E[X_n] \to E[X]$. (e.g. Let $X_n = n1_{(\frac{1}{n+1}, \frac{1}{n}]}$.) Incidentally, if $E[X_n] \not\to E[X]$, then it is clear that there exists a sub-σ-algebra C such that $E[X_n|C] \not\to E[X|C]$ μ-a.s. by taking $C = \{\phi,\Omega\}$.

Corollary 3.2.6 Under the same hypothesis as theorem 3.2.5 $\sup_n X_n \in L_1$ is not a necessary condition for the GLDCT.

Proof This follows immediately from the above theorem and comments. Theorem 3.2.5 also follows from the following easy lemma.

Lemma 3.2.7 If a σ-algebra F is generated by a countable partition of subsets of Ω , then any sub-σ-algebra $C \subset F$ is generated by a countable partition of subsets of Ω .

<u>Proof</u> Let $\mathcal{F} = \mathcal{T}(A_n , n = 1,2\ldots : \bigcup_{n=1}^{\infty} A_n = \Omega$ and $A_i \cap A_j = \emptyset , i \neq j)$. Then a proof follows from the fact that any member of $\mathcal{F}$ can be written as a union of A_n's and the intersection of $A_i \cap A_j = \emptyset$, $i \neq j$, and $\Omega \in \mathcal{C}$.

From the above lemma and $E[X_n|\mathcal{C}](\omega) = \frac{1}{\mu(F_n)} \int X_n \, d\mu$ for all $\omega \in F_n$, $\Omega = \bigcup_{n=1}^{\infty} F_n$ we have theorem 3.2.5.

We state the following two lemmas which can be found in Hahn and Rosenthal [6].

<u>Lemma 3.2.8</u> Let $(\Omega,\mathcal{F},\mu)$ be an atomless probability space, then for any $\varepsilon > 0$ there exists a non null set $A \in \mathcal{F}$ such that $\mu(A) < \varepsilon$.

<u>Proof</u> Proof follows from the fact that for any $A \in \mathcal{F}$ there exists $A' \in \mathcal{F}$ such that $\mu(A') < \frac{1}{2}\mu(A)$.

We now state the so called "Intermediate-value theorem" of a measure which can be found in Hahn and Rosenthal [6], pages 52-53.

<u>Theorem 3.2.9</u> Let $(\Omega,\mathcal{F},\mu)$ be a nonatomic probability space. Then for any $\Omega' \subset \Omega$, $\mu(\Omega') > 0$, if $\alpha \in (0, \mu(\Omega'))$ there exists a set $F \in \mathcal{F}$ and $F \subset \Omega'$ such that $\mu(F) = \alpha$.

§3 <u>Properties of probability spaces for which $\sup_n X_n \in L_1$ is a necessary condition of the Generalized Lebesgue Dominated Convergence Theorem (GLDCT).</u>

Here, we will give some properties of probability spaces such that

GLDCT will hold if and only if $\sup_n X_n \in L_1$.

The proof of our theorem is similar to the proof given by D. Blackwell and L. Dubins on theorem 1 of their paper [2]. However, the generators of the sub-σ-algebra C are subject to more restrictions since we are constructing the σ-algebra C on the original probability space (Ω,F,μ) .

Theorem 3.3.1 Let $X_n \geq 0$, X_n , $X \in L_1(\Omega,F,\mu)$, $X_n \to X$ μ-a.s. and $\sup_n X_n \notin L_1(\Omega,F,\mu)$. If the probability space (Ω,F,μ) is atomless then there exists a sub-σ-algebra $C \subset F$ such that

$$\mu(\{\omega: E[Xn|C](\omega) \not\to E[X|C](\omega)\}) = 1 .$$

Proof By following the reduction method employed by D. Blackwell and L. Dubins in their proof, we can reduce the above theorem to the case where each X_n has only two values, 0 , $\nu_n > 0$ and at every sample point ω exactly one X_n has a positive value ν_n . Thus, if $\mu(\{\omega: X_n(\omega) = \nu_n\}) = P_n$, we have $0 < P_n < 1$, $\sum_{n=1}^{\infty} P_n = 1$, $X \equiv 0$, $E[\sup_n X_n] = \sum_{n=1}^{\infty} P_n \nu_n = \infty$.

To attain this reduction, let $F_n(\omega) = \max((X_n - X)(\omega), 0)$, $G_n(\omega) = \min((X_n - X)(\omega), 0)$. Then $F_n \geq 0$, $F_n \in L_1(\Omega,F,\mu)$,

$$\sup_n F_n(\omega) \geq \sup_n X_n - X \notin L_1(\Omega,F,\mu) , \quad F_n \to 0 \quad \mu\text{-a.s.}$$

$$\sup_n |G_n(\omega)| \leq X(\omega) , \quad G_n(\omega) \to 0 \quad \mu\text{-a.s., } \sup |G_n(\omega)| \in L_1 .$$

For any sub-σ-algebra $C \subset F$, $E[G_n|C] \to 0$ μ-a.s. for the GLDCT. Thus,

$$\mu(\{\omega:\ E[X_n|C](\omega) \to E[X|C](\omega)\})$$

$$= \mu(\{\omega:\ E[X_n - X|C](\omega) \to 0\})$$

$$= \mu(\{\omega:\ E[F_n + G_n|C](\omega) \to 0\})$$

$$= \mu(\{\omega:\ E[F_n|C](\omega) \to 0\})\ .$$

Hence, if we produce a σ-algebra $C \subset F$ such that $\mu(\{\omega:\ E[F_n|C] \nrightarrow 0\}) = 0$, then $\mu(\{\omega:\ E[X_n|C] \nrightarrow E[X|C]\}) = 0$. Thus we have reduced the theorem to the special case of the F_n, i.e. to the case $X = 0$.

Before we continue with our reduction, we state the lemma which we require for our reduction.

<u>Lemma 3.3.2</u> Let X_n, Y_n non negative r.v., $X_n \to 0$ μ-a.s., $Y_n \to 0$ μ-a.s. Then for any sub-σ-algebra $C \subset F$, if $Y_n \leq X_n$ for each n and $\mu(\{\omega:\ E[Y_n|C](\omega) \nrightarrow 0\}) = 1$ we have $\mu(\{\omega:\ E[X_n|C](\omega) \nrightarrow 0\}) = 1$.

<u>Proof</u> This follows immediately from theorem 2.2.1 part iii which implies $0 \leq E[Y_n|C](\omega) \leq E[X_n|C](\omega)$.

We now construct a sequence $\{S_n, n = 1,2\ldots\}$ of simple functions such that $0 \leq S_n \leq X_n$ and $\sup_n S_n \notin L_1$ whenever $\sup_n X_n \notin L_1$.

Let $A_k = \{\omega:\ X_k(\omega) \geq \operatorname{Sup}_n X_n - 1,\ X_i < \operatorname{Sup}_n X_n - 1 \text{ for } i < k\}$, then from the defining properties of $\{A_k, k = 1,2\ldots\}$ they are disjoint and since $\Omega - A \subset \bigcup_{k=1}^{\infty} A_k$, $\sum_{k=1}^{\infty} \mu(A_k) = 1$, where A is a null set.

Choose a simple function S_k such that $0 \le S_k \le X_k$, $S_k 1_{A_k^c} = 0$ and $E[S_k] \ge \int_{A_k} X_k \, d\mu - \frac{1}{2^k}$. Then $\sup_k S_k = \sum_{k=1}^{\infty} S_k$ and we have $E[\sup_k S_k] = E[\sum_{k=1}^{\infty} S_k] = \Sigma_k E[S_k] \ge \Sigma_k \int_{A_k} X_k \, d\mu - 1 \ge \Sigma_k \int_{A_k} \sup_n X_n \, d\mu - 2 = E[\sup_k X_n] - 2 = \infty$. Hence, $\sup_n S_n \notin L_1$ whenever $\sup_n X_n \notin L_1$. From the above lemma, since $S_k \le X_k$ for each k we have

$$\mu(\{\omega: \ E[S_k|C](\omega) \to 0\}) = 0 \quad \text{implies}$$

$$\mu(\{\omega: \ E[X_k|C](\omega) \to 0\}) = 0$$

for any sub-σ-algebra $C \subset F$. Therefore, we have reduced the theorem to the case of S_k , i.e. if the theorem holds for S_k then it holds for X_k . We now consider the S_k to be a sum of finite number non negative functions, each having only two values, one of which is 0 , and no two of which are simultaneously positive.

We now rearrange these functions into a single sequence. If any of these functions are zero with probability one then we omit them from the sequence. If there exists the set B on which all of these functions vanish has positive probability, then we add the indicator 1_B as an additional member of the sequence. (This is done for the technical reason, and it is clear that we can do this, for example, we could let 1_B as the first member of our sequence.) Then we have a sequence which we will denote henceforth as X_1, X_2 with the properties stated at the beginning of the section and this completes our reduction.

We now prove the theorem in the above "special" case. For the construction of our σ-algebra C , we shall make a frequent use of theorem

3.2.9 (an intermediate-value theorem). Let us now assume we have this, i.e. $X_n = \nu_n 1_{D_n}$, $0 < \nu_n < 1$, $\mu(D_n) = p_n$, $0 < p_n < 1$, $\Sigma\, p_n = 1$, $X = 0$ and $E[\sup_n X_n] = \Sigma_n\, p_n\, \nu_n = \infty$.

Define:

For, $n = 1,2\ldots$ $I_n = \{i\colon$ i is an integer, $i \geq 2$ and $2^{n+k_0} \leq \nu_i < 2^{n+1+k_0}\}$, where k_0 is a positive integer such that $1 < 2^{k_0} p_1 \leq 2$. Let $I_0 = \{i\colon$ i is an integer, $i \geq 2$, $i \notin I_n$, $n = 1,2\ldots\}$ and $I = \bigcup_{n=1}^{\infty} I_n$,

$$r_n = \sum_{i \in I_n} p_i \ , \ n = 1,2\ldots \ ,$$

$$t_n = r_n + 2^{-(k_0+n)} \ , \ n = 1,2\ldots \ ,$$

$$t = \sum_{n=1}^{\infty} t_n = \sum_{i \in I} p_i + 2^{-k_0}$$

Note that $\{1,2\ldots\} = \{1\} \cup I_0 \cup I$ and the cardinality of I_n is less than or equal to $\aleph_0$, and that there does not exist a number $n_0 \varepsilon N$ such that $I_n = \phi$ for all $n \geq n_0$, otherwise a sequence $\{X_n$, $n = 1,2\ldots\}$ is dominated which means $\sup_n X_n \,\varepsilon\, L$.

By applying theorem 3.2.9, we next construct a sequence of events $A_n \,\varepsilon\, F$ as follows.

Let

$$A_0 \subset (\bigcup_{i \in I} D_i)^c \text{ be such that } \mu(A_0) = 1 - t \ .$$

For $n = 1,2\ldots$, let

$$A_n^* = \bigcup_{i\varepsilon I_n} D_i \text{ , and let}$$

$$*A_n \subset (A_0 \cup (\bigcup_{i\varepsilon I} D_i))^c \text{ be such that}$$

$$\mu(*A_n) = 2^{-(k_0+n)} .$$

For $n = 1,2\ldots$, then let

$$A_n = A_n^* \cup *A_n .$$

Then since $\mu(*A_n) = \mu(\bigcup_{i\varepsilon In} D_i) = rn$ and $\mu(A_n) = \mu(A_n^*) + \mu(*A_n) = \nu n+2^{-(k_0+n)} = t_n$, we have $\mu(A_n) = t_n$, $n = 1,2\ldots$, and also

$$\Omega = \bigcup_{n=0}^{\infty} A_n \cup M \text{ , where } \mu(M) = 0 .$$

Let $P = \{A_0 , A_1,\ldots\}$. We shall refine this "partition" sequentially in order to generate the σ-algebra C .

<u>The first refinement of $P = \{A_0 , A_1\ldots\}$</u>

We shall "partition" each member of P into $\bar{\eta}_0$ events, where $\eta_0 = 1 +$ the cardinality of I_0 . By employing theorem 3.2.9, we can construct for each $n = 0,1,2\ldots$ events $A_n^{i_0}$, $i_0 \varepsilon \{1\} \cup I_0$ as follows.

For each $i_0 \varepsilon \{1\} \cup I_0$, $j \varepsilon I_n$, $n = 1,2\ldots$, let

$$_j A_n^{i_0} \varepsilon F ,$$

$${}_j A_n^{i_0} \subset D_j \ ,$$

$${}_j A_n^i \cap {}_j A_n^k = \phi \quad \text{if} \quad i \neq k \ , \ i,k \in \{1\} \cup I_0 \ , \text{ with measure}$$

$$\mu({}_j A_n^1) = \frac{p_1 - 2^{-k_0}}{1 - t} \mu(D_j) \quad \text{and}$$

$$\mu({}_j A_n^{i_0}) = \frac{p_{i_0}}{1-t} \mu(D_j) \ , \ i_0 \in I_0 \ , \ j \in I_n \ , \ n = 1,2\ldots \ .$$

For each $i_0 \in \{1\} \cup I_0$, $n = 1,2\ldots$, let

$*A_n^{i_0} \in \mathcal{F}$, be such that

$$*A_n^{i_0} \subset *A_n \ ,$$

$$*A_n^i \cap A_n^k = \phi \ , \ i \neq k \ , \ i,k \in \{1\} \cup I_0 \ ,$$

$$\mu(*A_n^1) = \frac{p_1 - 2^{-k_0}}{1 - t} \ \mu(*A_n) \quad \text{and}$$

$$\mu(*A_n^{i_0}) = \frac{p_{i_0}}{1-t} \mu(*A_n) \ .$$

For each $i_0 \in \{1\} \cup I_0$, $n = 1,2\ldots$, let

$$A_n^{i_0} = \bigcup_{j \in I_n} {}_j A_n^{i_0} \cup *A_n^{i_0} \ , \quad \text{then we have}$$

$$A_n^i \cap D_j = {}_j A_n^i \quad \text{and} \quad A_n^i \cap *A_n = *A_n^i \ ,$$

$$\mu(A_n^1) = \frac{p_1 - 2^{-k_0}}{1 - t}\,\mu(A_n) ,$$

$$\mu(A_n^{i_0}) = \frac{p_{i_0}}{1-t}\,\mu(A_n) , \text{ and}$$

for each $i_0 \varepsilon \{1\} \cup I_0$, let

$A_0^{i_0} \varepsilon F$, be such that

$$A_0^{i_0} \subset A_0 ,$$

$$\mu(A_0^1) = \frac{p_1 - 2^{-k_0}}{1 - t}\,\mu(A_0) ,$$

$$\mu(A_0^{i_0}) = \frac{p_{i_0}}{1-t}\,\mu(A_0) . \quad \text{Then}$$

for each $n = 0,1,2...$

$$A_n = \bigcup_{i\varepsilon\{1\}\cup I_0} A_n^i \cup M_n , \quad \text{where} \quad \mu(M_n) = 0 .$$

The following calculation shows that the measures of ${}_jA_n^{i_0}$ and $A_n^{i_0}$ are compatible with the measure of $A_n^{i_0}$ and the measure of $A_n^{i_0}$ is compatible with the measure of A_n .

$$\mu(A_n^{i_0}) = \mu(\bigcup_{j\varepsilon I_n} {}_jA_n^{i_0} \cup {}^*A_n^{i_0})$$

$$= \sum_{j\varepsilon I_n} \mu({}_jA_n^{i_0}) + \mu({}^*A_n^{i_0})$$

$$= \sum_{j\varepsilon I_n} \frac{p_{i_0}}{1-t}\,\mu(D_j) + \frac{p_{i_0}}{1-t}\,\mu(*A_n)$$

$$= \frac{p_{i_0}}{1-t}\,[\mu(A_n^*) + \mu(*A_n)]$$

$$= \frac{p_{i_0}}{1-t}\,\mu(A_n) \ , \ n = 1,2\ldots \ .$$

If $i_0 = 1$, then let the coefficient be $\dfrac{p_1 - 2^{-k_0}}{1 - t}$.

$$\mu(A_n) = \mu(\bigcup_{i\varepsilon\{1\}\cup I_0} A_n^i) = \mu(\bigcup_{i\varepsilon\{1\}\cup I_0}(\bigcup_{j\varepsilon I_n \ j} A_n^i \cup *A_n^i))$$

$$= \sum_{i_0\varepsilon I_0} \frac{p_{i_0}}{1-t}\,\mu(A_n) + \frac{p_1 - 2^{-k_0}}{1 - t}\,\mu(A_n)$$

$$= (\frac{p_1 - 2^{-k_0}}{1 - t} + \sum_{i_0\varepsilon I_0} \frac{p_{i_0}}{1-t})\ t_n$$

$$= \frac{p_1 - 2^{-k_0} + \sum_{i_0\varepsilon I_0} p_{i_0}}{1 - t}\, t_n = t_n \ , \ n = 1,2\ldots \ ,$$

since $p_1 - 2^{-k_0} + \sum_{i_0\varepsilon I_0} p_{i_0} = 1 - t$.

Therefore, from the construction of A_n^i , $\{A_n^i\!: \ i \ \varepsilon \ \{1\} \cup I_0\}$, constitute a "partition" of A_n . Let us denote the collection of all A_n^i , $i \ \varepsilon \ \{1\} \cup I_0 \quad n = 0,1,2\ldots$ by P^{I_0} , the refined "partition" of Ω .

The second "refinement" of P

We now "partition" each member of P^{I_0} into η_1 events, where $\eta_1 = 1 +$ the cardinality of I_1 .

For each $i_0 \in \{1\} \cup I_0$, $i_1 \in \{1\} \cup I_1$, $n = 1,2\ldots$, let ${}_j A_n^{i_0,i_1} \in F$ be such that

$$ {}_j A_n^{i_0,i_1} \subset {}_j A_n^{i_0} , $$

$$ {}_j A_n^{i_0,i} \cap {}_j A_n^{i_0,k} = \phi \text{ if } i \neq k , \; i,k \in \{1\} \cup I_1 , $$

$$ \mu({}_j A_n^{i_0,1}) = \frac{2^{-(k_0+1)}}{t_1} \mu(A_n^{i_0}) , \text{ and} $$

$$ \mu({}_j A_n^{i_0,i_1}) = \frac{p_{i_1}}{t_1} \mu({}_j A_n^{i}) , \; n = 2,3\ldots . $$

For each $i_0 \in \{1\} \cup I_0$, $i_1 \in \{1\} \cup I_1$, $n = 2,3\ldots$, let ${}^*A_n^{i_0,i_1} \in F$ be such that

$$ {}^*A_n^{i_0,i_1} \subset {}^*A_n^{i_0} $$

$$ {}^*A_n^{i_0,i} \cap {}^*A_n^{i_0,k} = \phi , \; i \neq k , \; i,k \in \{1\} \cup I_1 , $$

$$ \mu({}^*A_n^{i_0,1}) = \frac{2^{-(k_0+1)}}{t_1} \mu({}^*A_n^{i_0}) \text{ and} $$

$$ \mu({}^*A_n^{i_0,i_1}) = \frac{p_{i_1}}{t_1} \mu({}^*A_n^{i_0}) . $$ Then let

$$ A_n^{i_0,i_1} = (\bigcup_{j \in I_n} {}_j A_n^{i_0,i_1}) \cup {}^*A_n^{i_0,i_1} \quad \text{for } n = 2,3\ldots . $$

For $n = 1$, let

$$ A_1^{i_0,i_1} = {}_{i_1} A_1^{i_0} , $$

$$A_1^{i_0,1} = {}^*A_1^{i_0} \;, \quad i_0 \in \{1\} \cup I_0 \;, \; i_1 \in I_1 \;.$$

Then for $n = 1,2\ldots$

$$\mu(A_n^{i_0,1}) = \frac{2^{-(k_0+1)}}{t_1} \mu(A_n^{i_0}) \;, \text{ and}$$

$$\mu(A_n^{i_0,i_1}) = \frac{p_{i_1}}{t_1} \mu(A_n^{i_0}) \;, \quad i_0 \in \{1\} \cup I_0 \;.$$

Note that $\mu(A_1^{i_0,i_1}) = \mu(i_1 \; A_1^{i_0}) = \frac{p_{i_0}}{1-t} \mu(D_{i_1}) = \frac{p_{i_0} p_{i_1}}{1-t} = \frac{p_{i_1}}{t_1} \mu(A_1^{i_0})$,

$i_0 \in I_0$, $i_1 \in I_1$ and $\mu(A_1^{1,i_1}) = \mu(i_1 \; A_1^1) = \frac{p_1 - 2^{-k_0}}{1 - t} \mu(D_{i_1}) =$
$\left(\frac{p_1 - 2^{-k_0}}{1 - t}\right) p_{i_1} = \frac{p_{i_1}}{t_1} \left(\frac{p_1 - 2^{-k_0}}{1 - t}\right) \mu(A_i) = \frac{p_{i_1}}{t_1} \mu(A_1^1)$.

For each $i_0 \in \{1\} \cup I_0$, $i_1 \in \{1\} \cup I_1$, let

$A_0^{i_0,i_1} \in F$ be such that

$$A_0^{i_0,i_1} \subset A_0^{i_0}$$

$$A_0^{i_0,i} \cap A_0^{i_0,k} = \phi \;, \quad i \neq k \;, \; i,k \in \{1\} \cup I_1 \;,$$

$$\mu(A_0^{i_0,1}) = \frac{2^{-(k_0+1)}}{t_1} \mu(A_0^{i_0}) \quad \text{and}$$

$$\mu(A_0^{i_0,i_1}) = \frac{p_{i_1}}{t_1} \mu(A_0^{i_0}) \;, \quad i_0 \in \{1\} \cup I_0 \;, \; i_1 \in I_1 \;.$$

The following calculation shows that the measures of $A_n^{i_0,i_1}$ are compatible with the measures of $A_n^{i_0}$, $i_0 \varepsilon \{1\} \cup I_0$, $i_1 \varepsilon \{1\} \cup I_1$.

$$\mu(A_n^{i_0,1}) = \mu(\bigcup_{j\varepsilon I_n} {}_j A_n^{i_0,1}) + \mu({}^*A_n^{i_0,1})$$

$$= \sum_{j\varepsilon I_n} \frac{2^{-(k_0+1)}}{t_1} \mu({}_j A_n^{i_0}) + \frac{2^{-(k_0+1)}}{t_1} \mu({}^*A_n^{i_0})$$

$$= \frac{2^{-(k_0+1)}}{t_1} \mu(A_n^{i_0}) \text{ , similarly,}$$

$$\mu(A_n^{i_0,i_1}) = \frac{p_{i_1}}{t_1} \mu(A_n^{i_0}) \text{ , } i_0 \varepsilon \{1\} \cup I_0 \text{ .}$$

Then $\mu(A_n^{i_0}) = \mu(\bigcup_{i_1\varepsilon\{1\} \cup I_1} A_n^{i_0,i_1}) = \frac{2^{-(k_0+1)}}{t_1} \mu(A_n^{i_0}) + \sum_{i_1\varepsilon I_1} \frac{p_{i_1}}{t_1} \mu(A_n^{i_0}) =$

$$\frac{2^{-(k_0+1)} + \sum_{i_1\varepsilon I_1} p_{i_1}}{t_1} \mu(A_n^{i_0}) \text{ .}$$

Therefore, from the construction of $A_n^{i_0,i_1}$, $i_0 \varepsilon \{1\} \cup I_0$, $i_1 \varepsilon \{1\} \cup I_1$, $A_n^{i_0,i_1}$ form a partition of A_n (i.e. within a null set). Let us denote the collection of all $A_n^{i_0,i_1}$, $i_0 \varepsilon \{1\} \cup I_0$, $i_1 \varepsilon \{1\} \cup I_1$ and the second "refinement" of P by P^{I_0,I_1} .

<u>The ℓ'th "refinement" of P</u>

For the ℓ-1'th refinement of P , let $P_n^{I_0,\ldots,I_{\ell-2}}$ be the collection of events which "partition" A_n . Let η_ℓ = 1 + the cardinality of I_ℓ , we shall "partition" each member of $P_n^{I_0,\ldots,I_{\ell-2}}$ into η_ℓ events as follows:

For each $i_0 \varepsilon \{1\} \cup I_0,\ldots,i_\ell \varepsilon \{1\} \cup I_\ell$, $j \varepsilon I_n$, $n = 1,2\ldots$, $n-1 \neq \ell$, let

$${}_j A_n^{i_0,\ldots,i_\ell} \in F \quad \text{be such that}$$

$${}_j A_n^{i_0,\ldots,i_\ell} \subset {}_j A_n^{i_0,\ldots,i_{\ell-1}} \ , \ i_\ell \in \{1\} \cup I_\ell \ ,$$

$${}_j A_n^{i_0,\ldots,i_{\ell-1},i} \cap {}_j A_n^{i_0,\ldots,i_{\ell-1},k} = \phi \ , \text{ if } \ i \neq k \ , \ i,k \in \{1\} \cup I_\ell \, .$$

$$\mu({}_j A_n^{i_0,\ldots,i_{\ell-1},1}) = \frac{2^{-(k_0+\ell)}}{t_\ell} \mu({}_j A_n^{i_0,\ldots,i_{\ell-1}}) \ , \text{ and}$$

$$\mu({}_j A_n^{i_0,\ldots,i_{\ell-1},i_\ell}) = \frac{p_{i_\ell}}{t_\ell} \mu({}_j A_n^{i_0,\ldots,i_{\ell-1}}) \ , \ i_\ell \in I_\ell \ .$$

For each $i_0 \in \{1\} \cup I_0,\ldots,i_\ell \in \{1\} \cup I_\ell$, $n = 1,2\ldots$, $n-1 \neq \ell$, let

$${}^*A_n^{i_0,\ldots,i_\ell} \in F \quad \text{be such that}$$

$${}^*A_n^{i_0,\ldots,i_\ell} \subset {}^*A_n^{i_0,\ldots,i_{\ell-1}} \ , \ i_\ell \in \{1\} \cup I_\ell \ ,$$

$${}^*A_n^{i_0,\ldots,i_{\ell-1},i} \cap {}^*A_n^{i_0,\ldots,i_{\ell-1},k} = \phi \ , \ i \neq k \ , \ i,k \in \{1\} \cup I_\ell \ ,$$

$$\mu({}^*A_n^{i_0,\ldots,i_{\ell-1},1}) = \frac{2^{-(k_0+\ell)}}{t_\ell} \mu({}^*A_n^{i_0,\ldots,i_{\ell-1}}) \quad \text{and}$$

$$\mu({}^*A_n^{i_0,\ldots,i_{\ell-1},i}) = \frac{p_{i_\ell}}{t_\ell} \mu({}^*A_n^{i_0,\ldots,i_{\ell-1}}) \ .$$

For $n - 1 \neq \ell$, $n = 1,2\ldots$ let

$$A_n^{i_0,\ldots,i_\ell} = (\bigcup_{j \in I_n} A_n^{i_0,\ldots,i_\ell}) \cup {}^*A_n^{i_0,\ldots,i_\ell} \quad \text{and}$$

for $n-1 = \ell$, let

$$A_\ell^{i_0,\ldots,i_\ell} = i_\ell\, A_\ell^{i_0,\ldots,i_{\ell-1}} \;,\; i_\ell \in I_\ell \;, \text{ and}$$

$$A_\ell^{i_0,\ldots,i_{\ell-1},1} = {}^*A_\ell^{i_0,\ldots,i_{\ell-1}} \;. \quad \text{Then}$$

if $i_m \neq 1$ for $m = 0,1,2\ldots,\ell$

$$\mu(i_\ell\, A_\ell^{i_0,\ldots,i_{\ell-1}}) = \frac{p_{i_{\ell-1}}}{t_{\ell-1}}\,\mu(i_\ell\, A_\ell^{i_0,\ldots,i_{\ell-2}})$$

$$= \frac{p_{i_{\ell-1}}}{t_{\ell-1}}\,\frac{p_{i_{\ell-2}}}{t_{\ell-2}}\,\mu(i_\ell\, A_\ell^{i_0,\ldots,i_{\ell-3}})$$

$$\vdots$$

$$= \frac{p_{i_{\ell-1}}}{t_{\ell-1}}\,\frac{p_{i_{\ell-2}}}{t_{\ell-2}}\cdots\frac{p_{i_0}}{1-t}\,\mu(D_{i\ell})$$

$$= \frac{p_{i_\ell}}{t_\ell}\,\frac{p_{i_{\ell-1}}}{t_{\ell-1}}\cdots\frac{p_{i_0}}{1-t}\,t_\ell = \mu(A_n^{i_0,\ldots,i_{\ell-1},i_\ell}) \;.$$

The above calculation shows that when $i_m = 1$ for $m = 0\ldots\ell$ of $A^{i_0,\ldots i_\ell}$ and $i_\ell\, A_\ell^{i_0,\ldots,i_{\ell-1}}$ are compatible. If any of $i_0,\ldots i_{\ell-1}$ is equal to 1 , then change the coefficient p_{i_0} to $p_1 - 2^{-k_0},\ldots,p_{i_{\ell-1}}$ to $2^{-(k_0+\ell)}$, and the compatibility still holds.

The reason we have a special case when the subscripts of A_{n-1} are the same as the stages of our "refinement" is so that we can easily recover the sets D_n .

Thus, we have partitioned each member of $\mathcal{P}_n^{I_0,\ldots I_{\ell-1}}$ into n_ℓ events. The verification of the other statements concerning the compatibility of the measures of events is similar to the previous stages. In this manner, we "refine" the partition $\mathcal{P}$ sequentially.

We are now in position to produce σ-algebra $\mathcal{C} \subset \mathcal{F}$ such that

$$\mu(\{\omega: \ E[X_n|\mathcal{C}](\omega) \not\to 0\}) = 1 \tag{3.1}$$

Let $J_n = (\{1\} \cup I_0) \times (\{1\} \cup I_1) \times \ldots \times (\{1\} \cup I_n)$. Then for $i \in I_n$, $n = 0,1,2\ldots$

Consider the event

$$C_n^i = \bigcup_{k=0}^{\infty}\Big(\bigcup_{(i_0,\ldots,i_{n-1})\in J_{n-1}} A_k^{i_0,\ldots,i_{n-1},i}\Big) \text{ and let}$$

$$\mathcal{C} = \mathcal{T}(C_n^i: \ i \in I_n \ , n = 0,1,2\ldots) \ ,$$

i.e. $\mathcal{C}$ is a σ-algebra generated by the sets C_n^i , $i \in I_n$, $n = 0,1,2\ldots$.

We now claim that $\mathcal{C}$ has the property (3.1).

We have

$$\mu(C_n^i) = \mu\Big(\bigcup_{k=0}^{\infty}\Big(\bigcup_{(i_0,\ldots,i_{n-1})\in J_{n-1}} A_k^{i_0,\ldots,i_{n-1},i}\Big)\Big)$$

$$= \sum_{k=0}^{\infty}\Big(\bigcup_{(i_0,\ldots,i_{n-1})\in J_{n-1}} A_k^{i_0,\ldots,i_{n-1},i}\Big)$$

$$= \sum_{k=0}^{\infty}\frac{p_i}{t_n}\Big(\bigcup_{(i_0,\ldots,i_{n-1})\in J_{n-1}} A_k^{i_0,\ldots,i_{n-1}}\Big)$$

$$= \sum_{k=0}^{\infty} \frac{p_i}{t_n} \mu(A_k) = \frac{p_i}{t_n}$$, since $i_0,\ldots,i_{n-1}$ run through all of their respective index sets and $\mu(A_n^{i_0,\ldots,i_{m-1},i}) = \frac{p_i}{t_m} \mu(A_n^{i_0,\ldots,i_{m-1}})$ and $A_n^{i_0,\ldots,i_{m-1}}$ is a partition of A_n .

We first note that C is independent of $\{A_0, A_1,\ldots\}$. To show this, consider the intersection of generators $\{C_n^i:\ n = 0,1,2\ldots,\ i \in I_n\}$ of the σ-algebra C and A_n ,

$$A_n \cap C_m^i = \bigcup_{(i_0,\ldots,i_{m-1})\in J_{m-1}} A_n^{i_0,\ldots,i_{m-1},i} .$$

Then

$$\mu(A_n \cap C_m^i) = \mu(A_n \cap \bigcup_{k=0}^{\infty} \bigcup_{(i_0,\ldots,i_{m-1})\in J_{m-1}} A_k^{i_0,\ldots,i_{m-1},i})$$

$$= \mu(\bigcup_{(i_0,\ldots,i_{m-1})\in J_{m-1}} A_n^{i_0,\ldots,i_{m-1},i})$$

$$= \frac{p_i}{t_m} \mu(A_n) = \mu(C_m^i) \cdot \mu(A_n)$$

since $A_\ell^{i_0,\ldots,i_{m-1},i} \cap A_k^{i_0,\ldots,i_{m-1},i} = \phi$ if $k \neq \ell$, and $\mu(A_n^{i_0,\ldots,i_{m-1},i}) = \frac{p_i}{t_m} \mu(A_n^{i_0,\ldots,i_{m-1}})$.

The pairwise independence of C_m^i and A_n's implies that C is independent of A_n from Loève [9], page 224.

For $i \in I_n$, $n = 1,2\ldots$, we have

$$D_i = \{\omega:\ X_i = \nu_i\} = A_n \cap \bigcup_{(i_0,\ldots,i_{n-1})\varepsilon J_{n-1}} A_n^{i_0,\ldots i_{n-1},i} = A_n \cap C_n^i \qquad (3.2)$$

(This is the reason why we have taken such care in partitioning A_{n+1} in the n'th refinement.)

We now show that for $i \varepsilon I_n$, $n = 1,2\ldots$ $E[X_i|C](\omega) = \nu_i t_n 1_{C_n^i}$.

$$E[X_i|C](\omega) = E[\nu_i 1_{D_i}|C](\omega)$$

$$= \nu_i E[1_{D_i}|C](\omega) \qquad \text{by (3.2)}$$

$$= \nu_i E[1_{A_n} \cap C_n^i|C](\omega) .$$

Then $\nu_i E[1_{A_n} \cap C_n^i|C](\omega)$

$$= \nu_i 1_{C_n^i} E[1_{A_n}|C](\omega)$$

from theorem 2.2.1, but then by independence and from theorem 2.2.2

$$\nu_i 1_{C_n^i} E[1_{A_n}|C](\omega) = \nu_i 1_{C_n^i} E[1_{A_n}] = \nu_i 1_{C_n^i} \cdot t_n$$

Thus $E[X_i|C](\omega) = \nu_i t_n 1_{C_n^i}$, $i \varepsilon I_n$, $n = 1,2\ldots$.

We now proceed to show that $E[X_i|C](\omega) \geq 1$ for infinitely many i with probability one. To achieve this, it suffices to show that the events C_n^i occurs for infinitely many n with probability one, since if the events C_n^i occurs for $i \varepsilon I_n$, then $E[X_i|C](\omega) = \nu_i t_n \geq 1$. To show that the

C_n^i occurs for infinitely many n with probability one, we shall make a use of the famous lemma called Borel-Cantelli lemma theorem 1.3.1.

Consider the events $\bigcup_{i\varepsilon I_n} C_n^i$, $n = 1,2\ldots$.

We now claim that they are independent and have the measure $\frac{r_n}{t_n}$.

It is easy to see that $\bigcup_{i\varepsilon I_n} C_n^i$ has the measure $\frac{r_n}{t_n}$, since

$\mu(C_n^1) = \frac{2^{-(k_0+n)}}{t_n} = 1 - \frac{r_n}{t_n}$ and the complement of C_n^1 is $\bigcup_{i\varepsilon I_n} C_n^i$.

For the independence of C_n^i , $n = 1,2\ldots$, consider the following events

C_n^1 , C_m^1 , $n > m$ with $\frac{2^{-(k_0+n)}}{t_n}$, $\frac{2^{-(k_0+m)}}{t_m}$ respectively. Then

$$\mu(C_n^1 \cap C_m^1)$$

$$= \sum_{k=0}^{\infty} \mu[(\bigcup_{(i_0,\ldots,i_{n-1})\varepsilon J_{n-1}} A_k^{i_0,\ldots,i_{n-1},1}) \cap (\bigcup_{(i_0,\ldots,i_{m-1})\varepsilon J_{m-1}} A_k^{i_0,\ldots,i_{m-1},1})]$$

$$= \sum_{k=0}^{\infty} \sum_{i_0\varepsilon\{1\}UI_0} \cdots \sum_{i_{n-1}\varepsilon\{1\}UI_{n-1}} \mu(A_k^{i_0,\ldots,i_{m-1},1,i_{m+1},\ldots i_{n-1},1})$$

$$= \sum_{k=0}^{\infty} \sum_{i_0\varepsilon\{1\}UI_0} \ldots \sum_{i_{n-1}\varepsilon\{1\}UI_{n-1}} \frac{2^{-(k_0+n)}}{t_n} \mu(A_k^{i_0,\ldots i_{m-1},1,i_{m+1},\ldots i_{n-1}})$$

$$= \sum_{k=0}^{\infty} \sum_{i_0\varepsilon\{1\}UI_0} \ldots \sum_{i_{m-1}\varepsilon\{1\}UI_{m-1}} \mu(A_k^{i_0,\ldots,i_{m-1},1}) \cdot \frac{2^{-(k_0+n)}}{t_n}$$

$$= \sum_{k=0}^{\infty} \sum_{i_0\varepsilon\{1\}UI_0} \cdots \sum_{i_{m-1}\varepsilon\{1\}UI_{m-1}} \frac{2^{-(k_0+n)}}{t_n} \cdot \frac{2^{-(k_0+m)}}{t_m} \mu(A_k^{i_0,\ldots,i_{m-1}})$$

$$= \sum_{k=0}^{\infty} \frac{2^{-(k_0+n)}}{t_n} \cdot \frac{2^{-(k_0+m)}}{t_m} \mu(A_k)$$

$$= \frac{2^{-(k_0+n)}}{t_n} \cdot \frac{2^{-(k_0+m)}}{t_m} .$$

Since this can be extended to any finite number of events, we have the independence of C_n^1 , $n = 1,2\ldots$ which implies the independence of $\bigcup_{i\in I_n} C_n^i$, $n = 1,2\ldots$. The independence of $\bigcup_{i\in I_n} C_n^i$ follows from the fact that if an event A is independent of an event B then A^c is independent of B^c .

To satisfy the hypothesis of the Borel-Cantelli lemma, it only remains to show that $\sum_{n=1}^{\infty} \frac{r_n}{t_n} = \infty$.

We shall consider the following two cases.

Case i) If $r_n \geq 2^{-(n+k_0)}$ for infinitely many n , then $\frac{r_n}{t_n} \geq \frac{1}{2}$ for infinitely many n , since $r_n \geq 2^{-(n+k_0)}$, $t_n \leq 2^{-(n+k_0-1)}$ which imply

$$\frac{r_n}{t_n} \geq \frac{2^{-(n+k_0)}}{2^{-(n+k_0-1)}} = \frac{1}{2} .$$

Consequently, we have

$$\sum_{n=1}^{\infty} \frac{r_n}{t_n} = \infty$$

if $r_n \geq 2^{-(n+k_0)}$ for infinitely many n .

Case ii) If $r_n < 2^{-(n+k_0)}$ for sufficiently large n , say $n \geq n_0$, then

$$\sum_{n} \frac{r_n}{t_n} \geq \sum_{n \geq n_0} \sum_{i \in I_n}{}' \frac{p_i \nu_i}{4} = \sum_{i \in T} \frac{p_i \nu_i}{4} ,$$

where $T = \{i \geq 2: \ \nu_i \geq 2^{n_0+k_0}\}$.

The first inequality follows from the fact that if $r_n < 2^{-(n+k_0)}$ for $n \geq n_0$, then $t_n < 2^{-(n-1+k_0)}$ for $n \geq n_0$,

$$\frac{r_n}{t_n} \geq r_n \cdot 2^{n-1+k_0} = \frac{r_n}{4} 2^{n+1+k_0} \geq \frac{r_n}{4} \nu_i$$

and $\sum_{i \in I_n} p_i = r_n$. From $\sum_{i=1}^{\infty} p_i \nu_i = \infty$ and $\sum_{i \notin T} p_i \nu_i \leq 2^{n_0+k_0} \Sigma p_i = 2^{n_0+k_0}$ we have

$$\sum_{i \in T} p_i \nu_i = \infty .$$

Therefore,

$$\sum_{n=1}^{\infty} \frac{r_n}{t_n} = \infty .$$

For both cases, we have

$$\sum_{n=1}^{\infty} \frac{r_n}{t_n} = \infty$$

and these two cases exhaust all the possible cases. Then from the Borel-Cantelli lemma

$$\mu(\overline{\lim_n} \bigcup_{i \in I_n} C_n^i) = 1 .$$

More intuitively, let $\{A_n\ , \text{i.o.}\} = \{\omega\colon\ \omega \in A_n$ for an infinitely many $n\}$ then $\mu(\{\bigcup_{i \in I_n} C_n^i \text{ i.o. }\}) = 1$. Then since $\omega \in \bigcup_{i \in I_n} C_n^i \Rightarrow \omega \in C_n^i$ for some $i \in I_n$, we have for some $i \in I_n$.

C_n^i occurs for an infinitely many n with probability one.

Therefore,

$$E[X_i|C](\omega) = \nu_i t_n 1_{C_n^i} \geq 1$$

for infinitely many times with probability one. Thus, we have

$$\mu(\{\omega\colon\ E[X_n|C](\omega) \not\to 0\}) = 1 \ .$$

This completes our investigation of establishing a condition on a p.s. (Ω,F,μ) under which $\sup_n X_n \in L_1$ is a necessary condition for the GLDCT.

We note that if (Ω,F,μ) is not an atomless probability space, then by theorem 3.2.1 we can consider the convergence of a sequence of r.v. on atomic and non atomic parts of the space, separately.

BIBLIOGRAPHY

[1] Birkhoff, G. Averaging operators, Symposium in Lattice Theory, A.M.S., 1960.

[2] Blackwell, D. and Dubins, L.E. A Converse to the Dominated Convergence Theorem, Illinois J. Math., Vol. 7 (1963) pp. 508-514.

[3] Brieman, L. Probability Theory, Addison Wesley Co. Inc., 1970.

[4] Doob, J.L. Stochastic Processes, New York: J. Wiley and Sons; London: Chapman and Hall, 1953.

[5] Douglas, R.G. Contractive Projections on a L_1 Space, Pacific J. Math., Vol. 15 (1965) pp. 443-462.

[6] Hahn, H. and Rosenthal, A. Set Function, The University of New Mexico Press.

[7] Halmos, P.R. Measure Theory, Princeton, N.J.: D. Van Nostrand Co., 1950.

[8] Hewitt, E. and Stromberg, K. Real and Abstract Analysis, Springer-Verlag, New York, Inc., 1965.

[9] Loéve, M. Probability Theory, 3rd ed., Princeton, J.J.: D. Van Nostrand Co., 1963.

[10] Moy, S.C. Characterization of Conditional Expectation as a Transformation on Function Spaces, Pacific J. Math., Vol. 4 (1954) pp. 47-63.

[11] Olson, M. A Characterization of Conditional Probability, Pacific J. Math., Vol. 15 (1965) pp. 971-983.

[12] Rényi, A. Foundations of Probability, Holden-Day, Inc., San Francisco, 1970.

[13] Rota, G.C. On the Representing of Averaging Operator, Rend. Padova, Vol. 30 (1960) pp. 52-64.

[14] Royden, H.L. Real Analysis, New York: The Macmillan Co.; London: Collier - Macmillan Ltd., 1963.

[15] Šidák, Z. On Relations Between Strict-Sense and Wide-Sense Conditional Expectation, Teor. Verojatnost. i Primenen; Vol. 2 (1957) pp. 283-288.