12

US ARMY MATERIEL COMMAND

AD

TECHNICAL REPORT BRL-TR-2769

AD-A175 192

SURVEY AND INTRODUCTION TO MODERN MARTINGALE THEORY (STOPPING TIMES, SEMI-MARTINGALES AND STOCHASTIC INTEGRATION)

Gerald R. Andersen

October 1986

US ARMY BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MARYLAND

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

AD-A175 192

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704-0188 Exp Date Jun 30 1986

1a REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT: Approved for Public Release, Distribution Unlimited.
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION: US Army Ballistic Research Laboratory; 6b OFFICE SYMBOL (If applicable): SLCBR-SE	7a NAME OF MONITORING ORGANIZATION
6c ADDRESS (City, State, and ZIP Code): Aberdeen Proving Ground, MD 21005-5066	7b ADDRESS (City, State, and ZIP Code)
8a NAME OF FUNDING/SPONSORING ORGANIZATION; 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61101A & 61102A	A91A AH43	-	DA306438 DA306672

11 TITLE (Include Security Classification)
SURVEY AND INTRODUCTION TO MODERN MARTINGALE THEORY
(Stopping times, Semi-martingales and Stochastic Integration)

12 PERSONAL AUTHOR(S)
ANDERSEN, GERALD R.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Technical Report	FROM 02/85 TO 05/86		

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Martingale, Semi-martingale, Martingale Transform, Stopping Time, Stochastic Integral, Stochastic Process, Dual Previsible Projection, Random Measure, Local Martingale.
12	01		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This document is primarily about the work of the Strasbourg group lead by P.A. Meyer and C. Delacherie and the development of the modern theory of semi-martingales and stochastic integration. These developments have occurred over the last two decades and extend the theories of J. Doob and K. Ito. Unlike the sources, the first chapter introduces the subject in terms of a stochastic calculus for discrete parameter processes. In particular, discrete parameter point processes are defined and the discrete form of stochastic calculus is applied to them to obtain the nonlinear filtering formulas. There are two reasons for including the material in Chapter 1: First, discrete parameter point processes provide a useful tool for modeling a wide variety of dynamical systems; second, this material provides an elementary introduction to the remaining chapters. Chapter 2 introduces the notion of a filtered probability space stopping times relative to a filtration, stochastic intervals and defines previsible, optional and progressive stochastic processes. Stochastic point processes are considered in Chapter 3 along with Lebesgue-Stieltjes stochastic integrals, the simplest of the stochastic

(continued on reverse side)

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL: GERALD R. ANDERSEN	22b TELEPHONE (Include Area Code): 278-6657; 22c OFFICE SYMBOL: SLCBR-SE-C

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

Block 19. ABSTRACT - continued

integrals. Chapter 4 contains one of the most basic concepts of modern martingale theory, the dual previsible projection and local martingales are defined in Chapter 5. The construction of the stochastic integral of bounded previsible processes relative to semimartingales is given in the final Chapter along with some applications to Brownian motion.

PREFACE

1. Introduction

1.1. **Background**: The theory of semi-martingales is a major part of the general theory of stochastic processes. This theory has undergone massive growth during the last two decades. Much of the impetus for the rapid advances in this branch of pure mathematics comes from efforts to solve applied problems. For example, the theory of stochastic integration relative to semi-martingales is the right tool for the analysis of stochastic dynamical systems and so for a large class of studies carried out by theoretical physicists, electronic engineers, system and control theorists, probabilists and statisticians. A semi-martingale is in fact a general model of the engineer's "signal plus noise" and the statistician's "trend plus random fluctuations".

1.2. **History**: Following the work of Paul Lévy and Joseph Doob, the epoch making works in the general theory of stochastic processes are due to Paul Andre Meyer [1967], all his papers in the 18 or so Strasbourg seminars in probability (especially, No. 10), Kunita and Watanabe [1967], Meyer [1973], Dellacherie [1972], Dellacherie and Meyer [1975, 1980] and Jacod [1979]. For anyone interested in reading into the last two decade's progress in the theory of semi-martingales, however, it must be understood that the principal original source is the collection of Strasbourg Séminaires. These seminars are published by Springer-Verlag in the Lecture Notes in Mathematics Series. The Université de Strasbourg Séminaire de Probabilités not only contain the modern theory of semi-martingales, but also retain the false starts, the subsequent alterations to the "correct" directions, the seemingly interesting and possibly uninteresting concepts and the "ripening" of proofs and techniques that are characteristic of any developing mathematical theory. For example, see the Université de Strasbourg Séminaire de Probabilités in 1967, 1970, 1975, and 1980 for successive accounts of stochastic integration; the first three were given by Meyer. It is a rare thing to be able to observe the evolution of a new theory and to see it mature in such a short period of time. Most of the credit for this rapid development probably belongs to a group of predominately French mathematicians led by Paul-Andre Meyer and loosely referred to as the "Strasbourg School".

Métivier [1982] gives a slightly different emphasis to the subject than the works of those previously mentioned. He starts his work with quasi-martingales and bases the entire subject from martingales to the stochastic integration of semi-martingales on the so-called Doleans measure. This is a very elegant development. The use of the Doleans measure, to some degree, brings stochastic integration within the domain of classical measure theory, a fact that will please a large number of mathematicians. At the same time, Métivier's approach retains the stopping time flavor of the Strasbourg School. The additional distinctive feature of this excellent work is that most results are formulated for Banach valued processes, thus providing a theory applicable to multi-dimensional processes.

An additional book, in the spirit of Métivier, has recently been published. Kai Li Chung [1983], together with Ruth Williams, has written a clear and concise work on stochastic integration. Since it is anchored in all of Chung's other works and those of J. Doob it is worth reading. The only shortcoming from the standpoint of this note is that it only considers martingales with continuous paths. Gopinath Kallianpur's 1980 work on stochastic filtering theory also skips the point process case. But it is worth reading, if only to appreciate the maturity of the continuous parameter filtering problem and the clarity of Kallianpur's style.

The principal study of point processes from the standpoint of martingales is 'Point Processes and Queues' by P. Brémaud, 1981. This is an excellent treatise on the theory of martingales applied to queuing and the filtering problem for point processes. Brémaud develops his theory from first principles, relying on Dellacherie's Dual Previsible Projection Theorem rather than the Doob-Meyer Decomposition theorem and the extensive recent developments in stochastic integration relative to semi-martingales. It is the best introduction to the subject from the standpoint of applications and much of what will follow in this note concerning filtering is borrowed, in one way or another, from the ground-breaking work of Brémaud since 1972.

Outside of some examples illustrating the methodology, a few simple results in Chapter 1 and a personal viewpoint, all of the mathematics in this note is known. The opening Chapter introduces a discrete parameter version of the martingale calculus that will be introduced in the remaining five chapters. The purpose of this Chapter, and its threads into the later sections where the continuous parameter model is studied, is to provide some intuition and background for the study of these technically difficult subjects. Starting from first principles, many of the hard to reach concepts of the continuous time model are almost trivial in the discrete model; certainly, the proofs and technical details are elementary. The case of discrete parameter point processes are of particular interest (Section 1.10). One can only wonder why this material is not written down somewhere. In most instances results about such processes follow from the general theory in a relatively straight forward manner (e.g., Section 4.7 of Chapter 4), but that does not replace the insight obtained from deriving these results directly. Moreover, having to go to the general theory of marked point processes in order to solve an applied problem involving discrete parameter point processes seems a bit excessive and would certainly inhibit applications of the basic concepts of the theory.

1.2.1. **Contents**: The six chapters contain foundation material on stopping times, filtrations, various types of function measurability, martingales, and a brief description of integration relative to martingales. These chapters are meant to constitute a brief survey and introduction to this material. Therefore, proofs are given only when they pass loose criteria based on brevity, insight and simplicity. Chapter 1 contains a brief introduction to nonlinear filtering.

There are two excellent surveys on martingales and stochastic integration. One, by C. Dellacherie [1978], concentrates on stochastic integration. The other, due to A.N. Shiryayev, is very broad. Both of these papers are true surveys in that they tell what has been accomplished in these areas and appropriately assume that the reader has some understanding of the area, especially probability theory and stochastic processes and is an active mathematician. The present note, on the other hand, is meant to be both a survey of recent developments in this area and an introduction to the basic theory. As such, definitions observe the mathematical traditions of such things, examples and counter examples are supplied to aid in the understanding of new objects defined, and Theorems, Corollaries and Lemmas are rigorously stated. But complete proofs, sketchs or indications of proofs are given only when they are relatively easy and informative, or when they illustrate the meaning of newly defined concepts. Chapter 6 is the chapter with the most proofs simply because it is impossible to have any kind of understanding of the stochastic integral without them. One of the reasons this is true is that most readers of this note will have a strong intuition built on classical theories of integration and this knowledge, combined with the fact that notationally most integrals look alike and have similar properties, will mislead rather than support their understanding of the stochastic integral.

1.2.2. **Purpose**: The primary purpose of the note is twofold: (i) To summarize a recently evolved theory and indicate how it might be applied to some BRL tasks; (ii) To form a foundational document, a common ground for an interdisciplinary group within the CSM branch of BRL-SECAD, all of whom are concerned with various mathematical aspects of stochastic network problems in Army Communication, Command and Control.

ACKNOWLEDGEMENTS

I would like to thank the Ballistic Research Laboratory for the opportunity to do this work and, in particular, Steve Wolff (BRL) for his guidance and encouragement. I also extend sincere thanks to many colleagues at BRL for numerous stimulating mathematical discussions and for reviewing this document, especially Richard Kaste, A. Brinton Cooper and Walter Egerland.

TABLE OF CONTENTS

Chapter 1: A Discrete Time Model of Martingale Calculus

Chapter 2: Continuous Parameter Stochastic Processes

Chapter 3: Increasing Processes

TABLE OF CONTENTS

SURVEY AND INTRODUCTION TO STOPPING TIMES, MARTINGALES AND STOCHASTIC INTEGRATION

Chapter 1. **A Discrete Time Model Of Martingale Calculus**

1.1. **Introduction**: This first section is meant to be a discrete time model for most of the of the topics in this note. The initial purpose of this section, though, was only to guide the reader (and writer) through the intricacies of stochastic integration by first studying martingale transforms (stochastic integrals for discrete time processes). This led to introducing the quadratic variation and variance processes, and the original Doob decomposition of submartingales. Before long it was clear that most of the subsequent topics would be much easier to discuss, in the sometimes sketchy manner appropriate to a survey and introduction, if one could lean on an intuition built on the sequences of random variables. Thus the present form of this section became an attempt to provide such an intuition or background before launching off into the much more sophisticated concepts required by processes indexed by a continuum.

This Chapter is not meant to be a summary of the theory of martingale sequences. This subject is huge. For an almost flawless treatment of this theory one would surely read Neveu's book [1975] or Meyer's [1973] Springer-Verlag monograph. For a treatment of martingale sequences that has a large number of examples and gives a very readable account of the theory, one should see Karlin and Taylor (1975). Rather, this Chapter is an attempt to give a brief description of a "discrete time martingale calculus", applicable to the study of discrete (stochastic) dynamical systems (Section 1.10).

It may also prove useful to see how a few of the concepts introduced here must be modified when "time" becomes non-denumerable, most notably the concept of "previsibility". It took many years for the role of such processes to be understood. In Meyer's 1967 book, he talks about "natural" processes instead of previsible ones. The connection between the two was made in an elegant paper by K.M. Rao (1969), but again only the Strasbourg Seminars (Meyer (1970)) show how the importance of the concept emerged. By the time one reads Dellacherie and Meyer (1980), previsible processes are referred to as the "Borel" functions of the general theory of stochastic processes.

1.2. **Filtrations and Stopping Times**: Let Z be the set of non-negative integers and let (Ω, H, P) denote a **probability space**, where H is a σ-algebra of subsets of Ω and P is a probability measure on H. A sequence, $G = (G_n, n \in Z)$, of sub σ-algebras of H is called a **filtration**, if (relative to set inclusion) the G_n are nondecreasing functions of n. We will assume that G_0 is complete, in the sense that it contains all subsets of events (i.e., members of H) which are assigned probability zero by P. G_∞ will denote the smallest σ-algebra containing all the G_n: $G_\infty = \sigma(\bigcup_{k \geq 0} G_k)$. ($G_\infty$ is a sub-σ-algebra of H.)

Perhaps the single most important concept in martingale theory is the notion of stopping time. This is more evident in the continuous case than here. But even here where all we are trying to do is lay a foundation of sorts for things to come, this notion plays a fundamental role. Stopping times are defined relative to filtrations, so to motivate the definition and at the same time give a concrete example of a filtration, we first consider the following

Example: Let $X = (X_n, n = 0,1,2,...)$ be a sequence of random variables executing a symmetric random walk on the real line, starting from the origin (hence, $X_0 = 0$). For definiteness, suppose that X_n represents the value of a game at its n^{th} trial and $X_n = X_{n-1} + 1$, and $X_n = X_{n-1} - 1$, *each with* probability $\frac{1}{2}$. Let $G_0 = \{ \phi, \Omega \}$ and G_n denote the smallest σ-algebra generated by the X_k, $0 \leq k \leq n$: $G_n := \sigma(X_k, 0 \leq k \leq n)$. G_1 is the family consisting of the empty set and unions of the partition $\{ w : X_1(w) = 1 \}$, $\{ w : X_1(w) = -1 \}$; G_2 is the family consisting of the empty set and unions of the partition

$$\{ w : X_1(w) = -1, \text{ and } X_2(w) = -2 \},$$

$$\{ w : X_1(w) = -1, \text{ and } X_2(w) = 0 \},$$

$$\{ w : X_1(w) = 1, \text{ and } X_2(w) = 0 \},$$

$$\{ w : X_1(w) = 1, \text{ and } X_2(w) = 2 \}.$$

Notice that the union of the first two of these events and then the union of the second two give the events that make up the partition defining G_1. Thus, we have that $G_0 \subset G_1 \subset G_2$. The remaining G_k are defined in a similar fashion and monotonicity continues to hold. $G = (G_n)$ is therefore a filtration. This is an example of a special type of filtration called variously the **natural filtration** or the **internal history** of the processes X, or the **filtration generated by X**.

Now, for each $w \epsilon \Omega$, let $T(w) := \min\{ k : |X_k(w)| = 2 \}$, if {...} is not empty and $T(w) := \infty$, if $\{ \cdots \} = \phi$. T is the first time that the process, X, takes on the value plus or minus 2. T is a mapping of Ω onto the extended, nonnegative, real line, $\overline{R}_+ := [0,\infty]$, with the property that the event $[T \leq n] := \{w : T(w) \leq n\}$ is a member of G_n. To see this it is enough to look at a couple of cases; the formal induction will be clear. Explicitly, $[T \leq 0] = [T \leq 1] = \phi$, so these two events are contained in G_0 and G_1. Since $[T \leq 2] = [X_1 = -1, X_2 = -2] \bigcup [X_1 = 1, X_2 = 2]$, we have $[T \leq 2] \epsilon G_2$. Viewing the family of events, G_n , as the history of the process up to "time" n, this means that the value of T at time n depends only on history of the process, X, up to and including time n. In this sense, the extended valued random variable T is said to be a 'stopping time' relative to the filtration (history) G. Contrast this with the variable, S, defined by setting $S(w) := \max\{k : 1 \leq k \leq 5, |X_k(w)| = 2\}$, if such a k exists and ∞ otherwise. Clearly, the values of S depend on the entire history of the paths, $n \rightarrow X_n(w)$, of the process X. Therefore, S is not a stopping time relative to the filtration G, according to the following

1.2.1. **Definition**: A mapping T from Ω to $\overline{Z} := Z \bigcup \{\infty\}$ is said to be a **G-stopping time (optional time)** *if*

$$\{ w \mid T(w) = n \} \epsilon G_n$$

for all n in Z. When T is a G-stopping time, the σ-algebra, G_T, of **events that occur prior to T**, is defined by setting

$$G_T = \{ B \epsilon G_\infty \mid B \bigcap [T = n] \epsilon G_n, \text{ for all n in Z } \}.$$

By definition, $[T=n]$ is in G_∞ for all n in Z, so that $[T = \infty]$, the complement of all these events, is also an event in G_∞. Consequently, the mapping $T: \Omega \rightarrow \overline{Z}$ is G_∞-measurable. Hence, T is a random variable on (Ω, G_∞) and so on (Ω, H).

Finally, it is immediate that **T is a G-stopping time** iff $[T \leq n] \epsilon G_n$ for all n in Z. Just notice that

$$[T \leq n] = (\bigcup_{k=0}^{n} [T = k]) \epsilon G_n,$$

for all n, since $[T = k] \epsilon G_k$ is contained in G_n, for all $k \leq n$. Conversely, $[T = n] = [T \leq n] - [T \leq n-1]$ is in G_n. More trivial, but of some interest for later comparison to the situation when the stopping times are R_+ valued is that $[T \leq n] \epsilon G_n$ iff $[T < n] \epsilon G_{n-1}$

We will return to the topic of stopping times in general after a few more definitions. In Chapter 2, where the model is more complex, we will give several more examples.

1.3. **Stochastic Processes, Previsibility and Optionality**: Let Z be the set of non-negative integers. A sequence, $X = (X_n, n \epsilon Z)$, of mappings of Ω into the set of real numbers is called a real valued **stochastic process** if for each n, the mapping, $w \rightarrow X_n(w)$, of Ω into R is H-measurable. That is, for each n in Z, $\{w \epsilon \Omega : X_n(w) \epsilon B\} \epsilon H$, for all real Borel sets, B. Of course, this is just the statement that for each n, X_n is a real valued, random variable on the measurable space (Ω, H).

Further, X is said to be **G-adapted**, if X_n is G_n-measurable for each n in Z. If X is adapted to G, then we also say X is **observable** relative to those processes which generate G. It is useful to realize that if X is G-adapted, then measurable functions of successive finite segments, $g(X_1, \cdots, X_n)$, *of X define G-adapted* processes.

Convention: Throughout this Chapter, whenever processes are discussed it will always be assumed that they are adapted relative to the same fixed filtration, unless stated otherwise. This is no restriction in generality since we have not excluded the **trivial filtration**, $(G_n, n \epsilon Z)$, where $G_n = H$ for all n.

For the discrete time processes, the important notion of "previsibility" takes on a very simple and intuitive meaning: $V = (V_n, n \text{ in } Z)$ is said to be **G-previsible**, if each random variable V_n is G_{n-1} measurable. This description of previsibility is intuitive since if a process, (V_n), is G-previsible, when $G_n = \sigma(X_k, k=0,1,...,n)$, for some process, X, then V_n is a Borel function of X_k, for $k=0,1,...,n-1$. Thus, the value of the process V at time n is completely determined by the value of X at the times 0,1,2,...,n-1. That is, just before time n (prior to n) the value of V_n is known; it is previsible. "Previsible" is the French term; in English it is usually translated to "predictable". We use the former term because the notion of predictability as a technical term carries too many possible meanings (e.g., in wide sense stationary time series analysis) and the English interpretation of the term "previsible", viz., "being visible before", rather precisely describes the intended technical meaning.

Later in this chapter we will need a reasonably precise understanding of the statement that a (discrete parameter) process, X, is "evaluated at a stopping time", $X_{T(w)}(w)$. An immediate difficulty that one might notice is that stopping times take values in $\overline{Z}_+$, while for any w in Ω, $n \rightarrow X_n(w)$ is defined only on Z_+. This can be overcome by setting, for example, the value of the process at "infinity" equal to zero, for all w in Ω. This is equivalent to writing $X_{T(w)}(w) 1_{[T<\infty]}$ in place of $X_{T(w)}(w)$. As it is convenient, we will use one or the other, or just qualify appropriate statements by saying " on $[T < \infty]$", while writing either $X_{T(w)}(w)$ or X_T. In any case, we then need to know what conditions must be imposed on X so that X_T is a random variable. Hence, we must first say what is meant for a random variable to be defined on a subset of Ω. So let Ω_o be a subset of Ω of the probability space, (Ω, H, P). The **trace** σ-algebra, denoted $H \bigcap \Omega_o$, is the family $\{ A \bigcap \Omega_o : A \epsilon H \}$ of subsets of Ω_o. Of course, Ω_o may not belong to H, but if it does, then the trace σ-algebra is just $\{ A : A \epsilon H,$ A a subset of $\Omega_o \}$. Now we are all set: X is a real valued **random variable** on a subset, Ω_o, of Ω (on $(\Omega_o, H \bigcap \Omega_o, P)$) iff $X^{-1}(B) \epsilon H \bigcap \Omega_o$ for all real Borel sets, B. Further, we can talk about a **G-measurable random variable or function** defined on a subset, Ω_o, where G is a sub σ-algebra of H, by replacing H by G in the definition above. Then the following result holds (Neveu [1975]).

1.3.1. **Lemma**: *If X is a G-adapted process and T is a G-stopping time, then the random variable* X_T, *defined on* $\{w: T(w) < \infty\}$ *by setting* $X_T(w) := X_{T(w)}(w)$ *is* G_T*-measurable.*

The random variable of this definition-theorem is obtained as a result of **evaluating the process at the stopping time T**. Perhaps the most important example is that of a stopped process. If T is a stopping time, then $T_n(w) := (T \wedge n)(w) := T(w) \wedge n$, the minimum of numbers T(w) and n, defines a stopping time for each $n \epsilon Z_+$. Let $X_n^T(w) := X_{T_n}(w)$. We define the process X **stopped at time T** by setting $X^T = (X_n^T, n \epsilon Z_+)$.

The paths, $n \rightarrow X^T(w)$, of the stopped process are constant to the right of the interval [0, T(w)]. Stopped processes are fundamental to modern martingale theory. Later in these notes, the notion of path-wise "localization" of a process is introduced, whereby properties such as "boundedness" are attributed to the process "locally" in the sense that the stopped process is bounded. For example, from elementary calculus, a path-wise continuous process is locally bounded. This technique becomes a powerful tool for extending certain results to more and more general classes of processes and will be used extensively in Chapter 6.

Another detail that we need throughout is a way to say two processes are "equal". That is, we need an equivalence relation on a set of processes. Let X and Y be two discrete parameter processes defined on the same probability space (Ω, H, P). Since any countable collection of events of P-measure zero is again an event of P-measure zero, the statements that

$$P(X_n = Y_n) = 1, \ n\epsilon Z_+ \ \text{and} \ P(X_n = Y_n, \ n\epsilon Z_+) = 1$$

are equivalent. This is not true in the continuous parameter case considered in Chapter 2, where processes having the first property are called "modifications" of one another and those having the second are called **indistinguishable**. These concepts being equivalent for discrete parameter processes, we will only use the latter for now. Clearly, indistinguishability determines an equivalence relation on the set of all processes defined on (Ω, H, P). So any processes or random quantities which are discussed in this chapter are only specified to within membership in a particular equivalence class. On occasion we will emphasize this point by writing "a.s.P", meaning "almost surely relative to the probability P", or, "with probability one" next to equalities and inequalities involving random quantities.

With an eye toward later chapters, we also remark at this point that a process which is indistinguishable from the process which is identically zero is said to be **evanescent**. Subsets of $Z_+ \times \Omega$, called **random sets**, are said to be **evanescent** if their indicator functions are evanescent.

1.4. **Transforms of Stochastic Processes**: Let $V = (V_n, n\epsilon Z)$ and $X = (X_n, n\epsilon Z)$ be two processes. Extend the time domain of processes on $Z \times \Omega$ by setting $X_{-1} = 0$ for all w in Ω. Set $\Delta X_k = X_k - X_{k-1}$, then in particular, $\Delta X_0 = X_0$

Given two processes X and V, define the process V.X on Ω by setting

$$(V.X)_n(w) := \sum_0^n V_k(w) \ \Delta X_k(w) \tag{1}$$

for all n in Z and each $w\epsilon\Omega$. V.X is called the **transform** of X by V. When we want to use the transform of X by V to anticipate results about stochastic integrals, we will sometimes call this transform a **discrete integral of V with**

respect to X. In this case we have in mind that $V_k = v_{t_k}$ and $X_k = x_{t_k}$, $k=1,2,\ldots,n$, where $0=t_0<t_1< \cdots <t_n=t$, for some continuous parameter processes v and x.

Equation (1) is also written in the forms

$$(V.X)_n(w) = V_0(w)\, X_0(w) + \sum_1^n V_k(w)\ \Delta X_k(w) \tag{2}$$

$$= (V.X)_{n-1}(w) + V_n(w)\, X_n(w). \tag{2.1}$$

As a discrete integral it is clear that the transform in (1) is nothing more than a particular form of a Darboux sum associated with a Riemann Stieltjes integral. As such, in later chapters of this note, it will become the major building block of stochastic integrals relative to various types of (continuous time) processes.

1.5. **The Quadratic Variation and Variance Processes:** We now introduce two more processes that play an important role in stochastic integration. These processes also form a link back to classical probability and statistics.

Again let $X = (X_n, n \epsilon Z)$ be any process and define the stochastic process, $[X,X]$, on Ω by setting

$$[X,X]_n(w) := X_0^2(w) + \sum_1^n (X_k(w) - X_{k-1}(w))^2 = \sum_0^n (\Delta X_k(w))^2, \tag{3}$$

for all n in Z and each $w \epsilon \Omega$. (Recall that $X_{-1} := 0$.) The increasing process $[X,X]$ is called the **quadratic variation** of X. Some writers ingeniously call it **square brackets** X.

If Y is any other process parameterized by Z, we define the **cross quadratic variation**, $[X,Y]$, by polarization

$$[X,Y] := \frac{1}{2}\left([X+Y, X+Y] - [X,X] - [Y,Y] \right). \tag{4}$$

By elementary manipulations, this definition is equivalent to setting

$$[X,Y]_n := \sum_0^n \Delta X_k \ \Delta Y_k . \tag{5}$$

Now, we assume that $E(X_n^2) < \infty$ for each n in Z; that is, $X \epsilon L_2(P)$. Let $G = (G_n, n \epsilon Z_+)$ be the underlying filtration for the processes in this section and define $G_{-1} = G_0$. Then set

$$<X,X>_n := \sum_0^n E\{(\Delta X_k)^2 \mid G_{k-1}\} \tag{6}$$

for each $n \geq 0$. For now, it is appropriate to call <X,X> the **variance process**. Clearly, both <X,X> and [X,X] are increasing processes. It is important to note, however, that [X,X] always exists, but <X,X>, as it has been defined, exists only when X has finite second moments. Finally, note that <X,X> is a G-previsible process, whereas [X,X] is only G-adapted, if X is G-adapted.

The **covariance process**, <X,Y> is defined by polarization, as in the case of the quadratic variation, and leads to a formula analogous to equation (5).

1.5.1. With the notational agreements made at the beginning of the section, notice that we can write any process in the form

$$X_n = \sum_0^n \Delta X_k = \sum_0^n d_k, \tag{7}$$

where the sequence $d_k := \Delta X_k$ is called the **difference process** associated with X.

1.5.2. **Example**: Assume, for this paragraph, that the d_k are independent of G_{k-1}, with d_0 independent of G_0, and have zero mean value and finite variance, σ_k^2. Then $E\{d_k^2 \mid G_{n-1}\} = Ed_k^2 = \sigma_k^2$, so that $<X,X>_n = \sum_0^n \sigma_k^2$. That is, <X,X> is the variance of the process X. Thus, if X_n is a sum of zero mean random variables which are independent of the "past" and have finite variance, then $<X,X>_n$ reduces to an increasing, deterministic process which is equal to the variance of X_n. For example, if X is the random walk of the previous example, then the (d_k) are independent, symmetric Bernoulli random variables.

Further, if Y is another process whose difference process has the same properties as those of X in this example, then it is easy to see that $<X,Y>_n$ is just the covariance of X and Y, at time n.

Of course, $<X,X>$ is in general not a deterministic process, but it is always an increasing stochastic process. As such, it is perhaps more honestly referred to as a "stochastic measure" in that its properties derive more from the fact that, on each path, its increments define a positive measure of the algebra of all subsets of Z.

1.6. Martingales:

1.6.1. **Definition**: Let $(\Omega, H, (G_n), P)$ be a filtered probability space. A **G-martingale** is a sequence $(M_n, n \epsilon Z)$ of random variables on Ω with the following properties:

(a) $M = (M_n(w))$ is adapted to G

(b) $E\{|M_n|\} < \infty$, for all $n \epsilon Z$

(c) $E\{M_n \mid G_{n-1}\} = M_{n-1}$, a.s.P,

for all n in Z.

By definition of conditional expectation, condition (c) is equivalent to requiring that for all $A \epsilon G_{n-1}$

$$(c') \qquad \int_A M_n \, dP = \int_A M_{n-1} \, dP.$$

If the equality in (c) or (c') is replaced by $\leq$, or $\geq$, then M is called a **G-supermartingale**, or a **G-submartingale**, respectively. When the underlying filtration, G, remains fixed in a particular discussion we will often drop the qualifier G and just write "martingale" or "supermartingale" or "submartingale".

It follows from the definition that a martingale satisfies

$$M_k = E(M_n \mid G_k),$$

for every pair, (k,n), of nonnegative integers with $k \leq n$, not just neighboring integers. Similar statements hold for supermartingales and submartingales. To see this in the case of supermartingales just use the fact that filtrations are increasing and conditional expectations are smoothing operators and proceed as follows: $M_n \geq E(M_{n+1} \mid G_n)$, so that if $k \leq n$

$$E(M_n \mid G_k) \geq E(E(M_{n+1} \mid G_n) \mid G_k) = E(M_{n+1} \mid G_k).$$

Thus, $(E(M_n \mid G_k), n \geq k)$ is a decreasing sequence. Hence, supermartingales decrease in conditional mean and so

$$M_k = E(M_k \mid G_k) \geq E(M_n \mid G_k).$$

Similarly, submartingales increase in conditional mean and martingales are constant in conditional mean with the same obviously being true in the case of the unconditional means.

1.6.2. Remark: There are some immediate results about martingales that are simple to verify and are used constantly. As usual, a single underlying filtration is assumed in each statement.

- o If M and N are martingales, then M + N is a martingale.

- o If ϕ is a real valued convex function defined on R_1, M is a martingale and $\phi(M_k)$ has finite expectation, then $(\phi(M_k))$ is a submartingale.

- o If M is a martingale which is square integrable relative to P, then $M^2 - [M,M]$ is a martingale. Also, $M^2 - \langle M,M \rangle$ is a martingale.

All but the second statement follows by straightforward computation using the definition of the quantities involved.

The second statement requires Jensen's inequality. This is based on a result about real valued convex functions which states that there exist affine maps, $\phi_n = a_n x + b_n$, such that $\phi = \sup \phi_n$. Using the monotonicity and linearity of the conditional expectation operators, we obtain

$$E(\phi(X) \mid G) \geq E(\phi_n(X) \mid G) = \phi_n(E(X \mid G)).$$

Jensen's inequality follows: $E(\phi(X) \mid G) \geq \phi(E(X \mid G))$. By replacing X and G by M_k and G_{k-1}, and using the fact that M is a G-martingale, we obtain the result. Our applications include the important case $\phi(x) = x^2$.

We have encountered some basic martingales earlier. Let $X = (X_n)$ be written as in equation (7), and give the sequences of differences (d_k) the assumptions in the paragraph following (7). Then X is a martingale, since

$$E\{ X_n \mid G_{n-1} \} = E\{ d_n \mid G_{n-1} \} + E\{ X_{n-1} \mid G_{n-1} \} \tag{8}$$

and

$$E\{ d_n \mid G_{n-1} \} = 0, \quad E\{ X_{n-1} \mid G_{n-1} \} = X_{n-1}, \text{ a.s.P.} \tag{9}$$

The first of these equations is due to the fact that we took the difference sequence to be independent of the past and have zero expectation (i.e., the difference sequence is centered at conditional expectations). The second is a result of the fact that X is adapted to G. Because then, X_{n-1} is G_{n-1}-measurable, and it is a property of conditional expectations that $E\{f|K\} = fE\{1|K\} = f$, a.s.P, whenever f is K-measurable. Putting equations (8) and (9) together verifies the claim that X is a martingale.

Finally, it should be clear that we didn't need the finite variance assumption on the difference sequence; this was only assumed in the original example because we wanted to give an example about the variance process. In fact from equation (8) it follows that if the d_k have finite expectations and are centered at expectations conditioned on G_{k-1}, then X is an F-martingale.

1.7. **Doob's Theorems**: Another example of a martingale does assume that the X_n has finite variance, but that is all. Then

$$[X, X] - \langle X,X \rangle \tag{10}$$

is a martingale relative to the filtration G. This follows directly from the explicit form for the quadratic variation and the variance process for exactly the reasons that our first example was a martingale. Although this is true in the continuous case also, it will follow from the Doob-Meyer decomposition and will constitute the definition of the process $\langle X,X \rangle$.

It will be beyond the scope of this note to even outline the proof of the continuous time Doob-Meyer decomposition. Therefore, we will give a proof of the decomposition theorem in the discrete case. This has the added advantage that it is simple to prove and its proof, along with the statement of the result, will allow us to introduce a number of concepts which are quite difficult in the continuous time analogue.

1.7.1. Lemma: (Uniqueness of the Doob-Meyer Decomposition)
If a process $X = (X_n,\ n \epsilon Z)$ can be written in the form $X = M + A$, where $M = (M_n)$ is a G-martingale and $A = (A_n)$ is a G-previsible process, with $A_0 = 0$, then the representation is unique (up to indistinguishability).

Proof: Suppose that two representations exist : $X = M + A = m + a$, where m and a have the same properties as M and A. Then M - m = A - a demands that M - m is a previsible martingale. This implies that $M_n - m_n = E\{ M_n - m_n \mid G_{n-1} \} = M_{n-1} - m_{n-1}$. Hence, $M_n - m_n = M_0 - m_0$. Finally, since the last quantity is equal to $A_0 - a_0 = 0$, $M_n = m_n$, a.s.P; hence, M - m is evanescent. This of course implies the same for A - a.

Notice that we have also proved the interesting and useful fact that **previsible martingales are constant a.s.P**. A similar statement is true in continuous time (Chapter 4), but requires an enormous amount of machinery to prove.

1.7.2. Theorem: (Doob Decomposition)
Let $X = (X_n)$ be an $L_1(P)$, G-adapted stochastic process. Then there exist processes M and A, where M is a martingale and A is previsible with $A_0 = 0$, such that $X = M + A$. This representation is unique (modulo indistinguishability).

Because of the previous Lemma the proof of this statement just consists in observing that we can write

$$X_n - X_{n-1} = X_n - E(X_n \mid G_{n-1}) + E(X_n - X_{n-1} \mid G_{n-1}). \tag{11}$$

It follows that X = M + A, where

$$M_n = X_0 + \sum_{k=1}^{n} (X_k - E(X_k \mid G_{k-1})) \tag{12}$$

and

$$A_n = \sum_{k=1}^{n} E(X_k - X_{k-1} \mid G_{k-1}), \qquad A_0 = 0. \tag{13}$$

Clearly, M is a martingale and A is previsible. Of course, all these equations hold with probability one only.

The process, A, of Doob's decomposition is called the "compensator" of the process, X, according to the following. Let X be a P-integrable process. Then the process, $\tilde{X}$, defined by setting $\Delta\tilde{X}_n = E(\Delta X_n \mid G_{n-1})$ for $n \geq 1$, $\tilde{X}_0 = 0$, is called the **compensator** of the process, X. If in addition to P-integrability, X is G-adapted, then $\tilde{X}$ is obviously characterized by the following three properties:

(a) $X - \tilde{X}$ is a G-martingale;

(b) $\tilde{X}$ is a G-previsible process;

(c) $\tilde{X}_0 = 0$.

Compensators will be examined in some detail in Chapter 4.

The following corollary is immediate and is of the form stated in the sequel, where the index set is a continuum:

1.7.3. **Corollary: (Doob-Meyer Decomposition Theorem)**
If X is a G-submartingale, then there exist processes M and A, where M is a G-martingale and A is an increasing, G-previsible process with $A_0 = 0$*, such that* $X = M + A$*, uniquely (modulo indistinguishability).*

The only part that now requires proof is the statement that A is an increasing process. Since X is a submartingale, this follows immediately from the definition of A in equation (13) written in the form $A_n = A_{n-1} + E(X_n|G_n) - X_{n-1} \geq A_{n-1}$, a.s.P. (When $A_0(w) = 0$ and $A_n(w) \geq A_{n-1}(w)$ for P almost all w in Ω and $n \geq 1$, the process A is said to be an **increasing process**.)

1.7.4. Remark: Immediately following the definition of martingales we pointed out that when M is an L_2 martingale (so that by Jensen's inequality, M^2 is a submartingale), both M^2 - [M,M] and M^2 - <M,M> are martingales. Since <M,M> is previsible, it follows from the uniqueness of the Doob-Meyer Decomposition that $M^2 = m + \langle M,M \rangle$ is the decomposition specified by the

Corollary, the Doob-Meyer decomposition of M^2. Thus, $<M, M>$ is a previsible process which "compensates" for M^2 not being a martingale, even though M is one. Indeed, the process $<M,M>$ is the **compensator** of M^2. This is because

$$E(\Delta(M_n^2) \mid G_{n-1}) = E((\Delta M_n)^2 \mid G_{n-1}) = <M,M>.$$

1.7.5. Another way of visualizing the DMD Theorem is to recall that on the average, submartingales rise. That is, $n \rightarrow EX_n$ is an increasing function on Z_+. The DMD Theorem says that A accounts for this proclivity to rise by previsibly compensating X to produce a martingale, X - A, which of course has constant expectation.

1.7.6. Remark: Processes of the form $X = M + A$, where M is a martingale and A is an increasing process, are special cases of a class of processes called **semi-martingales** in the sequel. When the decomposition is unique (to within distinguishability), then X is called a **special semi-martingale**. Hence, the Doob-Meyer Theorem states that submartingales are a particular form of special semi-martingale. This is a very convenient interpretation from the standpoint of applications since a semi-martingale is just a mathematical model for a dynamical system which consists of a "signal" or "trend" term, A, and a "noise" term, M.

It is easily seen that the Doob-Meyer Theorem also holds when X is a supermartingale. We need only write $X = M - A$ in order to maintain the property that A is an increasing processes. Again, on the average supermartingales fall and A previsibly compensates to produce $X + A$, which has constant averages.

1.7.7. Remark: Since engineers have been using the "signal plus noise" model for decades, it is probably worthwhile to take a moment to understand why they have been so successful (and to acknowledge the generality of their achievement). The DMD Theorem states that any discrete time process with finite mean that is observable relative to some filtration (**flow of information**, Wong [1973]) is a semi-martingale. In fact, if $X = (X_n, n \epsilon Z)$ is any finite mean process and (F_n) is any information flow, then the sequence (Y_n), where $Y_n = E(X_n \mid F_n)$, (i.e., what is observable about X relative to available information), can be shown to be a semi-martingale. It took mathematicians a while to understand all this and then to do what their discipline demands, namely, explain the reason why "signal plus noise" models were important, from a viewpoint other than "the model seems to work". This note is in some sense shows the lengths to which mathematicians have gone in the last 30 or so years to explain the full significance of semi-martingales (in continuous time), including their construction of a calculus to study these processes in their most general form and at the same time to provide scientists with the correct tools to model stochastic dynamical systems. Only time will tell whether or not the resulting mathematical theory is technically too difficult for applications.

1.7.8. We will now return to the initial reason for this chapter: to introduce and study transforms of martingales, called **martingale transforms**, in an effort to set an intuitive foundation for the development of stochastic integrals.

1.7.9. **Theorem**:
Let X be a martingale (supermartingale, submartingale). If V is a nonnegative, previsible process and the transform of X by V is P-integrable, then V.X is a martingale (supermartingale, submartingale).

The proof of this very important result is an immediate consequence of the second representation of a transform in equation (2):

$$\begin{aligned} E\{ (V.X)_n - (V.X)_{n-1} \mid G_{n-1} \} &= E\{ V_n (X_n - X_{n-1}) \mid G_{n-1} \} \\ &= V_n E(X_n - X_{n-1} \mid G_{n-1}). \end{aligned}$$

The right side of this equation $= 0$, ≤ 0 or ≥ 0, depending on whether X is a martingale, a supermartingale or a submartingale, respectively. The result follows since $(V.X)_{n-1}$ is G_{n-1}-measurable.

1.7.10. **Corollary**:
If T is a G-stopping time, and X is a martingale (supermartingale, submartingale), then the stopped process, X^T, *is a martingale (supermartingale, submartingale).*

It is easy to see that $X^T = V.X$, when $V_n = 1_{[n \leq T]}$. To show that V is G-previsible just write

$$[\, n \leq T \,] = (\bigcup_{k=1}^{n-1} [T = k])^c \in G_{n-1}.$$

Thus, the indicator function of $[n \leq T]$ is G_{n-1}-measurable, so that V is G-previsible. It only remains to show that X^T is P-integrable. Since $T(w) \wedge n \leq n$, this a consequence of

$$|X_n^T(w)| \leq \sum_0^n |X_k(w)|.$$

1.7.11. Remark: Subsets B of $Z_+ \times \Omega$ are called **random sets**. In the sequel such sets will be called previsible random sets if their indicator processes, $(n,w) \rightarrow 1_B(n,w)$, are previsible processes. Anticipating a concept that will be introduced in Chapter 2, we point out that if T is a stopping time, then random sets of the form $\{ (n,w) : n \leq T(w), (n,w) \in Z_+ \times \Omega \}$ are previsible random sets. This random set is a particular example of a **stochastic interval**, denoted $[[0,T]]$. In this instance, we would write $V = 1_{[[0,T]]}$ as a process defined on $Z_+ \times \Omega$. Notice that in the proof of the last theorem we wrote $V_n = 1_{[n \leq T]}$, thereby defining the process V by means of a sequence of random variables on Ω. These two ways of defining the same process leads to nothing new in discrete time, but once we enter the continuous time domain we will find that studying processes as families of random variables will not be adequate. It will turn out that stochastic intervals will provide an intuitive way of studying the measurability of such processes as a functions of two variables.

1.7.12. Remark: It is convenient at this point to add the following Corollary. This form of Doob's Optional Sampling (Stopping) Theorem (1953) is not stated in its most general form, but it is sufficient for our purposes. The boundedness condition imposed on the stopping times can be relaxed; such a form of Doob's theorem (in the continuous parameter case) will be stated in Chapter 2. Page 67 in Neveu [1975] contains the discrete parameter version.

1.7.13. **Theorem (Doob's Optional Sampling Theorem)**:
If X is a martingale, and S, T are bounded stopping times with $S \leq T$, then X_S and X_T are P-integrable and

$$E\{ X_T \mid G_S \} = X_S, \quad (\text{a.s.}P). \tag{14}$$

(T is a **bounded stopping time** if there exists a constant, K, such that $T(w) \leq K$ for all w in Ω.)

1.7.14. For the proof, just set $V_n = 1_{[S<n\leq T]}$, then $V_n = 1_{[n\leq T]} - 1_{[n\leq S]}$. Using the obvious linearity of transforms, and realizing as in the proof of the previous Corollary that V.X is P-integrable, this Corollary states that $Y := V.X$ is a martingale, which satisfies $Y_n = X_n^T - X_n^S$ and, in this case, satisfies $Y_0 = 0$. Because of the boundedness condition, we can choose a positive integer m such that $m \geq \max(S,T)$ on Ω. Then $Y_m = X_T - X_S$. Therefore, $0 = EY_0 = EY_m =$

$E(X_T - X_S)$. It is a simple exercise to show that $E(X_T) = E(X_S)$ is equivalent to equation (14). Let A be any element in G_S. Define the stopping times S' and T' by setting $S' = S1_A + m\,1_{A^c}$, and $T' = T1_A + m\,1_{A^c}$. Then $S' \leq T'$, and so we again have $0 = E(X_{T'} - X_{S'})$, which by definition of S' and T' can be written in the following form: $0 = E(1_A(X_T - X_S))$. Referring to the definition of conditional expectation, this last equation is exactly the statement in equation (14).

1.7.15. Remark: Recall the random walk example given at the beginning of this Chapter. The symmetric random walk, X_n, is a martingale and so $EX_0 = 0 = EX_n$. If we define the stopping times $T := \min(n{:}X_n = 1)$ and $S = 0$, then $S<T$, but since $P(X_T = 1) = 1$, we have that

$$EX_T = 1 \neq EX_S = 0.$$

The problem is that T is not a bounded stopping time. Of course. $P(T<\infty) = 1$ since the random walk is recurrent.

1.7.16. Recall the decomposition given in the first remark following the Doob Meyer Decomposition and apply the Optional Sampling Theorem to the martingale M. Then

$$E(X_T - X_S \mid G_S) = E(A_T - A_S \mid G_S),$$

where S and T are bounded stopping times with $S \leq T$. Aldous [1981] then gives the following partial converse to the Doob-Meyer Decomposition Theorem:

1.7.17. **Corollary:**
Let X be a submartingale with $X_0 = 0$ and A a previsible process. If $E\,X_T = E\,A_T$, for all bounded stopping times T, then A is the compensator of X.

For the proof, just set $M = X - A$. Then $EM_0 = 0$, so that $EM_T = 0$, for each stopping time T. As in the proof of the Optional Sampling Theorem. $E(M_T - M_S) = E(1_F(M_T - M_S))$, for all F in G_S. Hence, M is a martingale and, therefore, A is the compensator of X.

1.7.18. Remark: If X is a supermartingale, then the theorem continues to hold with the equality in equation (14) replaced by "$\leq$". Similarly, if X is a submartingale, then the equality is replaced by "$\geq$". To appreciate the importance of this result, it should be noted that Abraham Wald's theory of sequential testing is based on this theorem.

1.7.19. There is an extremely important collection of results on the convergence of martingale (super and submartingale) sequences together with the fundamental inequalities of Doob (the Maximal Inequality) and others, that could be mentioned at this point. The interested reader should consult Neveu, 1975. Some of these results will be mentioned in Chapter 2 and used in the sequel.

1.7.20. We now return to the martingale transform proper, and the quadratic variation and variance processes.

1.8. **Calculus of Martingale Transforms**: One of the simplest and most useful relationships involving transforms is **integration by parts**. Let $X = (X_n)$ and $V = (V_n)$ be processes on (Ω, F_∞). Define the process X_- from X by setting $(X_-)_k := X_{k-1}$. Then the integration by parts formula is

$$(V.X)_n(w) + (X_-.V)_n(w) = V_n(w)\, X_n(w). \tag{15}$$

The proof of (15) follows immediately from the definition of a transform by by writing down the formulae for the left side of equation (15) and verifying that the result is a telescoping sum that reduces to the product on the right side of (15).

Observing that transforms are bilinear and writing $V_k = V_{k-1} + \Delta V_k$, we see that $V.X = (V_-).X + \Delta V.X$. Therefore, we can write the integration by parts formula in the more symmetric form

$$X_n V_n = (V_- . X)_n + (X_- . V)_n + \sum_0^n \Delta V_k \Delta X_k.$$

We have already encountered the last term in this equation, namely the cross covariation process corresponding to V and X. Thus, for future reference we state the following

1.8.1. **Theorem (Integration by Parts):**

$$X_n V_n = (V_- . X)_n + (X_- . V)_n + [V,X]_n.$$

This form of integration by parts would coincide exactly with the familiar Riemann-Stieltjes or Lebesgue-Stieltjes form, if we had defined [X,X] as a summation from 1 to n instead of 0 to n. Then we would have the usual $X_n V_n - X_0 V_0$ on the left side of the last equation. We will return to this topic in Section 3.2.

1.8.2. **Examples**:

(1) $X_n = \sum_0^n d_k$, where the r.v.'s d_k are arbitrary. Then, using integration by parts,

$$X_n^2 = 2(X_{-}.X) + [X,X]_n,$$

By substitution, into this equation we obtain the following well-known formula from linear algebra:

$$(\sum_0^n d_k)^2 = 2 \sum_{0 \le j < k \le n} d_j d_k + \sum_0^n d_k^2,$$

a classical formula, but obtained here as the sum of the discrete stochastic integral of X relative to itself and the quadratic variation of X! When we complete Chapter 6 and have a stochastic integral for continuous parameter processes, we will realize that this formula in X continues to hold in exactly the same form. In particular, when X is the Brownian motion process, we will see that $[X,X](t)=t$. So the formula will read

$$X^2(t) = 2 \int_0^t X(s)\, dX(s) + t$$

and Ito's stochastic integral will not follow the "usual" rules of calculus. Kiyosi Ito, the creator of the stochastic integral relative to the Brownian motion process B, a martingale, designed this integral to have the property that the process $t \to \int_0^t g\, dB = (g.B)(t)$ is a martingale, for a useful class of processes g. This had the consequence that a number of the rules of ordinary calculus do not carry over to the Ito integral. The generalization of Ito's stochastic integral to one with a martingale integrator (Chapter 6) retains these characteristics.

A Russian mathematician, R. Stratonovich, modified Ito's definition slightly and produced a stochastic integral which followed the usual rules but necessarily lost the martingale property. This makes the above discrete application all the more interesting: the "Ito" integral may have its most natural setting in the discrete case.

It is an amusing exercise to define a discrete analogue to the Stratonovich stochastic integral: Let X and V be arbitrary discrete parameter processes and set

$$(V{:}X)_n \;:=\; \sum_0^n \left(\frac{V_k + V_{k-1}}{2}\right) \Delta X_k.$$

Then one can immediately find the following relationship between the Ito and Stratonovich transforms:

$$(V{:}X)_n \;=\; (V.X)_n - \frac{1}{2}[V,X]_n.$$

The same relationship continues to hold between the Stratonovich and Ito stochastic integrals in the case of continuous parameter processes.

In Chapter 6, after we have formally introduced the Brownian motion process and stochastic differential equations, we will see that another correction factor arises when one attempts to approximate an Ito stochastic differential equation by replacing the Brownian motion term with a member of a sequence of smooth processes. If this sequence converges to Brownian motion, then (under certain conditions) the corresponding sequence of differential equations converges in the mean to a process which satisfies a stochastic differential equation which differs from the original one by a term called the Wong-Zakai factor (see E. Wong and M. Zakai [1965]).

We conclude this example by illustrating that the (discrete) Stratonovich integral obeys the classical rules of calculus in a simple special case. Set V=X in the last equation and substitute $X_k = \Delta X_k + X_{k-1}$ into the integrand of our transform on the right of this equation to obtain

$$2(X{:}X)_n \;=\; 2(X_{-}.X)_n + [X,X]_n \;\equiv\; X_n^2.$$

The equality on the right is due to the integration by parts formula derived earlier. So, as in ordinary calculus, the discrete "Stratonovich integral" of X with respect to X is just X-squared over 2.

(2) The following process, N, is called a **discrete point process** and will be the subject of the end of this Chapter. Let $N_n := \sum_0^n d_k$, where the d_k are random variables with values in {0,1}, Bernoulli r.v.'s. Let (F_n) be a filtration and $\lambda_k = E(d_k \mid F_{k-1})$. A moment's reflection will lead one to conclude that

$$[N,N]_n = N_n$$

So, using the last theorem, we have the interesting, nonclassical formula

$$N^2 = 2(N_-.N) + N .$$

Notice that this also gives us an example of a simple process whose variance process is not deterministic:

$$<N,N>_n = \sum_0^n \lambda_k.$$

1.9. **Properties of Martingale Transforms**: We now collect some additional transform properties which will play an important role in the chapter on stochastic integration.

1.9.1. **Theorem**:
Let T be a stopping time and H, V, Y and X stochastic processes defined on the same filtered probability space. Then

$$\Delta(V.X) = V\,\Delta X; \tag{a}$$

$$[V.X,H.Y] = VH.[X,Y]; \tag{b}$$

$$H.(V.X) = (HV).X; \tag{c}$$

$$V, X \text{ previsible} \rightarrow V.X \text{ previsible}; \tag{d}$$

$$(V.X)^T = (V.X^T) = (V^T.X^T); \tag{e}$$

$$[V.X]^T = [V^T.X^T] = [V,X^T]; \tag{f}$$

$$(\tilde{X})^T = (\widetilde{X^T}) . \tag{g}$$

1.9.2. Remarks: These statements are important for later developments of the stochastic integral and its attendant calculus. In the discrete case the ease with which they can be proved belies their importance. But before demonstrating this fact we will say a few words about their meaning. If we interpret the first statement in continuous time, anticipating Chapter 6, with $\Delta Y_t = Y_t - Y_{t-}$, where $Y_{t-} := \lim_{s \to t-} Y_s$, and

$$(V.X)_t = \int_0^t V_s \, dX_s,$$

then $\Delta(V.X)_t = V_t \Delta X_t$ means that "jump" points of the integral are due entirely to the jump points of the integrator, not the integrand. The integrand only affects the magnitude of the jump. The same statements apply to the discrete parameter processes being considered in this chapter if we say that a transform has a "jump" at time n iff $\Delta(V.X)_n \neq 0$. The interesting thing to note, here and as we pass through the various types of processes on our way to the general stochastic integral, is that these and many other properties of transforms continue to hold at each step. This is very important, because after the Lebesgue-Stieltjes stochastic integral the definitions of "integral" may at first bear little resemblance to the traditional notions of such things.

As to the proofs of these statements in the context of this chapter, the first amounts to noting that $\Delta(V.X)_n$ is just the n^{th} term of the sum, $V.X$.

Part (b) of the theorem follows immediately from (a). For simplicity take V=H and X=Y. Since the general term of $[V.X,V.X]$ is $(\Delta(V.X)_n)^2$ and this equals $(V_n \Delta X_n)^2 = V_n^2 (\Delta X_n)^2$. Then (b) follows by observing that this is the general term of $V^2 . [X,X]$.

Parts (c) and (d) are immediate consequences of the definition of a transform. In particular, Part (d) has the corollary that if T is a stopping time and X is previsible, then X^T is previsible.

Now Part (c) can be used to prove Part (e). For instance, to verify this claim, take $H_n = 1_{[T \geq n]}$. Using Part (c), we obtain

$$(V.X)^T = H.(V.X) = (HV).X = (VH).X = V.(H.X) = V.X^T.$$

The rest of (e) is proved in a similar manner. Part (e) provides a mechanism by which the stochastic integrals introduced in Chapter 6 are extended to larger classes of integrators by localization and "pasting". It says that the transform of X by V stopped at T is the transform of X, stopped at T, by V.

The proof of (f) follows from similar observations. Again set $H_n = 1_{[T \geq n]}$. Then

$$\Delta[V^T, X^T] = \Delta[H.V, H.X] = (\Delta H.V)(\Delta H.X) = H\Delta V\,\Delta X = \Delta[V,X]^T,$$

since by its definition $H^2 = H$.

Finally, the proof of Part (g) uses the characterization of compensators given in Section 1.7.2, Part (d) and the fact that a stopped martingale is also a martingale.

1.10. **Discrete Parameter Point Processes**: We now introduce a discrete parameter stochastic point process theory which parallels the continuous parameter point process work done, primarily by Brémaud, from 1972 to the present. The latter material considers mainly the case where the martingale compensator of the continuous parameter point process is absolutely continuous relative to Lebesgue measure; most applied works involving martingale techniques treat this case. The necessary assumptions for the discrete parameter analogues of these results and the exact form of their conclusions can sometimes be deduced directly from this continuous parameter case and sometimes they cannot. In either case, discovering the correct form and supplying a direct proof in the discrete parameter case is usually quite simple (mathematically) and informative. As far as I can determine, however, such an approach does not appear explicitly in the literature. The basic mathematical foundation for the discrete case resides in a more general part of the theory (random measures) than point processes with absolutely continuous compensators and presents an unreasonable technical and intuitive hurdle for most applied probabilists, mathematicians and statisticians.

The only paper I am aware of that suggests the importance of working directly with discrete parameter point processes is by T. C. Brown [1983]. Brown's objective is to approximate continuous parameter point processes by the discrete case. One of his results says, roughly, that a large class of continuous parameter point processes can be approximated arbitrarily closely over intervals of random length by a discrete point process. In a later BRL report, it is our intent to use some of Brown's results together with the discrete point process calculus suggested here and the limit theory developed in Aldous [1981] to approximate stochastic network models.

1.10.1. **Definition**: An F-adapted process, $X = (X_n, (F_n))$, where $X_n : \Omega \rightarrow \{1,0\}$ and $X_0 = 0$, is called a **F-Discrete Point Process (DPP)**. $\lambda_n = E(X_n \mid F_{n-1})$ is called the **F-intensity** of the DPP.

1.10.2. Remarks: (1) Define $T_0 = 0$, and for $k \geq 1$, $k \epsilon Z_+$, set

$$T_k := \inf\{ n \epsilon Z_+ : X_n = 1, \ n > T_{k-1} \},$$

if $\{ \cdots \} \neq \phi$, and $+\infty$ otherwise.

Define $N_n := \sum_{k=0}^{n} X_k$. Then $N_n = \sum_{k \geq 1} 1_{[T_k \leq n]}$. It is immediate that (N_n) and (T_n) are equivalent representations of a DPP, (X_n). Note that (T_k) is a sequence of F-stopping times, since $[T_k \leq n] = [N_n \geq k] \epsilon F_n$ for all $k \geq 1$.

(2) Set $A_n = \sum_{0}^{n} \lambda_k$. Then $M = N - A$ is an F-martingale. The F-predictable process, A, is the **martingale compensator** of N. The concept of martingale compensator has been introduced earlier in Section 1.7.2.

The proofs of the following statements and additional results will appear in later BRL Reports, Andersen[I,II,1986].

Discrete parameter PP's are of interest here for at least three reasons: first, they present an insight into the continuous parameter version of DPP, second, they are applicable to time slotted, single channel communication networks (for example, packet radio networks) and third, as noted in the reference to T.C. Brown above, they can be used to approximate continuous parameter point processes.

1.10.3. **Theorem: (An Exponential Martingale of a Point Process)**
Let $N = (N_n, F_n)$ be an F adapted DPP with F-intensity λ, and define the process, $Y = (Y_n)$, by setting

$$Y_n = \frac{e^{aN_n}}{\prod_{0}^{n} (1 + \lambda_k (e^a - 1))} \tag{16}$$

for all real a and $n \epsilon Z_+$. Then Y is an F-martingale.

1.10.4. Remark: Assume λ_k is F_0-measurable for all k. Then

$$E(e^{a(N_n - N_m)} \mid F_m) = \prod_{m+1}^{n} (1 + \lambda_k (e^a - 1)). \tag{17}$$

We will call a process satisfying (17) a **Doubly Stochastic Bernoulli Process**. Notice that in this case $\lambda_k = E(X_k \mid F_0)$, for all k. E.g., if $F_0 = \sigma(\Lambda)$, where Λ is some r.v., then $\lambda_k = g_k(\Lambda)$, where g_k is an F_0-measureable function for each k. As in the Poisson case (Brémaud [1981]), the intensity will be said to be **driven** by Λ.

1.10.5. Remark: With Y as in the statement of the Theorem, Y satisfies

$$\Delta Y_n = Y_{n-1}\,\Delta B_n, \tag{18}$$

where $B_n = \sum_{0}^{n} (b_k - 1)$ is an F-martingale and $b_k = \dfrac{\exp(aX_k)}{1 + \lambda_k(\exp(a) - 1)}$. Equation (18) is an analogue of the continuous parameter differential equation $dY = YdB$.

1.10.6. Remark: The proof of the Theorem is almost trivial once one writes $Y_n = Y_{n-1}\, b_k$ and notices that $E(b_k \mid F_{k-1}) = 1$.

1.10.7. Remark: Using the fact that $e^{aX} = Xe^a + 1 - X$ when X takes only the values 0 and 1, it is easy to check that

$$b_k - 1 = \frac{(e^a - 1)}{1 + \lambda_k(e^a - 1)}(X_k - \lambda_k) = g_k\,\Delta m,$$

where m is the compensated martingale, $m = N - A$, with $A_n = \sum_{1}^{n}\lambda_k$. It then follows immediately from equation (18) that Y satisfies the following stochastic "integral" equation :

$$Y_n = 1 + ((g\,Y_-).m)_n,$$

where $g_k = \dfrac{(e^a - 1)}{1 + \lambda_k(e^a - 1)}$. This observation is a special case of a result due to Kabanov, Liptser, and Shiryayev [1983] for continuous parameter processes. In this sense it is also a special case of the results of C. Doleans-Dade [1970] and occurs in a similar form in P. Bremaud [1981] for continuous parameter point processes with absolutely continuous (relative to Lebesgue measure) compensators.

1.11. Introduction to Non-linear Filtering of Discrete Point Processes:

Earlier we showed (Doob's Decomposition) that an integrable (P), F-adapted sequence possessed a unique representation as the sum of an F-martingale and a predictable process. If (F_n) is an observable history, and X is F-adapted, the time evolution of X is **observable**. Therefore, Doob's result says that observable processes with finite mean values all behave as **semi-martingales**. As noted earlier this is a very general and far reaching theoretical result which becomes an important result for applications when it is noticed that a semi-martingale that is not adapted to an observable history can be projected onto an history of observation in such a way that its image is a semi-martingale, with signal and martingale parts adapted to the history of observation. This is the content of the Projection Theorem below.

If the observed history is generated by a discrete point process, then the martingale portion of this projected semi-martingale has an integral (transform) representation in terms of the observed point process. This result combined with the Projection Theorem leads directly to nonlinear filtering: the estimation of functionals of an unobservable process in terms of their projections onto an observable point process history.

1.12. *Integral Representation, Projection and Innovation*:

By a **discrete point process martingale** we mean a martingale which is adapted to the internal history of a discrete point process. An integral representation of such a martingale plays a crucial role in nonlinear filtering since it guarantees the existence of the "innovations gain", whose computation results in the construction of "filters".

The following theorem is proved in Brémaud [1981]; it is the only reference he makes to discrete PP's. However, there is a huge literature regarding the representation of continuous parameter point processes. We mention only Boel, Varaiya and Wong [1975], Davis [1976] and Chou, Meyer [1975].

1.12.1. Theorem: Integral Representation of DPP Martingales:

Let $N = (N_n, F_n)$ *be a DPP with* $F_n = \sigma(X_k, k \leq n)$ *and F-intensity* λ. *Then, if* $m = (m_n, F_n)$ *is an F-martingale, there exists an F-predictable process H, with* $E((|H|.\langle M,M\rangle)_n) < \infty$, *for all* $n \in Z_+$, *such that* $m = H.M$, *where* $M = N - A$, $A_n = \sum_0^n \lambda_k$

Because nonlinear filtering has its origins in engineering, we will follow the customary terminology of that field and refer to the value of a process at any time n as the "state" of the **dynamical system** represented by that process at time n. We have the following

1.12.2. **Theorem: (Projection of State)**:
Let Z be a semi-martingale adapted to a filtration F:

$$Z_n = Z_0 + \sum_0^n f_k + m_n,$$

where $E|Z_0| < \infty$ *and*

(1) $m = (m_n)$ *is a zero mean, F-martingale*

(2) $f = (f_n)$ *is an F-adapted process with finite mean*

(3) $O = (O_n)$, *a filtration with* O_n *contained in*

F_n *for all n and* $O_0 = \{\phi, \Omega\}$

Then there exists a zero mean, O-martingale, $\dot{m}$ *such that*

$$\hat{Z}_n = E(Z_n \mid O_n) = EZ_0 + \sum_0^n \hat{f}_k + \dot{m}_n,$$

with $\hat{f}_k = E(f_k \mid O_{k-1})$

1.12.3. Remark: In the continuous parameter case f must be taken to "progressively measurable" (Brémaud [1981]).

1.12.4. We now consider the important concept of **innovations**. Innovations were introduced by Kailath for Brownian motion processes and by Brémaud[1976, 1981] for the continuous parameter point processes. In our discrete parameter case, the following simple description of "innovations" is rigorous. This type of argument, not the concept itself, is only formal in continuous time.

Using the notation Theorem 1.12.2, suppose

(1) Let $O_n = \sigma(X_0, X_1, \ldots, X_n)$, then O_n is contained in

F_n , for all $n \geq 1$.

(2) Set $\hat{\lambda}_n = E\{\lambda_n \mid O_{n-1}\}$, where λ_n is the F_n intensity of X_n, and $\hat{A}_n = \sum_0^n \hat{\lambda}_n$. Then $\hat{M} = N - \hat{A}$ is a zero mean O-martingale.

(3)
$$\begin{aligned}\Delta\hat{M}_k &= \Delta N_k - \Delta\hat{A}_k \\ &= \Delta N_k - E(X_k \mid O_{k-1}) \\ &= \Delta N_k - E(\Delta N_k \mid O_{k-1}) \\ &= \text{Observed} - \text{Expected} \\ &\equiv \text{Innovative Information.}\end{aligned}$$

Therefore, the O-martingale, $\hat{M}$, is called the **innovation process** associated with the DPP N.

(4) *Using the DPP representation, the state projection of* Theorem 1.12.2 takes the form

$$\hat{Z}_n = EZ_0 + \sum_0^n \hat{f}_k + (K.\hat{M})_n \ .$$

The O-previsible process, K, is called the **innovations gain**.

After the following statement of the filtering problem we will show how to explicitly determine K.

1.13. The Non-Linear Filtering Problem for Discrete Point Processes:

We can now summarize the state equations and their projections by the following two systems of stochastic equations:

$$\begin{cases} Z_n = Z_0 + \sum_{k=0}^{n} f_k + m_n \;:\; m \text{ is an F-martingale.} \\ N_n = A_n + M_n \;:\; A_n = \sum_0^n \lambda_k \;;\; M \text{ is an F-martingale.} \end{cases}$$

$$\begin{cases} \hat{Z}_n = EZ_0 + \sum_0^n \hat{f}_k + \hat{m}_n \;;\; \hat{m} = K \,.\, \hat{M} \text{ is an O-martingale.} \\ N_n = \hat{A}_n + \hat{M}_n \;:\; \hat{M}_n \text{ is an O-martingale.} \end{cases}$$

The Problem: On the basis of observations on N_n construct a recursive estimator for

$$\hat{Z}_n = E(Z_n \mid O_n) .$$

All that remains is to determine K from the fact that the filtering error is orthogonal to the flow of information described by (O_n) :

$$E\{(Z_n - \hat{Z}_n)(H \,.\, \hat{M})_n\} = 0 , \tag{19}$$

for all O-predictable processes, H.

1.13.1. **An Application of Discrete Martingale Calculus**: We will illustrate the use of the martingale calculus given in the beginning of this chapter to determine the innovations gain.

Set $\phi = H \,.\, \hat{M}$, $F_n = \sum_0^n f_k$, and $\hat{F}_n = \sum_0^n \hat{f}_k$. Assume that (Z_n) is bounded. Then, using integration by parts,

$$\begin{aligned} Z_n\phi_n &= (Z_- \,.\, \phi)_n + (\phi_- \,.\, Z)_n + [Z \,,\, \phi]_n \\ &= ((HZ_-) \,.\, \hat{M})_n + (\phi_- \,.\, (F + m))_n + [F + m \,,\, \phi]_n \\ &= ((HZ_-) \,.\, M)_n + ((H \,.\, Z_-) \,.\, (A - \hat{A}))_n + (\phi_- \,.\, F)_n + (\phi_- \,.\, m)_n + \\ &\quad + (f \,.\, \phi) + H \,.\, [m \,,\, \hat{M}]. \end{aligned}$$

Making a similar calculation for $\hat{Z}_n \phi_n$, obtain

$$\hat{Z}_n \phi_n = (\hat{Z}_- . \phi)_n + (\phi_- . \hat{Z})_n + [\hat{Z} , \phi]_n$$

$$= ((H\hat{Z}_-) . \hat{M})_n + (\phi_- . (\hat{F} + \hat{m}))_n + [\hat{F} + \hat{m} , \phi]_n$$

$$= ((H\hat{Z}_-) . \hat{M})_n + (\phi_- . \hat{F})_n + (\phi_- . \hat{m})_n +$$

$$+ (\hat{f} . \phi) + KH . [\hat{M} , \hat{M}].$$

Now, taking expectations of both of these equations, using the fact that the expectations of the martingale transforms vanish, and using equation (19), we obtain

$$0 = E\{((HZ_-) . (A - \hat{A}))_n + (H(X - \hat{\lambda}) . (m + F)_n\}$$
$$- E\{((KH) . <\hat{M},\hat{M}>)_n\}.$$

It follows that $K_\nu(1 - \hat{\lambda}_\nu) = \Psi_{1,\nu} - \Psi_{2,\nu} + \Psi_{3,\nu} - \Psi_{4,\nu}$ where the processes Ψ_j (j = 1 , 2 , 3 , 4) are O-predictable and satisfy

$$E\sum^n C_\nu Z_{\nu-1} \lambda_\nu = E\sum^n C_\nu \Psi_{1\nu} \hat{\lambda}_\nu,$$

$$E\sum^n C_\nu Z_{\nu-1} \hat{\lambda}_\nu = E\sum^n C_\nu \Psi_{2\nu} \hat{\lambda}_\nu,$$

$$E\sum^n C_\nu X_\nu \Delta Z_\nu = E\sum^n C_\nu \Psi_{3\nu} \hat{\lambda}_\nu,$$

$$E\sum^n C_\nu \hat{\lambda}_\nu \Delta Z_\nu = E\sum^n C_\nu \Psi_{4\nu} \hat{\lambda}_\nu,$$

for all nonnegative, O-predictable processes, C, and $\Delta Z_\nu = f_\nu + \Delta m_\nu$.

1.13.2. These calculations follow most of the work in this area (Brémaud [1976, 1981], Davis [1978], and others; also see Yor [1977] and Van Schuppen [1977]). The formula for the gain given here, however, is slightly different from those of the listed sources because in the analogous continuous parameter, absolutely continuous compensator set-up $\Delta Z_s = \Delta m_s$.

Chapter 2. **Continuous Parameter Stochastic Processes**

2.1. **Introduction**: Stopping times (Optional times) are fundamental to the modern theory of martingales. They bring the spirit of (plane) geometry, with its attendant intuitions, to the study of these processes and they give the probabilist a way to replace the continuum with the countable. The development that we outline here is pure Claude Dellacherie [Capacites et processus stochastiques, 1972]. After J. Doob's original work, this is the next monumental work on stopping times and associated delineations of measurability. The introduction of graphs of stopping times and the notions of "previsible", "totally inaccessible", and "accessible" stopping times allow a classification of stochastic processes that is both natural and necessary for the productive development of the modern theory of semi-martingales, their applications and the general theory of stochastic processes.

For instance, a "previsible" time is one which is anticipated by the previous occurrence of a sequence of observable events. Accessible times are those whose graphs consist of pieces of the graphs of previsible times. Totally inaccessible times are therefore those times whose graphs are disjoint from the graphs of all previsible times. It then follows that the graph of every stopping time is the union of the graphs of accessible and totally inaccessible times.

Optional, accessible, and previsible times are used to construct "stochastic intervals", which in the manner of Borel are used to generate algebras of events with properties similar to those of the generators. Measurability relative to these algebras is then used to single out various classes of stochastic processes that form the building blocks of a stochastic calculus for semi-martingales which at the same time extends the classical Ito integral from Brownian motion to semi-martingale integrators and is maximal (cannot be extended further) in an intuitive, Cauchy sense.

These algebras also lead to a projection theory which yields a generalization of the conditional expectation operator for processes, and of the "infinitesimal generator" for measures.

The material in the following chapters is based primarily on Dellacherie [1972], Meyer [1973], Dellacherie and Meyer [1980], Meyer [1976], Doleans-Dade and Meyer [1970], Kunita and Watanabe [1967], Métivier [1982], Liptser and Shiryayev [1977,1978], Brémaud [1981], and most importantly, the Strasbourg Séminaires in Probability, published in the Springer-Verlag "Lecture Notes in Mathematics" from 1967 to the present.

Definitions that have been covered in the discrete parameter case and carry over with little change will be treated formally here. An attempt will be made to give some insight into others and compare some with the discrete case in the hopes of understanding each a little better.

2.2. **Filtrations**: Let (Ω,H,P) be a probability space. A family of sub-sigma algebras, F := (F(t),t≥0), of H is said to be a **filtration** on (Ω,H), iff

(i) $F(s) \subset F(t)$, when s<t.

If, in addition,

(ii) F(0) contains all P-null sets, and

(iii) $F(t+) \; (:= \bigcap_{h>0} F(t+h)\;) = F(t)$, for all t≥0,

then the filtration, F, is said to satisfy the **usual conditions** (Dellacherie, 1972). In this case, and with $H = \sigma(\bigcup_{t\geq 0} F(t)) := F(\infty)$, the structure (Ω,H,F,P) is called a **filtered probability space** satisfying the usual conditions. Finally, we note that if F satisfies (iii), F is said to be **right continuous**. The first and second conditions guarantee that each F(s) is **complete**. (As a reminder, a subset B of Ω is a **P-null set** if there exists an event A in H such that $B \subset A$ and P(A)=0.)

In addition to the σ-algebra, F(t+), we define $F(t-) := \sigma(\bigcup_{s<t} F(s))$. In general, theses algebras of events satisfy $F(t-) \subset F(t) \subset F(t+)$, for all $t \epsilon [0,\infty]$. F(t-) can be thought of as representing the history of observation "prior" to time t.

2.3. **Stochastic Processes**: A **stochastic process** is a mapping $X:[0,\infty)\times\Omega \rightarrow R$ such that, for each t ≥ 0, the mapping w→X(t,w) is **H-measurable**. (H-measurable means that the inverse image, $X_t^{-1}(B) = \{ w : X(t,w) \epsilon B \}$, under X(t) of real Borel set B, is contained in H.) More directly, in terms of familiar concepts, a stochastic process is a family of random variables, r.v.'s, indexed by t≥0.

2.3.1. The **trajectories** (**paths**) of a stochastic process, X, are the mappings t→X(t,w), indexed by w in Ω. Regularity properties attributed to a process, X, such as **continuity** or **right continuity** or **left limits** refer to the trajectories.

and will be said to hold **almost surely, relative to P** (a.s.P), if the set of all $w \in \Omega$ for which the property holds has P-measure equal to one. For example, X is continuous (a.s.P), if $P(\{w \in \Omega : (t \in R_+) \rightarrow X(t,w) \text{ is continuous}\}) = 1$; that is, if, relative to P, almost all trajectories are continuous functions on R_+. After a while the qualifier, (a.s.P.), will be taken to be understood and will only be mentioned occasionally.

However, even with all this explanation, statements about such things can be a little obscure; for example, a process which is a.s.P continuous at each t is not necessarily a.s.P continuous! Such an example is given below after Lemma 1.

2.3.2. Two processes, X and Y, are said to be **modifications** of each other if

$$P(\, w \in \Omega \colon X(t,w) = Y(t,w) \,) = 1,$$

for each $t \geq 0$. More strongly, if

$$P(\, w \in \Omega \colon X(t,w) = Y(t,w), \text{ for all } t \geq 0 \,) = 1,$$

the two processes are called **indistinguishable**.

Thus, two processes are indistinguishable if their paths coincide a.s.P. As in the discrete case, indistinguishability establishes an equivalence relation on the set of processes on the common probability space (Ω, H, P) indexed by R_+. In this sense we identify all indistinguishable processes. A process which is indistinguishable from the process that is identically zero is said to be **evanescent**. A subset B of $[0,\infty) \times \Omega$ is called a **random set**. Random sets are said to be evanescent if their indicator functions are evanescent processes. Equivalently, a random set, B, is evanescent if its projection into Ω is a P-null set. In the language of random sets two processes X and Y are indistinguishable iff the random set $\{ X \neq Y \} := \{ (t,w) : X(t,w) \neq Y(t,w),\ t \geq 0,\ w \in \Omega \}$ is evanescent.

Clearly, if X and Y are indistinguishable and X has continuous (a.s.P) paths, then Y also has continuous paths. If X and Y are modifications, unlike the discrete case, one cannot claim indistinguishability. However, we have the following (Dellacherie, 1972)

2.3.3. **Lemma**: *If X is a modification of Y and these processes are right continuous, then they are indistinguishable.*

2.3.4. Remark: Just use the modification property on the rationals, a countable set, whose union is P-null and, by right continuity, contains $(X(t) \neq Y(t))$ for all $t \geq 0$.

This Lemma is the first hint that path regularity in the form of right continuous paths will be an important assumption in this note.

2.3.5. Remark: We now give an example of two processes that are modifications of one another, but are not indistinguishable, one of them not being right continuous. Let $\Omega = R_+$, H be the real Borel sets of R_+, and P the probability measure induced on H by the standard exponential distribution. Let X be the diagonal process, X(t,w) equal to 1 on the diagonal of $R_+ \times \Omega$ and equal to 0 elsewhere. Set Y equal to 0 on $R_+ \times \Omega$. Then X is a modification of Y, since $P([X(t) \neq Y(t)]) = P(\{t\}) = 0$ for each t in R_+. To see that X is not right continuous, just note that the set of X trajectories which are not right (or left) continuous has P-measure 1: $P(\{w : w = t, \text{ for all } t \geq 0\}) = P(R_+) = 1$. Since this is the same as $P(\{w : X(t,w) \neq 0 \text{ for all } t \geq 0\}) = 1$ the two processes X and Y are certainly not indistinguisable.

What if we replace X by Z, where Z is one on the diagonal of $R_+ \times \Omega$ only when the coordinates are *rational numbers, and otherwise Z is zero? With the same* P, Z is again a modification of Y, but this time Z is a.s.P right continuous and so indistinguishable from Y.

2.3.6. We should point out that even though we have assumed that our processes are real valued, we could have been more abstract and taken the **state space** of the processes to be some measurable space (E,B(E)), where B(E) is the Borel sigma-algebra generated by the open sets in E. Much of what we will talk about here still holds in this more general case with a few qualifiers. For example, in the previous Lemma we would have had to assume that E is separable.

2.3.7. A stochastic process, X, is said to be **adapted** to the filtration F, or **F(t)-adapted**, if the mapping, $w \rightarrow X(t,w)$, is F(t)-measurable for each $t \geq 0$. Historically, adapted processes were said to be **nonanticipating**. A process X is always adapted to $F_X(t)$, the **filtration generated by X**, $F_X(t) := \sigma(\, X(s), 0 \leq s \leq t \,)$.

Clearly, under the **"usual conditions", modifications of adapted processes are adapted**.

In applications, when F(t) is interpreted as a history of the evolution of a collection of processes, an F-adapted process will be said to be **observable** relative to these processes.

2.3.8. It is important to realize that classical theories of martingales, Markov processes and stopping times were only concerned with internal histories. The modern theory on the other hand just assumes that there is a **single filtration**, a **reference family**, relative to which all processes are adapted. Stopping times are defined relative to this filtration and used to characterize several σ-algebras of random events. In the modern setting of a single filtration, applications generally involve several partially ordered families of filtrations. For example, in the nonlinear filtering encountered in Chapter 1, we have the filtrations corresponding to the state and observation processes with the "state filtration" containing the observation filtration.

2.3.9. There are several additional types of measurability that are necessary for the calculus of martingales with a continuous parameter (time) set that cannot be discerned in the discrete parameter case. For the moment, we only introduce measurability relative to the product spaces $B([0,\infty))\times F(\infty)$ and $B([0,t])\times F(t)$:

A process is said to be **measurable**, if the mapping $X:[0,\infty)\times\Omega \to R$ is $B([0,\infty))\times H$-measurable (i.e., measurable as a function of two variables). In most cases we will consider processes which are both measurable and adapted. That is, a measurable mapping of $([0,\infty)\times\Omega,\ B([0,\infty)\times H)$ into $(R,\ B(R))$ such that for each fixed t, $w\to X(t,w)$ is $F(t)$-measurable.

Notice that when $[0, \infty]$ is replaced by Z_+, as in the discrete parameter case, every process is measurable; in the first chapter adapted processes corresponded to adapted and measurable processes.

2.3.10. By restricting the notion of measurability to the time interval $[0,t]$, we obtain **measurability relative to the filtration**, or **progressive measurability relative to (F(t),t≥0)** : X is said to be **F-progressive** if the mapping $(s,w)\to X(s,w)$, restricted to $[0,t]\times\Omega$, is $(B[0,t]\times F(t))$-measurable, where $B[0,t]$ is the Borel σ-algebra of $[0,t]$. Random sets are called progressive if their indicator processes are progressive.

Clearly, if X is progressive then it is adapted and measurable. The example given after the following Lemma shows that X can be adapted without being progressive. Dellacherie and Meyer[1975, IV T15] show

2.3.11. **Lemma:** *If X is adapted and right continuous (left continuous), then X is progressive.*

2.3.12. Remark: Let X, Y be the processes given in the example after Lemma 2.3.3. Take the same probability space as in that example and define the filtration, F = (F(t)), by letting F(t) be the σ-algebra generated by the family, { {s} : s$\leq$t }. Then the diagonal process, X, is F-adapted, since [X(t) = 1] = {t}ϵF(t), but X is not F-progressive since {X = 1} equals the rectangle [0,t]$\times$[0,t] and this does not belong to B([0,t])$\times$F(t) because [0,t] does not belong to F(t), which contains only countable sets. We have already noted that X is not right continuous.

One of the consequences of the "usual conditions" is that every martingale has a modification that is right continuous and has left limits (**at each point of a path, a.s.P**). Processes which are right continuous and have left limits, are sometimes called **cadlag**, or **rcll** ; the French abbreviation, "cadlag" stands for "continu a' droite, limites a' gauche". Recently, some authors have begun referring to such processes as **Skorokhod processes**, after Russian mathematician A.N. Skorokhod [1956]. We will use the last descriptor. The full importance of the Skorokhod assumption will begin to emerge in Section 2.8. Essentially, all the processes considered in Chapters 5 and 6 will be taken to be Skorokhod.

2.4. **Stopping Times**: *Often in probability* we are interested in the time, T, at which a certain random phenomenon associated with a stochastic process, X, occurs. E.g., the first time, T(w), that the path, t $\rightarrow$ X(t,w), hits a particular level. In fact, if F(t), of the filtration F=(F(t)), is interpreted as the collection of all events associated with the evolution of a process, X, during the time interval [0,t], we can make precise the statement that this phenomenon occurred before time t by requiring that $[T\leq t] := \{w|T(w)\leq t\}$ belong to F(t), for every $t\geq 0$.

2.4.1. **Definition**: A positive r.v. T, finite or not, is called a **stopping time** (or **optional time**) relative to the filtration $F=(F(t),t\geq 0)$, if the event $[T \leq t] \epsilon F(t)$, for each $t\geq 0$. (Note: "positive" is meant in the sense of nonnegative.)

In Chapter 1 we saw that for non-negative, integer valued G-stopping times, $[T = n] \epsilon G_n$ iff $[T \leq n] \epsilon G_n$ iff $[T < n] \epsilon G_{n-1}$. Here the situation is a little different. To appreciate the difference, let T be an F-stopping time. Then

$$[T < t] = \Big(\bigcup_{\epsilon > 0} [T \leq t - \epsilon] \Big) \epsilon F(t),$$

since $[T \leq t - \epsilon]\epsilon F(t - \epsilon)\subset F(t)$, for $t\geq 0$, by monotonicity of filtrations. Therefore, if T is an F-stopping time, $[T < t]\epsilon F(t)$ and then so does $[T\geq t]$, for all

$t \geq 0$. But if all that we know about a mapping $T:\Omega \rightarrow \overline{R}_+$ is that $[T < t] \epsilon F(t)$ for all $t \geq 0$, then all we can conclude is that

$$[T \leq t] \epsilon F(t+) := \bigcap_{h>0} [T < t + h],$$

so T just an F(t+) stopping time. Therefore, if the filtration is right continuous then $[T<t] \epsilon F(t)$ for all $t \geq 0$, implies that T is an F-stopping time. Thus, under the "usual conditions" the two conditions are equivalent. As noted already we will generally assume that our filtrations are right continuous.

Any nonnegative constant is a stopping time relative to any filtration. For example, if $T(w) = c$, for all $w \epsilon \Omega$, then $[T \leq t] = \Omega$ when $c \leq t$ and $= \phi$, otherwise. If T is a stopping time and c is a nonnegative real number, then $c + T$, is also a stopping time: $[T + c \leq t] = [T \leq t - c] \epsilon F(t-c) \subset F(t)$, for all $t \geq 0$.

There are numerous interesting simple results concerning stopping times that are needed to develop an intuition about them, but covering them is beyond the scope of this short note. Probably the best treatments are given by Dellacherie(1972) and Métivier(1982). We will try to introduce only what will be needed to provide a *reasonable understanding of "previsibility"* and its role in the theory of martingales and stochastic integration.

2.4.2. We observe in passing then that the minimum and maximum of two F-stopping times are again F-stopping times. Also, the supremum of a sequence of F-stopping times is an F-stopping time:

$$[\sup\{T_n : n>0\} \leq t] = (\bigcap_{n=1}^{\infty} [T_n \leq t]) \epsilon F(t).$$

The infimum, S, of a sequence of F-stopping times is, however, an F(t+)-stopping time. That is, we can only claim $[S \leq t] \epsilon F(t+)$, for all $t \geq 0$:

$$[S \leq t] = \bigcap_{k=1}^{\infty} (\bigcup_{n=1}^{\infty} [T_n < t + \frac{1}{k}]) \epsilon \bigcap_{j=1}^{\infty} F(t + \frac{1}{j}) = F(t+).$$

But again, since we assume the "usual conditions", S is also an F-stopping time. Hence, the limsup and liminf of a sequence of F-stopping times are F-stopping times. Therefore, whenever the limit of a sequence of stopping times exists, the limit is a stopping time. Another simple fact is that the sum of any two F-stopping times is again an F-stopping time.

2.4.3. In the realm of sophisticated stopping times we mention the "hitting" time or **debut** of a random set, A, defined as $D_A(w) := \inf\{ t \epsilon R_+ : (t,w) \epsilon A \}$, or as $D_A(w) := +\infty$, if the **section** $A(w) = \{ t : (t,w) \epsilon A \}$ is empty. Dellacherie [1972] used capacity theory to show that when the filtration satisfies the usual conditions and *A is an F-progressive random set then the debut of A is an F-stopping time.* We will return to this example later in the chapter where we will introduce "k-debuts". For this purpose, notice that we can write

$$D_A(w) = \inf\{t \epsilon R_+ : [0,t] \cap A(w) \text{ contains at least one element } \}.$$

2.4.4. **Definition**: Given an F-stopping time, the family of events which occur **prior to T, denoted F(T)**, is defined as the set of all events $A \epsilon F(\infty) := \sigma(\bigcup_{t \geq 0} F(t))$, for which $A \cap [T \leq t] \; \epsilon F(t)$, for each $t \geq 0$.

2.4.5. If T is a.s.P equal to a constant time, t, then $F(T) = F(t)$. This justifies the notation F(T) when T is a stopping time. Further, it is easy to verify that F(T) is a sigma-algebra and T is F(T)-measurable. (For the latter, just observe $[T \leq t] = A \cap [T \leq t] \; \epsilon \; F(t)$, for all $t \geq 0$, where $A = [T \leq t]$, so $A \epsilon F(T)$, and consequently, T is F(T)-measurable.)

These σ-algebras are monotone at stopping times, in the sense of the next theorem.

2.4.6. **Theorem:**
Let S and T be F-stopping times. If $S \leq T$, *a.s.P, then* $F(S) \subset F(T)$.

Remark: $S \leq T$ implies $[T \leq t] \subset [S \leq t]$, so that for any $A \epsilon F(S)$

$$A \cap [T \leq t] = A \cap [S \leq t] \cap [T \leq t] \epsilon F(t).$$

Therefore, $A \epsilon F(T)$.

Remark: The following are just as easy to prove:

- o $A \epsilon F(S)$ implies $A \cap [S \leq T] \epsilon F(T)$
- o $F(\min(S,T)) = F(S) \cap F(T)$
- o $[S < T]$, $[S > T]$ and $[S = T]$ are in F(S) and F(T)

If S is a positive r.v. which is measurable relative to F(T), then S is not necessarily an F-stopping time. A sufficient condition is that $S \geq T$. Since this is simple and important, it is worth a proof. Just use $[T \leq t]$ to partition the sample space, Ω. Then since $S \geq T$, $[S \leq t] = [S \leq t] \bigcap [T \leq t]$ and, because S is F(T)-measurable, the right side of this equation is in F(t) for all $t \geq 0$. Hence, $[S \leq t] \epsilon F(t)$ for all $t \geq 0$, so that S is a stopping time.

This has the consequence that every F-stopping time, T, can be written as the limit of a decreasing sequence of F-stopping times, each taking a countable number of values: Just define T(n) by setting $N = N(n) = 2^n$ and

$$T(n) := \sum_{k>0} \frac{k}{N(n)} 1_{[k-1 \leq N(n)T<k]},$$

when T is finite, and ∞, otherwise. Then $T(n) \geq T$ and the previous result applies, making T(n) a stopping time. Also, on $\{w:T(w)<\infty\} = [T<\infty]$, we have $0 \leq (T(n,w) - T(w)) \leq 1/N(n)$; hence $T(n,w) \rightarrow T(w)$, as $n \rightarrow \infty$, for every w in Ω. Finally, if $T(n,w) = k/N(n)$, then we must either have $T(n+1,w) = (2k-1)/N(n+1) < k/N(n)$, or $T(n+1,w) = k/N(n)$. So, $T(n,w) \geq T(n+1,w)$, for all $w \epsilon \Omega$.

Notice that without the "countable valued" requirement, it is obvious that the sequence $S_n = T + \frac{1}{n}$ decreases to T a.s.P on $[T<\infty]$.

Observe carefully that one cannot make a symmetric statement relative to increasing sequences of stopping times. In the next section we will see that requiring this symmetry leads to the notion of "previsible stopping times".

2.5. **Stochastic Intervals**: Let S and T, with $S \leq T$, be two F-stopping times and set

$$[[S,T)) := \{ (t,w) \mid S(w) \leq t < T(w),\ 0 \leq t < \infty,\ w \epsilon \Omega \}.$$

[[S,T)) is called a **stochastic interval**; if we want to emphasize the underlying filtration, we will write **F-stochastic interval**. Stochastic intervals [[S,T]], ((S,T)), and so on, are defined in the same manner. If S=T, then [[T]] := [[T,T]] is called the **graph of T**.

2.5.1. F-stochastic intervals are F-progressive random sets. That is, the indicator function of an F-stochastic interval is an F-progressive process on $R_+ \times \Omega$:

First note that $(s,w) \to 1_{[[S,T))}(s,w)$ is F-adapted since $1_{[[S,T))}(s,w) = 1$, if $S(w) \leq s < T(w)$, 0 otherwise, and $[S \leq s] \bigcap [T > s] \epsilon F(s)$. Since $1_{[[S,T))}$ has right continuous paths by inspection, Lemma 2.3.11 applies. Similarly, $((S,T]]$ has an F-progressive indicator function. The other types of stochastic intervals are handled in the same way or as combinations of stochastic intervals whose regularity properties are known. E.g., $((S,T)) = ((S,T]] \bigcap [[S,T))$.

2.5.2. We can now show the converse of the result guaranteeing that a debut is a stopping time. That is, every stopping time, T, is the debut of a progressive random set: Just set $A = [[T,\infty))$. Then A is progressive and the statement follows by noting that $[D_A \leq t] = [T \leq t]$. Also note that if $A = [[S,T))$, then $D_A = S$ on $[S < T]$ and $= \infty$, on $[S = T]$.

When A is of the form $A = \{(t,w): X(t,w) \epsilon B\}$, where X is some stochastic process, the debut of A is called the **hitting time** or the **first entrance time of X** into B. By what has been said, if X is progressively measurable and B is a real Borel set, then the debut of A is a stopping time. The best discussion of this is given in Williams [1979].

2.5.3. The following example will be used later on as an example of a stopping time which is not a previsible time. (It is an exercise in Métivier [1982]). We will specialize it somewhat in order to have a simple example to illustrate the graph of a stopping time. Let A be a nonempty, proper subset of the interval $[0,1] := \Omega$. Set $F(0) = \{\phi, \Omega\}$ and $F(1) = \{A, A^c, \phi, \Omega\}$. Define the filtration $F(t) := F(0)$, if $t \epsilon [0, 1)$ and $:= F(1)$, if $t \geq 1$. Then $(F(t), t \geq 0)$ is a right continuous filtration. Set $T := 1 + 1_A$. Then T is an F-stopping time:

$$[T \leq t] = \begin{cases} \phi & \text{if } 0 \leq t < 1 \\ A^c & 1 \leq t < 2 \\ \Omega & 2 \leq t \end{cases}$$

so that $[T \leq t] \epsilon F(t)$, for all $t \geq 0$. If we take the usual two dimensional coordinate system with time (the range of T) as the horizontal axis and Ω as the interval [0,1] on the vertical axis, then with $A = [0,.5]$ the graph, $[[T]]$, is the following union of straight line segments:

$$[[T]] = \{(1,w): w \epsilon (.5,1]\} \bigcup \{(2,w): w \epsilon [0,.5]\}.$$

2.5.4. **Definition**: The family of events which **occur strictly prior to the stopping time, T**, is denoted by F(T-) and is defined as the sigma-algebra generated by F(0) and events of the form $A \cap [T > t]$ for all $A \epsilon F(t)$ and all $t \geq 0$.

As with F(T), T is F(T-)-measurable. Also, since the generators of F(T-) belong to F(T), $F(T-) \subset F(T)$. Also, if S is a stopping time with $S \leq T$ then $F(S-) \subset F(T-)$.

2.5.5. Remark: It is important to note that left continuity of F does not imply that F(T) = F(T-). For example, let $\Omega := [0, \infty)$, and B([a,b)) be the Borel σ-algebra of subsets of the interval [a, b). Set $F(t) := (B([0,t])) \bigcup \{(t,\infty), R_+\}$ for all $t \geq 0$ and note that F(t-) = F(t) = F(t+) for all non-negative t. Setting T(w) = w on Ω defines T to be an F-stopping time with $F(T) \neq F(T-)$. This is an exercise in Métivier [1982]. However, this is not meant to imply that the mathematical setup is simple. This setup, or a slight variation, is at the heart of numerous papers (e.g. Dellacherie [1970], Chou and Meyer [1975] and finally, with corrections, in Chapter 4 of Dellacherie and Meyer [1975]).

For example, in the last reference, a filtration, G, is taken to be

$$G(t) = \sigma(B\{ (0,t) \}, [t, \infty))$$

for all $t \epsilon [0, \infty]$. Then G(t+) contains { t } and (t, ∞), and these sets are not in G(t). Therefore, in this case $G(t) \neq G(t+)$ and G is not right continuous. It follows that T, the identity mapping as defined at the beginning of this remark, is a G(t+), but not a G(t), stopping time. We will return to this example at the end of the next section to illustrate the special classes of stopping times introduced there.

2.6. **Previsible, Accessible, Optional Times**: Recall again that, unless stated otherwise, we assume that the "usual conditions" hold on the underlying filtrations.

Earlier when we were approximating stopping times from above, we pointed out that they cannot in general be approximated from below by increasing sequences of stopping times. However, from the standpoint of the calculus of martingales, those stopping times that do have this property can be used to characterize the most important class of measurable processes. Dellacherie and Meyer(1980) point out that processes with this type of measurability (previsibility) play the same role in stochastic integration as Borel functions play in the classical theories of measure and integration. The construction of this class begins with the following definition.

2.6.1. **Definition**: An F-stopping time, T, is said to be **F-previsible (predictable)** if there exists a sequence (T(n)) of F-stopping times with the following properties:

(i) $T(n,w) < T(w)$, a.s.P, on $[T>0]$, $n>0$;

(ii) (T(n)) is increasing (a.s.P) and converges (a.s.P) to T.

Note: Generally, when there is no possibility of confusion, we will drop reference to the underlying filtration, F.

The sequence of stopping times, (T(n)), is said to **announce** T, and is called an **announcing sequence for T**. Clearly, if T is a stopping time and c is a positive real number, then T+c is previsible. Just take $T(n) := T + c(1 - (1/n))$, $n>0$.

Intuitively, if T is the first time an event can happen, then T is previsible if we are aware that the event is about to happen; a sequence of events takes place that foretell the occurrence of T. As a matter of fact, the announcing sequence is also called a **foretelling sequence**.

The traditional example of a previsible stopping time is first time that an adapted, continuous (hence progressive) process, X, ($X(0,w) \equiv 0$) hits a singleton set: For example,

$$T(w) := \inf\{ t : X(t,w) = 1 \}$$

and := ∞, when $\{...\} = \phi$. For definiteness, take X to be standard Brownian motion. To see that T is previsible, just take T_n, an announcing sequence of T, to be $T_n(w) := \inf\{ t : X(t,w) = 1 - \frac{1}{n} \}$.

A famous non-previsible stopping time, T, is the "time to the k^{th} event" of a Poisson process. The standard proof of this fact can be found in Liptser and Shiryayev [Vol II]. We will give a simpler but more sophisticated demonstration by Aldos [1981] that also yields a result useful later in this chapter. Let $N = (N(t), t \geq 0)$ be a Poisson process with parameter μt at time t. Then $s \rightarrow N(s+t)$ N(t) defines a Poisson process with parameter μs and so by the Strong Markov Property, $N(s+T) - N(T)$ is again Poisson, μs. Now, if we assume that T is previsible, then T+s is previsible and announcing sequences exist for both. Evaluating the Poisson increments at these announcing sequences and passing to the limit, we have that $N((s+T)-) - N(T-)$ is Poisson, μs. Remembering that N has right continuous paths and letting $s \rightarrow 0+$, we obtain $N(T) - N(T-) = 0$, a.s.P. This states that T is not a jump time of the process as originally supposed. Therefore, T cannot be previsible.

2.6.2. We now introduce a stopping time which is (a.s.P) never equal to any previsible time, appropriately, it will be called a totally inaccessible time. The time to the first jump (event) of a Poisson process is such a time. The "complement" of a totally inaccessible time will be said to be accessible. More formally, we give the following:

2.6.3. **Definition**: Let T be an F-stopping time. Then

(i) T is said to be **accessible** if there exists a sequence of previsible times, $(T(n))$, with the property $\bigcup_{n>0} [[T(n)]] \supset [[T]]$, up to an evanescent event.

(ii) T is said to be **totally inaccessible**, if the intersection, $[[T]] \bigcap [[S]]$, is empty, up to an evanescent event, for each previsible stopping time, S.

That is, the graph of an accessible time, T, is made-up of sections of graphs of previsible times, and the graph of a totally inaccessible time is disjoint with the graph of every previsible time. Parts (i) and (ii) of the definit can be written $P(\bigcup_n [T_n = T][T<\infty]) = 1$, and $P([T = S][T<\infty]) = 0$, respectively.

Remark: It is clear from the definition that if T is previsible then it is accessible and optional. The example of Dellacherie and Meyer at the end of the last section provides a case where a stopping time is nonprevisible and accessible. We state of some of their observations on this example:

We noted that T, the identity mapping as defined there, is a G(t+), but not a G(t) stopping time. Dellacherie and Meyer show further that every (G(t))-stopping time is G-predictable. Continuing with this example, a probability measure P is introduced and G is completed relative to P. Call the completed filtration G' . Dellacherie and Meyer show the following:

(a) If P is nonatomic, then G' satisfies $G'(S) = G'(S-)$ for all previsible times, S. Also, the identity mapping, T, is totally inaccessible.

(b) If P is purely atomic and nondegenerate, then the identity mapping, T, is a nonpredictable, accessible time.

2.6.4. **Definition**: If T is a stopping time and A is an event ($A \epsilon H$), set $T_A(w) = T(w)$, if $w \epsilon A$, and $= \infty$, otherwise. Then T_A is called the **restriction of the stopping time, T, to the event A**.

It follows immediately from the definition that $[T_A \leq t] = A \bigcap [T \leq t]$. Then, from the definition of F(T), T_A is a stopping time iff $A \epsilon F(T)$.

Remark: As can be guessed, the graph of any stopping time can be written (uniquely, a.s.P) as the union of the graphs of accessible and totally inaccessible times.

2.6.5. **Theorem (Dellacherie):**
Let T be a stopping time. Then there exist events A and B in F(T-) which constitute a unique, up to P-measure zero, partition of $[T<\infty]$, *such that* T_A *is accessible and* T_B *is totally inaccessible.*

We now mention a sequence of stopping time results that are useful in the study of stochastic processes. Our principal use will be in the last of these results which gives a characterization of previsibility of restrictions of stopping times.

o Let S and T be previsible (accessible, totally inaccessible) times. Then the minumun and the maximum of S and T are previsible (accessible, totally inaccessible).

o Let T be a stopping time and $A \epsilon F(t)$. If T is accessible (totally inaccessible) then T_A is accessible (totally inaccessible).

(This is immediate from $[[T_A]] \subset [[T]]$.)

o Let $T = \lim_{n \to \infty} T_n$.

(a) If (T_n) is an increasing sequence and each T_n is previsible, then T is also previsible.

(b) If (T_n) is a decreasing sequence and for each $w \in \Omega$ there exists a natural number $n = n(w)$ such that $T_{n(w)} = T(w)$, then T is previsible (accessible) whenever the T_n are previsible (accessible).

From this result it can be shown that

o If T is a stopping time then the collection of all $A \in F(T)$ such that T_A is previsible is closed under countable unions and countable intersections.

This result can then be used to show the following important result that will be used several times in the sequel:

o Let T be a stopping time and $A \in F(T)$. Then if T_A is previsible, $A \in F(T-)$. Conversely, if $A \in F(T-)$ and T is previsible, then T_A is previsible.

2.7. **Previsible, Accessible, Optional Processes**: Let X be a stochastic process on (Ω, H) and recall that H has been taken to be the smallest sigma algebra containing the union of all members of the filtration, F, and then denoted $F(\infty)$.

If T is a positive r.v. on (Ω, H), then by X(T) we mean the mapping $w \rightarrow X(T(w),w)$ of Ω into R. If X is $B[0,\infty) \times H$-measurable, then this mapping defines a r.v. since it is the composition of the measurable mappings $w \rightarrow (T(w),w)$ and $(t,w) \rightarrow X(t,w)$.

When X is a Skorokhod process, Meyer [1973] gives a simple method for approximating X(T), by $X(T_n)$, where for each n, T_n is a countable valued random variable and the sequence (T_n) decreases point-wise to T: Let $D_n = \{k/2^n : k \in Z_+\}$, and set $T_n(w)$ equal to the infimum of $D_n \bigcap (T(w),\infty)$. Then the right continuity of X gives $X(T_n) \rightarrow X(T)$, a.s.P.

X(T) is called the process **evaluated at time T**. In general, we will allow T to be an arbitrary stopping time. This means that T will be allowed to take the (nonreal) value ∞ of the extended set of positive real numbers, $\overline{R}_+$. Since we define processes X on $R_+ \times \Omega$ and not $\overline{R}_+ \times \Omega$, we write $X_T 1_{[T<\infty]}$ to denote X_T on the event $[T<\infty]$ and zero on the event $[T=\infty]$. We give the following sufficient condition for the F(T) measurability of X(T).

2.7.1. **Theorem**:
If X is F-progressively measurable and T is an F-stopping time, then $1_{[T<\infty]}X(T)$ *is F(T)-measurable.*

2.7.2. Remark: We will give a sketch of Dellacherie's proof of this result. Let A be any Borel set of the real line. We must show that $[X(T)\epsilon A]\cap[T\leq t]\epsilon F(t)$, for all nonnegative t. But the event formed by this intersection is equal to $[X(S(t))\epsilon A]\cap[T\leq t]$ where $S(t) = \min(T,t)$ is easily seen to be F(t)-measurable. But the process XoS is measurable relative to F(t), since it is obtained as the composition of the mappings $w\rightarrow(S(w),w)$ of $(\Omega,F(t))$ into $([0,t]\times\Omega,B[0,t]\times F(t))$ and $(s,t)\rightarrow X(s,w)$ of $([0,t]\times\Omega,B[0,t]\times F(t))$ into (R,B), and because of the definition of a progressive process.

2.7.3. Remark: As in Chapter 1, an important example is obtained when the process, X, is evaluated at the random time $S := T\wedge t$, where $T(w)\wedge t := \min(T(w),t)$ for $t\geq 0$. Then X(S) is called the **process stopped at time T** and denoted X^T. Thus, $X^T(t,w) = X(T(w)\wedge t,w)$, and $T\wedge t$ is sometimes called a **truncation** of T. The use of stopped processes is fundamental to the modern theory of martingales. As noted in Chapter 1, one reason for this is the Doob Optional Sampling (Stopping) Theorem and another is based on the concept of localization to be discussed at some length in Chapter 6.

Another important stopping time that can be constructed from T is the **translation** of T: $T_t = T + t$, $t\epsilon R_+$. Then $X(T_t)$ is called a **random shift** of X. (For more information see Chung, Doob [1965].)

2.7.4. Remark: For future use, we point out that the filtration F=(F(t)) is said to be **quasi-left continuous** iff F(T)=F(T-) for each previsible time, T. It can then be shown that quasi-left continuity is equivalent to accessible times being previsible.

2.7.5. In what follows we will often use the term "optional" time in place of "stopping" time.

2.7.6. **Definition** : PT(F) := "family of F-previsible times"; AT(F) := "family of F-accessible times"; OT(F) := "family of F-optional times", where F is the filtration (F(t)).

F will usually not be mentioned and in these cases we will just write PT, AT, and OT. We now define three sigma algebras of events generated by stochastic intervals from each of these families. Let K represent any one of the family of stopping times PT, AT, or OT and set

$$G(K) := \sigma\{ [[S,T)) : S\epsilon K,\ T\epsilon K \}$$

Remark: Since previsible times are accessible, G(PT) $\subset$ G(AT) and accessible times are, in particular stopping times, G(AT) $\subset$ G(OT). If we let G(Prog) denote the σ-algebra generated by progressive random sets, we have that G(OT) $\subset$ G(Prog) by applying Lemma 2.3.11 to $1_{[[S,T))}$. Hence,

$$G(PT) \subset G(AT) \subset G(OT) \subset G(Prog)$$

2.7.7. **Theorem**:

$$G(K) := \sigma\{ [[S,T]] : S\epsilon K,\ T\epsilon OT \},$$

where K is PT, AT, or OT.

To see this, let SϵK and TϵOT. Then notice that

$$[[S,T]] = \bigcap_{n>0} [[S,T+(1/n)))$$

and T + (1/n) is previsible, hence accessible and optional. Therefore the generators, [[S,T]], can be obtained from the defined generators of G(K), K=PT,AT,OT. Since

$$[[T]] = \bigcap_{n} [[T,T+(1/n)]]$$

and [[S,T)) = [[S,T]] - [[T]], the reverse is true and the proof is complete.

Remark: This demonstration also proves that [[T]]ϵG(K), if TϵK, K = PT, AT, OT. So, for example, previsible times have previsible graphs. Not surprising, but certainly comforting. Notice that although it is true, we have not proved the converse. The only proof that I am aware of requires the so-called "(Cross) Section Theorem" from Capacity theory. This will be mentioned in the Chapter on Previsible Projections (Section 4.4).

Now that we know that optional and previsible random sets are progressive, the next Theorem follows from Dellacherie's result stated earlier, saying that progressive random sets have optional debuts.

2.7.8. **Theorem:**
If the random set $A \epsilon G(OT)$ or $G(PT)$, *then the debut of* A *is a stopping time.*

2.7.9. **Definition** : The process, X, on (Ω, H) is said to be an **optional, accessible**, or, **previsible process** according to whether X is, respectively, G(OT)-measurable, G(AT)-measurable, or G(PT)-measurable.

We have already used the following

2.7.10. **Corollary:**
If X *is an optional process and* B *is a real Borel set then the hitting time of* B *is a stopping time.*

2.7.11. Remark: Let $0(w) = 0$ for all $w \epsilon \Omega$; 0 is called the **zero stopping time.** As with all boundary cases, it is instructive to satisfy oneself that 0 is a previsible stopping time. Further, 0_A is previsible, if $A \epsilon F(0)$. This is easy to show by constructing an announcing sequence. For example, let $T(n,w) := n\, 1_B(t,w)$, where B is the complement of A. Then $T(n,w) = n$ on $[0_A > 0] = [0_A = \infty] = B$, and $T(n,w) = 0$ on $[0_A = 0] = A$. Hence, $(T(n))$ is strictly increasing on $[0_A > 0]$, approaches ∞ where 0_A is infinite, and is identically zero where 0_A vanishes. *Finally, each* $T(n)$ *is a stopping time, since* $[T(n) \leq t] = A \epsilon F(0) \subset F(t)$ for all t, $0 \leq t < n$, and for $t > n$, $[T(n) \leq t] = A \bigcup B = \Omega$, which is in every $F(t)$.

The σ-algebras $G(K)$, $K \epsilon \{PT, AT, OT\}$, were defined by varying the type of stopping time in intervals of the form $[[S,T))$. This was reduced to just closed stochastic intervals with only the left end-point determining the type of measurability. The next result shows that intervals of the form $((S,T]]$, with S and T both optional, are sufficient to generate the previsible σ-algebra, G(PT), provided that we account for zero stopping times.

2.7.12. **Theorem**: (Dellacherie (1972, p.67 ff))
G(PT) is generated by $[[\, 0_A \,]]$, where $A \epsilon F(0)$, and by $((S,T]]$, where S and T are optional.

It follows immediately that G(PT) is also generated by the random sets, $B \times (s,t]$, where $B \epsilon F(s)$ and $s < t$ are any real numbers, together with $\{0\} \times B$, where $B \epsilon F(0)$. (We will call the indicator process of these sets the **kernel** process of G(PT).)

On the other hand, the indicator processes corresponding to $B \times (s,t]$ are left continuous. Hence, G(PT) is contained in the σ-algebra generated by left continuous processes. The converse statement and, consequently, part (i) of the next theorem follow from the fact that left continuous processes can be obtained as limits of linear combinations of the kernel processes of G(PT):

2.7.13. **Theorem**:

(i) G(PT) is generated by left continuous, adapted processes.

(ii) G(OT) is generated by Skorokhod, adapted processes.

2.7.14. The following example is standard. Set

$$X(t,w) := Z(w)\ 1_{[[S,T))}\ (t,w).$$

Then

(a) X is optional if $Z \epsilon F(S$, where S,T are optional;

(b) X is accessible if $Z \epsilon F(S)$, where S,T are accessible;

(c) X is previsible if $Z \epsilon F(S-)$, where S,T are previsible;

(d) $Y(t,w) = Z(w)\ 1_{((S,T]]}$ is previsible if S and T are optional and $Z \epsilon F(S)$.

For (a), first let $Z = 1_A$, $A \epsilon F(S)$. Then Z is F(S)-measurable and $Z\ 1_{[[S,T))} = 1_{[[S_A,T_A))}$, which is optional. Thus, the statement holds for indicators, hence for simple functions, hence for limits of sequences of non-negative simple functions, etc.

2.7.15. Remark: Part (i) of the last theorem might be stated more explicitly as follows: G(PT) is generated by mappings f from $[0,\infty) \times \Omega$ into R such that each function $t \rightarrow f(t,w)$ is left continuous and each function $w \rightarrow f(t,w)$ is F(t)-measurable.

2.7.16. Remark: Part (i) of the last theorem guarantees that every left continuous processes is previsible (hence also, every continuous process). However, not every previsible process is left continuous. For example, if T is a stopping time, then T+1 is previsible so that $1_{[[T+1]]}$ is a previsible process. But this process is not left continuous.

2.7.17. Remark: The "modern" (post Dellacherie [1972]) way of defining G(PT) is to use part (i) of the last Theorem as a definition, while not assuming that the filtration, F, is right continuous. Then, having defined previsible events, a stopping time is called previsible if and only if its graph, [[T]], is a previsible subset of $[0,\infty)\times\Omega$. Our definition, in these notes, then holds as a theorem with (F(t)) replaced by (F(t+)). For such a development see Métivier(1982). Under the "usual conditions" these two approaches are equivalent.

2.7.18. Remark: We have not said nearly enough about stopping times, neither their properties nor use in studying processes. So we will look with a little more detail at one small sequence of results that are important in the sequel.

First, letting A be a random set, we extend the last definition of debut (Section 2.4.8) by setting $D_A^{(1)} := D_A$ and defining, for each $n\epsilon Z_+$,

$$D_A^{(n)} := \inf\{t\epsilon R_+ : [0,t]\cap A(w) \text{ contains at least n elements}\},$$

where A(w) is the section of A at w. $D_A^{(n)}$, is called the **n-debut** of A. We have stated earlier that if A is progressive, then D_A, and so $D_A^{(1)}$, is a stopping time. Using this fact, we can show by induction on $n\epsilon Z_+$ that when A is progressive then each n-debut is a stopping time. To see this, just observe that we can write

$$D_A^{(n+1)} = D_{A\cap((D_A^{(n)},\infty))}.$$

Given that A is progressive, this equation exhibits $D_A^{(n+1)}$ as the debut of a progressive set, if $D_A^{(n)}$ is a stopping time. Observing that the 1-debut is a stopping time, and making an induction assumption that the n-debut is a stopping time, it follows then that the (n+1)-debut is a stopping time. Therefore, by induction $(D_A^{(k)}, k\epsilon Z_+)$ is a sequence of stopping times.

The following definition and Theorem are included here, not only because they will allow us to "prove" some results in Chapter 3 and beyond, but also because they give an indication of the spirit in which the use of stopping times give intuitive meaning to what could otherwise be a tedious litany of analytic conditions.

2.7.19. **Definition**: Let X be a Shorokhod, F-adapted process. Then X is said to **charge a stopping time**, T, if $P(T < \infty, X(T) \neq X(T-)) > 0$ and to have a **jump at a stopping time**, T, if $P(T < \infty, X(T) \neq X(T-)) = 1$. Further, a sequence, (T_n), of stopping times is said to **exhaust the jumps** of X if

(i) X has a jump at each T_n, $n\epsilon Z_+$,

(ii) $[[T_i]] \bigcap [[T_j]] = \phi$, $i \neq j$,

(iii) X does not charge any other stopping times.

2.7.20. Remark: Let X be an adapted, Skorokhod process, then X_- is a previsible process. Set $A = \{X \neq X_-\}$. For each $n \epsilon Z_+$, let $A_n = \{ | X - X_- | > \frac{1}{n}\}$. Then $A_n \epsilon G(OT)$ for all n and

$$A = \bigcup A_n.$$

Now, since X is Skorokhod the sections, $A_n(w)$, have no cluster points in R_+ for each n and all $w \epsilon \Omega_0$, where $P(\Omega_0) = 1$. It follows that each A_n is the union of the graphs of its k-debuts and so A is contained in the union of a countable number of graphs of stopping times. We need the following:

2.7.21. **Lemma (Dellacherie, 1972, IV T17):** *If* $A \epsilon G(K)$, *where* K *is either the class of previsible or accessible or optional times, and* $A \subset \bigcup [[S_n]]$ *for any sequence of stopping times, then there exists a sequence,* (T_n), *with* $T_n \epsilon G(k)$ *for each n and*

$$A = \bigcup [[T_n]],$$

and the graphs of the T_n *are pairwise disjoint.*

Combining this Lemma and the previous remarks we have

2.7.22. **Theorem: (Dellacherie, 1972, IV T30)**

(i) If X is any adapted, Skorokhod process, then there exists a sequence of stopping times, (T_n), *which exhaust the jumps of X*

(ii) If X is previsible (accessible), then the (T_n) *in part (i) are previsible (accessible).*

Again, let X be an adapted, Skorokhod process, then from part (ii) of this Theorem we see that if X is accessible, X cannot charge any totally accessible time. The converse of this statement is also true and is a result of the following observations. Let X be adapted and Skorokhod. Then since X does not charge any totally inaccessible time, we know from part (i) of the Theorem that the sequence (T_n) which exhausts the jumps of X must be accessible. Then $A = \bigcup [[T_n]]$ is accessible and its complement $B := \{X = X_-\}$ is accessible. Since $X = 1_B X_- + 1_A X$, and X_- is previsible, it follows that X is accessible.

Therefore, the following Corollary holds.

2.7.23. **Corollary**:
Let X be an adapted, Skorokhod process.

> *Then X is accessible iff X does not charge any totally inaccessible time.*

Remark: Earlier we introduced quasi-left continuous filtrations. The following result leads to an analogous class of processes:

2.7.24. **Theorem:**
Suppose that X is adapted and Skorokhod. Then the following statements are equivalent

> *(i) The jump times of X are totally inaccessible;*
>
> *(ii) X does not charge previsible times;*
>
> *(iii) If the stopping times $T_n \uparrow T$ then $\lim_n X(T_n) = X(T)$ on $[T<\infty]$, a.s.P.*

Remark: A process X satisfying any one of these conditions is said to be a **quasi-left continuous process**. Later in this chapter we will point out that each Shorokhod martingale is quasi-left continuous when the underlying filtration is quasi-left continuous.

From the previous theorem on the jumps of Skorokhod processes, we can see that if the Skorokhod process X is previsible, then X is quasi-left continuous iff X is (a.s.P) continuous. A more important result concerning Skorokhod previsible processes is given by

2.7.25. **Theorem: (Dellacherie, Meyer [1980])**
Let X be a Skorokhod process. Then X is previsible iff the following two conditions hold:

> *(a) $\Delta X_T = 0$, a.s.P, for all totally inaccessible stopping times T,*
>
> *(b) For every predictable stopping time T, X_T is F(T-)-*

measurable on $[T<\infty]$

Remark: Part (a) confirms the intuitive fact that previsible processes cannot jump at totally inaccessible times. Part (b) of this theorem can be strengthened as follows:

If X is previsible, then the random variable $1_{[T<\infty]}X_T$ *is F(T-) measurable for all stopping times T.* This exactly expresses the meaning of previsibility.

Since we are not going to prove the Theorem, it would be helpful to prove the last remark. Perhaps more importantly this can be accomplished by a relatively standard argument based on the Monotone Class Theorem (Appendix A). There are numerous places in this note where this device should be used, but isn't. So we will take a moment and at least show the setup. Since, in this case, X is previsible, we first look at the kernel processes $X_t(w) = 1_A(w)1_{(u,\infty)}(t)$, where $A \epsilon F(u)$ and $0 \leq u$. Then $X_T = 1_A 1_{[u<T]}$. Since T is F(T-) measurable and $A \epsilon F(u)$ $1_A 1_{[u<T]} 1_{[T<\infty]}$ is F(T-) measurable. But $A \cap [u<T]$ is a generator of F(T-). Therefore, when X has this simple form $1_{[T<\infty]}X_T$ is F(T-) measurable.

Now let H^* be the set of all such processes X such that $1_{[T<\infty]}X_T$ is F(T-) measurable for all stopping times T. Also, let L be the collection of all subsets of $(0,\infty)\times\Omega$ of the form $(u,\infty)\times A$, $u\geq 0$, $A \epsilon F(u)$. Then $1 \epsilon H^*$ and $1_B \epsilon H^*$ when B is in L. Next, it would have to be shown that if (X_n) is an increasing sequence of nonnegative functions in H^* such that $\sup_n X_n$ is finite, then $\sup_n X_n$ is in H^*.

The Monotone Class Theorem then states that H^* contains all processes measurable with respect to $\sigma(L) \equiv G(PT)$, as desired.

As a final remark about the meaning of the result itself, recall that if X is previsible, then it is progressive. Since it is progressive, we know that X(T) is F(T) measurable. Thus, we see that the more restrictive assumption of previsibility, produces the sharper result that X(T) is F(T-) measurable (as we would expect from the intuitive meaning of previsibility).

2.8. **Martingales**: This small section contains a list of some basic results on martingales that will be needed in the remaining parts of this note.

As in previous sections, all processes will be considered relative to a probability space (Ω,H,P) equipped with a filtration, $F=(F(t),t\geq 0)$. Unless stated otherwise, we assume that F satisfies the "usual conditions".

We have already discussed the martingale concept in Chapter 1, and will only note that if X is some stochastic process, $h > 0$, and we want to estimate the increment process, $t \to X(t+h) - X(t)$ on the basis of information that has accrued up to and including time t, then a reasonable estimator is $\psi(t) := E(X(t+h) - X(t) \mid F(t))$. If $\psi \equiv 0$ ($\psi \geq 0$, $\psi \leq 0$) then according to the following definitions, X is a martingale (submartingale, supermartingale).

2.8.1. **Definition**: A **F-martingale**, m, is a P-integrable process satisfying

$$E(m(t) \mid F(s)) = m(s) , \qquad \text{(a.s.P)},$$

for all $t \geq s \geq 0$.

From the properties of conditional expectation, martingales are F-adapted by their definition. **Supermartingales** are P-integrable, F-adapted processes, Y, such that $Y(s) \geq E(Y(t)|F(s))$, a.s.P, for all $t \geq s \geq 0$. Finally, X is an **F-submartingale**, if -X is a supermartingale. Clearly, a martingale is both a supermartingale and a submartingale.

2.8.2. It is proved in Meyer [1967] that

2.8.3. **Lemma:** *If the filtration F satisfies the "usual conditions", then an F-submartingale Y has a Skorokhod (right continuous with left limits) modification iff the mapping $t \to EY(t)$ is right continuous.*

2.8.4. Since we assume the "usual conditions" such modifications always exist for martingales. (This follows directly from the definition of martingale since $Em(t) = Em(0)$, $t \geq 0$. That is, martingales have constant mean value functions.) Combining Lemmas 2.3.3 and 2.8.3, we can and will always **identify a martingale with its Skorokhod modification**. Actually, Meyer proves that if a submartingale is right continuous then it has finite left limits a.s.P. and states Lemma 2.8.3 for right continuous submartingales.

However, if X is an F-submartingale which is not right continuous, all that can be said for it is that for each $t \geq 0$ and a.s.P all paths, right and left limits exist at t for the restriction of X to any countable dense subset of $[0,\infty)$. That is, letting Q be the set of nonnegative rationals,

$$P(\{ w : \lim_{s \to t+, s \in Q} X(s), \lim_{s \to t-, s \in Q} X(s) \text{ exist}\}) = 1,$$

for each t ≥ 0.

Therefore, we can define the process, Y, by setting $Y(t) := \lim_{s \to t+, s \in Q} X(s)$, for each nonnegative t on a subset C of Ω where $P(C) = 1$ and arbitrarily on $\Omega - C$, so that Y is right continuous. Further, Y is F-adapted by right continuity of our filtrations: $F(t+) = F(t)$. To see that Y is also an F-submartingale, let (h_n) be a sequence of nonnegative real numbers decreasing to zero. Then the sequence (Y_{t+h_n}) is a "reversed submartingale", due to the fact that the original process X is a submartingale, which can be shown to be uniformly integrable. We will spend some time in a few paragraphs discussing the uniform integrability condition, but for now it is enough to know that it is sufficient for a.s.P convergence to imply convergence in $L_1(P)$. Letting $A \in F(s)$ and $s \leq t$, and applying this result to

$$EY(s)1_A = \lim_{n \to \infty} EX(s+h_n)1_A \leq \lim_{n \to \infty} EX(t+h_n)1_A = EY(t)1_A,$$

we obtain $Y(s) \leq E(Y(t) \mid F(s))$; Y is an F-submartingale. Y is called the **right continuous modification** of X. Thus, under the "usual conditions" a right continuous modification of X always exists.

In the same manner, one shows that $EX(t)1_A \leq EY(t)1_A$, for all $A \in F(t)$, so that $X(t) \leq Y(t)$ a.s.P for all t. It follows from this last statement, that $X(t) = Y(t)$ a.s.P, for each t iff $EX(t) = EY(t)$ for each t. This is basically the content of Lemma 3.

2.8.5. Remark: There are a number of results from classical martingale theory that will be needed in the following chapters. One, Doob's Optional Sampling (Stopping) Theorem, has already been stated and proved in the discrete parameter case. The continuous parameter version of this theorem, and others to be stated later, follows in a relatively simple manner from the discrete version when the "usual conditions" obtain and the processes are Skorokhod. For brief, self-contained proofs see N. Ikeda and S. Watanabe [1981]. K. Chung [1974, 1983] is also an excellent source.

To state these theorems in a form convenient for application in Chapter 6, we first introduce some terminology for martingales which have finite moments of order p: M is said to be an $\mathbf{L_p}$ martingale iff M is a martingale and $M \in L_p$, where p belongs to $[1, \infty)$. A related classification, that we won't use very often until the last chapter, is L_p-bounded. A martingale M is said to be $\mathbf{L_p}$**-bounded** if

$\sup_{t \geq 0} E(|M(t)|^p) < \infty$. L_2 bounded martingales are also called **square integrable** martingales.

Theorem: (Doob's Inequality)

Let M be an L_p-martingale and $p \in [1, \infty]$. Set $M_t^* = \sup\{|M_s| : 0 \leq s \leq t\}$. Then

$$\lambda^p P(M_t^* \geq \lambda) \leq \int_{[M_t^* > \lambda]} |M_t|^p \, dP \tag{1}$$

and, if $p > 1$ then $M_t^* \in L_p$, for all $t \geq 0$, and

$$E\{(M_t^*)^p\} \leq \left(\frac{p}{p-1}\right)^p E\{|M_t|^p\}. \tag{2}$$

2.8.6. Remark: Clearly, $E|M_t|^p \leq E(M_t^{*\,p})$. So in terms of the $L_p(\Omega, H, P)$ norm, inequality (2) says that

$$\|M_t\|_p \leq \|M_t^*\|_p \leq \frac{p}{p-1} \|M_t\|_p.$$

Therefore, when $p > 1$, the mappings $m \rightarrow \|M_t\|_p$ and $m \rightarrow \|M_t^*\|_p$, **define equivalent norms**. This remark will be extremely important in Chapter 6, where the initial analysis will take place with L_2–bounded (square integrable) processes.

2.8.7. Remark: The inequality (2) is usually called Doob's inequality. Since this inequality is of great importance to us we will show how it can be deduced from (1): An application of integration by parts gives

$$E((M_t^*)^p) = p\int_0^\infty \lambda^{p-1} P(M_t^* \geq \lambda) d\lambda$$

$$\leq p\int_0^\infty \lambda^{p-2} \int_{[M_t^* > \lambda]} |M_t| \, dP \, d\lambda$$

$$= p\int_0^\infty \lambda^{p-2} \int_\Omega 1_{[M_t^*>\lambda]} \,|M_t|\, dP\, d\lambda$$

$$= p\int_0^\infty \lambda^{p-2} \int_0^\infty P[M_t^*>\lambda, \quad |M_t| > \mu]\, d\mu\, d\lambda$$

$$= \frac{p}{p-1}\int_0^\infty \{(p-1)\int_0^\infty \lambda^{p-2}\, P[M_t^* \geq \lambda, \quad |M_t| > \mu]\, d\lambda \;\}d\mu$$

$$= \frac{p}{p-1}\int_0^\infty E(1_{[\,|M_t|>\mu]} \;(M_t^*)^{p-1})\, d\mu$$

$$= \frac{p}{p-1}\; E((M_t^*)^{p-1}\, |M_t|\,)$$

$$\leq \frac{p}{p-1}\; (E(M_t^*)^p)^{\frac{p-1}{p}} \;(E(M_t^p))^{\frac{1}{p}}.$$

This last inequality is a consequence of Holder's inequality. The result follows by dividing both sides of the last inequality by the first term on the right; if it were zero, there would be nothing to prove.

2.8.8. Remark: Uniform integrability of a family of functions is a classical concept. (E.g., Meyer [1967], Loève [1960].) Since it plays a somewhat remarkable role in the theory of martingales, we are obliged to spend some time discussing the concept and its application to martingales. The principal use of this material will be to construct the Stochastic Integral in Chapter 6.

A family, Ψ, of P-integrable random variables on (Ω,H) is said to **uniformly integrable** iff

$$\lim_{a\to\infty} \sup \Big(\int_{[\,|X|>a\,]} |X(w)|\, P(dw) : X \in \Psi \Big) = 0.$$

Let $m = (m(t), t \geq 0)$ be a martingale. We will see that for $\Psi = \{m(t), t \epsilon R_+\}$, uniform integrability can be characterized by the requirement that there exists a P-integrable r.v. Z which **closes** the martingale in the sense $m(t) = E\{Z \mid F(t)\}$, for all $t \geq 0$. In this case it can be shown that $Em(t) = EZ$ for all $t \geq 0$.

$$\lim_{t \to \infty} m(t) = Z , \qquad \text{a.s.P},$$

and also in L_1. Z is usually denoted by $Z(\infty)$, and is called the **terminal random variable** of the process m.

We will now indicate how this result and some others are derived with the aid of uniform integrability. More exactly, we will discuss uniform integrability and its impact on supermartingale and martingale sequences. The transfer of these results to the "continuous" parameter case is simple for the processes under consideration in this note (they are Skorokhod processes).

We will quote two principal sources as we proceed and the interested reader can refer to these for complete details. However, an attempt will be made to supply the basic mathematical ideas that yield the results. First of all, Meyer[1967, p.17] points out that every finite family of processes is uniformly integrable and every family majorized by a P-integrable process is uniformly integrable. To understand his remark about finite families, consider $\Psi = \{h\}$, a family with only a single $L_1(P)$ random variable. Then

$$\int_{[|h| > a]} |h(w)| \; dP(w) \to 0, \text{ as } a \to \infty ,$$

since $h \epsilon L_1(P)$, $P([|h| > a]) \to 0$ as $a \to \infty$, and the measure determined by the map $B \to \int_B |h| \, dP$ of H into R_+ is absolutely continuous relative to P.

The case for finite Φ follows immediately, as does the case where a family is dominated by a single P-integrable function. These observations are essentially contained in a characterization given by Meyer[1967, IIT19] which states that uniform integrability is equivalent to the uniform boundedness of $E|f|$ for all $f \epsilon \Psi$ (i.e., $\sup_{f \epsilon \Psi} E|f| < \infty$) and the "uniform" absolute continuity of the measures $B \to \int_B |h| \, dP$, B in H, f in Ψ.

The uniform boundedness condition, $\sup_n E|f_n| < \infty$, implies that

$$\sup E\, f_n^- < \infty \quad \text{and} \quad \sup E\, f_n^+ < \infty\ . \tag{1}$$

Because of the monotonicity of their expectations, supermartingales with the first part of condition (1) are uniformly bounded; the same is true for submartingales with the second part of condition (1).

It is well known that condition (1) is a sufficient condition for a supermartingale (submartingale) (f_n, $F_n, n \epsilon Z_+$) to converge a.s.P to a **terminal** random variable, denoted f_∞, with the property that $E(f_\infty) \leq \lim_n E\, f_n$.

Another consequence of uniform integrability (Meyer[1976, II T21]) is that it extends Lebesgue's Theorem and tells us that the a.s.P convergence of f_n to f_∞ also takes place in $L_1(P)$. Therefore, $Ef_\infty = \lim_{n \to \infty} Ef_n$.

Thus, if (f_n, F_n) is a uniformly integrable supermartingale, then this supermartingale converges a.s.P and in L_1 to a terminal random variable. Consequently, (f_n , F_n , $n \epsilon \overline{Z}_+$), where $\overline{Z}_+ = Z_+ \bigcup \infty$, is also a supermartingale. That is, under uniform integrability, the time domain of a supermartingale can be extended to $\overline{Z}_+$ in an obvious manner and the resulting process continues to be a supermartingale.

In terms of martingales, this says that for every n, we can write $f_n = E(f_\infty \mid F_n)$. Moreover, the converse of this result is true in the following sense: If there exists an L_1 random variable, U, such that $f_n = E(U \mid F_n)$, then (f_n , F_n) is a uniformly integrable martingale. That ($E(U \mid F_n)$) is a martingale is obvious. That it is uniformly integrable follows from the following Lemma, which is of general interest.

2.8.9. **Lemma**

Let U be an L_1 (P) *random variable and C be a collection of sub-σ-algebras of the σ-algebra H. Then the family* { $E(U \mid G) : G$ *belongs to C* } *is uniformly integrable.*

2.8.10. Remark: This is quite easy to prove. Just use the Chebyshev inequality for positive random variables and Jensen's inequality to show that

$$\sup_{G \epsilon C} P(\ |E(U \mid G)| \geq a\ ) \leq \frac{1}{a} E\,|U|\,.$$

We now state three Theorems that are basic to the development carried out in Chapter 6.

2.8.11. **Theorem: (Martingale Convergence)**
Let M be an F, L_p-martingale, $p \epsilon [1,\infty]$ and suppose that $\sup\{ E|M_t|^p : 0 \leq t < \infty \} < \infty$. *Then there exists a random variable* $M_\infty \epsilon L^p$ *such that*

$$\lim_{t \to \infty} M_t = M_\infty \quad \text{a.s.P.}$$

Further, if p = 1, and M is uniformly integrable, or simply, if p>1, then (M_t) *is an F,* L_p*-martingale, where now* $t \epsilon [0,\infty]$*, the extended, positive real line, with* $F(\infty) = \sigma(\bigcup_{s \geq 0} F(s))$*, and* M_t *converges to* M_∞*, in* L_p*, as* $t \to \infty$.

Remark: See Chung-Williams [1983], or Meyer [1967].

2.8.12. **Theorem**:
If $m = (m_t, t \geq 0)$ *is a Skorokhod supermartingale (submartingale) and* $\sup_{s \leq t} E m_t^- < \infty$ $(\sup_{s \leq t} E m_t^+ < \inf)$, *then* $m_t \to m_\infty$ as $t \to \infty$, *a.s.P, and* $m_\infty \epsilon L_1(P)$.

2.8.13. Remark: Recalling condition (1) and remarks, it follows that any uniformly integrable martingale, m, has a terminal r.v. m_∞: $m_t \to m_\infty$, a.s.P, in $L_1(P)$, and $m_t = E(m_\infty \mid F(t))$. Conversely, if $Z \epsilon L_1(P)$, there exists a uniformly integrable martingale, m, such that $m_t = E(Z \mid F(t))$.

2.8.14. **Theorem (Doob's Optional Sampling Theorem)**:
Let X be a Skorokhod supermartingale and suppose that there exists a r.v. $Y \epsilon L_1(P)$ *such that* $X_t \geq E(Y \mid F(t))$, $t \geq 0$. *Let S and T be F-stopping times with* $S \leq T$, *then* X_S *and* X_T *are P-integrable, and* $X_S \geq E(X_T \mid F(T))$.

Chapter 3. **Increasing Processes**

3.1. **Point Processes**: The reader should recall the discussion on discrete point processes in Chapter 1. Let $(T(n), n>0)$ be a sequence of positive random variables defined on some filtered probability space, (Ω, H, F, P). This sequence is called a **point process** (PP) if $(T(n))$ is an increasing sequence on Ω with values in $(0,\infty]$ which satisfies $T(n,w) < T(n+1,w)$ for each natural number n and each w in Ω, if $T(n,w) < \infty$. We immediately extend the definition by setting $T(0,w) = 0$ and $T(\infty,w)$ equal to the limit of the $T(n,w)$ as n approaches infinity, for each w in Ω. Only on one or two occasions in this note will each $T(n)$ not be a stopping time relative to some non-trivial filtration. (As random variables, the $T(n)$ are always stopping times relative to the **trivial filtration**, which is defined as H for every "time" t.)

3.1.1. The "counting process", $N = (N(t), t\geq 0)$, associated with a point process $(T(n), n\geq 0)$, is the stochastic process defined by setting $N(t,w) := n$, if $T(n,w) \leq t < T(n+1,w)$, and $:= \infty$, if $T(\infty)\leq t$. It follows that for $t\geq 0$ and $w\epsilon\Omega$,

$$N(t,w) = \sum_{n\geq 1} 1_{[[T(n),\infty))}(t,w).$$

Since N and (T(n)) both contain the same information it is usual refer to each as a point process. We will adopt this custom and reserve the name **counting process** for those point processes which are **non-explosive**, in the sense that

$$N(t) < \infty \text{ , for all real t, } t\geq 0.$$

This condition is equivalent to

$$\lim_{n\to\infty} T(n) = \infty.$$

Notice that the non-explosive condition does not preclude either $N(\infty) = \infty$ or $N(\infty)<\infty$. In both cases $\lim T(n) = \infty$. If, for example, N only has a single jump, then T(n) is equal to ∞ for $n \geq 2$, by definition of the sequence as a point process. The jump times of nonexplosive point processes, our counting processes, do not have finite limit points.

Finally, in the Chapter on Dual Previsible Projections, it will be shown that corresponding to every point process, N, there is a unique (a.s.P), previsible, increasing process denoted $\tilde{N}$, called the **previsible compensator** or the **dual previsible projection**, such that

$$N(t) - \tilde{N}(t) \text{ is an F-martingale,} \tag{1}$$

and

$$\int_{[T(\infty),\infty]} d\,\tilde{N}(s) = 0. \tag{2}$$

Jacod(1975) shows that there is a version of $\tilde{N}$ satisfying (2) and having the property that $\Delta\tilde{N} \leq 1$, for each $t \geq 0$.

We note that the point process, N, is also called a **simple point process** by virtue of the fact that its jumps are always equal to 1.

When we want to remind the reader of the underlying filtered probability space, we will write (N,P) or (N,F) for the point process and often refer to the (P,F)-point process.

Although we will deal almost exclusively with counting processes, most of the important results holding for such processes carry over to point processes and the more general class of **marked point processes**. In order to take marked point processes into account and also to use these more general processes to understand the meaning (limitations) of the assumptions characterizing counting processes, we will introduce marked point processes here and give a few examples. These processes will be studied in more detail in Chapter 4 and again at the end Chapter 6.

We let $Z = (Z(n), n \geq 0)$ be an arbitrary sequence of random variables defined on Ω and taking values in a space E; let (E,ξ) be a measurable space. Then, with (T(n)) as above, the double sequence (T(n),Z(n)) is called a **marked point process** and E is called the **mark space**.

If we define the process $N^A = (N^A(t), t \geq 0)$ by

$$N^A(w,t) := \sum_{n \geq 1} 1_{[T(n) \leq t, Z(n) \epsilon A\,]}$$

for A in ξ. Then $N(t,w) := N^E(w,t)$ is the point process introduced earlier.

The mapping

$$A \rightarrow \mu(w,[0,t]\times A) \;:=\; N^A(w,t)$$

defines a **random measure** on $[0,\infty)\times E$.

This random measure is the primary object of study in the classical (*sans* martingales) approach to point processes. This is exemplified in the works of Kallenberg[1976,1982] and Matthes, et al. Recently, work has started to appear (Hoeven) combining martingale and random measure approaches. The first significant modern work on random measures by the martingale community is Jacod [1975]; we will return to this paper and random measures in Chapter 4.

We will end this little digression with some examples of marked point processes:

(a) $E = \{1\}$. Then $n(t,E)$ is just our original point process.

(b) $E = \{1,2,3,...,k\}$, then Z might be the number of messages arriving at a computer at some random time, T(n).

(c) Use E as in (b). Brémaud defines the multivariate counting process, $N = (N(t),t\geq 0)$ by setting

$$N(t,w,i) \;=\; \sum_{n>0} 1_{[[\,T(n),\infty]]}(t,w)1_{[\,Z(n)=i\,]}(w)$$

and then defining N(t) by $N(t) = (N(t,1),...,N(t,k))$.

Naturally, most univariate counting process results carry over to this multivariate process, including results on nonlinear filtering. We will not utilize this below. In applications to stochastic networks of queues it plays a significant role.

(d) This example is really about counting processes. Just note that when $E=\{1\}$, the study of $(T(n),n>0)$ includes the study of renewal processes as a special case, where the interoccurrence times

$$S_{n+1} \;:=\; T(n+1)-T(n)$$

are assumed to be independent and identically distributed. Hence, it includes the model for life testing, for example. Note that the counting process makes no assumptions of interoccurrence time independence, and certainly no distributional assumptions.

(e) It is probably clear that with the proper assumptions the marked point process (T(n),Z(n)) is also a model for countable state Markov processes! A characterization of such processes can be given using the notion of "dual previsible projection".

Though by no means exhaustive, these examples should convince the reader that point processes can be used in a wide variety of applications. We will look more carefully at one particular application in the sequel.

3.2. **Increasing Processes and Lebesgue-Stieltjes Stochastic Integrals**: In Chapter 6, where the major properties of stochastic integrals with respect to martingales are developed, we require some elementary facts about one of the simplest of stochastic integrals, namely those involving integration with respect to processes whose paths are of bounded variation. This theory alone would be sufficient for the nonlinear filtering problem if we were able to restrict our problems to those dynamical systems where the state process was of bounded variation.

We will assume that the reader recalls the definition of a real valued function of bounded variation defined on R. It is sufficient to recall that every such function can be characterized as the difference of two non-decreasing functions (Also see the Odds and Ends Appendix).

3.2.1. **Definition**: An F-adapted, P-integrable, nonnegative process $A = (A(t), t \geq 0)$ is said to be **increasing** if the paths, $t \rightarrow A(t,w)$, are increasing and right continuous, a.s.P (satisfying $A(t) < \infty$, a.s.P, for all $t \epsilon R_+$). Note that "increasing" does not mean "strictly" increasing.

Additionally, A is said to be **integrable** if $A(\infty) = \lim_{t \rightarrow \infty} A(t)$, which always exists, is P-integrable, that is, if $EA(\infty) < +\infty$. Then $EA(t) < +\infty$, for all $t \geq 0$.

3.2.2. Remark: Numerous authors talk about increasing processes on the extended real line, $[0,\infty]$, and not wanting to exclude "jumps" at ∞, write the limit of A(t) as $t \rightarrow \infty$ as $A(\infty-)$. The jump at infinity is then just $A(\infty) - A(\infty-)$. In this case, when A is defined on $[0,\infty]$, it is said to be integrable if $E\{A(\infty-)\} < \infty$.

3.2.3. Remark: By definition, an increasing process, A, has increasing, right continuous paths, a.s.P. In particular then, A is Skorokhod. Further, A is adapted, so every increasing process is optional relative to $(F(t))$. When A is an increasing process relative to the trivial filtration, $F(t) = H$ for all $t \geq 0$, A is said to be a **raw increasing process**. When a given discussion involves a nontrivial filtration, and we want to talk about a raw increasing process, we will often just say A is an increasing, not necessarily adapted, process. These distinctions become important in applications as well as in the theory, since the state of a process is "observable" if the process is adapted to the filtration representing the observable history.

3.2.4. Remark: Denote by $V^+ = V^+(F,P)$ the family of equivalence classes (under indistinguishability) of increasing processes. Set $BV := V^+ - V^+$. Then BV is called the space of processes of **bounded variation**, or **finite variation**. In particular, elements of BV have the property that almost every sample path of each process is of bounded variation on compact subsets of R_+.

Let IV^+ be that subset of V^+ consisting of increasing, integrable processes and IV be the set of differences of members of IV^+ . IV is then called the space of processes of **integrable variation**. $A \epsilon IV$ implies $E\int_0^\infty |\, dA(s)\,| < \infty$.

3.2.5. Let $X = (X(t), t \geq 0)$ be a measurable process, and $A \epsilon V^+$. Then with each path, $t \rightarrow A(t,w)$, we can associate a Lebesgue-Stieltjes integral

$$\int_0^t X(s,w)\, dA(s,w) \;:=\; \int_{(0,t]} X(s,w)\, dA(s,w)$$

where, as is the custom, $dA(s,w)$ represents the measure associated with A, for each w:

$$dA((a,b\,],w) := A\,(b,w) - A\,(a,w)$$

Now let X be a measurable process such that $E(\int_0^\infty |\,X(s)\,|\;|\,dA(s)\,|\;)$ is finite. Denote **the family of such processes, X, by** $L_1(A)$. Then for $X \epsilon L_1(A)$ the process

$$I(t,w) := \int_0^t X(s,w)dA(s,w)$$

is well defined up to a set of P-measure zero. This is because Fubini's theorem guarantees that

$$w \to \int_0^t X(s,w)dA(s,w)$$

is H-measurable (See Appendix A for details). Hence, I(t) is a random variable for each $t \geq 0$. Indistinguishable versions of the process $I = (I(t), t \geq 0)$ are identified and the resulting equivalence classes are denoted by $X.A = ((X.A)(t), t \geq 0)$ and called the (Lebesgue-Stieltjes) **stochastic integral** of X relative to A. As usual we have suppressed $w \epsilon \Omega$.

3.2.6. Let $X \epsilon L_1(A)$, with $A \epsilon V^+$. Then, the process $((X.A)(t), t \geq 0)$ is continuous on the right (continuous, if A is continuous), and therefore by Lemma 2 of Chapter 2, it is a progressive process. Hence, by an earlier remark, (X.A)(T) is F(T)-measurable for each stopping time, T.

Also, since we can write,

$$X.A = (X^+).A - (X^-).A \ ,$$

X.A is the difference of two increasing functions, and hence, is a function of bounded variation.

Further, if X is assumed to be F-progressive, then since A is F-adapted, Fubini's theorem tells us that the process (X.A)(t) is F(t)-measurable (F-adapted) for each $t \geq 0$. Hence, in this case X.A is an optional process (since we have already noted that X.A is a right continuous process).

3.2.7. *Remark: From Chapter 2, section 2.7,* we know that if A is an increasing process, and therefore Skorokhod and adapted, there exists a sequence of stopping times (T(n)) which exhaust the jumps of A and have the same measurability as A. Set $A^d(t) = \sum_n (A(T(n)) - A(T(n)-)) 1_{[[T(n),\infty))}(t)$. Then A^d is increasing and $A^c := A - A^d$ is continuous and increasing. Therefore, A^c is previsible and so if A is previsible, then A^d is also. Finally, the decomposition $A = A^c + A^d$ is unique, in the usual sense.

It is shown in Dellacherie [1972] that A can be written in the form

$$A = A^c + \sum_{n>0} a(n)\, 1_{[[T(n),\infty))},$$

where $a(n) > 0$ for all n. A^c is the **continuous part** of A, and the process A^d is called the **purely discontinuous** part of A.

It follows that

$$X.A = X.(A^c) + \sum_{n>0} a(n)\, X(T(n))\, 1_{[[T(n),\infty))}$$

and so X.A is previsible, if A and X are previsible or A is continuous (since, in the latter case, X.A is continuous).

This equation has the obvious consequence that when A is a counting process, N = (N(t), t≥0), where $\Delta N(t) = 1$ or 0 for all t,

$$X.N = \sum_{n\geq 0} X(T(n))\, 1_{[[T(n),\infty))}$$

3.2.8. The following Theorem is well known and easily proved. It was stated in Chapter 1 for discrete parameter processes and will be extended, in Chapter 6, to stochastic integrals with local martingale integrators.

3.2.9. **Theorem**:
Let A and B be two Skorokhod processes in BV. Then AB∊BV and

$$A(t)B(t) - A(0)B(0) = \int_0^t (A(s)\, dB(s) + B(s-)\, dA(s)), \tag{3}$$

and

$$A(t)B(t) = \int_{[0,t]} (A(s)\, dB(s) + B(s-)\, dA(s)), \tag{4}$$

for $t \geq 0$. *That is, in (3)*

$$A(t)B(t) - A(0)B(0) \;=\; (A \,.\, B)(t) + (B_{-} \,.A)(t)\ ,$$

where $B_{-}(t) := B(t-) := \lim_{s \to t-} B(s)$.

3.2.10. Remark: Equation (4) is the correct analogue to the Chapter 1 integration by parts formula.

3.2.11. Remark: The last equation is often written in the "differential" form $d(AB) = AdB + B_{-}dA$. Of course, this has meaning only through the Theorem.

As in discrete case, the integration by parts equation can be rewritten in a more symmetric form,

3.2.12. **Corollary**:

$$d(AB) \;=\; A_{-}dB + B_{-}dA + d[A,B], \tag{5}$$

where, the **square brackets**, or **cross quadratic variation** process is given by

$$[A,B](t) := \sum_{0 \leq s \leq t} \Delta A(t)\Delta B(t),$$

where the summation is taken over the countable number of common discontinuities of the bounded variation processes, A and B, and $\Delta\ A(t) := A(t) - A(t\text{-})$

Clearly, the equation for d(AB) is obtained from the Theorem by noticing that the Lebesgue Stieltjes integral, $(A - A_{-}).B$, is just [A,B].

The importance of the representation for d(AB) above will be recognized when it is demonstrated that the natural integrands for the stochastic integrals defined below are previsible processes (and for instance A_{-} is previsible).

3.2.13. Remark: Recall that any martingale can be taken to be Skorokhod. If our filtration $F = (F(t), t \geq 0)$ is quasi-left continuous, then any F-martingale, of BV, is continuous at previsible times:

$$M(T) = E(\ M(T)\ \mid\ F(T)\ ) = E(\ M(T)\ \mid\ F(T-)\ ) = M(T-).$$

Since under quasi left continuity, accessible times are previsible, we have that (BV) martingales can jump only at totally inaccessible times. Therefore, under this condition on the filtration, integration with respect to martingales of bounded variation is even permissible in the **Riemann-Stieltjes** sense when the integrands are continuous at totally inaccessible times. For example, when the integrand only jumps at previsible times, as in the case of Skorokhod previsible processes (Jacod,1979).

3.2.14. Remark: Liptser and Shiryayev[1978,Vol 2, p261] give a very informative example. Considering the LS integral of a Poisson process relative to the centered Poisson process, they demonstrate that (N.M)(t) is not a martingale, but that $(N_{-}.M)(t)$ is one, where $N_{-}(t)=N(t-)$ and $M(t) = N(t) - ct$, $t\geq 0$, c the Poisson parameter of N. Notice that $N_{-}(t)$ is previsible, because it is left continuous.

3.2.15. Remark: The following result is proved by Doleans and Meyer(1970,p.89).

3.2.16. **Theorem**:
If X is an F-previsible process, M is an F-adapted martingale which belongs to IV and $X \in L_1(M)$, then (X.M)(t), $t\geq 0$, is an F-martingale.

We have seen the analogous result for martingale transforms in Chapter 1. More such results will be seen in Chapter 6 as stochastic integrals are extended to wider and wider classes of integrators. Moreover, these stochastic integrals will agree with the Lebesgue Stieltjes stochastic integral when the "integrator" martingale is taken to be a member of IV.

3.2.17. Remark: It is easy to show that M is a F-martingale iff $E(X.M)(t) = 0$, for all F-previsible kernels, $X = 1_{B\times(s,t]}$, $t\geq s$ and $B\epsilon F(s)$. Brémaud [1981] points this out and observes that this is just one of the many reasons why previsible processes play a central role in the theory of stochastic integration. As we proceed we will meet numerous other instances to support this position.

Chapter 4. **Dual Previsible and Previsible Projections**

4.1. **Introduction**: As observed earlier in these notes, measurable processes are not necessarily adapted. If X is such a process (i.e., measurable, not adapted) and the filtration F=(F(t)) is interpreted as the history of observation connected with some experiment, then **path segments** of X, $X_{[0,t]} := (X(s), 0 \leq s \leq t)$, are not observable outcomes of this experiment for every time t. Therefore, if it is appropriate to derive information about the evolution of X=(X(t)) over a time interval [0,t] from this experiment, this information must be estimated with the aid of the history of outcomes of the experiment. One such estimate is $\hat{X}(t,w) := E(X(t) \mid F(t))(w)$, a.s.P. If one intends to use $\hat{X}$ to estimate the path segments, $X_{[0,t]}$, however, then one is faced with the seemingly impossible task of pasting together an uncountable number of the **versions** of $\hat{X}(s)$ to obtain $\hat{X}(s,w)$ for all $s \leq t$ and all w in some set K with the property that P(K) = 1. The results of this section show that this can be accomplished uniquely, provided that the estimating process is carried out at optional or previsible times.

In order to look at the results of this section from another direction, suppose that the process X is adapted to (F(t)). Then it is well known that X is determined by (F(t)) through $E\{ X(t) \mid F(t) \} = X(t)$, a.s.P $(\hat{X}(t) = X(t)$ a.s.P). That is, when X is F(t)-adapted, X(t) is determined by the integrals $E\{ X(t)\, 1_A \}$, *for all* $A \epsilon F(t)$. Results of the first part of this section show that previsible processes X are uniquely determined by the P - integrals of X_T on $[T < \infty]$, where T is previsible.

These two observations concern "previsible (optional) projections" of a stochastic process. The majority of results of this section concern the "dual previsible projection" of a process. This projection concerns increasing processes and it plays a fundamental role in the calculus of martingales. The dual previsible projection will be defined in terms of previsible projections and "admissible" measures, the latter coming up next.

We will assume throughout this section that the underlying filtration $F = (F(t), t \geq 0)$ satisfies the "usual conditions".

4.2. **Measures Generated by Increasing Processes:** Let μ be a measure on a sub σ algebra G of $B([0,\infty)) \times H$, where (Ω,H,F,P) is the underlying filtered probability space with $H = \sigma(\bigcup_{t \geq 0} F(t)) := F(\infty)$.

We will follow Métivier and call μ **admissible**, if for $B \epsilon G$ and B evanescent,

$\mu(B) = 0$. This is similar to absolute continuity of measures: Let $\Pi(B)$ be the projection of B onto Ω. Then μ is admissible if $P(\Pi(B)) = 0$ implies $\mu(B) = 0$.

As an example of such a measure, let A be an increasing, integrable ($A(\infty)\epsilon L_1(P)$) process and set

$$\mu(C) = E\{ \int_{[0,\infty)} 1_C(s,w)dA(s,w)\} , \tag{1}$$

where $C \epsilon G = B(R_+)\times H$. This measure, μ, is admissible and bounded on $B([0,\infty))\times H$.

With A as above, define μ_A, by setting

$$\mu_A(X) = E\{ \int_{[0,\infty)} X(s)dA(s)\} \tag{2}$$

on the space of bounded measurable processes X. Then μ_A is a linear functional on this space. Observe that $\mu_A(1_C) = \mu(C)$.

It is easy to show (e.g., first use simple processes, then pass to the limit) that

$$\mu_A (X) = \int_{R_+\times \Omega} X \, d\mu \tag{3}$$

for X measurable and bounded (or X positive) and μ as in (1). In a common abuse of the language, both μ and μ_A are often referred to as measures, μ_A as the **measure generated** by A.

4.2.1. Remark: Later in this Chapter, measures generated by increasing processes will be characterized and used to introduce and study the notion of "dual previsible projection". Prior to this development, such measures together with previsible projections will be used to state a criterion for the previsibility of raw increasing processes.

4.2.2. Now take μ to be as defined in (1) and let T be an F-Optional time. Set $(s,w) \rightarrow A(s,w) := 1_{[[T,\infty))}(s,w)$. Then A jumps at [[T]], the graph of T, and is equal to 1 to the "right" of [[T]], that is, on $[[T,\infty))$.

Therefore,

$$\int_{[0,\infty)} 1_C\,(s,w)dA(s,w) = 1_C\,(T(w),w)\; 1_{[T<\infty]}\,(w).$$

It follows that

$$\mu(C) = \int_\Omega 1_C\,(T(w),w)\; 1_{[T<\infty]}(w)\; P(dw) = E\{\; 1_C(T)\; 1_{[T<\infty]}\;\}.$$

and so, for any bounded measurable process X,

$$\int_{R_+\times\,\Omega} X\; d\mu = E\{X_T\; 1_{[T<\infty]}\}, \tag{4}$$

where we have written X_T in place of X(T). This is most easily seen by first taking X to be a simple process and then passing to the limit. For example,

$X := \sum\alpha_k\; 1_{C_k}$, where $X(s,w) = \alpha_k$ on C_k , (C_k) a finite partition of $B(R+)\times\Omega$.

Then

$$\int X d\mu = \int \sum\alpha_k\; 1_{C_k}\; d\mu = \sum\alpha_k\; \mu(C_k) = E\{\sum\alpha_K\; 1_{C_k}(T)\; 1_{[T<\infty]}\;\} = E\{\; X_T\; 1_{[T<\infty]}\;\}.$$

With $A = 1_{[[T,\infty))}$ as in the beginning of this paragraph, denote the admissible measure μ by μ_T.

The following Theorem establishes a mapping of bounded measurable processes into previsible processes. This mapping behaves much as a conditional expectation operator.

4.2.3. **Theorem**:*(Mètivier,1982)*
For every bounded measurable process, X, there exists a unique, previsible process, ^{p}X*, such that*

$$\int_U X \, d\mu_T = \int_U {}^{p}X \, d\mu_T, \tag{5}$$

where $U = R_+ \times \Omega$*, for every previsible time, T.*

4.3. **Previsible Projections:** The process, ^{p}X, is called the **previsible projection** of X onto G(PT), the F-σ-algebra of previsible events. By equation (4), this defining equation is equivalent to the requirement that

$$E\{ X_T 1_{[T<\infty]}\} = E\{ {}^{p}X_T 1_{[T<\infty]}\} \tag{6}$$

for every previsible stopping time, T. This equation in turn is equivalent to

$$E\{ X_T \, 1_{[T<\infty]} \mid F(T-) \} = {}^{p}X_T \, 1_{[T<\infty]}. \tag{7}$$

a.s.P.

4.3.1. Remark: The proof of this last statement is quite easy. The trick is to take the previsible time, T, to be the restriction, T_C, to any set C in F(T-). By the last result in Section 2.6.5 T_C is previsible and the previous equation for the previsible projection applies. Since T_C takes the value ∞ off the set C, a moment's thought gives this equation in the form

$$E\{ X_T 1_{[T<\infty]} 1_C \} = E\{ {}^{p}X_T 1_{[T<\infty]} 1_C \}.$$

Then using the fact that $^{p}X_T 1_{[T<\infty]}$ is F(T-)-measurable, the result follows from the definition of conditional expectation and the arbitrariness of C in F(T-).

4.4. **Section Theorems:** The proof of the Métivier Theorem itself, however, relies on one of the deeper parts of the general theory of stochastic processes, namely, the so-called Section Theorems. These are the result of applying the Theory of Choquet Capacity and Analytic Sets to measure theory. Of course this theory will not be discussed here, but to establish some frame of reference for the projection theory of this section, we will state one of the Section Theorems and two results that follow from this theorem. To some extent, this will further justify some of the grandiose claims about stopping times made in the introduction to Chapter 2.

4.4.1. **Theorem (Section Theorem)**:
If $(\Omega,H,(F(t),t\geq 0),P)$ *is a filtered probability space and the random set* A *is optional, then given any* $\epsilon > 0$*, there exists a optional time,* T*, such that*

(a) $[[T]]$ *is contained in* A*, and*

(b) $\epsilon + P\{w : T(w) < \infty\} \geq P\{\Pi(A)\}$,

where $A \rightarrow \Pi(A)$ *is the projection map of* $R_+ \times \Omega$ *onto* Ω*. Further, if* A *is previsible, then* T *can be taken to be previsible.*

The following is immediate

4.4.2. **Corollary (Dellacherie, 1972):**
Let X *and* Y *be optional (previsible) processes. Then* X *and* Y *are indistinguishable iff* $X(T) = Y(T)$*, a.s.P, for any optional (previsible) time.*

The proof given by Dellacherie will be paraphrased here because it is simple and indicates why the Section Theorems are important: Let $A = \{ (t,w) : X(t,w) \neq Y(t,w) \}$. Assume that optional X and Y are not indistinguishable, then A is not evanescent. Then there exists an optional time T, whose graph is contained in A and which is not evanescent (by (a) and (b) of the theorem). Hence, $X(T(w),w) \neq Y(T(w),w)$ on an event with positive probability. That is, $X(T) = Y(T)$, a.s.P, implies that X and Y are indistinguishable. Conversely, if X and Y are indistinguishable, then $P(\Pi(A)) = 0$ so that $X(T) = Y(T)$, a.s.P, for all optional times T.

A second application proves the uniqueness statement in Métivier's theorem on the existence of previsible projections:

4.4.3. **Corollary**:
Let X and Y be previsible bounded (or positive) processes. If for each previsible time T one has

$$E\{ X_T 1_{[T<\infty]}\} = E\{ Y_T 1_{[T<\infty]}\}, \tag{8}$$

then the processes X and Y are indistinguishable.

The proof of is similar to that of the last Corollary.

Once the uniqueness of the previsible projection is shown, a monotone class argument centering on processes of the form: $X := 1_{F\times(s,t]}$, with F in F(s) for s<t, is used to show the existence of previsible projections. Previsible projections for such processes will be given below in the Examples subsection.

On the way to proving the existence of previsible projections, Métivier proves that

$$^p(ZX) = Z\,^pX$$

for all bounded previsible processes Z. This gives another important property of previsible projections and one which again suggests that they behave like conditional expectations.

Letting X be a bounded measurable process, we briefly note several **properties of previsible projections**:

(a) If X is a previsible process, then $^pX = X$;

(b) The mapping $X \rightarrow {}^pX$ is linear;

(c) If (X_n) is an increasing sequence of bounded measurable processes, then the previsible projection of the supremum of the sequence is the supremum of the projections;

(d) If X is left continuous, then its previsible projection is left continuous.

4.5. **Optional Projections**: The optional projection, ^oX, of a bounded measurable process X also exist, are unique, and satisfy (see equation (7))

$$E\{ X_T 1_{[T<\infty]} \mid F(T) \} = {}^{o}X_T 1_{[T<\infty]}, \tag{9}$$

a.s.P. (See Dellacherie, 1972 and Dellacherie and Meyer, 1981.) Thus, as with the previsible projection, the optional projection can be written as a conditional expectation, but with the conditioning algebra equal to F(T), rather than F(T-).

The properties listed above for the previsible projections have obvious counterparts in the optional case. The following properties also hold:

(e) ${}^{p}X = {}^{p}({}^{o}X)$.

(f) $X \leq Y$, a.s.P, implies ${}^{o}X \leq {}^{o}Y$, ${}^{p}X \leq {}^{p}Y$.

The last property says that optional and previsible projections are order preserving. The next property says that optional and previsible projections are not very different.

(g) Ω sections, B_ω, of the random set, $B = \{ {}^{o}X \neq {}^{p}X \}$, are countable for all $\omega \epsilon \Omega$. This means that on any path of a process X, its optional and previsible projections differ at only a countable number of time points. In general, **random sets**, C, subsets of $R_+ \times \Omega$, whose sections, $C_\omega = \{ t \geq 0 : (t,\omega) \epsilon C \}$, are countable for each ω are said to be **thin** or "mince" in French literature. Therefore, $B = \{ {}^{o}X \neq {}^{p}X \}$ is a thin random set. In the case of the Poisson process see example (3) immediately following.

(h) An earlier remark, characterizing the previsibility of increasing processes, has the following analogue under optionality: A is an optional increasing process iff for all bounded measurable processes X, $\mu_A(X) = \mu_A({}^{o}X)$.

4.5.1. **Examples**:

(1) X of the form $X = Z\, 1_{((r,s]]}$, where r<s are positive real numbers and Z is a bounded measurable function.

Optional case: Set $Y(t) = E\{ Z \mid F(t) \}$. Y can be and is chosen to be a right continuous modification having left limits. Since Y is adapted it is then optional. Therefore,

$$E\{ X_T 1_{[T<\infty]} \mid F(T) \} = E\{ 1_{((r,s]]}(T) Z \mid F(T) \} = 1_{((r,s]]}(T) Y(T),$$

for any optional time T, where $Y(T) = E(Z|F(T))$ by Doob's Optional Stopping Theorem. Hence, ${}^{o}X = 1_{((r,s]]} Y$.

Previsible case: Take Y as before and let T be a previsible time. Let (T(n)) be a sequence of optional times announcing T. Then $Y(T(n)) = E\{ Z \mid F(T(n)) \}$ and $Y(T-) = \lim_{n\to\infty} Y(T(n)) = E\{ Z \mid \sigma(\bigcup F(T(n))) \}$ a.s.P, (V.T8 Dellacherie, 1972). But $F(T-) = \sigma(\bigcup F(T(n)))$, (III, T39.b Ibid). Therefore, $Y(T-) = E\{ Z \mid F(T-) \}$, for previsible T. Finally, since T and Y(T-) are F(T-)-measurable, it follows as in the last case that ${}^{p}X = 1_{((r,s]]}Y_{-}$.

(2) Let S be a totally inaccessible time and set $X = 1_{[[S]]}$. Then $X(T(w),w) = 1_{[[S]]}(T(w),w) = 1_{[T=S]}(w)$ a.s.P, for any previsible time, T. Therefore, $E(X_T 1_{[T<\infty]}) = E\, 1_{[T=S<\infty]}$. The latter quantity equals zero a.s.P, by definition of total inaccessibility. Hence, by the first Corollary to the Section Theorem, we then have that ${}^{p}X$ is evanescent. Setting all the details aside, this should be intuitively clear from the definition of total inaccessibility and any reasonable interpretation of projection.

(3) Let X be a Poisson process with parameter $c>0$, so that X is optional (it is right continuous) and, consequently, ${}^{o}X = X$. Then ${}^{p}X = X_{-}$, where $X_{-}(t) = X(t-)$, $t \geq 0$. This example can be used to illustrate the idea of thin random sets defined in the previous section. Notice that the sections B_w of $B = \{ {}^{o}X \neq {}^{p}X \}$ are just $B_w = \{ T_n(w) : n \geq 0 \}$, where (T_n) is the sequence of jump times of the process X.

4.6. **Dual Previsible Projections**: Consider the measure, μ_A, defined earlier in this Chapter by setting $\mu_A(X) = E\{(X.A)(\infty)\}$ for all bounded (or positive) measurable functions X, where A was an increasing process. As noted, μ_A is called the measure "generated" by the process A. Let X be any positive measurable process, define another measure $m(X) := E\{({}^{p}X.A)(\infty) \}$ and ask if there is a nondecreasing processes $\tilde{A}$ with the property that $m(X) = \mu_{\tilde{A}}(X)$. This question is the same as interpreting $\mu_A(X)$ as an ordered scalar product $<X,A>$ and asking about the dual, $\tilde{A}$, of ${}^{p}X$, in the sense that $< {}^{p}X,A> = <X,\tilde{A}>$.

We will drop the subscript A on μ_A for a while, but retain the above definition. Dellacherie shows that if μ^p is defined by setting $\mu^p(X) := \mu ({}^{p} X)$ for every positive measurable process X, then μ^p is a σ-finite measure on $B(R_+) \times F(\infty)$. The corresponding unique (up to indistinguishability) increasing process generating this measure is denoted by A^p and called the **dual previsible projection** of the process A. The measure μ^p, is referred to as the **dual previsible projection** of the measure μ. (E.g., Métivier, and Hoeven.)

4.6.1. Remark: Assume the usual conditions on the underlying filtered probability space (Ω,H,(F(t)),P). Dellacherie[1972, IV T41, T42] gives the following characterization of measures generated by increasing processes:

4.6.2. **Theorem:**
A σ-finite *measure* μ *on* $(R_+ \times \Omega, B(R_+) \times H$ *is generated by an integrable, increasing (not necessarily adapted) process, A, iff*

(a) $\mu([[0]]) = 0$ *and* $\mu([[0,t]]) < \infty$, $t \epsilon R_+$,

(b) μ *is P-admissible.*

Then A is unique up to P-indistinguishability.

Further, A is adapted iff

(c) $\mu([0,t] \times B) = \mu(E(1_B \mid F_t)\, 1_{[[0,t]]})$

for all $t \epsilon R_+$ *and* $B \epsilon H$.

4.6.3. Remarks: Recall that to avoid some complications in exposition we have assumed as part of the definition of increasing process in Chapter 3 that A(0)=0. This is Dellacherie's assumption also, but Jacod [1979] does not make this assumption here, nor do Dellacherie and Meyer [1980]. These latter works also do not assume that $\mu([[0]]) = 0$, but that the measure has finite mass. As defined in the beginning of this Chapter, condition (b) just says that μ assigns zero measure to evanescent random sets.

4.6.4. Remarks: The reader will notice that as we come to the end of this note, more proofs will be given. This is especially true in Chapter 6. For a number of reasons, we choose to prove the present Theorem:

Suppose then that μ is generated by an increasing process as specified in the statement of the theorem. We first note that the finiteness of μ in (a) follows from the integrability of A. The admissibility of μ is obvious, since P(B)=0 implies that $I \times B$, where I is an interval, is evanescent. Therefore, all that remains of the necessity portion of the proof is condition (c). We observe first that A is F-adapted iff for all $B \epsilon H$

$$E 1_B A_t = E(E(1_B A_t \mid F_t)) = E(E(1_B \mid F_t) A_t)$$

for all $t \epsilon R_+$. This is, iff A is orthogonal to all r.v.s of the form $1_B - E(1_B \mid F_t)$. But the last equation is just condition (c), since when $\mu = \mu_A$ is generated by A and A is a adapted

$$\begin{aligned}
\mu_A([0,t] \times B) &= E(\int_0^\infty 1_{[[0,t]]}(s,.) 1_B(s)\, dA_s \\
&= E(1_B A_t) = E(E(1_B \mid F_t) A_t) \\
&= E(\int_0^\infty E(1_B \mid F_t)\, 1_{[[0,t]]}\, dA_s) \\
&= \mu_A(E(1_B \mid F_t)\, 1_{[[0,t]]}).
\end{aligned}$$

Conversely, if the three conditions are satisfied, then for all $B \epsilon H$ define

$$Q_t(B) := \mu([0,t] \times B).$$

Then $Q_0(B) = 0$ for all B and Q_t is a bounded measure for all $t \geq 0$. Admissibility of μ shows that Q is absolutely continuous with respect to P on (Ω, H). Let A' be defined by setting $A_t' = \dfrac{dQ_t}{dP}$, a.s.P, the Radon-Nikodym derivative of Q_t relative to P. Then $A_0' = 0$, a.s.P, and A_t' is P-integrable. Since μ is a positive measure, $A_s' \leq A_t'$, if $s \leq t$.

By Lebesgue's Monotone Convergence Theorem, $A_t' = \lim_{n \to \infty} A_{t_n}'$ in $L_1(P)$, where (t_n) is any seqence decreasing to t. It follows that the convergence is also almost sure (P). Hence, we can define the process A as a right continuous, increasing, modification of the process A' by setting $A_t := \inf\{A_r' : r \text{ rational and } > t\}$

With this A, the measure generated by A satisfies

$$\mu_A([0,t] \times B) = E\int_0^t 1_B\, dA_s = E 1_B A_t.$$

But since A is a modification of A', A_t is, a.s.P, the Radon-Nikodym derivative of Q_t relative to P. Hence,

$$E 1_B A_t = \int_0^t 1_B \, dQ_s = \mu([0,t] \times B).$$

Therefore,

$$\mu_A([0,t] \times B) = \mu([0,t] \times B)$$

for all $t \epsilon R_+$ and $B \epsilon H$. It follows that $\mu_A = \mu$, since the sets of the form $[0,t] \times B$ are generators for the product σ-algebra, $B(R_+) \times H$.

Finally, we have already verified that A is adapted by condition (c). Uniqueness follows by noting that if G is another increasing process generating μ, then $G_t = A_t$, a.s.P, for each t, and G is a modification of A. Since G and A are right continuous, Lemma 2.3.3 guarantees that they are indistinguishable.

4.6.5. Remark: Let X be any positive, measurable process and set

$$\mu(X) = E({}^pX.A_\infty),$$

where A is a, not necessarily adapted, integrable, increasing process. Because of the properties of linearity, monotonicity and continuity of previsible projections, μ is a σ-finite measure on $B(R_+) \times H$. Therefore, by the last theorem, there exists a unique increasing process, denoted A^P, which generates μ. Hence, from the last equation

$$EX.A^p_\infty = E({}^pX.A_\infty).$$

We need the following lemma to conclude that the process A^p is previsible:

4.6.6. **Lemma: (Dellacherie[1972, V T26])** *An integrable, increasing process, A, is previsible iff for any two positive, measurable processes X and Y with the same previsible projection* $\mu_A(X) = \mu_A(Y)$

But if X and Y have the same previsible projections then from the preceding construction

$$EX.A^p_\infty = E({}^pX.A_\infty) = E({}^pY.A_\infty) = EY.A^p_\infty.$$

Hence, the Lemma shows that A^p is a previsible process. Therefore, we have the following

4.6.7. **Theorem**:(Dellacherie,1972,p107)
Let A be an integrable, increasing, not necessarily adapted, process with $A(0) = 0$. *For each positive measurable process, X, there exists a unique*

previsible, increasing process, A^p, *called the* **dual previsible projection** *of A, such that*

$$E\{ \int_0^\infty {}^pX_s \, dA_s \} = E\{ \int_0^\infty X_s \, dA_s^p \}. \tag{10}$$

4.6.8. Remark: Brémaud (1981) and Meyer (1973) state this result in a slightly different form which will be useful later on:

Let A be an integrable, increasing process with A(0) = 0. Then there exists a unique (to indistinguishability) an integrable, previsible, increasing process, A^p, *such that* $A^p(0) = 0$ *a.s.P and*

$$E\{ \int_0^\infty C(s) d\, A^p(s) \} = E\{ \int_0^\infty C(s) dA(s) \}$$

for all non-negative, previsible processes, $(C(s), s \geq 0)$.

As indicated in this result (with $C \equiv 1$), the duals of integrable processes are themselves integrable. However, it may be shown that the dual projections of increasing bounded processes are not necessarily bounded.

The following strengthens the definition of the dual previsible projection:

4.6.9. **Theorem**:
Let S, T be F-stopping times, with $S \leq T$, *and A an integrable, increasing process. Then*

$$E\{ \int_S^T {}^pX(t) \, d\, A(t) \mid F(S) \} = E\{ \int_S^T X(t) \, d\, A^p(t) \mid F(S) \} \tag{11}$$

for any bounded (or positive), measurable process X.

4.6.10. Remark: The proof follows from the definition of conditional expectation and the definition of dual previsible projection. Let $C \epsilon F(S)$, and set S_C and T_C equal to the restrictions of S and T to the event C. Then we know that the stochastic interval, $((S_C, T_C]]$, is previsible. Hence, from the properties of previsible projections, we have

$$1_{((S_C,T_C]]}\ {}^{p}X = {}^{p}(\ 1_{((S_C,T_C]]}\ X\ ),$$

and from the definition of A^p, the dual previsible projection of A, we have

$$\mu_A(\ {}^{p}(\ 1_{((S_C,T_C]]}\ X\ )\ ) = \mu_{A^p}(\ 1_{((S_C,T_C]]}\ X\ ).$$

Therefore,

$$\mu_A(\ 1_{((S_C,T_C]]}\ {}^{p}X\ ) = \mu_{A^p}(\ 1_{((S_C,T_C]]}\ X\ )\ ).$$

This last equation is the same as

$$E\{\ 1_C \int_S^T {}^{p}X\, dA\ \} = E\{\ 1_C \int_S^T X\, dA^p\ \}.$$

Since $C \epsilon F(S)$ is arbitrary, the last equation is equivalent to the statement of the theorem.

4.6.11. **Definition**: Two raw increasing processes, A and B, having the same dual previsible projection are said to be **associated**. If A and B are associated, then we write $A\ \rho\ B$.

4.6.12. Remark: Dellacherie shows that each equivalence class determined by the relation ρ contains one and only one previsible increasing process.

4.6.13. Remark: We now set down some results whose main object is to characterize adapted associated processes. This characterization will be extended by "localization" in Chapter 6.

4.6.14. **Theorem:**
Let $A \epsilon IV_0$. *Then the following statements are equivalent:*

(a) A *is a martingale;*

(b) A^p *is evanescent;*

(c) μ_A *vanishes on previsible random sets.*

Remark: There is very little to prove. First consider the equivalence of (a) and (c): Let $s \leq t$ and $B \epsilon F_s$. Then it is easy to see that $\mu_A(\ ((s_B,t_B]]\ ) = E(1_B(A_t - A_s))$. Recalling the generators of G(PT), it is clear that this equation entails the equivalence of (a) and (c). Let S and T be stopping times, $S \leq T$; then $((S,T]]$ is previsible. The equivalence of (b) and (c) follows in a manner similar to the last case by noting that $\mu_A(\,((S,T]]\,) = \mu_{A^p}(\,((S,T]]\,)$. (We have left off some details in the two pairs of equivalences concerning the generators $\{0\} \times B$ and $[[0_B]]$, $B \epsilon F_0$, respectively, but these are easy.)

The following Corollary is the desired characterization:

4.6.15. **Corollary**:
Increasing, integrable processes A and B are associated iff the process M = A - B is a martingale.

Although the implication (b) implies (a) gives the necessity of this Corollary, it is instructive to use the previous Theorem 4.6.9. The necessity of the Corollary follows from the previous Theorem by setting X equal to 1 and using the fact that A and B have the same dual previsible projections. This yields

$$E(\,A(t) - A(s) \mid F(s)\,) = E(\,B(t) - B(s) \mid F(s)\,)$$

for all real numbers s and t, with $s \leq t$. Since A and B are adapted (part of the definition of increasing process), it follows that A - B is a martingale.

Conversely, $A - B \epsilon IV_0$ and $A - B$ is a martingale. The last Theorem tells us that $0 = (A - B)^p$, so that linearity gives $A^p = B^p$. Therefore, $A \rho B$.

4.6.16. **Definition**: Let A be an integrable, increasing process. (Hence A is adapted.) The dual previsible projection of A is called the **(previsible) compensator** of A, and is denoted by $\tilde{A}$.

4.6.17. Remark: The previsible compensator of A is that previsible process that must be subtracted from A to obtain a martingale.

4.6.18. Remark: In the Chapter on martingale transforms, we noticed that in discrete time, if an increasing previsible process was a martingale, then it was a.s.P constant (and equal to zero if it took the value zero at the origin). A similar remark can be made for the continuous time analogue. Again this follows immediately from the last Theorem.

A direct proof repeats part of the proof of Theorem 4.6.14, perhaps more carefully. The argument goes as follows: If A is an integrable, increasing process which is also a martingale, then $E(1_D(A(t) - A(s)) \mid F(s)) = 0$, for all D in F(s). But this says that the measure generated by A vanishes on events of the form $1_{(s,t]\times D}$, $s \leq t$ and D in F(s). It is also obvious that this measure vanishes on $\{0\} \times D$, where $D \epsilon F(0)$. Since these events are generators of the σ-algebra of F-previsible events, the measure generated by A vanishes on the entire F-previsible algebra. Therefore, $\mu_{A^p}(X) = \mu_A({}^pX) = 0$, for all bounded, measurable X. It follows that A^p is evanescent. But if A is previsible, then $A = A^p$, and so A is evanescent. Therefore,

4.6.19. **Theorem**: *Integrable, increasing, previsible martingales are evanescent.*

4.6.20. Remark: This will be extended to local martingales in Chapter 6.

4.6.21. In the section on Lebesgue-Stieltjes stochastic integrals we have noted that if $X \epsilon L_1(A)$ is positive, then X.A is an increasing process. It is natural to consider the dual previsible projection of X.A when either X or A is previsible. Dellacherie, V T31, 1972, shows that

(1) If A is an increasing previsible process, then $(X.A)^p = {}^pX.A$;

(2) If X is a positive, previsible, $L_1(A)$ *process, then* $(X.A)^p = X.A^p$.

4.6.22. Remark: We will prove the second proposition using the ordered scalar product notation introduced above: Let Y be a positive, measurable process, then

$$<Y,(X.A)^p> = <Y^p,(X.A)> = <Y^pX,A> =$$
$$<(YX)^p,A> = <YX,A^p> = <Y,X.A^p>.$$

4.6.23. Remark: Dellacherie, 1972, discusses the notion of absolute continuity of increasing processes. This is of some importance in the analysis of counting processes. Let A and B be (raw) increasing processes. A is said to be **absolutely continuous** relative to B if $Y.B = 0$ implies $Y.A = 0$ for all positive measurable processes, Y. If μ and λ are the measures generated by B and A, respectively, then this is the same as saying that λ is absolutely continuous relative to μ. Thus, if f is the Radon-Nikodym density of λ relative to μ, then, using the notation from the beginning of this Chapter,

$$\int_U X\, f\, d\mu = \int_U X\, d\lambda$$

or, equivalently,

$$E(Xf.B) = E(X.(f.B)) = E(X.A)$$

for all bounded measurable processes X. The unicity (up to indistinguishability) of the generating processes then implies that $A = f.B$, where by definition, f is a positive measurable process in $L_1(B)$. When A and B are both previsible, $A = A^p = (f.B)^p = {}^pf.B$. Therefore, if all the assumptions of this paragraph hold and A and B are previsible, then there exists a previsible process $g \epsilon L_1(B)$ such that $A = g.B$. That is, f is previsible.

4.6.24. Remark: In nonlinear filtering of point processes, N, an important class of problems is covered by the case where A, the dual previsible projection of N, is absolutely continuous relative to the deterministic process, B(t,w) = t, a.s.P. In this case, f is called the **intensity** of the point process N. More precisely, recalling that we have suppressed the underlying filtration, $(F(t),t\geq 0)$, and recognizing that the dual previsible projection depends strongly on its filtration, and the underlying probability, P, f is called the **F-intensity**, or the **(P,F)-intensity** of N. The intensity is, in general, a previsible stochastic process.

4.6.25. Remark: Now let's cover the last paragraph again from a different starting point. We return to Theorem 4.6.9., with $X \equiv 1$, and take B to be the dual previsible projection of A. If it is further assumed that B is absolutely continuous relative to Lebesgue measure, with density λ, then as defined earlier, λ is the F-intensity of A and satisfies

$$E(\, A(t) - A(s)\,) \mid F(s)\,) = E(\, \int_s^t \lambda(y)\, dy \mid F(s)\,).$$

This equation becomes extremely important when A is a counting process. Brémaud, 1981, is concerned almost exclusively with this case and the sections below on nonlinear filtering will deal mostly with this case, following Brémaud, [1978,79,80,81]. For now we just consider the simple example when A is a one jump counting process: $A(t) = 1_{[T\leq t]}$, with T an F-stopping time. With this definition of A the last equation becomes

$$P(\, s < T \leq t \mid F(s)\,) = E(\, \int_s^t \lambda(y)\, dy \mid F(s)\,).$$

Using the "little o" notation, this statement has historically been written somewhat less exactly as

$$P(\, s < T \leq t \mid F(s)\,) = \lambda(s)(t-s) + o(t-s), \quad (t \to s+).$$

To justify this statement in a simple case, assume that λ is right continuous and use the identity

$$\int_s^t \lambda(y)\,dy = \lambda(s)(t-s) + (t-s)\left\{ \frac{1}{(t-s)} \int_s^t (\lambda(y) - \lambda(s))\,dy \right\}.$$

4.7. **Random Measures and Jacod's Formula:** In this section we will introduce a useful formula for calculating the dual previsible projection of a point process. The derivation of this formula in its most general form is due to Jacod in his 1975 paper on multivariate (marked) point process. (The origin of the formula is contained in the paper of Delacherie [1970] which considers a point process with a single jump. Also see Brown [1978] for a short proof of the formula in the case of simple point processes.) Although the results of Jacod's paper are extremely important and go far beyond just the formula and, as pointed out by Jacod, reading the paper does not require an enormous technical background we will not attempt to give a digest of its contents. Our goal is just to introduce Jacod's "hazard" function formula for the compensator of a marked point process, to do this without proofs, but with enough preliminary explanation to allow one to understand why the formula holds. To accomplish this we will first show how to develop a discrete parameter version of the Jacod formula in the case of a simple (unmarked) point processes. Then we will suggest its continuous time analogue, recall (and extend) the concept of a random measure from Chapter 3 and state the general Jacod formula together with some useful special cases. Examples of the use of the formula are given in Chapter 5.

Recall the discussion and notation of Section 1.10 on discrete point processes, $(N_n, F_n, n \epsilon Z_+)$. So $N_n = \sum_{k=0}^{n} X_k$ with the X_k being 0–1 Bernoulli random variables. As in Chapter 1, let $\lambda = (\lambda_k)$ be the F-intensity of the point process (X_k). Thus,

$$\lambda_k = E(X_k \mid F_{k-1}) = P(X_k = 1 \mid F_{k-1}).$$

for $k \geq 1$. $X_0 \equiv \lambda_0 = 0$.

Paralleling the assumptions of Jacod we take the filter F to be the internal history of the discrete point process X. Recall the definition of the "jump" times (T_n) of the counting process N. Then

$$N_m = \sum_{n \geq 1} 1_{[T_n \leq m]}.$$

It is clear that this is a finite sum since the stopping times (T_n) are integer valued. But we will continue the practice of writing it as an infinite sum.

We prove the following formula for the intensity of the discrete point process. (Convention: If α and β are two functions and $\beta(w) = 0$ implies that $\alpha(w) = 0$, then it is natural to define the quotient $\frac{\alpha}{\beta}$ as zero whenever β vanishes.)

4.7.1. **Theorem**:

Under the assumptions stated above

$$\lambda_k = \sum_{n \geq 1} \frac{P(T_n = k \mid F_{T_{n-1}})}{P(T_n \geq k \mid F_{T_{n-1}})} 1_{[[T_{n-1}, T_n))}, \tag{12}$$

where $F_{T_j} = \sigma(T_1, \cdots, T_j)$.

First note that

$$X_k = \Delta N_k = \sum_{n \geq 1} 1_{[T_n = k]},$$

so that

$$\lambda_k = E(X_k \mid F_{k-1}) = \sum_{n \geq 1} P(T_n = k \mid F_{k-1}).$$

The following relation holds for the trace σ-algebra on $[T_{n-1} \leq k\text{-}1 < T_n]$:

$$F_{k-1} \cap [T_{n-1} \leq k\text{-}1 < T_n] = F_{T_{n-1}} \cap [T_{n-1} \leq k\text{-}1 < T_n]. \tag{13}$$

Observe that $[T_{n-1} \leq k\text{-}1 < T_n] \epsilon F_{k-1}$ since the T's are F-stopping times and

$$[T_{n-1} \leq k\text{-}1 < T_n] = [T_{n-1} \leq k\text{-}1 < k \leq T_n],$$

since the stopping times are integer valued.

It follows from the last equality that

$$[T_n = k] = [X_k = 1] \cap [T_{n-1} \leq k\text{-}1 < k \leq T_n]. \tag{14}$$

We need the following variation of Bayes Theorem:

4.7.2. Lemma (Brown [1978]):
Let (Ω, H, P) *be a probability space. If* G *and* K *are sub* σ-algebras *of* H, $B \epsilon H$, $C \epsilon G$ *and*

$$G \cap C = K \cap C.$$

Then

$$P(B \cap C \mid G) = \frac{P(B \cap C \mid K)}{P(C \mid K)}$$

on C and = 0, on the complement of C, where $P(C \mid K)$ *is a version with* $P(C \mid K) \neq 0$ *on* C.

Remark: Brown does not give a proof of this Lemma, but it follows easily from the "quotient rule" for Radon-Nykodym derivatives by paying careful attention to the use of the restrictions of P to the various sub σ-algebras involved in the hypotheses.

Using this Lemma, which applies due to (13) through (14), we have

$$P(T_n = k \mid F_{k-1}) = \frac{P(T_n = k \mid F_{T_{n-1}})}{P(T_n \geq k \mid F_{T_{n-1}})} 1_{[[T_{n-1}, T_n))}.$$

and consequently formula (12).

Having obtained formula (12) for the "first difference" of the compensator of a discrete point process, it is natural of conjecture that the compensator A of a point process $N = (N_t, F_t, t \geq 0)$, where $F_t = \sigma(N_s, s \leq t)$ and $N(t) = \sum_{n \geq 1} 1_{[T_n \leq t]}$, satisfies

$$A(dt) = \sum_{n \geq 1} \frac{P(T_n \epsilon dt \mid F_{T_{n-1}})}{P(T_n \geq t \mid F_{T_{n-1}})} 1_{((T_{n-1}, T_n]]}. \tag{15}$$

This equation does indeed hold and occurs in various forms (e.g., Brown [1978], Liptser-Shiryayev [1978]) and in numerous applications (e.g., Jacobson [1982], Gill [1980]). We will come back to this case at the end of the present Section where it will occur as a consequence of the general Jacod formula. For this purpose we need to recall the concept of marked point processes and random measures which were mentioned briefly in Chapter 3.

4.7.3. Remark: Let $(T_n, n\epsilon Z_+)$ be a point process relative to the probability space (Ω,H,P): (T_n) is an increasing sequence of random variables on Ω with values in $\overline{R}_+$ and such that $T_n < T_{n+1}$ on $[T_n < \infty]$. Set $N_t = \sum_{n\geq} 1_{[T_n \leq t]}$.

Let (E,ξ) be a measurable space and $(Z_n, n\epsilon Z_+)$ a sequence of random variables on Ω with values in a space E. Hence, the Z_n are H-measurable relative to ξ. (Z_n) is called the sequence of marks and E the mark space. E is assumed to have the structure of a Borel subset of a complete metric space. (This assumption is sufficient for the existence of regular conditional distributions of random variables with values in E. (Shiryayev [1984].) The usual applications will have $E = R^n$, for some natural number n, or $E = R^{Z_+}$.)

The definition of the range of the sequence of marks is extended in the following way: Let ς be some point exterior to E and define $Z_n(w) = \varsigma$ iff $T_n(w) = \infty$ (the nth event "never occurs"). To understand why this extension is made see Jacod[1979, p.74]. Finally, let $Z_0(e) = \delta$ for all e in E and $T_0 = 0$, $S_{n+1} = T_{n+1} - T_n$, for $n\geq 0$. The sequence (T_n, Z_n) is called a **marked point process**.

Let

$$N_t^A = \sum_{n\geq 1} 1_{[T_n \leq t]} 1_{[Z_n \epsilon A]}.$$

Then N^A counts the number of times jumps of N have marks in A.

Set $E_\varsigma = E\bigcup\{\varsigma\}$, $\overline{E} = (0,\infty)\times E$, $\overline{E}_\varsigma = \overline{E}\bigcup\{\infty,\varsigma\}$ and $\overline{\Omega} = \Omega\times[0,\infty)\times E$. With an analogous meaning, let ξ_ς, $\overline{\xi}$, $\overline{\xi}_\varsigma$ be the usual σ-algebras on E_ς, $\overline{E}$, $\overline{E}_\varsigma$, respectively. (E.g., $\overline{\xi} = B((0,\infty))\times\xi$.)

Let (Ω,H,F,P) be a filtered probability space with the filtration $F = (F_t, t\geq 0)$. Then if $\Pi = \Pi(F)$ denotes the σ-algebra of F-previsible subsets of $\Omega\times[0,\infty)$, set $\overline{\Pi} := \overline{\Pi}(F) = \Pi\times\xi$, where $F = (F_t, t\geq 0)$.

With this structure, we call the family $\mu = \{\mu(w,.): w\epsilon\Omega\}$ of nonnegative functions a **random measure** on $(\overline{E},\overline{\xi})$ if μ is a positive transition measure (Appendix A). That is, if

(a) $w\rightarrow\mu(w,A)$ is H-measurable for each $A\epsilon\overline{\xi}$, and

(b) $A\rightarrow\mu(w,A)$ is a positive σ finite measure for each $w\epsilon\Omega$.

Further, a random measure μ is said to be an **integer valued measure** if

(c) the mapping $w \rightarrow \mu(w,A) \epsilon \overline{Z}_+$, for each $A \epsilon \overline{\xi}$, and

(d) $\mu(w,\{t\} \times E) \leq 1$, for all $w \epsilon \Omega$ and $t>0$.

Let μ be a random measure on $(\overline{E},\overline{\xi})$. If $W = (W(w,t,z), w\epsilon\Omega, t\geq 0, z\epsilon E)$ is a nonnegative $H \times B((0,\infty)) \times \xi$ measurable function on $\overline{\Omega}$, set

$$W*\mu_t(w) \ := \int_{(0,t]\times E} W(w,s,z)\mu(w,ds,dz).$$

By denoting the Radon-Nykodym derivative of W(w, .) relative to the measure $\mu(w,.)$ by $Wo\mu$, we can write the last definition in the form

$$W*\mu_t(w) \ := \ (Wo\mu)(w,(0,t]\times E).$$

A random measure η is said to be F-**previsible** if for each positive $\overline{\Pi}$-measurable process X, the process $X*\eta_t$ is F-previsible.

The marked point process $(T_n, Z_n, n\epsilon Z_+)$ is completely determinded by the random measure μ defined on $(\overline{E},\overline{\xi})$ by setting

$$\mu(w,B) \ = \ \sum_{n\geq 1} 1_B(T_n(w),Z_n(w))1_{[T_n<\infty]} \tag{16}$$

for all $B\epsilon\overline{\xi}$. We will often refer to such a measure as a **point process measure** or the **random measure of a point process**.

It will be convenient to also write μ in the form

$$\mu(w,dt,dz) \ = \ \sum_{n\geq 1} \epsilon_{(T_n(w),Z_n(w))}(dt,dz) \ 1_{[T_n<\infty]}, \tag{17}$$

where ϵ_a is the Dirac measure (unit mass concentrated) at the point a.

Following Jacod [1975], to each probability measure P on (Ω,H) and point process random measure μ we associate a nonnegative measure M_μ on $(\overline{\Omega},\overline{\Pi})$ defined by setting

$$M_\mu(W) \ := \ E((W*\mu)_\infty)$$

for any nonnegative $\overline{\Pi}$-measurable function W on $\overline{\Omega}$. Jacod then proves

4.7.4. **Lemma:**
If η is a random measure such that M_η is σ-finite, then there exists a unique (up to a P-null set) F-previsible random measure η' such that for each positive $X\epsilon\overline{\Pi}$,

$$M_\eta(X) = M_{\eta'}(X).$$

Remark: Comparing this result to Theorem 4.6.7., it is clearly appropriate to call the random measure η' of this Lemma the **dual previsible projection** of the measure η. In a moment, we will point out the more compelling reason that $\eta - \eta'$ is, in a natural sense, a martingale.

In order to apply this Lemma to the random measure μ of a marked point process (refer of equation (16)), Jacod shows that M_μ is σ-finite on $(\overline{\Omega},\overline{\Pi})$. Then he obtains

4.7.5. **Theorem:**
If (T_n,Z_n) *is a marked point process and* μ *is given by (17), then there exists a unique (up to modification on P-null events)* F*-previsible random measure* ν *such that for each positive previsible process* X $(X\epsilon\overline{\Pi})$,

$$E\int_{(0,\infty)\times E} X(t,z)\,\mu(dt,dz) = E\int_{(0,\infty)\times E} X(t,z)\,\nu(dt,dz). \tag{18}$$

Jacod then uses (18) and one of the Section Theorems to show that the dual previsible projection ν of μ in (18) can be chosen so that

$$\nu(\{t\}\times E) \leq 1 \tag{19}$$

and

$$\nu([T_\infty,\infty)\times E) = 0. \tag{20}$$

Remark: Set $A_t^C(w) := \nu(w,(0,t]\times C)$ and $A_t := A_t^E$. Then A is the compensator of $N_t^E = N_t = \sum_{n\geq 1} 1_{[T_n\leq t]}$ and inequality (19) says that the jumps of A have magnitude not exceeding one: $0\leq\Delta A_t\leq 1$. Equation (20) says that A does not charge the random set $[T_\infty,\infty)$.

In order to emphasize the connection between this and earlier results of the Chapter (when E is a singleton set), we note that the dual previsible projection of the marked point process measure μ is characterized by (19), and (20) together with the requirements that

(i) the process $(\nu((0,t]\times B),t\geq 0)$ is previsible for each $B\epsilon\xi$, and

(ii) $m_t^{(n)} := \mu((0,t\wedge T_n]\times B) - \nu((0,t\wedge T_n]\times B)$, defines a uniformly integrable process $m^{(n)} = (m_t^{(n)}, t \geq 0)$, for each $n > 0$ and $B \epsilon \xi$.

4.7.6. Remark: Set $G_t = \sigma(N_s^A, s \leq t, A \epsilon \xi)$. It can be shown (e.g., Itim [1980], Brémaud [1981]) that The filtration G is continuous on the right and $G_S = \sigma(N_{S \wedge t}^A, t \geq 0, A \epsilon \xi)$, where S is a G-stopping time and from this that for $n \geq 1$,

$$G_{T_n} = \sigma(T_k, Z_k, 1 \leq k \leq n) \tag{21}$$

and

$$G_{T_{n}-} = \sigma(T_k, Z_k, T_n, 1 \leq k \leq n-1).$$

Now if we take into account the probability measure P, and define F_t, $t \geq 0$, to be the smallest σ-algebra generated by the union of the family, Γ, of all P-null sets in H and G_t, then the family F retains the right continuity of G. Therefore, since

$$F_t = \sigma(\Gamma \bigcup G_t) \tag{22}$$

is complete, it is a filtration satisfying the "usual conditions".

For each $n \epsilon Z_+$, let $K_n(w,dt,dz)$ be a version of the regular conditional distribution of (S_n, Z_n) given $F_{T_{n-1}}$ and $H_n(w,dt) = K_n(w,dt,E_\varsigma)$, the conditional distribution of S_n. $K_n(w,.)$ is a probability on $\overline{E}_\varsigma$ while $H_n(w,.)$ is a probability on $(0,\infty]$.

We can now state the Jacod formula:

4.7.7. Theorem:
With the filtration $F = (F_t, t \geq 0)$ given by equation (22), the dual previsible projection ν of the random measure μ of (17) satisfies

$$\nu(dt,dz) = \sum_{n \geq 1} \frac{K_n(dt-T_{n-1},dz)}{H_n([t-T_{n-1},\infty])} 1_{[T_{n-1}<t \leq T_n]}. \tag{23}$$

4.7.8. Remark: Several examples of marked point processes were given in Chapter 3. But for the purposes of this section an informative example to keep in mind is that of a jump process. A **jump process**, $X = (X_n, t \geq 0)$, is a Skorokhod process all of whose paths are step functions (with only a finite number of jumps in any bounded interval of time, Appendix A). If we let the sequence $(T_n, n \geq 1)$ denote the sequence of jump times of such a processes X and $(Z_n, n \geq 1)$ the sequence of jump sizes of X at these jump times, $Z_n := \Delta X_n$, then with the proper conventions at time 0, (T_n, Z_n) is a marked point process and

$$X_t(w) = X_0(w) + \sum_{n\geq 1} Z_n(w)\, 1_{[[T_n,\infty))}(w,t). \qquad (24)$$

In this case the random measure μ in (17) is called the **saltus measure** or **jump measure** of the process X.

In this case, with the filtration as in (22), the Theorem of Jacod shows us that the dual previsible projection ν of the saltus measure of X can be written in the form

$$\nu(dt,dz) = \sum_{n\geq 1} \frac{P(T_n \epsilon dt, Z_n \epsilon dz \mid F_{T_{n-1}})}{P(T_n \geq t \mid F_{T_{n-1}})} 1_{((T_{n-1},T_n]]}. \qquad (25)$$

We have expressed the conditional laws of (23) in terms of the process (T_n) instead of the inter-occurrence times (S_n) so that we could make a direct comparison with the discrete point process case given in (12). As one can see, (25) could have be conjectured from (12).

4.7.9. Remark: From the standpoint of the creators of the General Theory of Stochastic Processes and existing literature, one would deduce (12) (similar remarks would hold for its marked analogue) from (23). To see how this can be accomplished, we will use the notation for discrete point processes given at the beginning of this Section and let [ ] denote the "greatest integer" function. Since it does not make the problem more difficult, we will assume in our discrete parameter case that there is a sequence of marks, (Z_n). Thus, we start with the integer valued times (T_n), the sequence of marks (Z_n) and the filtration $F = (F_n, n\geq 0)$. The filtration will be the one defined by

$$F_{T_n} = \sigma(T_k, Z_k, 1\leq k\leq n, \Gamma) \qquad (26)$$

and

$$F_{T_n-} = \sigma(T_k, Z_k, T_n, 1\leq k\leq n-1, \Gamma).$$

Once we define the continuous time filtration, this is enough information to construct the random measure μ and its dual previsible projection ν, as well as the continuous parameter point processes, $N^B = (N_t^B, t\geq 0)$ and $N := N^E$. The continuous parameter filtration $\bar{F}$ can be defined by setting $\bar{F}_t := F_{[t]}$. Since the times are integers, this gives in particular that $\bar{F}_{T_n} = F_{T_n}$.

Thus all the main features of an induced continuous parameter marked point process have been defined. For example, the dual previsible projection of N is $A_t := \nu((0,t]\times E)$. To recover (12) from (23), just define the F-intensity by $\lambda_{[t]} := \nu(\{[t]\}\times E)$ and the result follows. By Jacod's Theorem then, $0\leq \lambda_k \leq 1$.

Certainly this is the proper route to (12). But from the standpoint of building an intuition and gaining the interest of practitioners from other fields the discrete parameter approach has some worth.

4.7.10. Remark: We will close this Section and the Chapter by stating oft used forms of Jacod's formula for unmarked point processes. Assume that the filtration is given by (22) with $E = \{1\}$, so that G (in (22)) takes the obvious form. If the point process $N_t = \sum_{n \geq 1} 1_{[T_n \leq t]}$ has dual previsible projection A (with $A_0 \equiv 0$), then using the notation of Jacod's Theorem (23) we have $A_t = \nu((0,t])$, so that equation (15) holds.

Also, in the unmarked case, $H_n = K_n$, so that (assuming the point process is non-explosive) we can write the compensator in terms of the conditional inter-occurrence time distributions by integrating equation (23) over $(T_{n-1}(w),t]$, for $T_{n-1} < t \leq T_n(w)$ and then making a change of variable to obtain:

$$A_t(w) = A_{T_{n-1}}(w) + \int_0^{t - T_{n-1}(w)} dK_n(w,s)/(1-K_n(w,s-)), \tag{27}$$

when $T_{n-1}(w) < t \leq T_n(w)$, $n \geq 1$.

Here is a particular example of (27): Let $K_n(y) = 1 - e^{-\lambda_n y}$, for $y > 0$, zero otherwise; let $E = \{1\}$. Then, from (27),

$$A_t = A_{T_{n-1}} + \lambda_n(t - T_{n-1}),$$

when $(t,w) \epsilon ((T_{n-1},T_n]]$. This yields the interesting relationship

$$A_{T_n} - A_{T_{n-1}} = \lambda_n(T_n - T_{n-1}),$$

for Markovian systems.

Formula (27) can be rewritten in terms of the conditional distribution functions of the (T_n). (So can (23), of course.) For this purpose, let $L_n(w,s) = K_n(w,s-T_{n-1}(w))$, then L_n is the conditional distribution function of T_n, given $F_{T_{n-1}}$ and

$$A_t = A_{T_{n-1}} + \int_{T_{n-1}}^{t \wedge T_n} dL_n(s)/(1-L_n(s-)), \tag{28}$$

on $[T_{n-1} < t \leq T_n]$, $n \geq 1$. Simple direct proofs of equation (28) without the aid of (23) can be found in T. C. Brown [1978] and Liptser, Shiryayev [1978].

Finally, we point out that if the point process N has an F-intensity

$\lambda = (\lambda_s, s \geq 0)$ (i.e., if A is absolutely continuous with respect to Lebesgue measure, with λ as the Radon-Nikodym derivative), then

$$\lambda_t(w) = \sum_{n \geq 0} \frac{k^{(n+1)}(w, t - T_n(w))}{K^{(n+1)}(w, [t - T_n(w), \infty))} 1_{((T_n , T_{n+1}]]}(w, t) \qquad (29)$$

where $k^{(n+1)}$ is the conditional density of S_{n+1} give F_{T_n}. Hence, we have an interpretation of the intensity as a conditional hazard function.

Chapter 5. **Local Martingales and Semi-Martingales**

5.1. **Local Martingales:** The important concept of local martingales was introduced by K. Ito and S. Watanabe in an article titled *Transformation of Markov Processes by Multiplicative Functionals*, published in the Annals of Institute of Fourier in 1965. This concept provides a generalization of martingales which will be used to extend the stochastic integral developed in Chapter 6 beyond the class of square integrable martingales.

5.1.1. **Definition**: An adapted, Skorokhod process M is said to be an **F-local martingale** iff there exists a sequence of F-stopping times, $(T(n), n \geq 1)$, increasing to ∞ as $n \rightarrow \infty$, such that for each n, $m_n = (M_t^{T(n)}, t \geq 0)$ is a uniformly integrable F-martingale.

We also introduce the term **F-local L_p-martingale** as a process, M, for which there exists a sequence of F-stopping times, $S_n \uparrow \infty$, such that for each n, $m_n(t) = M^{S_n}(t)$, $t \geq 0$, defines an L_p-martingale.

5.1.2. Remark: The notation is attempting to say that for each n, the process defined by $t \rightarrow m_n(t) = M(T(n) \wedge t)$ is a uniformly integrable martingale.

The sequence (T(n)) is called the **localizing sequence** of the local martingale, or of the local L_p-martingale. This device of only requiring desirable properties such as boundedness and integrability locally (on stochastic intervals [[0,T(n)))), occurs frequently in the theory of martingales and will be discussed at some length in Chapter 6. Relative to paths, this particular form of localization is in the same spirit as truncation of functions in classical analysis and probability theory, with the further qualification that it is intended for use on processes that will occur as integrators in (stochastic) integrals. Another type of localization by stopping times for integrands will occur in Chapter 6. In the study of martingales, localization is a type of path-wise truncation that is mathematically tractable because of the Doob Optional Sampling (Stopping) Theorem.

5.1.3. Remark: The definition states that $(m_n(t), F(t), t \geq 0)$ is a uniformly integrable martingale for each $n > 0$. This can be proved equivalent to the same requirement on $(m_n(t), F(T(n) \wedge t), t \geq 0)$, Kalianpur [1980].

5.1.4. Remark: Actually, the definition of a local martingale does not have to include uniform integrability. This can always be achieved by replacing T(n) with $T(n) \wedge k$, for some fixed $k > 0$. This remark is made to highlight what is really being assumed. In this spirit, we remark that *if X is a bounded local F-*

martingale, then X is an F-martingale.

Just notice that $X(y \wedge T(n)) \to X(y)$, a.s.P, $y \geq 0$, as $n \to \infty$, and since $E(X(t \wedge T(n)) \mid F(s)) = X(s \wedge T(n))$, for $s<t$, the boundedness allows us to pass to the limit under the expectation as $n \to \infty$ to obtain $E(X(t) \mid F(s)) = X(s)$, a.s.P.

Similarly, using Fatou's Lemma (the liminf part), it is easy to show that *a positive,local martingale is a positive supermartingale.*

5.1.5. Remark: *Every martingale is a local martingale.* To see this, take $T(n)=n$ and let M be a martingale. Then

$$M(T(n) \wedge t) = M(n \wedge t) = E\{M(n) \mid F(t \wedge n)\} = E\{M(T(n)) \mid F(t \wedge T(n))\}.$$

It follows, from the characterization of uniform integrability given in Chapter 2, with $M(T(n)) = Z(\infty) = Z(\infty,n)$, that $t \to M(T(n) \wedge t)$ is a uniformly integrable martingale.

5.1.6. Finally, Chung and Williams give the following converse of sorts to the previous observations.

5.1.7. **Theorem**:
If M is a local L_p*-martingale and if for each* $t \geq 0$, $\{ \mid M_{t \wedge T(k)} \mid \}$ *is uniformly integrable, where (T(k)) is the localizing sequence of M, then M is an* L_p*-martingale.*

5.1.8. Remark: We observe the following fact, which will explain to some extent, the Ito-Kunita-Watanabe approach to stochastic integration, which builds on the class of square integrable martingales. Consider an almost surely continuous martingale, m. Define the sequence (T(n)), by $T(n) := \inf\{ t : |m(t)| \geq n \}$, and $= \infty$, if $\{...\} =$ empty, for each positive integer n. Each T(n) is a stopping time by results in Chapter 2 concerning debuts. By Doob's Optional Sampling Theorem, $m(t \wedge T(n))$ is a martingale, for each n, and $| m(t \wedge T(n))| \leq n$, for all $t \geq 0$. It follows that the stopped continuous martingale is bounded on the interval $[[0,T(n)]]$, and so "square integrable", in a sense to be made precise in Chapter 6. Therefore, once the stochastic integral has been defined for square integrable martingales, it is available for all continuous martingales by **localization**.

It should be noted that if the trajectories of m are not continuous then

$m(t \wedge T(n))$ is bounded only on $[[0, T(n)))$. We have no idea of the magnitude of any possible jump at $T(n)$. Providing for this, in extensions of the integral, is one of the difficult issues in the construction of a stochastic integration theory for arbitrary local martingales, rather than just for continuous local martingales.

5.1.9. Having hinted at one use of localization we will now formally state and prove a result (Chung-Williams[1983, p.21]) that is required for the work in Chapter 6:

5.1.10. **Lemma:**

Any continuous local martingale is a local L_p*-martingale for any* $p \in [1, \infty]$

The proof of this technical result depends strongly on the previously stated "converse" of Chung and Williams, so the idea of the proof is to figure out how to stop the local martingale in such a way that it defines a sequence of uniformly integrable local martingales. Let m be the continuous local martingale with localizing sequence (T_n) and set $S_n = \inf\{ t > 0 : |m(t)| > k \}$, a sequence of stopping times (Chapter 2).

Then $R_k^n := min(S_k, T_n)$ defines a double sequence of stopping times. So Doob's Optional Stopping Theorem tells us that $(m(t \wedge R_k^n))$ is a double sequence of martingales. Further, by definition of (S_n), for each k, this sequence of martingales is bounded by k for all n. Therefore, $m^k(t) := m(t \wedge S_k)$ defines a martingale for each fixed k. Hence, (S_k) is a localizing sequence for m such that for each k, m^k is bounded and so in L_p for any $p \geq 1$.

5.1.11. Remark: There is also a close relationship between martingale transforms and local martingales. Let X be adapted. It can be shown that X is a (discrete time) local martingale iff X is the transform of a martingale.(Meyer [1973]) It follows that if X is a P-integrable, local martingale then it is a martingale. (This is not true in continuous time.) So a local martingale, in discrete time, is not much of a generalization of a martingale.

5.2. **Semi-Martingales:** We have encountered the concept of semi-martingale several times in this note. We can now give a general definition of this concept.

5.2.1. **Definition**: A Skorokhod process, $X = (X(t), t \geq 0)$, is called a **semi-martingale** if it allows the following decomposition:

$$X(t) = X(0) + m(t) + A(t),$$

where m is an F-local martingale, null at time zero and A is a process of bounded variation (BV(F)).

5.2.2. Remark: In the last section of this chapter we will give a number of examples to illustrate how a wide variety of particular processes can easily be put into the form of a semi-martingale.

Recall the Doob Meyer Decomposition in Chapter 1. There are numerous varieties of this decomposition theorem. This particular form will be deduced from a much more restrictive and easily proved form in Chapter 6. Indeed, in our attempt to construct a stochastic integral relative to semi-martingales, we will spend a relatively large amount effort studying semi-martingales in Chapter 6.

5.2.3. **Theorem: (Doob-Meyer Decomposition)**
If X is a submartingale, then there exists a unique previsible increasing process A, A(0)=0 and a local martingale M, M(0)=0, such that

$$X(t) = X(0) + M(t) + A(t).$$

This decomposition is unique (a.s.P).

5.2.4. Remark: A is the previsible compensator of X, as defined in the Chapter on dual previsible projections. We will illustrate the Doob-Meyer Decomposition with counting processes.

We first note that since a counting process, N, always has nondecreasing sample paths, it is a submartingale. It follows from the Decomposition theorem that there exists an increasing, F-previsible, P-integrable, process, A, with $A(0)=0$, and an F-local martingale, m with $m(0) = 0$, such that $N = m + A$.

5.2.5. **Theorem:**
Let N *be a point process adapted to the filtration* $F = (F(t), t \geq 0)$. *Then there exists a unique,* F*-previsible, increasing process,* A, *with* $A(0) = 0$, *such that* $N(t) = M(t) + A(t)$, *where* M *is an* F*-local martingale,* $M(0) = 0$. *The localization sequence,* $(T_n, n \geq 1)$, *for* M *may be defined by setting* $T_n := \inf\{t \mid N(t) \geq n\}$, *if* "{"..."}" $\neq \phi$, *and* $= \infty$, *otherwise.*

It can be shown [e.g. Liptser and Shiryayev vol II] that A is continuous iff the counting process, N, only jumps at totally inaccessible times. Since applications in this note will concentrate primarily on counting processes with absolutely continuous compensators, it follows that in these cases the counting process jump times are always totally inaccessible. As noted earlier, the Poisson process is such a process. Its compensator is, of course, given by $A(t) = \lambda t$, where $\lambda > 0$.

5.2.6. We now give some examples to illustrate the dependence of A on the filtration, F. We need to recall the Jacod [1975] formula discussed in the section on dual previsible projections.

Assume that the filtration is the **internal history**, the σ-algebra generated by the counting process, N.

5.2.7. **Example(1)**: Except for some simple modifications, this example is given in Liptser and Shiryayev [1978]. Suppose that $X=(X(t),K(t),t\geq 0)$ is an adapted process with continuous paths and (K(t)) satisfies the "usual conditions". Define $T(n):=\inf\{t: X(t) > 1-(1/n)\}$, with $T(n,w) := \infty$ if $\{...\}$ is empty. Then we know from Chapter 2 that each T(n) is an K-optional time. Define the counting process, N=(N(t),K(t)), by setting $N(t) := 1_{[T(\infty)\leq t]}$. Then, since (T(n)) increases to $T(\infty)$ and the sequence is optional, we see that $T(\infty)$ is a previsible time. By definition then, N is also previsible. Hence, in the Doob-Meyer decomposition, the previsibility of A (and so, the uniqueness of the decomposition) implies that N=A. (Any process which is indistinguishable from zero is certainly a martingale.)

Now, changing histories, let N be defined as before, except that N=(N(t),O(t)), where O(t) is the sigma algebra generated by N. Let F be the distribution function of $T(\infty) := T$, and suppose that $1 - F(s-) > 0$ on $[0,\infty]$.

Then, using the Jacod result (see the section on Dual Previsible Projections),

$$A(t) = \int_0^{t \wedge T(\infty)} dF(s)/(1-F(s-)).$$

Clearly, $A(t) = -\ln(1 - F(t \wedge T)), t\geq 0$.

Thus, when A is K-previsible, A is the two valued counting process, N, but in the second example when A is O-previsible, $1 - \exp(-A(t,w)) = F(t \wedge T(w))$.

5.2.8. **Example(2)**: When A is absolutely continuous relative to Lebesgue measure and F(t) contains O(t), as defined in the last example, then,

with history (F(t)) we have

$$A(t) = \int_0^t r(s)ds.$$

with history (O(t)) we have

$$A(t) = \int_0^t \hat{r}(s)ds,$$

where $\hat{r}(s) := E\{ r(s) \mid O(t) \}$.

5.3. **Examples of Semi-Martingales:** In this section we will give several examples of semi-martingales. The last of these examples will demonstrate a procedure for writing a function of a point process as a semi-martingale.

5.3.1. **Example(1)**: Let N be a counting process adapted to the filtration F. By definition, N is finite for every $t \geq 0$. Assume that its previsible compensator is absolutely continuous relative to Lebesgue measure, with Radon-Nikodym density, λ, the F-intensity of N. (See the section on previsible projections for these definitions.) Then,

$$N(t) = \int_0^t \lambda(s)\, ds + M(t), \qquad (*)$$

where M is an F-local martingale and λ is a non-negative, measurable process.

For example, when λ is a constant, then (N,P) is the Poisson process with parameter λ. As the Poisson process is the baseline counting process, both historically and usefully, it is important to note that property (*) characterizes this process. That is, Watanabe [1964] proved that if $N(t) - t\lambda$ is a martingale, then (N,P) is Poisson. P. Brémaud [1975] subsequently showed that if A is any deterministic, right continuous increasing mapping of $(0,\infty)$ into itself, with $A(0) = 0$, and N - A is a (P,N)-local martingale then (N,P) is a generalized Poisson process in the sense that the characteristic function of (P,N) is given by

$$E(\exp(iu(N(t)-N(s)))) =$$

$$= \prod_{s<v\leq t} \{ e^{iu}\,\Delta A(v) + (1-\Delta A(v)\} \exp[e^{iu}-1]\,(A^c(t) - A^c(s)),$$

where A^c is the continuous part of A. Compare this and equation (*) with the Doubly Stochastic Bernoulli process of Section 1.10.4.

5.3.2. **Example**(2): A sequence of sums of independently distributed random variables, (Y_k), with finite expectation, can be used to construct a sequence of semi-martingales. Let $a_k = EY_k$ and for each $n \geq 1$, define $X_n(t)$ for $t \geq 0$, by setting

$$X_n(t) = \sum_{k=1}^{[nt]} (Y_k - a_k) + \sum_{k=1}^{[nt]} a_k = M(n,t)+B(n,t).$$

Then $X_n = (X_n(t), t \geq 0)$ is a sequence of semi-martingales .

It is worth noting that in this example the term, B, is purely deterministic and, under very general conditions, for large n, M has the characteristic properties of integrated **noise**.

5.3.3. **Example(3)**: Let

$$X(t) := \int_0^t f(s)\,ds + M(t) = B(t) + M(t),$$

where $(M(t),F(t))$ is a Wiener Process (Doob, [1953]), and $f(t)$ is an $F(t)$-measurable process such that $E(\int_0^t |f(s)|\,ds) < \infty$, for each $t \geq 0$. This is the classical model for **integrated signal** plus noise.

In this case, X(t) is a Wiener process with **drift** process B(t) (drift rate f). Further, if W denotes the standard Wiener process, and, if M(t) is the Ito-integral of g relative to W (see the section on Stochastic Integration), with the Lebesgue integral of g^2 having finite expectation, then X(t) is a Wiener process with drift rate, f, and **diffusion coefficient**, g.

5.3.4. Remark: The remaining examples in this section illustrate a technique for writing functions of point process as semi-martingales. This type of procedure will be extremely useful in any application of the theory to nonlinear filtering of point process.

5.3.5. **Example(4)**: Let (N,P) be a Poisson process with parameter c and define the stochastic processes X by setting $X(t) := \exp(qN(t))$, for every $t \geq 0$, where q is a fixed positive number. Although it won't play a distinctive role in this example, we will let F=(F(t)) denote a history of the process N = (N(t)). This example, like others, will be used again in this note as we illustrate the various stages of the filtering problem, and notation will be carried forward.

Clearly,

$$X(t) = X(0) + \sum_{0<s\leq t} \Delta X(s).$$

Since, X jumps at a point s only when N jumps at s, and then $N(s) = N(s-) + 1$, we can write

$$\Delta X(s) = 2\, X(s-)\, e^{\frac{q}{2}} \sinh(\frac{q}{2}).$$

at jump points. So, since $X(0) = 1$, we have

$$X(t) = 1 + e^{\frac{q}{2}} \sinh(\frac{q}{2}) \sum_{0\leq s\leq t} X(s-)\Delta N(s)$$

Then we can write X in the form

$$X(t) = 1 + \int_0^t c\, \lambda(s)ds + M(t) = X(0) + B(t) + M(t).$$

where $\lambda(s) := 2(\exp(q/2))\sinh(q/2)X(s-)$, and

$$M(t) := \int_0^t \lambda(s)\, d(N(s) - cs)).$$

Since the compensated point process, N(t) - ct, is a martingale, and λ is previsible, it follows from the theory of Lebesgue-Stieltjes stochastic integration that M(t) is a martingale. Thus, as B is a process of integrable variation, X is a semi-martingale.

5.3.6. **Example(5)**: (Brémaud [1977], [1981]) Consider a **queue** with the number of messages (customers) arriving during the interval [0,t] denoted by a(t) and the number of departures during this time period denoted by d(t). Let q(t) be the number of messages waiting for service (processing) or being served at time t. Assume that q(0) is a positive random variable. Set $q(t) = q(0) + a(t) - d(t)$ for all $t > 0$. Assume that $\Delta a(t)\Delta d(t) = 0$ for all $t \geq 0$ (i.e., a and b have no jumps in common). By definition $q(t) \geq 0$, for all $t \geq 0$. Let $z(t,w,n) = 1_{[q(t,w)=n]}$; later we will use this example to determine the conditional distribution of q given observations on the number of arrivals. This is because

$$E\,(\,z(t)\mid F(t)\,) \;=\; P\,(\,q(t) = n \mid F(t)\,).$$

As in the previous example, we begin by writing

$$z(t) = z(0) + \sum_{0<s\leq t} \Delta z(s) = z(0) + \sum_{0<s\leq t} \Delta z(s)\Delta a(s) + \sum_{0<s\leq t} \Delta z(s)\Delta d(s)$$

Fix $n \geq 1$; then if s is a point of increase of a, (that is, if $\Delta a(s)=1$), then $q(s) = q(s-) + 1$. Hence,

$$z(s,n) := 1_{[q(s)=n]} = 1_{[q(s-)+1=n]} = z(s-,n-1),$$

so that $\Delta z(s,n) = z(s-,n-1)\, 1_{[n\geq 1]} - z(s-,n)$.

Similarly, if $\Delta d=1$, then $\Delta z(s,n) = z(s-,n+1) - z(s-,n)$.

Assuming that the counting processes a and b have F-intensities, $t \rightarrow l(t,w)$ and $t \rightarrow u(t,w)$, we can accumulate the previous equations to write

$$z(t,n) - z(0,n) = \int_0^t \Delta\, z(s,n)\,(d\,a(s) + d\,d(s)).$$

Then, as in the last example, by adding and subtracting Lebesgue integrals of the intensities, we obtain

$$z(t,n) - z(0,n) = \int_0^t \Delta\, z(s,n)\,(l(s)ds + u(s)ds + d\,m(s) + d\,v(s))$$

$$= \int_0^t (z(s,n-1)1(n\geq 1) - z(s,n))\, l(s)\, ds) +$$

$$+ \int_0^t (z(s,n+1) - z(s,n)1(n\geq 0)u(s)ds) + M(t) + V(t).$$

Thus, using the linearity of the Lebesgue integral and the fact that the sum of two martingales, in this case M and V, is again a martingale, we have

$$z(t,n) - z(0,n) = \int_0^t f(s)\, ds + m(t) = B(t) + m(t),$$

as the semi-martingale representation of z, where

$$f(s) = (\, z(s,n-1)\, 1(n\geq 1) - z(s,n)\,)\, l(s) + (\, z(s,n+1) - z(s,n)\, 1(n\geq 0))u(s)$$

and

$$m(t) = M(t) + V(t) = \int_0^t (z(s-,n-1)1(n\geq 1) - z(s-,n))(da(s) - l(s)ds)$$

$$+ \int_0^t (z(s-,n+1) - z(s-,n)1(n\geq 0))(dd(t) - u(s)ds)).$$

Chapter 6. **Stochastic Integrals**

6.1. **Introduction**: N. Wiener [1923] defined a stochastic integral with Brownian motion integrators and deterministic integrands. K. Ito [1944,1951] developed a stochastic integral for a class of processes which are optional (non-anticipating) relative to Brownian motion.

In his construction, aside from the properties of any continuous martingale, Ito only used two properties of Brownian motion. Namely, that

$$(W(t), t \geq 0) \tag{1}$$

and

$$(W^2(t) - t, t \geq 0) \tag{2}$$

are martingales, where W is standard Brownian motion.

6.1.1. J. Doob [1953] extended stochastic integration of Ito to the class of square integrable martingales. In their important paper, "On Square Integrable Martingales", Kunita and Watanabe used the following result analogous to equation (2) for square integrable martingales: Since m is a square integrable martingale, m^2 is a submartingale, so by the Doob-Meyer Theorem, there is an increasing process, denoted $<m,m>$, such that

$$(m^2(t) - <m,m>(t), t \geq 0) \tag{3}$$

is a martingale. We will formally introduce $<m,m>$ below, but equation (3) has already occurred in Chapter 1 for discrete processes so the reader should not have difficulty with it. For continuous process, we will only point out that in the case of Brownian motion, $<m,m>(t) = t$, in which case equations (2) and (3) agree and, also, that the Kunita Watanabe stochastic integral reduces to the Ito integral when the martingale integrator is Brownian motion.

Stochastic calculus is still young enough, in terms of the length of time it takes for significant mathematical theories to develop, that it is almost always presented as it was developed historically. We will call this the traditional approach. Dellacherie's 1978 talk at Helsinki is an exception, and in some ways Jacod [1979] is also. We will follow Jacod.

In the traditional approach the stochastic integral is developed as outlined in the next Section. As described there, it is defined first for a particular class of intergrands consisting of linear combinations of (previsible) indicator processes relative to a square integrable martingale, m. The actual definition of this "elementary stochastic integral" is given in terms of the transforms of Chapter 1. The first extension, to the space of square integrable martingales, is accomplished using the path-wise stochastic Lebesgue Stieltjes integral relative to the increasing, previsible process, $<m,m>$. However, as we have pointed out earlier, this process does not exist when the underlying martingale does not have moments of the second order. To remedy this situation a second increasing process is created that does not require the existence of second order moments. It is the continuous parameter analogue of the optional quadratic variation process of Chapter 1 and is also denoted by $[m,m]$. As in Chapter 1, if the process m has second order moments, so that $<m,m>$ exists, $[m,m] - <m,m>$ is a martingale. Hence, when m is not in L_2, the $< , >$ process is defined as the dual previsible projection of the process $[m,m]$. Formally then, the development of the stochastic integral for the larger class of integrators proceeds as before in terms of a Lebesgue-Stieltjes integral relative to $[m,m]$.

This then becomes the procedure for extending the integral to an ever widening circle of families of processes, culminating with its final extension to semimartingale integrators and locally bounded previsible integrands. At each step, preparation for the next extension is made by first extending the increasing process, $[m,m]$, and then repeating the definition of the next more general stochastic integral in terms of, notationally, the same defining equation as utilized at the previous step.

6.1.2. Embedded in the procedure just sketched is a method of extending the (integrator) processes themselves. We have already encountered one example in going from martingales to local martingales. This is the **method of localization**. It is one of the most important applications of stopping times in the theory of martingales. It goes as follows.

Let $(\Omega , H, (F(t), t\epsilon R_+), P)$ be a filtered probability space satisfying the "usual conditions". Let C be a family of processes (equivalence classes of indistinguishable processes) defined on this probability space. Denote by $C_{loc} = C_{loc}(F,P)$ the family of processes, X, defined on the same probability space for which there exists an increasing sequence, $(T_n , n\epsilon Z_+)$, of stopping times, $T_n \uparrow \infty$ a.s.P., such that each stopped process $X^{T_n} \epsilon C$. For example, letting M_u be the set of uniformly integrable martingales, its localization, is $(M_u)_{loc}$, the family of local martingales. We have seen that $M_u \subset (M_u)_{loc}$ in Chapter 5.

C_{loc} is called a **localized class**. Jacod [1979] proves a number of interesting results on the algebra of localized classes. For instance, he shows that if a class, C, a vector space of processes, is closed under the operation of stopping (called **stable** under optional stopping), then $(C_{loc})_{loc} = C_{loc}$. The reader should ponder this result in relation to the family M_u.

On our way to extending stochastic integrals, we will apply localization to a number of classes of processes. This will be carried out a little differently for integrands than for integrators, for obvious reasons. In any case, the class of bounded, previsible processes becomes the class of locally bounded, previsible processes, and the class of processes of integrable variation becomes the class of processes of locally integrable variation. We will prove that the class, S, of semi-martingales cannot be extended by localization: $S = S_{loc}$ (Jacod [1979]).

The stochastic integral will not be extended beyond the class S of integrators. The reason for this is simple. It cannot be done. That is, it can't be done if we want sequences of stochastic integrals to have the following natural Cauchy property:

> Let $(h_t^{(n)})$ be a uniformly bounded sequence of previsible processes. Then the point-wise convergence of this sequence to 0 with $n \to \infty$, implies that
>
> $$\int_{[0,t]} h_s^{(n)} \, d X_s \to 0, \text{ in probability, as } n \to \infty$$
>
> for all t, where X is in S.

Bichtler [1981] proved that if an integrator is Skorokhod, adapted and the corresponding stochastic integral possessed this Cauchy property, then the integrator is necessarily a semi-martingale.

The material in this chapter is based primarily on Jacod [1979], Kunita and Watanabe [1967], Doleans-Dade and Meyer [1970], Meyer [1976], Rogers [1981], and Dellacherie and Meyer [1978]. Dellacherie [1978], Chung and Williams [1983], and Ikeda and Watanabe [1981] were also used.

6.2. An Outline of the Construction of Stochastic Integrals:

6.2.1. **Introduction**: In this Section, we will attempt to outline the traditional approach to constructing the stochastic integral. In succeeding Sections, we will mostly follow the development of Jacod [1979]. Although Jacod's development does not begin with elementary processes and simple integrals, there is much in common with the outline given here. The principal reason for following Jacod is that it leans more heavily on martingale methods (the Strasbourg variety), than on the methods of classical functional analysis. It therefore appears to be more succinct and self-contained than the traditional approach.

What we are referring to as the traditional approach begins the way most of us would expect. However, as observed by Rogers [1981], some very clever new ideas were required to successfully carry out the original development of the stochastic integral. As noted earlier, this was done by Kiyosi Ito in the 1950's for Brownian Motion and extended to square integrable martingales in the 1960's by Kunita and Watanabe. P-A. Meyer and the Strasbourg School of probabilists are mostly responsible for the final extension to semi-martingales in the late 1960-70 time frame.

6.2.2. **Outline**: The first integral to be introduced in this Section is called the **elementary stochastic integral**. In French literature, it is called the **triviale stochastic integral**, translated as the "obvious stochastic integral". As demonstrated in an example by Rogers, aside from starting with processes whose sample paths are simple step functions and defining their integral as a finite sum, the definition of the elementary stochastic integral is neither trivial nor obvious.

Let $(\Omega, H, (F(t)), P)$ be a filtered probability space with $F(\infty) = \sigma(\bigcup_{s \geq 0} F(s))$ contained in H and assume the "usual conditions".

6.2.3. Let the family Ξ designate the vector space of linear combinations of indicator functions of rectangular subsets of $(s, t] \times \Omega$ of the form $(s, t] \times A$, with $A \epsilon F(s)$, and $s<t$, s,t in R_+; in other words, Ξ consists of linear combinations of the kernel processes which generate the F-previsible σ-algebra. More precisely, let Ξ be the family of processes $H=(H(t), t \geq 0)$ such that H is adapted, left continuous, bounded and for which there exists a finite set $\{t_k : k = 0,1,2,\dots,n,n+1\}$ which partitions $[0,\infty]$,

$$t_0 = 0 < t_1 < \cdots < t_n < t_{n+1} = \infty,$$

and such that $t \rightarrow H(t,w)$ is constant on each subinterval of the partition. Further, assume that each r.v. $H_i = H(t_i)$ is $\mathbf{F(t_i)}$ **measurable**, $i=0,1,\dots,n,n+1$.

6.2.4. **Definition**: Let M be a bounded martingale and $H \in \Xi$. Then the **elementary stochastic integral**, H.M, is

$$(H.M)(t) := \sum_{k \geq 0} H_k \Delta M_{t_k}^{t} \tag{4}$$

$$:= H(0)M(0) + \sum_{k=1}^{n} H(t_k)\,(M(t_{k+1} \wedge t) - M(t_k \wedge t)).$$

6.2.5. Remark: Based on Chapter 1, this is nothing more than the martingale transform of an integrable, stopped, bounded martingale M by the bounded previsible process H, hence we have immediately that H.M is a martingale.

To ease the notational burden for the reader, we will often write $\int_0^t H(s)\, dM(s) := (H.M)(t)$. Notice that with the definition as in (4), there should be no ambiguity in meaning if we set the upper limit in the last expression equal to the symbol ∞.

In general, here and in the sequel, when the notation H(s) becomes too cumbersome because of superscripts and such we will write H_s for *H(s)*. *Though perhaps* ambiguous here, this should not be the case in actual usage.

Now, since M is a square integrable martingale, $M^2 = (M_t^2, t \geq 0)$ is an F-submartingale and so the Doob-Meyer decomposition theorem of Chapter 5 guarantees the existence of an increasing, previsible process, A, with the property that $M^2 - A$ is an F-martingale. (We have already used the notation $\langle M,M \rangle = A$.)

Then it is easy to see that for a simple process H,

$$E\Big(\int_0^\infty H_s\, dM_s\Big)^2 = E\Big(\sum_{k \geq 0} H_k^2 (\Delta M_k)^2\Big)$$

$$= E\Big(\sum_{k \geq 0} H_k^2\, E(\,(\Delta M_k)^2 \mid F_k\,)\Big)$$

$$= E(\,(H^2.\langle M,M \rangle)^\infty\,) = E\int_0^\infty H_s^2\, d\langle M,M \rangle_s\,.$$

where we have used the previsibility of H and the adaptiveness of M to obtain

$$E(H_jH_k\Delta M_j\Delta M_k) = E(H_jH_k\Delta M_j E(\Delta M_k \mid F_{k-1})) = 0$$

for $0 \leq j < k$, and so the first equation. To go from the first equation to the second use either the definition of $<M,M>$ from Chapter 1, or a simple relationship between M and $<M,M>$ (in fact, the reason for the name "quadratic variation") that will be derived later in this Chapter.

Since the F-previsible σ-algebra is generated by the kernel processes, any previsible process is the limit of a sequence of simple processes. Let

$$L_2(M) := \{H : H \text{ previsible}, \; E\int_0^\infty H_s^2 d<M,M>_s < \infty\},$$

where the last integral is a Stochastic Lebesgue-Stieltjes integral (Chapter 3) relative to the nondecreasing process $<M,M>$, the dual previsible projection of M.

Let m be a square integrable martingale and set $\|H\|_{L_2(m)} := E(\int_0^\infty H_s^2 d<m,m>)^{\frac{1}{2}}$. Then $\|H\|_{L_2(m)}$ is a norm on $L_2(m)$. Notice that since H is previsible, $L_2(m)$ is independent of the choice of martingale compensator of m^2. (Just recall the results in Chapter 4 on dual previsible projection.)

Define K^2 to be the space of square integrable martingales, m. ($\sup_{t \in R_+} Em_s^2 < \infty$). The phrase "square integrable" is a result of the fact that as a submartingale, m^2, has an increasing mean-value so that when the supremum is taken over compact intervals, $[0,T]$, rather than over R_+, "square integrability" indeed just means $Em_T^2 < \infty$, or existence of second order moments.

From Chapter 2, we know that if m is square integrable, then it has a terminal r.v., m_∞. Let $\|m_\infty\|_2 := (Em_\infty^2)^{\frac{1}{2}}$ be the norm on K^2.

The following Theorem is then proved in almost every presentation of the stochastic integral. It establishes an isometry between $L_2(m)$ and K^2.

6.2.6. **Theorem:**
The mapping $H \rightarrow H.m$, *from* Ξ *to bounded martingales, can be extended uniquely as a norm preserving operator from* $L_2(m)$ *onto* K^2, *and will continue to be denoted by* $H \rightarrow H.m$.

6.2.7. Next, it may be verified (as in Chapter 1) that

o $X^T = 1_{[[0,T]]}.X$, for all optional T.

o $\Delta(H.X)_t = H_t \Delta X_t$, a.s.P, $t \geq 0$.

Seemingly, in all integration theories the difficult work begins with K^2, the space of square integrable martingales. Much more will be said about this space in the next Sections. Kunita and Watanabe [1967] give fundamental results on the decomposition of the space K^2. To prepare for this we need to know that a "stable" subspace, Q, of K^2, is essentially just a closed subspace of K^2, that is closed under stopping. We need the following

6.2.8. **Definition:**: Processes $m,n \epsilon K^2$ are said to be orthogonal if the process $mn = (m_t n_t, t \geq 0)$ is a martingale.

Remark: It will be shown later in this Chapter that if m and n are square integrable and vanish at the origin, $m_0 = 0 = n_0$, then m and n are orthogonal if $Em_T n_T = 0$, for every stopping time T.

Kunita and Watanabe [1967] prove: *If Q is a stable subspace of* K^2, *then every* martingale, m, in K^2 can be uniquely decomposed into a sum, $m = x + y$, where x belongs to Q and y is orthogonal to every element of Q.

If one recalls (Chapter 2) that the norm in K^2 is equivalent to the $L_2(P)$ norm of the supremum process, $m_t^* = \sup_{s \leq t} m_s$, it is easy to show that the space of continuous, square integrable processes is stable. If we call this space Q, and observe the convention that $m_{0-} = 0$, we have $Q \subset K_0^2$, the latter being the set of square integrable martingales that vanish at the origin.

Applying the decomposition theorem of Kunita and Watanabe, this Q yields ***a unique decomposition of any square integrable martingale, m, into a continuous, square integrable martingale,*** m^c, ***and a "purely discontinuous" square integrable martingale,*** m^d, ***which is orthogonal to every element of Q.***

The space of purely discontinuous martingales will be described in a later Section. For now it is sufficient to know that this space is the closure in K^2 of a relatively simple class of martingales whose paths are of bounded variation, a.s.P. But not every purely discontinuous martingale is of bounded variation. In contrast to this, every nonconstant, continuous (nonzero, by the convention $m_0 = 0$), martingale has (a.s.P) paths of unbounded variation. This follows easily from the Doob Meyer decomposition theorem by assuming that such a (necessarily previsible) martingale is of bounded variation.

For these reasons the construction of a stochastic integral, even within the space of square integrable martingales, is a formidable affair.

Kunita and Watanabe also define a process, $<m,n>$, for square integrable martingales (recall Chapter 1 for the discrete analogue), as the unique, previsible process with the property that mn - $<m,n>$ is a martingale. Note that $<m,n> = 0$ then becomes a sufficient condition for orthogonality of m and n. This new process, which is of bounded variation, is used by Kunita and Watanabe to characterize the process, H.m, as opposed to the operator H→H.m. But it is clear that existence problems arise when m and n are not square integrable. To cope with this difficulty, Meyer introduced a process denoted by [m,n] which exists even when m and n are not square integrable and, like the process $<m,n>$, is a process of bounded variation.

Finally, Kunita and Watanabe created a type of Schwarz inequality in terms of the process $<m,n>$, (given later in this chapter in terms of the process [m,n]) and used Stochastic Lebesgue-Stieltjes integrals (introduced in Chapter 3) with previsible integrands to establish the following characterization of the stochastic process, H.m:

6.2.9. **Theorem:**
If $m,n \in K^2$ *and* $H \in L_2(m)$, *then*

$$E\left(\int_0^\infty |H_s| \; |d<m,n>| \right) < \infty$$

and the stochastic integral, H.m, *is the unique element of* K^2 *(up to indistinguishability) which satisfies the equation*

$$[H.m, n] = H.[m,n]$$

for every n in K^2.

This rest of the development, as noted in the introduction to this Chapter, consists of a succession of extensions of the quadratic variation processes and of the stochastic integral which culminate in the stochastic integral of locally bounded, previsible processes relative to local martingales and thence to semi-martingale integrators.

The Jacod development starts with the definition of the stochastic integral of a local martingale, albeit a continuous one. The attractive feature of his approach is that it focuses on semi-martingales from the beginning. With some minor exceptions (occurring with the treatment of purely discontinuous processes and in the preparation for the definition of quadratic variation) Jacod's approach is followed in the remainder of this chapter.

6.3. **Some Extensions to Chapters 3-5:** In this Section we will bring together and extend some of the material in Chapters 3, 4 and 5. As usual, let $(\Omega, F(\infty), (F(t), t \geq 0), P)$, where $F(\infty) = \sigma(\bigcup_{t\geq 0} F(t))$, be the underlying filtered probability space. Recall (Section 3.2) the definitions of increasing processes and the notation for the classes of increasing processes, V^+, of processes of bounded (finite) variation, $BV = V^+ - V^+$, and of processes of integrable variation, $IV = IV^+ - IV^+$. (Of course, we mean $V^+ = V^+(F,P)$, $BV = BV(F,P)$, and so on.) Let C be a class of processes. We write C_0 for the set of all $A \epsilon C$ with $A(0) = 0$.

If $A \epsilon BV$, then $B(t) = \int_{[0,t]} |dA(s)|$ denotes the **variation** of the process A. It is the unique (to indistinguishability) process of V^+ such that the measure $(0,t] \rightarrow dB(t,w)$ on R_+ is the total variation of the signed measure $(0,t] \rightarrow dA(t,w)$.

6.3.1. Using the notation introduced in Section 6.1.2, an increasing process, A, ($A \epsilon V^+$), with $A(0) = 0$ is said to be **locally integrable** if $A \epsilon (IV_0^+)_{loc}$. That is, if there exists an increasing sequence, $T_n \uparrow \infty$, a.s.P, of stopping times such that $EA^{T_n} < \infty$. Since $A(t \wedge T_n) \leq A(T_n)$ we can and will write the condition as $EA_{T_n} \leq \infty$. If $A(0) \neq 0$, then the definition applies to the process $t \rightarrow A(t) - A(0)$ and $E(A(0) \mid F(0)) < \infty$ is required. In this case we write $A \epsilon (IV)^+_{loc}$.

A is said to be of **local integrable variation** if $A \epsilon BV$ and the process $t \rightarrow \int_{[0,t]} |dA(s)| := B(t)$ is locally integrable. In this case we write $A \epsilon (IV)_{loc}$. More succinctly, $A \epsilon (IV)_{loc}$ iff $B \epsilon (IV)^+_{loc}$.

When the local integrability of the variation process, B, is at issue we can use the fact that if $EB(T) < \infty$ and $EB(S) < \infty$ for two stopping times, then

$EB(\max(S,T)) < \infty$. Therefore, for local integrable variation, we don't have to require that $T_n \uparrow \infty$; $\sup_n T_n = \infty$ is good enough.

6.3.2. The first two results in this Section concern local variation. Simply stated, increasing previsible processes are locally integrable, and optional processes of BV are of local integrable variation iff they differ from local martingales by a previsible process of bounded variation. The proofs may be found in Dellacherie and Meyer[1982, VI 80] or Jacod[1979, p.17].

6.3.3. **Theorem:**
Let A be a process of bounded variation.

(1) If A is previsible then A is of local integrable variation.

(2) A is of locally integrable variation iff there exists a unique, previsible process $B \in BV$ *such that* $A - B$ *is a local martingale.*

B is unique, modulo indistinguishability, and is called the **dual previsible projection**, or the previsible **compensator** of the A.

This extends the Chapter 4 notion of the dual previsible projection of increasing, integrable processes to bounded variation processes of local integrable variation.

6.3.4. Remark: We will sketch the proof the first part of this Theorem with the aim of giving the reader a feeling for the use and force of these new definitions. In part (1), since A is of bounded variation and previsible, the total variation process of A is previsible. Therefore, without loss of generality, take A to be increasing and $A(0) = 0$. Set $S_n = \inf(\, t : A(t) \in [n,\infty)\,)$, where $1 \le n \in Z_+$. Then by the results concerning the debut of random sets in Chapter 2, S_n is a previsible stopping time. It is strictly positive since $A(0) = 0$, and it can be shown to be previsible since A is previsible. Therefore, for each n, S_n has an announcing sequence, $(S_n^k ,\ k \in Z_+)$ and since $S_n^k < S_n$ for all k (since S_n is strictly positive), we have by the definition of S_n that $A(S_n^k) < n$ for all k. Therefore, $EA(S_n^k) < \infty$. Since $\sup\{S_n^k : n,k\} = \infty$, A is locally integrable.

The following Corollary is useful and obvious. It is a generalization of the result in Chapter 4 which said that IV_0 martingales have evanescent dual previsible projections.

6.3.5. **Corollary:**

If A is of bounded variation, then A is a local martingale iff A is of local integrable variation and the previsible compensator of A is evanescent.

6.4. **Some Spaces of Martingales:** Let $(\Omega, F(\infty), F, P)$, where $F(\infty) = \sigma(\bigcup_{t \geq 0} F(t))$ and $F = (F(t), t \in R_+)$, be a filtered probability space. Let $M_u = M_u(F,P)$ be the space of uniformly integrable F-martingales. Of course M_u is a set of equivalence classes of processes under indistinguishability.

As noted in Chapter 2, if X belongs to M_u, then $(X(t))$ converges a.s.P, and in L_1, as $t \to \infty$, to a terminal random variable $X(\infty)$ and we can write $X(t) = E(X(\infty) \mid F(t))$. The converse was also discussed, so we know that if $Z \in L_1(\Omega,H,P)$ there exists a unique X in M_u such that $X(t) = E(Z \mid F(t))$ and $X(\infty) = E(Z \mid F(\infty))$. It follows that M_u **is mapped bijectively onto** $L_1(\Omega,F(\infty),P)$.

6.4.1. **Definition**: A (right continuous) supermartingale, $X = (X(t), F(t))$, is said to belong to the **class D** iff the family of random variables, $\{X(T) : T \text{ any finite F-stopping time}\}$, is uniformly integrable.

6.4.2. We now characterize M_u as a subset of $(M_u)_{loc}$. For convenience we will write $M_{loc} := (M_u)_{loc}$ throughout this chapter.

6.4.3. **Lemma:** *Let* $X \in M_{loc}$. *Then* $X \in M_u$ *iff* X *belongs to the class D.*

Remark: We will indicate the proof. Let X belong to M_u. Then by our previous remarks, there exists $X(\infty) \in L_1$ such that $X(T) = E(X(\infty) \mid F(T))$, for each optional T. As noted in Chapter 2 (Doob's Optional Stopping Theorem), if we let T range over the set of finite (i.e. real) valued optional times, then the family $\{X(T)\}$ is uniformly integrable and X belongs to the class D.

Conversely, let X be a local martingale in the class D. Then, in particular the family $\{X(t), t \in R_+\}$ is uniformly integrable. Thus, it remains to show that $(X(t), t \in R_+)$ is a martingale. Let (T_n) be a localizing sequence for the local martingale X. Then for s and t real numbers, $s \leq t$, the families $\{X(t \wedge T_n), n \in Z_+\}$ and $\{X(s \wedge T_n), n \in Z_+\}$ are uniformly integrable and the corresponding sequences (s and t are fixed) converge a.s.P. and in L_1 to the random variables X(t) and X(s), respectively. By definition of X as a local martingale, $X^{T_n}(s) = E(X^{T_n}(t) \mid F(s))$ for each n. It follows (using Jensen's inequality) that

$| E((X^{T_n}(t)-X(t)) \mid F_s) | \leq E(| X^{T_n}(t)-X(t) | \mid F_s)$. Taking expectations of both sides, we obtain $E(| E(X^{T_n}(t) | F_s) - E(X(t) | F_s) |) \leq E | X^{T_n}(t) - X(t) |$. Having noted that $(X^{T_n}(t),\ n \geq 1)$ converges to $X(t)$ in L_1, and similarly $(X^{T_n}(s),\ n \geq 1)$ to $X(s)$, we have $X(s) = E(X(t) \mid F_s)$, hence $X \epsilon M_u$.

6.4.4. Remarks: Now let $m^*(t) = \sup_{s \leq t} | m(s) |$ for any process $m = (m(s), s \epsilon R_+)$. Let $p \epsilon [1,\infty]$ and $\| Y \|_p^p = E(| Y |^p)$, the $L_p = L_p(\Omega, F, P)$ norm of Y of order p. Recall that if $p=\infty$, L_∞ denotes the family of F-measurable, bounded functions. Set

$$K^p = \{ m \epsilon M_{loc} : \| m^*(\infty) \|_p < \infty \}.$$

Then $K^p \subset M_u$, for $p \geq 1$. This is because $K^p \subset K^{p'}$ for all $p' \leq p$, by Holder's inequality, and so in particular $K^p \subset K^1$. But if $m \epsilon K^1$, then $m^*(\infty)$ is P-integrable and so $\{m\}$ is in the class D. Therefore, by the last Lemma, $m \epsilon M_u$. That is, $K^1 = M_u$.

6.4.5. As noted in Chapter 2, if $p > 1$ then the norms $\|m(\infty)\|_p$ and $\|m^*(\infty)\|_p$ are equivalent. Therefore if $p \epsilon (1,\infty]$, then K^p can be equipped with the norm defined by the mapping $m \rightarrow \|m(\infty)\|_p$. In this manner, K^p *is identified with the space* $L_p(\Omega, F(\infty), P)$ *through the bijection* $(m \epsilon K^p) \longleftrightarrow m(\infty) \epsilon L_p$, *for* $p > 1$. (Recall also that there exists a bijection between M_u and L_1.)

The space K^2 is called the space of **square integrable martingales** or the space of L_2 - bounded martingales. (These were also defined in Chapter 2.) By remarks in the previous paragraph, the space K^2 is identified with the Hilbert space L_2, having norm $m \rightarrow \|m(\infty)\|_2$ and scalar product $(m,n) \rightarrow E(m(\infty)n(\infty))$.

Set M_0 equal to the space of martingales such that $m \epsilon M_u$ and $m(0) = 0$. We will write $(M_0)_{loc}$ as $M_{0,loc}$.

6.4.6. Following Jacod [1979] we state the following.

Definition: Let m and n be local martingales. Then m and n are said to be (strongly) **orthogonal** if the product mn is a local martingale which vanishes at the origin.

Remark: In the traditional source, Meyer [1975], first defines orthogonality for square integrable martingales; he then extends this to local martingales as above.

For square integrable martingales he defines m to be orthogonal to n if $m(0) = 0$ and $Em(T)n(T) = 0$ for all stopping times, T. He then proves that $m,n \epsilon K^2$ are orthogonal iff the product $mn \epsilon K^1$ and $m(0) = 0$. Jacod's definition restricted to the space K^2. To show this characterization of orthogonality, one needs the following interesting characterization of M_u. One should refer back to the proof of Doob's Optional Sampling Theorem (in Chapter 1) for the genesis of this theorem. Actually, that the equality of expectations of an integrable process at different finite stopping times is equivalent to the Chapter 1 form of Doob's Theorem for martingales is sometimes called Komatsu's Lemma. This Lemma will also be used in the proof of the Theorem characterizing stochastic integrals relative to a continuous local martingale.

6.4.7. **Lemma:**
Let L *be an adapted Skorokhod process for which* $\lim_{t \to \infty} L(t)$ $(=: L(\infty))$ *exists. Then* L *is in* M_u *iff* L(0) *is P-integrable and* $E(L(T)) = E(L(0))$ *for all stopping times,* T.

Remark: If L is in M_u this follows from Doob's Optional Stopping Theorem. For the converse take $T = t_A$, the restriction of the constant stopping time t to some A in F(t). Then $E(L(T)) = E(L(0))$. Since $E(L(\infty)) = E(L(0))$ ($S = \infty$ is an optional time), decomposing both expectations over A and A^c, it follows that $E(L(t)1_A) = E(L(\infty)1_A)$. That is, the last equation holds because

$$EL(0) = EL(T) = \int_A L(t)\,dP + \int_{A^c} L(\infty)\,dP$$

and

$$EL(0) = EL(\infty) = \int_A L(\infty)\,dP + \int_{A^c} L(\infty)\,dP.$$

Since A is an arbitrary event in F(t), we have $L(t) = E(L(\infty) \mid F(t))$ and so L belongs to M_u.

Remark: It follows easily then that this martingale definition of orthogonality is stronger than the natural orthogonality in the Hilbert space $L_2 \longleftrightarrow K^2$ under the inner product condition $Em(\infty)n(\infty) = 0$:

6.4.8. **Theorem:**
If the square integrable martingales m and n are strongly orthogonal, then $Em(\infty)n(\infty) = 0$ *and the product mn is a martingale in* K_0^1.

Remark: We could equally well claim that for all stopping times T, m_T and n_T are orthogonal in L_2. This follows since m_∞^* and n_∞^* in L_2, implies that the product $m_\infty^* n_\infty^* \epsilon L_1$. But then $(mn)_\infty^* \leq m_\infty^* n_\infty^*$ so that $mn \epsilon K_0^1$. It follows that $Em_T n_T = Em_0 n_0 = 0$, by the Lemma.

A converse also holds: *If* $m_0 n_0 = 0$ *and* m_T, n_T *are orthogonal in* L_2, *for all stopping times T, then m and n are orthogonal in the sense of the definition of strong orthogonality.* This is also a consequence of the Lemma.

6.4.9. A **continuous local martingale** (CLM) is a local martingale whose paths are continuous (for P-almost all paths). Let M_{loc}^c be the family of continuous local martingales. On occasion, we will also (following Jacod) use the notations $K^{p,c}$, $M_{0,loc}^c$ and so on, with the same meaning being carried by the superscript c; namely, to denote various subfamilies of M_{loc}^c satisfying the additional requirement of path continuity.

6.4.10. *If the local martingale m is (strongly) orthogonal to each* $n \epsilon M_{0,loc}^c$, *then m* is said to be a **(purely) discontinuous**, or a **compensated jump martingale**. The first name is widely used, but is misleading since, for example, the compensated Poisson process $(N(t) - \lambda(t),\ t \geq 0)$, is such a martingale and its paths are continuous between jumps.

Let M_{loc}^d be the subset of M_{loc} consisting of compensated jump martingales. M_{loc}^d is called the space of **compensated jump martingales**.

Lemma:
Let $p \epsilon [1,\infty]$. *Then* K_0^p, $K^{p,c}$ *and* $K^{p,d}$ *are closed subspaces of* K^p.

Remark: The proof uses the equivalence of $\|m^*(\infty)\|_p$ and $\|m(\infty)\|_p = \|m\|_{K^p}$. Since then if $\|m^{(n)} - m\|_p$, converges to 0, as $n \to \infty$, there exists a subsequence (n_k) such that $\sup_{t \epsilon R_+} | m_t^{(n_k)} - m_t | \to 0$ a.s.P. That is, $m_t^{(n_k)}(w) \to m_t(w)$, uniformly on $[0,\infty]$, for all w in some set of P-measure 1. Therefore, sequences of continuous processes converge to continuous processes, so $K^{p,c}$ is closed.

If $m^k \epsilon K^{p,d}$, for each k and n is any bounded continuous process with $n(0) = 0$, then $En_T m_T^k = 0$ for each k. Again, if m^k converges to m, then $En_T m_T = 0$

and so m is in $K^{p,d}$.

We will return to discuss the structure of $K^{2,d}$ at the end of Section 6.5.

6.5. **Semi-Martingales Revisited:** We now return to the definition of a semi-martingale, introduce some convenient notation and state some results that are crucial to the development of stochastic integrals. This Section and the remainder of the chapter follow Jacod [1979] very closely. In order to remind the reader that the notation should not obscure the simplicity of the semi-martingale concept, we state its definition as follows:

6.5.1. **Definition:** A Skorokhod process, X, is a **semi-martingale** relative to a filtered probability space, (Ω, H, F, P), if there exists a sequence of F-stopping times, $T_n \uparrow \infty$, such that for each n, there exists a sequence of F-martingales $M^{(n)}$ with $M^{(n)}(0) = 0$ and an F-adapted process of bounded variation, $A^{(n)}$, such that $X(t,w) = M^{(n)}(t,w) + A^{(n)}(t,w)$ for all $(t,w) \epsilon [[0, T_n))$.

6.5.2. **Remark:** Of course, this is equivalent to the requirement that there exist processes $m \epsilon M_{0,loc}$ and $A \epsilon (BV)_{loc}$ such that $X = m + A$. Notice that the condition $m(0) = 0$ is no restriction, since if $m(0) \neq 0$ then we can write $X = (m - m(0)) + (A + m(0))$ obtaining a representation of X that satisfies the requirement. Notice also that the requirement that X be Skorokhod is redundant since both $M^{(n)}$ and $A^{(n)}$ in the definition are Skorokhod.

6.5.3. **Remark:** Let S = S(F) = S(F,P) denote the collection of equivalence classes of semi-martingales on (Ω, F(∞), F, P).

If $X \epsilon S$, the representation $X = m + A$ is in general not unique. For a simple, but artificial, example let $A(t) = t$ and m be any local martingale of bounded variation. Then another representation of this X is $X = 0 + A'$, where $A' = m + A$.

A semi-martingale for which the decomposition $X = m + A$ is unique is called a **special semi-martingale**. The subfamily of S consisting of all special semi-martingales will be denoted by S_p.

M. Yor is credited by Dellacherie and Meyer with the following example of a semi-martingale which is not special. Start with the probability space, ([0,1],H,L), where L denotes a complete Lebesgue measure and A is a positive random variable. Define the filtration (F(t)) on H by setting $F(t) = \{\phi,[0,1]\}$, for $0 \leq t < 1$, and $F(t) = H$, for $t \geq 1$. Set $X(t) := A\, I_{[1,\infty)}(t)$. Then X is an increasing process and so is a semi-martingale. This process X is a special semi-martingale iff $A \epsilon L_1$, according to the next Theorem.

The Doob-Meyer Decomposition shows that submartingales are contained in S_p. The following theorem sheds further light on the structure of S_p.

6.5.4. **Theorem:(Characterization of Special Semi-martingales)**
Let $X \epsilon S$ *and*

$$X = m + A. \tag{8}$$

The following statements are equivalent:

> *(1) If there exists a decomposition* (8) *with* A *previsible and in* $(IV)_{loc}$, *then* $X \epsilon S_p$.
>
> *(2) There exists a decomposition* (8) *with* A *in* $(IV)_{loc}$.
>
> *(3) Each decomposition* (8) *satisfies* A *in* $(IV)_{loc}$.
>
> *(4) The increasing process* $X^*(t) = \sup_{t \geq s} |X(s)|$ *belongs to* $(IV^+)_{loc}$.

Remark: The decomposition specified in (1) is called the **canonical decomposition of an element of** S_p.

The following Lemma is needed in the proof of this Theorem:

6.5.5. **Lemma:**

> *(1) X is both a local martingale and a process of bounded variation iff X is a local martingale of local integrable variation.*
>
> *(2) If* $X \epsilon M_{0,loc}$ *and is a previsible process of bounded variation, then X is evanescent.*

Remark: Jacod's sensible way of expressing (1) of the Lemma is to just write

$$M_{loc} \cap BV = M_{loc} \cap (IV)_{loc}.$$

Part (2) has the obvious consequence that the only continuous local martingales of bounded variation are constant processes. In plain language, non-constant, continuous local martingales are of unbounded variation (have paths that are of unbounded variation).

To obtain some exercise with the definitions, we indicate the proof of (1). We only need to show the inclusion in one direction. Let X belong to the left side of the last equation and (T_n) be the localizing sequence for X as a local martingale. Set

$$S_n = \inf(t : \int_{[0,t]} |\, dX(s) \,| \geq n).$$

We will use the notation X(s) and X_s interchangeably. In any case, X^T is still the process stopped at T. Then $\int_{[0,\ t]} |\, dX(s)\,| \leq n + |\,\Delta X_{S_n}^{T_n}\,|$. Since $|\,\Delta X_{S_n}^{T_n}\,| \leq n + |\,X_{S_n}^{T_n}\,|$ and $\min(S_n, T_n) \uparrow \infty$, as $n \rightarrow \infty$, we have that X belongs to $(IV)_{loc}$.

The second statement of the Lemma is an extension of the same result in Chapter 4, where the process was a previsible, increasing martingale vanishing at the origin. The result here follows from the first Theorem of the third Section of this Chapter and its Corollary.

Remark: (Proof of the Theorem 6.5.4 characterizing S_p) : Following Jacod[1979,p.29], assume that statement (2) of the theorem holds and so $X = m + A$, with A of local integrable variation ($A \epsilon (IV)_{loc}$). We show that (1) holds. Write $X = m + A = m + A - A^p + A^p$. Since $A \epsilon (IV)_{loc}$ we know by Theorem 6.3.3 that $A - A^p$ is a local martingale. Therefore, $X = m' + A^p$, where $m' \epsilon M_{loc}$ and A^p is a previsible process of bounded variation. But again by Theorem 6.3.3, it follows then that A^p is in $(IV)_{[loc]}$. This takes care of statement (1), except for uniqueness. But this follows easily from part (2) of the last Lemma. That is, just assume that $X = m' + A^p$ has a second representation $X = n + B$. Then the process $n - m' = A^p - B$ satisfies the conditions of part (2) of the Lemma and so is evanescent. Therefore the representation is unique up to indistinguishability, and (1) holds.

The next step is to show that statement (4) follows from (1). Let $X = m + A$ be the "canonical decomposition" of (1) and $A^*(t) = \sup_{t \geq s} |\,A(s)\,|$, for all $t \geq 0$. Then $A^* \epsilon IV_{loc}^+$, since $A \epsilon IV_{loc}$. Let (T_n) be the localizing sequence for m and

$S_n := \inf(t{:}m^*(t)\geq n)$, where the * again indicates the supremum process. Then we can assume that $S_n\uparrow\infty$, so that $\min(S_n, T_n)\uparrow\infty$. Therefore, $m(S_n\wedge T_n)$ is P-integrable (just recall that m^{T_n} belongs to M_u). Hence, $m^*(S_n\wedge T_n)$ is bounded above by $n + |m(S_n\wedge T_n)|$, so that m^* is of local integrable variation. Since A^* is also in this family of processes we have that X^* is of local integrable variation and (4) holds.

The remaining parts, showing that (4) implies (3) and (3) implies (2), are, respectively, straightforward and trivial.

6.5.6. The following Corollary shows that any semi-martingale can be transformed into a special semi-martingale with uniformly bounded jumps:

6.5.7. **Corollary:**
Let X *be any semi-martingale,* $a>0$ *and* X^a *the process defined by setting*

$$X^a(t) = \sum_{s\leq t}\Delta X(s)\, 1_{[\,|\Delta X(s)|>a]}.$$

Then $X^a\epsilon BV$ *and* $X - X^a$ *is a special semi-martingale whose canonical decomposition,* $X - X^a = m + A$, *satisfies* $|\Delta m| \leq 2a$ *and* $|\Delta A| \leq a$.

Remark: Since this result will allow a second Corollary that is central to the construction of the stochastic integral, we will give its proof: By definition of semi-martingale, X is Skorokhod, so that the paths $t\rightarrow X^a(t,w)$ have only a finite number of jumps in any finite interval (Section A 1.1.2). Consequently, X^a is of bounded variation. Since adding a process of bounded variation to a semi-martingale returns a semi-martingale, $Y := X - X^a\epsilon S$. By construction $|\Delta Y| \leq a$. We use this fact to show that the supremum process corresponding to Y is an increasing process of local integrable variation, which by the Theorem demonstrates that Y is a special semi-martingale. Set $T_n = \inf(t{:}Y^*(t)\geq n)$. Then we can choose $T_n\uparrow\infty$, for if not then $0\leq Y^*(t)\leq n_0$ for some n_0 and all $t\geq 0$, and then the process is of increasing, integrable variation, so certainly of local integrable variation. On the other hand when $T_n\uparrow\infty$, then $0\leq Y^*(T_n)\leq n+a$, so that Y^* is an increasing process which is locally of integrable variation. Y is therefore a special semi-martingale (6.5.4).

Let $Y = m + A$ be the canonical decomposition of Y. We have $\Delta Y = \Delta m + \Delta A$. The idea of the proof is as follows: We know from Chapter 4, on previsible projections, that ${}^p\Delta Y = {}^p\Delta m + {}^p\Delta A$. Since A is previsible,

${}^p\Delta A = \Delta A$. It will be shown below that ${}^p\Delta m = 0$. Consequently, ${}^p\Delta Y = \Delta A$. Boundedness of the jumps of A follows from the Chapter 4 result that previsible projections preserve order: $|\Delta Y| \leq a$ implies ${}^p|\Delta Y| \leq {}^p a = a$. This gives the result that $|\Delta A| \leq a$. Immediately then $|\Delta m| \leq |\Delta Y| + |\Delta A|$, so that $|\Delta m| \leq 2a$, and the proof is complete except for justifying ${}^p\Delta m = 0$.

To see that the previsible projection of the jump process of m (i.e., $(\Delta m_t, t \geq 0)$) is evanescent, let T be a previsible time with announcing sequence (T_n). Then $T_n \uparrow T$, and $T_n < T$ on $[T > 0]$. Doob's Optional Stopping Theorem supplies the fact that $E((m(T) - m(T_n)) \mid F(T_n)) = 0$ so that (heuristically) letting $n \to \infty$, we obtain $E((m(T) - m(T-)) \mid F(T-)) = 0$ on $[0 < T < \infty]$, which says that the jump process has an evanescent previsible projection.

Now the much anticipated and important result.

6.5.8. **Corollary:**
If M *is a local martingale, then it has a decomposition,* $M = m' + m''$, *where* m' *is a local martingale vanishing at the origin, with uniformly bounded jumps,* $|\Delta m'| \leq 1$ *and* m'' *is a local martingale of local integrable variation.*

6.5.9. Remark: This decomposition is not unique. Since local martingales are, of course, semi-martingales, we can apply the last corollary to $M \in M_{loc}$ with $a = 1/2$. The result is $M = M^a + m + A$, where m and A are as in the definition of a special semi-martingale, $|\Delta m| \leq 2a = 1$ and A is previsible and locally of integrable variation. Set $m = m'$ and $m'' = M^a + A = M - m'$. Since M^a is of bounded variation, $m - m' = m''$ is of bounded variation. Since m'' is also a local martingale, Lemma 6.5.5 guarantees us that m'' is of local integrable variation.

Remark: We now discuss the structure of the class $K^{2,d}$ and obtain, as a consequence, information about the sums of jumps of any local martingale. Such results are needed in order to define the quadratic variation of local martingales and thence semi-martingales in the next Section.

Finally, we reformulate the previous decomposition theorem for local martingales into one whose summands are continuous and purely discontinuous local martingales. For us this will complete the geometrical picture of local martingales as sums of orthogonal processes. The main purpose for including it here, however, is that it can be used to obtain a corollary which gives us the important fact that the continuous part of any semi-martingale is unique.

6.5.10. Remarks: A rigorous discussion on the structure of the space of purely discontinuous square integrable martingales would require more space than is appropriate in this note. But certain facts can be explained. We start with martingales which are of bounded variation. Let m be such a martingale. Then we can show that

$$m_t = m_0 + \sum_{0<s\leq t} \Delta m_s - (\sum_{0<s\leq t} \Delta m_s)^p,$$

where the symbol p indicates that the last term on the right is the dual previsible compensator of the sum of the jump process $t \rightarrow \Delta m_t$ over the interval $(0,t]$. Thus, m is represented as a sum of **compensated jump martingales**.

The proof of this statement is quite easy and it also shows that the compensator (the dual previsible projection) is continuous: Just set $X = m - m_0 - J$, where $J := \sum_{0<s\leq t} \Delta m_s$. Then it is clear that X is continuous and so previsible. So $X^p = X$. Also, since $m - m_0$ is an IV_0 martingale, its dual previsible projection is evanescent (Theorem 1.6.14). Therefore, using the linearity of the dual previsible projection operator, we have that $X = X^p = -J^p$, which says that J^p is continuous and $m = m_0 + J - J^p$, which is the stated result.

It turns out that such martingales are dense in $K^{2,d}$. The usual way to establish this fact (Meyer [1976]) is to let T be a stopping time and define the subspace M[T] of $K^{2,d}$ to be those martingales which are continuous outside of the graph of T. In order to state the basic results, first consider the case where $T = 0$, a.s.P (remember that we are still under the "usual conditions", so T is indistinguishable from the zero stopping time). If $m \epsilon M[0]$, then $m - m_0$ is a square integrable, purely discontinuous martingale which is also continuous. Therefore, $m - m_0$ must be the zero martingale, and so m_t is the constant martingale equal to the random variable m_0 for all $t \geq 0$. (Remember the convention stated in the Outline, which set $m_{0-} = 0$, so that all members of the space of continuous martingales must satisfy the condition $m_0 = 0$.)

Therefore, suppose from now on that $T > 0$, a.s.P. So if $m \epsilon M[T]$, we must have $m_0 = m_{0-} = 0$ and so $M[T] \subset K_0^d$, when $T > 0$, a.s.P.

Now, let $m := A - A^p$, where $A := g\, 1_{[[T,\infty))}$, where g is a random variable with finite second moment, so that A is in IV and m is a martingale (Chapter 4), a **compensated jump martingale**.

There are two cases to treat: (i) T totally inaccessible and (ii) T previsible.

Consider case (i): Since A^p is previsible, its discontinuities, if any, are exhausted by a sequence of previsible times and by definition (of the term "exhaust", Chapter 2) it cannot charge any other stopping times; in particular, it cannot charge T, since T is totally inaccessible. Further, we now show that A^p cannot even charge any previsible time.

To see this, recall the language of Chapter 4, and the fact that the measures generated by A and A^p agree on G(PT), the σ-algebra of previsible events. Then notice that here, the support of the measure μ_A is [[T]]. So if A^p charges a previsible time, U, then the random set $[[U]] \bigcap [[T]]$ is not evanescent and so T is not totally inaccessible. This contradiction therefore tells us that A^p does not charge any previsible time. Hence, A^p must be continuous and so m is continuous outside of the graph of T.

Finally in case (i), with more effort than we want to expend here, it can be shown that $A_\infty^p \epsilon L_2$, so that this compensated sum of jumps martingale is square integrable. Therefore, $m \epsilon M[T] \subset K_0^{2,p}$.

Now in case (ii), with T previsible and a.s.P. positive, and with the additional assumption on g that $E(g \mid F(T)) = 0$, it can be shown that $(g\, 1_{[[T,\infty))})^p = 0$. Hence the compensated jump martingale, m, has the form $m = g\, 1_{[[T,\infty))}$ and belongs to M[T].

Therefore, with either of the assumptions on T and the corresponding assumptions on g, $m = g\, 1_{[[T,\infty))} - (g\, 1_{[[T,\infty))})^p$ (called a **compensated jump martingale**) is a martingale in M[T].

Further, it can be shown that for every n, $n \epsilon K^2$, the process

$$L = mn - \Delta m_T \Delta n_T 1_{[[T,\infty))}$$

is a uniformly integrable martingale which vanishes at the origin (belongs to M_0).

When n is continuous at T, this shows that $mn \epsilon M_0$, so that m is orthogonal to every n which is continuous at T. Since our compensated jump martingale is in K^2, we also have $m^2 - (\Delta m_T)^2\, 1_{[[T,\infty))} \epsilon M_0$. From this and the properties of uniformly integrable martingales,

$$Em_{\infty}^2 = E(\Delta m_T)^2. \tag{9}$$

We now apply these observations to an arbitrary n in K^2. Set $m = \Delta n_T 1_{[[T,\infty))} - (\Delta n_T 1_{[[T,\infty))})^p$. Then m is a compensated jump martingale with the property that n - m is continuous at T, and consequently, orthogonal to M[T]. m is therefore the projection of n onto m[T].

We can now state the principal result concerning the structure of purely discontinuous, square integrable martingales.

6.5.11. **Theorem:**
If $m \epsilon K_0^{2,d}$, *then m is the sum of a series of compensated jump martingales and m is orthogonal to every martingale* $n \epsilon K^2$ *which does not charge a jump of m (so m is orthogonal to every member of* $K^{2,c}$*).*

Remarks: By a Theorem in Chapter 2 there is a sequence of stopping times, (T_n), that exhaust the jumps of m. (Recall that the definition of exhaust includes the fact that the graphs of these stopping times are pairwise disjoint.) Further, since each stopping time can be decomposed into the sum of a totally inaccessible and an accessible stopping time (whose graphs are disjoint), and by definition each *accessible time is included in the union of a sequence of previsible times,* we can assume that each T_n is either totally inaccessible or previsible.

For each k, let $m^{(k)}$ be the compensated jump martingale associated with the stopping time T_k. Since the graphs, $[[T_n]]$, are pairwise disjoint, the $m^{(k)}$ are pairwise orthogonal (in $L_2(P)$, if you like).

Letting $U^{(k)} := \sum_1^k m^{(j)}$, we have that $m - U^{(k)}$ is continuous at the stopping times $T_1, \cdots, T_k$ and, therefore, orthogonal to $m^{(1)}, \cdots, m^{(k)}$. It follows that $m - U^{(k)}$ is orthogonal to $U^{(k)}$.

Therefore, if we write $m = U^{(k)} + (m - U^{(k)})$, square both sides and then take expectations of the result we have

$$\begin{aligned} Em_{\infty}^2 &= \sum_1^k E(m_{\infty}^{(j)})^2 + E(m_{\infty} - U_{\infty}^{(k)})^2 \\ &= \sum_1^k E(\Delta m_{T_n})^2 + E(m_{\infty} - U_{\infty}^{(k)})^2. \end{aligned}$$

(We have used equation (9).) It follows that $U_{(k)}$ converges to an element U of

$K^{2,d}$. (Recall a previous Lemma stating that $K^{2,d}$ is closed.) It is a simple matter to conclude therefore that $m - U \epsilon K^{2,c}$, and m - U is orthogonal to U. But since m is purely discontinuous, it follows that m - U is orthogonal to itself. Hence, m = U, and so

$$Em_{\infty}^{2} = \sum_{1}^{\infty} E(\Delta m_{T_n})^2.$$

This completes the "proof" of the Theorem.

Now if we take any element in K^2, not necessarily continuous or purely discontinuous, then we can carry out the same construction and write $m = U + (m - U)$. Therefore, we have a unique decomposition of m into its "continuous" and " purely discontinuous" parts. This decomposition also yields, as before,

$$Em_{\infty}^{2} = \sum_{1}^{\infty} E(\Delta m_{T_n})^2 + E(m_{\infty} - U_{\infty})^2.$$

But now m is not necessarily equal to U, so

$$Em_{\infty}^{2} \geq \sum_{1}^{\infty} E(\Delta m_{T_n})^2, \tag{10}$$

with equality holding iff $m \epsilon K^{2,d}$.

Returning to the decomposition of $m \epsilon K^2$, and writing $m^c = m - U$ and $m^d = U$, we can say that there exist $m^c \epsilon K^{2,c}$, $m^d \epsilon K^{2,d}$ such that m is uniquely decomposed into the sum $m = m^2 + m^d$.

This is a essentially a special case of a more general result about local martingales. To prove the more general statement directly, without the last version, Jacod first notes that any local martingale of bounded variation is in the family M_{loc}^{d} and, using the decomposition given in Corollary 6.5.8 above, reduces the proof of the decomposition of local martingales into their continuous and purely discontinous parts to a proof of this statement for members of K_0^2.

We state this result:

6.5.12. **Theorem:**
Let m *be a local martingale. Then there exist martingales* m^c *and* m^d *in* M^c_{loc} *and* M^d_{loc}*, respectively, such that* $m = m^c + m^d$ *. This decomposition of* m *into its continuous and discontinuous parts is unique.*

Remark: We have already mentioned the first of the following two results. The second will be needed in the construction of the stochastic integral for local martingales.

> (1) *Any local martingale of bounded variation is in the family* M^d_{loc}.
>
> *(2) Any two members of* M^d_{loc} *which have indistinguishable jumps processes are indistinguishable. The latter means simply that* $\Delta m = \Delta n$ *implies* $m = n$ *for* $m, n \epsilon M^d_{loc}$.

Remark: For the second statement, let $X = m - n$. Then the hypothesis of (2) says that the jump process of X is the zero process. By unicity of the decomposition theorem, X is then a continuous process which takes the value zero at the origin. Since X is purely discontinuous, this means that X is the zero process, or what is the same, $m = n$.

6.5.13. **Corollary**
Let any semi-martingale, X, have the representations

$$X = m + A \quad \text{and} \quad X = n + B$$

where $m, n \in M_{0,loc}$ *and* $A, B \in BV$. *Then* $m^c = n^c$.

Remark: Just note that $m - n = B - A \subset M^d_{0,loc}$.

Therefore, $0 = (m - n)^c = m^c - n^c$.

If the semi-martingale X decomposes as $X = m + A$, then we write $X^c = m^c$ and call X^c the **continuous part** of the semi-martingale. By the Corollary, X^c is independent of the decomposition. If $X \epsilon M_{loc}$, we set $X^c = m^c$ where m^c is given by the decomposition $m = m^c + m^d$ of the theorem itself.

Now, return to the inequality (10). This immediately yields the following.

6.5.14. Theorem

If $m \epsilon K^2$, *then* $\sum_{s \leq t} \Delta m_s^2 < \infty$, *a.s.P, for all* $t \geq 0$.

Remark: Following custom, we have used the following abbreviation: $\Delta m_s^2 := (\Delta m_s)^2$.

The extension of this result to local martingales is an immediate consequence of applying this theorem to the decomposition given in Corollary 6.5.8 and using an obvious property (explicitly stated below) of processes of integrable variation.

The following Corollary is necessary to prove the existence of the quadratic variation of a semi-martingale:

6.5.15. Corollary:

If $m \epsilon M_{loc}$, *then* $\sum_{s \leq t} \Delta m_s^2 < \infty$, *a.s.P, for all* $t \geq 0$.

6.5.16. Corollary:

If $X \epsilon S$, $X = m + A$, *then* $\sum_{s \leq t} \Delta X_s^2 < \infty$, *a.s.P, for all* $t \geq 0$.

Remark: Since A is in BV,

$$\sum_{s \leq t} \Delta A_s^2 \leq C \sum_{s \leq t} \Delta | A_s | < \infty$$

a.s.P, for some positive constant C, for all $t \geq 0$. The result follows from the previous Corollary by noting that

$$\Delta X_s^2 \leq 2(\Delta m_s^2 + \Delta A_s^2).$$

We will now give a result of Jacod which says that localization does not extend semi-martingales.

6.5.17. Theorem:

(1) S_p *is not extended by localization:* $S_p = (S_p)_{loc}$.

(2) S is not extended by localization: $S = S_{loc}$.

Remark: We will only prove (1). The main purpose is to illustrate what Dellacherie and Meyer [1981] call "pasting": a procedure for constructing a single process from segments of a sequence of processes.

As always, $S_p \subset (S_p)_{loc}$. So let $X \epsilon (S_p)_{loc}$ and (T_n) be a localizing sequence of X which **reduces** X to S_p, that is, such that $X^{T_n} \epsilon S_p$ for each n. Let the canonical decomposition be $X^{T_n} = m^{(n)} + A^{(n)}$ for each n. Since the T_n are nondecreasing, $T_{n+1} \wedge T_n = T_n$, so that $(X^{T_{n+1}})^{T_n} = X^{T_n}$. The uniqueness of the canonical decomposition allows the summands of the decomposition to inherit this property:

$$(m^{(n+1)})^{T_n} = m^{(n)},$$

$$(A^{(n+1)})^{T_n} = A^{(n)}.$$

The required local martingale, m, and previsible process of bounded variation, A, are obtained by pasting these path segments together, path by path, over all paths. Geometrically, it might help the reader to realize that equation (9), for instance, means that on $[0, T_n(w)]$, $m^{(n)}(w) = m^{(n+1)}(w)$. Thus, m and A with the required properties exist such $m^{T_n} = m^{(n)}$ and $A^{T_n} = A^{(n)}$, and $X = m + A$. Therefore, $X \epsilon S_p$, and so $S_p = (S_p)_{loc}$.

Remark: Thus, we have reached the end of the line in extending our processes by localization. That this is exactly the right place to stop in order to develop the stochastic integral will only be apparent after we complete the construction of the stochastic integral.

6.5.18. Remark: In Chapter 5 we gave a very general form of the Doob-Meyer Decomposition Theorem. We will now state this important result in a more restrictive and more easily proved form (see for example Ikeda and Watanabe). Then, using this and the results developed so far in this Section, we prove the Theorem as stated in Chapter 5. This will to some extent explain the central role this result plays in the modern theory of semi-martingales.

6.5.19. **Lemma (Doob-Meyer Decomposition for Class D submartingales):**

Let X be a supermartingale of the class D. Then there exists a unique, previsible, integrable increasing process, A, such that $X + A \epsilon M_0$ (is a uniformly integrable martingale which vanishes at the origin). Further, A is continuous iff X is quasi-left continuous.

6.5.20. Remark: Note that the Lemma does not involve localization. With Theorem 6.5.10 and Lemma 6.5.12, we can now state and prove the DMD

Theorem in a form equivalent to that of Chapter 5:

6.5.21. **Theorem (Doob-Meyer Decomposition):**
Every supermartingale (submartingale) is a special semi-martingale.

Remark: Because the proof (Jacod) is very elegant and gives application of some basic martingale results, we will give its outline. Set $T_n := \inf\{t : |X(t)| \geq n\}$ and $S_n := \min(n, T_n)$. Let $F = (F(t))$ be the underlying filtration. For each n, consider the stopped process, X^n, and notice that since X is an F-supermartingale, when t>n, $X_n = E(X_n \mid F_t)$, and when $n \geq t$, then $X_t \geq E(X_n \mid F_t)$. These two statements can be combined by writing $X_t^n \geq Y_t^{(n)} := E(X_n \mid F_t)$; thus, for each n, the uniformly integrable martingale, $Y^{(n)}$ is a minorant for the stopped process X^n. Therefore, Doob's Optional Sampling Theorem applies to the stopped process, X^{S_n}. That is, the process $X^{S_n} = (X^n)^{T_n}$ is an F-supermartingale. Further, since this process is majorized, for each n, by the random variable $n + (X^n_{S_n})^+$, it is a class D supermartingale. The previously stated class D form of the Doob-Meyer decomposition theorem then applies and X^{S_n} is a special semi-martingale. Since we know that the class of special semi-martingales is closed under localization, we have that X is a special semi-martingale.

The following also holds:

6.5.22. **Theorem:**
Every special semi-martingale is the difference of two local supermartingales (submartingales).

Doob's class D Lemma also applies to submartingales. The only change would be that we would have $X = m + A$ with A increasing.

6.6. **The Quadratic Variation Processes of a Semi-Martingale:** In various forms, we have mentioned that if m is a square integrable martingale, then m^2 is in the class D and, hence by Doob's Theorem, a previsible, increasing process of integrable variation exists which compensates m^2 into a martingale. In Chapter 1 we denoted this increasing process by $<m,m>$.

It may have escaped notice, but we have also proved this for the family of martingales, m, that are only locally square integrable. If $m \in K^2_{loc}$, then m^2 is a special semi-martingale. This is because $m \in K^2_{loc} \rightarrow m^2 \in (S_p)_{loc} = S_p$, by Theorem 6.5.10 Letting the associated previsible process, A, of the canonical decomposition be denoted by $<m,m>$ and observing that it is an increasing process since m^2 is a local submartingale, we have

6.6.1. **Lemma:**
If m *is a locally square integrable martingale, then there exists a locally integrable, increasing, previsible process,* $<m,m>$, *such that* $m^2 - <m,m> \in M_{0,loc}$.

As in Chapter 1, we call process $<m,m> = (<m,m>(t), t \in R_+)$ the **previsible quadratic variation** of m. In our notation, $<m,m> \in (IV^+)_{loc}$.

When m and n belong to K^2_{loc}, the process $<m,n>$ is defined by polarization, as in Chapter 1. It is immediate that the mappings $m \rightarrow <m,n>$ and $n \rightarrow <m,n>$ are linear. Indeed, for m,n,l,k in K^2_{loc} and a,b,c,d real numbers,

$$<am+bn,cl+dk> = ac<m,l> + ad<m,k> + bc<n,l> + bd<n,k>.$$

6.6.2. **Theorem**:
If $m,n \in K^2_{loc}$, *then* $<m,n>$ *is the unique previsible process in* $(IV)_{loc}$ *such that* $mn-<m,n>$ *belongs to* $M_{0,loc}$.

6.6.3. Remark: It might be worthwhile to first consider the case where m and n are square integrable martingales ($m,n \in K^2$). The Theorem then follows by noting that $(m+n)^2 - <m+n,m+n>$ is a martingale and equals the sum of the following three martingales:

$$m^2 - <m,m>,\ n^2 - <n,n>,\ 2\{mn - (<m+n,m+n> - <m,m> - <n,n>)/2\},$$

so that the last term in the braces must a martingale. The conclusion that $mn - <m,n>$ is a martingale follows from the uniqueness of the Doob-Meyer decomposition.

The reader should note that the conclusion of the remark says that if you start with martingales you end up with a martingale, not just a local martingale.

A proof (Jacod [1979]) that gives the generality of the Theorem follows by recognizing the product mn as a special semi-martingale. This is because writing

$$mn = \frac{1}{2}((m+n)^2 - (m-n)^2)$$

expresses the product mn as the difference of two submartingales. Hence, mn is in S_p, by Theorem 6.5.15.

As in Chapter 1, the process, $<m,n>$, is called the **covariance process** of m and n.

6.6.4. Remark: Again take m and n to be square integrable martingales. An easy computation, based on the last Remark, yields

$$E\{m(t)n(t) \mid F(s)\} - m(s)n(s) = E\{<m,n>(t) - <m,n>(s) \mid F(s)\}.$$

This equation states that *the product mn is a martingale iff* $<m,n>(t) = 0$ *for all* $t \geq 0$. Recall the earlier discussion on orthogonality of martingales and store for later purposes the fact that $<m,n> = 0$ if m is continuous and n is a compensated jump martingale.

6.6.5. We now define the (optional) **quadratic variation** and the (optional) **cross quadratic variation** of semi-martingales, X and Y.

$$[X,X]_t := <X^c, X^c>_t + \sum_{0 \leq s \leq t} (\Delta X(s))^2 \tag{11}$$

$$[X,Y]_t := <X^c, Y^c>_t + \sum_{0 \leq s \leq t} \Delta X(s)\, \Delta Y(s) \tag{11.1}$$

for all t in R_+.

6.6.6. Remark: From the definition of $< , >$ on K^2_{loc}, $<m,n>$ is well-defined for m and n in M^c_{loc}, since $M^c_{loc} \subset K^2_{loc}$. Section 5.1.11 gives a proof of the fact that any continuous local martingale is an L_p local martingale, for any $p \geq 1$. Hence, the first terms on the right side of equations (11) and (11.1) are well-defined. Corollary 6.5.16 of the last Section then shows that [X,X] is well-defined. It follows easily that [X,Y] makes sense.

Having observed that $M^c_{loc} \subset K^2_{loc}$, the following example due to C. Stricker (Dellacherie and Meyer [1980]) of a local martingale that is not locally square integrable is probably worth the interruption. Define the filtration (F(t)) on the probability space, (Ω,H,P), as $F(t) = \{\phi,\Omega\}$, for $0 \leq t < 1$, and $F(t) = H$, for $t \geq 1$. Then $X(t) = E(h|F(t))$, where $h \epsilon (L_1 - L_2)$, is such an example.

6.6.7. It can be shown that $<X^c,X^c>$ is always continuous. (Actually, the previsible compensator of the submartingale in the DMD Theorem is continuous iff the process is quasi-left continuous, which is true if the process is continuous.) Therefore, [X,X] will be a continuous process iff the sum vanishes, that is, iff X is continuous.

Of course, in the cross quadratic variation, the sum is the zero process if X and Y have no common jumps and clearly $[X,Y] = 0$ if one of the factors is continuous and the other is purely discontinuous, since then $<X^c,Y^c> = 0$.

We will not take time to prove the following important Theorem.

6.6.8. **Theorem:** *Let m and n be local martingales. Then*

(1) $[m,n]$ *is a process of bounded variation and*

(2) $mn - [m,n] \in M_{0,loc}$

6.6.9. Remark: We see from this Theorem that m and n are (strongly) orthogonal iff $[m,n]$ is a member of $M_{0,loc}$, for then $mn \in M_{0,loc}$.

Our main purpose, however, in stating this Theorem is that it is but a short step to the result that $[X,X]$ is a member of V^+. For let $X = m + A$, with m in $M_{0,loc}$ and A in BV. Then, as shown in the proof of Corollary 6.5.16, the series with terms $(\Delta A)^2$ converges, so that the process defined by $\sum_{0 \leq s \leq t} (\Delta A(s))^2$ belongs to V^+. By (1) of the previous Theorem, $[m,m]$ is in V^+. Hence, $[X,X] \in V^+$, and we have the following Theorem.

6.6.10. **Theorem:**
If X is a semi-martingale, then $[X,X]$ *is an increasing process.*

6.6.11. Remark: We will list a few of the consequences stemming from this result. Let X and Y be semi-martingales. Then

(1) $[X,Y] \in (BV)$;

(2) $(X,Y) \to [X,Y]$ is bilinear;

(3) If T is a stopping time, then
$$[X,Y]^T = [X^T,Y^T] = [X^T,Y] = [X,Y^T]$$

6.6.12. Remark: If $m,n \in K^2_{loc}$, then we have seen that $mn - <m,n>$ and $mn - [m,n]$ both belong to $M_{0,loc}$. Therefore, $[m,n] - <m,n> \in M_{0,loc}$ also. Since $[m,n] \in (IV)_{loc}$, we have $<m,n> = [m,n]^p$. That is, $<m,n>$ is the previsible projection of $[m,n]$. For $X,Y \in S$, we therefore can extend the definition of $< \, , >$ by setting $<X,Y> := [X,Y]^p$, whenever $[X,Y] \in (IV)_{loc}$.

6.6.13. As might be expected, to complete the construction of the stochastic integral we require inequalities analogous to the Cauchy-Schwartz inequality. The form of the factors on the right side of the second of these inequalities should be noted in order to understand the selection of a norm for $L_p(m)$, defined below. The following is due to Kunita and Watanabe [1967]:

6.6.14. **Theorem (Kunita-Watanabe Inequality):**
If H *and* K *are optional processes and* $m,n \epsilon K^2$, *then*

$$\int_0^\infty |H(s)|\,|K(s)|\,|d[m,n]_s| \leq \left(\int_0^\infty H^2(s)\,d[m,m]_s\right)^{\frac{1}{2}} \left(\int_0^\infty K^2(s)\,d[n,n]_s\right)^{\frac{1}{2}}.$$

If $p>1$ *and* q *is the conjugate of* p, *then*

$$E\left(\int_0^\infty |H(s)|\,|K(s)|\,|d[m,n]_s|\right) \leq \|(H^2 \cdot [m,m])_\infty^{\frac{1}{2}}\|_p \, \|(K^2 \cdot [n,n])_\infty^{\frac{1}{2}}\|_q.$$

Remark: If n and m are continuous, then we can replace [,] with $< , >$. In fact, the inequality was originally proved in terms of $< , >$.

The following remarkable Theorem shows that with $p>1$ the norm $\| m^*(\infty) \|_p$ and the norm $\| [m,m]^{\frac{1}{2}}(\infty) \|_p$ are equivalent. In particular, this means that we can define the space K^2 of square integrable processes in terms of the L_2 norm of $\sqrt{[m,m]}$.

6.6.15. **Theorem (Davis, Burkholder, Grundy):**
Let $p\epsilon[1,\infty)$. *Then there exist positive constants* c_p *and* c_p' *such that for each* $m \epsilon M_{loc}$

$$c_p \|m_\infty^*\|_p \leq \| [m,m]_\infty \|_p \leq c_p' \| m_\infty^* \|_p.$$

6.7. **Stochastic Integrals Relative to Continuous Local Martingales:** Let m be a local martingale and $p\epsilon[1,\infty]$. As usual $\|.\|_p$ denotes the L_p norm: $\| f \|_p^p := E(|f|^p)$. Set

$$\| H \|_{p,m} := \| (H^2 .[m,m](\infty))^{\frac{1}{2}} \|_p$$

and $L_p(m) := \{ H \text{ previsible} : \| H \|_{p,m} < \infty \}$. The Kunita and Watanabe Inequality shows that $L_p(m) \subset L_q(m)$ if $q \leq p$.

6.7.1. After our discussion on localization of integrators in the introduction to this Chapter, we noted that localization for integrands would be carried out differently than for integrators. Let $p \epsilon [1,\infty)$, and $L_{p,loc}(m)$ denote the set of all previsible processes for which there exists an increasing sequence, $T_n \uparrow \infty$, of optional times such that $H \ 1_{[[0,T_n]]} \epsilon L_p(m)$. Notice that this type of localization is a natural choice for integrands. Attempting to integrate constants other than zero over unbounded sets relative to σ-finite measures tends to produce undesirable results.

6.7.2. Suppose that $m \epsilon M^c_{0,loc}$. Then the Lebesgue-Stieltjes stochastic integral $H^2.[m,m]$ is continuous, and the increasing processes $t \rightarrow \sqrt{H^2.[m,m](t)}$ is continuous and vanishes at the origin. Therefore, this process is of bounded variation iff it is locally bounded. Therefore, under the assumption that $m \epsilon M^c_{0,loc}$, it can be shown that $L_{p,loc}(m) = L_{1,loc}(m)$ for all $p \geq 1$.

To define the stochastic integral for H in $L_{1,loc}(m)$ relative to $m \epsilon M^c_{0,loc}$ it is therefore sufficient to define it for $L_2(m)$.

6.7.3. Let $H \epsilon L_2(m)$. Consider the linear transformation on K^2 defined by

$$n \rightarrow C(n) := E((H.[m,n])(\infty)). \tag{12}$$

Set

$$||| n ||| = \| n(\infty) \|_2,$$

the norm on the Hilbert space, K^2, equipped with the inner product $(m,n) \rightarrow Em(\infty)n(\infty)$. Then according to the Kunita-Watanabe Inequality with K = 1 and using equivalence of norms, we have that $| C(n) |$ is bounded above by $\| H \|_{2,m} \ ||| n |||$. This shows that C is continuous on K^2.

But as a continuous, linear functional on a Hilbert space, there exists a unique process $Y \epsilon K^2$ with the property that $E(Y(\infty)n(\infty)) = C(n)$ for all K^2. Recalling equation (12), this remark justifies the following

6.7.4. **Definition:** If $H \epsilon L^2(m)$, we call the **stochastic integral** of H relative to $m \epsilon M_{0,loc}^c$, denoted by H.m, the unique element of K^2 such that

$$E((H.m)_\infty)\, n_\infty) = E(H.[m,n])_\infty, \tag{13}$$

for all n in K^2.

Having justified the definition of the stochastic integral, we now give a result which characterizes it.

6.7.5. **Theorem (Characterization of H.m on** $M_{0,loc}^c$**):**
If $H \epsilon L^2(m)$ *and* $m \epsilon M_{0,loc}^c$, *then* $H.m \epsilon K_0^{2,c}$ *and* H.m *is the unique element of* K^2 *such that*

$$[H.m, n] = H.[m, n], \tag{14}$$

for all $n \epsilon K^2$.

6.7.6. Remark: Thus, if Y is a solution of [Y,n]=H.[m,n], an equation equating a process of bounded variation and a Lebesgue-Stieltjes integral, then Y = H.m . If we define $\int_0^t H(s)\, dm(s) = (H.m)(t)$, for all $t \geq 0$, the equation (14) takes on the following form:

$$[\int_0^{\cdot} H(s)\, dm(s),\ n](t) = \int_0^t H(s)\, d[m,n](s). \tag{14.1}$$

6.7.7. Remark: This Theorem is due to Kunita and Watanabe. The proof given here is from Jacod. From earlier remarks, we know that $Y = H.m \epsilon K^2$. Letting C be as defined above, $C(Y^d) = E(Y(\infty) Y^d(\infty))$ and, since the discrete and continuous parts of Y are orthogonal, we have

$$C(Y^d) = E((Y^c(\infty) + Y^d(\infty))\, Y^d(\infty)) = E((Y^d(\infty))^2).$$

But since m is continuous, we have $[m, Y^d] = 0$, the zero process. Therefore, by definition of stochastic integral $C(Y^d) = E(H.[m,Y^d]) = 0$ and consequently $E((Y^d(\infty))^2) = 0$. Therefore, $Y^d = 0$, the zero process (it is clear that we are working with equivalence classes), and so Y is continuous and $Y_0 = 0$ (the latter with the convention $Y_{0-} = 0$). That is, $Y \epsilon K_0^{2,c}$.

Next, for each optional time, T, and each $n \epsilon K^2$ we know that $Y n^T - [Y,n^T] \epsilon K_0^1$ and $[Y,n^T] = [Y,n]^T$; hence,

$$\begin{aligned} E([Y,n](T)) &= E\, Y(T)\, n(T) = E(E(Y(\infty) \mid F_T))\, n(T)) \\ &= E(Y(\infty)\, n^T(\infty)) = C(n^T(\infty)) = E(H.[m,n^T](\infty) \\ &= E(H.[m,n]^T(\infty)) = E(H.[m,n](T)). \end{aligned}$$

That is,

$$E[Y,n](T) = EH.[m,n](T)$$

for any $n \epsilon K^2$ and optional time T.

It follows from Theorem 6.6.8 that $[Y, n] - H.[m,n] \epsilon M_0$. But $[Y, n] - H.[m,n]$ is a process of bounded variation which is previsible (the latter since Y and m are continuous), so that $[Y, n] - H.[m,n] = 0$, the zero process. This proves the theorem in one direction.

Conversely, let $Y \epsilon K^2$ and satisfy $[Y, n] = H.[m,n]$ for all $n \epsilon K^2$. Then $E[Y,n](\infty) = E(H.[m,n](\infty)) = C(n) = E(Y(\infty)n(\infty)$. Therefore, $E[Y,n](\infty) = E(Y(\infty)n(\infty))$ for all $n \epsilon K^2$. Then, by definition, $Y = H.m$, completing the proof.

Remark: We have observed that when m is continuous, H.m is continuous. Hence, $[H.m,n] = \langle H.m,n \rangle$ and we remarked earlier that m continuous gave us $[m,n] = \langle m,n \rangle$ so under the assumption of the theorem, equation (14) can be expressed as

$$\langle H.m,n \rangle = H.\langle m,n \rangle. \qquad (14.2)$$

Further, from the properties of [,], we can show

6.7.8. **Corollary:**
$(H.m)^T = H.m^T = H\, 1_{[[0,T]]}.m$ *for all optional times T.*

6.7.9. Remark: Let $H \epsilon L_{1,loc}(m)$ and $T_n \uparrow \infty$, a localization of H relative to $L_2(m)$. (No, the 2 is not a mistake: recall that $L_{1,loc} = L_{2,loc}$.) That is, $H\, 1_{[[0,T_n]]} \epsilon L_2(m)$. The previous Corollary provides us with a way of extending the definition of H.m to H in $L_{1,loc}$ by setting $(H.m)^{T_n} = (H\, 1_{[[0,T_n]]}).m$, for each n. The result is called the **stochastic integral of H relative to m** . Thus, the stochastic integral has been defined for $H \epsilon L_{1,loc}$ and $m \epsilon M^c_{0,loc}$. It satisfies a characterization analogous to that stated in the last Theorem and the same equations as given in the Corollary.

6.8. Stochastic Integrals Relative to Local Martingales:

6.8.1. **Definition**: If $H \epsilon L_{1,loc}(m)$ and $m \epsilon M_{loc}$, then the **stochastic integral** H.m of H relative to m is the unique element of M_{loc} which satisfies $(H.m)^c = H.m^c$, $\Delta(H.m) = H\Delta m$.

6.8.2. Remark: Recall that if m is a continuous local martingale then the paths of m are of unbounded variation. So, in this case, H.m should never be mistaken for the Lebesgue-Stieltjes (pathwise) integral of H relative to m. We know that such objects do not exist.

Hence, until this Section, either $m \epsilon M^c_{loc}$ or $m \epsilon M_{loc} \bigcap BV$ and *so the stochastic* integral H.m was either that of the last Section or the Lebesgue-Stieltjes stochastic integral, respectively. The last definition considers $m \epsilon M_{loc}$ so now the possibility of an inconsistency in our definition of H.m arises. Jacod shows that Y = H.m as just defined cannot have two distinct meanings. He argues as follows: If $m \epsilon M_{loc} \bigcap BV$, $H \epsilon L_1(m)$, and $n(t) = \int_0^t H(s)\, dm(s)$ exists as a Lebesgue-Stieltjes integral, then $n \epsilon (IV)_{loc}$. This can be shown to imply that $n \epsilon M^d_{loc}$. But $m \epsilon M_{loc} \bigcap B \subset M^d_{loc}$ so that $m^c = 0$. But then, by definition, $Y^c = 0$ ($Y^c = H.m^c$). So $Y \epsilon M^d_{loc}$ also. Now recall the "you know them by their jumps" description of M^d_{loc} given earlier. Using the facts that $\Delta n = H\Delta m$ and, by definition, $\Delta Y = H\Delta m$, we conclude that $Y = n$.

Remark: We have just noted that the stochastic integral of this Section reduces to the Lebesgue-Stieltjes stochastic integral when $m \epsilon IV_{loc}$ so that it is important to realize that contrary to the case of the Lebesgue Stieltjes integral, H.m is not defined pathwise; its definition depends on the underlying probability and filtration.

6.8.3. **Theorem (Characterization of H.m on M_{loc}):**

> *(1) Let* $m \epsilon M_{loc}$ *and* $H \epsilon L_{1,loc}(m)$. *Then* H.m *is the unique element of* M_{loc} *which satisfies* $[H.m,n] = H.[m,n]$ *for all* $n \epsilon M_{loc}$.
>
> *(2) In order that* $H.m \epsilon K^p$ *(respectively* K^p_{loc}*) it is necessary and sufficient that* $H \epsilon L_p(m)$ *(respectively* $L_{p,loc}$*)*

6.8.4. Remark: Part 1 echoes the characterizations of stochastic integrals on M^c_{loc}. Thus, the definition of H.m on M_{loc} is consistent with the definition on M^c_{loc}. Part 2 says that the "size" of the integral is directly related to the "size" of the integrand. This result is not surprising when the definition of $L_p(m)$ is recalled.

6.8.5. Remark: A sketch of the proof that $[H.m,n] = H.[m,n]$ is as follows. By definition of [,], $[H.m,n] = \langle (H.m)^c, n^c \rangle + \sum (\Delta H.m)\Delta n$. By definition of H.m, $(H.m)^c = H.m^c$, so that $\langle (H.m)^c, n^c \rangle = \langle H.m^c, n^c \rangle$. The latter equals $[H.m^c, n^c]$, which by the last characterization for continuous local martingales equals $H.[m^c,n^c] = H.\langle m^c,n^c \rangle$. Finally, since $\Delta H.m = H\Delta m$, we have $[H.m,n] = H.\langle m^c,n^c \rangle + \sum H\Delta m \Delta n = H.(\langle m^c,n^c \rangle + \sum \Delta m \Delta n) = H.[m,n]$. For the converse we must show that if $Y \epsilon M_{loc}$ and $[Y.n] = H.[m,n]$ for all $n \epsilon M_{loc}$, then $Y^c = (H.m)^c = H.m^c$ and $\Delta Y = H\Delta m$. The interested reader should just write $Y = Y^c + Y^d$, and $m = m^c + m^d$ and proceed, or see Jacod.

6.9. **Stochastic Integrals Relative to Semi-Martingales**

6.9.1. For simplicity, let H be a bounded previsible process. Let $X \epsilon S$ and have the decompositions $X = m + A = n + B$ with the usual meanings. By the previous two Sections, the stochastic integrals H.m, H.n and the Lebesgue Stieltjes integrals H.A, H.B are well-defined. Since, $m - n = B - A$ is a local martingale of bounded variation, we know by the consistency of the stochastic and Lebesgue Stieltjes stochastic integrals that $H.(m - n) = H.(B - A)$. Therefore, the formula $H.X = H.m + H.A$ defines the stochastic integral of a semi-martingale and this definition is independent of the choice of the decomposition. The resulting expression H.X is called the **stochastic integral** of H relative to X.

6.9.2. Remark: Properties specifically derived or implied in these Sections are summarized (with a minimum of special notation) in the following Portmanteau Theorem. The first part of the Theorem contains a result referred to in the introduction concerning the extension of the stochastic integral beyond the class of semi-martingales. It also includes a stochastic integral version of Lebesgue's Dominated Convergence Theorem and relates the elementary stochastic integral discussed in the Outline in Section 6.2.2 to the integral developed in this Chapter.

For additional details on the construction of the stochastic integral the reader should consult Dellacherie-Meyer [1982,313], Jacod [1979], and Dellacherie [1978].

Let the space Ξ of elementary processes introduce in 6.2.3. and equip Ξ with the topology of uniform convergence. Denote by $L_0 = L_0(F,P)$ the space of finite measurable functions equipped with the topology of convergence in probability, P.

6.9.3. **Theorem (Portmanteau)**
(1) Let X be fixed Skorokhod process and H.X denote the elementary stochastic integral of H relative to X. Then the mapping H→H.X, from Ξ to L_0, defined by

$$H(t) \rightarrow \int_0^t H(s)\ dX(s) := H.X(t) \qquad (*)$$

for each non-negative t, is continuous iff X is a semi-martingale.

(2) Let X be a semi-martingale. The mapping from Ξ into L_0 defined by (*) can be extended uniquely to the space of all bounded, previsible processes in such a way that (retaining the notation H.X), the mapping H→H.X is linear, the process H.X is Skorokhod and the following properties hold:

> *(a) (Lebesgue): If the sequence (Y_n) of bounded measurable processes converges pointwise to a process, Y, and the Y_n are dominated in absolute value by a bounded previsible process, then Y is a bounded previsible process and the sequence $Y_n.X$ converges in probability to Y.X .*
>
> *(b) For every bounded, previsible H, H.X is a semi-martingale. Also, if X is a special semi-martingale, then H.X is a special semi-martingale.*
>
> *(c) For every H in Ξ, the (extended) stochastic integral, H.X, is an elementary integral.*

(d) If X is of bounded variation and H is bounded and previsible then the (extended) stochastic integral, H.X, is indistinguishable from the stochastic integral, H.X, defined pathwise by $H.X(t,w) = \int_{[0,t]} H(s,w)\, dX(s,w)$, *for each* $w \in \Omega$, *as a Lebesgue-Stieltjes integral.*

(e) If H and K are bounded previsible processes, then $K.(H.X) = (KH).X$, *and* $\Delta(H.X) = H\, \Delta X$.

(f) If T is a stopping time,

$$(H.X)^T = (H\, 1_{[[0,T]]}.X) = (H.(1_{[[0,T]]}.X)) = H.X^T$$

(h) If H is a bounded, previsible process, then H.X is a martingale, local martingale or process of bounded variation, if X is one of these processes.

6.10. **Local Characteristics of Semi-Martingales:** In this Section we will only add a few remarks to what has already been written with the aim of showing some relationships between several of these concepts and with a portion of the classical theory stochastic processes (processes with independent increments). Recall Corollaries 6.5.7 and 6.5.8. If $X \in S$ and we define

$$Y_t := X_0 + \sum_{s \leq t} \Delta X_s 1_{[|\Delta X_s| > 1]}, \tag{15}$$

then X-Y is a semi-martingale with bounded jumps and hence a special semi-martingale. Therefore,

$$X - Y = m + \alpha,$$

where m is a local martingale with uniformly bounded jumps and α is a previsible process of integrable variation. Both m and α vanish identically at time zero. Decompose m into its continuous and purely discontinuous parts, $m = m^c + m^d$, and recall that $X^c := m^c$. It follows that

6.10.1. **Lemma**:
If $X \in S$, *then X can be written in the form*

$$X_t = X_0 + \alpha_t + X_t^c + Y_t + m_t^d \tag{16}$$

and this representation is unique.

Remark: Let μ be the saltus measure of X (Chapter 4):

$$\mu(w,dt,dz) = \sum_s 1_{[\Delta X_s \neq 0]}\epsilon_{(s,\Delta X_s)}(dt,dz), \tag{17}$$

and ν be the dual previsible projection of the random measure, μ. Setting $\beta = \langle X^c, X^c \rangle$, the triple (α,β,ν) is called the **triple of P-local characteristics** of the semi-martingale X. This triple is uniquely determined by the semi-martingale X, to within a P-null set. But while β and ν are intrinsic characteristics of X, the component α depends on the "truncation point" in the definition of Y in (15). Therefore, the triple does not characterize the semi-martingale X.

In Chapter 4 integration relative to a random measure was taken in the sense of a Lebesgue-Stieltjes integral. But we also noticed there that, $\mu((0,t]\times B) - \nu((0,t]\times B)$ is a local martingale, for each $B\epsilon\xi$, $t\geq 0$. In fact, it is a purely discontinuous local martingale. So, if we want to integrate relative to $\mu-\nu$, we need to at least recognize the fact that yet another stochastic integral is required. We will not go into the construction of this type of stochastic integral, but recommend Part I of the 1978 paper by Kabanov, Liptser and Shiryayev in the Sbornik or Jacod[1979,p.96].

With the aid of this stochastic integral which we will denote by $\int \cdots d(\mu-\nu)$ and with m^d as in (16), Kabanov et al show that

$$m^d = \int_0^t \int_{\{x:|x|\leq 1\}} x\,d(\mu-\nu). \tag{18}$$

We can also write the process Y in (15) (as a path integral) in terms of the saltus measure μ of X:

$$Y_t - X_0 = \int_0^t \int_{\{x:|x|>1\}} x\mu(ds,dx). \tag{19}$$

Thus, we can state the following

6.10.2. **Theorem:**
If $X\epsilon S$, with saltus measure μ and local characteristics (α,β,ν), then

$$X_t = X_0 + \alpha_t + \beta_t + \int_0^t \int_{\{x:|x|>1\}} x\mu(ds,dx) + \int_0^t \int_{\{x:|x|\leq 1\}} x\,d(\mu-\nu). \tag{20}$$

and this representation is a.s.P unique.

This representation of a semi-martingale allows one to relate semi-martingales to P. Levy's remarkable theory of processes with independent increments

(Lévy[1937] and Loève[1960]).

6.10.3. **Definition**: A process with **independent increments** (II) on a filtered probability space (Ω,H,F,P) is a Skorokhod process X adapted to F such that for each pair (s,t) with $0 \leq s \leq t < \infty$ the random variable $X_t - X_s$ is probabilistically independent of F_s. Further, a process X with II is said to be a process with **stationary independent increments** (SII) if $X_0 = 0$ and $X_t - X_s$ has the same distribution as X_{t-s}, $0 \leq s \leq t < \infty$.

Remark: The most famous examples of processes with stationary independent increments are Brownian motion and the Poisson process. **Standard Brownian motion**, also called the **standard Wiener process**, is a process B with the properties that B is F adapted and for each pair of numbers (s,t), $0 \leq s \leq t < \infty$, the random variable $B_t - B_s$ has a normal distribution with zero expectation, variance $t - s$ and is independent of F_s.

Since

$$E(B_t^2 - B_s^2) \mid F_s) = E(B_t - B_s)^2 \mid F_s) = E(B_t - B_s)^2 = t-s,$$

it follows that $\langle B,B \rangle_t = t$, $t \geq 0$. Notice that this also shows that $(B_t^2 - t, t \geq 0)$ is an F-martingale. B is obviously an F-martingale also and it can be shown that P-almost all of its paths are continuous. So the paths of Brownian motion are of unbounded variation with probability one.

Any reference to a Brownian motion process will mean a process X such that $X_t = mt + \sigma B_t$, where m is any real number, $\sigma > 0$, and B is standard Brownian motion.

Not only is B a process with stationary independent increments, but if X is any SII process which is a.s.P continuous then X is Brownian motion ($X = mt + \sigma B$). That is, every a.s.P continuous process with stationary independent increments is a Brownian motion process.

Since standard Brownian motion is a martingale it follows that every Brownian motion process is a semi-martingale. Poisson processes are submartingales, so they are also semi-martingales by the Doob-Meyer Decomposition Theorem. But not every process with independent increments is a semi-martingale. Jacod[1979] shows that a process with independent increments is a semi-martingale iff the function $t \rightarrow Ee^{iuX_t}$, u, t real, has finite variation on compact sets.

Remark: At this point it might be of some interest to readers of this note to glance back at Loève's 1960 book on probability. Specifically, refer to Section 22 where the classical Central Limit Problem is defined and recall the role played by "infinitely divisible" random variables in the solution of this problem. Then turn to Section 37 and look at the definition of a "decomposable" random function (stochastic process); this is a process with independent increments. Some of the principal results there indicate the beginnings of the modern theory of semi-martingales and random measures.

Remark: Jacod (Jacod[1979, 90-95]) shows that semi-martingales which are processes with independent increments have deterministic local characteristics (i.e., there exists a version of the triple (α,β,ν) that does not depend on $w\epsilon\Omega$) and conversely only semi-martingales with II have this property. When the local characteristics have the additional property that α and β are linear in t and ν is a particular product measure on $(0,\infty)\times R$, then these processes are also stationary. This provides a useful link between the classical and modern theories of stochastic processes.

6.11. **Ito's Formula and Applications to Brownian Motion:** We will limit our discussion of Ito's formula to processes with continuous paths. Stochastic integrals relative to this type of process are the most studied because of their close connection to Brownian motion and stochastic differential equations.

6.11.1. Remark: Let the function $K:R\rightarrow R$ have continuous second order derivatives. Let m be a continuous function on R_+. Then using a finite Taylor series expansion applied to the increments of K, we have

$$K(m_t) - K(m_0) = \sum_{k=0}^{n-1}(K(m_{t_k}) - K(m_{t_{k-1}}))$$

$$= \sum^{n-1} K'(m_{t_{k-1}})\Delta m_{t_k} + \frac{1}{2}\sum^{n-1} K''(m_{t_{k-1}})(\Delta m_{t_k})^2 + r_n^{(2)}.$$

(E.g., $t_k = t_k^{(n)} = tk/n$, so that $0 = t_0 < t_1 < \ldots < t_n = t$.)

If m is of finite variation (in addition to being continuous), the remainder $r_n^{(2)}$ and $\sum_{k=0}^{n-1}(\Delta m_{t_k})^2$ converge to zero and we have the usual change of variable formula for Stieltjes integrals:

$$K(m_t) - K(m_0) = \int_0^t K'(m_s)dm_s$$

or symbolically $dK(m_t) = K'(m_t)dm_t$.

Now, if we replace the function m_t by a standard Brownian motion, B, a continuous process of unbounded variation, it can be shown that

$$\sum_{k=0}^{n-1} (\Delta B_{t_k})^2 \rightarrow <B,B>_t \; = \; t, \text{ a.s.P,} \qquad (*)$$

and $r_n^{(2)} \rightarrow 0$ as $n \rightarrow \infty$. Further, both the sequence of sums $\sum^{n-1} (\Delta B)^\nu$ and the remainders r_n^ν converge to zero for each $\nu \geq 3$. Therefore we would expect the change of variable formula for stochastic integrals with Brownian integrators to be of the form

$$K(B_t) - K(B_0) \equiv \int_0^t K'(B_s)dB_s + \frac{1}{2}\int_0^t K''(B_s)ds.$$

This is Ito's original formula for Brownian motion. When B is replaced by continuous local martingale M, equation (*) continues to hold but the limit is the process $<M,M>$, the compensator of the submartingale M^2. We will show below that the process $<M,M>$ is distinguishable from $<B,B>$ unless M=B. So for any continuous local martingale one would expect that the change of variables formula for stochastic integrals would become

$$K(M_t) - K(M_0) = \int_0^t K'(M_s)dM_s + \frac{1}{2}\int_0^t K''(M_s)d<M,M>_s.$$

This is the claim of the next Theorem.

Let (Ω,H,F,P) be a filtered probability space. Take m to be a continuous local martingale and recall that $M_{loc}^c \subset K_{loc}^2$, so that $<m,m>$ exists. We will say that X is a **continuous semi-martingale** if X = m + A, with m as specified above and A a continuous process of bounded variation on finite intervals. Then the following form of Ito's change of variables formula holds (Kunita, Watanabe[1967]; Meyer[1976]):

6.11.2. **Theorem:**
Let X be a continuous semi-martingale and K be a function mapping $R \rightarrow R$ and having continuous second derivatives on R. Then the process Y=K(X) is a semi-martingale and (up to indistinguishability)

$$K(X_t) - K(X_0) = \int_0^t K'(X_s)\,dX_s + \frac{1}{2}\int_0^t K''(X_s)\,d<X,X>_s. \qquad (21)$$

This change of variables formula is often written in the purely symbolic form of "differentials": $dK(X) = K'(X)dX + \frac{1}{2}K''\, d<X>$, but this only has meaning in terms of the integral equation in (21).

Remark: Although we will consider only continuous processes in this Section, it is informative to see how theorem changes in the case of an arbitrary semi-martingale:

$$K(X_t) - K(X_0) = \int_0^t K'(X_{s-})\, dX_s + \frac{1}{2}\int_0^t K''(X_{s-})\, d<X^c,X^c>_s \qquad (21*)$$
$$+ \sum_{0<s\leq t} (K(X_s) - K(X_{s-}) - K'(X_{s-})\Delta X_s).$$

Remark: When the semi-martingale is purely discontinuous and of bounded variation, it is clear from the application of Taylor's Theorem above that a change of variable formula should only involve the first derivative of K. Ito's formula, as given in the last equation, verifies and extends this to show that in the case of an arbitrary purely discontinuous semi-martingale, the formula also involves only the first order derivative of K.

6.11.3. Remark: The Theorem immediately extends to vector valued continuous, semi-martingales (a finite dimensional vector whose components are continuous continuous semi-martingales): $X = (X^1, X^2, \cdots, X^n)$. Let K be a function from R_n to R having continuous second order partial derivatives. Let D^iK denote the first order derivative of K relative to its i^{th} component with the obvious meaning for $D^{ij}K$, the formula takes the form

$$K(X_t) - K(X_0) = \sum_i^n \int_0^t (D^iK)(X_s)\, dX_s^i + \frac{1}{2}\sum_{i,j}^n \int_0^t (D^{ij}K)(X_s)\, d<X^i,X^j>_s. \qquad (22)$$

6.11.4. Remark: When $K(u)=u^2$ in (21), and $m \epsilon M_{loc}^c$, Ito's formula gives

$$m_t^2 - m_0^2 = 2\int_0^t m_s\, dm_s + <m,m>_t.$$

When $K(u,v) = uv$ and we use (22) with $m,n \epsilon M_{loc}^c$, we obtain

$$m_t n_t - m_0 n_0 = \int_0^t m_s dn_s + \int_0^t n_s dm_s + <m,n>_s. \qquad (23)$$

This integration by parts formula is of course the continuous parameter (continuous process) analogue of the one in Chapter 1. It can be extended to general semi-martingales.

Remark: It is useful to allow the map K in Ito's Theorem to be a complex-valued function. For this purpose, the expectation, conditional expectation, and so on, of complex valued processes are defined in terms of their real counterparts via the real and imaginary parts of the process. For instance, a complex valued martingale is one whose real and imaginary parts are martingales.

Remark: The following is the canonical first application of Ito's Theorem. It is due to P. Lévy. As proved in Doob[1953], it assumes that X is a continuous martingale with the property that $(X_t^2 - t)$ is a martingale. The statement and proof given here is due to Kunita and Watanabe[1967]. It uses their extension of Ito's formula and assumes for the proof of Lévy's Theorem only that X is a continuous local martingale satisfying the condition that $\langle X,X\rangle_t = t$. When X is a martingale, this latter condition is equivalent to the requirement that $(X_t^2 - t)$ is a martingale as in Doob's statement of Lévy's Theorem. Our presentation of the Kunita-Watanabe proof is due in part to Chung and Williams[1983].

6.11.5. **Theorem:**
X is standard Brownian motion relative to the filtration F if, and only if, $X \epsilon M_{0,loc}^c(F)$ *and* $\langle X,X\rangle_t = t, t \geq 0$.

Remark: The condition is necessary by a remark in the last Section. In order to prove that the condition is sufficient, define K_u on R by $K_u(x) = e^{iux}$, for each u in R and apply the Ito formula. From (21), since $X_0 = 0$, we obtain

$$K_u(X_t) - 1 = \int_0^t K_u'(X_s)dX_s - \frac{1}{2}\int_0^t K_u''(X_s)ds.$$

That is,

$$e^{iuX_t} - 1 = iu\int_0^t e^{iuX_s}dX_s - \frac{u^2}{2}\int_0^t e^{iuX_s}ds. \tag{24}$$

The second integral on the right of (24) results from (21) and $\langle X,X\rangle_s = s$. The first integral on the right of (24) is a martingale, because its integrand is a bounded, previsible process and the stopped process X^t is a martingale for any t. Therefore, $E(\int_s^{t+s} e^{iuX_v}dX_v \mid F_s) = 0$, for $s,t>0$. Then, from the definition of conditional expectation, if $B \epsilon F_s$, $E(1_B \int_s^{t+s} e^{iuX_v}dX_v) = 0$. It follows that

$$E(1_B(e^{iuX_{t+s}} - e^{iuX_s})) = -\frac{u^2}{2}E1_B\int_s^{t+s} e^{iuX_v}dv$$

$$= -\frac{u^2}{2}\int_s^{t+s} E(1_B e^{iuX_v})dv, \tag{25}$$

with an application of Fubini's Theorem. If we define $g_s(t) := E(1_B e^{iuX_{t+s}})$ equation (25) becomes

$$g_s(t) - g_s(0) = -\frac{u^2}{2}\int_0^t g_s(v)dv.$$

It can be shown from this equation that g_s must satisfy $g_s(t) = g_s(0)e^{-\frac{u^2}{2}t}$.

Therefore, again using the definition of c.exp. and the fact that $g_s(0)$ is F_s-measurable, we obtain

$$E(e^{iu(X_{t+s} - X_s)} \mid F_s) = e^{-\frac{u^2}{2}t}, \tag{26}$$

and so if Y is an arbitrary bounded F_s-measurable random variable,

$$E(Y\, e^{iu(X_{t+s} - X_s)} \mid F_s) = Y\, e^{-\frac{u^2}{2}t}. \tag{27}$$

Hence,

$$E(Y\, e^{iu(X_{t+s} - X_s)}) = (EY)\, e^{-\frac{u^2}{2}t}.$$

But from (26), this is the same as

$$E(Y\, e^{iu(X_{t+s} - X_s)}) = (EY)\, E(e^{iu(X_{t+s} - X_s)}).$$

It follows that the random variables $e^{iu(X_{t+s} - X_s)}$ and Y are independent. Hence, $(X_{t+s} - X_s)$ is independent of F_s. Again, by (26), $E(e^{iu(X_{t+s} - X_s)}) = e^{-\frac{u^2}{2}t}$, the characteristic function of a Normal zero mean random variable with variance t. Therefore, X is standard Brownian motion.

6.11.6. Remark: We have already noticed in the previous section that Brownian motion is the only continuous process with stationary independent increments. The following observation is a much stronger indication of the importance of Brownian motion in the General Theory of Stochastic Processes. It says that a large class of continuous local martingales are but "a time change away from being Brownian motion".

Let M be a continuous local martingale and suppose that $\langle M,M\rangle_\infty = \infty$.

Then, if we define

$$\eta_t = \inf\{s:<M,M>_s > t\},$$

the process β defined by setting $(\beta_s := M_{\eta_s}, s \geq 0)$ can be shown to be an (F_{η_s})-Brownian motion process and $M_t = \beta_{<M,M>_t}$, $t \geq 0$. (Dubins, Schwartz[1965].)

Remark: We now give a very simple application of Ito's formula that will be extended to vector valued processes later. Let $K:R \rightarrow R$ have two continuous derivatives and introduce the differential operator, L, by setting

$$LK := mK' + \frac{1}{2}\sigma^2 K''.$$

Let X be a Brownian motion process:

$$X_t = mt + \sigma B,$$

where $m \epsilon R$, $t \geq 0$, $\sigma > 0$ and B is standard Brownian motion. Then, in differential form, Ito's formula gives

$$\begin{aligned} dK(X) &= K'(X)dX + \frac{1}{2}K''(X)d<X,X> \\ &= K'(X)(mdt + \sigma dB) + \frac{1}{2}K''(X)\sigma^2 dt \\ &= \sigma K'(X)dB + (mK'(X) + \frac{1}{2}\sigma^2 K''(X))dt \end{aligned}$$

Therefore,

$$dK(X) = \sigma K'(X)dB + (LK)(X)dt.$$

Since K' is continuous, and so previsible, and B is a martingale, it follows that

$$K(X_t) - K(X_0) - \int_0^t (LK)(X_s)ds = \int_0^t \sigma K'(X_s)dB_s,$$

is a martingale.

6.11.7. Remark: We conclude this Section and the Chapter with a brief look at **stochastic differential equations**. To be consistent with the generality of the stochastic integral introduced in this Chapter, we will start with the development of C. Doleans-Dade [1976]. However, our main intent is to introduce stochastic differential equations "driven" by Brownian motion processes, Ito diffusions, and relate these to A.N. Kolmogorov's original description of a diffusion. We will use the Ito formula and a generalization of the operator L defined in the last Remark to very briefly describe the connection with the Stroock-Varahdan theory [1979].

Let (Ω,H,F,P) be a filtered probability space with the filtration F satisfying the "usual conditions". Suppose that σ and b are two functions mapping $R_+\times\Omega\times R$ into R, which are left continuous with right limits in the first factor, F-adapted relative to the second and satisfy the following uniform Lipschitz condition:

$$| b(s,w,x)-b(s,w,y) | + | \sigma(s,w,x)-\sigma(s,w,y) | \leq K | x-y | \tag{L}$$

for some constant K and all (s,w,x), (s,w,y) in the domains of σ and b. C. Doleans-Dade [1976] proves the following:

Theorem:
If M *is an* F*-local martingale and* A *is a process in* BV *and* σ *and* b *satisfy the conditions stated above, then there exists one and only one adapted Skorokhod process* X *satisfying the stochastic integral equation*

$$X_t = X_0 + \int_0^t \sigma(s,X_{s-})dM_s + \int_0^t b(s,X_{s-})dA_s.$$

Remark: As pointed out by Doleans-Dade, the uniform Lipschitz condition in x implies that the mappings $(w,x)\rightarrow\sigma(t,w,x)$ and $(w,x)\rightarrow b(t,w,x)$ are $F_t\times B(R)$-measurable. Consequently, the functions $w\rightarrow\sigma(t,w,X_{s-})$ and $w\rightarrow b(t,w,X_{s-})$ are F-adapted, if we assume that the process X is Skorokhod and adapted. By the assumed left continuity and existence of right limits for σ and b, we have therefore that the processes $(\sigma(t,X_{t-}),t\geq0)$ and $(b(t,X_{t-}),t\geq0)$ are adapted, left continuous and have right limits. Hence, these processes are F-previsible and locally bounded. Therefore, if X is any adapted Skorokhod process the integrals on the right side of the equation in the Theorem exist by earlier results in this Chapter.

This Theorem can be extended in several ways. One is that it can be restated for M as a d-dimensional vector valued process, with σ a matrix valued function of order (n,d) and b a vector valued process with values in R^n. The condition (L) can be modified in an obvious way and if we agree that vector valued processes are adapted, Skorokhod, etc when their components have these properties, an existence and uniqueness Theorem analogous to the one above continues to apply. We will consider this type of structure in paragraph 6.11.8 below, with the components of M being independent Brownian motion processes.

A very interesting paper that we mentioned in the first Chapter (Doleans-Dade [1970]) treats a special case of the stochastic integral given above. Suppose that

$\sigma(t,w,x) = b(t,w,x) = x$, then this stochastic integral takes the form

$$X_t = X_0 + \int_0^t X_{s-}dZ_s,$$

where Z is a semi-martingale. In her 1970 paper, C. Doleans-Dade finds the explicit solution to this equation. It is called the **exponential of Z** when $X_0 \equiv 1$ and is given by

$$X_t = \exp(Z_t - \frac{1}{2}<Z^c,Z^c>_t)\prod_{s\leq t}(1 + \Delta Z_s)e^{-\Delta Z_s}.$$

The proof that this process satisfies the previous stochastic integral equation is a simple application of the general Ito formula. If we set $e(Z) = X$ in the last equation, a two line application of integration by parts to evaluate the product $e(Y)e(Z)$ for $Y,Z \epsilon S$ yields

$$e(Y)e(Z) = e(Y + Z + [Y,Z])$$

and not the expected $e(Y+Z)$. The expected happens, of course, when $[Y,Z] = 0$. For example, this occurs when Y and Z are counting processes representing the number of arrivals and departures (respectively) at a particular queueing station when arrivals and departures from the queue never occur at the same time.

6.11.8. Remark: Now, we will specialize the local martingale in the previous Theorem to Brownian motion, the integral relative to the process A to an integral relative to Lebesgue measure, allow the processes σ and b to depend on t only through X_t and, in the other direction, consider multi-dimensional processes. Thus, let $B = (B_t)$ be a d-dimensional F-Brownian motion process. That is, $B_t = (B_t^1, \cdots, B_t^d)$, where the B^i are P-independent F-Brownian motion processes; so in particular the distribution of $B_t - B_s$ $(t>s)$ is normal $(0,(t-s)I)$, where $I = d \times d$ is the identity matrix.

Thus, $\sigma: R^n \rightarrow R^n \times R^d$, $b: R^n \rightarrow R^n$ and X satisfies the equation

$$X_t = X_0 + \int_0^t \sigma(X_s)dB_s + \int_0^t b(X_s)ds. \quad (28)$$

With $X_0 = x$, the process X is called an **Ito diffusion**, and is said to satisfy the stochastic differential equation

$$dX_t = \sigma(X_t)dB_t + b(X_t)dt.$$

for $t>0$, and $X_0 = x$. $X = (X_t)$ is then a strong Markov process with a.s.P continuous paths. (For an easy to read account on stochastic differential equations see Oksendal [1986], in particular, his Theorem 5.5 for an existence and uniqueness result that covers this case.)

The k^{th} component, X^k, of equation (28) is given in differential form by

$$dX^k = \sum_{j=1}^{d} \sigma_{kj} dB^j + b_k dt.$$

An application of (22) to (28) yields

$$dK(X) \equiv \sum_{i=1}^{n} \sum_{j=1}^{d} (D^{ij}K)(X)\sigma_{ij} dB^j + \qquad (29)$$

$$+ \sum_{i}^{n} (D^iK)(X) b_i dt + \frac{1}{2} \sum_{i,k}^{n} D^{ik}K)(X) a_{ik} dt$$

where $a = \sigma\sigma^*$, σ^* the transpose of σ.

The extension of the previously defined operator L to functions K on R^n is

$$(LK)(x) = \sum_{i}^{n} (D^iK)(x) b_i + \frac{1}{2} \sum_{i,j}^{n} (D^{ij}K)(x) a_{ij}.$$

Then from (29) we see that

$$C_t^K(X) = K(X_t) - K(x) - \int_0^t (LK)(X_s) ds \qquad (30)$$

is a martingale, as in the one dimensional case treated earlier.

Thus, starting with Brownian motion on a given filtered probability space, and an Ito process X satisfying (28), we associated the operator L with the property that $C^K(X)$ was a martingale for a large class of functions, K, defined on R^n.

There is a "converse" to this result due to Stroock and Varahdan [1979] which is extremely important in the study of vector valued diffusions and, further, can be used to define diffusions on more general manifolds than R^n. We can not say much about the Stroock-Varahdan approach in this note, but highly recommend the paper by D. Williams [1981] for an introduction to this subject and its relationship to the Ito method.

Roughly speaking, view a process, X, as a member of the space, W, of continuous functions from R_+ to R^n. Take A_t^* to be the σ-algebra of subsets of W generated by $\{X_s, s \leq t\}$ and set $A^* = A_\infty^*$. Let $x \epsilon R^n$ and L be an operator of the form

$$(LK)(x) = \sum_{i}^{n} (D^iK)(x) b_i + \frac{1}{2} \sum_{i,j}^{n} (D^{ij}K)(x) a_{ij},$$

where the matrix valued function "a" and the vector valued function b are defined on R^n.

Suppose that P_x is a probability measure on (W, A^*), with the property that $P_x(X_0 = x) = 1$ and C^K, defined by (30), is an $(W, A^*, (A_t^*), P^x)$-martingale, for all twice differentiable functions K on R^n having compact support. (Then P_x is said to solve the **martingale problem** for L starting from x.)

Finally, if "a" can be written in the form $a_{ij} = (\sigma\sigma^*)_{ij}$, then X is continuous and there exists a Brownian motion process B such that X_s is independent of $B_t - B_s$ and

$$X_t = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s.$$

BIBLIOGRAPHY

Aldous, D.J. (1981). Weak Convergence and the General Theory of Processes. *Draft Monograph* Dept of Statistics, Univ of Calif, Berkeley, CA.

Andersen, G. (1986). An Application of Discrete Parameter Martingale Calculus to Discrete Point Processes. Part I: Girsanov transformations, and compensator robustness. *In draft.*

Andersen, G. (1986). An Application of Discrete Parameter Martingale Calculus to Discrete Point Processes. Part II: Nonlinear Filtering. *In draft.*

Bichteler, K. (1981). Stochastic Integration and Lp-Theory of Semimartingales. *The Annals of Probability* Vol.9 No.21 49-89.

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley & Sons.

Boel, R., Varaiya P. and Wong, E. (1975). Martingales on Jump Processes. I: Representation Results. *SIAM J Control* Vol 13 No.5, 999-1021.

Boel, R., Varaiya P. and Wong, E. (1975). Martingales on Jump Processes. II: Applications. *SIAM J Control* Vol 13 No.5, 1022-1061.

Bremaud, P. (1975). An Extension of Watanabe's Theorem of Characterization of Poisson Processes over the Positive Real Half Line. *J. Appl. Prob.* 12 396-399.

Bremaud, P. (1975). The Martingale Theory of Point Processes over the Real Half Line Admitting an Intensity. *Lect. Notes in Economics and Math. Systems (Control Theory)* 107 Springer-Verlag.

Bremaud, P. (1976). La methode des semi-martingales en filtrage lorsque l'observation est un processus ponctel. *Sèminaire Proba.* X, *Universitè de Strasbourg*, Lect. Notes in Math. 511, Springer-Verlag, 1-18.

Bremaud, P. (1978). On the Output Theorem of Queueing Theory via Filtering. *J. Appl. Prob.* 15 397-405.

Bremaud, P. (1981). *Point Processes and Queues Martingale Dynamics.* Springer-Verlag Series in Statistics.

Bremaud, P. and Jacod, J. (1977). Processus Ponctuels et Martingales: Resultats Recents sur la Modelisation et le Filtrage. *Adv. Appl. Prob.* 9 362-416.

Bremaud, P. and Yor, M. (1978). Changes of Filtrations and of Probability Measures. *Z. Wahr. verw. Gebiete* 45 269-295.

Brown, T.C. (1978). A martingale approach to the Poisson convergence of simple point processes. *The Annals of Probability* Vol. 6 726-744.

Brown, T.C. (1983). Some Poisson Approximations Using Compensators. *The Annals of Probability* Vol. 11 No.3 726-744.

Chou, C.S. and Meyer, P.A. (1975). Sur la Representation des Martingales Comme Integrales Stochastiques Dans les Processus Ponctuels. *Séminaire Proba. IX, Université de Strasbourg*, Lect. Notes in Math. 465, Springer-Verlag, 226-236.

Chung, K.L. (1982). *Lectures from Markov Processes to Brownian Motion.* Springer-Verlag.

Chung, K.L. and Doob, J.L (1965). Fields optionality and measurability. *Amer. J. Math.* 397-424.

Chung, K.L. and Williams, R.J. (1983). *Introduction to Stochastic Integration.* Birkhauser.

Davis, M. (1976). The Representation of Martingales of Jump Processes. *SIAM J Control and Opt.* Vol 14.

Dellacherie, C. (1970). Un Exemple de la Theorie Generale des Processus. *Séminaire Proba. IV, Université de Strasbourg*, Lect. Notes in Math. 124, Springer-Verlag, 60-70.

Dellacherie, C. (1972). *Capacites et processus stochastiques.* Springer-Verlag.

Dellacherie, C. (1978). Un survol de la theorie de l'integrale stochastique. *Proc. of the International Congress of Mathematicians Helsinki* 2 733.

Dellacherie, C. and Meyer, P.-A. (1975). *Probabilities and Potential.* North Holland Mathematical Studies No. 29.

Doleans-Dade, C. (1970). Quelques applications de la formule de changement de variables pout les semi-martingales. *Z. Wahr. verw. Giebiete* 16 181-194.

Doleans-Dade, C. (1976). On the Existence and Unicity of Solutions of Stochastic Integral Equations. *Z. Wahr. verw. Giebiete* 36 93-101.

Dellacherie, C. and Meyer, P.-A. (1980). *Probabilities and Potential B Theory of Martingales.* North Holland Mathematical Studies No. 72.

Doleans-Dade, C. and Meyer, P.A. (1970). Integrales Stochastiques Par Rapport Aux Martingales Locales. *Sèminaire Proba. IV, Universitè de Strasbourg,* Lect. Notes in Math. 124, Springer-Verlag, 77-107.

Doob, J.L. (1953). *Stochastic Processes.* John Wiley & Sons.

Dubins, L., Schwarz, G. (1965). On Continuous Martingales. *Proc. Nat. Acad. Sci. U.S.A.* 53.

Gill, R.D. (1980). *Censoring and Stochastic Integrals.* Mathematical Centre Tracts 124.

Helland, I. (1981). Central Limit Theorems for Martingales. *Scand. J. Statistics* 9.

Ikeda, N. and Watanabe, S. (1981). *Stochastic Differential Equations and Diffusion Processes.* North-Holland Publishing Company.

Itmi, M. (1980). Histoire Interne des processus ponctuels marqués stochastiques. Etude d'un probléme de filtrage. Thése de 3éme cycle U. de Rouen.

Ito, K. (1944). Stochastic Integral. *Proc. Imp. Acad. Tokyo.* 20.

Ito, K. (1951). Multiple Wiener Integral. *J. Math. Soc. Japan* 3.

Jacobsen, M (1982). *Statistical Analysis of Counting Processes.* Lect. Notes in Statistics, No.12, Springer-Verlag.

Jacod, J. (1975). Multivariate Point Processes: Predictable Projection Radon-Nikodym Derivatives Representation of Martingales. *Z. Wahr. verw. Giebiete* 32 235-253.

Jacod, J. (1976). Un théorème de représentation pour les martingales discontinues. *Z. Wahr. verw. Giebiete* 34 225-244.

Jacod, J. (1979). *Calcul Stochastique et Problemes de Martingales.* Lect. Notes in Math. 714 Springer-Verlag.

Kabanov, J., Lipcer, R., Sirjaev, A. (1979). Absolute Continuity and Singularity of Locally Absolutely Continuous Probability Distributions. I. *Math USSR Sbornik* Vol. 35, No 5, 631-680.

Kabanov, Y.M. and Liptser, R.S. and Shiryaev, A.N. (1983). Weak and Strong Convergence of the Distributions of Counting Processes. *Theory of Probability and its Applications* Vol XXVIII 303-336.

Kallianpur, G. (1980). *Stochastic Filtering Theory.* Springer-Verlag.

Karlin, S. and Taylor, H.M. (1975). *A First Course in Stochastic Processes (Second Edition).* Academic Press.

Kunita, H. and Watanabe, S. (1967). On Square Integrable Martingales. *Nagoya Math. J.* 30 209-245.

Lévy, P. (1937). *Theorie de l'addition des variables aleatories.* Paris.

Lipster, R. and Shiryayev, A. (1977). *Statistics of Random Processes I General Theory.* Springer-Verlag.

Lipster, R. and Shiryayev, A. (1978). *Statistics of Random Processes I General Theory.* Springer-Verlag.

Loeve, M. (1960). *Probability Theory.* Van Nostrand.

Metivier, M. (1977).
Reele and Vektorwertige Quasimartingale und die Theorie der Stochastischen Integration. Lect. Notes in Math. 607, Springer-Verlag.

Metivier, M. (1982). *Semimartingales a Course on Stochastic Processes.* Walter de Gruyter Berlin NY.

Meyer, P.A. (1967). Integrales Stochastiques (4 exposes). *Sèminaire Proba. I, Universitè de Strasbourg,* Lect. Notes in Math. 39, Springer-Verlag.

Meyer, P.A. (1969). Les Inegalites de Burkholder en Theorie des Martingales d"apres R. Gundy. *Sèminaire Proba. III, Universitè de Strasbourg,* Lect. Notes in Math. 88, Springer-Verlag.

Meyer, P.A. (1973). *Martingales and Stochastic Integrals I.* Lect. Notes in Math. 284, Springer-Verlag.

Meyer, P.A. (1976). Un Cours Sur les Integrales Stochastiques. *Sèminaire Proba. X, Universitè de Strasbourg,* Lect. Notes in Math. 511, 245-400, Springer-Verlag.

Meyer, P.A. (1966). *Probability and Potential.* Blaisdell.

Neveu, J. (1975). *Discrete-Parameter Martingales.* North Holland.

Oksendal, B. (1986). *Stochastic Differential Equations.* Springer-Verlag.

Rao, K.M. (1969). On Decomposition Theorems of Meyer. *Math. Scand.* 24 66-78.

Rogers, L.C.G. (1981). Stochastic Integrals: Basic Theory. *Stochastic Integrals, Proc LMS Durham Symposium*Lect. Notes in Math. 851, Springer-Verlag.

Segall, A., Davis, M. and Kailath, T. (1975). Nonliear Filtering with Counting Observations. *IEEE Transactions on IT* Vol IT-21 No. 2 143.

Shiryayev, A.N. (1982). Martingales: Recent Developments Results and Applications. *Scand J. Statistics* 10.

Shiryayev, A.N. (1984). *Probability.* Springer-Verlag.

Skorokhod, A.V. (1956). *Studies in the Theory of Random Processes.* Addison-Wesley.

Stroock, D., Varadhan, S.R.S. (1956). *Multidimensional Diffusion Processes.* Springer-Verlag.

Stratonovich, R. L. (1964). A New Form of Representing Stochastic Integrals and Equations. *Vestnik Moskov. Univ. Ser. I. Math. Meh.* 1 3-12.

Van Schuppen, J.H. (1977). Filtering, Prediction and Smoothing for counting processes, a martingale approach. *SIAM J App. Math.* 32 552-520.

Van der Hoven, P.C.T. (1980). *On Point Processes.* Mathematical Centre Tracts.

Watanabe, S. (1964). Additive Functionals of Markov Processes and Levy Systems. *Jap. J. Math.* 34 53-79.

Wiener, N (1923). Differential Space. *J. Math. and Phys.* Vol 2 131-174.

Williams, D (1979). *Diffusions, Markov Processes, and Martingales. Vol I Foundations.* Wiley & Sons.

Williams, D (1981). To begin at the beginning:. ... . *Stochastic Integrals, Proc LMS Durham Symposium* Lect. Notes in Math. 851, Springer-Verlag.

Wong, E. (1973). Recent Progress in Stochastic Processes (Applications of Stochastic Processes, 1968-72) *IEEE Trans. IT* .

Wong, E. and Zakai, M. (1965). On the Convergence of Ordinary Integrals to Stochastic Integrals. *Ann.Statist.* 36 5 1560-1564.

Yor, M. (1977). Sur les theories du filtrage et de la prediction. *Sèminaire Proba. XI, Universitè de Strasbourg,* Lect. Notes in Math. 581, Springer-Verlag.

Yor, M. (1977). Sur quelques approximations d'integrales stochastiques. *Sèminaire Proba. XI, Universitè de Strasbourg,* Lect. Notes in Math. 581, Springer-Verlag.

Yor, M. (1979). Les Inegalites de Sous-Martingales Comme Consequences de la Relation de Domination. *Stochastics* Vol 3 1-15.

LIST OF SYMBOLS

$X := (X(t), t \epsilon R_+)$: Stochastic process

$X(t,w) := X_t(w)$: Process evaluated at (t,w)

$[X_t \epsilon A] := \{w : X_t(w) \epsilon A\}$

$\{X \epsilon A\} := \{(t,w) : X_t(w) \epsilon A, t \geq 0, w \epsilon \Omega\}$

X^T: Process X stopped at time T.

V.X: Stochastic integral of V relative to X.

[X,X]: (Optional) quadratic variation of X.

$<X,X>$: (Previsible) quadratic variation of X.

$X_{t-} := \lim_{s \to t-} X_s$; $(X_-)_t := X_{t-}$, $t \geq 0$.

$\Delta X_t := X_t - X_{t-}$, $t \epsilon R_+$.

$(X_-)_n := X_{n-1}$, $n \epsilon Z_+$; $\Delta X_n := X_n - X_{n-1}$, $n \epsilon Z_+$.

$\sigma(G)$: Sigma-algebra generated by the collection of sets, G.

T_A: Restriction of the stopping time T to the set A.

1_A: Indicator function of the set A.

[[T]]: Graph of the stopping time T.

[[S,T]]: A stochastic interval, S and T stopping times.

${}^pX$ (${}^oX$): Previsible (Optional) projection of X.

X^p: Dual previsible projection of X.

$X^*(t) := \sup_{s \leq t} | X(s) |$: Supremum process.

G(PT): σ-algebra of previsible sets.

G(OT): σ-algebra of optional sets.

G(AT): σ-algebra of accessible sets.

$\overline{R}_+$: Extended non-negative real line, $[0,\infty]$.

$\overline{Z}_+$: Extended non-negative integers.

$a \wedge b$: Minimum of the numbers a and b.

$\tilde{X}$: Compensator of the process X.

Spaces of stochastic processes:

M_u: Uniformly integrable martingales.

$M_0 := (M_u)_0$: Members of M_u with $m(0) = 0$.

$M_{loc} := (M_u)_{loc}$: Local martingales.

$M_{0,loc} := (M_0)_{loc}$.

K^2: Square integrable martingales.

$K^{2,c}$: Continuous square integrable martingales.

$K^{2,d}$: Purely discontinuous square integrable martingales.

M[T]: Square integrable martingales, continuous outside of [[T]].

V^+: Increasing processes.

BV: Processes of bounded variation.

IV^+: Integrable increasing processes.

IV: Processes of integrable variation.

INDEX OF DEFINITIONS

APPENDIX A.

ODDS AND ENDS, INCLUDING FUBINI'S THEOREM.

Appendix A

A 1. Odds and Ends, including Fubini's Theorem.

A 1.1. Some Useful Definitions and Results:

A 1.1.1. **Conditional Expectation**:
Let (Ω,H,P) be a probability space and G be a sub σ-algebra of H. Let X be a P-integrable random variable and define the measure μ on G by setting

$$\mu(A) := \int_A X(w)P(dw) \equiv \int_A X dP,$$

for all A in G.

Then μ is a finite measure on G which is absolutely continuous relative to the restriction of P to G. The Radon-Nikodym derivative of μ with respect to this restriction is called the **conditional expectation** of X given G. Therefore, $E(X|G)$ is an a.s.P unique G-measurable integrable random variable Z which is characterized by

$$\int_A Z\, dP = \int_A X\, dP, \tag{1}$$

for all A in G, since P and its restriction agree on G.

The following is a list of some of the more important properties of conditional expectation. These properties together with equation (1) are constantly (and silently) used in Chapters 1 through 6.

Let X and Y be P-integrable random variable and a,b real numbers. Then

(i) $E(aX + bY \mid G) = aE(X \mid G) + bE(Y \mid G)$, a.s.P.

(ii) If Y is G-measurable and XY is P-integrable, then $E(XY \mid G) = YE(X \mid G)$, a.s.P.

(iii) If J is a sub σ-algebra of G, then $E(X \mid J) = E(E(X \mid G) \mid J)$, a.s.P.

(iv) $E(1 \mid G) = 1$.

(v) If $X \geq 0$ a.s.P, then $E(X \mid G) \geq 0$.

(vi) If $X_n \epsilon L_1(P)$, for all $n \epsilon Z_+$, and $X_n \rightarrow X$, in $L_1(P)$, then $E(X_n \mid G) \rightarrow E(X \mid G)$, in $L_1(P)$.

(vii) If $X_n \epsilon L_1(P)$, for all $n \epsilon Z_+$, $X_n \uparrow X$, a.s.P, and $X \epsilon L_1(P)$, then $E(X_n \mid G) \rightarrow E(X \mid G)$, a.s.P.

(viii) If $h: R_1 \rightarrow R_1$ is convex, and $h(X) \epsilon L_1(P)$, then $h(E(X \mid G)) \leq E(h(X) \mid G)$, a.s.P.

Remark: Properties (ii) and (iv) combine to yield $Y = E(Y \mid G)$, a.s.P, when Y is G-measurable and P-integrable.

A 1.1.2. **Skorokhod Processes**

A function is called Skorokhod if it is right continuous with left limits at each point in its domain. Some basic results on such functions can be found in Billingsley [1968]. Billingsley considers real-valued Skorokhod functions defined on compact intervals. In Chapters 2 to 6 in the body of the present note, the usual domain for functions (as paths of stochastic processes) is the interval $[0,\infty)$. The results from Billingsley that we quote here carry over in an obvious way to this domain. For this purpose, let f be a Skorokhod function defined on $[0,\infty)$. Then

(i) f has at most a countable number of discontinuities;

(ii) On any compact interval, f has at most a finite number of discontinuities where the magnitude of the corresponding jumps exceed a specified fixed positive number;

(iii) f is bounded on compact intervals.

A 1.2. **Fubini's Theorem**:

A 1.2.1. **Definition** (Transition Probability): Let $(v_w, w \epsilon \Omega)$ be a family of probability measures on the measure space (E,G). Let (Ω,H) be a measure space If the mapping $w \rightarrow v_w(B)$ is H-measurable, for each B in G, then the family $(v_w, w \epsilon \Omega)$ is called a **transition probability** from (Ω,H) to (E,G).

A 1.2.2. **Theorem** (Fubini):
Let $U = E \times \Omega$, $V = G \times H$, and f be a real valued, V-measurable function (a random variable on (U,V)).

(i) Then, for each $w \epsilon \Omega$,

$$(x \rightarrow f(x,w)) \text{ is G-measurable}$$

and, for each $x \epsilon E$,

$$(w \rightarrow f(x,w)) \text{ is H-measurable.}$$

(ii) Further, let P be a probability measure on (Ω,H) and $(v_w, w \epsilon \Omega)$ a transition probability from (Ω,H) into (E,G). Then there exists a unique probability measure, μ, on (U,V) such that

$$\mu(C \times D) = \int_D v_w(C)\, P(dw)$$

for all $C \epsilon G$ and $D \epsilon H$.

(iii) If f is non-negative, then

$$(w \rightarrow \int_E f(x,w)\, v_w(dx)) \text{ is H-measurable}$$

and

$$\int_U f\, d\mu = \int_\Omega \int_E f(x,w)\, v_w(dx) P(dw). \qquad (2)$$

If $f \epsilon L_1(\mu)$, then equation (2) holds and $(x \rightarrow f(x,w)) \epsilon L_1(v_w)$, a.s.P.

1.2.3. Remark: In the special case that v_w is independent of w, μ is called the product measure.

1.2.4. Remark: We have stated Fubini's Theorem in terms of transition probabilities. It holds also, and will be applied, when the indexed family of probability measures in the definition of transition probability is replaced by an indexed family of σ-finite measures, satisfying the measurability condition of the definition. The result is called a **transition measure**.

APPENDIX B.

LEBESGUE-STIELTJES STOCHASTIC INTEGRALS.

Appendix B

B 1. Lebesgue-Stieltjes Stochastic Integrals:

B 1.1. On the Existence of a Lebesgue-Stieltjes Stochastic Integral:

We will now give a detailed explanation of the the existence of the stochastic Lebesgue-Stieltjes integral induced by an increasing stochastic process, A.

Let $\mathbf{B}=B([0,\infty])$ be the σ algebra of subsets generated by intervals of the form $(a,b]$, a and b non-negative. Let $C \rightarrow v(C,w)$, $C\epsilon\mathbf{B}$, be the measure on $\mathbf{B}$ induced by the right continuous, increasing function, A by setting $v((a,b],w) := A(b,w) - A(a,w)$, for all non-negative a and b ($a<b$) and each $w\epsilon\Omega$.

Let $V := \mathbf{B}\times H$ be the product algebra on $U := [0,\infty)\times\Omega$. Since $t\rightarrow A(t,w)$ is increasing, A is V-measurable. Therefore, the mapping $w \rightarrow v(C,w)$ is H-measurable for each C in $\mathbf{B}$.

By Fubini's Theorem, given the family of σ–finite transition measures $\{v(.,w) : w\epsilon\Omega \}$ and the probability measure, P, on H, there corresponds a unique σ-finite measure, μ, from V into $[0,\infty]$, such that

$$\mu(C\times D) = \int_D v(C,w)P(dw)$$

for all $C\epsilon\mathbf{B}$ and $D\epsilon H$, and for any μ-integrable real function, f, defined on U, the mapping

$$w \rightarrow \int_{[0,\infty)} f(s,w)v(ds,w)$$

is H-measurable, and

$$\int_U f\,d\mu = \int_\Omega \int_{[0,\infty)} f(s,w)v(ds,w)P(dw).$$

Let X be any V-measurable process such that $E\int_0^\infty |X(s)|\,dA(s) < \infty$, where dA denotes the integration relative to the measure v.

Then

$$\int_U X d\mu = E\int_0^\infty X(s)\,dA(s)$$

and Fubini's theorem states that the pathwise Lebesgue Stieltjes integral

$$(X.A)_\infty(w) = \int_0^\infty X(s,w)\, dA(s,w)$$

exists, a.s.P. The process $((X.A)_t, t \geq 0)$ is then defined by setting $(X.A)_t := (1_{[0,t]}X.A)_\infty$, for each $t \geq 0$.

B 1.2. **Monotone Class Theorem:**

B 1.2.1. **Theorem**: Let O be a set and C a collection of subsets of O which is closed under finite intersection.

1) Let S(C) be the smallest collection of subsets of O which contains C and satisfies

a) $O \in S(C)$;

b) If $A, B \in S(C)$, with A a subset of B, then $B-A \in S(C)$;

c) S(C) is closed under countable unions of increasing sequences of its members.

Then S(C) is the smallest σ algebra containing C.

2) Let H^* be a vector space of real-valued functions defined on the set O and satisfying

a) $1 \in H^*$ and if $A \in C$ then $1_A \in H^*$;

b) If $(f_n, n > 0)$ is an increasing sequence of nonnegative members of H^*, with bounded supremum, then $\sup\{f^n : n > 0\}$ is also a member of H^*.

Then H^* contains all bounded real-valued functions, defined on O, which are measurable relative to the σ algebra generated by C.

DISTRIBUTION LIST

No. of Copies	Organization
1	Deputy Assistant Secretary of the Army (Requirements & Programs) ATTN: Dr. P. C. Dickenson Pentagon, Room 2E673 Washington, D.C. 20310
1	Superintendent Naval Postgraduate School ATTN: Dir of Lib Monterey, CA 93940
2	Commander U.S. Army Materiel Command ATTN: AMCDRA-ST (B. Bartholow) 5001 Eisenhower Avenue Alexandria, VA 22333-0001
1	Commander Fort Leavenworth Scientific Advisor to CAC ATTN ATZL-SCI (Jim Fox) Fort Leavenworth, Kansas 66027
1	Commander Fort Leavenworth ATTN ATZL-SCI (Dr. E Inselman) Fort Leavenworth, Kansas 66Q27
1	Commander USAOTEA ATTN STE-ZS 5600 Columbia Pike Falls Church, Virginia 22041
1	Departm:nt of the Army ATTN: SAUS-OR (Dan Willard) Washington, D.C. 20301-0102
1	Dr. R. J. Heaston OUSDRE(R&AT/MST) Room 3D1089, Pentagon Washington, D.C. 20301-3080
1	Dr. S. Wolff Office of Advanced Scientific Computing National Science Foundation Room 533 1800 G ST. NW Washington, DC.
1	Director US Army Research Office ATTN: SLCRO-MA (Dr. R. Launer) PO Box: 12211 Durham, NC 27709-2211
1	Director US Army Research Office ATTN: SLCRO-MA (Dr. J. Chandra) PO Box: 12211 Durham, NC 27709-2211
1	Director US Army Research Office ATTN: SLCRO-MA (Dr. T. Dressel) PO Box: 12211 Durham, NC 27709-2211
1	Commander US Army Missile Command Research, Development & Engineering Center ATTN: AMSMI-RD-DE-TS (Dr. B Henriksen) Redstone Arsenal, AL 35898

DISTRIBUTION LIST

No. of Copies	Organization
1	US Army RDSG ATTN: Dr. J. Gault Box 65 FPO New York, NY 09510
1	The Catholic University of America ATTN: Math. Department (Dr. Daniel Gallo) 620 Michigan Ave NE Washington, DC
1	George Washington University ATTN: Operations Research Dept (Dr. N. Singpurwalla) 707 22 St NW Washington, DC 20052
1	Mathematical Sciences Institute Cornell University ATTN: Dr. N. U. Prabhu Caldwell Hall Ithaca, NY 14853-2602
1	University of Maryland ATTN: Computer Science Dept (Dr. R. Austing) College Park, MD 20742
1	Columbia University in the City of New York ATTN: Math. Stat. Dept (Dr. I. Karatzas) Mathematics Building New York, NY 10027
1	George Mason University ATTN: Math. Dept (Dr. Saperstone) 4400 University Drive Fairfax, Va 22030
1	University of Maryland ATTN: Statistics Dept (Dr. Grace Yang) College Park, MD 20742
1	Dr. Tatsuo Kawata Tsutsujigaoka 31-18 Midori-ku Yokohama City, 227 Japan
1	Dr. B. Ramachandra Indian Statistical Institute Staff Qtrs A-8 7, S.J.S. Sansanual May New Delhi, 110016, India
1	Bowling Green University ATTN: Math. Stat. Dept. (Dr. R. G. Laha) Bowling Green, Ohio
1	George Mason University ATTN: OR & Statistics Dept (Dr. K. Jo) 4400 University Drive Fairfax, Va 22030
1	University of Alabama ATTN: Dept of Mgmt Sciences and Statistics (Dr. Badrig Kurkjian) Tuscolusa, Alabama 35486

DISTRIBUTION LIST

No. of Copies	Organization
1	Director Night Vision & Electro Optics Laboratory ATTN: AMSEL-NV-ACD (Dr. V. Mirelli) Building 357 Ft. Belvoir, VA 22060-5677
1	Dr. E. Lukacs 3727 Van Ness St. NW Washington, DC
1	Dr. B. J. McCabe Daniel H. Wagner, Associates Station Square One Paoli, PA 19301
1	John Gubner University of Maryland Electrical Engineering Dept College Park, MD 20742

DISTRIBUTION LIST

No. of
Copies   Organization

Aberdeen Proving Ground

Dir, USAAMSA
ATTN:
AMXSY-CS, H. Burke
AMXSY-MP, H. Cohen
AMXSY-LR, M. Rosenblatt

Dir, USAHEL
ATTN: AMXHE-D B. Cummings

US Army Ballistic Research
Laboratory

G. R. Andersen, SECAD
W. Baker, SECAD
B. Bodt, SECAD
H. Breaux, SECAD
P. Broome, SECAD
S. Chamberlain, SECAD
A. B. Copper, SECAD
P. Dykstra, SECAD
W. Egerland, SECAD
J. Groff, SECAD
J. Grynovicki, SECAD
C. Hansen, SECAD
T. Harkins, SECAD
G. Hartwig, SECAD
R. Kaste, SECAD
V. Kaste, SECAD
R. McGee, SECAD
M. Muuss, SECAD
R. Natalie, SECAD
T. Perkins, SECAD
H. L. Reed, SECAD
B. Reichard, SECAD
R. Reschly, SECAD
M. S. Taylor, SECAD
J. Thomas, SECAD
W. Winner, SECAD
S. Wolff, SECAD
C. R. Zoltani, IBD

Copies	Organization
12	Administrator Defense Technical Info Center ATTN: DTIC-DDA Cameron Station Alexandria, VA 22304-6145
1	HQDA DAMA-ART-M Washington, D.C. 20310
1	Commander U.S. Army Materiel Command ATTN: AMCDRA-ST 5001 Eisenhower Avenue Alexandria, VA 22333-0001
1	Commander Armament R&D Center U.S. Army AMCCOM ATTN: SMCAR-TSS Dover, NJ 07801
1	Commander Armament R&D Center U.S. Army AMCCOM ATTN: SMCAR-TDC Dover, NJ 07801
1	Director Benet Weapons Laboratory Armament R&D Center U.S. Army AMCCOM ATTN: SMCAR-LCB-TL Watervliet, NY 12189
1	Commander U.S. Army Armament, Munitions and Chemical Command ATTN: SMCAR-ESP-L Rock Island, IL 61299
1	Commander U.S. Army Aviation Research and Development Command ATTN: AMSAV-E 4300 Goodfellow Blvd St. Louis, MO 63120
1	Director U.S. Army Air Mobility Research and Development Laboratory Ames Research Center Moffett Field, CA 94035

Copies	Organization
1	Commander U.S. Army Communications-Electronics Command ATTN: AMSEL-ED Fort Monmouth, NJ 07703
1	Commander ERADCOM Technical Library ATTN: DELSD-L (Reports Section) Fort Monmouth, NJ 07703-5301
1	Commander U.S. Army Missile Command Research, Development & Engineering Center ATTN: AMSMI-RD Redstone Arsenal, AL 35898
1	Director U.S. Army Missile & Space Intelligence Center ATTN: AIAMS-YDL Redstone Arsenal, AL 35898-55
1	Commander U.S. Army Tank Automotive Cmd ATTN: AMSTA-TSL Warren, MI 48397-5000
1	Director U.S. Army TRADOC Systems Analysis Activity ATTN: ATAA-SL White Sands Missile Range, NM 880
1	Commandant U.S. Army Infantry School ATTN: ATSH-CD-CSO-OR Fort Benning, GA 31905
1	Commander U.S. Army Development and Employment Agency ATTN: MODE-TED-SAB Fort Lewis, WA 98433
1	AFWL/SUL Kirtland AFB, NM 87117
1	Air Force Armament Laboratory ATTN: AFATL/DLODL Eglin AFB, FL 32542-5000

Copies	Organization
10	Central Intelligence Agency Office of Central Reference Dissemination Branch Room GE-47 HQS Washington, D.C. 20502

ABERDEEN PROVING GROUND

Dir, USAMSAA
ATTN: AMXSY-D
AMXSY-MP, H. Cohen

Cdr, USATECOM
ATTN: AMSTE-TO-F

Cdr, CRDC, AMCCOM
ATTN: SMCCR-RSP-A
SMCCR-MU
SMCCR-SPS-IL

USER EVALUATION SHEET/CHANGE OF ADDRESS

This Laboratory undertakes a continuing effort to improve the quality of the reports it publishes. Your comments/answers to the items/questions below will aid us in our efforts.

1. BRL Report Number ______________________ Date of Report __________

2. Date Report Received ______________________

3. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which the report will be used.) __________

4. How specifically, is the report being used? (Information source, design data, procedure, source of ideas, etc.) __________

5. Has the information in this report led to any quantitative savings as far as man-hours or dollars saved, operating costs avoided or efficiencies achieved, etc? If so, please elaborate. __________

6. General Comments. What do you think should be changed to improve future reports? (Indicate changes to organization, technical content, format, etc.)

CURRENT ADDRESS

Name

Organization

Address

City, State, Zip

7. If indicating a Change of Address or Address Correction, please provide the New or Correct Address in Block 6 above and the Old or Incorrect address below.

OLD ADDRESS

Name

Organization

Address

City, State, Zip

(Remove this sheet along the perforation, fold as indicated, staple or tape closed, and mail.)

END 2-87