The evolution of stochastic mathematics that changed the financial world.

P.-C.G. VASSILIOU

Department of Statistical Sciences, University College London, and Mathematics Department, Aristotle University of Thessaloniki.

ABSTRACT

In this invited paper, for the plenary session of the conference of the Greek Statistical Institute, we study the evolution of stochastic mathematics that changed the financial world. We discuss what is thought to be its genesis i.e Bachelier's thesis at Sorbone and the decisive steps i.e. Measure Theory, Martingale Theory, Stochastic Integration, Black-Merton-Scholes partial differential equation and Girsanov's theorem. We briefly refer to their interrelation with finance and we discuss the present flow of research in Stochastic Finance. We also refer briefly to the life and work of Bachelier the father of the use of Brownian motion in Finance problems and also the life and work of the tragic Wolfgang Doeblin one of the founders of Ito-Doeblin formula.

1. Introductory thoughts

Let us think of the first act that the human race made that could be attributed to the area of finance . Then it is easy to trace it , in the first day that a group of humans decided to swap an asset with an another group and not to kill or steal to get it. Since then, million of years have gone by before the area of finance reached the point where it was thought to be a scientific discipline.

Let us now consider a rather fuzzy random process $\{X_t\}_{t\geq 0}$ which expresses the scientific progress achieved in Mathematics and its applications in the time interval $[t, t + dt)$. Then intuitively one would feel that a good model for $\{X_t\}_{t\geq 0}$ would be a diffusion process and more specifically a mean reverting one but with jumps. In what follows we will try to pin-point the jumps in the random evolution of stochastic mathematics that led to the fascinating and important scientific discipline now days known as Mathematical Finance or Theory of Finance or Stochastic Finance. Naturally,

655 Avenue of the Americas, New York, NY 10010 0024-3795/94/$6.00

this is not an easy task and there is a great danger that some important moments might be omitted in this small space. However the jumps that we will mention are ones that surely had a great impact in the evolution which led to today's dense research and applications of Stochastic Finance.

In section 2 we discuss what is thought to be the genesis of stochastic finance i.e. the Bachelier thesis at Sorbone. Interesting details of the life and work of Bachelier are presented taking into account the limited time and space. In section 3 we discuss the decisive steps in the progress of stochastic mathematics that led to the nowdays enormous flow of research on Mathematical Finance. These are Measure theory, Martingale theory, Stochastic Integration, Girsanov's theorem and the Black , Merton and Scholes partial differential equation. In this path we also refer briefly to their interrelation with financial problems.We also mention the tragic story of Wolfgang Doeblin who discovered stochastic integration in the barracks of the second World War and this has been a secret over 60 years. Finally, in section 3 we refer to the main areas of research which are presently active in Stochastic Finance.

2. Genesis

The modelling of risky asset prices begin with Brownian motion, so let us begin there too. The first thing is to define Brownian motion. We assume given some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

DEFINITION 2.1. A real valued stochastic process $\{B_t\}_{t\geq 0}$ is a Brownian motion if it has the properties : (i) the map $t \longrightarrow B_t(\omega)$ is a continuous function of $t \in \mathbb{R}^+$ for all ω; (ii) for every $t, h \geq 0$, $B_{t+h} - B_t$ is independent of $\{B_s : 0 \leq u \leq t\}$, and has a Gaussian distribution with mean 0 and variance h.

Brownian motion is a rich and beautiful object in its own right (Rogers and Williams(1994)). Brownian motion is a martingale, a Gaussian process, a diffusion, a Levy process, a Markov process,...; Brownian motion is sufficiently concrete that one can do explicit calculations, which are impossible for more general objects; Brownian motion can be used as a building block for other processes.

The earliest attempts to model Brownian motion mathematically (Jarrow and Protter(2004)) can be traced to three sources, each of which knew nothing about the others: the first was that of T.N. Thiele of Copenhagen , who effectively created a model of Brownian motion while studying time series in 1880, (Thiele (1880)).; the second was that of L. Bachelier , who created a model of Brownian motion while deriving the evolution of the Paris asset prices, in 1900, (Bachelier (1900a,b)).; and the third was that of Einstein, who proposed a model of the motion of small particles suspended in a liquid, in an attempt to convince other physicists of the molecular nature of matter , in 1905, (Einstein (1905)).

The date March 29, 1900, should be considered as the birthdate of Mathematical Finance. On that day, a French postgraduate student, Louis Bachelier, successfully defended at the Sorbone his thesis *Theorie de la Speculation.* This work together with his subsequent was for many years neglected by the economic community but not by the probabilists such as Kolmogorov. In the present as a testimony of his great contribution , the international Finance Society is named after him. At this point we go into a little detail about what happened to Bachelier and to have a glance at the environment into which his discoveries took place.

Bachelier was born in Le Havre to a well-to-do family on March 11, 1870, (see Taqqu (2001)). His father , Alphonse Bachelier, was a wine dealer at Le Harve and his mother Cecile Fort-Meu, was a banker's daughter. But he lost his parents in 1889 and was forced to abandon his studies in order to earn a livelihood. It is known however that he register in Sorbonne in 1892. The Paris Stock Exchange, had become by 1850, the world market for the rentes, which are perpetual government bonds. It all began with " the emigrants' billion" (le milliard des emigres). During the French revolution , the nobility left and their holdings were sold as national property. When they returned in 1815, it was necessary to make restitution. Through the bonds the French state took a loan of a billion francs at the time, which was a considerable sum. The securities had a nominal value of 100 francs, but once a bond was issued, its price fluctuated. The sums that went through Paris were enormous . The French state paid always the interest but never paid the capital. When finally default appeared considerable fortunes were made and lost. These extreme fluctuations were not addressed by Bachelier in his thesis, he was merely concerned with the ordinary day-by-day fluctuations. Bachelier's subject of thesis was out of the ordinary. The "appropriate" thesis of the era for Sorbone were theses on the theory of functions (Borel, Baire, Lebesgue). Therefore, it was not an acceptable thesis topic. We must not forget that Probability as a mathematical discipline dates from after 1925 , see the special invite paper by Cramer Harald (1976) in the Annales of Probability . As usual the thesis went to Poincare, where all the thesis that at first glance did not seemed interesting, ended. The beginning of the report is as follows:

> *The subject chosen by Mr. Bachelier is somewhat removed from those which are normally dealt with our applicants. His thesis is entitled " Theory of Speculation" and focuses on the application*

of probability to the stock market. First, one may fear that the author had exaggerated the applicability of probability as is often done. Fortunately, this is not the case . In his introduction and further in the paragraph entitled " Probability in Stock Exchange Operations", he strives to set limits within which one can legitimately apply this type of reasoning. He does not exaggerate the range of his results, and I do not think he is deceived by his formulas.

It must be said that, Poincare was after the Dreyfus Affair, very doubtful that probability could be applied to anything in real life. He took a different view in 1906 after the articles of Emile Borel. Bachelier did not took the highest possible grade in his thesis and that influenced badly his academic carrier. The other factor of Bachelier's misfortune was the wrong estimate by Paul Levy on one of his research findings. Later in life Levy apologized for that but it was rather late for Bachelier. However , it was Bachelier (1906) and its extension to the multidimensional case Bachelier (1910), that prompted Kolmogorov toward the end of the 20s, to develop his theory, the analytical theory of Markov processes, Kolmogorov (1931and 1991) .

3. The decisive steps

Measure theory started with Lebesgues thesis in 1902, (see Doob (1996)) , which extended the definition of volume in $\mathbb{R}^N$ to the Borel sets. Radon (1913) made the further step to general measures of Borel sets of $\mathbb{R}^N$ (finite on compact sets). In 1913 Daniell's approach to measure theory appeared, and it was these ideas , combined with Fourier series, that N. Wiener used in 1923 to construct Brownian motion. Indeed, Wiener used the ideas of measure theory to construct a measure on the path space of continuous functions, giving the canonical path projection process the distribution of what we now know as Brownian motion.

It must be said however (Williams (1991)) that measure theory , that most arid of subjects when done for its own sake, becomes amazingly more alive when used in probability, not only because it is then applied, but also because it is immensely enriched. In Finance we need a way to mathematically model the information on which future decisions can be based . There is no other model than the appropriate σ−algebra.

You cannot avoid measure theory: Think! An *event* in probability is a measurable set, a *random variable* is a measurable function on the sample space, the expectation of a random variable is its integral with respect to the probability measure and so on. Stochastic Finance really enriches and enlivens things in the sense that we deal with lots of different $\sigma-$algebras, not just the one $\sigma-$algebra which is the concern of measure theory. Of course, intuition in the use of measure theory is much more important than the actual knowledge of technical results.

Wiener and others proved many properties of the paths of Brownian motion , an activity that continuous to this day. Two key properties are that

(1). The paths of Brownian motion have a non zero finite quadratic variation, such that on an interval (s, t) , the quadratic variations $(t-s)$ and

(2). The paths of Brownian motion have infinite variation on compact time intervals, almost surely.

In recognition of his work, his construction of Brownian motion is often referred to as the *Wiener process*. It might worth noting that the original terminology suggested by Feller (1957) in his famous treatise *An introduction to Probability Theory and its Applications* was the *Wiener-Bachelier* process.

The next Step was the creation of **Martingale theory**. Martingales are an important class of stochastic processes. The roots of the study of Martingales is in gambling. Their name comes from an old strategy used around 1815 where one at each stage doubles the stakes in any game until he wins for the first time. The name Martingale is due to J.Ville(1939). Martingales were extensively studied by Paul Levy(1886-1971) and Doob (1911- 2002), see Doob(1953).

DEFINITION 3.1. Let $\mathcal{G}$ be a filtration on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\{X_t\}$ be a stochastic process which is adapted to $\mathcal{G}$. Then $\{X_t\}$ is a martingale if for all t

(i) $\mathbb{E}(\mid X_t \mid) < \infty$

(ii) $\mathbb{E}(X_{t+1} \mid \mathcal{G}_t) = X_t$

The theory of Martingales now plays one of the most important roles in Stochastic Finance. In order to give an example of the importance of the theory of Martingales in Finance we have to explain the notion of arbitrage. Consider any market of some assets that could be traded together with a savings account with an interest rate process. We define *arbitrage* as a trading strategy that begins with no money, has zero probability of loosing money, and has a positive probability of making money. When arbitrage is present in a market then pandora's box has been opened within the market. Real markets sometimes exhibit arbitrage, but this is necessary fleeting; as soon as someone discovers it , trading takes place that removes it. It is worth mentioning at this point that all over the globe there are hunters of arbitrage called arbitrageurs who are paid to look for arbitrage in any market. The first fundamental theorem of asset pricing declares the following:

THEOREM 3.1. ***(First fundamental theorem of asset pricing).*** *If a market model has an equivalent martingale measure, then it does not admit arbitrage.*

It is very important to note that under the equivalent martingale measure all asset prices discounted by the current interest rate process are martingales. This is a result used extensively in the evaluation of any derivatives. Note that another theorem from the theory of martingales play a most decisive role in finding a hedging strategy for a trade of derivatives. In its rather simplified form this theorem states that in a model with one asset and one Brownian motion modelling the evolution of its price the existence of a hedging strategy depends on the following Theorem:

THEOREM 3.2. ***(Martingale representation , one dimension).*** *Let* B_t , $0 \leq t \leq T$, *be a Brownian motion on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, *and let* $\mathcal{B}(t)$, *be a filtration generated by this Brownian motion. Let* $M(t)$,$0 \leq t \leq T$, *be a martingale with respect to this filtration (i.e., for every* $t, M(t)$ *is* $\mathcal{B}(t)$ *−measurable and for* $0 \leq s \leq t \leq T$, $\mathbb{E}[M(t) \mid \mathcal{B}(s)] = M(s)$*). Then there is is an adapted process* $\Gamma(u)$, $0 \leq u \leq T$, *such that*

$$M(t) = M(0) + \int_0^t \Gamma(u)\, dW(u) \ \ , 0 \leq t \leq T.$$

It is finding the adapted process $\Gamma(u)$ that creates the great mathematical difficulty for any proposed model of Brownian motion.

Another almost simultaneous big step in the groundwork was **stochastic integration.** Stochastic integration was independently discovered by Kiyosi Itô and the tragic Wolfgang Doeblin .

Kiyosi Itô attempted to establish a true *stochastic differential* to be used in the study of Markov processes and with this motivation being the primary one he studied what is known as stochastic integrals. Independently the same was studied by Doeblin before him, although of course Doeblin's work was secret, hidden away in the safe of the French Academie of Science. This is a story with many messages in its own existence and worth taking the time and space to mention it briefly. Thus, in what follows we will refer to the life and mathematical legacy of Wolfgang Doeblin (for more details see Bru and Yor(2002) from where the following story was taken).

The procedure of a "Pli cachete" goes back to the very origin of the Academie des Sciences. One of the first known examples was that of the deposit by Johann Bernouli, on February 1st, 1701, of a "sealed parcel containing the problems of Isoperimetrics so that it be kept and be opened only when the solutions of the same problems by his brother, Mr. Bernouli from Basle, will appear" . A "Pli cachete", since that time, allows an author to establish a priority in the discovery of a scientific result, when he/she is momentarily unable to publish it its entirely, in a manner which prevents anybody form exercising any control, and/or asking for some paternity, over the result. This procedure continued after the creation in 1835 of the Comtes Rendus de l'Academie des Sciences which play a comparable role(to the pli cachetes) , but which, to some degree, are submitted to the judgments of peers and referees, while they do not allow in general the development of methods and proofs. This procedure is still in use today and is the subject to rules updated in 1990. These stipulate that a Pli can only be opened one hundred years after its deposit unless the author or his/her relatives explicitly demand it. Once the century has elapsed, a special commission of the Academy opens the Pli in the order of its registering and decides whether to publish it or not.

In May 2000, the sealed envelope sent in February 1940 by Wolfgang Doeblin from the front line in Lorrain to the Academy of Sciences in Paris, was finally opened. This was a long -awaited event for researchers in probability, with some interest in the history of their field, and who had in the past been struck by the modernity of the ideas of Wolfgang Doeblin .

The pli has now been published in its entirety in the *Comptes Rendus* of the Academie des Science as a Special Issue, dated December 2000, and this seems to have awakened interest in both Wolfgang Doeblin's life and work.

Wolfgang Doeblin was born on the 17th of March 1915, in Berlin. His father Alferd Doblin(1878-1957), who belonged to Jewish family, was a physician and was starting to get a name in the vanguard of German literature. He became famous in 1929 once his novel *Berlin Alexanderplatz* was published. The Doeblin family was forced into exile in March 1933 and after a short time in Zurich , the Doeblins settled in Paris. At the end of 1935, he carried out research about theory of Markov chains under the guidance of Maurice Frechet. The young Doeblin very quickly obtained some most remarkable results. Lindvall (1993, pp. 55-56) quotes K.L. Chung's review of (Levy 1955) in the Math. Reviews:

> *After all there can be no greater testimony of a man's work than its influence on others. Fortunately , for Doeblin, this influence has been visible and is still continuing. On limit theorems his work has been complemented by Gnedenko and other Russian authors. On Markov processes it has been carried on mostly in the United States byDoob, T.E.*

Harris and the reviewer. Here his mine of ideas and techniques is still being explored.

At the age of 23 years and with only two years of active research behind him, Doeblin's performance must be considered unique, probably since Laplace (see Bru and Yor (2002)).

Wolfgang Doeblin, together with his parents and his two younger brothers Claude and Stephan, acquired French citizenship in 1936. After defending his famous thesis in Mathematics , Doeblin (1938) in Spring 1938 , he was enlisted for two years military service which had been deferred for the duration of his studies. Getting depressed by the barracks routine life, he stopped all his mathematical work for four months. After that he was trying very hard , as he wrote to Frechet, to "fight against depression. As I am not interested in alcohol, I cannot resort to getting drunk." Mathematics as a therapeutic against the blues, a nice Pascalian theme. In any case , the possibilities of intensive intellectual work were quite limited . In a letter dated November 12th ,1939 , Doeblin informed Frechet that he had started work again " oh! not much, about one hour every day" during the night when the others went to sleep. Doeblin had no scientific document at hand and no place to work apart from the telephone booth.

During the first days of November 1939 in a small village of the Ardennes, went out to by a school exercise book of 100 pages and began to write down the development of his note "Sur l equation de Kolmogorov" . The first pages of the Pli indicate that this was a form of therapy which the author imposed upon himself. In the middle of January 1940, the dream of an early end of the war was brutally replaced by reality, with the "alert on Belgium". It may well have been in Athienville, probably around the middle of February, that Doeblin finished writing the Pli. He would then have sent it to the Academie. At the same time as the Pli a second paper was send which was presented by Borel on March 4th, Doeblin (1940). His spirits remained high, one reason being that, at long last, he may possibly have obtained a leave in the middle of March, which he may then have put to profit by going to the Institute Henri Poincare (IHP) to look for the memoirs of Hostinsky he needed. Doeblin continued sending papers to the Comptes Rendus de la Academie de Sciences as the German offensive progressed.

During the night of June 20th to 21st, as the remains of his decimated regiment are in Vosges, completely encircled by German troops and surrender is imminent , the already decorated soldier Doeblin who, according to the opinion of his superiors, has always been a "constant model of bravery and devotion" , leaves his company and tries to escape on his own. After walking all night long , he finds himself inside the German net in the village of Housseras. Wolfgang enters a farm, which belongs to the Triboulot family. There, without saying a word, he burns all his papers in the Kitchen stove. He then comes out of the farm building , enters the barn and shoots himself in the head.

Thus, if we lend the conclusion , from Bernard Bru and Mark Yor, Wolfgang Doeblin wanted to disappear in silence. Among his burnt papers, there may have been his "research note book" in which he had always jotted down new questions to study, ideas to develop... and which has not been found. The Nazis has burnt the works of his father and had forced the family into exile. For Wolfgang Doeblin , there remained the ultimate freedom to burn his papers himself and to kill himself in order to preserve his ideal of life and the beauty of his work.

We now turn to Kiyosi Itô who's first paper on stochastic integration was published in 1944 , Itô (1944). Itô has explained his motivation himself Itô (1987), and we let him express it: "In the papers by Kolmogorov(1931) and Feller(1936), I saw a powerful analytic method to study the transition probabilities of the process, namely Kolmogorov's parabolic equation and his extension by Feller. But I wanted to study the paths of Markov processes in the same way as Levy observed differential processes. Observing the intuitive background in which Kolmogorov derived his equation, I noticed that a Markovian particle would perform a time homogeneous differential process for infinitesimal future at every instant, and arrived at the notion of a stochastic differential equation governing the paths of a Markov process that could be formulated in terms of the differentials of a single differential process".

Let us now spent some time and space in order to understand some of the basic problems of stochastic integration and its interrelation with financial problems. We fix a positive number T and we are looking to find

$$\int_0^T \Delta(t)\, dW(t),$$

where $W(t)$, $t \geq 0$ is a Brownian motion or a Wiener process together with a filtration $\mathcal{B}(t)$ for this Wiener process. We will let the integrand $\Delta(t)$ be an adapted stochastic process. Our reason for doing this is that $\Delta(t)$ will eventually be the position we take in an asset at time t, and this typically depend on the price path of the asset up to time t. Requiring $\Delta(t)$ to be adapted means that we require $\Delta(t)$ to be $\mathcal{B}(t)$-measurable for each $t \geq 0$. Recall that increments of the Brownian motion after time t are independent of $\mathcal{B}(t)$ and since $\Delta(t)$ is $\mathcal{B}(t)$-measurable , it must also be independent of these future Brownian increments. Positions we take in assets may be independent on the price history of those assets , but they must be independent of the future increments of the Brownian motion that drives those prices. One of the problems we face when trying to assign meaning to the Ito integral is that Brownian motion paths cannot be differentiated with respect to time. The other basic problem is that if we consider a partition of $[0, T]$; i.e.,

$$0 = t_0 \leq t_1 \leq .... \leq t_n = T,$$

and take the Riemann sum

$$\sum_{i=0}^{n-1} \Delta(t_i) [W(t_i + 1) - W(t_i)]$$

then given the $\sigma-$algebra $\mathcal{B}(t_i)$, $W(t_i + 1)$ still remains a random variable and that makes the above Riemann sum a random variable. At this point to resolve the problem Ito made the logical step for a probabilists. Instead of taking the limit of the Riemann sum as the partition grows larger in number of points, which was not possible in this case , he took the convergence in mean square and thus he defined stochastic integration. Naturally , some conditions where necessary to guaranty its existence and these are given in the next Theorem.

THEOREM 3.3. *Let T be a positive constant and let $\Delta(t)$, $0 \leq t \leq T$, be an adapted stochastic process that satisfies the condition*

$$\mathbb{E}\left[\int_0^T \Delta^2(t)\, dt \mid \mathcal{B}(t)\right] < \infty.$$

Then

$$I(t) = \int_0^T \Delta(t)\, dW(t),$$

has the following properties:

(i) (Continuity) As a function of the upper limit of integration t, the paths of $I(t)$ are continuous.

(ii)(Adaptivity) For each t, $I(t)$ is $\mathcal{B}(t)-$measurable.

(iii) (Linearity) If $I(t) = \int_0^T \Delta(t)\, dW(t)$ and $J(t) = \int_0^T \Gamma(t)\, dW(t)$, then $I(t) \pm J(t) = \int_0^T [\Delta(t) \pm \Gamma(t)]\, dW(t)$; furthermore, for every constant c, $cI(t) = \int_0^T c\Delta(t)\, dW(t)$.

(iv) (Martingale) $I(t)$ is a martingale.

(v)(Isometry) $\mathbb{E}\left[I^2(t)\right] = \mathbb{E}\left[\int_0^T \Delta^2(t)\, dt\right]$.

(vi)(Quadratic Variation) $[I, I](t) = \int_0^T \Delta^2(t)\, dt$.

Naturally, it is not possible to find the stochastic integral of various integrands as limits of expected mean squares. For that goal the following Ito Doeblin formula is used:

THEOREM 3.4. ***(Ito-Doeblin formula for Brownian motion).*** *Let $f(t,x)$ be a function for which the partial derivatives $f_t(t,x)$, $f_x(t,x)$,and $f_{xx}(t,x)$ are defined and continuous , and let $W(t)$ be a Brownian motion. Then, for every $T \geq 0$,*

$$f(T, W(T)) = f(0, W(0)) + \int_0^T f_t(t, W(t))dt + \int_0^T f_x(t, W(t))\, dW(t) + \frac{1}{2}\int_0^T f_{xx}(t, W(t))\, dt.$$

J.L.Doob realized that Ito's construction of his stochastic integral for Brownian motion did not use the full strength of the independence of the increments of Brownian motion(Jarrow and Protter(2004)). In his highly influential 1953 book he extended Ito's stochastic integral for Brownian motion first to processes with orthogonal increments (in the L^2 sense) , and then to processes with conditionally orthogonal increments,that is, martingales. What he needed, however , was a martingale M such that $M^2(t) - F(t)$ is again a martingale, where the increasing process F is non-random. He established the now famous Doob decomposition theorem for submartingales:

THEOREM 3.5. *If X_n is a (discrete time) submartingale, then there exists a unique decomposition $X_n = M_n + A_n$ where M is a martingale, and A is a process with non-decreasing paths, $A_0 = 0$, and with the special measurability property that A_n is $\mathcal{F}_{n-1}$ measurable.*

Since M^2 is a submartingale when M is a martingale, he needed an analogous decomposition theorem in continuous time in order to extend further his stochastic integral. As it was , however, he extended Ito's isometry relation as follows:

$$\mathbb{E}\left[\left(\int_0^T H_t dM_t\right)^2\right] = \mathbb{E}\left(\int_0^T H_t^2 dF(t)\right),$$

where F is a non-decreasing and non-random, $M^2 - F$ is again a martingale, and also the stochastic integral is also a martingale.(See Doob(1953). Thus it became an interesting question, if only for the purpose of extending the stochastic integral to martingales in general, to see if one could extend Doob's decomposition theorem to submartingales indexed by continuous time.

The issue was resolved in two papers by the (then) young French mathematician P.A. Meyer in 1962 (Jarrow and Protter (2004). Indeed, as if to underline the importance of probabilistic potential theory in the development of the stochastic integral, Meyer's first paper, establishing the existence of the Doob decomposition for continuous time submartingales (Meyer(1962)) , is written in the language of potential theory. Meyer showed that the theorem is false in general, but true if and only if one assumes that the submartingale has a uniform integrability property when indexed by stopping times, which he called " Class (D) " , clearly in honor of Doob. Ornstein (see for example Meyer (2000) had shown that there were submartingales not satisfying the Class(D) property, and G. Johnson and Helms (1963) quickly provided an example in print , using three dimensional Brownian motion. Also in 1963, Meyer established the uniqueness of the Doob decomposition , which today is known as the Doob-Meyer decomposition Theorem. In addition, in this second paper Meyer provides an analysis of the structure of L^2 martingales, which later will prove essential to the full development of the theory of stochastic integration. Two years later, in 1965, Ito and Watanabe, while studying multiplicative functionals of Markov processes, define *local martingales(1965).* This turns out to be the key object needed for Doob's original conjecture to hold. That is, any submartingale X, whether it is of Class (D) or not , has a unique decomposition

$$X_t = M_t + A_t,$$

where M is a local martingale, and A is a non- decreasing, predictable process with $A_0 = 0$.

Important parallel developments were occuring in the Soviet Union (Jarrow and Protter (2004)) . The books of Dynkin on Markov processes appeared early, in 1960 and in English as Springer Verlag books in 1965. A **decisive step** was the work by Girsanov(1960) on transformation of Brownian motion which extends the much earlier work of Cameron and Martin (1949) and Maruyama (1954). It was not until VanSchuppen and Wong (1974) that these results were extended to martingales, followed by Meyer(1976) and Lengart (1977) for the current modern versions. The version which more often is applied in financial problems is the following.

THEOREM 3.6. ***(Girsanov one dimension).*** *Let $W(t)$, $0 \leq t \leq T$, be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{B}(t)$, $0 \leq t \leq T$, be a filtration for this Brownian motion. Let $\Theta(t)$, $0 \leq t \leq T$, be an adapted process. Define*

$$Z(t) = \exp\left\{ -\int_0^t \Theta(u)\, dW(u) - \frac{1}{2}\int_0^t \Theta^2(u)\, du \right\},$$

$$\hat{W}(t) = W(t) + \int_0^t \Theta(u)\, du,$$

and assume that

$$\mathbb{E}\left[\int_0^T \Theta^2(u)\, Z^2(u)\, du < \infty\right].$$

Then

$$\mathbb{E}[Z(t)] = 1$$

and under the probability measure $\hat{\mathbb{P}}$ *given by*

$$\hat{\mathbb{P}}(A) = \int_A Z(\omega)\, d\mathbb{P}(\omega) \quad \text{for all } A \in \mathcal{F},$$

the process $\hat{W}(t)$, $0 \leq t \leq T$, *is a Brownian motion.*

It was the work of Doleans-Dade and Meyer (1970) that removed the

assumption that the underlying filtration of σ- algebras was quasi left continuous or alternatively stated as saying that the filtration had no fixed times of discontinuity thus making the theory a pure martingale theory. This can now be seen as **a key step** that led to the **fundamental papers in finance** of Harrison and Kreps (1979) and Harrison and Pliska (1981,1983). Harrison and Kreps paper was referred 1462 times according to the web of science and Harrison and Pliska's paper 1291 times. Last, in the same paper Doleans-Dade and Meyer coined the modern term *semimartingale*, to signify the most general process for which one knew (at that time) there existed a stochastic integral.

In 1969, Robert Merton introduced stochastic calculus into the study of Finance. Merton was motivated by the desire to understand how prices are set in financial markets, which is the classical economics question of "equilibrium" , and in later papers used the machinery of stochastic calculus to begin investigation of this issue. The fact that the world had seen the emergence of a new scientific discipline, *Mathematical Finance* , *Stochastic Finance, or Theory of Finance* was reflected by awarding Harry Markowitz, William Sharpe, and Merton Miller the 1990 Nobel Prize in Economics. The genesis of this science has been verified by the awarding of the 1997 Nobel Prize in Economics the formal press release of which from the Royal Academy of Sciences was the following:

For a new method to determine the value of derivatives. Robert C. Merton and Myron S. Scholes have in collaboration with the late Fisher Black, developed a pioneering formula for the valuation of stock options. Their methodology has paved the way for economic valuations in many areas. It has also generated new types of financial instruments and facilitated more efficient risk management in *society.*

The 1997 Nobel price was awarded for their papers Black and Scholes (1973) which has 2793 citations in the web of science and Merton (1973) which was cited 1323 times, followed by Merton (1974) seminal paper which introduced the theory of credit risk. The formal press release although true , is just the proverbial of the iceberg (see Jarrow (1999)). The impact of the Black-Merton-Scholes model , is greater than most people realize. Their work on option pricing has not only provided a technique for valuation, but has also created a new field within finance, known as derivative, and offered a new perspective on related areas including corporate finance, capital budgeting, and financial markets and institutions. In mathematics and computer science , the direction of study in probability theory and numerical methods has been influenced by problems arising from the use of option pricing technology. In private industry, the Black-Merton-Scholes option pricing theory has generated not just "new types of financial instruments" , but also new organizational structures within corporations to help manage risks. Research in stochastic processes and numerical methods has been financed within large investment corporations, the results of which are not known since they are highly classified by them. Mathematics and Engineering departments have recently introduced masters programs specializing in derivatives and mathematical finance. In the last fifteen years mathematicians and theoretical physicists can now find alternate and high-paying demand for their skills in the financial world. Note though , that competition for these jobs is fierce and the better your skills on mathematics the better are your chances. In addition , there is no limit in the working hours per week a young researcher has to provide and as for job security the policy is hire and fire in correlation with the many turbulences of the international market.

There are four basic types of option contracts: European calls, European puts, American calls and American puts. Perhaps surprisingly, the option prefixes have nothing to do with geographical considerations. A European call option gives its holder the right, but not the obligation, to purchase an asset at a fixed price -called the "strike" or "exercise" price- at a fixed future date -called "maturity" or "expiration" date. A rational holder will, therefore, only exercise the option to purchase at the maturity date if the asset price at that time exceeds the exercise price. A European put option differs from a European call option in that it gives the right to sell, rather than the right to buy, the underlying stock. An American call option differs from a European call option in that it gives the right to buy at an time after entering the contract and up until and including the maturity date.

Consider an agent who at each time t has a portfolio valued at $X(t)$. This portfolio invests in a money market account paying a constant rate of interest r and in an asset modeled by the geometric Brownian motion

$$dS(t) = aS(t)\,dt + \sigma S(t)\,dW(t).$$

Suppose at each time t, the investor holds $\Delta(t)$ shares of stock. The position $\Delta(t)$ can be random but must be adapted to the filtration associated with the Brownian motion $W(t), t \geq 0$. According to the Itô-Doeblin formula the differentials of the discounted asset price and the discounted portfolio value is

$$d(e^{-rt}S(t)) = (a-r)e^{-rt}S(t)\,dt + \sigma e^{-rt}S(t)\,dW(t),$$

$$d(e^{-rt}X(t)) = \Delta(t)\,d\left(e^{-rt}S(t)\right).$$

Consider a European call option that pays $(S(T)-K)^+$at time T. Let the stochastic process $c(t, S(t))$ be the value of the option at time t. Then Black-Merton-Scholes proved (Shreve(2004)), that should satisfy the partial differential equation

$$rc(t, S(t)) = c_t(t, S(t)) + rS(t)\,c_x(t, S(t)) + \frac{1}{2}\sigma^2 S^2(t)\,c_{xx}(t, S(t))$$

for all $t \in [0, T]$ and that satisfies the *terminal condition*

$$c(T, S(T)) = (S(T)-K)^+.$$

4. A brief glance to the flow of research paths

As mentioned earlier two are the basic assumptions underlying the Black-Merton-Scholes model, the constant risk -free interest rates and a constant volatility for the underlying asset. In April 1973 , around the time of the publication of the Black -Merton- Scholes model , the Chicago Board Options Exchange began trading the first listed options in the United States. Since that time, the growth in exchange traded and over the counter traded options on equities, indices, foreign currencies, commodities, and interest rates has been phenomenal. In response to these new derivatives markets , new firms were created and new departments in existing firms and banks were formed to take advantage of these new trading opportunities.

From that point of time i.e. since 1973 , we have witnessed a tremendous acceleration in research efforts aimed at better comprehending, modeling and hedging all risks involved. Later through the machinery of the Theory of Martingales and Gyrsanov's theorem, martingale methods have been constructed which generalized considerably these assumptions (see Musiela and Rutkowski (1998) and Schreve (2004)) . Generalizations included models in which volatility was random and models in which asset prices jumped , rather than moving smoothly. In the 1980's increased interest rate volatility occurred due to double-digit inflation . That created a new demand for interest rate derivatives for both motives insurance and speculation. In this type of problems the seminal paper is that of Heatth, Jarrow and Morton (1992) a paper which has about 1286 citations in the web of science. Various stochastic process models have been created which we will briefly mention in what follows. The book by Brigo and Mercurio (2004) on interest rate models is one that combines a strong mathematical background with expert knowledge of practice. This simultaneous attention is difficult to find in other available literature. Local volatility models have been introduced as a straight forward analytical extensions of a geometric Brownian motion that allow skews in the implied volatility. Another excellent book in the area is that of Rebonato (1998) . The more flexible models of this type , allowing for smile-shaped implied volatilities, have been proposed by Brigo, Mercurio and Sartorelli (2003) and Brigo and Mercurio (2003). The already briefly mentioned stochastic volatility models where the volatility is assume to follow a diffusion process have as main representatives the works of Hull and White (1987). and Heston (1993), with the related application to the LIBOR market model developed by Wu and Zhang (2002). Another class of models are the Jump- Diffusion models which have been introduced to model discontinuities in the underlying stochastic process, namely the possibility of finite changes in the value of the related financial variable over infinitesimal time intervals. Discontinuous dynamics seem ideally suited for the interest rate market, where short-term rates can suddenly jump due to central banks interventions. The first example of Jump- Diffusion models in the financial literature is due to Merton. Jump diffusion Libor models have been developed by Glasserman and Merener (2001) and Glasserman and Kou (2003). Finally another interesting class of models are the Levy-driven models. These have been designed to allow for stochastic evolutions governed by general Levy processes. A book with the applications of Levy Processes in Finance for pricing financial derivatives is Schoutens (2003).

It is interesting to stress the common standard technical assumptions for the above mentioned models and the ones to follow:

(i). All reference probability spaces are assumed to be complete (with respect to the reference probability measure).

(ii). All filtrations satisfy the *usual conditions* of right -continuity and completeness (see Karatzas and Shreve (1991) and (1998)).

(iii). The sample paths of all stochastic processes are right- continuous functions, with finite left-limits, with probability one; in other words, all stochastic processes are assumed to be cadlag.

(iv). All random variables and stochastic processes satisfy suitable integrability conditions, which ensure the existence of considered conditional expectations, deterministic or stochastic integrals, etc.

Another large area of the Theory of Finance is the one that deals with *default risk* (see Bielecki and Rutkowski 2004). A default risk is a possibility that a counterpart in a financial contract will not fulfill a contractual commitment to meet her/his obligations stated in the contract. If this actually happens, we say that the party defaults, or that the default event occurs. More generally, by credit risk we mean the risk associated with any kind of credit - linked events, such as : changes in the credit quality (including downgrades or upgrades in credit ratings), variations of credit spreads, and the default event. There are two kinds of credit risks the *reference credit risk and the counterpart credit risk.* In the reference credit risk the two parties of the contract are default-free but some reference entity in the contract which plays an important role appears to produce a default risk. *Credit derivatives* are recently developed financial instruments that allow market participants to isolate and trade the reference credit risk. In counterpart credit risk each counterpart is exposed to the default risk of the other party. The counterpart risk emerges in a clear way in such contracts as *vulnerable claims* and *default swaps*. In both of these cases one needs to quantify the default risk of both parties in order to correctly assess the contracts value. A corporate bond is an example of a defaultable claim.

A vast majority of mathematical research is devoted to the credit risk is concerned with the modeling of the random time when the default event occurs, i.e. the default time. Two competing methodologies have emerged in order to model the default/ migration times and the recovery rates: *the structural approach and the reduced-form approach.* Structural models are concerned with modeling and pricing credit risk that is specific to a particular corporate obligor. Credit events are triggered by movements of the firm's value relative to some (random or non random) credit -event-triggering threshold (or barrier) . From the long list of works devoted to structural approach, let us mention in here: Merton (1974) , Black and Cox (1976) , Ericson and Renedy (1998) , Ericson (2000).

In the Reduced-form models approach, the value of the firm's asset and its capital structure are not modeled at all, and the credit events are specified in terms of some exogenously specified jump process. We can distinguish between the reduced form models that are concerned with the modeling of the default time, and that are henceforth referred to as *intensity - based models*, and the reduced form models with migrations between credit rating classes, called the *credit migration models.* The main emphasis is put on the modeling of the random time of default as a hazard process, as well as evaluating conditional expectations under risk-neutral probability of functionals of the default time and the corresponding cash flows. Interesting works in this respect that pioneered the area are Pye

(1974) , Ramaswamy and Sundaresan (1986), Jarrow and Turnbull (1995) , Lando (1997, 1998).

The credit migration models assume that the credit quality of corporate debt is quantified and categorized into a finite number of disjoint credit rating classes. Each credit class is represented by an element in a finite set one of which is the default state. The assumed process for the evolution of the credit quality is referred to as the migration process. The main issue in this approach is is the modeling of the transition intensities under the real world probabilities , the equivalent martingale measure and the forward measure. The next step is the evaluation of conditional expectations under the equivalent martingale measure and the forward measure of certain functionals , typically related to the default time. The most highly cited papers in the area are those of Jarrow and Turnbull (1995) , and Jarrow , Lando and Turnbull (1997). References dealing with the stochastic modeling of credit migrations include Duffie and Singleton (1998), Kijima (1998), Thomas et al. (1998) , Huge and Lando (1999), Bielecki and Rutkowski (2000), Lando (2000) ,Schonbucher (2000) and Vasileiou and Vassiliou(2006).

REFERENCES

BACHELIER, L. (1900). Theorie de la speculation. *Annales Scientifiques de l'Ecole Normale Superieure, 21-86.*

BACHELIER, L. (1900). *Theorie de la speculation.* Gautiers-Villars. Paris.

BACHELIER, L. (1906). Theorie des probabilites continues. *Journal de Mathematiques Pures et Appliquees, 2: 259-327.*

BACHELIER, L. (1910). Les probabilites a plusiers variables. *Annales Scientifiques de l' Ecole Normale Superieure, 27: 339-360.*

BIELECKI, T.R. AND RUTKOWSKI, M. (2000). Multiple ratings model of defaultable term structure. *Math. Finance* , 10, 125-139.

BIELECKI, T.R. AND RUTKOWSKI, M. (2004). *Credit Risk: Modeling, Vluation and Hedging.* Springer.

BLACK , F. AND SCHOLES (1973). The pricing of options and corporate liabilities. *Journal of Political Economy. 81, pp.637-59.*

BLACK , F. AND COX, J.C. (1976). Valuing corporate securities: Some effects of bond indenture provisions. *J.Finance* 31, 351-367.

BRIGO, D. AND MERCURIO, F. (2000). *Interest Rate Models- Theory and Practice . With Smile, Inflation and Credit.* Part IV-VIII. Second edition. Springer.

BRIGO, D. AND MERCURIO, F. AND SARTORELLI, G. (2003). Alternative Asset-price Dynamics and Volatility Smile. *Quantitative Finance* 3(3), 173-183.

BRU, B. AND YOR, M. (2002). Comments on the life and mathematical legacy of Wolfgang Doeblin. *Finance and Stochastics.* 6, 3-47.

CAMERON, R.H. AND MARTIN, W.T. (1949). Transformation of Wiener integrals by non-linear transformations. *Transactions of the American Math. Society* 66, 253-283.

CRAMER, H. (1976). Half a century with probability theory: Some personal recollections . *The Annals of Probability* . Vol 4, 4, 509-546.

DOEBLIN, W. (1938). Sur les proprietes asymptotiques de mouvement regis par certains types de chaine simples. *These Sci. Math. Paris, Bucarest: Imprimerie Centrale,* 1938, *et Bull. Soc. Math. Roumaine Sci.* 39 (1), 57-115, 39 (2) 3-61.

DOEBLIN, W. (1940). Sur l'equation de Kolmogoroff. *C.R. Acad. Sci. Paris,* 210 pp.365-367.

DOLEANS-DADE, C. AND MEYER, P.A. (1970). Integrales stochastiques par raport aux martingales locales, Seminaires de Probabilites IV, Lecture Notes in Mathematics, 124, 77-107.

DOOB, J.L. (1953). *Stochastic Processes.* John Wiley.

DOOB, J.L. (1996). The development of rigor in Mathematical Probability (1900- 1950). *The American Mathematical Monthly.* Vol. 103, No, 7, pp. 586-595.

DUFFIE, D. AND SINGLETON, K.J. (1998). Rating based term structures of credit spreads. Working paper , Stanford University.

DYNKIN, E. (1960). Theory of Markov Processes. Pergamon Press, Oxford.

DYNKIN, E. (1965). Markov Processes (two volumes). Springer-Verlag, Berlin.

EINSTEIN, A. (1905). On the movement of small particles suspended in stationary liquid demanded by the molecular-kinetic theory of heat, *Ann. d. Physic* **17.**

ERICSSON, J. AND RENEDY, J. (1998). A framework for valuing corporate securities. *Appl. Math. Finance* 5, 143-163.

ERICSSON, J. (2000). Asset substitution, debt pricing, optimal leverage and maturity. *Finance* 21 (2), 39-70.

FELLER, W. (1957). *An Introduction to Probability and its Applications*, vol. 2. New York: Wiley.

GLASSERMAN, P. AND MERENER, N. (2001). Numerical solutions of Jump-Diffusion LIBOR market models. Working paper, Columbia University.

GLASSERMAN, P. AND KOU, S.G. (2003). The term structure of simple forward rates with jump risk. Mathematical Finance. 13 ,(3), 383-410.

GIRSANOV, I.V. (1960). On transforming a certain class of stochastic processes by absolutaly continuous changes of measures. *Theory Probab. Appl.*, 5, 285-301.

HARRISON, J.M. AND KREPS, D.M. (1979). Martingales and Arbitrage in multiperiod security markets. *J. Econ. Theory* . 20, 381-408.

HARRISON, J.M. AND PLISKA, S.R. (1981). Martingales and stochastic integrales in the theory of continuous trading. *Stochastic Process Appl.* 11, 215-260.

HARRISON, J.M. AND PLISKA, S.R. (1983). A stochastic calculus model of continuous trading : complete markets. *Stochastic Process Appl.* 15, 313-316.

HEATH, D., JARROW, R. AND MORTON, A. (1992). Contingent Claim Vluation with Random evolution of Interest Rates. *Econometrica.* 60 :1 pp. 77-105.

HESTON , S.L. (1993). A Closed -Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies. 6*, 327-343.

HUGE, B. AND LANDO, D. (1999). Swap pricing with two sided default risk in a rating based model. *European Finance Review*,3, 239-268.

HULL, J. AND WHITE, A. (1987). The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Financial and Quantitative Analysis 3*, 281-300.

ITO, K. (1944). Stochastic integral. *Proc. Imp. Acad. Tokyo* 20, 519-524.

ITO, K. (1987). *Forward, K. Itô Collected Papers.* Springer-Verlag, Heidelberg.

ITO , K. AND WATANABE, S. (1965). Transformation of Markov processes by multiplicative functionals. *J. Math. Kyoto Univ.* 4, 1-75.

JARROW, R.A. AND TURNBULL, S.M. (1995). Pricing derivatives on financial securities subject to credit risk. *J. Finance* 50, 53-85.

JARROW, R.A. LANDO, D. AND TURNBULL, S.M. (1997). A Markov model for the term structure of credit risk spreads. *Rev. Finan. Stud.* 10, 481- 523.

JARROW, R.A. AND PROTTER, P. (2004). A short history of stochastic integration and mathematical finance the early years, 1880-1970 . *IMS Lecture Notes Monograph*, Vol 45 (2004) 1-17.

JOHNSON, G. AND HELMS, L.L. (1963). Class (D) Supermartingales. *Bull. American Math. Society 69*, 59-62.

KARATZAS, I., SHREVE, S. (1991). *Brownian motion and Stochastic Calculus.* 2^{nd} edition. Springer-Verlag, Berlin.

KARATZAS, I., SHREVE, S. (1998). *Methods of Mathematical Finance.* Springer-Verlag, Berlin.

KIJIMA, M. (1998). Monotonicity in a Markov chain model for valuing coupon bond subject to credit risk. *Mathematical Finance* 8, 229-247.

KOLMOGOROV, A.N. (1931). Uber die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematishe Annalen*, 104: 415-458.

KOLMOGOROV, A.N. (1991). *Selected works.* Kluer Academic Publishers, Dordrecht.

LANDO, D. (1997). Modelling bonds and derivatives with credit risk. In: *Mathematics of Derivative securities*, M. Dempster, S. Pliska, eds. Cambridge University Press, Cambridge, pp. 369-393.

LANDO, D. (1998). On Cox processes and credit -risky securities. *Rev. Derivatives Res.* 2, 99-1210.

LANDO, D. (2000). Some elements of rating-based credit risk modeling. In : Advanced Fixed -Income ValuationTools, N. Jegadeesh, B. Tuckman, eds., J.Wiley.

LENGLART, E. (1977). Transformation des martingales locales par chagement absolument continu de probabilites, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 39, 65-70.

LINDVALL, T. (1993). Sannolikheten och odet. Om matematiken Wolfgang Doeblin, Ord and Bild 6, 48-57.

LEVY, P (1955). Wolfgang Doeblin (V. Doblin) (1915-1940) . *Rev. Histoire Sci.* , pp. 107-115.

MARUYAMA, G. (1954) . On the transition probability functions of Markov processes, *Nat. Sci. Rep. Ochanomizu Univ.* 5, 10-20.

MERTON, R.C. (1969). Lifetime portofolio selection under uncertainty:the continuos -time case, *Rev. Econ. Stat.* 51, 247-257.

MERTON, R.C. (1973). Theory of rational option pricing. *Bell. Journal of Economics and Management Science.* 4,pp. 141-83.

MERTON, R.C. (1974). On the pricing of Corporate debt: The risk structure of interest rates. *Journal of Finance. 29:2* pp.449-70.

MERTON, R.C. (1976). Option Pricing When Underlying Stock Returns are Distcontinuous. *Journal of Financial Economics* 3, 125-144.

MEYER, P.A. (1962). A decomposition theorem for supermartingales, III. *J. Math.* 6, 193-205.

MEYER, P.A. (1963). Decomposition of supermartingales: the uniqueness therom, III. *J. Math.* 7, 1-17.

MEYER, P.A. (1976). Un cours sur les integrales stochastiques. Seminaire de Probabilites X, *Lecture Notes in Mathematics*, 511, 246-400.

MEYER, P.A. (2000). Les Processus Stochastiques de 1950 a Nos Jours, in Development of Mathematics 1950-2000, edited by Jean-Paul Pier ; Birkhauser, Boston,MA. 813-848.

MUSIELA, M. AND RUTKOWSKI, M. (1998). *Martingale Methods in Financial Modeling.* Springer.

REBONATO, R. (1998). *Interest rate option models.* Second Edition. John Wiley.

ROGERS, L.C.G. AND WILLIAMS, D. (1994). Diffusions, Markov Processes and Martingales, Volume 1, Foundations. John Wiley.

SCHONBUCHER, P.J. (2000). *Credit Risk Modeling and Credit Derivatives.* Ph.D. disertation, University of Bonn.

SCHOUTENS, W. (2003). Levy Processes in Finance. Pricing Financial Derivatives. Wiley.

SHREVE , S (2004). *Stochastic Calculus for Finance II. Cintinuous - time models.* Springer .

SUR L' EQUATION DE Kolmogorov , Pli cachete depose le 26 fevrier 1940, ouvert le 18 mai 2000 . C.R. Acad. Sci. Paris, Serie I,331, 1031-1087 (2000).

TAQQU, M.S. (2001). Bachelier and his Times: A Conversation with Bernard Bru (2001). *Finance and Stochastics*, 5(1): 3-32, 2001.

THIELE, T.N. (1880). Sur la compensation de quelques erreurs quasi-systematiques par la method des moindres carres. Reitzel, Copenhagen.

PYE, G. (1974). Gauging the default premium. *Finan. Analysts. J.* 30(1), 49-52.

VAN SCHUPPEN, J.H. AND WONG, E. (1974). Transformations of local martingales under a change of law. *Annals of Probability 2*, 879-888.

VASILEIOU, AGLAIA AND VASSILIOU, P.-C.G. (2006). An inhomogeneous semi-Markov model for the term structure of credit risk spreads. *Advances in Applied Probabilities.* 38,1, 171-198.

VILLE, J. (1939). *Etude critique de la notion de collectif.* Gauthier-Villars, Paris.

WILLIAMS, D. (1991). Probability with Martingales. Cambridge Mathematical Textbooks.

WU, L. AND ZHANG, F. (2002). Libor Market Model: from Deterministic to Stochastic Volatility. Working Paper. Claremont Graduated University and Hong Kong University of Science and Technology.

Chapter 2.

Background in Probability and Stochastic Processes

2.1. Introduction.

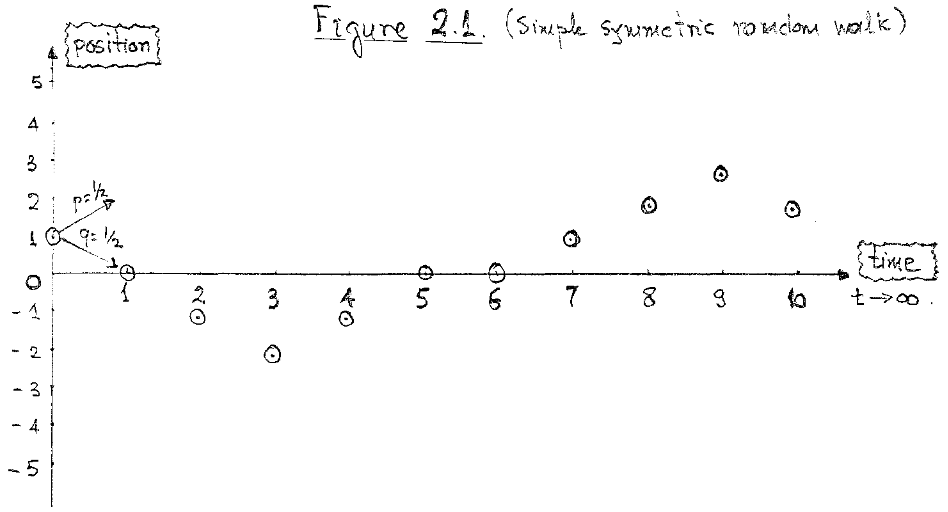
This chapter serves as a memory refresher for some, or a quick preview for others but not as an introduction to probability theory or stochastic processes. It contains all the known results in probability and stochastic processes which are used, sometimes without further reference, in the main body of the book.

Special weight will be attached to infinite probability spaces which are used to model situations in which a random experiment with infinitely many possible outcomes is conducted. The sample spaces which we are interested to study are not only infinite but are uncountably infinite (i.e. it is not possible to list their elements in a sequence). The first problem we face with an uncountably infinite sample space is that, for most interesting experiments, the probability of any particular outcome is zero. Consequently, we cannot determine the probability of a subset A of the sample space, a so called event, by summing up the probabilities of the elements in A. We must instead define the probabilities of events directly. But in infinite sample spaces there are infinitely many events. Even though we may understand well what random experiment we want to model, some of the events may have such complicated descriptions that it is not obvious what their probabilities should be. It would be hopeless to try to give a formula that determines the probability for every subset of an uncountably infinite sample space. We instead give a formula for the probability of certain simple events and the appeal to the properties of probability measures to determine the probability of more complicated events. For the purpose of gaining in depth for the concepts that will follow in respect with infinite probability spaces, there are two such experiments to keep in mind

(i) Choose a number from the unit interval $[0,1]$.

(ii) A discrete random walk with steps $+1$ and -1 with equal probabilities which goes on almost indefinitely (figure 2.1)

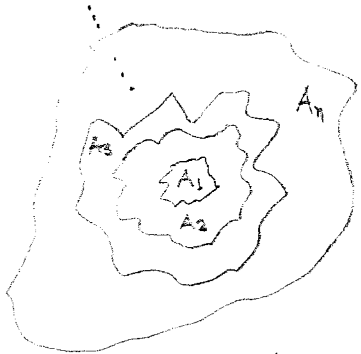
Figure 2.1. (Simple symmetric random walk)



2.2. Some Set Theory

Let Ω be the set of all possible outcomes of an experiment which we call the sample space and let $\omega \in \Omega$ be the generic element. Let $\mathcal{F}(\Omega)$ be the collection of all subsets of Ω including set ϕ (empty). The union of A and B in $\mathcal{F}(\Omega)$ is denoted by $A \cup B$, the intersection by $A \cap B$, the complement of A by A^c or $\bar{A}$. When A and B are disjoint ($A \cap B = \phi$), the union $A \cup B$ is also denoted $A + B$ and is called the sum of A and B. When A is included in B ($A \subset B$), the strict difference $B - A$ is the set of elements ω of Ω belonging to B and not belonging to A. For arbitrary subsets A and of Ω their symmetric difference, $A \Delta B$, is the set $(A \cup B) - (A \cap B)$. For any collection $(A_i, i \in I)$ of subsets of Ω, the union of all its terms is denoted $\bigcup_{i \in I} A_i$; similarly, for the intersection one writes $\bigcap_{i \in I} A_i$.

Let $(A_n, n \geq 1)$ be a countable sequence in $\mathcal{F}(\Omega)$. It is said to be increasing if $A_{n+1} \supset A_n$, $n \geq 1$.



Increasing sequence $(A_n, n\geq 1)$.

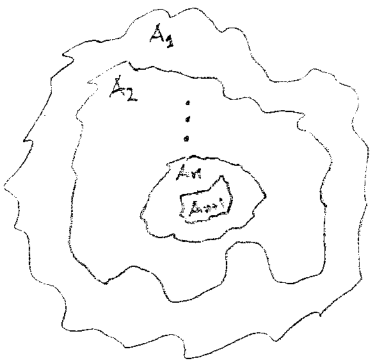
The union

$$A = \bigcup_{n\geq 1} A_n$$

is called the <u>increasing limit</u> of $(A_n, n\geq 1)$ and we write

$$A = \lim\uparrow A_n \quad \text{or} \quad A_n \uparrow A.$$

Similarly, for any decreasing sequence $(A_n, n\geq 1)$ in $\mathcal{F}(\Omega)$ i.e if



Decreasing sequence $(A_n, n\geq 1)$

$$A_{n+1} \subset A_n, \; n\geq 1$$

$$A = \bigcap_{n\geq 1} A_n$$

is called the <u>decreasing limit</u> of $(A_n, n\geq 1)$ and we write

$$A = \lim\downarrow A_n \quad \text{or} \quad A_n \downarrow A.$$

Let $(A_n, n\geq 1)$ be an arbitrary countable sequence in $\mathcal{F}(\Omega)$. Then the sequence $(B_n, n\geq 1)$ defined by

$$B_n = \bigcup_{m\geq n} A_m$$

has the property

$$B_1 \supset B_2 \supset \cdots \supset B_{n+1} \supset B_n$$

i.e. B_n is decreasing. It's decreasing limit is denoted by

$$\limsup A_n = \bigcap_{n\geq 1} B_n = \bigcap_{n\geq 1} \bigcup_{m\geq n} A_m$$

Also the sequence $(C_n, n\geq 1)$ defined by

$$C_n = \bigcap_{m\geq n} A_m$$

has the property

$$C_1 \subset C_2 \subset \cdots \subset C_n$$

i.e. C_n is increasing. It's increasing limit is denoted by

$$\liminf A_n = \bigcup_{n\geq 1} C_n = \bigcup_{n\geq 1} \bigcap_{m\geq n} A_m$$

It is easy to see that in general

$$\liminf A_n \subseteq \limsup A_n$$

When

$$\limsup A_n = \liminf A_n = A,$$

one says that the sequence $(A_n, n \geq 1)$ admits a limit denoted

$$\lim A_n = A.$$

The notions of lim sup and lim inf have an intuitive meaning in terms of outcomes and events as the following two results will show

Result 1:

$\omega \in \limsup A_n \Leftrightarrow$ for all $n_0 \geq 1$ there exists $n \geq n_0$ such that $\omega \in A_n$

Result 2:

$\omega \in \liminf A_n \Leftrightarrow \omega \in A_n$ for all but a finite number of n's.

If $\omega \in \limsup A_n$ then we say that ω occurs infinitely often i.e $\{\omega \mid A_n \text{ i.o.}\}$.

- **π-system**

Let $\mathcal{C} \subset \mathcal{F}(\Omega)$, then $\mathcal{C}$ is a π-system iff the intersection of any finite collection of elements in $\mathcal{C}$ is an element of $\mathcal{C}$.

Semi-algebra

Let $\mathcal{C} \subset \mathcal{F}(\Omega)$ then $\mathcal{C}$ is a semialgebra iff:

(a) $\Omega, \phi \in \mathcal{C}$

(b) For any $A_1, A_2, \ldots, A_n$ finite, $\bigcap_{i=1}^{n} A_i \in \mathcal{C}$

(c) If $A \in \mathcal{C}$ then A^c is the union of a finite collection of pairwise disjoint sets of $\mathcal{C}$

d-system

Let $\mathcal{C} \subset \mathcal{F}(\Omega)$ then $\mathcal{C}$ is d-system iff:

(a) $\Omega, \phi \in \mathcal{C}$

(b) For any finite collection of elements in $\mathcal{C}$ their strict difference is an element in $\mathcal{C}$.

(c) For any increasing limit $\{A_n\}$ of elements in $\mathcal{C}$ then its limit is an element in $\mathcal{C}$.

Algebra

Let $\mathcal{C} \subset \mathcal{F}(\Omega)$ then $\mathcal{C}$ is an algebra iff

(a) $\Omega, \phi \in \mathcal{E}$

(b) If $A \in \mathcal{E}$ then also $A^c \in \mathcal{E}$

(c) For any finite collection of elements $A_1, A_2, \ldots, A_n$ in $\mathcal{E}$ then their intersection

$$\bigcap_{i=1}^{n} A_i \in \mathcal{E}.$$

<u>σ - algebra</u>

Let $\mathcal{E} \subset \mathcal{F}(\Omega)$ is called a σ-algebra, iff it is an algebra and moreover for any countable collection of elements in $\mathcal{E}$ their intersection is also an element in $\mathcal{E}$ i.e.

For any $A_1, A_2, \ldots, A_n, \cdots \in \mathcal{E}$ then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{E}$.

<u>Result 3.</u>

If $\mathcal{E} \subset \mathcal{F}(\Omega)$ is a d-system and a "π-system, then it is a σ-algebra.

<u>Result 4.</u>

If $\mathcal{E}_i, i \in I$ is an arbitrary collection of σ-algebras then their intersection

$$\bigcap_{i=1} \mathcal{E}_i$$

is a σ-algebra.

<u>Result 5</u>

In the definition of the σ-algebra show that it is equivalent to say" iff it is an algebra and moreover for any countable collection of elements in $\mathcal{E}$ their union is also an element in $\mathcal{E}$ i.e.

For any $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{E}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{E}$.

▲

Let $\mathcal{E} \subset \mathcal{F}(\Omega)$ and consider the collection of all σ-algebras defined on Ω and containing $\mathcal{E}$. By Result 4 the intersection of all these σ-algebras is a σ-algebra clearly containing $\mathcal{E}$, and obviously is the smallest σ-algebra with this property. We denote such a σ-algebra by $\sigma(\mathcal{E})$ and we call it the σ-algebra generated by $\mathcal{E}$.

<u>Result 6.</u>

If $\mathcal{E}_1 \subset \mathcal{E}_2 \subset \mathcal{F}(\Omega)$ then $\sigma(\mathcal{E}_1) \subset \sigma(\mathcal{E}_2)$

▲

Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two σ-algebras defined on Ω. It is easy to see that in general $\mathcal{F}_1 \cup \mathcal{F}_2$ is not a σ-algebra.

We denote by $\mathcal{F}_1 \vee \mathcal{F}_2 = \sigma(\mathcal{F}_1 \cup \mathcal{F}_2)$ the σ-algebra generated by $\mathcal{F}_1 \cup \mathcal{F}_2$. Let $\mathcal{C}$ be

$$\mathcal{C} = \{ S \subset \Omega : S = A \cap B \text{ with } A \in \mathcal{F}_1 \text{ and } B \in \mathcal{F}_2 \}.$$

It is easy to see that

$$\mathcal{F}_1 \cup \mathcal{F}_2 \subset \mathcal{C} \subset \mathcal{F}_1 \vee \mathcal{F}_2 .$$

Now by Result 6

$$\sigma(\mathcal{F}_1 \cup \mathcal{F}_2) \subset \sigma(\mathcal{C}) \subset \sigma(\mathcal{F}_1 \vee \mathcal{F}_2) = \mathcal{F}_1 \vee \mathcal{F}_2$$

and therefore

$$\sigma(\mathcal{C}) = \mathcal{F}_1 \vee \mathcal{F}_2$$

Therefore $\mathcal{F}_1 \vee \mathcal{F}_2$ is generated by intersections of events of $\mathcal{F}_1$ with events of $\mathcal{F}_2$.

We continue with one of the fundamental tools of probability theory: the indispensable monotone class theorem.

In general a σ-algebra $\mathcal{F}$ will be defined as the smallest σ-field generated by a class $\mathcal{C}$ of elementary events. Sometimes one has to show that a certain property is true for all elements $A \in \mathcal{F} = \sigma(\mathcal{C})$, and this might be impossible in view of the definition of $\mathcal{F}$ as $\sigma(\mathcal{C})$. However if $\mathcal{C}$ is a π-system, and if one can show that the property in question is verified for all A in $d(\mathcal{C})$, then it is also verified for all A in $\sigma(\mathcal{C})$.

Theorem 2.1. (Monotone Class Theorem).

Let $\mathcal{C}$ be a π-system defined on Ω. Then $d(\mathcal{C}) = \sigma(\mathcal{C})$. In other words the d-system generated by a π-system is a σ-algebra.

Proof: It is easy to see that $d(\mathcal{C}) \subset \sigma(\mathcal{C})$, it suffices to show that for any element $A \in d(\mathcal{C})$ also $A \in \sigma(\mathcal{C})$.

If we show that $d(\mathcal{C})$ is a σ-algebra then immediately we have that $d(\mathcal{C}) = \sigma(\mathcal{C})$. By result 3 it is enough to show that $d(\mathcal{C})$ is a π-system. Define by

$$\mathcal{D}_1 = \{ B \in d(\mathcal{C}) \mid B \cap A \in d(\mathcal{C}) \text{ for all } A \in \mathcal{C} \}$$

Then $\mathcal{D}_1$ is a d-system, and $\mathcal{D}_1 \supset \mathcal{C}$.

Verify that if $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{F}(\Omega)$ then $d(\mathcal{C}_1) \subset d(\mathcal{C}_2)$

Thus $\mathcal{D}_1 \supset d(\mathcal{E})$. By the definition of $\mathcal{D}_1$, $\mathcal{D}_1 \subset d(\mathcal{E})$, so that $\mathcal{D}_1 = d(\mathcal{E})$. Define now

$$\mathcal{D}_2 = \{B \in d(\mathcal{E}) \mid B \cap A \in d(\mathcal{E}) \text{ for all } A \in d(\mathcal{E})\}$$

Then $\mathcal{D}_2$ is a d-system. Also, if $A \in \mathcal{E}$, then $B \cap A \in d(\mathcal{E})$ for all $B \in \mathcal{D}_1 = d(\mathcal{E})$, and therefore $\mathcal{E} \subset \mathcal{D}_2$. Therefore $d(\mathcal{E}) \subset d(\mathcal{D}_2) = \mathcal{D}_2$. Also by definition of $\mathcal{D}_2$, $\mathcal{D}_2 \subset d(\mathcal{E})$, so that finally $\mathcal{D}_2 = d(\mathcal{E})$, which expresses that $d(\mathcal{E})$ is a π-system.

Borel σ-algebras.

Let (X, T) be a topological space, the topology of which is defined by the collection of its open sets, let say $\mathcal{E}$. By definition, the σ-algebra generated by $\mathcal{E}$ is called the Borel σ-algebra of X and it is denoted by $\mathcal{B}(X)$. If $X = \mathbb{R}^n$ then $\mathcal{B}(\mathbb{R}^n)$ sometimes in denoted by $\mathcal{B}^n$ and if $X = \mathbb{R}$ by $\mathcal{B}^1$.

For any subset of a topological space X we define

$$\mathcal{B}(A) = \{C \mid C = A \cap B, B \in \mathcal{B}(X)\}$$

Equivalently $\mathcal{B}(\mathbb{R})$ is the smallest σ-algebra containing all the closed intervals $[a,b] \in \mathbb{R}$.

The σ-algebra $\mathcal{B}^1$ is the most important of all σ-algebras. Every subset of $\mathbb{R}$ which you meet in everyday use is an element of $\mathcal{B}^1$.

Stochastic Finance and σ-algebras.

In stochastic Finance, we need a way to mathematically model the information on which our future decisions can be based. For example, we are told the stock price at some time in the past and nothing else. This information is in many cases enough to narrow down the possibilities. Let us try to clarify how such a role is played by the σ-algebras in the following important example which will be used for many purposes in what follows.

Example 2.1: Let that the stock of corporation CIL is being studied closely by its share holders for three days. The share holders assign number one at a particular day if the stock has gone up. Otherwise they assign the number 0 for the same day. Let Ω be the sample space of the just described random experiment and ω its element. Then Ω is the following

$$\Omega = \{(1,1,1),(1,1,0),(1,0,1),(0,1,1),(1,0,0),(0,1,0),(0,0,1),(0,0,0)\}.$$

Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then $\mathcal{F}_0$ is a σ-algebra since

a) $\Omega \in \mathcal{F}_0$

b) The complement of $\emptyset$ is $\Omega \in \mathcal{F}_0$ and the complement of Ω is $\emptyset \in \mathcal{F}_0$.

c) Apparently
$$\Omega \cup \emptyset = \Omega \in \mathcal{F}_0.$$

The σ-algebra $\mathcal{F}_0$ is contains no information for the true outcome ω for any day since the set $\emptyset$ and the whole space Ω are always resolved, even without any information; the true ω does not belong to $\emptyset$ and does belong to Ω.

Let $\mathcal{F}_1 = \{\emptyset, \Omega, [(1,1,1),(1,1,0),(1,0,1),(1,0,0)], [(0,1,1),(0,1,0),(0,0,1),(0,0,0)]\}$.

Then $\mathcal{F}_1$ is a σ-algebra since

a) $\Omega \in \mathcal{F}_1$.

b) The complement of $\emptyset$ is $\Omega \in \mathcal{F}_1$ and the complement of Ω is $\emptyset \in \mathcal{F}_1$. The same applies for the remaining elements of $\mathcal{F}_1$.

c) Any union of any number of elements of $\mathcal{F}_1$ belongs to $\mathcal{F}_1$.

Now consider the elements of $\mathcal{F}_1$

$$A_1 = \{(1,1,1),(1,1,0),(1,0,1),(1,0,0)\}$$

and

$$A_0 = \{(0,1,1),(0,1,0),(0,0,1),(0,0,0)\}$$

We say that the σ-algebra $\mathcal{F}_1$ is containing the information learned by observing the stock in the first day. More precisely, if instead of being told the increase or decrease of the price of the stock the first day, we are told, for each set in $\mathcal{F}_1$, whether or not the true ω belongs to the set, we know the outcome of the first day. Specifically, if $\omega \in A_1$ we know that in the first day we had an increase in the stock price. In fact we know nothing more than that.

Now consider the sets

$$A_{11} = [(1,1,1),(1,1,0)] \quad , \quad A_{10} = [(1,0,1),(1,0,0)],$$

$$A_{01} = [(0,1,1),(0,1,0)] \quad ; \quad A_{00} = [(0,0,1),(0,0,0)]$$

and consider the 6-algebra

$$\mathcal{F}_2 = \left\{ \begin{array}{l} \emptyset, \Omega, A_1, A_2, A_{11}, A_{10}, A_{01}, A_{00}, A_{11}^c, A_{10}^c, A_{01}^c, A_{00}^c \\ A_{11} \cup A_{01}, A_{11} \cup A_{00}, A_{10} \cup A_{01}, A_{10} \cup A_{00} \end{array} \right\}$$

then $\mathcal{F}_2$ is a 6-algebra. We say that the 6-algebra is containing the information learned by observing the stock the first two days. More specifically, if we are told for each set in $\mathcal{F}_2$, whether or not the true w belongs to the set, we know the outcome of the first two days.

It is important to note that $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2$.

Now consider the 6-algebra $\mathcal{F}_3$ to be the set of all subsets of Ω. There are 256 subsets of Ω and if we are told for each set in $\mathcal{F}_3$, whether or not the true w belongs to the set, we know the outcomes of the three days.

We observe again that $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$. This collection of 6-algebras $\mathcal{F}_0$, $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ is an example of filtration. The now provide the continues-time formulation of this case in the following definition

Definition 2.1 Let Ω be a nonempty set. Let T be a fixed positive number, and assume that for each $t \in [0,T]$ there is a 6-algebra $\mathcal{F}(t)$. Assume further that if $s \leq t$, then every set in $\mathcal{F}(s)$ is also in $\mathcal{F}(t)$. Then we call the collection of 6-algebras $\mathcal{F}(t)$, $0 \leq t \leq T$, a filtration.

A 6-algebra $\mathcal{F}(s)$ tels us all the information for the true value of w from 0 to time s. The way $\mathcal{F}_2$ was constructed from the four sets A_{11}, A_{10}, A_{01}, A_{00} suggests that the 6-algebras in a filtration can be built by taking unions and complements of certain fundamental sets. If this were the case, it would be enough to work with these atoms (indivisible sets in the 6-algebra) and not consider all the other sets. In uncountable sample spaces, however, there are sets that cannot be constructed as countable unions of atoms and uncountable unions are forbidden because we cannot add probabilities of such unions.

2.3. Measure Spaces.

Let Ω be a set and $\mathcal{F}$ is a σ algebra on Ω.

Definition 2.2.

The pair $(\Omega, \mathcal{F})$ is called a measurable space. An element of $\mathcal{F}$ is called a $\mathcal{F}$-measurable subset of Ω.

Let $(\Omega, \mathcal{A})$ be a pair where Ω is a set and $\mathcal{A}$ is an algebra. Let also μ a non-negative set function

$$\mu : \mathcal{F} \to [0, \infty],$$

then μ is called additive if $\mu(\phi) = 0$ and for $A, B \in \Omega$ then

$$A \cap B = \phi \rightsquigarrow \mu(A \cap B) = \mu(A) + \mu(B).$$

The set function μ is called countably additive or σ-additive if $\mu(\phi) = 0$ and for any sequence $\{A_n\}_{n=1}^{\infty}$ of disjoint sets in $\mathcal{F}$ with union

$$A = \bigcup_n A_n \in \mathcal{F}$$

then

$$\mu(A) = \mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n).$$

Note that we need the assumption that $A \in \mathcal{F}$ since we assumed that $\mathcal{A}$ is an algebra and not the strouger assumption that $\mathcal{A}$ is a σ-algebra.

Definition 2.3

Let $(\Omega, \mathcal{F})$ be a measurable space, i.e. $\mathcal{F}$ is a σ-algebra on S.

A map

$$\mu : \mathcal{F} \to [0, \infty]$$

is called a measure on $(\Omega, \mathcal{F})$ if μ is countably additive.

The tripple $(\Omega, \mathcal{F}, \mu)$ is then called a measure space.

Definition 2.4.

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Then μ or indeed the measure space

$^{+}(\Omega,\mathcal{F},\mu)$ is called finite if
$$\mu(\Omega)<\infty$$
and σ-finite if there is a sequence $\{A_n\}_{n=1}^{\infty}$ of elements of $\mathcal{F}$ such that
$$\mu(A_n)<\infty \ \forall n\in\mathbb{N} \text{ and } \bigcup_n A_n=\Omega.$$
▲

Let $(\Omega,\mathcal{F},\mathbb{P})$ be a measure space, then $\mathbb{P}$ is called a probability measure if
$$\mathbb{P}(\Omega)=1.$$
and then $(\Omega,\mathcal{F},\mathbb{P})$ is then called a probability triple.

An element $A\in\mathcal{F}$ is called $\mathbb{P}$-null if $\mathbb{P}(A)=0$.

A statement S about events ω of Ω is said to hold almost everywhere (a.e) if
$$A:=\{\omega : S(\omega) \text{ is false}\}\in\mathcal{F} \text{ and } \mathbb{P}(A)=0.$$

Let $(\Omega,\mathcal{F},\mu)$ be a measure space. Then it is not difficult to prove the following

(a) For any $A,B\in\mathcal{F}$ then $\mu(A\cup B)\leq\mu(A)+\mu(B)$

(b) For every $A_1,A_2,\dots,A_n\in\mathcal{F}$ we have
$$\mu\Big(\bigcup_{i\leq n}A_i\Big)\leq\sum_{i\leq n}\mu(A_i)$$
(c) For any $A,B\in\mathcal{F}$ if $\mu(\Omega)<\infty$ then
$$\mu(A\cup B)=\mu(A)+\mu(B)-\mu(A\cap B)$$
(d) For $A_1,A_2,\dots,A_n\in\mathcal{F}$
$$\mu\Big(\bigcup_{i\leq n}A_i\Big)=\sum_{i\leq n}\mu(A_i)-\sum\sum_{i<j\leq n}\mu(A_i\cap A_j)$$
$$+\sum\sum\sum_{i\leq j\leq k\leq n}\mu(A_i\cap A_j\cap A_k)$$
$$-\dots+(-1)^n\mu(A_1\cap A_2\cap\dots\cap A_n).$$

Monotone convergence properties of measures.

Proposition 2.1. If A_n is an increasing sequence in $\mathcal{F}$ i.e. $A_1\subset A_2\subset\dots\subset A_n$ and in addition $A_n\uparrow A$ then $\mu(A_n)\uparrow\mu(A)$.

Proof: Refresh the fact that $A=\bigcup_n A_n$.

Write $B_1=A_1$ and $B_n=A_n-A_{n-1}$ for $n\geq 2$.

Then the sets B_n, $n\in\mathbb{N}$ are disjoint, and

$$\mu(A_n) = \mu(B_1 \cup B_2 \cup \ldots \cup B_n) = \sum_{k=1}^{n} \mu(B_k)$$

Now

$$\mu(A_n) \uparrow \sum_{k=1} \mu(B_k) = \mu(A)$$

▲

Proposition 2.2. If $C_n \in \mathcal{F}$ and C_n is a decreasing sequence in $\mathcal{F}$ with $\mu(C_k) < \infty$ for some K and $C_n \downarrow C$ then $\mu(C_k) \downarrow \mu(C)$.

Now we are in a position to prove the following crucial lemma for measures.

Lemma 2.1

Let S be a set. let I be a π-system on S, and let $\mathcal{F} := \sigma(I)$. Suppose that μ_1 and μ_2 are measures on the measurable space $(S, \mathcal{F})$ such that

$$\mu_1(S) = \mu_2(S) < \infty$$

and that in addition

$$\mu_1 = \mu_2 \text{ on } I.$$

Then we have

$$\mu_1 = \mu_2 \text{ on } \mathcal{F}$$

Proof: let the set of elements of $\mathcal{F}$ for which $\mu_1 = \mu_2$ be

$$D = \{A \in \mathcal{F} : \mu_1(A) = \mu_2(A)\}$$

We will prove that D is a d-system Firstly we know that $S \in D$ since by hypothesis we have $\mu_1(S) = \mu_2(S)$

Now if $A, B \in D$ then

$$\mu_1(B - A) = \mu_1(B) - \mu_1(A) = \mu_2(B) - \mu_2(A) = \mu_2(B - A)$$

hence

$$B - A \in D$$

Finally, if $A_n \in D$ is an increasing sequence with $A_n \uparrow A$ then by Proposition 2.1 we have

$$\mu_1(A) = \lim \mu_1(A_n) \uparrow = \lim \mu_2(A_n) \uparrow = \mu_2(A)$$

hence

$$A \in D$$

and D is a d-system. We will now use a lemma known also as Dynkin's lemma which will be given without proof but certainly is an important one:

Lemma 2.2. Dynkin's Lemma

Any d-system which contains a π-system contains the σ-algebra generated by that π-system.

From its construction $D \supseteq I$ thus by Dynkin's Lemma we have that

$$D \supseteq \sigma(I) = \mathcal{F}$$

hence

$$\mu_1 = \mu_2 \text{ on } \mathcal{F}$$

▲

There is an immediate corollary of the just proven Lemma 2.1 which is more often applied that the actual Lemma.

Corollary 2.1

If two probability measures agree on a π-system, then they agree on the σ-algebra generated by that π-system.

We know provide without proof a well known theorem by Caratheodory.

Theorem 2.2 Caratheodory's extension Theorem.

Let S be a set and let $\mathcal{A}$ be an algebra on S and let

$$\mathcal{F} := \sigma(\mathcal{A})$$

If μ_0 is a countably additive map

$$\mu_0 : \mathcal{A} \rightarrow [0, \infty],$$

then there exists a measure μ on $(S, \mathcal{F})$ such that

$$\mu = \mu_0 \text{ on } \mathcal{A}$$

If $\mu_0(S) < \infty$ then this extension is unique.

Probability measures on uncountable sample spaces.

We now provide examples of construction of probability measures on uncountable sample spaces. However, before that it will be helpfull to rewrite Propositions 2.1 and 2.2 in a more helpfull way for probability measures.

Proposition 2.3.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $A_1, A_2, A_3, \ldots$

be a sequence of sets in $\mathcal{F}$

(i) If $A_1 \subset A_2 \subset A_3 \subset \cdots$, then

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$$

(ii) If $A_1 \supset A_2 \supset A_3 \supset \cdots$, then

$$\mathbb{P}\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Example 2.2

Consider the random experiment for choosing a number at random from the unit interval $[0,1]$. We want to construct a mathematical model so that the probability is distributed uniformly over the interval.

Let us define by

$$\mathbb{P}[a,b] = \text{prob}\{\text{the number chosen is between } a \text{ and } b\}.$$

We define the probability of any close interval $[a,b]$ to be

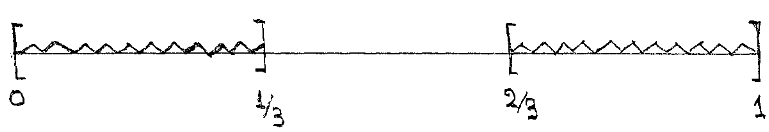
$$\mathbb{P}[a,b] = b-a \;, \; 0 \le a \le b \le 1.$$

This particular measure is called Lebesgue measure and is denoted by $\mathcal{L}$. The Lebesgue probability measure is the same for the open interval (a,b) as well since the probability of a single point is $\mathbb{P}[a,a] = a-a = 0$.

Obviously the σ-algebra appropriate for this random experiment is the Borel σ-algebra on $[0,1]$. The sets in this σ-algebra are called Borel sets. This σ-algebra is constructed by beginning with the closed intervals and adding all other sets in order to have a σ-algebra. These so called Borel set could be complicated and sometimes difficult to intuitivaly think about them. Such a Borel set is the well known Cantor set which we will now describe.

Consider the interval $[0,1]$, (figure (2.3)).

Figure 2.3



Remove the middle third i.e. the open interval $(1/3, 2/3)$. Now define the set C_1 to be

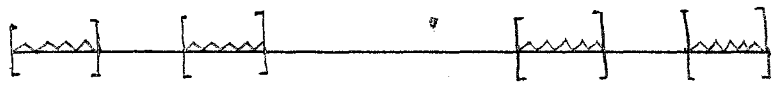
$$C_1 = [0, 1/3] \cup [2/3, 1],$$

then

$$\mathbb{P}[C_1] = \mathbb{P}\{[0, 1/3] \cup [2/3, 1]\} = \mathbb{P}[0, 1/3] + \mathbb{P}[2/3, 1]$$
$$= (\text{Lebesgue measure}) = 2/3.$$

Consider again the interval $[0,1]$ and remove from the interval $[0, 1/3]$ the middle third and from the interval $[2/3, 1]$ again the middle third (figure 2.4).

Figure 2.4



Now define the set C_2 to be

$$C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup \left[\frac{8}{9}, 1\right]$$

then

$$\mathbb{P}[C_2] = \mathbb{P}[0, 1/9] + \mathbb{P}[2/9, 1/3] + \mathbb{P}[2/3, 7/9] + \mathbb{P}\left[\frac{8}{9}, 1\right]$$

$$= \frac{4}{9} = \left(\frac{2}{3}\right)^2.$$

We continue this process and at stage k it is easy to see that the set C_k will have 2^k pieces and

$$\mathbb{P}[C_k] = \left(\frac{2}{3}\right)^k \quad \text{for } k = 1, 2, 3, \ldots$$

Now we define the Cantor set C to be

$$C = \bigcap_{k=1}^{\infty} C_k$$

From Proposition 2.3 we have that

$$\mathbb{P}(C) = \lim_{k \to \infty} \mathbb{P}(C_k) = \lim_{k \to \infty} \left(\frac{2}{3}\right)^k = 0.$$

Hence, although that the Cantor set certainly contains the points $0, 1/3, 2/3, 1, 1/9, 2/9, 7/9, 8/9, \cdots$ which are the end points of the intervals appearing at the successive stages and they are never removed it has probability zero. In fact

• In fact the Cantor set does not have countably many points as the above remark implied but it could be proved that it has uncountably infinite many points.

The observation that a set that has uncountably infinite many points has a probability zero highlights a paradox in uncountable probability space in the sence that is counterintuitive. We would like to say that something that has probability zero cannot happen. Mathematicians have created a terminology that equivocates. We say that an event is <u>almost sure</u>, meaning it has probability one, even though it may not include every possible outcome. The outcome or set of outcomes not included, taken all together, has probability zero.

<u>Completeness and product spaces.</u>

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Any event A which has zero probability, that is $\mathbb{P}(A)=0$, is called <u>null</u>. It may seem reasonable to suppose that any subset B of a null set A will itself be null, but this may be without meaning since B may not be an event, and thus $\mathbb{P}(B)$ may not be defined.

<u>Definition 2.4</u>. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called <u>complete</u> if all subsets of null sets are events.

Any incomplete space can be completed. Let $\mathcal{N}$ be the collection of all subsets of null sets in $\mathcal{F}$ and let $\mathcal{G} = \sigma(\mathcal{F} \cup \mathcal{N})$ be the smallest σ-algebra which contains all sets in $\mathcal{F}$ and $\mathcal{N}$. It can be shown that the domain of $\mathbb{P}$ may be extended in an obvious way from $\mathcal{F}$ to $\mathcal{G}$; Then $(\Omega, \mathcal{G}, \mathbb{P})$ is called the <u>completion</u> of $(\Omega, \mathcal{F}, \mathbb{P})$.

The probability spaces discussed so far in the present notes are usually assuming that the sample space Ω contains all the possible outcomes of one random experiment. However there are cases which they occur naturally where we need to combine the outcomes of several independent experiments into one space.

Suppos two experiments have associated probability spaces $(\Omega, \mathcal{F}_1, \mathbb{P}_1)$

and $(\Omega, \mathcal{F}_2, \mathbb{P}_2)$ respectively. The sample space of the pair of experiments, considered jointly, is the collection

$$\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

of ordered pairs.

Now suppose that Ω_1 and Ω_2 are finite, and their σ-algebras contain all their subsets. Then consider the family of all sets

$$\mathcal{F}_1 \times \mathcal{F}_2 = \{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\} = \mathcal{G}$$

then $\mathcal{F}_1 \times \mathcal{F}_2$ is a σ-algebra.

However if Ω_1 and Ω_2 are not finite then $\mathcal{F}_1 \times \mathcal{F}_2$ is not in general a σ-algebra. In this case $\mathcal{F}_1 \times \mathcal{F}_2$ is replaced by $\sigma(\mathcal{F}_1 \times \mathcal{F}_2)$ which is a unique smallest σ-algebra which contains $\mathcal{F}_1 \times \mathcal{F}_2$.

The are now many probability measures that can be applied to $(\Omega_1 \times \Omega_2, \mathcal{G})$. As an important example we provide the probability measure

$$\mathbb{P}_{12} : \sigma(\mathcal{F}_1 \times \mathcal{F}_2) \longrightarrow [0,1]$$

given by

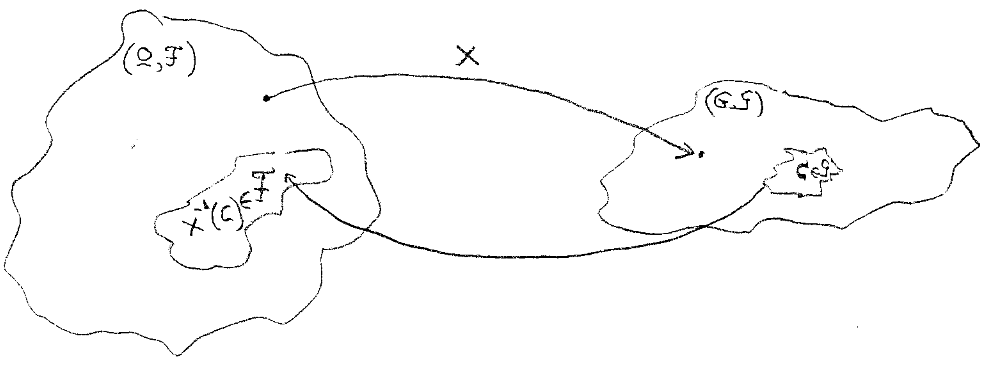
$$\mathbb{P}_{12}(A_1 \times A_2) = \mathbb{P}_1(A_1)\, \mathbb{P}_2(A_2) \quad \text{for } A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2$$

The probability measure $\mathbb{P}_{12}$ is sometimes called the <u>product measure</u> since its defining equation assumed that two experiments are independent. The probability space $(\Omega_1 \times \Omega_2, \mathcal{G}, \mathbb{P}_{12})$ is called the <u>product space</u> of $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$.

2.4 Random variables

Let $(\Omega, \mathcal{F})$ and $(G, \mathcal{G})$ be two measurable spaces, and let $X : \Omega \to G$

Figure 2.5



be a mapping such that

$$X^{-1}(C) \in \mathcal{F} \quad \text{for all } C \in \mathcal{G}.$$

Then X is called a measurable mapping from $(\Omega, \mathcal{F})$ to $(G, \mathcal{G})$.

We now provide without proof the following interesting theorem:

Theorem 2.3.: Let Ω be some set on which is defined a family $(\mathcal{F}_i, i\in I)$ of σ-fields. Let $\mathcal{F} = \sigma(\mathcal{F}_i, i\in I)$. Then for any event A of $\mathcal{F}$, there exists a countable subset J of I such that A is $\sigma(\mathcal{F}_i, i\in J)$-measurable.

There exists an a logical consequence of the definition of measurable mapping which we provide in the following definition.

Definition 2.5. Let $(\Omega, \mathcal{F})$ be a measurable function and let a function $f: \Omega \to \mathbb{R}$ Let all $\mathcal{B}(\mathbb{R})$ the Borel σ-algebra and for any $A \subseteq \mathbb{R}$ let

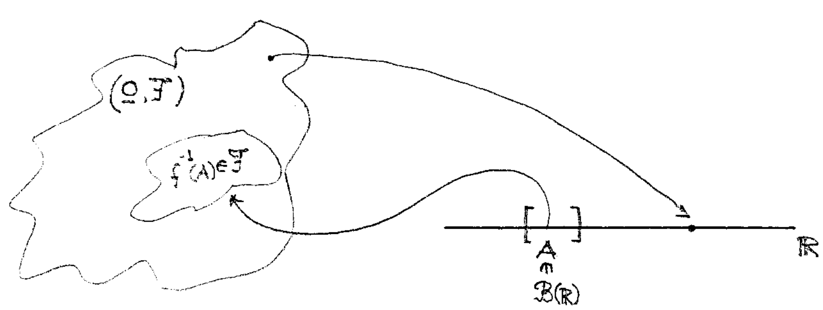
$$f^{-1}(A) := \{\omega \in \Omega : f(\omega) \in A\}.$$

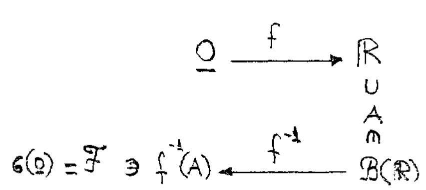
Then we say that f is $\mathcal{F}$-measurable iff:

For every $A \in \mathcal{B}(\mathbb{R})$ then $f^{-1}(A) \in \mathcal{F}$.

Picture definition 2.5 in figure 2.6

Figure 2.6





We write $(m\mathcal{F})_\Omega$ for the class of $\mathcal{F}$-measurable functions on Ω, and $(m\mathcal{F})^+_\Omega$ for the class of non-negative elements in $m\mathcal{F}$. Finally we denote by $(b\mathcal{F})_\Omega$ the class of bounded $\mathcal{F}$-measurable functions on Ω.

Result 6

Let $(\Omega, \mathcal{F})$ be a measurable space then $(m\mathcal{F})_\Omega$ is an algebra over $\mathbb{R}$.

Result 7.

Let $(\Omega, \mathcal{F})$ be a measurable space, $h \in (m\mathcal{F})_\Omega$ and $f \in (m\mathcal{B})_\mathbb{R}$ then $f \circ h \in (m\mathcal{F})_\Omega$, where $f \circ h$ is the composition of the two functions.

Definition 2.6

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable is an element of $(m\mathcal{F})_\Omega$. More explicitly a random variable is a real-valued function X defined on Ω with the property that for every Borel subset of $\mathbb{R}$, the subset of Ω given by

$$\{X \in \mathcal{B}\} = \{\omega \in \Omega ; X(\omega) \in \mathcal{B}\}$$

is in the σ-algebra $\mathcal{F}$.

Example 2.3

Let the Example 2.1 and define the random variable

$Y :=$ {the number of times the stock has gone up the first two days}

Then the random variable Y is $\mathcal{F}_2$-measurable in Ω i.e. is a member of the class $(m\mathcal{F}_2)_\Omega$.

It is apparent from Table 2.1 that Y is a function from $\Omega \to \mathbb{R}$.

Table 2.1.

Possible values of r.v. Y.

Event	Y
$\omega_1 = (1,1,1)$	2
$\omega_2 = (1,1,0)$	2
$\omega_3 = (1,0,1)$	1
$\omega_4 = (0,1,1)$	1
$\omega_5 = (0,1,0)$	1
$\omega_6 = (1,0,0)$	1
$\omega_7 = (0,0,1)$	0
$\omega_8 = (0,0,0)$	0

Now let take the closed interval $A = [0,1] \in \mathcal{B}(\mathbb{R})$. Then we have that

$$Y^{-1}(A) = \{(1,0,1), (0,1,1), (1,0,0), (0,1,0), (0,0,1), (0,0,0)\}$$

It is easy to verify that

$$Y^{-1}(A) \in \mathcal{F}_2 .$$

The same is true for every $A \in \mathcal{B}(\mathbb{R})$ for the random variable Y.

We now provide the definition of a σ-algebra generated by a random variable X which we denote by $\sigma(X)$.

<u>Definition 2.7</u>. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be a random variable i.e.

$$X : \Omega \to \mathbb{R}.$$

We call the σ-algebra generated by X and denote it by $\sigma(X)$ the smallest σ-algebra in Ω which is generated by the set $X^{-1}(\mathbb{R})$.

From the above definition it is obvious that in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ any random variable X is $\mathcal{F}$-measurable in Ω since $\mathcal{F}$ is the largest σ-algebra in Ω. However, in general there also other σ-algebras smaller than $\mathcal{F}$ in Ω for which the random variable X is measurable. The smallest of these is the σ-algebra generated by X.

We will now provide a definition which is a sequel of the above definition for two random variable X and Y.

<u>Definition 2.8</u>: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X and Y be two random variables

$$X : \Omega \to \mathbb{R} \quad \text{and} \quad Y : \Omega \to \mathbb{R}.$$

We call the σ-algebra generated from X and Y and we denote it by $\sigma(X,Y)$ the smallest σ-algebra in Ω to which X, Y are measurables in Ω.

Let X a random variable with values in the set $\mathbb{N}$, i.e a discrete random variable. Define the sets

$$A_i = \{\omega : X(\omega) = i\} \; ; \; i = 1, 2, \dots$$

then the sets $A_i, i = 1, 2, \dots$ is a partition of the sample space Ω since

$$\Omega = \bigcup_{i=1}^{\infty} A_i \quad \text{and} \quad A_i \cap A_j = \emptyset \quad \text{for} \quad i \neq j$$

In this case $\sigma(X)$ consists of the sets $A_i, (i=1,2,\ldots)$ the sets ϕ and Ω and every set which is the union of some sets from the $A_i, (i=1,2,\ldots)$.

For example let the random variable Y as defined in example 2.3. Then

$$A_0 = \{\omega : Y(\omega)=0\}$$

and from table 2.1 we get that

$$A_0 = \{(0,0,1), (0,0,0)\}.$$

Similarly we get that

$$A_1 = \{(1,0,1), (0,1,1), (1,0,0), (0,1,0)\}$$

$$A_2 = \{(1,1,1), (1,1,0)\}$$

We have that

$$\Omega = A_0 \cup A_1 \cup A_2 \quad \text{and}$$

$$A_0 \cap A_1 = \phi \quad , \quad A_0 \cap A_2 = \phi \quad \text{and} \quad A_1 \cap A_2 = \phi$$

Thus $\sigma(Y)$ is

$$\{\phi, \Omega, \{(0,0,1),(0,0,0)\}, \{(1,0,1),(0,1,1),(1,0,0),(0,1,0)\},$$
$$\{(1,1,1),(1,1,0)\}\}$$

and all possible unions of the above. However this is the σ-algebra $\mathcal{F}_2$ of example 2.1. Thus we have $\sigma(Y)=\mathcal{F}_2$ as we have seen in example 2.2.

Let now that we have two random variables X taking values $x_1, x_2, \ldots$ and Y taking values $y_1, y_2, \ldots$ i.e. two discrete random variables. Define the sets

$$A_{ij} = \{\omega : X(\omega)=i, Y(\omega)=j\}, \quad i,j=1,2,\ldots$$

then

$$\Omega = \bigcup_{i,j=1}^{\infty} A_{ij} \quad \text{and} \quad A_{ij} \cap A_{mn} = \phi \quad \text{if } (i,j) \neq (m,n)$$

In this case the σ-algebra $\sigma(X,Y)$ contains all the sets A_{ij} $(i,j=1,2,\ldots,k)$, the ϕ, Ω and all the possible unions of the sets A_{ij} $(i,j=1,2,\ldots,k)$.

<u>Example 2.4</u>:

In example 2.3. we define in addition the following random variable

$$X = \{\text{the number of times the stock has gone up the first two days}\}$$

In table 2.2 we provide the different values of X, Y in the sample space $\underline{\Omega}$

Table 2.3

Sample space	Y	X
$\omega_0 = (1,1,1)$	2	0
$\omega_1 = (1,1,0)$	2	0
$\omega_2 = (1,0,1)$	1	1
$\omega_3 = (0,1,1)$	1	1
$\omega_4 = (1,0,0)$	1	1
$\omega_5 = (0,1,0)$	1	1
$\omega_6 = (0,0,1)$	0	2
$\omega_7 = (0,0,0)$	0	2

We have that

$$A_{02} = \{\omega : X(\omega)=0 ,\ Y(\omega) = 2\} = \{(1,1,1), (1,1,0)\}$$

$$A_{01} = \{\omega : X(\omega)=0 ,\ Y(\omega) = 1\} = \{\phi\}$$

$$A_{00} = \{\omega : X(\omega)=0 ,\ Y(\omega) = 0\} = \{\phi\}$$

$$A_{10} = \{\omega : X(\omega)=1 ,\ Y(\omega) = 0\} = \{\phi\}$$

$$A_{11} = \{\omega : X(\omega)=1 ,\ Y(\omega) = 1\} =$$

$$= \{(1,0,1), (0,1,1), (1,0,0), (0,1,0)\}$$

$$A_{12} = \{\omega : X(\omega)=1 ,\ Y(\omega) = 2\} = \{\phi\}$$

$$A_{20} = \{\omega : X(\omega)=2 ,\ Y(\omega) = 0\} = \{(0,0,1), (0,0,0)\}.$$

$$A_{21} = \{\omega : X(\omega)=2 ,\ Y(\omega) = 1\} = \{\phi\}$$

$$A_{22} = \{\omega : X(\omega)=2 ,\ Y(\omega) = 2\} = \{\phi\}.$$

It is easy to verify that

$$\underline{\Omega} = \bigcup_{i,j=1}^{2} A_{ij} \text{ with } A_{ij} \cap A_{nm} = \phi \text{ for every } (i,j) \neq (m,n).$$

Now

$$\sigma(X,Y) = \{\phi, \underline{\Omega}, \{(1,1,1), (1,1,0)\}, \{(1,0,1), (0,1,1), (1,0,0), (0,1,0)\},$$

$\{(0,0,1),(0,0,0)\}\}$

and all possible unions of these sets. Thus $\sigma(X,Y) = \mathcal{F}_2$ of example 2.1 which verifies the fact contains all the necessary information up to time 2.

Now we are in a position to understand more deeply the following.

If we have a collection $(Y_a : a \in C)$ of maps $Y_a : \Omega \to \mathbb{R}$, then

$$\Psi := \sigma(Y_a : a \in C)$$

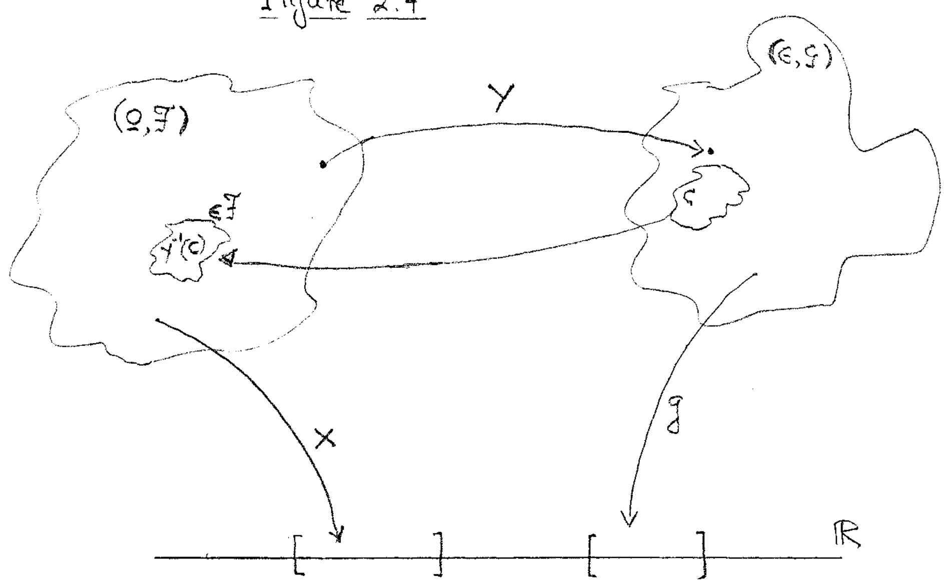
is defined to be the smallest σ-algebra Ψ such that each map Y_a $(a \in C)$ is Ψ-measurable. Clearly,

$$\sigma(Y_a : a \in C) = \sigma(\{\omega \in \Omega : Y_a(\omega) \in B\} : a \in C, B \in \mathcal{B}(\mathbb{R})).$$

We now provide the following theorem which by some is called a Characterization of Measurability.

Theorem 2.4: Let $(\Omega,\mathcal{F})$ and $(G,\mathcal{G})$ be two measurable spaces (figure 2.7), and Y a measurable mapping from $(\Omega,\mathcal{F})$ into $(G,\mathcal{G})$. Let X be a real random variable defined on $(\Omega,\mathcal{F})$ which is $\sigma(X)$-measurable. Then there exists a real random variable g defined on $(G,\mathcal{G})$ such that $X = g \circ Y$.

Figure 2.7



We now provide the following two useful definition for the applications in finance.

Definition 2.9

Let X be a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$

Let $\mathcal{G}$ be a σ-algebra of subsets of Ω. If every set in $\sigma(X)$ is also in $\mathcal{G}$ we say that X is $\mathcal{G}$-measurable.

A random variable X is $\mathcal{G}$-measurable if and only if the information in $\mathcal{G}$ is sufficient to determine the value of X. If X is $\mathcal{G}$-measurable, then $f(X)$ is also $\mathcal{G}$-measurable for any Borel-measurable function f; if the information in $\mathcal{G}$ is sufficient to determine the value of X, it will also determine the value of $f(X)$. If X and Y are $\mathcal{G}$-measurable, then $f(X,Y)$ is $\mathcal{G}$-measurable for any Borel-measurable function $f(x,y)$ of two variables. A portofolio position $\Delta(t)$ taken at time t must be $\mathcal{F}(t)$-measurable i.e. must depend only available to the investor at time t.

<u>Definition 2.10</u>. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space equipped with a filtration $\mathcal{F}(t)$ $0\leq t\leq T$. Let $X(t)$ be a collection of random variables indexed by $t\in[0,T]$. We say this collection of random variables is an adapted stochastic process, if for each t, the random variable $X(t)$ is $\mathcal{F}(t)$-measurable.

In the continuous-time models of this text, asset prices, portofolio processes and wealth processes will be adapted to a filtration that we regard as a model of the flow of public information.

For a random variable it is important to know, if it is discrete, how much probability mass is being assigned to a particular value. In other words it is very useful to have a measure of how probable is a particular valued compared with others. When the random variable is continuous we need a measure to spread a unit of probability mass is spread over the various sections of the real line. We call such a probability measure the distribution of the random variable, which is a measure on subsets of $\mathbb{R}$ rather than subsets of Ω.

<u>Definition 2.11</u> Let X be a random variable on a probability space $(\Omega,\mathcal{F},\mathbb{P})$. The distribution measure of X is the probability measure μ_X that assigns to each Borel subset B of $\mathbb{R}$ the mass

$$\mu_X(B) = \mathbb{P}\{X\in B\}$$

Obviously the above definition applies for both discrete and continuous random variables.

Random variables have distributions but there is not an unbreakable one to one mapping between them. Two different random variables can have the same distribution. In fact many random variables can have the same distribution. But for our purposes it is more important to have in mind that a single random variable can have two different distributions. Consider the following example.

Example 2.5.

Let $\mathbb{P}$ be the uniform measure on $[0,1]$ described in example 2.2, i.e the Lebesgue measure. Define the following random variables

$$X(\omega)=\omega \quad \text{and} \quad Y(\omega)=1-\omega, \text{ for all } \omega\in[0,1].$$

Then the distribution measure of X, is uniform i.e

$$\mu_X[a,b]=\mathbb{P}\{\omega;\, a\leq X(\omega)\leq b\}=\mathbb{P}[a,b]=b-a \quad ,0\leq a\leq b\leq 1$$

Although the random variable Y takes different values than X has the same distribution as X.

$$\mu_Y[a,b]=\mathbb{P}\{\omega;\, a\leq Y(\omega)\leq b\}=\mathbb{P}\{\omega;\, a\leq 1-\omega\leq b\}=\mathbb{P}[1-b,1-a]$$

$$=(1-a)-(1-b)=b-a=\mu_X[a,b] \quad ,0\leq a\leq b\leq 1.$$

Now let us assume that we define another probability measure $\tilde{\mathbb{P}}$ on $[0,1]$ by specifying

$$\tilde{\mathbb{P}}[a,b]=\int_a^b 2\omega\,d\omega=b^2-a^2 \quad ,\ 0\leq a\leq b\leq 1.$$

The above equation and the proberties of probability measures determine $\tilde{\mathbb{P}}(B)$ for every Borel subset B of $\mathbb{R}$. Note that $\tilde{\mathbb{P}}[0,1]=1$ so in fact $\tilde{\mathbb{P}}$ is a probability measure. Under $\tilde{\mathbb{P}}$, the random variable X no longer has the uniform distribution. Denoting the distribution measure of X under $\tilde{\mathbb{P}}$ by $\tilde{\mu}_X$, we have

$$\tilde{\mu}_X[a,b]=\tilde{\mathbb{P}}\{\omega;\, a\leq X(\omega)\leq b\}=\tilde{\mathbb{P}}[a,b]=b^2-a^2 \quad ,\ 0\leq a\leq b\leq 1.$$

Let us now find the distribution measure of Y under $\tilde{\mathbb{P}}$. We have

$$\tilde{\mu}_Y[a,b]=\tilde{\mathbb{P}}\{\omega;\, a\leq Y(\omega)\leq b\}=\tilde{\mathbb{P}}\{\omega;\, a\leq 1-\omega\leq b\}=$$

$$=\tilde{\mathbb{P}}\{1-b,1-a\}=(1-a)^2-(1-b)^2 \quad ,\ 0\leq a\leq b\leq 1.$$

There are also other ways to record the distribution of a random variable rather than specifying the distribution measure μ_X. We can describe the distribution of a random variable in terms of its cumulative distribution function (cdf)

$$F(x) = \mathbb{P}\{X \le x\}$$

Note that

$$F(x) = \mu_X[-\infty, x]$$

and

$$\mu_X[x,y] = F(y) - F(x).$$

A special case is when there is a density function $f(x)$, a nonnegative function defined for $x \in \mathbb{R}$ such that

$$\mu_X[a,b] = \mathbb{P}\{a \le X \le b\} = \int_a^b f(x)\,dx \quad , -\infty \le a \le b < \infty .$$

In particular, because the closed intervals have union $\mathbb{R}$, we must have

$$\int_{-\infty}^{\infty} f(x)dx = \lim_{n\to\infty} \int_{-n}^{n} f(x)dx = \lim_{n\to\infty} \mathbb{P}\{-n \le X \le n\}$$
$$= \mathbb{P}\{X \in \mathbb{R}\} = \mathbb{P}(\Omega) = 1.$$

A second special case is the discrete random variable, let say X, where with probability one it takes values on a set with a finite number of elements $x_1, x_2, \dots, x_N$ or with on a set with a countable number of elements $x_1, x_2, \dots$. We then define

$$p_i = \mathbb{P}\{X = x_i\}$$

with each p_i nonnegative, and

$$\sum_i p_i = 1$$

The probability $\mathbb{P}\{X = x_i\}$ is called the probability mass function.

The mass assigned to a Borel set $B \subset \mathbb{R}$ by the distribution measure of X is

$$\mu_X(B) = \sum_{\{i,\, x_i \in B\}} p_i \quad , \quad B \in \mathcal{B}(\mathbb{R}).$$

There are random variables whose distribution is given by a mixture of a density and a probability mass function. Also there are random variables whose distributions has no lumps of mass but neither does it have a density. Random variables of this last type have applications in finance.

2.5 Independence of Random Variables.

Independence of two events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a very well known concept. In an analogous way we want to define independence of two random variables

<u>Definition 2.12</u>: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}$ and $\mathcal{E}$ be sub-σ-algebras of $\mathcal{F}$. We say these two σ-algebras are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B) \text{ for all } A \in \mathcal{G},\ B \in \mathcal{E}.$$

Let X and Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that these random variables are independent if the σ-algebras the generate, $\sigma(X)$ and $\sigma(Y)$, are independent. We say that the random variable X is independent of the σ-algebra $\mathcal{G}$ if $\sigma(X)$ and $\mathcal{G}$ are independent.

In other words we say that X and Y are independent if and only if

$$\mathbb{P}\{X \in A \text{ and } Y \in B\} = \mathbb{P}\{X \in A\}\,\mathbb{P}\{Y \in B\}$$

for every $A, B \in \mathcal{B}(\mathbb{R})$.

The above definition generalizes straightforwardly for n-random variables and also for a sequence of random variables.

<u>Definition 2.13</u>: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \ldots$ be a sequence of sub-σ-algebras of $\mathcal{F}$. For a fixed n, we say that the n σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ are independent if

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2) \cdots \mathbb{P}(A_n)$$

$$\text{for all } A_1 \in \mathcal{G}_1,\ A_2 \in \mathcal{G}_2, \ldots, A_n \in \mathcal{G}_n.$$

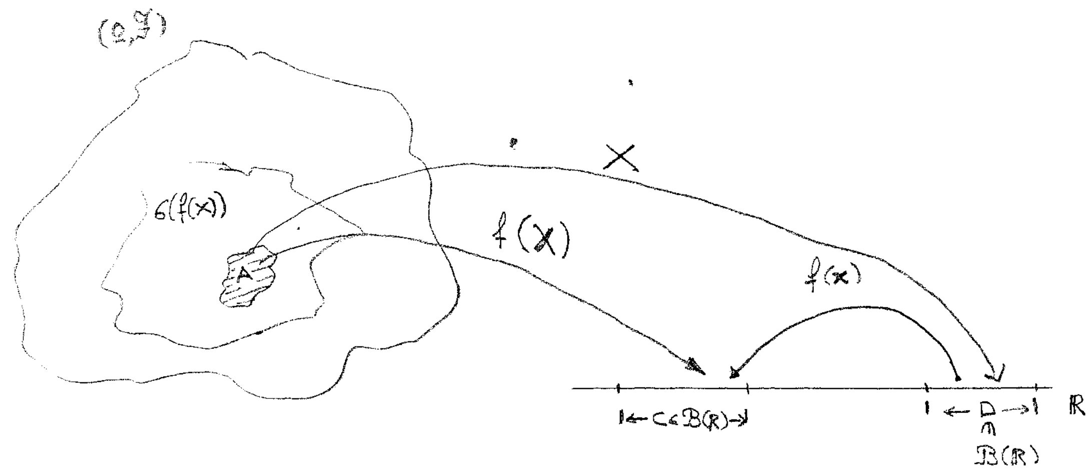
Let $X_1, X_2, X_3, \ldots$ be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say the n random variables $X_1, X_2, \ldots, X_n$ are independent if the σ-algebras $\sigma(X_1), \sigma(X_2), \ldots, \sigma(X_n)$ are independent. We say the full sequence of σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots$ is independent if for every positive integer n, the σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ are independent. We say the full sequence of random variables $X_1, X_2, X_3, \ldots$ is independent if, for every positive integer n, the n random variables $X_1, X_2, \ldots, X_n$ are independent.

The above definitions although precise and rigorous are very difficult to be used in practice except in very special cases. Instead the following results are more often used to check independence of random variables.

Theorem 2.5. Let X and Y be independent random variables, and let f and g be Borel-measurable functions on $\mathbb{R}$. Then $f(X)$ and $g(Y)$ are independent random variables.

Proof: Since X is a random variable then so is $f(X)$ and let $\sigma(f(X))$ be the σ-algebra generated by $f(X)$. We will show that $\sigma(f(X)) \subset \sigma(X)$ To show this (Figure 2.8) it is sufficient to show that if $A \in \sigma(f(X))$ then also $A \in \sigma(X)$.

Figure 2.8.



Since $A \in \sigma(f(X))$ then A is of the type

$$A = \{\omega \in \Omega : f(X(\omega)) \in C \in \mathcal{B}(\mathbb{R})\}$$

Now define by

$$D = \{x \in \mathbb{R} : f(x) \in C \in \mathcal{B}(\mathbb{R})\} \in \mathcal{B}(\mathbb{R})$$

Obviously for each $x \in \mathbb{R}$ there is an $f(x) \in C$ for which there is at least an event $\omega \in A$ for which $f(X(\omega)) = f(x)$. Vice versa for each $\hat{\omega} \in A$ then by definition $f(X(\hat{\omega})) \in C \in \mathcal{B}(\mathbb{R})$ and there exists at least one $\hat{x} \in D \in \mathcal{B}(\mathbb{R})$ such that $f(x) = f(X(\hat{\omega}))$. Thus

$$A = \{\omega \in \Omega : f(X(\omega)) \in C \in \mathcal{B}(\mathbb{R})\} = \{\omega \in \Omega, X(\omega) \in D\}.$$

The set on the right hand side of the above equation belongs to $\sigma(X)$ since $D \in \mathcal{B}(\mathbb{R})$ thus $A \in \sigma(X)$.

Now let B be in the σ-algebra generated by $g(Y)$. Then with the same argument we have that $B \in \sigma(Y)$. Since X and Y are independent we have

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$$

and that conclude the proof.

Definition 2.14. Let X and Y be random variables. The pair of random variables (X,Y) takes values in the plane $\mathbb{R}^2$ and the joint distribution measure is given by

$$\mu_{X,Y}(C) = \mathbb{P}\{(X,Y)\in C,\ C\in\mathcal{B}(\mathbb{R}^2)\}$$

[One way to generate the σ-algebra of Borel subsets of $\mathbb{R}^2$ is to start with the collection of closed rectangles $[a_1,b_1]\times[a_2,b_2]$ and then add all other sets necessary in order to have a σ-algebra.].

The joint cumulative distribution function of (X,Y) is

$$F_{X,Y}(a,b) = \mu_{X,Y}((-\infty,a]\times(-\infty,b]) = \mathbb{P}\{X\le a,\ Y\le b\},\ a\in\mathbb{R},\ b\in\mathbb{R}.$$

We say that a nonnegative, Borel-measurable function $f_{X,Y}(x,y)$ is a joint density for the pair of random variables (X,Y) if

$$\mu_{X,Y}(C) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbb{I}_C(x,y)\,f_{X,Y}(x,y)\,dx\,dy \quad \text{for all } C\in\mathcal{B}(\mathbb{R}^2),$$

where $\mathbb{I}_C(x,y)$ is defined to be

$$\mathbb{I}_C(x,y) = \begin{cases} 1 & \text{if } x,y\in C \\ 0 & \text{otherwise}\end{cases}$$

The previous equation for $\mu_{X,Y}(C)$ holds if and only if there exist $f_{X,Y}(x,y)$:

$$F_{X,Y}(a,b) = \int_{-\infty}^{\infty}\int_{-\infty}^{b} f_{X,Y}(x,y)\,dy\,dx \quad \text{for all } a\in\mathbb{R},\ b\in\mathbb{R}$$

The <u>marginal distribution measures</u> of X and Y are

$$\mu_X(A) = \mathbb{P}\{X\in A\} = \mu_{X,Y}(A\times\mathbb{R}) \quad \text{for all } A\in\mathcal{B}(\mathbb{R})$$

$$\mu_Y(B) = \mathbb{P}\{X\in B\} = \mu_{X,Y}(\mathbb{R}\times B) \quad \text{for all } B\in\mathcal{B}(\mathbb{R}).$$

If the joint density exists, then the <u>marginal densities</u> exist and are given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

The marginal densities, if they exist, are nonnegative, Borel measurable functions that satisfy

$$\mu_X(A) = \int_A f_X(x)\,dx \quad \text{for all } A\in\mathcal{B}(\mathbb{R}).$$

$$\mu_Y(B) = \mathbb{P}\{Y \in B\} = \mu_{X,Y}(\mathbb{R} \times B) \quad \text{for all } B \in \mathcal{B}(\mathbb{R})$$

The marginal cumulative distributions functions are

$$F_X(a) = \mu_X(-\infty, a] = \mathbb{P}\{X \leq a\} \quad \text{for all } a \in \mathbb{R}$$

$$F_X(b) = \mu_Y(-\infty, b] = \mathbb{P}\{Y \leq b\} \quad \text{for all } b \in \mathbb{R}.$$

If $f_{X,Y}(x,y)$ exists then we have also the existence of the marginal densities $f_X(x)$ and $f_Y(y)$ which are given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx.$$

In this case the marginal distribution measures are given by

$$\mu_X(A) = \int_A f_X(x)\,dx \quad \text{for every } A \in \mathcal{B}(\mathbb{R}).$$

$$\mu_Y(A) = \int_A f_Y(y)\,dy \quad \text{for every } B \in \mathcal{B}(\mathbb{R}).$$

The above relatis hold also if and only if

$$F_X(a) = \int_{-\infty}^{a} f_X(x)\,dx \quad \text{for all } a \in \mathbb{R}$$

$$F_Y(b) = \int_{-\infty}^{b} f_Y(y)\,dy \quad \text{for all } b \in \mathbb{R}.$$

We now provide the following theorem with the use of which we can check if two random variables X and Y are independent

Theorem 2.6 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X and Y be random variables. The following statements are equivalent

(i) X and Y are independent

(ii) The joint distribution measure is the product of the two marginal distribution measures

$$\mu_{X,Y}(A \times B) = \mu_X(A)\,\mu_Y(B) \quad \text{for all } A, B \in \mathcal{B}(\mathbb{R}).$$

(iii) The joint cumulative distribution measure is equal with the product of the marginal cummulative distribution functions

$$F_{X,Y}(a,b) = F_X(a)\, F_Y(b) \text{ for all } a,b \in \mathbb{R}.$$

(iii) The joint density is equal with the product of the marginal densities

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y) \text{ for almost every } x,y \in \mathbb{R}.$$

(v) The expectation of the produt XY is equal with the product of the expectations of each random variable

$$E(XY) = E(X)\,E(Y)$$

provided that

$$E[XY) < \infty.$$

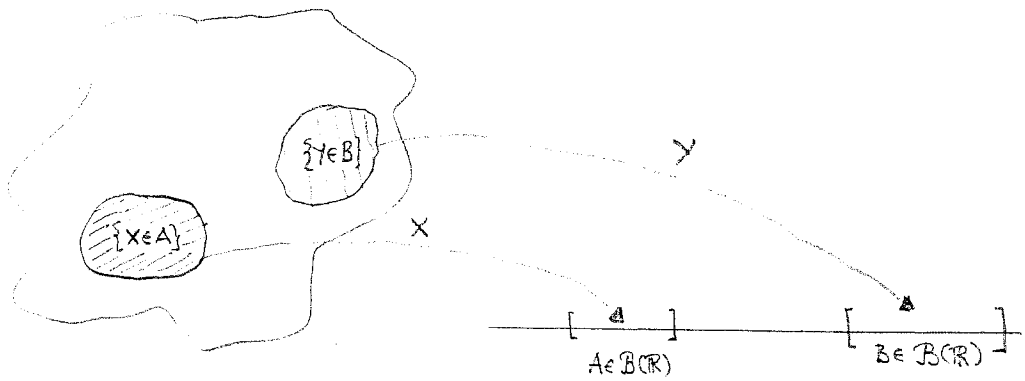
(iv) The joint moment-generating function is given by

$$E[e^{uX+vY}] = E[e^{uX}] \cdot E[e^{vY}]$$

for all $u \in \mathbb{R}$, $v \in \mathbb{R}$ for which the expectations are finite.

Proof: We start with (i) $\rightsquigarrow$ (ii) i.e we assume that X and Y are independent. Then (see also figure 2.9)

Figure 2.9



$$\mu_{X,Y}(A\times B) = \mathbb{P}\{X\in A \text{ and } Y \in B\}$$
$$= \mathbb{P}[\{X\in A\} \cap \{Y\in B\}] = (X,Y \text{ are independent})$$
$$= \mathbb{P}\{X\in A\} \cdot \mathbb{P}\{Y \in B\} = \mu_X(A)\cdot \mu_Y(B).$$

(ii) $\rightsquigarrow$ (i).

It is sufficient to show that for any element of $\sigma(X)$ and any element of $\sigma(Y)$ the probability of their intersection is equal with the product of the probabilities of each respective element. Let $A \in \mathcal{B}(\mathbb{R})$ an arbitrary element, then $\{X\in A\} \in \sigma(X)$

is equivalent of selecting arbitrarily an element of $\sigma(X)$. The same applies for $\{Y\in B\} \in \sigma(Y)$ for an arbitrarily selected set $B\in\mathcal{B}(\mathbb{R})$. Then we have

$$\begin{aligned}\mathbb{P}(\{X\in A\}\cap\{Y\in B\}) &= \mathbb{P}\{X\in A \text{ and } Y\in B\}\\ &= \mu_{X,Y}(A\times B) = \mu_X(A)\mu_Y(B) =\\ &= \mathbb{P}\{X\in A\}\,\mathbb{P}\{Y\in B\}.\end{aligned}$$

(ii) $\rightsquigarrow$ (iii) This is quite straightforward

$$\begin{aligned}F_{X,Y}(a,b) &= \mu_{X,Y}((-\infty,a]\times(-\infty,b])\\ &= \mu_X(-\infty,a]\,\mu_Y(-\infty,b]\\ &= F_X(a)\,F_Y(b).\end{aligned}$$

We will not go into the remaining proofs for equivalence of the various parts some of which are straightforward and can be taken as exersices and others are beyond the scope of the present.

2.6. Expectation.

Let X be a random variable defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ When Ω is finite or countably finite then the expected value of X, denoted by $\mathbb{E}(X)$, is well known as a kind of weighted mean, where weights are the respective probabilities of the various values of the random variable. Difficulty arises, however, if Ω is uncountably infinite. Naturally, in such a case, we must think in terms of integrals. The difficulty that arises however is that the Riemann integral cannot be used since the upper Riemann sum and the lower Riemann sum cannot be defined in the customery method of partitioning the x-axis. This is so due to the fact that all too often Ω is not a subset of $\mathbb{R}$. There is no natural way to partition the set Ω as we partition a closed interval in $\mathbb{R}$. To overcome this difficulty assume for the moment that

$$0\le X(\omega)<\infty \quad \text{for every } \omega\in\Omega$$

and let the partition Π with elements

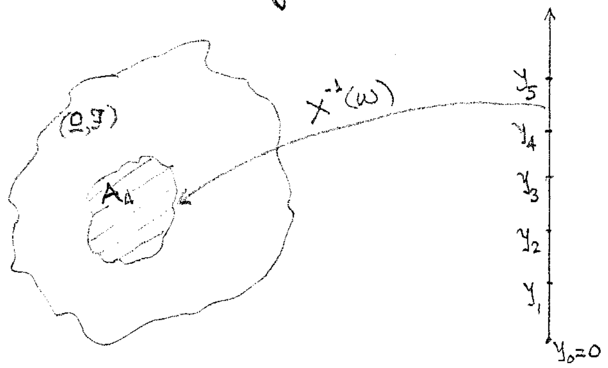
$$\Pi=\{y_0,y_1,y_2,\cdots\} \quad \text{where } 0=y_0<y_1<y_2<\cdots$$

Now for each interval $[y_k,y_{k+1}]$ define

$$A_k=\{\omega\in\Omega\,;\ y_k\le X(\omega)<y_{k+1}\}$$

see figure 2.10

Figure 2.9



Then we define the lower Lebesgue sum to be

$$LS_{\pi}^{-}(X) = \sum_{k=1}^{\infty} y_k \mathbb{P}(A_k) \quad .$$

This lower sum converges as the maximal distance between the y_k partition points, approaches zero, and we define this limit to be the Lebesgue integral

$$\int_{\Omega} X(\omega)\, d\mathbb{P}(\omega) \text{ or simply } \int_{\Omega} X\, d\mathbb{P}$$

We assumed so far that $0 \le X(\omega) < \infty$ for every $\omega \in \Omega$. If the set of $\omega \in \Omega$ that violates this condition has zero probability, there is no effect on the integral we just defined. If $\mathbb{P}\{\omega : X(\omega) \ge 0\} = 1$ but $\mathbb{P}\{\omega ; X(\omega) = \infty\} > 0$, then we define

$$\int_{\Omega} X(\omega)\, d\mathbb{P}(\omega) = \infty \quad .$$

It is not difficult to consider random variables X that can take both positive and negative values. Define by

$$X^{+}(\omega) = \max\{X(\omega), 0\} \quad , X^{-}(\omega) = \max\{-X(\omega), 0\}$$

Both $X^{+}(\omega)$ and $X^{-}(\omega)$ are non-negative random variables and

$$X = X^{+} - X^{-}$$

and we can define

$$\int_{\Omega} X(\omega)\, d\mathbb{P}(\omega) = \int_{\Omega} X^{+}(\omega)\, d\mathbb{P}(\omega) - \int_{\Omega} X^{-}(\omega)\, d\mathbb{P}(\omega)$$

If

$$\int_{\Omega} X^{+}(\omega)\, d\mathbb{P}(\omega) < \infty \quad \text{and} \quad \int_{\Omega} X^{-}(\omega)\, d\mathbb{P}(\omega) < \infty$$

we say that X is <u>integrable</u> and $\int_{\Omega} X(\omega)\, d\mathbb{P}(\omega)$ is finite

If

$$\int_\Omega X^+(\omega)\, d\mathbb{P}(\omega) = \infty \quad \text{and} \quad \int_\Omega X^-(\omega)\, d\mathbb{P}(\omega) < \infty,$$

then

$$\int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \infty .$$

If

$$\int_\Omega X^+(\omega)\, d\mathbb{P}(\omega) < \infty \text{ and } \int_\Omega X^-(\omega)\, d\mathbb{P}(\omega) = \infty,$$

then

$$\int_\Omega X(\omega)\, d\mathbb{P}(\omega) = -\infty$$

In the remaining cases the Lebesgue integrale $\int_\Omega X(\omega) d\mathbb{P}(\omega)$ is not defined.

We provide now without proof some useful properties of the Lebesgue integral.

Properties: Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

a) The random variable X is integrable if and only if

$$\int_\Omega |X(\omega)|\, d\mathbb{P}(\omega) < \infty$$

b) Let also Y a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$

If $X \le Y$ almost surely i.e $\mathbb{P}\{X \le Y\} = 1$ and if

$\int_\Omega X(\omega)\, d\mathbb{P}(\omega)$ and $\int_\Omega Y(\omega)\, d\mathbb{P}(\omega)$ are defined, then

$$\int_\Omega X(\omega)\, d\mathbb{P}(\omega) \le \int_\Omega Y(\omega)\, d\mathbb{P}(\omega).$$

c) If $X = Y$ almost surely and one of the integrals is defined, then they are both defined and

$$\int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \int_\Omega Y(\omega)\, d\mathbb{P}(\omega)$$

d) If X and Y are integrable random variables and $a, b \in \mathbb{R}$

OR

$a, b \geq 0$ and $X \geq 0$ and $Y \geq 0$

then

$$\int_\Omega (a X(\omega) + b Y(\omega))\, d\mathbb{P}(\omega) = a \int_\Omega X(\omega)\, d\mathbb{P}(\omega) + b \int_\Omega Y(\omega)\, d\mathbb{P}(\omega)$$

e) If A and B are disjoint sets in $\mathcal{F}$ then

$$\int_{A \cup B} X(\omega)\, d\mathbb{P}(\omega) = \int_A X(\omega)\, d\mathbb{P}(\omega) + \int_B X(\omega)\, d\mathbb{P}(\omega).$$

We are now in a position to define the expectation of a random variable X in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ in terms of the Lebesgue integral.

Definition 2.15. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If X is integrable i.e if

$$E[|X|] = \int_\Omega |X(\omega)| \, d\mathbb{P}(\omega) < \infty$$

or if $X \geq 0$ almost surely, then the expected value of X is defined to be

$$E[X] = \int_\Omega X(\omega) \, d\mathbb{P}(\omega) .$$

The following is an interesting result with applications in Finance.

Theorem 2.7: If φ is a convex, real-valued function defined on $\mathbb{R}$, and if X is an integrable random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ then the following holds

$$\varphi(E[X]) \leq E[\varphi(X)]$$

This result is well known as the Jensen inequality. In addition if φ is concave then we have that

$$\varphi(E[X]) \geq E[\varphi(X)]$$

Proof: We start by refreshing our knowledge for convex functions $\varphi : \mathbb{R} \to \mathbb{R}$. A function φ is convex in an interval $[a,b] \subseteq \mathbb{R}$ if for every $x_1, x_2 \in [a,b]$ and $0 < \lambda < 1$ we have that

$$\lambda \varphi(x_1) + (1-\lambda)\varphi(x_2) \geq \varphi(\lambda x_1 + (1-\lambda)x_2).$$

Assuming that $\varphi(x)$ for $x \in [a,b]$ is twice differentiable then $\varphi(x)$ is convex if and only if

$$\varphi''(x) = \frac{d^2\varphi(x)}{dx^2} \geq 0 .$$

Let $h(\omega) = X(\omega) - E[X]$. We use the Taylor expansion up to the second derivative to get that

$$\varphi(X(\omega)) = \varphi[E(X)] + [X(\omega) - E(X)]\varphi'[E(X)] + \tfrac{1}{2}[X(\omega) - E(X)]^2 \varphi''[X(\omega^*)]$$

where $X(\omega^*)$ is a point in the interval $[X(\omega), X(\omega) + h(\omega)]$.

We now take expectation on both sides of the above equation

$$E[\varphi(X)] = \int_\Omega \varphi(X(\omega))\, d\mathbb{P}(\omega) =$$

$$= \int_\Omega \varphi[E(X)]\, d\mathbb{P}(\omega) + \int_\Omega [X(\omega) - E(X)]\varphi'[E(X)]\, d\mathbb{P}(\omega) +$$

$$+ \tfrac{1}{2}\int_\Omega [X(\omega) - E(X)]^2 \varphi''[X(\omega^*)]\, d\mathbb{P}(\omega)$$

$$= \varphi[E(X)] \int_\Omega d\mathbb{P}(\omega) + \varphi'[E(X)] \int_\Omega [X(\omega) - E(X)]\, d\mathbb{P}(\omega) +$$

$$+ \tfrac{1}{2}\int_\Omega [X(\omega) - E(X)]^2 \varphi''[X(\omega^*)]\, d\mathbb{P}(\omega)$$

$$= \left(\text{since } \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = E(X) \text{ and } \int_\Omega d\mathbb{P}(\omega) = 1\right)$$

$$= \varphi[E(X)] + \tfrac{1}{2}\int_\Omega [X(\omega) - E(X)]^2 \varphi''[X(\omega^*)]\, d\mathbb{P}(\omega)$$

$$= (\text{since the integrand is always nonnegative})$$

$$\geqslant \varphi[E(X)].$$

The proof for the concave case is similar with the difference being that since $\varphi(x)$ is concave then $\varphi''(x) \leq 0$. ▲

Remark.

It can be proved that the Riemann integral $\int_a^b f(x)dx$ and Lebesgue integral

$$\int_{[a,b]} f(x)\, d\mathcal{L}(x)$$

have a close interelation. In fact

"If the Riemann integral $\int_a^b f(x)dx$ **exists** then f is Borel measurable so the Lebesgue integral

$$\int_{[a,b]} f(x)\, d\mathcal{L}(x)$$

is also defined and then the Riemann integral and the Lebesgue integral agree. Because of that reason we shall use the familiar notation $\int_a^b f(x)dx$ to denote also the Lebesgue integral. If the set A over which we wish to integrate is not an interval, we shall write $\int_A f(x)dx$.

When we are developing theory, we shall understand $\int_A f(x)dx$ to be a Lebesgue integral; when we need to compute, we will use techniques learned in calculus for computing Riemann integrals.

However the abstract space Ω is not a pleasant enviroment in which to actually compute integrals. For computations, we often need to rely on densities of the random variables under consideration, and we integrate these over $\mathbb{R}$ than on Ω.

We now provide the following Theorem the proof of which has an important methodology which is called the <u>standard machine</u>.

<u>Theorem 2.8</u>: Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, X a random variable and let g be a Borel-measurable function on $\mathbb{R}$. Then if

$$E[|g(X)|] = \int_{\mathbb{R}} |g(x)| \, d\mu_X(x) < \infty$$

then

$$E[g(X)] = \int_{\mathbb{R}} g(x) \, d\mu_X(x).$$

<u>Proof</u>: <u>Step 1</u>: Indicator functions

Suppose the function $g(x) = \mathbb{I}_B(x)$ the indicator function of $B \in \mathcal{B}(\mathbb{R})$.

Since $\mathbb{I}_B(x) \geq 0$ both the above integrals reduce to the integral

$$E[\mathbb{I}_B(X)] = \int_{\mathbb{R}} \mathbb{I}_B(x) \, d\mu_X(x)$$

We have that

$$E[\mathbb{I}_B(X)] = 1 \; \mathbb{P}\{X \in B\} + 0 \; \mathbb{P}\{X \notin B\} = \mathbb{P}\{X \in B\}$$

Also since for a particular value of $X = x$, $\mathbb{I}_B(x)$ takes the value 1 if $x \in B$ and 0 if $x \notin B$ we get that

$$\int_{\mathbb{R}} \mathbb{I}_B(x) \, d\mu_X(x) = 1 \; \mu_X\{x : \mathbb{I}_B(x) = 1\} + 0 \; \mu_X\{x : \mathbb{I}_B(x) = 0\} = \mu_X(B)$$

However by definition $\mu_X(B) = \mathbb{P}\{X \in B\}$ and that concludes the proof of step 1.

<u>Step 2</u>: Nonnegative simple functions

A simple function is a linear function of a finite number of Indicator functions i.e we assume that

$$g(x) = \sum_{k=1}^{n} a_k \mathbb{I}_{B_k}(x). \quad \text{with } a_i \geq 0 \text{ and } B_k \in \mathcal{B}(\mathbb{R}).$$

We easily get that

$$E[g(X)] = E\left[\sum_{k=1}^{n} a_k \mathbb{I}_{B_k}(X)\right] = \sum_{k=1}^{n} a_k E[\mathbb{I}_{B_k}(X)] =$$

$$= \sum_{k=1}^{n} a_k \int_{\mathbb{R}} \mathbb{I}_{B_k}(x) \, d\mu_X(x) =$$

$$= \sum_{k=1}^{n} a_k \int_{\mathbb{R}} \mathbb{I}_{B_k}(x)\, d\mu_X(x) = \int_{\mathbb{R}} \Big(\sum_{k=1}^{n} a_k \mathbb{I}_{B_k}(x) \Big) d\mu_X(x) = \int_{\mathbb{R}} g(x)\, d\mu_X(x)$$

which concludes the proof of step 2.

Step 3. Nonnegative Borel-measurable functions.

Let $g(x)$ be an arbitrary nonnegative Borel-measurable function defined on $\mathbb{R}$. For each positive integer n define the sets

$$B_{k,n} = \left\{ x : \frac{k}{2^n} \leq g(x) \leq \frac{k+1}{2^n} \right\}, \quad \text{for } k = 0, 1, 2, \ldots, 4^n - 1.$$

For each fixed n, the sets

$$B_{0,n},\ B_{1,n},\ \ldots,\ B_{4^n-1,n}$$

correspond to the partition

$$0 < \frac{1}{2^n} < \frac{2}{2^n} < \cdots < \frac{4^n}{2^n} = 2^n$$

At the next stage $n+1$, the partition points include all those at stage n and new partition points at the midpoints between the old ones. Because of this fact, the simple functions

$$g_n(x) = \sum_{k=0}^{4^n-1} \frac{k}{2^n} \mathbb{I}_{B_{k,n}}(x)$$

satisfy

$$0 \leq g_1 \leq g_2 \leq \cdots \leq g.$$

Furthermore

$$g_n \approx g \quad \text{as } n \text{ becomes larger}$$

with

$$\lim_{n \to \infty} g_n(x) = g(x)$$

From step 2, we know that

$$E[g_n(X)] = \int_{\mathbb{R}} g_n(x)\, d\mu_X(x) \quad \forall n$$

We now provide a well known theorem known as the Monotone convergence theorem which we will use in order to conclude step 3.

Theorem 2.9.

Let $g_1, g_2, g_3, \ldots$ be a sequence of Borel-measurable functions on $\mathbb{R}$ converging almost everywhere (everywhere except to a set of points with Lebesgue measure zero) to a function f. If in addition

$$0 \le f_1 \le f_2 \le f_3 \le \cdots \text{ almost everywhere}$$

then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx$$

▲.

Using now the above theorem 2.9, we get

$$E[g(X)] = \lim_{n\to\infty} E[g_n(X)] = \lim_{n\to\infty} \int_{\mathbb{R}} g_n(x)\,d\mu_X(x) = \int_{\mathbb{R}} g(x)\,d\mu_X(x),$$

which concludes the proof of step 3.

Step 4 : General Borel-measurable function.

Let $g(x)$ be a general Borel measurable function, which can take both positive and negative values. The functions

$$g^+(x) = \max\{g(x), 0\} \text{ and } g^-(x) = \max\{-g(x), 0\}.$$

are both nonnegative and from Step 3 we have

$$E[g^+(X)] = \int_{\mathbb{R}} g^+(x)\,d\mu_X(x) \text{ and } E[g^-(X)] = \int_{\mathbb{R}} g^-(x)\,d\mu_X(x)$$

Now we have - from our initial condition - that

$$E[|g(X)|] = \int_{\mathbb{R}} |g(x)|\,d\mu_X(x) = \int_{\mathbb{R}} g^+(x)\,d\mu_X(x) + \int_{\mathbb{R}} g^-(x)\,d\mu_X(x) < \infty$$

From which we get

$$\int_{\mathbb{R}} g^+(x)\,d\mu_X(x) < \infty$$

and

$$\int_{\mathbb{R}} g^-(x)\,d\mu_X(x) < \infty.$$

By substracting the above equations we get

$$\int_{\mathbb{R}} g^+(x)\,d\mu_X(x) - \int_{\mathbb{R}} g^-(x)\,d\mu_X(x) = \int_{\mathbb{R}} (g^+(x) - g^-(x))\,d\mu_X(x) =$$

$$\int_{\mathbb{R}} g(x)\,d\mu_X(x) = E[g(X)].$$

▲

Remark.

The above theorem provides us with the alternative that in order to compute the Lebesgue integral

$$E[X] = \int_{\Omega} X(\omega)\,d\mathbb{P}(\omega)$$

+ over the abstract space Ω, it suffices to compute the integral

$$\int_{\mathbb{R}} g(x)\, d\mu_X(x)$$

over the set of real numbers. However the integrator is the distribution measure μ_X with which altough for theoretical work we can go far, to actually perform a computation, we need some ferther tools. Depending on the nature of the random variable X, the distribution measure μ_X can have different forms.

The most common case for continuous-time models in finance is when X has a density. This means that there is a nonnegative, Borel-measurable function f defined on $\mathbb{R}$ such that

$$\mu_X(B) = \int_B f(x)\, dx \quad \text{for every } B \in \mathcal{B}(\mathbb{R}).$$

If X has a density, we can use this density to compute expectations as shown by the following theorem.

<u>Theorem 2.10</u>: Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable on it. Let g be a Borel-measurable function on $\mathbb{R}$. Suppose that X has a density f, then if

$$E[|g(X)|] = \int_{-\infty}^{\infty} |g(x)|\, f(x)\, dx < \infty$$

we have that

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)$$

<u>Proof</u>: The proof is similar with the proof of Theorem 2.8 using the methodology of the "standard machine".

2.7. Conditional Expectation.

Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let also $\mathcal{G}$ a σ-algebra with $\mathcal{G} \subseteq \mathcal{F}$. Consider a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ and assume that $\mathcal{G}$ does not contain all the necessary information to know X but at the same time it is not independent of X. In other words it contains "some" information about X. The major question of this section how we can use the informatin in $\mathcal{G}$

to get advantage of this information. This is in fact all to often the case in Finance and also in many respects in life. The conditional expectation of X given $\mathcal{G}$ is usually an answer to this problem.

To start with more familiar grounds let that X takes the distinct values

$$x_1, x_2, \dots, x_n$$

and let also Z be a random with variable with distinct values.

$$z_1, z_2, \dots, z_m$$

Now consider the elementary conditional expectation

$$Y_j = E[X \mid z_j]$$

Then if we call $Y = E[X|Z]$ then actually Y is a random variable with values

$$Y_1 = E[X|Z=z_1],\ Y_2 = E[X|Z=z_2], \dots, Y_m = E[X|Z=z_m]$$

and as is well known

$$E[X|Z=z_j] = \sum_{i=1}^{n} x_i\, \mathbb{P}[X=x_i \mid Z=z_j].$$

Now Y being a random variable will almost surely have an expected value

$$E[Y] = E[E[X|Z]] = \sum_{j=1}^{m} E[X|Z=z_j] \cdot \mathbb{P}[Y=Y_j] =$$

$$= \sum_{j=1}^{m} \sum_{i=1}^{n} x_i\, \mathbb{P}[X=x_i \mid Z=z_j] \cdot \mathbb{P}[Z=z_j]$$

$$= \sum_{j=1}^{m} \sum_{i=1}^{n} x_i\, \mathbb{P}[X=x_i, Z=z_j] =$$

$$= \sum_{i=1}^{n} x_i \sum_{j=1}^{m} \mathbb{P}[X=x_i, Z=z_j] =$$

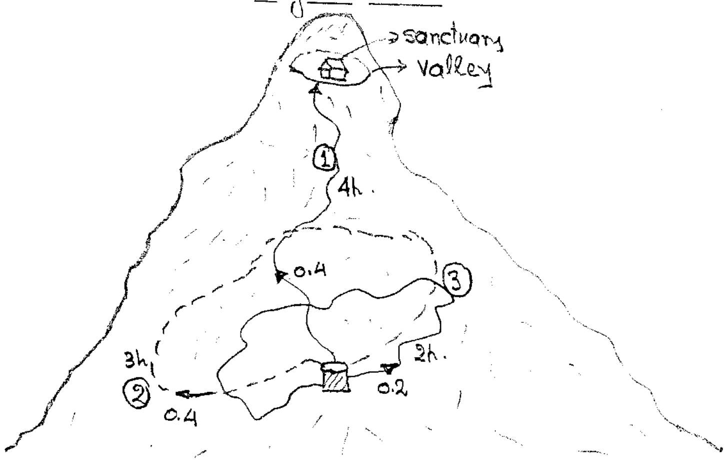
$$= \sum_{i=1}^{n} x_i\, \mathbb{P}[X=x_i] = E[X]$$

Now we will give an example which will help to gain some more depth on the concept of conditional expectation.

<u>Example 2.6</u>

A mountaineer has started climbing on mountain Olympus (home for the Greek gods in ancient Greece) from the altitude of 950 meters where all possible transportation of our century stops. The weather is cloudy and there are sporadic showers which makes a beautiful sound falling on the leafs of the forest trees. After a 2,5 hours of walk, thunders start to strike in the forest in a distance not far than 300-400 meters from the mountaineer. This new development increases his fears which together with tiredness and loneliness are a very bad company for him. He arrives at this moment in a clearing at the altitude of 1500 meters and in front of him is there is a sabor well behind of which he can see three pathways. In his condition he is not able to judge which pathway is the correct one which will lead him in the valley of the Godesses in an expected time of 4 hours. Let us denote by 1 the correct pathway and 2 and 3 the two wrong ones. He chooses pathway 2 with probability 0.4 and in such a case he does a cycle which brings him again at his starting point. This cycle has an expected duration of 3 hours. He chooses pathway 3 with probability 0.2 and again he makes a cycle with expected duration of 2 hours. The time is 1.30pm, nightfall is at 9.00pm, he has no flashlight with him and the sanctuaries are in the valey of the Godesses. Do you expect him to arrive on time and survive?

Figure 2.10



Define the random variable Y as follows

$Y=1$ if the moutaineer chooses the correct pathway

$Y=2$ if the mountaineer chooses pathway 2

$Y=3$ if the mountaineer chooses pathway 3.

- We have that

$$\mathbb{P}[Y=1]=0.4\ ;\ \mathbb{P}[Y=2]=0.4\ ;\ \mathbb{P}[Y=3]=0.2$$

From our last reasoning we have that

$$E[X]=E[E[X|Y]]=\sum_y E[X|Y=y]\,\mathbb{P}[Y=y]=E[X|Y=1]\,\mathbb{P}[Y=1]+$$
$$E[X|Y=2]\,\mathbb{P}[Y=2]+E[X|Y=3]\,\mathbb{P}[Y=3]$$

From the data we have we get

$E[X|Y=1]=E$[the time needed for the mountaineer to go from the jabor well to the valley of Godesses | the mountaineer is on the correct pathway] $=4^h$.

Also we get that

$E[X|Y=2]=E$[the time needed for the mountaineer to go from the jabor well to the valley of Godesses | the mountaineer is on the wrong pathway 2] $=$

$=E$[the time needer for the mountaineer to complete the cycle of pathway 2 and return to the jabor well]$+$

E[the time needer for the mountaineer to go from the jabor well to the valley of Godesses] $=3+E[X]$

It is thus implied that after so many hours of hardship he has forgotten which pathway he took last and consequently pathway 2 is among his options with the same probability.

The third step is similar with the second

$E[X|Y=3]=E$[the time needed for the mountaineer to go from the jabor well to the valley of Godesses | the mountaineer is on the wrong pathway 3] $=$

$=E$[the time needed for the mountaineer to complete the cycle of pathway 3 and return to the jabor well]$+$

$+ E[$the time need for the mountaineer to go from the jabor well to the valley of the Godesses$] = 2 + E[X]$

Thus from all the above arguments we get that

$$E[X] = 4 \cdot 0.4 + (3 + E[X])\,0.4 + (2 + E[X]) \cdot 0.2$$

Solving this equation we get

$$E[X] = 8^{h}.$$

Thus he is expected to arrive at the valley of the Godesses at 9.30 pm.
Do we know what will actually happen?

▲

We now provide a general and precise definition for the conditional expectation of a random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ given that the information we have on X is not complete but it is contained in a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$

<u>Definition</u>. <u>2.16</u>. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X a random variable on it that is either nonnegative or integrable, and let $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra.
The conditional expectation of X given $\mathcal{G}$, denoted $E[X|\mathcal{G}]$, is any random variable that satisfies

(i) $E[X|\mathcal{G}]$ is $\mathcal{G}$-measurable

(ii) $E[|X||\mathcal{G}] < \infty$

(iii) $$\int_A E[X|\mathcal{G}](\omega)\,d\mathbb{P}(\omega) = \int_A X(\omega)\,d\mathbb{P}(\omega) \quad \text{for all } A \in \mathcal{G}$$

The first logical question that arises is if such a random variable $E[X|\mathcal{G}]$ as the one defined in Definition 2.16 exists. The immediate question that comes to mind since the answer is affirmative is if it is unique. The answer is given in the form of a theorem without its proof.

<u>Theorem</u>. <u>2.11</u> Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X a random variable on it that is either nonnegative or integrable, and let $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. Then

there exists a random variable $E[X|\mathcal{G}]$ which satisfies (i), (ii) and (iii) of definition 2.16 and it is unique.

Proof: The uniqueness is easy and can be done as an exercise. The existence is hard. $\blacktriangle$

Property (i) of definition 2.16 captures the fact that the estimate $E[X|\mathcal{G}]$ of X is based on the information in $\mathcal{G}$. In addition it guarantees that, although the estimate of X based on the information in $\mathcal{G}$ is itself a random variable, the value of the estimate $E[X|\mathcal{G}]$ can be determined from the information in $\mathcal{G}$. The (iii) property ensures that $E[X|\mathcal{G}]$ is an unbiased estimate of X.

We will now provide some properties of conditional expectations some of which look quite familiar but they are in a more general context.

(i). Linearity of conditional expectations

If X and Y are integrable random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra and $a_1, a_2 \in \mathbb{R}$ then

$$E[a_1 X + a_2 Y | \mathcal{G}] = a_1 E[X|\mathcal{G}] + a_2 E[Y|\mathcal{G}] .$$

Firstly since $E[X|\mathcal{G}]$ and $E[Y|\mathcal{G}]$ are $\mathcal{G}$-measurables the same applies with $E[a_1 X + a_2 Y|\mathcal{G}]$ and so property (i) is satisfied. Similarly we arrive that the (ii) property is satisfied. We now go for property (iii):

We have for every $A \in \mathcal{G}$ that

$$\int_A E[a_1 X + a_2 Y | \mathcal{G}] \, d\mathbb{P}(\omega) = \int_A (a_1 E[X|\mathcal{G}] + a_2 E[Y|\mathcal{G}]) \, d\mathbb{P}(\omega) =$$

$$= a_1 \int_A E[X|\mathcal{G}] \, d\mathbb{P}(\omega) + a_2 \int_A E[Y|\mathcal{G}] \, d\mathbb{P}(\omega)$$

$$= a_1 \int_A X(\omega) d\mathbb{P}(\omega) + a_2 \int_A Y(\omega) \, d\mathbb{P}(\omega)$$

$$= \int_A (a_1 X(\omega) + a_2 Y(\omega)) \, d\mathbb{P}(\omega)$$

Thus $E[a_1 X + a_2 Y|\mathcal{G}]$ as defined satisfies the three properties needed.

(ii) Taking out what is known.

If X and Y are integrable random variables on a probability space $(\Omega,\mathcal{F},\mathbb{P})$, $\mathcal{G}\subseteq\mathcal{F}$ is a σ-algebra, XY is integrable and X is $\mathcal{G}$-measurable then

$$E[XY|\mathcal{G}] = X E[Y|\mathcal{G}].$$

Firstly $E[XY|\mathcal{G}]$ as defined in the previous equation is $\mathcal{G}$-measurable since X is $\mathcal{G}$-measurable by hypothesis and $E[Y|\mathcal{G}]$ is $\mathcal{G}$-measurable by definition. Similar arguments apply on integrability. We must now verify property (iii).

The appropriate methodology is that of the standard machine. For our purposes it is sufficient to prove property (iii) when X is an indicator function. The rest is given as an exercise.

Now according to our hypothesis X is a $\mathcal{G}$-measurable indicator random variable. That means $X=\mathbb{I}_B$ for any set $B\in\mathcal{G}$. Now we have that

$$\begin{aligned}
\int_A E[XY|\mathcal{G}](\omega)\,d\mathbb{P}(\omega) &= \int_A X(\omega)\,E[Y|\mathcal{G}](\omega)\,d\mathbb{P}(\omega)\\
&= \int_A \mathbb{I}_B(\omega)\,E[Y|\mathcal{G}](\omega)\,d\mathbb{P}(\omega)\\
&= \int_{A\cap B} E[Y|\mathcal{G}](\omega)\,d\mathbb{P}(\omega)\\
&= \int_{A\cap B} Y(\omega)\,d\mathbb{P}(\omega)\\
&= \int_A \mathbb{I}_B(\omega)\,Y(\omega)\,d\mathbb{P}(\omega)\\
&= \int_A X(\omega)\,Y(\omega)\,d\mathbb{P}(\omega)
\end{aligned}$$

<u>(iii) Iterated conditioning</u>

Let $(\Omega,\mathcal{F},\mathbb{P})$ a probability space and X an integrable random variable on it. Let $\mathcal{G}$ and $\mathcal{J}$ be σ-algebras such that $\mathcal{J}\subset\mathcal{G}\subseteq\mathcal{F}$. Then the following is true

$$E\big[E[X|\mathcal{G}]\,\big|\,\mathcal{J}\big] = E[X|\mathcal{J}]$$

This is probably the most useful property of conditional integration in Stochastic Finance. Observe the first part of the equation, there we have the random variable $E[X|\mathcal{G}]$ and we take its conditional expectation with respect to σ-algebra $\mathcal{J}$. Thus as defined in order to satisfy property (i) it should be $\mathcal{Y}$-measure. However this is true since $E[X|\mathcal{Y}]$ is $\mathcal{Y}$-measurable by definition. The same reasoning applies for the integrability property (ii).

To prove property (iii) we should show that

$$\int_A E[E[X|\mathcal{G}]|\mathcal{Y}](\omega)\,d\mathbb{P}(\omega) = \int_A E[X|\mathcal{Y}](\omega)\,d\mathbb{P}(\omega).$$

Showing in fact that we can replace $E[E[X|\mathcal{G}]|\mathcal{Y}]$ in the first equation by $E[X|\mathcal{Y}]$.

On the left hand side of the previous equation the random variable is $E[X|\mathcal{G}]$ thus from property (iii) for that we get

$$\int_A E[E[X|\mathcal{G}]|\mathcal{Y}](\omega)\,d\mathbb{P}(\omega) = \int_A E[X|\mathcal{G}](\omega)\,d\mathbb{P}(\omega)$$

Since $A\in\mathcal{Y}$ and $\mathcal{Y}\subset\mathcal{G}$ we have that $A\in\mathcal{G}$ also and thus.

$$\int_A E[X|\mathcal{G}](\omega)\,d\mathbb{P}(\omega) = \int_A X(\omega)\,d\mathbb{P}(\omega).$$

Now since $E[X|\mathcal{Y}]$ is a conditional expectation and $A\in\mathcal{Y}$ we get that

$$\int_A E[X|\mathcal{Y}](\omega)\,d\mathbb{P}(\omega) = \int_A X(\omega)\,d\mathbb{P}(\omega).$$

That concludes the proof of Iterated conditioning.

(iv) <u>Independence</u>.

If X is integrable and independent of $\mathcal{G}$, then

$$E[X|\mathcal{G}] = E[X]$$

First we have to show that

$$E[X] \text{ is } \mathcal{G}\text{-measurable}.$$

However $E[X]$ is a constant and so it is measurable with respect to every σ-algebra.

What is left, is to prove that we can replace $E[X|\mathcal{G}]$ in property (iii) of conditional expectation with $E[X]$, i.e. to prove that

$$\int_A E[X]\,d\mathbb{P}(\omega) = \int_A X(\omega)\,d\mathbb{P}(\omega).$$

This is done with the methodology of the standard machine.

We start with the assumption that X is an indicator random variable which according to (iv) is independent of $\mathcal{G}$. In this respect let $X = \mathbb{I}_B$ where naturally the set B is indepent of $\mathcal{G}$. Thus for all $A \in \mathcal{G}$ we have

$$\int_A X(\omega)\, d\mathbb{P}(\omega) = \int_A \mathbb{I}_B(\omega)\, d\mathbb{P}(\omega) = \int_{A\cap B} d\mathbb{P}(\omega)$$

$$= \mathbb{P}(A\cap B) = (\text{due to the independence of } B \text{ from all } A \in \mathcal{G}) =$$

$$= \mathbb{P}(A) \cap \mathbb{P}(B) = \mathbb{P}(A) \cdot E[\mathbb{I}_B]$$

$$= \mathbb{P}(A) \cdot E[X] = \left(\int_A d\mathbb{P}(\omega)\right) \cdot E[X] =$$

$$= \int_A E[X]\, d\mathbb{P}(\omega).$$

Which completes the part of the proof for X being an indicator function in the methodology of the standard machine.

(v). Conditional Jensen's inequality

If φ is a convex, real valued function defined on $\mathbb{R}$, and if X is an integrable random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra then the following is true

$$\varphi(E[X|\mathcal{G}]) \leq E[\varphi(X)|\mathcal{G}].$$

In addition if φ is concave then

$$\varphi(E[X|\mathcal{G}]) \geq E[\varphi(X)|\mathcal{G}]$$

The proof is similar with the unconditional case.

(vi). Redundant Conditioning

Let X be an integrable random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}_1, \mathcal{G}_2$ be σ-algebras with $\mathcal{G}_1 \subset \mathcal{F}$ and $\mathcal{G}_2 \subset \mathcal{F}$. Suppose that $\sigma(X)$ and $\mathcal{F}_1$ are independent of $\mathcal{G}_2$. Then

$$E[X|\mathcal{G}_1 \vee \mathcal{G}_2] = E[X|\mathcal{G}_1]$$

Firstly we should have that $E[X|\mathcal{G}_1]$ is $\mathcal{G}_1 \vee \mathcal{G}_2$-measurable which is an immediate consequence of the fact that it is $\mathcal{G}_1$-measurable and the independence assumption. The same applies for the integrability property.

We will now prove property (iii). Firstly for $\mathcal{G}_1 \vee \mathcal{G}_2$ from page 2.6 we have that it is the σ-algebra generated by $\mathcal{G}_1 \cup \mathcal{G}_2$ and more importantly for our purpose it is the σ-algebra of all the sets S which are of the type

$$S = A \cap B \quad \text{with } A \in \mathcal{G}_1 \text{ and } B \in \mathcal{G}_2 .$$

Thus $A \cap B$ is a typical atom set of the σ-algebra $\mathcal{G}_1 \vee \mathcal{G}_2$ with $A \in \mathcal{G}_1$ and $B \in \mathcal{G}_2$.

Now we have to prove that

$$\int_{A\cap B} E[X \mid \mathcal{G}_1 \vee \mathcal{G}_2](\omega)\, d\mathbb{P}(\omega) = \int_{A\cap B} E[X \mid \mathcal{G}_1](\omega)\, d\mathbb{P}(\omega) .$$

We have that

$$\int_{A\cap B} E[X \mid \mathcal{G}_1 \vee \mathcal{G}_2](\omega)\, d\mathbb{P}(\omega) = \int_{A\cap B} X(\omega)\, d\omega =$$

$$= \int_\Omega \mathbb{I}_A(\omega)\, \mathbb{I}_B(\omega) X(\omega) d\mathbb{P}(\omega) = (\text{By the independence hypothesis}) =$$

$$= \int_\Omega \mathbb{I}_A(\omega) X(\omega) d\mathbb{P}(\omega) \int_\Omega \mathbb{I}(\omega)\, d\mathbb{P}(\omega) = \int_A X(\omega)\, d\mathbb{P}(\omega) \int_\Omega \mathbb{I}_B(\omega)\, d\mathbb{P}(\omega) =$$

$$= \int_A E[X \mid \mathcal{G}_1](\omega)\, d\mathbb{P}(\omega) \int_\Omega \mathbb{I}_B(\omega)\, d\mathbb{P}(\Omega) = (\text{By the independence}$$

$$\text{hypothesis}) = \int_{A\cap B} E[X \mid \mathcal{G}_1](\omega)\, d\mathbb{P}(\omega)$$

That completes the proof of Redundant Conditioning.

<u>(vii). Gross Conditioning.</u>

Let X be an integrable random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let the trivial σ-algebra $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then

$$E[X \mid \mathcal{F}_0] = E[X] .$$

We have that $E[X]$ is a constant and so is measurable with respect to any σ-algebra and so is $\mathcal{F}_0$-measurable. It is integrable by hypothesis.

Also let $A = \Omega$ or $\emptyset$ then obviously

$$\int_A E[X]\, d\mathbb{P}(\omega) = \int_A X(\omega)\, d\mathbb{P}(\omega)$$

Now using the properties of Redundant Conditioning and Gross Conditioning one could prove as an exercise the following:

Let X an integrable random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ a σ-algebra and $\mathcal{G} \subseteq \mathcal{F}$. Let also that the σ-algebra $\sigma(X)$ is independent of $\mathcal{G}$ then

$$E[X|\mathcal{G}] = E[X].$$

2.8. Change of Measure.

Let us start with a finite sample space Ω in order to grasp more easily the intuition behind the concept of the change of measure.

Consider a random variable X which takes values

$$x_1, x_2, \dots, x_n$$

with probabilities

$$p_1 = \mathbb{P}(X = x_1), \quad p_2 = \mathbb{P}(X = x_2), \quad \dots, \quad p_n = \mathbb{P}(X = x_n).$$

In such a case we know that the expected value of X is given by

$$E(X) = \sum_{i=1}^{n} x_i p_i$$

It is very well known that sometimes we transform a random variable by introducing a new random variable in order to change its expected value. Among such cases the most famous one is when $X \sim N(\mu, \sigma^2)$ and we tranforme X into $Y = X - \mu/\sigma$ which is distributed as a $N(0,1)$.

However there is in front of us another way of doing so which in stochastic Finance has proved to be very useful.

Assume that instead, we change the probabilities p_i $(i = 1, 2, \dots, n)$ creating a one to one correspondence

$$\begin{array}{cccc} p_1, & p_2, & \dots, & p_n \\ \Updownarrow & \Updownarrow & & \Updownarrow \\ \tilde{p}_1 & \tilde{p}_2 & & \tilde{p}_n \end{array}$$

and let that

$$\tilde{p}_1 = \tilde{\mathbb{P}}(X = x_1), \quad \tilde{p}_2 = \tilde{\mathbb{P}}(X = x_2), \quad \dots, \quad \tilde{p}_n = \tilde{\mathbb{P}}(X = x_n).$$

We do this with the constraint that if $p_i > 0$ for any i then $\tilde{p}_i > 0$ also, i.e the p_i and $\tilde{p}_i$ $(i = 1, 2, \dots, n)$ agree which values of X are impossible.

Naturally we demand that

$$\sum_{i=1}^{n} \tilde{p}_i = 1,$$

so that $\tilde{p}_i$ will be proper probability distribution for X. Then we say that $\tilde{\mathbb{P}}$ is a new probability measure in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We in addition say that $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are equivalent since they agree which values of X are impossible. Under the equivalent probability measure $\tilde{\mathbb{P}}$ we now have a new expected value for X given by

$$E_{\tilde{\mathbb{P}}}(X) = \sum_{i=1}^{n} x_i \, \tilde{\mathbb{P}}(X = x_i).$$

It is easy to see that there is an infinite number of equivalent probability measures that could be devised given the "real" one. However in Stochastic Finance we will desire that under the new equivalent probability measure some of our random variables will have some desired properties. The most common such case is that we want the prices of the assets, portofolios e.t.c. to be martingales. We usuall call such a probability measure an equivalent martingale measure. Its existence and uniqueness have direct implications about the market within which we trade.

Let us remain to finite sample space Ω but make things more rigorous and formal. Without much loss of generality we assume that we have two probability measures $\mathbb{P}$ and $\tilde{\mathbb{P}}$ that both give positive probability to every element of Ω, so we can write the quotient

$$Z(\omega) = \frac{\tilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}$$

Obviously $Z(\omega)$ is a random variable since a map from $\Omega \to \mathbb{R}$.

We call Z the <u>Radon-Nikodým derivative of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$</u> due to the fact that for infinite sample spaces this quotient will look closer to a derivative. In order for $\tilde{\mathbb{P}}$ to be an equivalent probability measure to $\mathbb{P}$ then the random variable Z must satisfy the following:

Theorem 2.11. Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be probability measures on a finite sample space Ω, assume that $\mathbb{P}(\omega)>0$ and $\tilde{\mathbb{P}}(\omega)>0$ for every $\omega\in\Omega$, and define the random variable

$$Z(\omega)=\frac{\tilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}.$$

Then we have the following

(i) $\mathbb{P}(Z>0)=1$;

(ii) $E_{\mathbb{P}}(Z)=1$;

(iii) for any random variable X

$$E_{\tilde{\mathbb{P}}}(X)=E_{\mathbb{P}}(ZX).$$

Proof: (i). Since $\mathbb{P}(\omega)$ and $\tilde{\mathbb{P}}(\omega)$ are positive for every ω, so is their quotient, with probability 1.

(ii). We have that

$$E_{\mathbb{P}}(Z)=\sum_{\omega\in\Omega}Z(\omega)\,\mathbb{P}(\omega)=\sum_{\omega\in\Omega}\frac{\tilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}\,\mathbb{P}(\omega)=\sum_{\omega\in\Omega}\tilde{\mathbb{P}}(\omega)=1.$$

(iii). It is straightforward to see that

$$E_{\tilde{\mathbb{P}}}(X)=\sum_{\omega\in\Omega}X(\omega)\,\tilde{\mathbb{P}}(\omega)=$$

$$=\sum_{\omega\in\Omega}X(\omega)\,\frac{\tilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}\cdot\mathbb{P}(\omega)$$

$$=\sum_{\omega\in\Omega}X(\omega)\,Z(\omega)\,\mathbb{P}(\omega)=\mathbb{E}(XZ)$$

▲

In other words in finite sample spaces Ω to change from $\mathbb{P}$ to $\tilde{\mathbb{P}}$, we need to reassign probabilities in Ω using Z to tell us where in Ω we should revise the probability upward (where $Z>1$) and where we should revise the probability downward (where $Z<1$). Note also that it is important that $E_{\mathbb{P}}(Z)=1$ in order that $\tilde{\mathbb{P}}$ is a probability measure in $(\Omega,\mathcal{F},\mathbb{P})$.

Let us now move into infinite sample spaces Ω through an example in order to grasp some of the insight before we move into formal and rigorous definitions.

Let a stochastic process $\{X_t\}_t$ and assume that for a specific value of t then the random variable X_t follows the standard normal distribution i.e

$$X_t \sim N(0,1).$$

It is known in this that the density is

$$f(x_t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_t^2}$$

Let us denote with $d\mathbb{P}(x_t)$ the probability

$$d\mathbb{P}(x_t) = \mathbb{P}\left\{x_t - \tfrac{1}{2}dx_t < x_t < x_t + \tfrac{1}{2}dx_t\right\} =$$

$$= \int_{x_t - \frac{1}{2}dx_t}^{x_t + \frac{1}{2}dx_t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_t^2} \, dx_t$$

Since we choose the interval $[x_t - \frac{1}{2}dx_t, x + \frac{1}{2}dx_t]$ to be very small we could assume that within this interval the function

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_t^2}$$

is a constant and thus we may write that

$$d\mathbb{P}(x_t) = \frac{1}{\sqrt{2\pi}} \int_{x_t - \frac{1}{2}dx_t}^{x_t + \frac{1}{2}dx_t} e^{-\frac{1}{2}x_t^2} dx_t$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_t^2} \left[x_t + \tfrac{1}{2}dx_t - x_t + \tfrac{1}{2}dx_t\right]$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_t^2} \, dx_t$$

It is easy to see that

$$\int_{-\infty}^{\infty} d\mathbb{P}(x_t) = 1.$$

Now define the function

$$Z(x_t) = \exp\left(x_t \mu - \tfrac{1}{2}\mu^2\right)$$

We then have that

$$d\mathbb{P}(x_t)\, Z(x_t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_t^2} dx_t \exp\left(x_t\mu - \tfrac{1}{2}\mu^2\right)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}x_t^2 + x_t\mu - \tfrac{1}{2}\mu^2\right\} dx_t$$

$$= \frac{1}{\sqrt{2\pi}} \exp\{-½(x_t-\mu)^2\} dx_t$$

We set

$$d\tilde{\mathbb{P}}(x_t) = d\mathbb{P}(x_t)\, Z(x_t)$$

It is easy to see that $d\tilde{\mathbb{P}}(x_t)$ is a probability measure since

$$\int_{-\infty}^{\infty} d\tilde{\mathbb{P}}(x_t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-½(x_t-\mu)^2\} dx_t = 1.$$

In addition $d\tilde{\mathbb{P}}(x_t)$ is driven by the normal distribution $N(\mu,1)$.

Thus by multiplying $d\mathbb{P}(x_t)$ with the function $Z(x_t)$ we arrived to a new probability measure $d\tilde{\mathbb{P}}(x_t)$ driven by the same distribution i.e the normal but with a different mean value and the same variance.

Observe that if we take the function

$$Z^{-1}(x_t) = \exp\{½\mu^2 - x_t\mu\}$$

then we have

$$Z^{-1}(x_t)\, d\mathbb{P}(x_t) = \exp\{½\mu^2 - x_t\mu\} \frac{1}{\sqrt{2\pi}} \exp\{-½(x_t-\mu)^2\} dx_t$$

$$= \frac{1}{\sqrt{2\pi}} \exp\{-½ x_t^2\} dx_t = d\mathbb{P}(x_t)$$

We could conclude the above observations by writting

$$\frac{d\tilde{\mathbb{P}}(x_t)}{d\mathbb{P}(x_t)} = Z(x_t)$$

This relation is called the Radon-Nikodym deridative of the probability measure $\tilde{\mathbb{P}}$ with respect to the probability measure $\mathbb{P}$.

The formal definition of the Radon-Nikodym derivative of the probability measure $\tilde{\mathbb{P}}$ with respect to the probability measure $\mathbb{P}$ is given in the next theorem

Theorem 2.12. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let Z be an almost surely nonnegative random variable with $E_{\mathbb{P}}(Z) = 1$

For any set $A \in \mathcal{F}$, define

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega)\, d\mathbb{P}(\omega).$$

Then $\tilde{\mathbb{P}}$ is a probability measure. Furthermore, if X is a nonnegative random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, then

$$E_{\tilde{\mathbb{P}}}[X] = E_{\mathbb{P}}[XZ]$$

If Z is almost surely strictly positive, then for every nonnegative random variable Y, we also have

$$E_{\mathbb{P}}[Y] = E_{\tilde{\mathbb{P}}}\left[\frac{Y}{Z}\right]$$

Proof: We start proving that $\tilde{\mathbb{P}}$ is a probability measure on $(\Omega, \mathcal{F}, \mathbb{P})$. We have that

$$\tilde{\mathbb{P}}(\Omega) = \int_\Omega Z(\omega)\, d\mathbb{P}(\omega) = E_{\mathbb{P}}(Z) = 1.$$

The second step is to proove that $\tilde{\mathbb{P}}$ is countably additive. In this respect let $A_1, A_2, \dots$ be a sequence of disjoint sets in $\mathcal{F}$.

Define by

$$B_n = \bigcup_{k=1}^{n} A_k \quad \text{and} \quad B_\infty = \bigcup_{k=1}^{\infty} A_k.$$

Since obviously

$$\mathbb{I}_{B_1} \leq \mathbb{I}_{B_2} \leq \mathbb{I}_{B_3} \leq \dots$$

and $\lim_{n\to+\infty} \mathbb{I}_{B_n} = \mathbb{I}_{B_\infty}$ we may use monotone converge to write

$$\tilde{\mathbb{P}}(B_\infty) = \int_\Omega \mathbb{I}_{B_\infty}(\omega) Z(\omega)\, d\mathbb{P}(\omega)$$

$$= \lim_{n\to+\infty} \int_\Omega \mathbb{I}_{B_n}(\omega) Z(\omega)\, d\mathbb{P}(\omega)$$

At the same time

$$\int_\Omega \mathbb{I}_{B_n}(\omega) Z(\omega)\, d\mathbb{P}(\omega) = \sum_{k=1}^{n} \int_\Omega \mathbb{I}_{A_k}(\omega) Z(\omega)\, d\mathbb{P}(\omega) =$$

$$= \sum_{k=1}^{n} \int_{A_k} Z(\omega)\, d\mathbb{P}(\omega) = \sum_{k=1}^{n} \tilde{\mathbb{P}}(A_k).$$

From the above arguments we get

$$\tilde{\mathbb{P}}\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(A_k) = \sum_{k=1}^{\infty} \tilde{\mathbb{P}}(A_k)$$

It remains to show that

$$E_{\tilde{\mathbb{P}}}(X) = E_{\mathbb{P}}(XZ)$$

The proof of this part follows the standard machine methodology. Thus we start assumming that X is an indicator function let say $\mathbb{I}_A$, then

$$E_{\tilde{\mathbb{P}}}[X] = E_{\tilde{\mathbb{P}}}[\mathbb{I}_B] = \tilde{\mathbb{P}}(A) = \int_\Omega \mathbb{I}_A(\omega) Z(\omega) d\mathbb{P}(\omega)$$
$$= E_{\mathbb{P}}[\mathbb{I}_A Z] = E_{\mathbb{P}}[XZ].$$

The rest of the proof follows the standard machine methology with no new concept difficulties ▲

We usually use the first equation of Theorem 2.12 in the form

$$Z(\omega) = \frac{d\tilde{\mathbb{P}}(\omega)}{d\mathbb{P}(\omega)}.$$

This is expressed formally in the following definition.

<u>Definition 2.17</u>. Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\tilde{\mathbb{P}}$ be another probability measure on $(\Omega, \mathcal{F})$ that is equivalent to $\mathbb{P}$, and let Z be an almost surely positive random variable that relates $\mathbb{P}$ and $\tilde{\mathbb{P}}$ with the relation

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega)$$

Then we call the random variable Z the Radon-Nikodým derivative of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$ and we write

$$Z = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}$$

<u>Important Remark</u>

Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\tilde{\mathbb{P}}$ and $\mathbb{Q}$ be two other probability measures on $(\Omega, \mathcal{F})$ that are equivalent to $\mathbb{P}$.

Let Z be an almost surely positive random variable that relates $\mathbb{P}$ and $\tilde{\mathbb{P}}$ i.e.

$$Z = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}.$$

Let also W an almost surely positive random variable that relates $\mathbb{P}$ and $\mathbb{Q}$ i.e

$$W = \frac{d\mathbb{Q}}{d\mathbb{P}} .$$

Then there exists an almost surely positive random variable Y that relates $\tilde{\mathbb{P}}$ and $\mathbb{Q}$ and is given by

$$Y = \frac{Z}{W} = \frac{d\tilde{\mathbb{P}}}{d\mathbb{Q}} ,$$

which is called the Radon-Nikodym derivative of $\tilde{\mathbb{P}}$ with respect to $\mathbb{Q}$. It is important to note that in this case

$$E(Y) = 1 .$$

Interesting Result

Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let X be a nonnegative random variable with the exponential distribution and parameter $\lambda > 0$.

Let $\tilde{\mathbb{P}}$ an equivalent probability measure on $(\Omega, \mathcal{F})$ to $\mathbb{P}$. Define by Z an almost surely positive random variable that relates $\mathbb{P}$ and $\tilde{\mathbb{P}}$ i.e

$$Z = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} .$$

Then let Z be

$$Z = \frac{\tilde{\lambda}}{\lambda} e^{-(\tilde{\lambda} - \lambda) X} \quad \text{with } \tilde{\lambda} > 0$$

Find the cumulative distribution function

$$\tilde{\mathbb{P}}\{X \leq x\} \quad \text{for all } x \geq 0$$

for the random variable X under the probability measure $\tilde{\mathbb{P}}$.

References

1. Brémaud, P. (1981). Point Processes and Queues. Martingale Dynamics. Springer Series in Statistics.
2. Grimmett, G. and Strirzaker (2001). Probability and Random Processes. Oxford University Press.
3. Shreve, S. (2004). Stochastic Calculus for Finance. Vol I. The Binomial Asset Pricing Model. Springer.
4. Shreve, S. (2004). Stochastic Calculus for Finance. Vol II. Continuous-Time Models. Springer.
5. Williams, D. (1991). Probability with Martingales. Cambridge Morthewatical textbooks.

Chapter 3

Martingales.

3.1. Introduction.

Martingales play a central role in the modern theory of stochastic processes. Martingales have a constant expected value, which remain the same under random stopping. Martingales converge almost surely Stochastic integrals are martingales. These proporties hold under certain conditions and are the most important contributions of these creatures to the evolution of stochastic Mathematics. Martingales play also an important part in the foundation of the theory of Stochastic Finance.

The roots of the study of Martingales are in the gambling area. In fact the world Martingale comes from an old strategy around 1815 where in any game one whenever he looses he invests double the money in the next game until he wins. The terminology martingale was given to this category of stochastic processes by J. Ville (1939). Martingales where extensively studied by Paul Lévy (1886-1971) from (1934) and by J.L Doob (1910-2002) from 1940.

3.2. Martingales with respect to a filtration.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_n\}_{n=0}^{\infty}$ be a sequence of σ-algebras on Ω such that

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n \subset \cdots \subset \mathcal{F}$$

i.e. they are a filtration on Ω.

We define by $\mathcal{F}_\infty$ the σ-algebra generated by the union of the σ-algebras $\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_n, \dots$ i.e

$$\mathcal{F}_\infty = \sigma\left(\bigcup_n \mathcal{F}_n\right) \subseteq \mathcal{F}$$

It is known that $\mathcal{F}_\infty$ is the minimum σ-algebra which is such that

$$\bigcup_n \mathcal{F}_n \in \mathcal{F}_\infty \quad \text{for every } n.$$

As it is evident we start with discrete time before to proceed with continuous time.
Let $\{X_n\}_{n=0}^\infty$ a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ then we say that $\{X_n\}_{n=0}^\infty$ is adapted to the filtration $\{\mathcal{F}_n\}_{n=0}^\infty$ if for every n

$$X_n \text{ is } \mathcal{F}_n\text{-measurable in } \Omega$$

Example 3.1.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space and let $\{Y_n\}_{n=0}^\infty$ a sequence of random variables. Define by $\mathcal{F}_n$ to be the σ-algebra generated from the random variables $\{Y_0, Y_1, \dots, Y_n\}$ on Ω. In other words we have that

$$\mathcal{F}_0 = \sigma(Y_0)$$

$$\mathcal{F}_1 = \sigma(Y_0, Y_1)$$

$$\dots\dots\dots$$

$$\mathcal{F}_{n-1} = \sigma(Y_0, Y_1, \dots, Y_{n-1})$$

$$\mathcal{F}_n = \sigma(Y_0, Y_1, \dots, Y_n)$$

Then apparently we have the filtration

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_{n-1} \subseteq \mathcal{F}_n \subset \dots \subset \mathcal{F}.$$

Let the sequence of random variables $\{X_n\}_{n=0}^\infty$ defined by

$$X_0 = g_0(Y_0),$$

$$X_1 = g_1(Y_0, Y_1),$$

$$X_2 = g_2(Y_0, Y_1, Y_2),$$

$$\dots\dots\dots$$

$$X_n = g_n(Y_0, Y_1, \dots, Y_n),$$

where the functions $g_0(\cdot), g_1(\cdot), \dots, g_n(\cdot)$ are $\mathcal{B}(\mathbb{R})$-measurable. Then it is easy to see that the stochastic process $\{X_n\}_{n=0}^{\infty}$ is adapted to the sequence of σ-algebras $\{\mathcal{F}_n\}_{n=0}^{\infty}$.

We will now provide the definition of a Martingale in relation with a filtration of σ-algebras.

Definition 3.1 Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{X_n\}_{n=0}^{\infty}$ a sequence of random variables i.e a stochastic process. Let $\{\mathcal{F}_n\}_{n=0}^{\infty}$ be a sequence of σ-algebras in Ω, which are a filtration i.e.

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n \subseteq \dots \subseteq \mathcal{F}$$

Then $\{X_n\}_{n=0}^{\infty}$ is called a martingale in relation with the filtation $\{\mathcal{F}_n\}_{n=0}^{\infty}$ if the following conditions are satisfied

a) The stochastic process $\{X_n\}_{n=0}^{\infty}$ is adapted to the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$.

b) $E[|X_n|] < \infty$ for every n.

c) $E[X_n \mid \mathcal{F}_{n-1}] = X_{n-1}$ almost surely.

Example 3.2. We will now provide an example which clarifies the above definition but at the same time highlights a classical application of the concept of martingale in real problems.

Let the players A, B and C contest the following game. At each stage two of them are randomly chosen in sequence, with the first one chosen being required to give one unit of money to the other. Define by X_n, Y_n and Z_n be the random variables that denote, respectively, the amount of money that players A, B and C have after the nth stage play. All of the possible choices are equally likely and successive choises are independent of the past. This continues until one of the players has no remaining coins. At this point that player departs and the other two continue playing until one of them has all the units of money. Assume $X_0 = x$, $Y_0 = y$ and $Z_0 = z$.

If we denote by T the duration of the game, then T is the number of games needed that two of the values X_n, Y_n and Z_n to be equal to 0. We are looking for a way to find $E[T]$.

This is when we are trying to solve a problem which involves the use of martingales Being familiar with conditional expectation it is not difficult to make progress with stochastic arguments involving martingales. The great difficulty is to find the correct random variable which needs to be proved that it is a martingale. Naturally all too often this difficulty is not measured accuretly when looking at a solved problem.

Let in this respect the random variable

$$M_n = X_n Y_n + X_n Z_n + Y_n Z_n + n .$$

Define by

$$\mathcal{F}_n^{XYZ} = \sigma(X_i, Y_i, Z_i, i=0,1,\dots,n),$$

then it is easy to see that $\{M_n\}_{n=0}^{\infty}$ is adapted to the filtration $\{\mathcal{F}_n^{XYZ}\}_{n=0}^{\infty}$.

Apparently also $E[|M_n|] < \infty$ since x, y, z the initial capital of the three players is finite. Thus what remains to be shown is that

$$E[M_{n+1} \mid \mathcal{F}_n] = M_n \quad \text{for every } n \in \mathbb{N} .$$

Since $\mathcal{F}_n$ is known then we know the values of X_n, Y_n and Z_n.

If $X_n Y_n Z_n > 0$ then all three players are in the game and we know their capital at time n, let

$$X_n = x_n ,\ Y_n = y_n \text{ and } Z_n = z_n .$$

Hence at time $n+1$, X_{n+1} is possible to take three values x_n+1 or x_n or x_n-1. according to results of chosing two out of three players. The same applies for Y_{n+1} and Z_{n+1} whose possible values are respectively y_n+1, y_n or y_n-1 and z_n+1, z_n or z_n-1. Thus we have

$$E[X_{n+1} Y_{n+1} \mid X_n = x_n, Y_n = y_n] = (\text{two players only are chosen})$$

$$= \left[(x_n-1)(y_n+1) + x_n(y_n-1) + x_n(y_n+1) + (x_n-1)y_n + \right.$$

$$\left. + (x_n+1)(y_n-1) + (x_n+1)y_n \right] \cdot \tfrac{1}{6} =$$

$$= x_n y_n - 1/3 .$$

In a similar way we find that

$$E[Y_{n+1} Z_{n+1} \mid Y_n = y_n , Z_n = z_n] = y_n z_n - 1/3 ,$$

$$E[X_{n+1} Z_{n+1} \mid X_n = x_n , Z_n = z_n] = x_n z_n - 1/3$$

and so we get that

$$E[M_{n+1} / \mathcal{F}_n] = E[X_{n+1} Y_{n+1} / \mathcal{F}_n] + E[X_{n+1} Z_{n+1} | \mathcal{F}_n] + E[Y_{n+1} Z_{n+1} | \mathcal{F}_n] + n+1$$

$$= E[X_{n+1} Y_{n+1} / X_n = x_n , Y_n = y_n] + E[X_{n+1} Z_{n+1} | X_n = x_n , Z_n = z_n]$$

$$+ E[Y_{n+1} Z_{n+1} \mid Y_n = y_n , Z_n = z_n] + n+1$$

$$= x_n y_n + x_n z_n + y_n z_n + n = M_n .$$

Hence in the case of $X_n Y_n Z_n > 0$, M_n is a martingale.

The alternative is that by stage n one of the players has been eliminated from the game. Without any loss of generality assume that player A has left the game and so in this case $X_n = 0$. Then

$$E[Y_{n+1} Z_{n+1} \mid Y_n = y_n , Z_n = z_n] = [(y_n - 1)(z_n + 1) + (y_n + 1)(z_n - 1)] \cdot 1/2$$

$$= y_n z_n - 1 .$$

Thus it is easy to see that also in this case we have

$$E[M_{n+1} / \mathcal{F}_n] = M_n$$

Therefore under any possible outcome of the game M_n is a martingale.

Now since M_n is a martingale we know by the application of the optional stopping theorem which we will present it latter on that

$$E[T] = M_0 = xy + xz + yz.$$

▲

We now provide the definition of a <u>supermartingale</u> with respect to a filtration. Define by $X_n^- = \min\{X_n, 0\}$.

<u>Definition 3.2.</u> Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\{X_n\}_{n=0}^{\infty}$ a stochastic process. Let $\{\mathcal{F}_n\}_{n=0}^{\infty}$ be a sequence of σ-algebras in Ω, which are a filtration. Then $\{X_n\}_{n=0}^{\infty}$ is called a supermartingale in relation with

with the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$ if the following conditions are satisfied.

a) The stochastic process $\{X_n\}_{n=0}^{\infty}$ is adapted to the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$.

b) $E[X_n^-] > -\infty$

c) $E[X_n/\mathcal{F}_{n-1}] \leq X_{n-1}$ almost surely.

Let us define by

$$X_n^+ = \max\{X_n, 0\}$$

Definition 3.3: Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\{X_n\}_{n=0}^{\infty}$ be a stochastic process. Let $\{\mathcal{F}_n\}_{n=0}^{\infty}$ be a sequence of σ-algebras in Ω, which are a filtration. Then $\{X_n\}_{n=0}^{\infty}$ is called a submartingale in relation with the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$ if the following conditions are satisfied

a) $E[X_n^+] < \infty$

b) The stochastic process $\{X_n\}_{n=0}^{\infty}$ is adapted to the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$.

c) $E[X_n/\mathcal{F}_{n-1}] \geq X_{n-1}$.

Now the construction of a mathematical model of uncertainty and of flow of information in continuous time follows the same ideas as in discrete time, but it is much more complicated. Let $S(t)$ be the price of an asset when time changes continuously between 0 and T. We take then as Ω the set of all possibilities of movements of the process. A first logical reaction is to this of Ω as the set of all continuous functions on $[0,T]$ which is a rich space. However we will assume that the observed processes that belong to Ω are of a more general type

Definition 3.4. A stochastic process $\{X_t\}_{t\geq 0}$ is said to be càdlàg if it almost surely has sample paths which are right continuous with left limits. The nonsensical words àdlàg are acronyms from the French for continu à droite, limites à gauche. Similarly, a stochastic process $\{X_t\}_{t\geq 0}$ is said to be càglad if it almost surely has sample paths which are left continuous with right limits.

For our purposes in what follows we assume that we are given a complete probability $(\Omega, \mathcal{F}, \mathbb{P})$. In addition we are given a filtration $(\mathcal{F}_t)_{0\le t\le\infty}$ and we will alway assume that the following usual hypothesis hold.

<u>Usual Hypothesis</u>

(i) $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets of $\mathcal{F}$

(ii) $\mathcal{F}_t = \bigcap_{u>t} \mathcal{F}_u$ all t, $0\le t<\infty$; that is, the filtration $\mathbb{F} = \{\mathcal{F}_t\}$ is right continuous.

Let the sample space $\Omega = D[0,T]$ be the set of all cádlàg function on $[0,T]$. An element of this set, ω is a càdlàg function from $[0,T]$ into $\mathbb{R}$. First we must decide what kind of sets of these functions are measurable. The simplest sets for which we would like to calculate probabilities are sets of the form $\{a \le S(t_1) \le b\}$. for some t_1. We are also interested in how the price at time t_1 affects the price at another time t_2. Thus we need to find the joint distribution of asset prices $S(t_1)$ and $S(t_2)$. This means that we need to have in $\mathcal{F}$ sets of the form $\{S(t_1)\in B_1,\ S(t_2)\in B_2\}$ where $B_1, B_2 \in \mathcal{B}(R)$. More generally we would like to have finite-dimensional distributions of the process $S(t)$, that is, probabilities of the sets $\{S(t_1)\in B_1, S(t_2)\in B_2, \dots, S(t_n)\in B_n\}$ for any possible choise of $0\le t_1 \le t_2, \dots, t_n \le T$., where $B_i \in \mathcal{B}(\mathbb{R})$, $i=1,2,\dots,n$.

The sets of the form

$$\{\omega(\cdot)\in D[0,T] : \omega(t_1)\in B_1, \dots, \omega(t_n)\in B_n\}$$

were $B_i \in \mathcal{B}(\mathbb{R})$; $i=1,2,\dots,n$ are intervals on the line, are called *cylinder sets* or *finite-dimentional rectangles*. Probability is defined first on the cylinder sets, and then extended to the σ-field $\mathcal{F}$ generated by the cylinders, that is, the smallest σ-field containing all cylinder sets. One needs to be careful with consistency of probability defined on cylinder sets, so that when one cylinder contains another no contradiction of probability assignment is obtained. In this respect the Caratheodory's extension theorem plays a crusial part.

Let $S(t)$ be defined for $0 \le t \le T$, then for a fixed ω it is a function in t, called the sample path of $\{S(t)\}_{t \ge 0}$. Finite-dimensional distributions do not determine the continuity of sample paths.

<u>Example</u> <u>3.2</u> Let $X(t) = 0$ for all t, $0 \le t \le 1$. Also let τ be a uniformly distributed random variable on $[0,1]$.

Define the stochastic process $\{Y(t)\}_{t \ge 0}$ as follows

$$Y(t) = \begin{cases} 0 & \text{for } t \ne \tau \\ 1 & \text{for } t = \tau \end{cases}$$

Now we have

$$\mathbb{P}\{Y(t) \ne 0\} = \mathbb{P}\{\tau = t\} = 0$$

hence

$$\mathbb{P}\{Y(t) = 0\} = 1 - \mathbb{P}\{\tau = t\} = 1.$$

Also

$$\mathbb{P}\{X(t) = Y(t)\} = 1 - \mathbb{P}\{X(t) \ne Y(t)\} =$$
$$= 1 - \mathbb{P}\{t = \tau\} = 1.$$

So all one-dimensional distributions of $X(t)$ and $Y(t)$ are the same. Similarly all finite-dimensional distributions of X and Y are the same. However, the sample paths of the process $\{X(t)\}_{t \ge 0}$ are continuous in t, whereas every sample path $Y(t)_{0 \le t \le 1}$ has a jump at the point τ. In fact that leads us to the following distribution

<u>Definition</u> <u>3.5.</u> Two stochastic versions are called <u>versions</u> (<u>modifications</u>) of one another if

$$\mathbb{P}(X(t) = Y(t)) = 1 \quad \text{for all } t,\ 0 \le t \le T.$$

The two processes in Example 3.2 are versions of one another, one has continuous paths and the other does not. If we agree to pick any version of the process we want, then we can pick the continuous version when it exists. In general we choose the smoothest possible version of the process.

For two processes $\{X(t)\}_{t \ge 0}$ and $\{Y(t)\}_{t \ge 0}$ denote by

$$\mathcal{N}(t) = \{X(t) \neq Y(t)\}, \quad 0 \leq t \leq T.$$

In example 3.2 we have

$$\mathbb{P}\{\mathcal{N}(t)\} = 0 \quad \text{for any } t \text{ in } 0 \leq t \leq 1.$$

However,

$$\mathbb{P}\Big(\bigcup_{0 \leq t \leq 1} \mathcal{N}(t)\Big) = \mathbb{P}\big(\tau = t \text{ for some } t \text{ in } [0,1]\big) = 1$$

Thus although, each of $\mathcal{N}(t)$ is a $\mathbb{P}$-null set, the union

$$\mathcal{N} = \bigcup_{0 \leq t \leq 1} \mathcal{N}(t)$$

contains uncountably many null sets and in this case it is a set of probability one. That leads us to the following definition.

<u>Definition 3.6</u>. Let $\{X(t)\}_{t \geq 0}$ and $\{Y(t)\}_{t \geq 0}$ be two stochastic process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Also let

$$\mathcal{N}(t) = \{X(t) \neq Y(t)\} \quad \text{and} \quad \mathcal{N} = \bigcup_{0 \leq t \leq T} \mathcal{N}(t) \quad \text{for} \quad 0 \leq t \leq T.$$

Then the two stochastic processes $\{X(t)\}_{t \geq 0}$ and $\{Y(t)\}_{t \geq 0}$ are called <u>indistinguisable</u> if $\mathbb{P}(\mathcal{N}) = 0$ or equivalently

$$\mathbb{P}\big(\{\omega : \exists\, t : X(t) \neq Y(t)\}\big) = \mathbb{P}\Big(\bigcup_{0 \leq t \leq T} \{X(t) \neq Y(t)\}\Big) = 0$$

or

$$\mathbb{P}\Big(\bigcap_{0 \leq t \leq T} \{X(t) = Y(t)\}\Big) = \mathbb{P}\big(X(t) = Y(t) \text{ for all } t \in [0,T]\big) = 1.$$

<u>Result</u> : Let $\{X(t)\}_{t \geq 0}$ and $\{Y(t)\}_{t \geq 0}$ be two stochastic processes on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that versions of one another. If in addition they are right continuous then they are indistinguisable.

<u>Result</u> : Let $\{X(t)\}_{t \geq 0}$ and $\{Y(t)\}_{t \geq 0}$ be two stochastic processes which are càdlàg. If $\{X(t)\}_{t \geq 0}$ is a version of $\{Y(t)\}_{t \geq 0}$ then they are indinstinguisable.

We now provide the definition of a stopping time.

<u>Definition 3.7</u> : A random variable $\tau : \Omega \to [0, \infty]$ is a <u>stopping time</u> if the event $\{T \leq t\} \in \mathcal{F}_t$, every t, $0 \leq t \leq \infty$.

One important consequence of the "Usual Hypothesis" is the following Theorem.

<u>Theorem 3.1</u>. The event $\{T < t\} \in \mathcal{F}_t$, $0 \leq t \leq \infty$, if and only if T is a stopping time.

Proof: For any $\epsilon > 0$ we have that

$$\{T \le t\} = \bigcap_{t+\epsilon > u > t} \{T < u\},$$

hence

$$\{T \le t\} \in \bigcap_{u>t} \mathcal{F}_u = \mathcal{F}_t,$$

so T is a stopping time.

Conversely,

$$\{T < t\} = \bigcup_{t > \epsilon > 0} \{T \le t - \epsilon\},$$

and thus $\{T \le t - \epsilon\} \in \mathcal{F}_{t-\epsilon}$, hence also in $\mathcal{F}_t$. ▲

Result. Let T_1 and T_2 be stopping times. Then the following are stopping times:
(i) $\min(T_1, T_2)$, (ii) $\max(T_1, T_2)$, (iii) $T_1 + T_2$, (iv) $a T_1$, where $a > 1$.

We now provide the definition of a hitting time

Definition 3.8 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a filtered probability space, $\{X(t)\}_{t \ge 0}$ a càdlàg stochastic process, and let $B \in \mathcal{B}(\mathbb{R})$. Define

$$T(\omega) = \inf\{t > 0 : X(t) \in B\}$$

Then T is called a hitting time of B for $\{X(t)\}_{t>0}$.

Result: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a filtered probability space, $\{X(t)\}_{t \ge 0}$ be an adapted càdlàg stochastic process, and let B be an open set in $\mathcal{B}(\mathbb{R})$. The hitting time of B for $\{X(t)\}_{t>0}$ is a stopping time.

Result: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a filtered probability space, $\{X(t)\}_{t \ge 0}$ be an adapted càdlàg stochastic process and B be a closed set. Then the random variable

$$T(\omega) = \inf\{t > 0 : X_t(\omega) \in B \text{ or } X_{t^-}(\omega) \in B\},$$

where

$$X_{t^-}(\omega) = \lim_{s \uparrow t} X_s(\omega)$$

is a stopping time.

The σ-algebra $\mathcal{F}_t$ can be thought of as representing all observable events up to and including time t. We would like to have an analogous notion of events that are observable before a stopping time.

Definition 3.9: Let T be a stopping time. The stopping time σ-algebra $\mathcal{F}_T$ is defined to be

$$\{A \in \mathcal{F} : A \cap \{T \le t\} \in \mathcal{F}_t, \text{ all } t \ge 0\}.$$

More intuitive though is the following result

Result: Let T be a finite stopping time. Then $\mathcal{F}_T$ is the smallest σ-algebra containing all càdlàg processes sampled at T. That is

$$\mathcal{F}_T = \sigma\{X_T ; X \text{ all adapted càdlag processes}\}.$$

We are now in a position to provide the definition of a martingale for a stochastic process $\{M(t)\}_{t\geq 0}$ in continuous time

Definition 3.10. A stochastic process $\{M(t)\}_{t\geq 0}$ in a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is a martingale in relation to a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ if

(i) $\{M(t)\}_{t\geq 0}$ is adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$.

(ii) $\{M(t)\}_{t\geq 0}$ is integrable, that is, $E[|M(t)|] < \infty$

(iii) For any t and s with $0\leq s < t \leq T$

$$E[M(t)/\mathcal{F}_s] = M(s) \text{ almost surely}.$$

In an analogous way follows the definition of a supermartingale and a submartingale.

Definition 3.11. A stochastic process $\{M(t)\}_{t\geq 0}$ in a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is a supermartingale (submartingale) in relation to a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ if

(i). $\{M(t)\}_{t\geq 0}$ is adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$

(ii) $\{M(t)\}_{t\geq 0}$ is integrable.

(iii) For any t and s with $0\leq s < t \leq T$

$$E[M(t)/\mathcal{F}_s] \leq M(s) \text{ and } (E[M(t)/\mathcal{F}_s] \geq M(s)) \text{ almost surely}.$$

Result. A supermartingale $\{M(t)\}_{t\geq 0}$, $0\leq t\leq T$, is a martingale if and only if $E[M(T)] = E[M(0)]$.

A special role in the theory of integration is played by squared integrable martingales.

Definition 3.12. A random variable X is square integrable if $E(X^2) < \infty$. A process $X(t)$ on the time interval $[0,T]$, where T can be infinite, is square integrable if

$$\sup_{t\in[0,T]} E[X^2] < \infty .$$

Another basic class of martingales for which useful theorems in Finance hold is the class of uniformly integrable martingales.

It helps to get some insight in the definition that follows if we have in mind that X is an integrable random variable if and only if

$$\lim_{n\to\infty} E[\,|X|\,I(|X|>n)] = 0$$

Definition 3.13. A process $\{X(t)\}_{t\geq 0}$ is called uniformly integrable if

$$E[\,|X(t)|\,I(|X(t)|>n)]$$

converges to zero as $n\to\infty$ uniformly in t, that is

$$\lim_{n\to\infty} \sup_{t} E(|X(t)|\,I(|X(t)|>n)) = 0$$

where the supremum is over $[0,T]$ in the case of a finite time interval and $[0,\infty)$ if the process is considered on $0\leq t<\infty$.

We now provide in the form of two theorem sufficient conditions for uniform integrability.

Theorem 3.2.: Let $\{X(t)\}_{t\geq 0}$ be a stochastic process and let that there exists an integrable random variable Y such that

$$|X(t)| < Y \quad \text{and} \quad E(Y) < \infty$$

then $\{X(t)\}_{t\geq 0}$ is uniformly integrable.

Proof: Exercise

Theorem 3.3: If the stochastic process $\{X(t)\}_{t\geq 0}$ is square integrable, then it is uniformly integrable.

The following theorem provides a way of constructing uniform integrable martingales

Theorem 3.4: (Doob's Levy's martingale). Let Y be an integrable random variable and define

$$M(t) = E(Y\,|\,\mathcal{F}_t),$$

then $M(t)$ is a uniformly integrable martingale

Example 3.3. Although we will define later on Brownian motion and Itô's stochastic integral it is worth mentioning at the present stage that.

a) The stochastic process $\{B(t)\}_{t\geq 0}$ known as Brownian motion is a square integrable martingale, consequently is also a uniformly integrable martingale. Similarly, $B^2(t)-t$ is a square integrable martingale. Both stochastic process are not square integrable at $T=\infty$.

b) Let $f(x)$ be a bounded and continuous function on $\mathbb{R}$ the Itô integrals

$$\int_0^t f(B(s))\,dB(s) \quad \text{and} \quad \int_0^t f(s)\,dB(s)$$

are square integrable martingales.

We will now provide some important theorems without proof which will be useful in what follows.

Theorem 3.5 (Martingale Convergence Theorem). If $M(t)$, $0\leq t<\infty$, is an integrable martingale (supermartingale or submartingale), then there exists an almost sure limit $\lim_{t\to\infty} M(t)=X$ and X is an integrable random variable.

A corollary of the above theorem is the following

Corollary 3.6

1. Uniformly integrable martingales converge almost surely.
2. Square integrable martingales converge almost surely.
3. Positive martingales converge almost surely.
4. Submartingales bounded from above (negative) converge almost surely.
5. Supermartingales bounded from below (positive) converge almost surely.

The next three theorems belong to the class of optional stopping theorems.

Theorem 3.6 If $\{M(t)\}_{t\geq 0}$ is a martingale and τ is a stopping time, then the stopped process $M(\tau\wedge t)$ is a martingale. In addition

$$E[M(\tau\wedge t)] = E[M(0)].$$

Theorem 3.7. Let $M(t)$ be a martingale.

a) If $\tau \le K < \infty$ is a bounded stopping time then $E[M(\tau)] = E[M(0)]$.

b) If $M(t)$ is uniformly integrable, then for any stopping time τ, $E[M(\tau)] = E[M(0)]$.

Theorem 3.8. Let $M(\tau)$ be a martingale and τ a finite stopping time. If $E[|M(\tau)|] < \infty$, and

$$\lim_{t\to\infty} E[M(t)\, I(\tau > t)] = 0$$

then

$$E[M(\tau)] = E[M(0)].$$

The converse in some sence is also true as the next Theorem states.

Theorem 3.9. Let $\{X(t)\}_{t\geq 0}$, $t \geq 0$, be such that any bounded stopping time τ, $X(\tau)$ is integrible and $E[X(\tau)] = E[X(0)]$. Then $X(t), t\geq 0$ is a martingale.

The following result is sometimes known as the Optional Sampling Theorem

Theorem 3.10. (Optional Sampling Theorem). Let $M(t)$ be a uniformly integrable martingale, and $\tau_1 \le \tau_2 \le \infty$ two stopping times. Then

$$E[M(\tau_2)/\mathcal{F}_{\tau_1}] = M(\tau_1) \text{ almost surely}.$$

Example 3.4

Let the stochastic process $\{X_n\}_{n=0}^{\infty}$ be a simple random walk with two absorbing barriers at 0 and C. In this respect assume that Z_i $(i=1,2,\dots)$ are independent identically distributed random variables and let

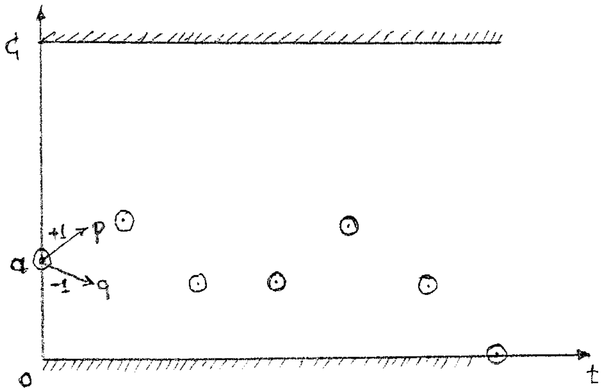
$$\mathbb{P}(Z_i = 1) = p \quad \text{and} \quad \mathbb{P}(Z_i = -1) = q$$

Let that $X_0 = a$ then it is known that

$$X_n = X_0 + Z_1 + Z_2 + \cdots + Z_n.$$

It is of interest to find the probability of absorption in any of the absorbing barries having in mind (figure 3.1) that X_n is the position of a particle at time starting at $0 < a < C$.

Figure 3.1



This problem is historically known as the gambler's ruin problem since in fact could be interpreted that two players have together the sum of C units of money. Player A has initial capital a units of money. Now $\{X_n\}_{n=0}^{\infty}$ is the capital of player A after n games. Absorption at C means that player A wins the entire capital available in both pockets while absorption at 0 means that player A has lost all available units of money.

Abraham de Morgan proposed the following stochastic process which helps the problem of finding the probability of absorption in either barriers. Let

$$M(n) = (q/p)^{X_n} \qquad n = 0, 1, 2, \dots$$

Then

$$E[M(n+1)/\mathcal{F}_n] = E[(q/p)^{X_n + Z_{n+1}}/\mathcal{F}_n]$$

$$= E[(q/p)^{X_n}/\mathcal{F}_n] \times E[(q/p)^{Z_{n+1}}/\mathcal{F}_n] =$$

$$= (q/p)^{X_n} \cdot [p(q/p) + q(q/p)^{-1}] = (q/p)^{X_n} = M(n), \; n = 0, 1, 2, \dots$$

Hence $\{M(n)\}_{n=0}^{\infty}$ is a martingale. Now it is easy to see that the probability of absorption is one. Thus if we define by p_c the probability of absorption at C then the probability of absorption at 0 is $1 - p_c$.

Now the time $\tau_1 = 0$ could be thought as a trivial stopping time and let T be the time of absorption at either of the absorption barriers. Then it is easy to see that T is also a stopping time for the martingale $\{M(n)\}_{n=0}^{\infty}$.

From theorem 3.10 we get

$$E[M(T)/\mathcal{F}_{T_1}] = M(T_1) = M(0) = (q/p)^a .$$

also

$$E[M(T)/\mathcal{F}_{T_1}] = E[M(T)/X_0 = a] = (q/p)^0 (1-p_c) + (q/p)^c p_c ,$$

hence

$$(q/p)^a = (1-p_c) + (q/p)^c p_c = 1 - p_c\left(1-(q/p)^c\right),$$

thus

$$p_c = \frac{1-(q/p)^a}{1-(q/p)^c} .$$

Once more we remark that the vital point in the solution was the fact that we guessed that $\{(q/p)^{X_n}\}_{n=0}^{\infty}$ is a martingale.

References

1. Grimmet, G and Stirzaker (2001). Probability and Random Process. Third ed. Oxford University Press.
2. Karlin, S. and Taylor H. (1975). A first course in stochastic process. Academic Press.
3. Klebaner, F. (2006). Introduction to Stochastic Calculus with Applications. Second Ed. Imperial College Press.
4. Protter, P. (2003). Stochastic Integration and Differential Equations. Second Ed. Springer.
5. Ross, S. (1996). Stochastic Processes. Wiley.
6. William, D. (2000). Probability with Martingales. Cambridge University Press.

Chapter 4

Brownian Motion
Stochastic Differential Equations.

4.1. Introduction.

In this chapter we will study the Brownian motion the history of which was presented in chapter 1. Brownian motion had a great impact in the applications of all the branches of physical sciences and also in many branches of social sciences. As a mathematical entity it has an unusual beauty. This stochastic process has so many properties that resembles an ingenuity which is nearer a Law of Nature than a product of the human mind. Any researcher in Mathematics who has endured succesfully for many years, he sences very strongly that Nature knows more advanced and more deep Mathematics and the human mind simply approximates locally the operation in applications of a small portion of the "Mathematics of Nature" as we may call them.

4.2. The genesis of the Brownian motion.

Let an interval $[0,T]$ and a partition of this interval

$$0=t_0<t_1<\dots<t_n=T$$

with $t_k=t_0+k\Delta t$ for $k=1,2,\dots,n$ with the step Δt a very small number.

Consider now the stochastic process $\{X_t\}_{t=0}^{\infty}$ in discrete time which expresses the price of an asset and let $\Delta h>0$ be a very small interval $\in \mathcal{B}(\mathbb{R})$.

Assume that the random variables

$$(X_{t_1}-X_{t_0}),\ (X_{t_2}-X_{t_1}),\ \dots,\ (X_{t_n}-X_{t_{n-1}})$$

are independent and identically distributed random variables with probability distribution:

$$X_{t_k} - X_{t_{k-1}} = \begin{cases} +\Delta h & \text{with probability } p \\ -\Delta h & \text{with probability } 1-p. \end{cases}$$

Let $X_0 = 0$ with no loss of generality then the stochastic process $\{X_t\}_{t\in[0,T]}$ at time T will take the value

$$X_T = (X_{t_1} - X_{t_0}) + (X_{t_2} - X_{t_1}) + \cdots + (X_{t_n} - X_{t_{n-1}})$$

Let us denote by

$$\Delta X(t_k) = X_{t_k} - X_{t_{k-1}} \quad \text{for any } k = 1, 2, \ldots$$

the ΔX expresses the increment of the stochastic process $\{X_t\}_{t\in T}$ and since the random variables

$$X_{t_1} - X_{t_0},\ X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$$

are independent and identically distributed then ΔX has the same expected value and variance at any point in time. For this reason there is a meaning in writting $E[\Delta X]$ and $\text{var}[\Delta X]$. We easily see that

$$E[\Delta X] = p\Delta h - (1-p)\Delta h = (2p-1)\Delta h$$

$$E[(\Delta X)^2] = p(\Delta h)^2 + (1-p)(-\Delta h)^2 = (\Delta h)^2$$

and then

$$\text{var}[\Delta X] = E[(\Delta X)^2] - (E[\Delta X])^2 = 4p(1-p)(\Delta h)^2$$

We will now find

$$E[X_T] = E[(X_{t_1} - X_{t_0}) + (X_{t_2} - X_{t_1}) + \cdots + (X_{t_n} - X_{t_{n-1}})]$$

$$= E[X_{t_1} - X_{t_0}] + E[X_{t_2} - X_{t_1}] + \cdots + E[X_{t_n} - X_{t_{n-1}}]$$

$$= n(2p-1)\Delta h = (\text{since } T = n\Delta t)$$

$$= T(2p-1)\frac{\Delta h}{\Delta t}$$

Similarly since the $\Delta X(t_k)$, $k = 0, 1, 2, \ldots$ are indepent we get

$$\text{var}[X_T] = \text{var}[(X_{t_1} - X_{t_0}) + (X_{t_2} - X_{t_1}) + \cdots + (X_{t_n} - X_{t_{n-1}})]$$

$$= \text{var}[X_{t_1} - X_{t_0}] + \text{var}[X_{t_2} - X_{t_1}] + \cdots + \text{var}[X_{t_n} - X_{t_{n-1}}]$$

$$= 4npq(\Delta h)^2 = 4Tp(1-p)\frac{(\Delta h)^2}{\Delta t}$$

In a partition of time and space in order to make more apparent the shape of the realization of the stochastic process we should choose the partition of time to be thiner. In this respect we choose

$$\Delta h = \sigma\sqrt{\Delta t}\,,$$

where σ^2 is the variance of X_T for $T=1$ i.e.

$$var[X_1] = \sigma^2.$$

The reason of choosing $\Delta h = \sigma\sqrt{\Delta t}$ is because in this way $var(X_T)$ depends on the T and not on Δt the step of the partition i.e.

$$E[X_T] = T(2p-1)\frac{\sigma}{\sqrt{\Delta t}}$$

and

$$var[X_T] = 4Tp(1-p)\sigma^2.$$

So far all our steps in our progress were independent of p, thus they will hold if we choose

$$p = \frac{1}{2}\left[1 + \frac{\mu}{\sigma}\sqrt{\Delta t}\right]$$

where μ is chosen to be the expected value of X_T when $T=1$ i.e. $E[X_1]=\mu$.

Now with the above selection of p we have

$$E[X_T] = T\frac{\mu}{\sigma}\sqrt{\Delta t}\,\frac{\sigma}{\sqrt{\Delta t}} = \mu T$$

$$var[X_T] = T\left(1 - \frac{\mu^2}{\sigma^2}\Delta t\right)\sigma^2$$

Finally allowing $\Delta t \to 0$ we move into the continuous time version of the process which is equivalent to $n \to \infty$ and according to the central limit theorem X_T is normally distributed with expected value μT and variance

$$\lim_{\Delta t \to 0} T\left(1 - \frac{\mu^2}{\sigma^2}\Delta t\right)\sigma^2 = \sigma^2 T$$

and thus finally

$$(X_T - X_0) \approx N(\mu T, \sigma^2 T).$$

As always when we have a normal distribution $N(\mu,\sigma^2)$ we are looking for the $N(0,1)$ to represent the entire class. In this case by selecting

$$p = 1/2 \quad \text{and} \quad \Delta h = \sqrt{\Delta t}$$

then we get

$$E[X_T] = 0 \quad \text{and} \quad var[X_T] = T$$

and due to the central limit theorem

$$X_T - X_0 \approx N(0, T).$$

A stochastic process as the one above with $p = 1/2$ and $\Delta h = \sqrt{\Delta t}$ is called the Wiener process. Wiener process has many ways of defining it but we started with the most natural one as we believe i.e. as the limit of a symmetric random walk with $\Delta t \to 0$.

We denote by $\{W_t\}_{t \geq 0}$ the Wiener stochastic process then since we have assumed that $\Delta h = \sqrt{\Delta t}$ then we get

$$(dW_t)^2 = dt$$

and from what we have so far established we get

$$E[dW_t] = 0 \quad \text{and} \quad var[dW_t] = E[(dW_t)^2] = dt$$

Since in addition

$$\lim_{\Delta t \to 0} \frac{\Delta h}{\Delta t} = \infty$$

it is evident that we can not define the derivative of the Wiener process.

Wiener founded mathematically the process which is also known as the Brownian motion from the botanologyst Brown who has observed the motion in a liquid.

We will now proceed with two equivalent definitions of the Brownian motion

<u>Definition 4.1</u>: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{B(t)\}_{t \geq 0}$ a stochastic process with real values. Then $B(t)$ is a Brownian motion if it satisfies the following conditions.

a) $B(0)(\omega) = 0 \quad \forall \omega \in \Omega$.

b) The mapping $t \to B(\omega)$ is a continuous function of $t \in \mathbb{R}^+$ for every $\omega \in \Omega$.

c) For every $t, \Delta t \geq 0$ the increment

$$B(t+\Delta t) - B(t)$$

is independent from the random variables

$$\{B(s) : 0 \leq s \leq t\}$$

and is normally distributed with expected value 0 and variance Δt, i.e.

$$E[B(t+\Delta t) - B(t)] = 0 \quad \text{and} \quad E[(B(t+\Delta t) - B(t))^2] = \Delta t .$$

▲

An equivalent definition for the Brownian motion is the following.

<u>Definition 4.2</u>: Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\{B(t)\}_{t\geq 0}$ be a stochastic process with real values. Then $\{B(t)\}_{t\geq 0}$ is a Brownian motion if it satisfies the following conditions.

1. $B(0) = 0$.
2. $B(t)$ is a continuous function of t.
3. The stochastic process $\{B(t)\}_{t\geq 0}$ has independent increments that are normally distributed.

   In other words if we have a partition of the interval $[0,T]$

   $$0 \leq t_0 < t_1 < t_2 < \cdots < t_n = T.$$

   and we denote by

   $$X_1 = B(t_1) - B(t_0),\ X_2 = B(t_2) - B(t_1), \ldots, X_n = B(t_n) - B(t_{n-1})$$

   then

   a) $X_1, X_2, \ldots, X_n$ are independent random variables.

   b) $E[X_k] = 0$ for every $k = 1, 2, \ldots, n$.

   c) $var[X_k] = t_k - t_{k-1}$

▲

We will now prove that the Brownian motion is a martingale. In fact as we have seen it is a square integrable martingale and also a uniformly integrable martingale.

In this respect let us denote by $\mathcal{B}_t$ the σ-algebra generated from by the Brownian motion $\{B(t)\}_{t\geq 0}$ for $s\leq t$ i.e.

$$\mathcal{B}_t = \sigma\{B(s) \text{ for all } s\leq t\}.$$

We will now prove that the Brownian motion is a martingale in relation to the above filtration.

<u>Theorem 4.1</u>. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{B(t)\}_{t\geq 0}$ be a Brownian motion, then $\{B(t)\}_{t\geq 0}$ is a martingale in relation to $\mathcal{B}_t$

<u>Proof</u>:

a) We should prove that $B(t)$ is $\mathcal{B}_t$-measurable in Ω. This is true by definition since $\mathcal{B}_t$ is the σ-algebra generated by $B(t)$ for $s\leq t$.

b) $\{B(t)\}_{t\geq 0}$ is an integrable stochastic process. It is easy to show as an exercise that $E[|B(t)|]<\infty$.

c) We have to prove that

$$E[B(t+r)/\mathcal{B}_t] = B(t) \text{ almost surely with } r>0,\ t+r\in\mathbb{R}^+$$

From the definition of the Brownian motion we get that

$$E[B(t+r)-B(t)/\mathcal{B}_t] = 0 \quad \text{almost surely}$$

or equivalently

$$E[B(t+r)\,|\,\mathcal{B}_t] = E[B(t)\,|\,\mathcal{B}_t] =$$

$$= (\text{according to definition 4.1 } B(t) \text{ is independent of } \mathcal{B}_t)$$

$$= B(t)$$

▲

The following proposition is a useful exercise

<u>Proposition 4.1</u> Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{B(t)\}_{t\geq 0}$ is a stochastic process which is a Brownia motion then

1) The stochastic process $\{-B(t)\}_{t\geq 0}$ is a Brownian motion.

2) For every $r>0$ the stochastic process

$$\{B(t+r)-B(t)\}_{t,r\in\mathbb{R}^+}$$

is a Brownian motion

3) For every $r \neq 0$ the stochastic process

$$\{r B(t/r^2)\}_{t \in \mathbb{R}^+}$$

is a Brownian motion

4) The stochastic process $\{\hat{B}(t)\}_{t \in \mathbb{R}^+}$ defined as

$$\hat{B}(0) = 0 \quad \text{and} \quad \hat{B}(t) = t B(1/t) \text{ for } t \in \mathbb{R}^+$$

is a Brownian motion.

We will now prove a theorem which will eventually lead us to the important Levy theorem.

<u>Theorem 4.2</u> Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{B(t)\}_{t \in \mathbb{R}^+}$ a Brownian motion. Then the stochastic process

$$[B(t)]^2 - t$$

is a martingale in relation with $\mathcal{B}_t$.

<u>Proof</u>. It is immediately apparent that $[B(t)]^2 - t$ is $\mathcal{B}_t$-measurable in Ω for all $t \in \mathbb{R}^+$. Now we will show that $[B(t)]^2 - t$ is integrable. We have that

$$E\left[\left|[B(t)]^2 - t\right|\right] \leq E\left[\,|B(t)|^2 + |t|\,\right]$$

$$\leq E\left[[B(t)]^2\right] + |t|$$

$$\leq (\text{since } B(t) \text{ is integrable for } t \text{ finite})$$

$$< \infty.$$

It remains to prove that

$$E\left[[B(t)]^2 - t / \mathcal{B}_s\right] = [B(s)]^2 - s \quad \text{for all } s \leq t \in \mathbb{R}^+.$$

Since $\{B(t)\}_{t \in \mathbb{R}^+}$ is a Brownian motion we have that

$$E\left[[B(t) - B(s)]^2 / \mathcal{B}_s\right] = t - s$$

From this relation we get

$$E\left[[B(t) - B(s)]^2 / \mathcal{B}_s\right] = E\left[[B(t)]^2 - 2[B(t)][B(s)] + [B(s)]^2 / \mathcal{B}_s\right] =$$

$$= E\left[[B(t)]^2 / \mathcal{B}_s\right] - 2E\left[[B(t)B(s)] / \mathcal{B}_s\right] + E\left[[B(s)]^2 / \mathcal{B}_s\right] =$$

$$= E\left[[B(t)]^2 / \mathcal{B}_s\right] - 2B(s)E\left[[B(t)] / \mathcal{B}_s\right] + [B(s)]^2$$

$$= E\left[[B(t)]^2 / \mathcal{B}_s\right] - [B(s)]^2$$

Thus we have that

$$E[[B(t)]^2/\mathcal{B}_s] - [B(s)]^2 = t-s$$

From which we get that

$$E[[B(t)]^2/\mathcal{B}_s] - t = [B(s)]^2 - s$$

and finally

$$E[[B(t)]^2 - t/\mathcal{B}_s] = [B(s)]^2 - s$$

$\blacktriangle$

<u>Theorem 4.3</u> (Levy). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and a stochastic process $\{X(t)\}_{t\in\mathbb{R}^+}$ which is such that the mapping $t \to X(t)(\omega)$ be continuous for all $\omega\in\Omega$. Let also the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$ in Ω. Then if the stochastic process

$$[X(t)]^2 - t$$

is a martingale in relation to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$ then the stochastic process $\{X(t)\}_{t\geq 0}$ is a Brownian motion.

We will now provide a theorem in which we will prove that an exponential function of a Brownian motion which is useful as a model for asset pricing is a martingale.

<u>Theorem 4.4</u>. Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{B(t)\}_{t\in\mathbb{R}^+}$ a Brownian motion. Then the stochastic process

$$\exp\{\lambda B(t) - \tfrac{1}{2}\lambda^2 t\} \quad , \lambda\in\mathbb{R}$$

is a martingale in relation with the filtration $\mathcal{B}_t$.

<u>Proof</u>. a) The stochastic process

$$\exp\{\lambda B(t) - \tfrac{1}{2}\lambda^2 t\} \text{ is } \mathcal{B}_t\text{-measurable in } \Omega \text{ for all } t\in\mathbb{R}^+$$

b) It is easy to prove that

$$E[|\exp\{\lambda B(t) - \tfrac{1}{2}\lambda^2 t\}|] < \infty \quad \text{for all finite } t\in\mathbb{R}^+$$

c) We have to prove that

$$E[\exp\{\lambda B(t) - \tfrac{1}{2}\lambda^2 t\}/\mathcal{B}_s] = \exp\{\lambda B(s) - \tfrac{1}{2}\lambda^2 s\} \text{ a.s.}$$

Since $\{B(t)\}_{t\geq 0}$ is a Brownian motion we know that

$$B(t) - B(s) = \mathcal{N}(0, t-s).$$

Since $B(t)-B(s)$ is normally distributed with expected value 0 and variance $t-s$ we immediately know that for $\lambda \in \mathbb{R}$

$$M_{B(t)-B(s)}(\lambda) = E\left[\exp\{\lambda B(t)-B(s)\} \mid \mathcal{B}_s\right] = \exp\left\{(t-s)\frac{\lambda^2}{2}\right\}$$

from which we get

$$E\left[\exp\{\lambda B(t)\}\exp\{-\lambda B(s)\} / \mathcal{B}_s\right] = \exp\{\tfrac{1}{2}\lambda^2 t\}\exp\{-\tfrac{1}{2}\lambda^2 s\}.$$

Now due to the independence property of the Brownian motion at different time points we get

$$E\left[\exp\{\lambda B(t)\} / \mathcal{B}_s\right] \cdot E\left[\exp\{-\lambda B(s)\} / \mathcal{B}_s\right] = \exp\{\tfrac{1}{2}\lambda^2 t\}\exp\{-\tfrac{1}{2}\lambda^2 s\},$$

from which it is apparent that

$$E\left[\exp\{\lambda B(t)\} / \mathcal{B}_s\right]\exp\{-\lambda B(s)\} = \exp\{\tfrac{1}{2}\lambda^2 t\}\exp\{-\tfrac{1}{2}\lambda^2 s\}$$

and finally

$$E\left[\exp\{\lambda B(t) - \tfrac{1}{2}\lambda^2 t\} / \mathcal{B}_s\right] = \exp\{\lambda B(s) - \tfrac{1}{2}\lambda^2 s\}$$

▲

We will now establish that a Brownian process is a Markov property and discuss its transition density.

We first start by refreshing the definition of a Markov process.

<u>Definition 4.3</u>: Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{F}_t$ a filtration of $\mathcal{F}$. Let $X(t)$ be an adapted stochastic process to the filtration $\mathcal{F}_t$ in the interval $0 \le t \le T$ where T is a fixed positive number.

Assume that for all f- Borel measurable function f, there is another Borel - measurable function g such that

$$E\left[f(X(t)) / \mathcal{F}_s\right] = g(X(s)).$$

We will now provide a Lemma named the independence Lemma. We will not supply a proof of this Lemma since it requires some measure theoretic ideas beyond the scope of these notes.

<u>Lemma 4.1</u>. Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Suppose that the random variables $X_1, X_2, \dots, X_m$ are $\mathcal{G}$-measurable and the random variable $Y_1, Y_2, \dots, Y_n$ are independent $\mathcal{G}$. Let $x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n$ be any set of values of the above random variables. Let also

$$f(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n)$$

be any function of the x's and y's. Define by

$$g(x_1, x_2, \dots, x_m) = \mathbf{E}\left[f(x_1, x_2, \dots, x_m, Y_1, Y_2, \dots, Y_n)\right]$$

Then

$$E\left[f(X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n) \mid \mathcal{G}\right] = g(X_1, X_2, \dots, X_m)$$

<u>Remark 4.1</u>. The methodology which steems out of Lemma 4.1 is that since the random variables $Y_1, Y_2, \dots, Y_n$ are independent of $\mathcal{G}$ then we should by keeping $X_1, X_2, \dots, X_n$ constants integrate them out without regard to the information in $\mathcal{G}$.

<u>Example 4.1</u>. Suppose we want to estimate some function $f(X, Y)$ of some random variables X, Y based on knowledge of X.

We cannot use the Independence Lemma directly because X and Y are not independent. However, we can write Y as

$$Y = \frac{\rho \sigma^2}{\sigma_1} X + W,$$

where W is independent of X and is normally distributed with mean μ_3 and variance σ_3.

The Independence Lemma tells us how to compute $E[f(X,Y) \mid X]$. We should take the expectation replacing X by any possible value x, i.e.

$$g(x) = \mathbf{E}\left[f\left(x, \frac{\rho \sigma_1}{\sigma_1} x + W\right)\right]$$

$$= \frac{1}{\sigma_3 \sqrt{2\pi}} \int_{-\infty}^{\infty} f\left(x, \frac{\rho \sigma_1}{\sigma_2} x + w\right) dw$$

Then

$$E\left[f(X,Y) \mid X\right] = \frac{1}{\sigma_3 \sqrt{2\pi}} \int_{-\infty}^{\infty} f\left(X, \frac{\rho \sigma_1}{\sigma_2} X + w\right) dw.$$

We are now in a position to proove the following Theorem.

<u>Theorem 4.5</u> : Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $B(t)$ a Brownian motion and let $\mathcal{B}_s$ be a filtration for this Brownian motion. Then $B(t)$ is a Markov process in continuous time and continuous state space.

Proof: Let $0 \leq s \leq t$ and f a Borel measurable function. Then we have to show that there is a function g which is Borel measurable such that

$$E\left[f(B(t)) / \mathcal{B}_s \right] = g(B(s)).$$

In this respect we have that

$$E\left[f(B(t)) / \mathcal{B}_s \right] = E\left[f((B(t) - B(s)) + B(s)) / \mathcal{B}_s \right]$$

Now $B(t) - B(s)$ is independent of $\mathcal{B}_s$ and $B(s)$ is $\mathcal{B}_s$-measurable and that satisfies the conditions of Lemma 4.1. In addition it is known that

$$B(t) - B(s) \approx N(0, t-s)$$

Now let x be any value of $B(s)$ and y any value of $B(t) - B(s)$. Then applying Lemma 4.1 we have

$$g(x) = E\left[f(B(t) - B(s) + x \right]$$
$$= \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{\infty} f(y+x)\, e^{-\frac{y^2}{2(t-s)}}\, dy$$

Then by Lemma 4.1

$$E\left[f(B(t)) / \mathcal{B}_s \right] = g(B(s)),$$

with

$$g(B(s)) = \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{\infty} f(y + B(s))\, e^{-\frac{y^2}{2(t-s)}}\, dy.$$

▲

Let us define by $p(t-s, B(s), z)$ be the transition density for the Brownian motion $B(t)$, where $B(s)$ is the position (value) of the Brownian motion at time s and z the increment $B(t) - B(s)$ in the time interval $t-s$. Then we define the transition density for the Brownian motion to be

$$p(t-s, B(s), z) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(z-B(s))^2}{2(t-s)}}$$

Then it is easy to see that

$$E[f(B(t))/\mathcal{F}_s] = g(W(s)) = \int_{-\infty}^{\infty} f(z)\, p(t-s, B(s), z)\, dz.$$

This equation expresses the fact that given the σ-algebra $\mathcal{F}_s$ which contains the information on the value of $B(s)$ then the increment $B(t)-B(s)$ in the time interval $t-s$ is normally distributed with mean zero and variance $t-s$.

4.3. A Class of stochastic differential equations.

We now use the Wiener process as the basis to define a large class of models appropriate for asset pricing. These models are given in a differential form and we name them stochastic differential equations. It doesn't exist a general method of solution for all stochastic differential equations. In the next chapter with the introduction of Itô-Doeblin formula we will solve some of them as applications of the more famous formula in Stochastic Calculus and Finance. The general form of stochastic differential equations appropriate as models for asset pricing is the following.

$$dB(t) = \mu(B(t),t)\,dt + \sigma(B(t),t)\,dW(t),$$

where $\mu(B(t),t)$ is called the drift parameter and is defined to be

$$\mu(B(t),t) = \lim_{h \downarrow 0} \frac{1}{h} E\left[B(t+h) - B(t)/\mathcal{B}_t\right],$$

and $\sigma(B(t),t)$ is called the diffusion parameter and is defined to be

$$\sigma^2(B(t),t) = \lim_{h\downarrow 0} \frac{1}{h} E\left[\{B(t+h) - B(t)\}^2/\mathcal{B}_t\right]$$

The part of the equation which is multiplied by the diffusion parameter $dW(t)$ is the one which brings the element of randomness into the equation since $dW(t)$ is an infinitesimal increment of the Wiener process

a) The geometric Brownian motion

The geometric Brownian motion is the most oftenly used model for asset pricing. In this case

$$\mu(B(t),t) = \mu B(t) \quad \text{and} \quad \sigma(B(t),t) = \sigma B(t),$$

and thus the stochastic differential equation of the geometric Brownian motion is given by

$$dB(t) = \mu B(t)\,dt + \sigma B(t)\,dW(t)$$

An equivalent definition of the geometric Brownian motion is the following.

Definition 4.4. Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{\tilde{B}(t)\}_{t\geq 0}$ a stochastic process with real values. Then $\{\tilde{B}(t)\}_{t\geq 0}$ is a geometric Brownian motion if the following conditions are satisfied:

a) $\tilde{B}(0)(\omega) = 0 \quad \forall\, \omega \in \Omega$.

b) The mapping $t \to \tilde{B}(t)(\omega)$ is a continuous function of $t \in \mathbb{R}^+$ for every $\omega \in \Omega$.

c) For every $t, \Delta t \geq$ the increment's ratio

$$\frac{\tilde{B}(t+\Delta t)}{\tilde{B}(t)}$$

is independent of the random variables

$$\{B(s) : 0 \leq s \leq t\}$$

and is lognormally distributed with mean $\mu \Delta t$ and variance $\sigma^2 \Delta t$.

In other words the increment's ratio

$$\log \frac{\tilde{B}(t+\Delta t)}{\tilde{B}(t)}$$

is normally distributed $\approx N(\mu t, \sigma^2 \Delta t)$. ▲

We will find a close analytic solution for the geometrical Brownian motion in the next chapter.

b) The square root Brownian motion.

All too often in practice due to the kind of the asset we want to follow the evolution of its price the following stochastic differential equation is is appropriate as a model

$$dB(t) = \mu B(t)dt + \sigma\sqrt{B(t)}\, dW(t), \; t \in \mathbb{R}^+.$$

The stochastic process which is the solution of the above stochastic differential equation is called the square root Brownian motion. It is apparent that the drift parts of the geometric Brownian motion and the square root Brownian motion are the same and they only differ in their diffusion parameters. It is apparent also that the diffusion parameter of the square root Brownian motion is less than that of geometric Brownian motion. Thus in any realization of the square root Brownian motion the volatility will almost surely be less than the corresponding volatility of the geometric Brownian motion.

c) Mean Reverting Brownian motion

There are assets that their price seems to revert towards a "normal" value. This is a result of the dynamics present in the market that when the price of the asset exceeds the "normal" value after an interval of time it reverts it very near the "normal" value. The analoguous responce is observed when the price of the asset becomes smaller than the "normal" value.
An appropriate model for an asset the price of which presents this physical characteristic is given by the following stochastic differential equation

$$dB(t) = \lambda(\mu - B(t))\,dt + \sigma B(t)\, dW(t)$$

In this stochastic differential equation μ is the "normal" value of the asset and λ is the rate of reverting of the price of the asset.
The stochastic process $\{B(t)\}_{t\geq 0}$ which is the solution of the above stochastic differential equation is called the mean reverting Brownian motion.
In the next chapter we will provide the solution of the present stochastic differential equation and will find its mean and variance. In addition we will discuss their asymptotic behaviour as $t\to\infty$.

d) Other forms of mean reverting Brownian motions.

Another form of mean reverting Brownian motion with the physical characteristic of having smaller variance is provided as the solution of the following stochastic differential equation.

$$dB(t) = \lambda(\mu - B(t))dt + \sigma\sqrt{B(t)}\,dW(t)$$

Another form of mean reverting Brownian motion with the physical characteristic of having a larger rate of reverting the larger the price of the asset is provided as the solution of the following differential equation

$$dB(t) = \lambda B(t)(\mu - B(t))dt + \sigma\sqrt{B(t)}\,dW(t).$$

The last form of reverting process is the known as the Ornstein-Uhlenbeck stochastic process which is the solution of the following stochastic differential equation

$$dB(t) = -\mu B(t)dt + \sigma dW(t).$$

e) <u>Stochastic differential equations with stochastic volatility.</u>

All the above stochastic processes are solutions of stochastic differential equations with variance or volatility as is called in the economic cycles constant. However there exist a class of stochastic processes in the literature which are solutions of stochastic differential equations which have stochastic volatility i.e volatility which is a stochastic process given as a solution of a stochastic differential equation. References to such work are Scott (1987), Wiggins (1987), Hull and White (1987, 1988), Stein and Stein (1991) and Heston (1993).

The basic stochastic process used in these studies was the geometrical Brownian motion with stochastic volatility in its stochastic differential equation form i.e.

$$d\tilde{B}(t) = \mu\tilde{B}(t)dt + \sigma(t)\tilde{B}(t)dW(t)$$

The stochastic volatility $\sigma(t)$ in this case is given in the form of a stochastic differential equation that models the physical characteristics of the volatility of the asset. The stochastic differential equations that are usually used as models for stochastic volatility are the following:

1. $$d\sigma(t) = \beta\sigma(t)dt + a\sigma(t)dW(t),$$
where obviously $\sigma(t)$ is a geometric Brownian motion.

2. In this case the stochastic volatility has the form of a mean reverting process. i.e it is the solution of the stochastic differential equation
$$d\sigma(t) = \sigma(t)\left(a - \beta\sigma(t)\right)dt + \gamma\sigma(t)dW(t)$$

3. In this case the following stochastic differential equation is proposed as a model for the volatility
$$d\sigma(t) = \left(\frac{\delta}{\sigma(t)} - \beta\sigma(t)\right)dt + \gamma dW(t).$$

4.5. Multi-dimensional Brownian motion

So far we have studied the one-dimesional Brownian motion. However Brownian motion can be defined also in more than one dimension.

Definition 4.5: An n-dimensions stochastic process
$$\underset{\sim}{B}(t) = \left(B_1(t), B_2(t), \ldots, B_n(t)\right)$$
is an n-dimentional stochastic process if the following conditions are satisfied.

a) Every one of the stochastic processes
$$\{B_k(t)\}_{t\geq 0}, k = 1, 2, \ldots, n \text{ is a one dimensional stochastic process}$$

b) If $i \neq j$ then the stochastic processes
$$\{B_i(t)\}_{t\geq 0} \text{ and } \{B_j(t)\} \text{ are independent.}$$

c) Let $\underset{\sim}{\mathcal{B}}_t$ be the σ-algebra generated by the family of random variables
$$\underset{\sim}{\mathcal{B}}_t = \sigma\left(\underset{\sim}{B}(s) \text{ for all } s \leq t\right)$$
then if if we consider a time interval $[t, T]$ and a partition of this interval
$$t = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_n = T$$
then the increments
$$\underset{\sim}{B}(t_1) - \underset{\sim}{B}(t_0),\ \underset{\sim}{B}(t_2) - \underset{\sim}{B}(t_1),\ \ldots,\ \underset{\sim}{B}(t_n) - \underset{\sim}{B}(t_{n-1})$$

are independent of the σ-algebra $\mathcal{B}_t$.

We will now prove the following useful theorem.

<u>Theorem</u>. Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{\underset{\sim}{B}(t)\}_{t\geq 0}$ an n-dimensional Brownian motion

$$\underset{\sim}{B}(t) = \left(B_1(t), B_2(t), \ldots, B_n(t)\right).$$

If $i \neq j$ then we have that

$$dB_i(t)\, dB_j(t).$$

<u>Proof</u>: Let an interval $[0,T]$ and let a partition of this interval

$$0 \leq t_0 \leq t_1 \leq \cdots \leq t_n = T.$$

We define the function $V_n(B)$ to be

$$V_n(B) = \sum_{k=0}^{n-1} \left[B_i(t_{k+1}) - B_i(t_k)\right]\left[B_j(t_{k+1}) - B_j(t_k)\right]$$

Then it is apparent that as $n \to \infty$

$$\max_k \left[t_{k+1} - t_k\right] \to 0 \quad , k = 0, 1, \ldots, n$$

and

$$\lim_{n\to\infty} V_n(B) = dB_i(t)\, dB_j(t) \quad \text{with } t \in [0,T].$$

We have that

$$\begin{aligned}
E\left[V_n(B)\right] &= E\left[\sum_{k=0}^{n-1} \left[B_i(t_{k+1}) - B_i(t_k)\right]\left[B_j(t_{k+1}) - B_j(t_k)\right]\right] \\
&= \sum_{k=0}^{n-1} E\left\{\left[B_i(t_{k+1}) - B_i(t_k)\right]\left[B_j(t_{k+1}) - B_j(t_k)\right]\right. \\
&= \sum_{k=0}^{n-1} E\left[B_i(t_{k+1}) - B_i(t_k)\right] E\left[B_j(t_{k+1}) - B_j(t_k)\right] \\
&= \Big(\text{since } B_i(t), B_j(t) \text{ are Brownian motions we have} \quad E\left[B_i(t_{k+1}) - B_i(t_k)\right] = 0\Big) \\
&= 0.
\end{aligned}$$

Thus we arrive at our first result that $E\left[V_n(B)\right] = 0$. It remains to show that

$$\lim_{n\to\infty} \operatorname{var}\left[V_n(B)\right] = 0.$$

$$E\{[V_n(B)]^2\} = E\Big\{\sum_{k=0}^{n-1}[B_i(t_{k+1})-B_i(t_k)]^2 + [B_j(t_{k+1})-B_j(t_k)]^2 +$$

$$2\sum_{m<k}^{n-1}[B_i(t_{m+1})-B_i(t_m)][B_j(t_{m+1})-B_j(t_m)]$$

$$[B_i(t_{k+1})-B_i(t_k)][B_j(t_{k+1})-B_j(t_k)]\Big\} =$$

$=$ (since the various increments are independent and since $B_i(t), B_j(t)$ are Brownian motion we have that

$$E[B_i(t_{k+1})-B_i(t_k)]^2 = t_{k+1}-t_k) =$$

$$= \sum_{k=0}^{n-1}(t_{k+1}-t_k)^2.$$

Since we have that

$$\begin{aligned}
var[V_n(B)] &= E\{[V_n(B)]^2\} - \{E[V_n(B)]\}^2 \\
&= \sum_{k=0}^{n-1}(t_{k+1}-t_k)^2 \le \\
&\le \max_k (t_{k+1}-t_k)\sum_{k=0}^{n}(t_{k+1}-t_k) \\
&\le \max_k (t_{k+1}-t_k)\, T
\end{aligned}$$

we get that as $n\to\infty$ then

$$\max_k (t_{k+1}-t) \to 0 \quad \text{and} \quad var[V_n(B)] \to 0$$

and consequently

$$dB_i(t)\, dB_j(t) = 0$$

▲

REFERENCES.

1. Karatzas, I and Shreve S.E. (1991). Methods of Mathematical Finance. Springer
2. Karatzas, I and Shreve S.E (1991). Brownian Motion and Stochastic Calculus. Springer
3. Oksendal, B (2005). Stochastic Differential Equations. Sixth edition. Springer.
4. Rogers, L and William, D. (2000). Diffusions, Markov Process and Martingales. Vol. 1 : Foundation. Cambridge University Press.
5. Shreve, S.E. (2004) Stochastic Calculus for Finance II. Continuous-Time Models. Springer.

---

a. Hull, J. and White, A. (1988). An analysis of the bias in option pricing caused by stochastic volatility. Advances in Futures and Options Research. 3, 29-61

b. Henston, S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies. 6, 327-343.

c. Scott, L.O. (1987). Option pricing when the variance changes randomly: Theory estimation and an application. Journal of Financial and Quantitative Analysis. 22, 419-438.

d. Stein, E.M. and Stein, J.C. (1991). Stock price distributions with stochastic volatility: an analytic approach. Review of Financial Studies 4, 727-752.

e. Wiggins, J.B. (1987). Option values under stochastic volatility. Theory and empirical estimates. Journal of Financial Economics.

Chapter 5

Stochastic Calculus

- Itô's Integral.
- Itô - Döeblin formula.
- Applications.

5.1 Introduction.

Let $\{\Delta(t)\}_{t\geq 0}$ be a adapted to a filtration $\{\mathcal{F}(t)\}_{t\geq 0}$ stochastic process, which will usually in Finance will be the position we take in an asset at time t, which typically depends on the price path at time t. $\mathcal{F}(t)$ will usually be the σ-algebra which at time t contains all the information available to the traders.

The price of an asset as we have seen in the previous chapter is usually modeled by some form of a Brownian motion. Naturally, if $B(t)$ is that Brownian motion it will be adapted to the filtration $\mathcal{F}(t)$. However, recall that increments of the Brownian motion after time t are independent of $\mathcal{F}(t)$ and since $\Delta(t)$ is $\mathcal{F}(t)$ - measurable, it must also be independent of the increment of these future Brownian increments. Positions we take on assets may depend on the price history of assets, but they must independent of the future increments of the Brownian motion that drives these prices.

If this is the case then part of the return in the interval $[0,T]$ is the integral

$$\int_0^T \Delta(t)\, dW(t).$$

First we review the definitions of the Riemann integral in calculus and the Riemann-Stieltzes integral in advanced calculus.

(a) Riemann Integral.

A bounded function f defined on a finite closed interval $[a,b]$ is called Riemann integrable if the following limit exists

$$\int_a^b f(t)\,dt = \lim_{\|\Pi_n\|\to 0} \sum_{i=1}^{n} f(\tau_i)(t_i - t_{i-1})$$

where $\Pi_n = \{t_0, t_1, \dots, t_n\}$ is a partition of $[a,b]$ with the convention

$$a = t_0 < t_1 < t_2 < \cdots < t_n = b,$$

$\|\Pi_n\| = \max_{1\le i\le n}(t_i - t_{i-1})$ and τ_i is an evaluation point in the interval $[t_{i-1}, t_i]$. If f is a continuous function on $[a,b]$, then it is Riemann integrable. Moreover, it is well known that a bounded function on $[a,b]$ is Riemann integrable if and only if it is continuous almost everywhere with respect to the Lebesgue measure.

(b) Riemann-Stieltjes Integral

Let g be a monotonically increasing function on a finite closed interval $[a,b]$. A bounded function f defined on $[a,b]$ is said to be Riemann-Stielzes integrable with respect to g if the following limit exists

$$\int_a^b f(t)\,dg(t) = \lim_{\|\Pi_n\|} \sum_{i=1}^{n} f(\tau_i)\big(g(t_i) - g(t_{i-1})\big).$$

It is a well-known fact that continuous functions on $[a,b]$ are Riemann-Stietjes integrable with respect to any monotonically increasing function on $[a,b]$.

The problem we face when we are trying to assign a meaning to the Itô integral as a Riemann-Stieltjes integral is that Brownian motion paths cannot be differentiated with respect to time.

Let us now consider the Reimann sum

$$\sum_{i=0}^{n-1} \Delta(t_{i+1}) \left[W(t_{i+1}) - W(t_i)\right]$$

or even since $\Delta(t)$ is $\mathcal{F}_t$-measurable we may equivalenty use the Riemann sum

$$\sum_{i=0}^{n-1} \Delta(t_i) \left[W(t_{i+1}) - W(t_i)\right]$$

However, even so the above Riemann sum contains the random variables $W_{t_{i+1}}$ and consequently is a random variable.

This fact creates the following problems

a) It is not possible to use the limit of the Riemann sum since there are random variables in it.

b) We have thus to use one of the known types of convergence for random variables.

c) Does it always exist such a limit whatever type of convergence we choose?

d) If it does not always exist under what conditions does it exists and what properties has the random variable to which it converges.

Itô is a six pages paper choose the known limit for a sequence of random variables $\{X_n\}_{n=0}^{\infty}$

$$\lim_{n\to\infty} E[X_n - X] = 0 .$$

The formal definition is given below:

Definition of Itô Integral.

Let $\{W_t\}_{t\geq0}$ be a Wiener process and let that it is adopted to the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$. Let also $\{\Delta(t)\}_{t\geq0}$ be a stochastic process adapted to the filtration $\{\mathcal{F}(t)\}_{t\geq0}$. Let also that

(i) $\{\Delta(t)\}_{t\geq0}$ is a square integrable stochastic process in the time interval $[0,T]$,

$$E\left[\int_0^T \Delta(t)\, dt\right] < \infty \quad \forall\, T,$$

then if we consider the partition

$$\Pi_n = \{0 = t_0 < t_1 < \dots < t_n = T\}$$

with $\|\Pi_n\| = \max_i \{t_{i+1} - t_i\}$, then the Itô integral

$$I(T) = \int_0^T \Delta(t)\, dW(t)$$

is defined to be

$$\lim_{\|\Pi_n\| \to 0} \mathbf{E}\left[\sum_{i=0}^{n-1} \Delta(t_i)\left[W(t_{i+1}) - W(t_i)\right] - \int_0^T \Delta(t)\, dW(t)\right]^2 = 0$$

▲

Example 5.1.

We will calculate the integral

$$\int_0^T W(t)\, dW(t)$$

following the above definition. Consider the above partition and define the Riemann sum

$$I_n(T) = \sum_{i=0}^{n-1} W(t_i)\left[W(t_{i+1}) - W(t_i)\right]$$

We have that

$$2W(t_i)\left[W(t_{i+1}) - W(t_i)\right] = \left(W(t_i) + W(t_{i+1}) - W(t_i)\right)^2 - W^2(t_i) - \left(W(t_{i+1}) - W(t_i)\right)^2 =$$

$$= \left(\text{Let } \Delta W(t_i) = W(t_{i+1}) - W(t_i)\right) =$$

$$= W^2(t_{i+1}) - W^2(t_i) - (\Delta W(t_i))^2$$

and thus we get

$$I_n(T) = \frac{1}{2}\sum_{i=0}^{n-1} W^2(t_{i+1}) - \frac{1}{2}\sum_{i=0}^{n-1} W^2(t_i) - \frac{1}{2}\sum_{i=0}^{n-1} (\Delta W(t_i))^2 =$$

$$= \frac{1}{2} W_T^2 - \frac{1}{2}\sum_{i=0}^{n-1} \Delta W^2(t_i)$$

According to the definition of the Itô integral we get

$$\lim_{n \to \infty} \mathbf{E}\left[I_n(T) - \int_0^T W(t)\, dW(t)\right]^2 = 0$$

or equivalently

$$\lim_{n \to \infty} \mathbf{E}\left[\tfrac{1}{2} W_T^2 - \tfrac{1}{2}\sum_{i=0}^{n-1} (\Delta W(t_i))^2 - \int_0^T W(t)\, dW(t)\right] = 0 .$$

Firstly, we have that

$$\mathbf{E}\left[\sum_{i=0}^{n-1} (\Delta W(t_i))^2\right] = \sum_{i=0}^{n-1} \mathbf{E}\left[(\Delta W(t_i))^2\right] =$$

$$= \left(\text{Since } W(t) \text{ is a Wiener process } \mathbf{E}\left[\Delta W(t_i)\right]^2 = t_{i+1} - t_i\right)$$

$$= \sum_{i=0}^{n-1} (t_{i+1} - t_i) = T.$$

Since $\frac{1}{2}W_T^2$ is independent of n and in addition since if $W(t)$ was not a random variable the integral $\int_0^T W(t)\,dW(t)$ would be a Riemann integral and

$$\int_0^T W(t)\,dW(t) = \tfrac{1}{2}W^2(t)$$

intuitively from what we have established so far we intuitively arrive at the possible result which in what follows we will prove that it is true.

$$\int_0^T W(t)\,dW(t) = \tfrac{1}{2}\left[W^2(T) - T\right].$$

We have that

$$\lim_{n\to\infty} E\left[\frac{1}{2}W^2(T) - \frac{1}{2}\sum_{i=0}^{n-1}\left[\Delta W(t_i)\right]^2 - \tfrac{1}{2}W^2(T) - \frac{T}{2}\right]^2 = 0$$

or equivalently

$$\lim_{n\to\infty} \tfrac{1}{2}E\left[\sum_{i=0}^{n-1}\left[\Delta W(t_i)\right]^2 - T\right]^2 = 0$$

Let us first start with

$$E\left[\sum_{i=0}^{n-1}(\Delta W(t_i))^2 - T\right]^2 = E\left\{\left[\sum_{i=0}^{n-1}(\Delta W(t_i))^2\right]^2 - 2T\sum_{i=0}^{n-1}(\Delta W(t_i))^2 + T^2\right\}$$

$$= E\left[\sum_{i=0}^{n-1}(\Delta W(t_i))^2\right]^2 - 2T\sum_{i=0}^{n-1}E\left[(\Delta W(t_i))^2\right] + T^2$$

$$= E\left[\sum_{i=0}^{n-1}(\Delta W(t_i))^2\right]^2 - 2T\sum_{i=0}^{n-1}(t_{i+1} - t_i) + T^2$$

$$= E\left[\sum_{i=0}^{n-1}(\Delta W(t_i))^2\right]^2 - T^2.$$

Now we have that

$$E\left[\sum_{i=0}^{n-1}(\Delta W(t_i))^2\right]^2 = E\left[\sum_{i=0}^{n-1}(\Delta W(t_i))^4 + 2\sum_{i=0}^{n-1}\sum_{j<i}(\Delta W(t_i))^2(\Delta W(t_j))^2\right]$$

$$= \sum_{i=0}^{n-1}E\left[(\Delta W(t_i))^4\right] + 2\sum_{i=0}^{n-1}\sum_{j<i}E\left\{(\Delta W(t_i))^2(\Delta W(t_j))^2\right\} =$$

= (Since W_t is a Wiener process the increments $\Delta W(t_i)$ and $\Delta W(t_j)$ are independent for $i \neq j$)

$$= \sum_{i=0}^{n-1}E\left[(\Delta W(t_i))^4\right] + 2\sum_{i=0}^{n-1}\sum_{j<i}E\left[\Delta W(t_i)\right]^2 E\left[\Delta W(t_j)\right]^2 =$$

$$= \sum_{i=0}^{n-1}E\left[(\Delta W(t_i))^4\right] + 2\sum_{i=0}^{n-1}\sum_{j<i}(t_{i+1} - t_i)(t_{j+1} - t_j).$$

Now we have to find $E[\Delta W(t_i)]^4$. Since $\Delta W(t_i)$ has the normal distribution we have that

$$E[\Delta W(t_i)^4] = 3(t_{i+1}-t_i)^2$$

Thus we have that

$$E\left[\sum_{i=0}^{n-1}(\Delta W(t_i))^2 - T\right]^2 = 3\sum_{i=0}^{n-1}(t_{i+1}-t_i)^2 + 2\sum_{i=0}^{n-1}\sum_{j<i}(t_{i+1}-t_i)(t_{j+1}-t_j) - T^2.$$

Let

$$m = \max_i (t_{i+1}-t_i),$$

then

$$E\left[\sum_{i=0}^{n-1}(\Delta W_{t_i})^2 - T\right] \le 3nm^2 + n(n-1)m^2 - T^2 \le$$
$$\le (\text{since } T \le nm) \le$$
$$\le 3nm^2 + n(n-1)m^2 - n^2m^2 \le$$
$$\le m^2(3n + n^2 - n - n^2) \le 2nm^2,$$

and consequently

$$\lim_{n\to\infty} E\left[\sum_{i=0}^{n-1}(\Delta W(t_i))^2 - T\right]^2 \le \lim_{n\to\infty} 2nm^2 = 0.$$

Thus we have proved that the stochastic integral $\int_0^T W(t)\,dW(t)$ is equal with

$$\int_0^T W(t)dW(t) = \tfrac{1}{2}(W_T^2 - T).$$

It is apparent that the above integral is different from the corresponding Riemann integral $\int_0^T x^2 dx = \frac{1}{2}x^2$. It is also more than obvious from the above example that we can not rely on the definition of the Itô integral as a methodology to calculate stochastic Integrals. Fortunately, for this purpose the Itô-Doeblin formula has been inveted and we will study it later on. We will now without proof provide in the form of a Theorem some of the properties of the Ito integral.

<u>Theorem</u> <u>5.1</u>. Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, $\{\mathcal{F}(t)\}_{t\ge 0}$ be a filtration. Let $\Delta(t)$, $0\le t\le T$ be an adaptive stochastic process to the filtration $\{\mathcal{F}(t)\}_{t\ge 0}$,

$$E\left[\int_0^T \Delta^2(t)\,dt\right] < \infty.$$

Then the Itô integral

$$I(t) = \int_0^t \Delta(u)\,dW(u)$$

has the following properties:

(i). (Continuity). As a function of the upper limit of integration t, the paths of $I(t)$ are continuous.

(ii). (Adaptivity). For each t, $I(t)$ is $\mathcal{F}(t)$-measurable.

(iii). (Linearity). If $\{A(t)\}_{t\geq 0}$, $0\leq t\leq T$ is an adapted stochastic process to the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$ then

$$\int_0^t (A(u)\pm\Delta(u))\,dW(u) = \int_0^t A(u)\,dW(u) + \int_0^t \Delta(u)\,dW(u).$$

(iv) (Martingale). The stochastic process

$$\int_0^t \Delta(u)\,dW(u)$$

is a martingale in relation to the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$

(v) (Itô isometry).

$$E\left[\int_0^t \Delta(u)\,dW(u)\right]^2 = E\int_0^t \Delta^2(u)\,du$$

(vi). (Quadratic variation). The quadratic variation of the stochastic process

$$\int_0^t \Delta(u)\,dW(u)$$

is equal to

$$\int_0^t \Delta^2(u)\,du.$$

Proof: (Exercise).

5.2 The Itô - Doeblin Formula.

We start with the definition of the quadratic variation of a function and by proving that the quadratic variation of the Brownian motion over an interval $[0,T]$ is equal with T. Together with two remarks these results help a lot in the proof of the above Theorem and of the Itô-Doeblin formula that follows.

Definition 5.1. Let $f(t)$ be a function defined for $0\leq t\leq T$. The quadratic variation of f up to time T is

$$[f,f](T) = \lim_{\|\Pi_n\|\to 0}\sum_{j=0}^{n-1}\left[f(t_{j+1})-f(t_j)\right]^2,$$

where $\Pi_n=\{t_0,t_1,\dots,t_n\}$ a partition of the interval $[0,T]$ and $\|\Pi\|=\max\{t_{k+1}-t_k\}$.

Remark 5.1.

Assume that $f\in C^1$, i.e. it has a continuous derivative. Then

$$\sum_{j=0}^{n-1}[f(t_{j+1})-f(t_j)]^2=\sum_{j=0}^{n-1}|f'(t_j^*)|^2(t_{j+1}-t_j)^2\le\|\Pi_n\|\sum_{j=0}^{n-1}|f'(t_j^*)|^2(t_{j+1}-t_j),$$

and thus

$$[f,f](T)\le\lim_{\|\Pi_n\|\to0}\left\{\|\Pi_n\|\cdot\sum_{j=0}^{n-1}|f'(t_j^*)|^2(t_{j+1}-t_j)\right\}$$

$$=\lim_{\|\Pi_n\|\to0}\|\Pi_n\|\cdot\lim_{\|\Pi_n\|\to0}\sum_{j=0}^{n-1}|f'(t_j^*)|^2(t_{j+1}-t_j)$$

$$=\lim_{\|\Pi_n\|\to0}\|\Pi_n\|\cdot\int_0^T|f'(t)|^2dt=0,$$

since from the fact that $f\in C^1$ we know that $\int_0^T|f'(t)|^2dt$ is finite.

Theorem 5.2. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $W(t)$ a Wiener process. Then $[W,W](T)=T$ for all $T\ge0$ almost surely.

Proof: Let $\Pi_n=\{t_0,t_1,\dots,t_n\}$ be a partition of $[0,T]$ Define the sampled quadratic variation corresponding to this partition to be

$$[W_{\Pi_n}]=\sum_{j=0}^{n-1}(W(t_{j+1})-W(t_j))^2$$

In order to prove the theorem it is sufficient to show that $\lim\limits_{\|\Pi_n\|\to0}[W_{\Pi_n}]=T$

Since $[W_{\Pi_n}]$ is apparently a random variable it suffices to show that

$$\lim_{\|\Pi_n\|\to0}E\{[W_{\Pi_n}]\}=T\quad\text{and}\quad\lim_{\|\Pi_n\|\to0}\operatorname{var}\{[W_{\Pi_n}]\}=0$$

We have that

$$E\{[W_{\Pi_n}]\}=\sum_{j=0}^{n-1}E\left[(W(t_{j+1})-W(t_j))^2\right]=\sum_{j=0}^{n-1}\operatorname{var}\left[W(t_{j+1})-W(t_j)\right]$$

$$=\sum_{j=0}^{n-1}(t_{j+1}-t_j)=T.$$

Moreover, we have that

$$\begin{aligned}
&\operatorname{var}\left[(W(t_{j+1}) - W(t_j))^2\right] = \\
&= E\left[\left\{(W(t_{j+1}) - W(t_j))^2 - E\left[(W_{t_{j+1}}(t_{j+1}) - W(t_j))^2\right]\right\}^2\right] \\
&= E\left[\left\{(W(t_{j+1}) - W(t_j))^2 - (t_{j+1} - t_j)^2\right\}^2\right] \\
&= E\left[(W(t_{j+1}) - W(t_j))^4 - 2(t_{j+1} - t_j)\, E\left[(W(t_{j+1}) - W(t_j))^2\right]\right. \\
&\quad + (t_{j+1} - t_j)^2 .
\end{aligned}$$

Since $W(t_{j+1}) - W(t_j) \approx N(0, t_{j+1} - t_j)$ as we have seen before we have

$$E\left[(W(t_{j+1}) - W(t_j))^4\right] = 3(t_{j+1} - t_j)^2,$$

hence

$$\begin{aligned}
\operatorname{var}\left[(W(t_{j+1}) - W(t_j))^2\right] &= 3(t_{j+1} - t_j)^2 - 2(t_{j+1} - t_j)^2 + (t_{j+1} - t_j)^2 \\
&= 2(t_{j+1} - t_j)^2
\end{aligned}$$

Consequently, we get that

$$\begin{aligned}
\operatorname{var}\left\{[W_{\Pi_n}]\right\} &= \sum_{j=0}^{n-1} \operatorname{var}\left[(W(t_{j+1}) - W(t_j))^2\right] = \sum_{j=0}^{n-1} 2(t_{j+1} - t_j)^2 \\
&\leq \sum_{j=0}^{n-1} 2\|\Pi_n\|(t_{j+1} - t_j) = 2\|\Pi_n\| T .
\end{aligned}$$

Thus

$$\lim_{\|\Pi_n\| \to 0} \operatorname{var}\left\{[W_{\Pi_n}]\right\} = 0 .$$

▲

<u>Remark 5.2</u>: All too often in our proofs we will use the approximation

$$\frac{(W(t_{j+1}) - W(t_j))^2}{t_{j+1} - t_j} \approx 1 .$$

or equivalently

$$(W(t_{j+1}) - W(t_j))^2 \approx t_{j+1} - t_j$$

this is apparent that it is compatible with the fact that

$$dW(t)\, dW(t) = dt$$

a relation which we established earlier, and which we write informally.

We are now in a position to state and prove the Itô - Doeblin formula.

Theorem 5.3 Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space and let $W(t)$ be a Brownian motion. In addition let $f(t,x)$ be a function for which

$$\frac{\partial f(t,x)}{\partial t}, \quad \frac{\partial f(t,x)}{\partial x}, \quad \frac{\partial^2 f(t,x)}{\partial x^2}$$

exist and are continuous. Then, for every $T>0$

$$f(T, W(T)) = f(0, W(0)) + \int_0^T \frac{\partial f(t, W(t))}{\partial t} dt + \int_0^T \frac{\partial f(t, W(t))}{\partial W(t)} dW(t) + $$
$$+ \frac{1}{2} \int_0^T \frac{\partial^2 f(t, W(t))}{\partial W^2(t)} dt .$$

Proof: We write Taylor's Theorem up to second derivative for the function $f(t,x)$. In this respect let x_{j+1}, x_j values of x and t_{j+1}, t_j be values of t then

$$f(t_{j+1}, x_{j+1}) - f(t_j, x_j) = \frac{\partial f(t_j, x_j)}{\partial t} (t_{j+1} - t_j) + \frac{\partial f(t_j, x_j)}{\partial x} (x_{j+1} - x_j)$$
$$+ \frac{1}{2} \frac{\partial^2 f(t_j, x_j)}{\partial x} (x_{j+1} - x_j)^2 +$$
$$+ \frac{\partial^2 f(t_j x_j)}{\partial t \partial x} (t_{j+1} - t_j)(x_{j+1} - x_j).$$
$$+ \frac{1}{2} \frac{\partial^2 f(t_j, x_j)}{\partial^2 t} (t_{j+1} - t_j)^2$$

We thus get that

$$f(T, W(T)) - f(0, W(0)) = \sum_{j=0}^{n-1} \left[f(t_{j+1}, W(t_{j+1})) - f(t_j, W(t_j)) \right]$$
$$= \sum_{j=0}^{n-1} \frac{\partial f(t_j, W(t_j))}{\partial t} \cdot (t_{j+1} - t_j) + \sum_{j=0}^{n-1} \frac{\partial f(t_j, W(t_j))}{\partial W(t)} \cdot (W(t_{j+1}) - W(t_j))$$
$$+ \frac{1}{2} \sum_{j=0}^{n-1} \frac{\partial^2 f(t_j, W(t_j))}{\partial W^2(t)} (W(t_{j+1}) - W(t_j))^2$$
$$+ \sum_{j=0}^{n-1} \frac{\partial^2 f(t_j, W(t_j))}{\partial t \partial W(t_j)} (t_{j+1} - t_j)(W(t_{j+1}) - W(t_j)).$$
$$+ \frac{1}{2} \sum_{j=0}^{n-1} \frac{\partial^2 f(t_j, W(t_j))}{\partial t^2} (t_{j+1} - t_j)^2.$$

In the previous equation in both parts we take the limit as $\|\Pi_n\| \to 0$. The left hand part of the equation remains unaffected. From the right hand side of the equation we have that

a) $$\lim_{\|\Pi_n\|\to 0} \sum_{j=0}^{n-1} \frac{\partial f(t_j, W(t_j))}{\partial t}(t_{j+1}-t_j) = \int_0^T \frac{\partial f_t(t, W(t))}{\partial t}\, dt.$$

b) $$\lim_{\|\Pi_n\|\to 0} \sum_{j=0}^{n-1} \frac{\partial f(t_j, W(t_j))}{\partial W(t)}\left(W(t_{j+1}) - W(t_j)\right) = \int_0^T \frac{\partial f(t, W(t))}{\partial W(t)}\, dW(t).$$

c) $$\lim_{\|\Pi_n\|\to 0} \sum_{j=0}^{n-1} \frac{\partial^2 f(t_j, W(t_j))}{\partial W^2(t)}\left(W(t_{j+1}) - W(t_j)\right)^2 = (\text{Due to Remark } 5.2) =$$

$$= \lim_{\|\Pi_n\|\to 0} \sum_{j=0}^{n-1} \frac{\partial^2 f(t_j, W(t_j))}{\partial W^2(t)}\left(t_{j+1} - t_j\right)^2 = \frac{1}{2}\int_0^T \frac{\partial^2 f(t, W(t))}{\partial W^2(t)}\, dt$$

d) $$\lim_{\|\Pi_n\|\to 0} \left| \sum_{j=0}^{n-1} \frac{\partial^2 f(t_j, W(t_j))}{\partial t\, \partial W(t)}(t_{j+1}-t_j)\left(W(t_{j+1}) - W(t_j)\right)\right| \leq$$

$$\leq \lim_{\|\Pi_n\|\to 0} \sum_{j=0}^{n-1} \left|\frac{\partial^2 f(t_j, W(t_j))}{\partial t\, \partial W(t)}\right| (t_{j+1}-t_j) \left|W(t_{j+1}) - W(t_j)\right| \leq$$

$$\leq \lim_{\|\Pi_n\|\to 0} \cdot \max_{0\leq i\leq n-1} \left|W(t_{i+1}) - W(t_i)\right| \cdot \lim_{\|\Pi_n\|\to 0} \sum_{j=0}^{n-1} \left|\frac{\partial^2 f(t_j, W(t_j))}{\partial t\, \partial W(t)}\right| (t_{j+1}-t_j) =$$

$$= 0 \cdot \int_0^T \left|\frac{\partial^2 f(t, W(t))}{\partial t\, \partial W(t)}\right| dt = 0.$$

e) $$\lim_{\|\Pi_n\|\to 0} \left| \frac{1}{2}\sum_{j=0}^{n-1} \frac{\partial^2 f(t_j, W(t_j))}{\partial^2 t}(t_{j+1}-t_j)^2 \right| \leq$$

$$\lim_{\|\Pi_n\|\to 0} \frac{1}{2}\sum_{j=0}^{n-1} \left|\frac{\partial^2 f(t_j, W(t_j))}{\partial^2 t}\right| (t_{j+1}-t_j)^2 \leq$$

$$\leq \frac{1}{2} \lim_{\|\Pi_n\|\to 0} \max_{0\leq i\leq n-1} (t_{i+1}-t_i) \lim_{\|\Pi_n\|\to 0} \sum_{j=0}^{n-1} \left|\frac{\partial^2 f(t_j, W(t_j))}{\partial^2 t}\right| (t_{j+1}-t_j)$$

$$= \frac{1}{2}\, 0 \cdot \int_0^T \left|\frac{\partial^2 f(t, W(t))}{\partial t^2}\right| dt = 0$$

▲.

Remark 5.3: From the above theorem it is evident that we could informally write the Itô-Doeblin formula in differential form as follows:

$$df(t,W(t)) = \frac{\partial f(t,W(t))}{\partial t}dt + \frac{\partial f(t,W(t))}{\partial W(t)}dW(t) + \frac{1}{2}\frac{\partial^2 f(t,W(t))}{\partial W^2(t)}dt$$

We have also established that

$$dW(t)dW(t) = dt \quad , \; dt\,dW(t) = dW(t)dt = 0 \quad , \; dt\,dt = 0 .$$

Example 5.2

Let that the value of a derivative is given by

$$f(W(t),t) = 2W^3(t) + t$$

We have that

$$\frac{\partial f(t,W(t))}{\partial W(t)} = 6W^2(t) \quad , \quad \frac{\partial f(t,W(t))}{\partial W(t)} = 1 \quad \text{and} \quad \frac{\partial^2 f(t,W(t))}{\partial^2 W(t)} = 12W(t)$$

From the above we arrive at the stochastic differential equation

$$df(t,W(t)) = 6W^2(t)dW(t) + (6W(t)+1)dt .$$

Example 5.3.

Let us now find the Itô integral of example 5.1 i.e

$$\int_0^T W(t)\,dW(t)$$

and have a chance to compare the contribution of the Itô-Doeblin formula to the solution of the problem

In such cases we need to find a function of $W(t)$ let say $f(t,W(t))$ which is such that

$$\frac{\partial f(t,W(t))}{\partial W(t)} \quad \text{is equal with the integrand } W(t).$$

In this respect we choose

$$f(t,W(t)) = \frac{1}{2}W^2(t) .$$

From the differential form of Itô-Doeblin Formula we get that

$$df(t,W(t)) = \frac{\partial f(t,W(t))}{\partial W(t)}\, dW(t) + \frac{\partial f(t,W(t))}{\partial t}\, dt + \frac{1}{2}\frac{\partial^2 f(t,W(t))}{\partial^2 W(t)}\, dt$$

with

$$\frac{\partial f(t,W(t))}{\partial W(t)} = W(t), \quad \frac{\partial f(t,W(t))}{\partial t} = 0, \quad \frac{\partial^2 f(t,W(t))}{\partial^2 W(t)} = 1.$$

Hence,

$$df(t,W(t)) = W(t)\,dW(t) + \tfrac{1}{2}\,dt.$$

We integrade both sides of this equation to get

$$f(t,W(t)) = f(W(0),0) + \int_0^T W(t)\,dW(t) + \tfrac{1}{2}\int_0^T dt.$$

Since $W(0)=0$ we arrive at

$$\tfrac{1}{2}W^2(T) = \int_0^T W(t)\,dW(t) + \tfrac{T}{2},$$

from which we easily get the known result that

$$\int_0^T W(t)\,dW(t) = \tfrac{1}{2}\left(W^2(T) - T\right).$$

▲

<u>Example 5.4</u>: We want to find the Itô integral

$$\int_0^T e^{W(t)}\,dW(t).$$

We choose as $f(t,W(t))$ the function.

$$f(t,W(t)) = e^{W(t)}$$

Now, we have that

$$\frac{\partial f(t,W(t))}{\partial W(t)} = e^{W(t)}, \quad \frac{\partial f(t,W(t))}{\partial t} = 0, \quad \frac{\partial^2 f(t,W(t))}{\partial^2 W(t)} = e^{W(t)}$$

Hence we get

$$df(t,W(t)) = e^{W(t)}\,dW(t) + \tfrac{1}{2}e^{W(t)}\,dt$$

Integrading both sides of this stochastic differential equation we get that

$$f(t,W(t)) = f(0,W(0)) + \int_0^T e^{W(t)}\,dW(t) + \tfrac{1}{2}\int_0^T e^{W(t)}\,dt$$

and since $W(0)=0$

$$\int_0^T e^{W(t)}\,dW(t) = e^{W(t)} - 1 - \tfrac{1}{2}\int_0^T e^{W(t)}\,dt$$

▲

5.3. Itô Processes.

In 4.3 we presented a specific class of stochastic differential equations usually used as models for various assets. We will now introduce a more general class of stochastic differential equations known also as Itô processes. The final aim of this section is to provide the Itô-Doeblin formula for Itô processes.

Definition 5.2. Let $W(t), t \geq 0$, be a Wiener process, and let it be adapted to the filtration $\mathcal{F}(t)$. An Itô process is a stochastic process of the form

$$X(t) = X(0) + \int_0^t \Theta(u)\,du + \int_0^t \Delta(u)\,dW(t)$$

with

$$\int_0^t |\Theta(u)|\,du < \infty \quad \text{and} \quad E\left\{\int_0^t \Delta^2(u)\,du\right\} < \infty$$

where $\Theta(t)$ and $\Delta(t)$ are adapted stochastic processes to the filtration $\mathcal{F}(t)$, and $X(0)$ is the intial value of $X(t)$ assumed to be known.

The informal differential form of a Itô process is given by

$$dX(t) = \Theta(t)dt + \Delta(t)dW(t).$$

Now using the already established results that

$$dW(t)\,dW(t) = dt \;,\; dt\,dW(t) = dW(t)\,dt = 0 \;\text{ and } dt\,dt = 0,$$

we get that

$$\begin{aligned} dX(t)dX(t) &= \Delta^2(t)\,dW(t)\,dW(t) + 2\Delta(t)\,\Theta(t)\,dW(t)\,dt + \Theta^2(t)\,dt\,dt \\ &= \Delta^2(t)dt. \end{aligned}$$

This denotes in fact that the process $X(t)$ accummulates in the interval $[t, t+dt)$ quadratic variation $\Delta^2(t)dt$. Hence the total quadratic accumulation in the interval $[0,t]$ is given by

$$[X,X](t) = \int_0^t \Delta^2(u)\,du.$$

This quadratic variation is solely due to the quadratic variation of the Itô integral

$$I(t) = \int_0^t \Delta(u)\,dW(u).$$

The ordinary integral

$$\int_0^t \Theta(u)du$$

has zero quadratic variation.

Intuitively it helps to think of $\int_0^t \Theta(u)du$ is like investing money in a market account where $\Theta(t)$ is the random interest rate. Then at time t since $\Theta(t)$ is adapted to $\mathcal{F}(t)$ we know $\int_0^t \Theta(u)du$ and in a small interval $[t,t+h]$ we have

$$\int_0^{t+h} \Theta(u)du \approx \int_0^t \Theta(u)du + \Theta(t)h$$

which apparently doesn't allow

$$\int_0^{t+h} \Theta(u)du$$

to be much volatile. In contrast

$$\int_0^{t+h} \Delta(u)dW(u) = \int_0^t \Delta(u)\, dW(u) + \Delta(t)\big(W(t+h)-W(t)\big)$$

where in addition to the fact that we do not know $W(t+h)-W(t)$ it is indepedent to the information contained to the filtration $\mathcal{F}(t)$. Since we model assets as stochastic differentials of Wiener process it is like investing in an asset. We know from real life that this investment is a lot more volatile.

We shall also need integrals with respect to Itô processes. This is easily done by separating the integral into one with respect to the Wiener process $dW(t)$ and one with respect to dt

Definition 5.3. Let $W(t)$ be a Wiener process adapted to a filtration $\mathcal{F}(t)$. Let also $X(t), t\geq 0$, be an Itô process which in differential form is given by

$$dX(t) = \Theta(t)dt + \Delta(t)dW(t)$$

Let $\Gamma(t), t\geq 0$ be an adapted process to the filtration $\mathcal{F}(t)$ such that

$$\int_0^t |\Gamma(u)\Theta(u)|du < \infty \quad , \quad E\left\{\int_0^t \Gamma^2(u)\Delta^2(u)dW(u)\right\} < \infty$$

Then we define the integral with respect to $X(t)$ by

$$\int_0^t \Gamma(u)dX(t) = \int_0^t \Gamma(u)\Delta(u)dW(u) + \int_0^t \Gamma(u)\Theta(u)du \qquad \blacktriangle$$

Now following the methodology 5.3 it easy to arrive at the following theorem:

Theorem 5.4. Let $W(t)$ be a Wiener process adapted to a filtration $\mathcal{F}(t)$. Let $X(t)$, $t \geq 0$, be an Itô process and let the conditions of definition 5.3 be satisfied. Let also a function $f(t,x)$ be a function for which the partial derivatives

$$\frac{\partial f(t,x)}{\partial x}, \quad \frac{\partial^2 f(t,x)}{\partial x^2}, \quad \frac{\partial f(t,x)}{\partial t}$$

exist and are continuous. Then, for every $T \geq 0$,

$$f(T,X(T)) = f(0,X(0)) + \int_0^T \frac{\partial f(t,X(t))}{\partial t} dt + \int_0^T \frac{\partial f(t,X(t))}{\partial X(t)} \Delta(t) dW(t)$$
$$+ \frac{1}{2}\int_0^T \frac{\partial f(t,X(t))}{\partial X(t)} \Theta(t) dt + \frac{1}{2}\int_0^T \frac{\partial^2 f(t,X(t))}{\partial X^2(t)} \Delta^2(t) dt.$$

▲

The above Theorem provides the Itô-Doeblin formula for an Itô process.

The differentian form of the Itô-Doeblin formula for an Itô process $X(t)$ given as in definition 5.3 is given by:

$$df(t,X(t)) = \frac{\partial f(t,X(t))}{\partial t} dt + \frac{\partial f(t,X(t))}{\partial X(t)} \Delta(t) dW(t) +$$
$$+ \frac{\partial f(t,X(t))}{\partial X(t)} \Theta(t) dt + \frac{1}{2} \frac{\partial^2 f(t,X(t))}{\partial X^2(t)} \Delta^2(t) dt.$$

The above formula is very useful in finding Itô-integrals and solving stochastic differential equations which are branches of what is known as Itô calculus. However their more important applications are in finance where usually f is a derivative and $X(t)$ is a model for the price of an asset.

5.4. Applications

$\boxed{1.}$ We start by demostrating the solution of the stochastic differential equation the solution of which is known as geometrical Brownian motion i.e.

$$dB(t) = \mu B(t) dt + \sigma B(t) dW(t).$$

Apparently the above is an Itô process with:

$$\Theta(t) = \mu B(t) \quad \text{and} \quad \Delta(t) = \sigma B(t).$$

Let us define the function

$$f(t,x) = \ln x,$$

and apply the Itô-Doeblin formula to obtain

$$d(\ln B(t)) = \frac{1}{B(t)} dB(t) + \frac{1}{2}\left(-\frac{1}{B^2(t)}\right)(dB(t))^2$$

$$= \frac{dB(t)}{B(t)} - \frac{1}{2B^2(t)} \cdot \sigma^2 B^2(t) dt =$$

$$= \frac{dB(t)}{B(t)} - \frac{1}{2}\sigma^2 dt .$$

Hence

$$\frac{dB(t)}{B(t)} = d(\ln B(t)) + \tfrac{1}{2}\sigma^2 dt$$

Now from the initial stochastic differential equation we get

$$\frac{dB(t)}{B(t)} = \mu dt + \sigma dW(t)$$

Hence

$$d(\ln B(t)) + \frac{1}{2}\sigma^2 dt = \mu dt + \sigma dW(t)$$

or

$$d(\ln B(t)) = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW(t).$$

Integrating both sizes of the above equation we get:

$$\ln \frac{B(t)}{B(0)} = \left(\mu - \tfrac{1}{2}\sigma^2\right)t + \sigma W(t)$$

or

$$B(t) = B(0) \exp\left(\left(\mu - \tfrac{1}{2}\sigma^2\right)t + \sigma W(t)\right)$$

Thus we have obtained a close analytic solution for the geometric Brownian motion. The geometric Brownian motion is the most often used stochastic model for asset pricing.

It is of interest to look also at the asymptotic behaviour of the geometric Brownian motion.

(i) If $\mu > \frac{1}{2}\sigma^2$ then $B(t) \to \infty$ as $t \to \infty$, almost surely.

(ii) If $\mu < \frac{1}{2}\sigma^2$ then $B(t) \to 0$ as $t \to \infty$, almost surely.

(iii) If $\mu = \frac{1}{2}\sigma^2$ then $B(t)$ will flactuate between arbitrary large and arbitrary small values as $t \to \infty$. ▲

$\boxed{2.}$ As another application let us know find the solution for a type of mean reverting Ornestein - Ulenbeck stochastic process

$$dB(t) = (\mu - B(t))\,dt + \sigma\, dW(t)$$

The reader may apply Itô-Doeblin formula on the function

$$f(t, B(t)) = e^{t}(\mu - B(t))$$

and prove that

$$B(t) = \mu + (B(0) - \mu)e^{-t} + \sigma \int_0^t e^{s-t}\, dW(s)$$

We are now in a position to find $E[B(t)]$. Since the expected value of an Itô integrale is always zero, it is easy to get that

$$E[B(t)] = \mu + (B(0) - \mu)\, e^{-t}.$$

Also we are able to find the $\operatorname{var}[B(t)]$, since we have

$$\begin{aligned}
\operatorname{Var}[B(t)] &= E[(B(t) - E(B(t)))^2] \\
&= E\left[\left[\mu + (B(0) - \mu)e^{-t} + \sigma\int_0^t e^{s-t}\, dW(s) - \mu - (B(0) - \mu)e^{-t}\right]^2\right) \\
&= E\left(\left[\sigma \int_0^t e^{s-t}\, dW(s)\right]^2\right) = \sigma^2 e^{-2t}\, E\left[\left(\int_0^t e^{s}\, dW(s)\right)^2\right] \\
&= (\text{by Itô's integral, Isometry property}) \\
&= \sigma^2 e^{-2t}\, E\left[\left(\int_0^t e^{2s}\, ds\right)\right] \\
&= \frac{1}{2}\sigma^2 (1 - e^{-2t}).
\end{aligned}$$

$\boxed{3.}$ We now provide an application of the Itô - Doeblin formula in the calculation of the following Itô integral

$$\int_0^T t\,W(t)\,dW(t).$$

The methodology we follow in such cases is to find a function $F(t, W(t))$ such that its partial derivative

$$\frac{\partial F(t,W(t))}{\partial W(t)}$$

will be equal to the integrand. Such a function is apparently

$$F(t, W(t)) = t\,W^2(t)$$

We have that

$$\frac{\partial F(t,W(t))}{\partial W(t)} = 2W(t)\ ,\quad \frac{\partial F(t,W(t))}{\partial t} = W^2(t)\ ,\quad \frac{\partial^2 F^2(t,W(t))}{\partial^2 W(t)} = 2t.$$

Hence from the Itô-Doeblin formula we get

$$dF(t,W(t)) = 2t\,W(t)\,dW(t) + W^2(t)\,dt + t\,dt$$

Integrading both sides of this equation we get

$$\begin{aligned} F(T, W(T)) &= F(0,W(0)) + \int_0^T 2tW(t)\,dW(t) + \int_0^T W^2(t)\,dt + \int_0^T t\,dt \\ &= F(0,W(0)) + 2\int_0^T t\,W(t)\,dW(t) + \int_0^T W^2(t)\,dt + \int_0^T t\,dt \end{aligned}$$

Since $W(0)=0$ we arrive at the relation

$$TW_T^2 = 2\int_0^T t\,W(t)\,dW(t) + \int_0^T W^2(t)\,dt + T$$

or

$$\int_0^T t\,W(t)\,dW(t) = \frac{1}{2}T\left(W^2(T) - 1\right) - \int_0^T W^2(t)\,dt.$$

$\boxed{4}$ Generalized geometric Brownian motion.

Let $W(t), t\geq 0$ be a Wiener process adapted to a filtration $\mathcal{F}(t), t\geq 0$ and let $\mu(t)$ and $\sigma(t)$ be adapted stochastic processes to the filtration $\mathcal{F}(t)$. Define the Itô process

$$X(t) = \int_0^t \sigma(s)\,dW(s) + \int_0^t \left(\mu(s) - \tfrac{1}{2}\sigma^2(s)\right)ds$$

In differential form the above Itô process could be written as

$$dX(t) = \sigma(t)\,dW(t) + \left(\mu(t) - \frac{1}{2}\sigma^2(t)\right)dt.$$

Consider now an asset price model which is of the form

$$S(t) = S(0)e^{X(t)} = S(0)\exp\left\{\int_0^t \sigma(s)\,dW(s) + \int_0^t \left(\mu(s) - \frac{1}{2}\sigma^2(s)\right)ds\right.$$

where $S(0)$ is known and positive being the initial value of the asset price.
With the use of the differential form of the Itô-Doeblin formula we arrive at the following stochastic differential equation.

$$dS(t) = \mu(t)S(t)\,dt + \sigma(t)S(t)\,dW(t).$$

This process is called the generalized Geometric Brownian motion and represents in fact a large class of asset pricing models. Its characteristics are that it is always positive, it has no jumps and is driven by a single Brownian motion.

Note that the distribution of $S(t)$ is not lognormal any more as it is in the ordinary geometric Brownian motion. This is due to the fact that $\mu(t)$ and $\sigma(t)$ are adapted stochastic processes. In the case of $\mu(t)=\mu$ and $\sigma(t)=\sigma$ constants we get

$$S(t) = S(0)\exp\left\{\sigma W(t) + (\mu - \tfrac{1}{2}\sigma^2)t\right\}.$$

An important remark is that the above $S(t)$ is not a martingale as one might intuitively wrongly might believe. This is mainly due to the fact that $e^{\sigma W(t)}$ is not a martingale since the convexity of $e^{\sigma x}$ imparts an upward drift. However if $\mu=0$ we have that

$$\exp\{\sigma W(t) - \tfrac{1}{2}\sigma^2 t\} \text{ is a martingale.}$$

Actually, we have

$$E\left[\exp\{\sigma W(t) - \tfrac{1}{2}\sigma^2 t\} / \mathcal{F}(s)\right] =$$

$$= E\left[\exp\{\sigma[W(t)-W(s)]\}\cdot\exp\{\sigma W(s) - \tfrac{1}{2}\sigma^2 t\} / \mathcal{F}(s)\right]$$

$$= \exp\{\sigma W(s) - \tfrac{1}{2}\sigma^2 t\}\cdot E\left[\exp\{\sigma[W(t)-W(s)]\} / \mathcal{F}(s)\right]$$

$$= \left(W(t) - W(s) \text{ is independent of } \mathcal{F}(s)\right).$$

$$= \exp\{\sigma W(s) - \tfrac{1}{2}\sigma^2 t\} E[\exp\{\sigma(W(t)-W(s))\}]$$

$=$ ($W(t)-W(s)$ is normally distributed with mean 0 and variance $t-s$)

$$= \exp\{\sigma W(s) - \tfrac{1}{2}\sigma^2 s\}.$$

Note that in the generalized geometric Brownian motion when $\mu(t)=0$ we get

$$dS(t) = \sigma(t) S(t)\, dW(t)$$

Integration of both sides yields

$$S(t) = S(0) + \int_0^t \sigma(s) S(s)\, dW(s)$$

The right hand side of the previous equation is an Itô intergal plus a constant hence it is a martingale. Consequently its expicit solution

$$S(t) = S(0) \exp\left\{ \int_0^t \sigma(s)\, dW(s) - \tfrac{1}{2}\int_0^t \sigma^2(t)\, ds \right\}$$

is a martingale.

This fact goes on to prove that the term $\sigma(t) S(t) dW(t)$ of the generalized geometrical Brownian motion, contributes no drift, just pure volatility, to the asset price.

The present application leads us to the following interesting theorem.

<u>Theorem 5.1</u> (Itô integral of a deterministic integrand).

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $W(t), t \geq 0$ a Brownian motion, and let $\Delta(s)$ be a nonrandom function of time. Let

$$\int_0^t \Delta(s)\, dW(s) \quad \text{for each } t \geq 0,$$

be the Itô integral of the deterministic integrand $\Delta(t)$. Then the above Itô integral apparently is a stochastic process and for each specific value of t is normally distributed with expected value 0 and variance $\int_0^t \Delta^2(s)\, ds$.

Proof: The initial value for $t=0$ of $\int_0^t \Delta(s)\, dW(s)$ is apparently equal to zero. However, $\int_0^t \Delta(s)\, dW(s)$ is a martingale since it is an Itô integral and thus

$$E\left[\int_0^t \Delta(s)\, ds\right] = 0$$

Consequently, we have

$$\mathrm{Var}\Big[\int_0^t \Delta(s)dW(s)\Big] = \mathbb{E}\Big[\int_0^t \Delta(s)dW(s)\Big]^2 = (\text{Itô's integral isometry property}) =$$

$$= \mathbb{E}\Big[\int_0^t \Delta^2(s)\,ds\Big] = \int_0^t \Delta^2(s)\,ds .$$

We will now show that $\int_0^t \Delta(s)\,dW(s)$ for a specific value of t is normally distributed with the above mean and variance We shall do this by establishing that $\int_0^t \Delta(s)dW(s)$ has the moment-generating function of a normal random variable with mean zero and variance $\int_0^t \Delta^2(s)\,ds$, i.e we have to show that

$$\mathbb{E}\Big\{\exp\Big[u\int_0^t \Delta(s)dW(s)\Big]\Big\} = \exp\Big\{\tfrac{1}{2}u^2\int_0^t \Delta^2(s)ds\Big\}.$$

Due to the fact that $\Delta(t)$ is deterministic we have that the above equation is equivalent with

$$\mathbb{E}\Big\{\exp\Big[u\int_0^t \Delta(s)dW(s) - \tfrac{1}{2}u^2\int_0^t \Delta^2(s)ds\Big]\Big\} = 1 .$$

However, we have already established in our study of the generalized Brownian motion that

$$\exp\Big[u\int_0^t \Delta(s)dW(s) - \frac{1}{2}u^2\int_0^t \Delta^2(s)ds\Big],$$

is a martingale. All we have to do is to put $\sigma(s) = u\Delta(s)$ in the expression

$$\exp\Big[\int_0^t \sigma(s)dW(s) - \tfrac{1}{2}\int_0^t \sigma^2(t)\,ds\Big],$$

which we proved to be a martingale

In addition this process takes at $t=0$ the value 1, and hence its expectation is always 1.

An interesting application of the above theorem is the following known also as Vasicek interest rate model. In this model it is assumed that the interest rate process $\{r(t)\}_{t\geq 0}$ satisfies the stochastic differential equation

$$dr(t) = (\alpha - \beta r(t))dt + \sigma dW(t)$$

where α, β and σ are positive constants and is a mean reverting stochastic process.

With similar methods presented to previous stochastic differential equations we arrive at the solution

$$r(t) = e^{-\beta t}R(0) + \frac{\alpha}{\beta}\big(1 - e^{-\beta t}\big) + \sigma e^{-\beta t}\int_0^t e^{\beta s}\,dW(s) .$$

From theorem 5.1 we get that for specific t the random variable

$$\int_0^t e^{\beta s} dW(s)$$

is normally distributed with mean zero and variance

$$\int_0^t e^{2\beta s} ds = \frac{1}{2\beta}\left(e^{2\beta t} - 1\right)$$

Hence, $r(t)$ is normally distributed with

$$E[r(t)] = e^{-\beta t} r(0) + \frac{a}{\beta}\left(1 - e^{-\beta t}\right) \;;\; var[r(t)] = \frac{\sigma^2}{2\beta}\left(1 - e^{-2\beta t}\right)$$

Apparently, whatever the values of the parameters $a>0$, $\beta>0$ and $\sigma>0$, there is positive probability that $r(t)$ is negative, an undesirable property. ▲

$\boxed{5.}$ (Cox-Ingersoll-Ross interest rate model).

The Cox-Ingersoll-Ross model for the interest rate process $r(t)$ is

$$dr(t) = \left(a - \beta r(t)\right)dt + \sigma\sqrt{r(t)}\, dW(t),$$

where $a>0$, $\beta>0$ and $\sigma>0$.

Unfortunately, the Cox-Ingersoll-Ross stochastic differential equation does not have a closed analytic solution. However, its advantage over the Vasicek model is that the interest rate in the Cox-Ingersoll-Ross model does not become negative. If $r(t)$ reaches zero, the term multiplying $dW(t)$ vanishes and the positive drift term $a dt$ drives the interest rate back into positive territory.

We will now find the expected value of the interest rate process for specific t under the Cox-Ingersoll-Ross model. In this respect apply the Itô-Doeblin formula to the following function

$$d\left(e^{\beta t} \cdot r(t)\right) = \beta e^{\beta t} r(t) dt + e^{\beta t}\left(a - \beta r(t)\right)dt + e^{\beta t}\sigma\sqrt{r(t)}\, dW(t)$$

$$= a e^{\beta t} dt + \sigma e^{\beta t}\sqrt{r(t)}\, dW(t)$$

We now integrate both sides of the above equation

$$e^{\beta t} r(t) = r(0) + a \int_0^t e^{\beta s} ds + \sigma \int_0^t e^{\beta s} \sqrt{r(s)}\, dW(s).$$

$$= r(0) + \frac{a}{\beta}\left(e^{\beta t} - 1\right) + \sigma \int_0^t e^{\beta s} \sqrt{r(s)}\, dW(s).$$

Taking expectation on both sides we have

$$e^{\beta t} E[r(t)] = r(0) + \frac{a}{\beta}\left(e^{\beta t} - 1\right) + \sigma\, E\left[\int_0^t e^{\beta s} \sqrt{r(s)}\, dW(s)\right]$$

$$= \text{(the expectation of an Itô-integral is zero)}.$$

$$= r(0) + \frac{a}{\beta}\left(e^{\beta t} - 1\right)$$

or equivalently

$$E[r(t)] = e^{-\beta t} r(0) + \frac{a}{\beta}\left(1 - e^{-\beta t}\right)$$

Note that it coinsides with the expected value under the the **Vasicek** model.

We now proceed to find the variance of $r(t)$ for a specific t. In this respect we apply the Itô-Doeblin formula for the function $[e^{\beta t} r(t)]^2$.

$$d\left(e^{2\beta t} r^2(t)\right) = 2a e^{2\beta t} r(t)\, dt + 2\sigma e^{2\beta t} [r(t)]^{3/2} dW(t) + \sigma^2 e^{2\beta t} r(t)\, dt$$

Integrating both parts yields

$$e^{2\beta t} r^2(t) = r^2(0) + \left(2a + \sigma^2\right) \int_0^t e^{2\beta s} r(s)\, ds + 2\sigma \int_0^t e^{2\beta s} [r(s)]^{3/2} dW(s).$$

Taking expectation in both parts of the above equation we have two unknowns. The first is

$$E\left[\int_0^t e^{2\beta s} [r(s)]^{3/2} dW(s)\right] = 0$$

due to the fact that is the expectation of an Itô integral,

$$E\left[\int_0^t e^{2\beta s} r(s)\, ds\right] = \int_0^t e^{2\beta s} E[r(s)]\, ds$$

for which we already found $E[r(t)]$. Hence

$$E[r^2(t)] = e^{-2\beta t} r^2(0) + \frac{(2a+\sigma^2)}{\beta}\left(r(0) - \frac{a}{\beta}\right)\left(e^{-\beta t} - e^{-2\beta t}\right)$$

$$+ \frac{a(2a+\sigma^2)}{2\beta^2}\left(1 - e^{-2\beta t}\right)$$

It is by now straight forward to find $\operatorname{var}[r(t)]$.

$$\operatorname{var}(r(t)) = E\left[r^2(t)\right] - E\left[r(t)\right]^2$$
$$= \frac{\sigma^2}{\beta} r(0)\left(e^{-\beta t} - e^{-2\beta t}\right) + \frac{\alpha\sigma^2}{2\beta^2}\left(1 - 2e^{-\beta t} + e^{-2\beta t}\right).$$

5.5. Itô - Doeblin Formula for Multidimensional Processes.

We will limit our presentation in a two dimensional space and for higher dimension things follow similarly. We have already defined a multi-dimensional Brownian motion to be a vector of independent one-dimensional Brownian motions i.e.

$$\underset{\sim}{W}(t) = \left[W_1(t), W_2(t), \ldots, W_n(t)\right]$$

For each component we have $[W_i, W_i](t) = t$, which we write informally as

$$dW_i(t)\, dW_i(t) = dt$$

The independence factor is expressed by the fact that $[W_i, W_j] = 0$ for $i \neq j$ or

$$dW_i(t) \cdot dW_j(t) = 0 \quad \text{for } i \neq j.$$

Now let $X(t), Y(t)$ be two Itô processes driven by a two-dimensional Brownian motion i.e. they may be written in the form:

$$X(t) = X(0) + \int_0^t \mu_1(u)\, du + \int_0^t \sigma_{11}(u)\, dW_1(t) + \int_0^t \sigma_{12}(u)\, dW_2(u)$$
$$Y(t) = Y(0) + \int_0^t \mu_2(u)\, du + \int_0^t \sigma_{21}(u)\, dW_1(u) + \int_0^t \sigma_{22}(u)\, dW_2(u)$$

where $\mu_1(t), \mu_2(t), \sigma_{11}(t), \sigma_{12}(t), \sigma_{21}(t), \sigma_{22}(t)$ are adapted stochastic process to a filtration $\mathcal{F}(t)$ that $\underset{\sim}{W}(t)$ is also adapted too. In differential form we write.

$$dX(t) = \mu_1(t)dt + \sigma_{11}(t)\, dW_1(t) + \sigma_{12}(t)\, dW_2(t)$$
$$dY(t) = \mu_2(t)dt + \sigma_{21}(t)\, dW_1(t) + \sigma_{22}(t)\, dW_2(t)$$

Now consider

$$dX(t)dX(t) = \left(\mu_1(t)dt + \sigma_{11}(t)\, dW_1(t) + \sigma_{12}(t)\, dW_2(t)\right) \times$$
$$\left(\mu_1(t)\, dt + \sigma_{11}(t) dW_1(t) + \sigma_{12}(t)\, dW_2(t)\right) =$$

$$= \left(dt\,dt = 0,\ dt\,dW_i(t) = 0,\ dW_i(t)\,dW_i(t) = dt,\ dW_i(t)\,dW_j(t) = 0 \text{ for } i \neq j \right)$$

$$= \left[\sigma_{11}^2(t) + \sigma_{12}^2(t)\right] dt$$

Which in fact means that

$$[X,X](t) = \int_0^t \left(\sigma_{11}^2(u) + \sigma_{11}^2(u)\right) du.$$

Similarly, we arrive at the differential formulas

$$dY(t)\,dY(t) = \left(\sigma_{21}^2(t) + \sigma_{22}^2(t)\right) dt$$

$$dX(t)\,dY(t) = \left(\sigma_{11}(t)\,\sigma_{21}(t) + \sigma_{12}(t)\,\sigma_{22}(t)\right) dt$$

The following theorem is the two-dimensional version of the Itô-Doeblin formula the proof of which is similar to the one-dimensional one

__Theorem 5.2.__ Let $f(t,X(t),Y(t))$ be a function whose partial derivatives

$$f_t = \frac{\partial f(t,X(t),Y(t))}{\partial t},\ f_x = \frac{\partial f(t,X(t),Y(t))}{\partial X(t)},\ f_y = \frac{\partial f(t,X(t),Y(t))}{\partial Y(t)},$$

$$f_{xx} = \frac{\partial^2 f(t,X(t),Y(t))}{\partial X^2(t)},\ f_{yy} = \frac{\partial^2 f(t,X(t),Y(t))}{\partial Y^2(t)},\ f_{xy} = \frac{\partial^2 f(t,X(t),Y(t))}{\partial X(t)\partial Y(t)},$$

$$f_{yx} = \frac{\partial^2 f(t,X(t),Y(t))}{\partial Y(t)\partial X(t)},$$ exist and are continuous,

where $X(t)$ and $Y(t)$ are the above presented Itô process. Then the two dimensional Itô-Doeblin formula in differential form is given by:

$$df(t,X(t),Y(t)) = f_t\,dt + f_x\,dX(t) + f_y\,dY(t) + \frac{1}{2} f_{xx}\,dX(t)\,dX(t)$$
$$+ f_{xy}\,dX(t)\,dY(t) + \frac{1}{2} f_{yy}\,dY(t)\,dY(t)$$

and replacing the differentials given above we arrive at.

$$df(t,X(t),Y(t)) = \left[\sigma_{11}(t)\,f_x + \sigma_{12}\,f_y\right] dW_1(t) + \left[\sigma_{12}(t)\,f_x + \sigma_{22}(t)\,f_y\right] dW_2(t)$$
$$+ \Big[f_t + \mu_1(t)\,f_x + \mu_2(t)\,f_y + \frac{1}{2}\left(\sigma_{11}^2(t) + \sigma_{12}^2(t)\right) f_{xx} + \left(\sigma_{11}(t)\,\sigma_{21}(t) + \sigma_{12}(t)\,\sigma_{22}(t)\right) \times$$
$$f_{xy} + \frac{1}{2}\left(\sigma_{21}^2(t) + \sigma_{22}^2(t)\right) f_{yy}\Big] dt$$

A very useful corollary of the above theorem is the following.

Corollary 5.1. (Itô's product rule). Let $X(t)$ and $Y(t)$ be Itô-processes. Then

$$d(X(t)Y(t)) = X(t)dY(t) + Y(t)dX(t) + dX(t)dY(t)$$

Proof: A direct application of Itô-Doeblin two dimensional formula.

5.6. Characterizing a Brownian motion.

The following theorem proved by P. Lévy is a useful characterization of a Brownian motion.

Theorem 5.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{F}(t)$ a filtation in it. Let $\{M(t)\}_{t\geq 0}$ be a martingale relative to the filtration $\mathcal{F}(t)$. In addition assume that

$$M(0)=0 \quad , \quad M(t) \text{ has continuous paths and } [M,M](t)=t \text{ for all } t\geq 0 .$$

Then $M(t)$ is a Brownian motion.

Proof. It is well known that a characterization of a Brownian motion is that it is a martingale whose increments are normally distributed. Since $M(t)$ is a martingale in order to prove that it is a Brownian motion it suffices to prove that it's increments are normally distributed.

Since by hypothesis we have that $M(t)$ has continuous paths and in addition $[M,M](t)=t$ then those where the only conditions needed for to prove the Itô formula for a Brownian motion. Thus assuming that $f(t,x)$ is a function whose partial derivatives exist and are continuous we have:

$$df(t,M(t)) = \frac{\partial f(t,M(t))}{\partial t}dt + \frac{\partial f(t,M(t))}{\partial M(t)}dM(t) + $$
$$+ \frac{1}{2}\frac{\partial^2 f(t,M(t))}{\partial M^2(t)} dM(t)\,dM(t).$$

However by hypothesis we have that $[M,M](t)=t$ for all $t\geq 0$ which is equivalent with $dM(t)\,dM(t)=dt$. Thus integrating both sides of the previous equation we get

$$f(t,M(t)) = f(0,M(0)) + \int_0^t \left[\frac{\partial f(s,M(s))}{\partial t} + \frac{1}{2}\frac{\partial^2 f(s,M(s))}{\partial M^2(s)}\right] ds +$$

$$+ \int_0^t \frac{\partial f(s,M(s))}{\partial M(s)}\, dM(s)$$

Since $M(t)$ is a martingale it could be proved in a similar way as for any Itô integral that the stochastic integral

$$\int_0^t \frac{\partial f(s,M(s))}{\partial M(s)}\, dM(s)$$

is a martingale. However, the value of this integral at $t=0$ is apparently equal to zero. Thus we have that

$$E\left\{ \int_0^t \frac{\partial f(s,M(s))}{\partial M(s)}\, dM(s) \right\} = 0.$$

So we have

$$E\left[f(t,M(t))\right] = f(0,M(0)) + E\left\{\int_0^t \left[\frac{\partial f(s,M(s))}{\partial t} + \frac{\partial^2 f(s,M(s))}{\partial M^2(s)}\right] ds\right\}$$

Choose the function $f(t,M(t))$ to be

$$f(t,M(t)) = \exp\left\{uM(t) - \tfrac{1}{2}u^2 t\right\} \quad \text{for any } u \in \mathbb{R}$$

Then we get

$$E\left[f(t,M(t))\right] = \exp\{0\} + E\left\{\int_0^t \left[-\frac{1}{2}u^2 f(t,M(t)) + \frac{1}{2}u^2 f(t,M(t))\right] ds\right\}$$

$$= 1,$$

thus,

$$E\left[\exp\left\{uM(t) - \tfrac{1}{2}u^2 t\right\}\right] = 1,$$

from which immediately arrive that for specific t

$$E\left[e^{uM(t)}\right] = e^{\frac{1}{2}u^2 t}.$$

However, this is the moment-generating function for a random variable which is normally distributed with mean zero and variance t. Thus for any $t \geq 0$, $M(t)$ is normally distributed with mean zero and variance t. ▲

The previous theorem is often called the one dimension Lévy theorem while for apparent reasons the following one is called the two dimension Lévy theorem.

Theorem 5.4 Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space and $\{\mathcal{F}(t)\}_{t\geq 0}$ be a filtration in it. Let $M_1(t)$ and $M_2(t)$, $t\geq 0$, be martingales relative to the filtration $\mathcal{F}(t)$. Assume that for $i=1,2$, we have:

$$M_i(0)=0 \ , \ M_i(t) \text{ has continious paths} , \ [M_i, M_i](t)=t$$

$$\text{and } \ [M_1, M_2](t) = 0 \quad \text{for all } t \geq 0 ,$$

then $M_1(t)$ and $M_2(t)$ are independent Brownian motions

Proof: Since for each $i=1,2$ we have

$$M_i(0) \ , \ M_i(t) \text{ has continuous paths} , \ [M_i, M_i](t) = t$$

due to the one dimension Lévy theorem we have that

$$M_1(t) \text{ and } M_2(t) \text{ are Brownian motions.}$$

Let $f(t, M_1(t), M_2(t))$ be any function f for which all possible partial differcial derivatives exist and are continuous then from the two-dimention Ito-Doeblin formula (Theorem 5.2) and the the fact that:

$$[M_i, M_i](t) = t \ \text{ for each } i=1,2 \text{ and } [M_1, M_2](t) = 0 \text{ for all } t\geq 0,$$

equivalently could be written in differential form as:

$$dM_1(t)\, dM_1(t) = dt \ , \ dM_2(t)\, dM_2(t) = dt \ \text{ and } \ dM_1(t)\, dM_2(t) = 0 ,$$

we have that

$$d f(t, M_1(t), M_2(t)) = f_t\, dt + f_{M_1} dM_1(t) + f_{M_2} dM_2(t) + \frac{1}{2} f_{M_1M_1} dt + \tfrac{1}{2} f_{M_2M_2} dt .$$

We integrade both sides to obtain.

$$f(t, M_1(t), M_2(t)) = f(0, M_1(0), M_2(0) +$$

$$+ \int_0^t \Big[f_t(s, M_1(s), M_2(s)) + \frac{1}{2} f_{M_1M_1}(s, M_1(s), M_2(s))$$

$$+ \frac{1}{2} f_{M_2M_2}(s, M_1(s), M_2(s)) \Big] ds +$$

$$+\int_0^t f_{M_1}(s, M_1(s), M_2(s))\,dM_1(s) + \int_0^t f_{M_2}(s, M_1(s), M_2(s))\,dM_2(s).$$

We take expectations on both sides of the above equation. Since $M_1(t)$ and $M_2(t)$ by hypothesis are martingales we have that

$$E\left[\int_0^t f_{M_1}(s, M_1(s), M_2(s))\,dM_1(s)\right] = 0 \quad \text{and}$$

$$E\left[\int_0^t f_{M_2}(s, M_1(s), M_2(s))\,dM_2(s)\right] = 0 ,$$

therefore

$$E\left[f(t, M_1(t), M_2(t))\right] = f(0, M_1(0), M_2(0)) + E\left[\int_0^t \left\{ f_t(s, M_1(s), M_2(s)) + \right.\right.$$

$$\left.\left. + \frac{1}{2} f_{M_1M_1}(s, M_1(s), M_2(s)) + \frac{1}{2} f_{M_2M_2}(s, M_1(s), M_2(s)) \right\} ds \right]$$

Choose the function $f(t, M_1(t), M_2(t))$ to be

$$f(t, M_1(t), M_2(t)) = \exp\left\{ u_1 M_1(t) + u_2 M_2(t) - \frac{1}{2}(u_1^2 + u_2^2) \right\} \quad \text{for } u_1, u_2 \in \mathbb{R},$$

Then it is easy to see that

$$f_t(t, M_1(t), M_2(t)) = -\frac{1}{2}(u_1^2 + u_2^2) f(t, M_1(t), M_2(t)) \; ; \; f_{M_1}(t, M_1(t), M_2(t)) = u_1 f(t, M_1(t), M_2(t))$$

$$f_{M_2}(t, M_1(t), M_2(t)) = u_2 f(t, M_1(t), M_2(t)) , \quad f_{M_1M_1}(t, M_1(t), M_2(t)) = u_1^2 f(t, M_1(t), M_2(t)),$$

$$f_{M_2M_2}(t, M_2(t), M_2(t)) = u_2^2 f(t, M_1(t), M_2(t)) .$$

Thus, we obtain

$$E\left[\exp\left\{ u_1 M_1(t) + u_2 M_2(t) - \frac{1}{2}(u_1^2 + u_2^2) \right\}\right] = 1$$

from which we get that

$$E\left[\exp\{u_1 M_1(t) + u_2 M_2(t)\}\right] = \exp\{\tfrac{1}{2} u_1^2\} \exp\{\tfrac{1}{2} u_2^2\} .$$

Hence, the joint moment-generating function factors into the product of moment-generating functions, $M_1(t)$ and $M_2(t)$ must be independent. ▲

Although we have defined a multidimensional Brownian motion to be a vector of independent one dimensional Brownian motions, we shall see in what follows how to build correlated Brownian motions from this. More specifically we shall see how to build a model for two asset prices that both follow the geometric Brownian motion but their respective Wiener processes are correlated. In this respect let $S_1(t)$ and $S_2(t)$ be the stochastic processes which respresent two different asset prices which are such that we know they are correlated. Suppose that

$$dS_1(t) = \mu_1 S_1(t)\,dt + \sigma_1 S_1(t)\,dW_1(t),$$

where $W_1(t)$ is a Wiener process, i.e $S_1(t)$ is modeled as a geometric Brownian motion. Suppose also that

$$dS_2(t) = \mu_2 S_2(t)\,dt + \sigma_2 S_2(t)\left[\rho\, dW_1(t) + \sqrt{1-\rho^2}\, dW_2(t)\right],$$

where $W_2(t)$ is a Wiener process independent of $W_1(t)$ and $\sigma_1>0$, $\sigma_2>0$, and $-1\leq\rho\leq 1$. We will now prove that

$$W_3(t) = \rho\, W_1(t) + \sqrt{1-\rho^2}\, W_2(t),$$

is a Wiener process, which in fact states that $S_2(t)$ is also modeled as a geometric Brownian motion and in addition $W_1(t)$ and $W_3(t)$ have correlation ρ.

Apparently, $W_1(0)=0$ and $W_2(0)=0$ and has continuous paths, thus $W_3(0)=0$ and have continuous paths. In addition

$$\begin{aligned} dW_3(t)\,dW_3(t) &= \rho^2\, dW_1(t)\,dW_1(t) + 2\rho\sqrt{1-\rho^2}\, dW_1(t)\,dW_2(t) \\ &\quad + (1-\rho^2)\,dW_2(t)\,dW_2(t). \\ &= \rho^2 dt + (1-\rho^2)\,dt = dt. \end{aligned}$$

Consequently,

$$[W_3(t), W_3(t)] = t,$$

and according to Lévy one-dimension Theorem $W_3(t)$ is a Wiener process. Thus if we write

$$dS_2(t) = \mu_2 S_2(t)\,dt + \sigma_2 S_2(t)\,dW_3(t),$$

which in fact shows that $S_2(t)$ is modeled as a geometric Brownian motion.

We will now show that $W_1(t)$ and $W_2(t)$ are corellated with corellation coefficient ρ.

By the Itô's product we have

$$d(W_1(t)W_3(t)) = W_1(t)dW_3(t) + W_3(t)dW_1(t) + dW_1(t)dW_3(t)$$

$$= W_1(t)dW_3(t) + W_3(t)dW_1(t) + \rho dt.$$

Integrating both sides of the above we get

$$W_1(t)W_3(t) = \int_0^t W_1(t)dW_3(t) + \int_0^t W_3(t)dW_1(t) + \rho t.$$

We take expectations in both sides of the above equation to obtain

$$E[W_1(t)W_3(t)] = \rho t$$

Because both $W_1(t)$ and $W_3(t)$ have standard deviation $\sqrt{t}$, the number ρ is the correlation between $W_1(t)$ and $W_3(t)$.

REFERENCES

1. Karatzas, I. and Shreve, S.E (1991). Methods of Mathematical Finance. Springer.
2. Karatzas, I and Shreve, S.E (1991). Brownian motion and Stochastic Calculus Springer.
3. Klebaner, F. (2006). Introduction to Stochastic Calculus with Applications. Second Ed. Imperial College Press.
4. Hui-Hsiung Kuo. (2000). Introduction to Stochastic Integration. Springer.
5. Protter, P. (2003). Stochastic Integration and Differential Equations. Second Ed. Springer.
6. Oksendal, B (2005). Stochastic Differential Equations. Sixth Edition Springer.
7. Shreve, S.E. (2004). Stochastic Calculus for Finance II. Continuous-Time Models. Springer.

Chapter 6.

Risk - Neutral (Martingale) Evaluation and Hedging of Vanila Derivatives

- Girsanov's and Martingale Representation Theorem.
- Fundamental Theorems of asset pricing.
- The Black - Scholes - Merton Formula.
- Forwards and Futures.

6.1. Introduction.

In this chapter we establish the famous partial differential equation of Black - Scholes and Merton and we provide a clever way to solve it using an equivalent to the real world probability measure called the risk-neutral or martingale measure. After we provide a way for hedging the short position of a European Derivative.

These steps rely on Girsanov's theorem which provides a method to construct the risk-neutral or martingale measure in a model with a single underlying security. Risk - neutral pricing is a powerful method for computing prices of derivative securities, but is fully justified only when it is accompanied by a hedge for a short position in the security being priced. This will be done with the use of the martingale representation theorem. Furthermore, we provide conditions that guarantee that such models does not admit arbitrage and that every derivative security in the model can be hedged.

6.2 Risk Neutral or Martingale Measure

It is really useful for the reader to refresh his understanding of change of measure by reading once more section 2.8. In the present section, we will study a similar change of measure in order to change the mean, but this time for a whole process rather than for a single random variable.

Assume that we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\{\mathcal{F}(t)\}_{t\geq 0}$ for the interval $0 \leq t \leq T$, where T is a fixed final time.

In addition let Z be an almost surely positive random variable with $E[Z]=1$ and let $\tilde{\mathbb{P}}$ be an equivalent to $\mathbb{P}$ probability measure defined by

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega)\, d\mathbb{P}(\omega), \quad \text{for all } A \in \mathcal{F}$$

or in differential form

$$\frac{d\tilde{\mathbb{P}}(\omega)}{d\mathbb{P}(\omega)} = Z(\omega) \quad ,$$

i.e. Z is the Radon-Nikodým derivative of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$.

We now define the Radon-Nikodým derivative process to be

$$Z(t) = \mathbf{E}\left[Z \,|\, \mathcal{F}(t) \right] \quad , \quad 0 \leq t \leq T.$$

The Radon-Nikodým derivative process is a martingale. Actually

$$E[Z(t) \,/\, \mathcal{F}(s)] = E\left[E\left[Z \,|\, \mathcal{F}(t) \right] / \mathcal{F}(s) \right] = E\left[Z \,|\, \mathcal{F}(s) \right] = Z(s).$$

We will now examine some of the properties of the above defined Radon-Nikodým derivative processes.

<u>Proposition 6.1</u>. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}(t)\}_{t\geq 0}$ be a

filtration, $Z(t)$ the Radon-Nikodým derivative as defined above and finally let Y be an $\mathcal{F}(t)$-measurable random variable. Then

$$\mathbb{E}_{\tilde{\mathbb{P}}}[Y] = \mathbb{E}_{\mathbb{P}}[Z(t)Y].$$

Proof. From section 2.8 we have

$$\mathbb{E}_{\tilde{\mathbb{P}}}[Y] = \mathbb{E}_{\mathbb{P}}[YZ] = \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}}[YZ]/\mathcal{F}(t)\right] =$$

$$= \left(Y \text{ is an } \mathcal{F}(t) \text{ measurable random variable}\right)$$

$$= \mathbb{E}_{\mathbb{P}}\left[Y\,\mathbb{E}_{\mathbb{P}}[Z/\mathcal{F}(t)]\right] =$$

$$= \left(\mathbb{E}[Z|\mathcal{F}(t)] = Z(t) \text{ by definition}\right)$$

$$= \mathbb{E}_{\mathbb{P}}[Y\,Z(t)]$$

▲

Proposition 6.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}(t)\}_{t\geq 0}$ be a filtration, $Z(t)$ the Radon-Nikodým derivative as defined above and Y be an $\mathcal{F}(t)$-measurable random variable. Let also that $0\leq s\leq t\leq T$ then

$$\mathbb{E}_{\tilde{\mathbb{P}}}[Y|\mathcal{F}(s)] = \frac{1}{Z(s)}\,\mathbb{E}_{\mathbb{P}}[YZ(t)/\mathcal{F}(s)]$$

Proof: By definition 2.16 we have to prove that

$$\frac{1}{Z(s)}\,\mathbb{E}_{\mathbb{P}}[YZ(t)/\mathcal{F}(s)] \text{ is } \mathcal{F}(s)\text{-measurable}$$

since then $\mathbb{E}_{\tilde{\mathbb{P}}}[Y/\mathcal{F}(s)]$ satisfies the first condition of the definition of a conditional expectation However, this is apparent since $\mathbb{E}_{\mathbb{P}}[YZ(t)/\mathcal{F}(s)]$ is a conditional expectation and thus it is $\mathcal{F}(s)$-measurable.

Now we must prove the condition of partial average i.e

$$\int_A \mathbb{E}_{\tilde{\mathbb{P}}}[Y/\mathcal{F}(s)]\,d\tilde{\mathbb{P}} = \int_A \frac{1}{Z(s)}\,\mathbb{E}_{\mathbb{P}}[YZ(t)/\mathcal{F}(s)]\,d\tilde{\mathbb{P}} = \int_A Y\,d\tilde{\mathbb{P}} \quad \text{for all } A\in\mathcal{F}.$$

In this respect we have that

$$\int_A \frac{1}{Z(s)}\,\mathbb{E}_{\mathbb{P}}[YZ(t)/\mathcal{F}(s)]\,d\tilde{\mathbb{P}} = \mathbb{E}_{\tilde{\mathbb{P}}}\left[\mathbb{I}_A \frac{1}{Z(s)}\,\mathbb{E}_{\mathbb{P}}[YZ(t)/\mathcal{F}(s)]\right] =$$

$$= (\text{By Proposition } 6.1) =$$

$$= \mathbb{E}_{\mathbb{P}}\left[\mathbb{I}_A \, \mathbb{E}_{\mathbb{P}}[Y Z(t)/\mathcal{F}(s)]\right] = (A \in \mathcal{F}(s)) =$$

$$= \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}}[\mathbb{I}_A Y Z(t)/\mathcal{F}(s)]\right] =$$

$$= \mathbb{E}_{\mathbb{P}}[\mathbb{I}_A Y Z(t)] = (\text{By Proposition } 6.1) =$$

$$= \mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{I}_A Y] = \int_A Y d\tilde{\mathbb{P}} .$$

▲

We now provide the one-dimension Girsanov's Theorem which plays a vital role in what follows:

<u>Theorem 6.1</u> (Girsanov, one dimension). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\{\mathcal{F}(t)\}_{t\geq 0}$ be a filtration, in the interval $0 \leq t \leq T$, $\{W(t)\}_{t\geq 0}$ a Wiener process adapted to $\mathcal{F}(t)$ and $\Theta(t)$ be a stochastic process adapted to $\mathcal{F}(t)$ for $0 \leq t \leq T$. Define

$$Z(t) = \exp\left\{-\int_0^t \Theta(s) dW(s) - \frac{1}{2}\int_0^t \Theta^2(s) ds\right\}$$

$$\tilde{W}(t) = W(t) + \int_0^t \Theta(s) ds .$$

and assume that

$$\mathbb{E}\left[\int_0^T \Theta^2(s) Z^2(s) \, ds\right] < \infty .$$

Then $\mathbb{E}[Z(T)] = 1$ and under the probability measure

$$\tilde{\mathbb{P}}(A) = \int_A Z(T)(\omega) \, d\mathbb{P}(\omega) ,$$

the process

$$\tilde{W}(t) , \quad 0 \leq t \leq T, \text{ is a Brownian motion.}$$

<u>Proof</u>: We have that $\tilde{W}(t)$ in differential form could be written as

$$d\tilde{W}(t) = dW(t) + \Theta(t) dt ,$$

hence,

$$d\tilde{W}(t)\, d\tilde{W}(t) = (dW(t) + \Theta(t)dt)^2 = dW(t)\, dW(t) = dt.$$

That means that $[\tilde{W}(t), \tilde{W}(t)] = t$. Also $\tilde{W}(0) = W(0) = 0$ and since $\tilde{W}(t)$ has continuous paths according to Levý's Theorem in order to prove that $\tilde{W}(t)$ is a Brownian motion it remains to prove that $\tilde{W}(t)$ is a martingale, under the probability measure $\tilde{\mathbb{P}}$.

We apply the Itô-Doeblin formula to $Z(t)$ to obtain

$$dZ(t) = -\Theta(t)\,Z(t)\,dW(t).$$

We integrate both sides of the above equation to obtain

$$Z(t) = Z(0) - \int_0^t \Theta(s) Z(s)\,dW(s).$$

By hypothesis the Itô integral above exists and is a martingale. Hence $Z(t)$ is also a martingale.

Since $Z(t)$ is a martingale we have that

$$E[Z(T)] = Z(0) = 1.$$

and

$$Z(t) = E[Z(T)/\mathcal{F}(t)] \text{ with } E[Z(T)] = 1,$$

which shows that $Z(t)$ as defined in the hypothesis is Radon-Nikodým derivative process.

We will now show that $\widetilde{W}(t)Z(t)$ is a martingale under $\widetilde{\mathbb{P}}$. We use Itô's product rule to obtain the differential form:

$$\begin{aligned}
d(\widetilde{W}(t)\,Z(t)) &= \widetilde{W}(t)\,dZ(t) + Z(t)\,d\widetilde{W}(t) + dZ(t)\,d\widetilde{W}(t) \\
&= -\widetilde{W}(t)\Theta(t)Z(t)\,dW(t) + Z(t)\,dW(t) + Z(t)\Theta(t)\,dt \\
&\quad + \left(dW(t) + \Theta(t)\,dt\right)\left(-\Theta(t)Z(t)\,dW(t)\right) \\
&= \left(1 - \widetilde{W}(t)\Theta(t)\right)Z(t)\,dW(t)
\end{aligned}$$

If we integrate both sides of the above equation we obtain

$$\widetilde{W}(t)\,Z(t) = \int_0^t \left(1 - \widetilde{W}(s)\Theta(s)\right)Z(s)\,dW(t).$$

The right hand side of the above equation is an Itô integral and thus it is a martingale, hence $\widetilde{W}(t)Z(t)$ is a martingale under the martingale measure $\widetilde{\mathbb{P}}$.

From proposition 6.2 we get for $0 \le s \le t \le T$:

$$E_{\tilde{\mathbb{P}}}\left[\tilde{W}(t)/\mathcal{F}(t)\right] = \frac{1}{Z(s)} E_{\mathbb{P}}\left[\tilde{W}(t)\,Z(t)/\mathcal{F}(s)\right] =$$
$$= \left(\tilde{W}(t)Z(t) \text{ is a martingale under } \mathbb{P}\right)$$
$$= \frac{1}{Z(s)} \cdot \tilde{W}(s)\,Z(s) = \tilde{W}(s),$$

thus $\tilde{W}(t)$ is a martingale under the probability measure $\tilde{\mathbb{P}}$. ▲

6.3. Evaluation of the Stock and portofolio process under the Risk-Neutral Measure.

Let that we have a market which consists of a money market account and an asset. Also let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration $\{\mathcal{F}(t)\}_{t\geq 0}$ in the interval $0 \leq t \leq T$, with T a fixed final time.

Let $\{r(t)\}_{t\geq 0}$ be the interest rate process of the money market account which is assumed to be adapted to the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$. It is very well known that if at time 0 we invest in a share of this money market account with price 1 then at time t the price of that share will be

$$e^{\int_0^t r(s)ds}.$$

We define the discount process to be

$$D(t) = \exp\left\{-\int_0^t r(s)ds\right\}.$$

or in differential form, using the Itô-Doeblin formula

$$dD(t) = -r(t)\,D(t)\,dt.$$

Also it is easy to see that

$$dD(t)\,dD(t) = 0,$$

which means that although $D(t)$ is random it's quadratic variation

$$[D(t), D(t)](t) = 0.$$

Now, let $\{S(t)\}_{t\geq 0}$ be the price of the asset of the market which we assume to be a generalized geometric Brownian motion which in differential form could be written as:

$$dS(t) = \mu(t)\, S(t)\, dt + \sigma(t)\, S(t)\, dW(t),$$

where $\mu(t)$ the mean return and $\sigma(t)$ the volatility are adapted processes to the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$. We assume that, for all $t \in [0,T]$, $\sigma(t)$ is almost surely not zero. It is known that a close analytic solution for $S(t)$ is given by

$$S(t) = S(0) \exp\left\{ \int_0^t \sigma(s)\, dW(s) + \int_0^t \left(\mu(s) - \tfrac{1}{2}\sigma^2(s)\right) ds \right\}$$

Now we will find the quadratic variation of $S(t)$:

$$\begin{aligned} dS(t)\, dS(t) &= \mu^2(t)\, S^2(t)\, dt\, dt + \sigma^2(t)\, S^2(t)\, dW(t)\, dW(t) \\ &\quad + 2\,\mu(t)\,\sigma(t)\, S^2(t)\, dW(t)\, dt \\ &= \sigma^2(t)\, S^2(t)\, dt, \end{aligned}$$

which shows that $S(t)$ has quadratic variation almost surely positive. Hence, unlike the price of the money market account, the stock price is susceptible to instantaneous unpredictable changes and is, in this sence, "more random" than $D(t)$.

The discounted asset price process is defined to be $\{D(t)\, S(t)\}_{t\geq 0}$ the physical meaning of which is apparent. Its differential form is easily found by applying Itô's product rule

$$\begin{aligned} d(D(t)\, S(t)) &= D(t)\, dS(t) + S(t)\, dD(t) + dD(t)\, dS(t) \\ &= D(t)\left[\mu(t)\, S(t)\, dt + \sigma(t)\, S(t)\, dW(t)\right] + S(t)\left[-r(t)\, D(t)\right] dt \\ &\quad + \left[-r(t)\, D(t)\right] \cdot \mu(t)\, S(t)\, dt\, dt + \left[-r(t)\, D(t)\right] \sigma(t)\, S(t)\, dW(t)\, dt \\ &= D(t)\,\mu(t)\, S(t)\, dt + \sigma(t)\, D(t)\, S(t)\, dW(t) - r(t)\, S(t)\, D(t)\, dt \\ &= \left[\mu(t) - r(t)\right] D(t)\, S(t)\, dt + \sigma(t)\, D(t)\, S(t)\, dW(t). \end{aligned}$$

Now define the market price of risk to be

$$\Theta(t) = \frac{\mu(t) - R(t)}{\sigma(t)}.$$

then the differential of the discounted stock price process could be written as:

$$d(D(t)S(t)) = \sigma(t)D(t)S(t)\left[\Theta(t)dt + dW(t)\right]$$

Now having in mind Girsanov's, one dimension theorem we define by

$$d\widetilde{W}(t) = \Theta(t)\,dt + dW(t).$$

then since $\Theta(t)$ is adapted to the filtration $\mathcal{F}(t)$, $d\widetilde{W}(t)$ is a Brownian motion under the the probability measure $\widetilde{\mathbb{P}}$, which is given by

$$\widetilde{\mathbb{P}}(A) = \int_A Z(T)(\omega)\,d\mathbb{P}(\omega) \quad \text{for all } A \in \mathcal{F},$$

where the Radon-Nikodým process $Z(t)$ is given by

$$Z(t) = \exp\left\{-\int_0^t \Theta(s)dW(s) - \tfrac{1}{2}\cdot\int_0^t \Theta^2(s)\,ds\right\}$$

with the assumption that

$$E\left[\int_0^T \Theta^2(s)Z^2(s)\,ds\right] < \infty.$$

We call the probability measure $\widetilde{\mathbb{P}}$ the risk-neutral probability measure or the martingale probability measure. The name ~~mane~~ martingale probability measure is due to the fact that under $\widetilde{\mathbb{P}}$ the discounted asset price process is a martingale. To prove that we write

$$d(D(t)S(t)) = \sigma(t)D(t)S(t)\,d\widetilde{W}(t).$$

Integrating both parts of the above equation we get

$$\begin{aligned} D(t)S(t) &= D(0)S(0) + \int_0^t \sigma(s)D(s)S(s)\,d\widetilde{W}(s) \\ &= S(0) + \int_0^t \sigma(s)D(s)S(s)\,d\widetilde{W}(s), \end{aligned}$$

where the integral on the right hand side of the equation is an Itô integral under the martingale measure $\widetilde{\mathbb{P}}$ and thus a martingale under $\widetilde{\mathbb{P}}$. Hence the discounted asset process is a martingale under the risk-neutral probability measure - or martingale probability measure.

The asset price $\{S(t)\}_{t\geq 0}$ under the martingale probability measure is written as a stochastic differential equation:

$$dS(t) = r(t)S(t)dt + \sigma(t)S(t)\,d\widetilde{W}(t).$$

Observe that in continuous time, the change from the actual probability measure, to the risk-neutral measure $\tilde{\mathbb{P}}$ changes the mean return rate to $r(t)$. However the volatility remains the same.

Now consider the stochastic process $\{X(t)\}_{t\geq 0}$ to be the price of a portofolio which consists of $\Delta(t)$ pieces of the asset and the money market account from which we borrow or invest with interest rate process $r(t)$. The in the time interval $[t, t+dt]$ the changes in the price of the portofolio come from two sources: a) The difference in the value of the $\Delta(t)$ pieces of the asset which is apparent given by

$$\Delta(t)\, dS(t)$$

b). The interest on the amount $(X(t)-\Delta(t)S(t))$ that we had to borrow (or invest) in order to buy $\Delta(t)$ pieses's of the asset at time t given the price of the portofolio at time t being $X(t)$. Note that if $(X(t)-\Delta(t)S(t))>0$ we assume that we invest this amount in the market money accoent with interest rate $r(t)$ which is assumed to be the same with borrowing in case $(X(t)-\Delta(t)S(t))<0$.

Hence we have that the differential $dX(t)$ of the price of the portofolio will be

$$dX(t) = \Delta(t)\, dS(t) + r(t)\left(X(t)-\Delta(t)S(t)\right)dt.$$

Now assuming that $S(t)$ is the generalized geometric Brownian motion we get that

$$\begin{aligned} dX(t) &= \Delta(t)\left(\mu(t)S(t)\,dt + \sigma(t)S(t)\,dW(t)\right) + r(t)\left(X(t)-\Delta(t)S(t)\right)dt \\ &= r(t)X(t)\,dt + \Delta(t)\left(\mu(t)-r(t)\right)S(t)\,dt + \sigma(t)\Delta(t)S(t)\,dW(t) \\ &= r(t)X(t)\,dt + \Delta(t)\sigma(t)S(t)\left[\Theta(t)\,dt + dW(t)\right]. \end{aligned}$$

Now similarly as previously we define

$$d\tilde{W}(t) = \Theta(t)\,dt + dW(t)$$

where $d\tilde{W}(t)$ according to Girsanov's one dimension theorem is a Brownian is a Brownian motion under the martingale probability measure. Then

$$dX(t) = r(t)X(t)\,dt + \Delta(t)\sigma(t)S(t)\,d\tilde{W}(t).$$

Consider now the discounted price of the portofolio process $D(t)X(t)$ then by Itô's product rule we have

$$d(D(t)X(t)) = D(t)\,dX(t) + X(t)\,dD(t) + dX(t)\,dD(t)$$

$$= D(t)\left[r(t)X(t)\,dt + \Delta(t)\sigma(t)S(t)\,d\tilde{W}(t)\right] + X(t)\left[-R(t)D(t)\right]dt$$

$$+ \left[r(t)X(t)dt + \Delta(t)\sigma(t)S(t)\,d\tilde{W}(t)\right]\cdot\left[-R(t)D(t)\right]dt$$

$$= \Delta(t)\sigma(t)D(t)S(t)\,d\tilde{W}(t).$$

Integrating both sides of the above equation

$$D(t)X(t) = X(0) + \int_0^t \Delta(t)\sigma(t)D(t)S(t)\,d\tilde{W}(t).$$

Since the above integral is an Itô integral under $\tilde{\mathbb{P}}$, then it is a martingale under $\tilde{\mathbb{P}}$ and thus the discounted portofolio price process is a martingale under the martingale probability measure $\tilde{\mathbb{P}}$.

6.4 European Call Option

A European Call option is a derivative security which is written on an asset. It is an agreement signed at time 0 which allows the holder the right to by at the expiration time $T>0$ the unit of the asset at the strike price $K \geq 0$.

The underlying asset it is possible to be anything that may be traded in the market. At the expiration time T the value of the European call option is

$$(S(T)-K)^+ = \max\left(S(T)-K, 0\right),$$

in the sence that if $(S(T)-K)>0$ then the holder of the European call option acquires the unit of the asset at time T at the price K while its actual price $S(T)>K$ and thus gains $(S(T)-K)$. Conversely, if $(S(T)-K)<0$ then he/she is not oblidged to buy.

Let $V(T)$ be an $\mathcal{F}(T)$-measurable random variable that represents at time T the payoff of a derivative security in this case the European call option. An investment cooperation which trades derivative securities will create a portofolio $\{X(t)\}_{t\geq 0}$ with initial value $X(0)$, the amount of money that we sold the derivative security at time 0, and the underlying asset with price $\{S(t)\}_{t\geq 0}$, holding at time t, $\Delta(t)$, $(0\leq t\leq T)$ units of the asset. With this strategy wishes to hedge a short position in this derivative security i.e.

$$X(T) = V(T) \text{ almost surely.}$$

Two are the important questions.

a) What should be the price, $[X(0)=V(0)]$, of the derivative security

b) What is the hedging strategy $\Delta(t)$ such that $X(T)=V(T)$ almost surely.

We shall see in what follows that these questions are possible to be answered. Once they are answered the fact that $D(t)X(t)$ is a martingale under $\widetilde{\mathbb{P}}$ implies

$$D(t)X(t) = E_{\widetilde{\mathbb{P}}}[D(T)X(T)/\mathcal{F}(t)] = \widetilde{E}_{\widetilde{\mathbb{P}}}[D(T)V(T)/\mathcal{F}(t)]$$

The value $X(t)$ of the hedging portofolio in the above equation is the capital needed at time t in order to succesfully complete the hedge of the short position in the derivative security with payoff $V(T)$. Hence, we can call this the price $V(t)$ of the derivative security at time t i.e.

$$D(t)V(t) = E_{\widetilde{\mathbb{P}}}[D(T)V(T)/\mathcal{F}(t)], \quad 0\leq t\leq T.$$

Now $D(t)$ is apparently $\mathcal{F}(t)$-measurable and hence

$$V(t) = E_{\widetilde{\mathbb{P}}}\left[\exp\left\{-\int_t^T R(s)ds\right\} \cdot V(T)/\mathcal{F}(t)\right], \quad 0\leq t\leq T.$$

We shall refer to the above equation as the risk-neutral pricing formula.

6.5. The Black-Scholes-Merton Equation.

In chapter 1 we presented the history of the Black-Scholes-Merton equation which revolutionized the financial world. The conditions of this problem are the following

- We are dealing with a European call option with an underlying asset whose price is represented by the stochastic process $\{S(t)\}_{t\geq 0}$.
- The interest rate process is a constant r.
- The price of the undelying asset of the European call option is assumed to be modeled by the geometrical Brownian motion

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t).$$

- The European call option is written at zero, has expiration date T and strike price K. Hence, the payoff at time T for the European call option will be

$$V(T) = [S(T)-K]^+$$

The price $V(t)$ of the European call option at time t will be according to the previous section

$$V(t) = E_{\tilde{P}}\left[e^{-r(T-t)}(S(T)-K)^+/\mathcal{F}(t)\right]$$

where $\tilde{P}$ is the risk-neutral measure or martingale probability measure. From the bottom of page 6.8 we know that under the risk neutral measure $S(t)$ is the solution of the stochastic differential equation

$$dS(t) = r(t)S(t)dt + \sigma(t)S(t)d\tilde{W}(t)$$

which under the conditions of the Black-Scholes-Merton model becomes

$$dS(t) = rS(t)dt + \sigma S(t)d\tilde{W}(t).$$

We solve the above stochastic differential equation to obtain

$$S(t) = S(0)\exp\left\{\sigma \tilde{W}(t) + \left(r - \frac{1}{2}\sigma^2\right)t\right\}$$

which lead us to write

$$S(T) = S(t)\exp\left\{\sigma\left(\tilde{W}(T) - \tilde{W}(t)\right) + \left(r - \tfrac{1}{2}\sigma^2\right)(T-t)\right\}$$

It is known that the distribution of $\tilde{W}(T) - \tilde{W}(t)$ is normal with mean zero and variance $T-t$ and that it is independent of $\mathcal{F}(t)$. We introduce the standard normal variable

$$Y = -\frac{\tilde{W}(T) - \tilde{W}(t)}{\sqrt{T-t}},$$

then, we write

$$S(T) = S(t)\exp\left\{-\sigma\sqrt{T-t}\,Y + \left(r - \tfrac{1}{2}\sigma^2\right)(T-t)\right\},$$

where, since Y is independent of $\mathcal{F}(t)$ then the random variable

$$\exp\left\{-\sigma\sqrt{T-t}\,Y + \left(r - \tfrac{1}{2}\sigma^2\right)(T-t)\right\} \text{ is independent of } \mathcal{F}(t)$$

Hence, for any $0 \le t \le T$ we have

$$E_{\tilde{\mathbb{P}}}\left[e^{-r(T-t)}\left(S(T) - K\right)^+ / \mathcal{F}(t)\right] = \left(Y \text{ is independent of } \mathcal{F}(t), \text{ } t\text{ is } \mathcal{F}(t)\text{-measurable}\right)$$

$$= E_{\tilde{\mathbb{P}}}\left[e^{-r(T-t)}\left(S(t)\exp\left\{-\sigma\sqrt{T-t}\,Y + \left(r - \tfrac{1}{2}\sigma^2\right)(T-t)\right\} - K\right)^+\right]$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-r(T-t)}\left(S(t)\exp\left\{-\sigma\sqrt{T-t}\,y + \left(r - \tfrac{1}{2}\sigma^2\right)(T-t)\right\} - K\right)^+ e^{-\frac{1}{2}y^2}\,dy.$$

Apparently we need to find for which values of y in $[-\infty, \infty]$ the integrand is positive which is equivalent in finding for which values of y the function

$$S(t)\exp\left\{-\sigma\sqrt{T-t}\,y + \left(r - \tfrac{1}{2}\sigma^2\right)(T-t)\right\} - K > 0$$

It is easy to see that this is so if and only if

$$y < \frac{1}{\sigma\sqrt{T-t}}\left[\log\frac{S(t)}{K} + \left(r - \tfrac{1}{2}\sigma^2\right)(T-t)\right] = d_-(T-t, S(t)).$$

Therefore, we get

$$V(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_-(T-t,S(t))} e^{-r(T-t)} \left[S(t) \exp\{-\sigma\sqrt{T-t}\, y + (r - \tfrac{1}{2}\sigma^2)(T-t)\} - K \right] e^{-\frac{1}{2}y^2} dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_-(T-t,S(t))} S(t) \exp\left\{ -\frac{y^2}{2} - \sigma\sqrt{T-t}\, y - \frac{\sigma^2(T-t)}{2} \right\} dy$$

$$- \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_-(T-t,S(t))} e^{-r(T-t)} K e^{-\frac{1}{2}y^2} dy$$

$$= \frac{S(t)}{\sqrt{2\pi}} \int_{-\infty}^{d_-(T-t,S(t))} \exp\left\{ -\frac{1}{2}\left(y + \sigma\sqrt{T-t}\right)^2 \right\} dy - e^{-r(T-t)} K N(d_-(T-t,S(t)))$$

$$= \frac{S(t)}{\sqrt{2\pi}} \int_{-\infty}^{d_-(T-t,S(t)) + \sigma\sqrt{T-t}} \exp\left\{ -\frac{1}{2} z^2 \right\} dz - e^{-r(T-t)} K N(d_-(T-t,S(t)))$$

$$= S(t)\, N(d_+(T-t,S(t))) - e^{r(T-t)} K N(d_-(T-t,S(t))),$$

where

$$d_+(T-t,S(t)) = d_-(T-t,S(t)) + \sigma\sqrt{T-t} =$$

$$= \frac{1}{\sigma\sqrt{T-t}} \left[\log\frac{S(t)}{K} + (r + \tfrac{1}{2}\sigma^2)(T-t) \right]$$

For future reference we denote the payoff price of a European call option under the model of Black - Scholes and Merton by $BSM(T-t, S(t); K, r, \sigma)$ and we have just shown that

$$BSM(T-t, S(t); K, r, \sigma) = S(t)\, N(d_+(T-t,S(t))) - e^{-r(T-t)} K N(d_-(T-t,S(t))).$$

where $N(x)$ denotes the value of the standard Normal distribution at x

6.6. Martingale Representation Theorem - Hedging Strategy.

In section 6.4 we answered the first important question in derivative securities, that is which is the value $V(0)$ which will work us an intial capital in order to create a portofolio $\{X(t)\}_{t\geq 0}$. This portofolio will be builded by holding at any time $t \in [0,T]$, $\Delta(t)$ units of the underlying asset whose price is $S(t)$ and the money market accoument with interest rate $r(t)$. The goal of this portofolio is to have

$$X(T) = V(T) \quad \text{almost surely}$$

The existence of such a hedging portofolio in the model with one underlying asset and one source of uncertainty driven by one Brownian motion is closely depended on the following theorem, which we state without proof.

<u>Theorem 6.2. (Martingale representation, one dimension).</u> Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, $W(t)$, $0 \le t \le T$ be a Brownian motion on it, and let $\mathcal{W}(t)$ be the filtration generated by this Brownian motion. Let $M(t)$, $0 \le t \le T$, be a martingale with respect to $\mathcal{W}(t)$. Then there is an adapted process $\Gamma(t)$, $0 \le t \le T$, such that

$$M(t) = M(0) + \int_0^t \Gamma(s)\, dW(s) \quad , 0 \le t \le T.$$

<u>Remark 6.1.</u> The above theorem asserts the existence of a process $\Gamma(t)$ adapted to the filtration $\mathcal{W}(t)$ such that any martingale $M(t)$ with respect to the filtration $\mathcal{W}(t)$ i.e.

a) $M(t)$ is $\mathcal{W}(t)$ measurable

b) $E_P[M(t)/\mathcal{W}(s)] = M(s)$ for any $0 \le s \le t \le T$

Unfortunately there is not any general result how for a specific martingale we will find $\Gamma(t)$. This fact is our basic difficulty in finding the correct hedging strategy for any derivative security.

<u>Remark 6.2.</u> Theorem 6.2 assumes that there is one only source of uncertainty which is represented by the Brownian motion $W(t)$, $0 \le t \le T$. Very important is the restriction that $M(t)$ is a martingale with respect to $\mathcal{W}(t)$ and not a general filtration containing more information. Note at this point that this assumption is more restrictive than the assumption of Girsanov's theorem where the filtration can be larger than the one generated by

by the Brownian motion. If we include this extra restriction in Girsanov's theorem, then we obtain the following Corollary.

Corollary 6.1. Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $W(t)$ be a Brownian motion on this space. Let $\mathcal{W}(t)$ be the filtration generated by this Brownian motion. Let $\Theta(t)$, $0 \le t \le T$ be an adapted process to $\mathcal{W}(t)$ and define

$$Z(t) = \exp\left\{-\int_0^t \Theta(s)\,dW(s) - \tfrac{1}{2}\int_0^t \Theta^2(s)\,ds\right\}$$

$$\tilde{W}(t) = W(t) + \int_0^t \Theta(s)\,ds .$$

Assume that

$$E_{\tilde{\mathbb{P}}}\left[\int_0^T \Theta^2(s)\,Z^2(s)\,ds\right] < \infty ,$$

where $\tilde{\mathbb{P}}$ is the equivalent probability measure

$$\tilde{\mathbb{P}}(A) = \int_A Z(T)\,d\mathbb{P}(\omega) \quad \text{for any } A \in \mathcal{W}(t).$$

Then

$$E_{\tilde{\mathbb{P}}}[Z(T)] = 1$$

and the process $\tilde{W}(t)$ is a Brownian motion under $\tilde{\mathbb{P}}$.

Now let $\tilde{M}(t)$, $0 \le t \le T$ be a martingale under $\tilde{\mathbb{P}}$ with respect to $\mathcal{W}(t)$. Then there exists a process $\tilde{\Gamma}(t)$ adapted to the filtration $\mathcal{W}(t)$ such that

$$\tilde{M}(t) = \tilde{M}(0) + \int_0^t \tilde{\Gamma}(s)\,d\tilde{W}(s)$$

Remark 6.3.

Note that in here $\tilde{\Gamma}(t)$ is adapted to the filtration $\mathcal{W}(t)$ of the Brownian motion $W(t)$ and not to the filtration of the Brownian motion $\tilde{W}(t)$ as an immediate application of the Martingale Representation Theorem would suggest. The proof is left as an exercise to the reader.

Let us now go back in section 6.4 and assume that the general filtration $\{\mathcal{F}(t)\}_{t\geq 0}$ that is referred in there in the filtration generated by the Brownian motion $\{W(t)\}_{t\geq 0}$ i.e $\{\mathcal{W}(t)\}_{t\geq 0}$.

We assume that the adapted to the feltration $\mathcal{W}(t)$ stochastic volatility process $\sigma(t)$ of the generalized geometric Brownian motion is almost surely different than zero.

Recall (page 6.11) that the discounted payoff price of the European call is given by

$$D(t)V(t) = E_{\tilde{\mathbb{P}}}\left[D(T)V(T)/\mathcal{W}(t)\right] \;,\; 0\leq t\leq T.$$

Since we create our hedging portofolio in order to have $X(t)=V(t)$ almost surely and in page 6.10 we proved that that the discounted portofolio process $\{D(t)X(t)\}_{t\geq 0}$ is a martingale under $\tilde{\mathbb{P}}$, it is natural to expect that also $\{D(t)V(t)\}_{t\geq 0}$ is a martingale under $\tilde{\mathbb{P}}$. Actually, it easy to see that.

$$E_{\tilde{\mathbb{P}}}\left[D(t)V(t)/\mathcal{W}(s)\right] = E_{\tilde{\mathbb{P}}}\left[E_{\tilde{\mathbb{P}}}\left[D(T)V(T)/\mathcal{W}(t)\right]/\mathcal{W}(s)\right]$$

$$= E_{\tilde{\mathbb{P}}}\left[D(T)V(T)/\mathcal{W}(s)\right] = D(s)V(s).$$

Hence, from Corollary 6.1 we get that there exist an adapted to the filtration $\mathcal{W}(t)$ process $\tilde{\Gamma}(t)$ such that

$$D(t)V(t) = D(0)V(0) + \int_0^t \tilde{\Gamma}(s)\,d\tilde{W}(s) \;,\; 0\leq t\leq T.$$

$$= \left(\text{Since } D(0)=1\right) =$$

$$= V(0) + \int_0^t \tilde{\Gamma}(s)\,d\tilde{W}(s)$$

In page 6.10 we have proved that for any hedging portofolio $X(t)$ which is constructed with $\Delta(t)$ units of the undelying asset of the European call option we have that the discounted price $D(t)X(t)$ is given by

$$D(t)X(t) = X(0) + \int_0^t \Delta(s)\,\sigma(s)\,D(s)\,S(s)\,d\tilde{W}(s).$$

In order to have $X(t)=V(t)$ almost surely for all t starting with initial capital $X(0) = V(0)$, $\Delta(t)$ should be such as to satisfy the equation

$$\Delta(t)\sigma(t)D(t)S(t) = \tilde{\Gamma}(t) \quad \text{for all } 0 \le t \le T.$$

Since $\sigma(t) \neq 0$ almost surely for all t, the above equation is equivalent to

$$\Delta(t) = \frac{\tilde{\Gamma}(t)}{\sigma(t)D(t)S(t)}.$$

Which in practice means that at all times the investment bank should keep at hand the price of $\Delta(t)$ units of the underlying asset of the European call option for hedging it's short position.

In a market where a derivative security can be hedged i.e it is possible to find the hedging strategy $\Delta(t)$ or equivalently to create a replicating portofolio is said to be complete.

Note the that the two key assumptions that made possible to find the hedging strategy $\Delta(t)$ where the following

- The volatility process $\sigma(t)$ is different than zero almost surely for every t.
- The filtration $\{\mathcal{F}(t)\}_{t\ge0}$ which is assumed in all relevant theorems is the one generated only by the Brownian motion $W(t)$ which represents the source of uncertainthy in the price process $\{S(t)\}_{t\ge0}^{\infty}$ of the underlying asset.

6.7. Girsanov's and Martingale Representation Theorem in higher dimensions.

Let a Brownian motion which is n-dimensional on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e.

$$\underset{\sim}{W}(t) = \left[W_1(t), W_2(t), \dots, W_n(t) \right].$$

Let also $\{\mathcal{F}(t)\}_{t\ge0}$ a filtration associated with this Brownian motion i.e.

is a filtration to which $\underset{\sim}{W}(t)$ is adapted and for which each increment after time t is independent of the σ-algebra $\mathcal{F}(t)$. We are assuming that we work within an interval $[0,T]$ where T is the end of all trades for the derivative securities and moreover $\mathcal{F} = \mathcal{F}(T)$.

Theorem 6.3. (Girsanov's multidimensional version).

Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let the stage be the one described above. In addition let

$$\underset{\sim}{\Theta}(t) = [\Theta_1(t), \Theta_2(t), \ldots, \Theta_n(t)]$$

be an adapted n-dimensional stochastic process to the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$.

Define the Euclidean norm to be

$$\|\underset{\sim}{\Theta}(t)\| = \left[\sum_{i=1}^{n} \Theta_i^2(t)\right]^{1/2}.$$

Define by

$$Z(t) = \exp\left\{-\int_0^t \underset{\sim}{\Theta}(s) \cdot d\underset{\sim}{W}(s) - \frac{1}{2}\int_0^t \|\underset{\sim}{\Theta}(s)\|^2 ds\right\}.$$

$$\underset{\sim}{\tilde{W}}(t) = \underset{\sim}{W}(t) + \int_0^t \underset{\sim}{\Theta}(s) ds.$$

and assume that

$$E_{\mathbb{P}}\left[\int_0^T \|\underset{\sim}{\Theta}(s)\|^2 Z^2(s)\, ds < \infty\right.$$

Then

$$E_{\tilde{\mathbb{P}}}[Z(T)] = 1$$

where the equivalent martingale measure $\tilde{\mathbb{P}}$ is given by

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega)\, d\mathbb{P}(\omega) \quad \text{for all } A \in \mathcal{F}.$$

and $\tilde{W}(t)$ is an n-dimensional Brownian motion

$$\tilde{W}(t) = [\tilde{W}_1(t), \tilde{W}_2(t), \ldots, \tilde{W}_n(t)],$$

where the Brownian motions

$$\tilde{W}_i(t) = W_i(t) + \int_0^t \Theta_i(s) ds \quad \text{for } i = 1, 2, \ldots, n$$

are independent under the probability measure $\tilde{\mathbb{P}}$.

Proof Similar with the one-dimension version ▲

Remark 6.4 The Itô integral in Girsanov's multidimensional version is

$$\int_0^t \Theta(s) \cdot dW(s) = \int_0^t \sum_{i=1}^n \Theta_i(s)\, dW_i(s) = \sum_{i=1}^n \int_0^t \Theta_i(s)\, dW_i(s).$$

Remark 6.5 The component processes $W_i(t)$, $(i = 1, 2, \dots, n)$ of $\underset{\sim}{W}(t)$ are independent under $\mathbb{P}$. However, for the n-dimensional stochastic process $\underset{\sim}{\Theta}(t)$ each component process $\Theta_i(t)$ can depend in a path-dependent, adapted way on all Brownian motions $W_1(t), W_2(t), \dots, W_n(t)$. Therefore under $\mathbb{P}$, the component processes $\tilde{W}_i(t)$, $(i=1,2,\dots,n)$ of $\underset{\sim}{\tilde{W}}(t)$ can be far from independent. Even so, after the change of measure $\tilde{\mathbb{P}}$, these component processes are independent under $\tilde{\mathbb{P}}$.

We now proceed in the Martingale representation theorem for n-dimensions. Assume that we keep the same stage as in Girsanov's theorem and we will stress any made change

Theorem 6.4. (Martingale representation, n-dimensional).

Let $\underset{\sim}{\mathcal{W}}(t)$ be the filtration generated by the n-dimensional Brownian motion $\underset{\sim}{W}(t)$, $0 \le t \le T$, and let $\mathcal{F} = \mathcal{W}(T)$.

Let $M(t)$, $0 \le t \le T$, be a martingale with respect to $\underset{\sim}{\mathcal{W}}(t)$ under $\mathbb{P}$.

Then there exists an adapted to the filtration $\underset{\sim}{\mathcal{W}}(t)$, n-dimensional stochastic process

$$\underset{\sim}{\Gamma}(t) = \left(\Gamma_1(t), \Gamma_2(t), \dots, \Gamma_n(t)\right), \quad 0 \le t \le T$$

such that

$$M(t) = M(0) + \int_0^t \Gamma(s) \cdot dW(s), \quad 0 \le t \le T$$

If, in addition, we assume the notation and assumptions of Theorem 6.3 and if $\tilde{M}(t)$, $0 \le t \le T$ is a martingale under $\tilde{\mathbb{P}}$, then there is an adapted n-dimensional stochastic process $\tilde{\Gamma}(t) = (\Gamma_1(t), \Gamma_2(t), \dots, \Gamma_n(t))$ such that

$$\tilde{M}(t) = \tilde{M}(0) + \int_0^t \tilde{\Gamma}(s) \cdot d\underset{\sim}{\tilde{W}}(s), \quad 0 \le t \le T.$$

The Multidimensional Market.

In a multidimensional market we assume that we have m assets which we trade and a market account with an interest rate process $r(t)$ adapted to a filtration $\{\mathcal{F}(t)\}_{t\geq 0}$.

Let $\{S_i(t)\}_{t\geq 0}$ be the price process for each $i=1,2,\ldots,m$ asset and assume that they are model as the following stochastic differential equations

$$dS_i(t) = \mu_i(t)\, S_i(t)\, dt + S_i(t) \sum_{j=1}^{n} \sigma_{ij}(t)\, dW_j(t)\ , \quad i=1,2,\ldots,m$$

where $\mu_i(t)$, for $i=1,2,\ldots,m$ and $\sigma_{ij}(t)$ for $i=1,2,\ldots,m$; $j=1,2,\ldots,n$ are adapted stochastic processes to the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$.

The price processes $\{S_i(t)\}_{t\geq 0}$ for $i=1,2,\ldots,m$ as modelled by the above stochastic differential equations are correlated. In fact, our first step is to find their correlation process.

We define by

$$\sigma_i(t) = \left(\sum_{j=1}^{n} \sigma_{ij}^2(t) \right)^{1/2}, \quad i=1,2,\ldots,m$$

which we assume that almost surely is different than zero.

Now, we define the processes

$$B_i(t) = \sum_{j=1}^{n} \int_0^t \frac{\sigma_{ij}(s)}{\sigma_i(s)}\, dW_j(s)\,, \quad i=1,2,\ldots,m\ .$$

For each $B_i(t)$, $(i=1,2,\ldots,m)$ we have that $B_i(0)=0$ and $B_i(t)$ is a continuous martingale since it is the sum of Itô - integrals. In addition it is easy to see that

$$dB_i(t)\, dB_i(t) = \sum_{j=1}^{n} \frac{\sigma_{ij}^2(t)}{\sigma_i^2(t)}\, dt = dt\,, \quad i=1,2,\ldots,m$$

which is equivalent to

$$[B_i, B_i](t) = 0\,, \quad i=1,2,\ldots,m$$

Hence, according to the one-dimension Lévy's theorem $B_i(t)$ is a Brownian motion. Now we may rewrite the stochastic differential equations for the price processes $\{S_i(t)\}_{t=0}^{\infty}$ as

$$dS_i(t) = \mu_i(t) S_i(t)\,dt + \sigma_i(t) S_i(t)\,dB_i(t),$$

where $\sigma_i(t)$ apparently is the volatility of $S_i(t)$.

Now, let $i \neq k$ and consider the differential $d(B_i(t)B_k(t))$. According to Itô's product rule we have

$$d(B_i(t)B_k(t)) = B_i(t)\,dB_k(t) + B_k(t)\,dB_i(t) + dB_i(t)\,dB_k(t)$$

We first note that

$$dB_i(t)\,dB_k(t) = \left[\sum_{j=1}^{n} \frac{\sigma_{ij}(t)}{\sigma_i(t)}\,dW_j(t)\right]\left[\sum_{j=1}^{n} \frac{\sigma_{kj}(t)}{\sigma_k(t)}\,dW_j(t)\right]$$

$$= \sum_{j=1}^{n} \frac{\sigma_{ij}(t)\,\sigma_{kj}(t)}{\sigma_i(t)\,\sigma_k(t)}\,dt = \rho_{ik}(t)\,dt \quad ,$$

where

$$\rho_{ik}(t) = \frac{1}{\sigma_i(t)\,\sigma_k(t)} \sum_{j=1}^{n} \sigma_{ij}(t)\,\sigma_{kj}(t).$$

Hence, if we integrate both parts of the differential $d(B_i(t)B_k(t))$ we get

$$B_i(t)B_k(t) = \int_0^t B_i(s)\,dB_k(s) + \int_0^t B_k(s)\,dB_i(s) + \int_0^t \rho_{ik}(s)\,ds.$$

Taking expectation on both sides of this equation we get:

$$E\left[B_i(t)B_k(t)\right] = E\left[\int_0^t \rho_{ik}(s)\,ds\right]$$

or equivalently, since $E[B_i(t)] = 0$ for $i = 1, 2, \dots, m$

$$\mathrm{cov}\left[B_i(t), B_k(t)\right] = E\left[\int_0^t \rho_{ik}(s)\,ds\right]$$

We call $\rho_{ik}(t)$ the instantaneous correlation between $B_i(t)$ and $B_k(t)$. At a specific value of t along a particular path, $\rho_{ik}(t)$ is the conditional correlation between the next increments of B_i and B_k over the time interval $(t, t+dt]$.

In addition, it is easy to see that

$$dS_i(t)\,dS_k(t) = \sigma_i(t)\,\sigma_k(t)\,S_i(t)\,S_k(t)\,dB_i(t)\,dB_k(t)$$

$$= \rho_{ik}(t)\,\sigma_i(t)\,\sigma_k(t)\,S_i(t)\,S_k(t)\,dt.$$

The discounted asset prices as we have already discussed in previous paragraphs are $D(t)\,S_i(t)$ for $i=1,2,\dots,m$, where

$$D(t) = \exp\left\{-\int_0^t r(s)ds\right\}.$$

It is easy to verify that

$$\begin{aligned} d\left(D(t)\,S_i(t)\right) &= D(t)\left[dS_i(t) - r(t)S_i(t)dt\right] \\ &= D(t)\,S_i(t)\left[\left(\mu_i(t) - R(t)\right)dt + \sum_{j=1}^{n} \sigma_{ij}(t)\,dW_j(t)\right] \\ &= D(t)\,S_i(t)\left[\left(\mu_i(t) - R(t)\right)dt + \sigma_i(t)\,dB_i(t)\right] \quad \text{for } i=1,2,\dots,m. \end{aligned}$$

In analogy with the one-dimensional case we want to use the Girsanov's theorem the multidimensional version in order to change the real world probability measure $\mathbb{P}$ into an equivalent martingale measure $\tilde{\mathbb{P}}$ which will be such that the discounted asset prices $D(t)\,S_i(t)$, $(i=1,2,\dots,m)$ will be martingales under $\tilde{\mathbb{P}}$. In this respect we are looking for the adapted processes $\Theta_j(t)$ which are such that

$$d\left(D(t)\,S_i(t)\right) = D(t)\,S_i(t)\sum_{j=1}^{n} \sigma_{ij}(t)\left[\Theta_j(t)\,dt + dW_j(t)\right]$$

If the $\Theta_j(t)$ are found then denote

$$\underset{\sim}{\Theta}(t) = \left[\Theta_1(t), \Theta_2(t), \dots, \Theta_n(t)\right].$$

and define

$$Z(t) = \exp\left\{-\int_0^t \Theta(s)\cdot dW(s) - \tfrac{1}{2}\int_0^t \|\underset{\sim}{\Theta}(s)\|^2 ds\right\}$$

then the equivalent martingale measure is given by

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega)\,d\mathbb{P}(\omega) \quad \text{for all } A \in \mathcal{F}$$

and

$$\tilde{W}(t) = W(t) + \int_0^t \Theta(s)ds$$

is an n-dimensional Brownian motion under $\mathbb{P}$ or equivalently

$$\tilde{W}_j(t) = W_j(t) + \int_0^t \Theta_j(s)ds \quad \text{for } j=1,2,\dots,n$$

are Brownian motion under $\tilde{\mathbb{P}}$.

Hence,

$$d\left(D(t)S_i(t)\right) = D(t)\,S_i(t)\sum_{j=1}^{n} \sigma_{ij}(t)\,d\tilde{W}_j$$

and $D(t)S_i(t)$ is a martingale under $\tilde{\mathbb{P}}$ for $i=1,2,\dots,m$.

Thus the problem is reduced in finding $\Theta_j(t)$ for $j=1,2,\dots,n$. It is easy to see that $\Theta_j(t)$ must satisfy the equations

$$\mu_i(t) - r(t) = \sum_{j=1}^{n} \sigma_{ij}(t)\,\Theta_j(t) \quad \text{for } i=1,2,\dots,m.$$

The equations are called the market price of risk equations. If one cannot solve the market price of risk equations, then there is an arbitrage lurking in the model; the model is bad and should not be used for pricing.

When there is no solution to the market price of risk equations, the arbitrage in the model may not be as obvious but it does exist. If there is a solution to the market price of risk equations, then there is no arbitrage.

At this point we will provide a more formal definition of an arbitrage

Definition 6.1 An arbitrage is a portofolio value process $X(t)$ satisfying $X(0)=0$ and also satisfying for some time $T>0$

$$\mathbb{P}[X(T)\geq 0] = 1, \quad \mathbb{P}[X(T)>0]>0$$

In other words arbitrage is a way of trading so that one starts with zero capital and at some later time T is sure not to have lost money and furthermore has a positive probability of having made money.

Again, in analogy with the market with one asset and the market account we create a portofolio with initial capital $X(0)$, the price with which we sell the derivative security and by holding $\Delta_i(t), (i=1,2,\dots,n)$ units of the asset that has price $S_i(t), (i=1,2,\dots,n)$. Hence, $\Delta_i(t)$ is a decision taken at time t and is an adapted process to the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$. It is not difficult to see that the differential of the portofolio $X(t)$ is given by

$$\begin{aligned} dX(t) &= \sum_{i=1}^{n} \Delta_i(t)\,dS_i(t) + R(t)\left(X(t) - \sum_{i=1}^{n} \Delta_i(t)\,S_i(t)\right)dt \\ &= r(t)X(t)\,dt + \sum_{i=1}^{n} \Delta_i(t)\left(dS_i(t) - r(t)S_i(t)\,dt\right) \end{aligned}$$

$$= r(t)X(t)\,dt + \sum_{i=1}^{n} \frac{\Delta_i(t)}{D(t)}\, d\big(D(t)S_i(t)\big)$$

The differential of the discounted portofolio value is

$$d\big(D(t)X(t)\big) = D(t)\big(dX(t) - r(t)X(t)\,dt\big)$$
$$= \sum_{i=1}^{n} \Delta_i(t)\, d\big(D(t)S_i(t)\big).$$

We have already seen that $D(t)S_i(t)$ is a martingale under the equivalent martingale measure $\tilde{\mathbb{P}}$ and thus from the above equation we get

$D(t)X(t)$ is a martingale under the equivalent martingale probability measure or risk-neutral probability measure $\tilde{\mathbb{P}}$. •

We are now in a position to state and prove what is called the First fundamental theorem of asset pricing.

Theorem 6.5. (First fundamental theorem of asset pricing). Let a market which is modeled by some stochastic processes. If for this model there exist an equivalent risk neutral probability measure or a martingale probability measure, then there is no arbitrage in the model.

Proof: From the hypothesis of our theorem we have that there exists a risk neutral probability measure let say $\tilde{\mathbb{P}}$. Let $\{X(t)\}_{t\geq 0}$ be any portofolio process then the discounted portofolio process will be a martingale under the risk neutral probability measure $\tilde{\mathbb{P}}$. Hence

$$E_{\tilde{\mathbb{P}}}\big[D(T)X(T)\big] = X(0)D(0) = X(0).$$

Now let that $X(0)=0$. In such a case we want to prove that it is impossible for $X(T)$ to satisfy the two conditions

$$\mathbb{P}\{X(T)\geq 0\} = 1 \quad \text{and} \quad \mathbb{P}\{X(T)>0\} > 0.$$

Suppose that $X(T)$ satisfies the first of the two i.e. $\mathbb{P}\{X(T)\geq 0\}=1$ then

$$\mathbb{P}[X(T)<0] = 1 - \mathbb{P}[X(T)\geq 0] = 0$$

In such a case, since $\tilde{\mathbb{P}}$ is an equivalent probability measure to $\mathbb{P}$, we have also

$$\tilde{\mathbb{P}}[X(T)<0] = 0$$

Hence, we simultaneously have

$$E_{\tilde{\mathbb{P}}}[D(T)X(T)] = 0 \text{ and } \tilde{\mathbb{P}}[X(T)<0] = 0$$

The above two equations imply that

$$\tilde{\mathbb{P}}[X(T)>0] = 0,$$

for otherwise we would have $\tilde{\mathbb{P}}[D(T)X(T)>0]>0$, which would imply $E_{\tilde{\mathbb{P}}}[D(T)X(T)]>0$. Since $\tilde{\mathbb{P}}$ and $\mathbb{P}$ are equivalent probability measures we will also have

$$\mathbb{P}[X(T)>0] = 0$$

Hence, there cannot exist an arbitrage since every portofolio value process $X(t)$ satisfying $X(0)=0$ cannot be an arbitrage. ▲

We now proceed to define formally the concept of a complete market.

<u>Definition</u> <u>6.2</u>. A market model is complete if every derivative security can be hedged.

We need necessary and sufficient conditions in order to have a complete market. This is provided by the second fundamental theorem of asset pricing which we will prove in what follows. Let us though start by describing the market and the hedging procedure.

We are assuming a market which consists of m assets with price processes $\{S_i(t)\}_{t\geq 0}$, $i=1,2,\ldots,m$ and a money market account with interest rate process $\{r(t)\}_{t\geq 0}$.

Suppose that for the asset price processes there are n-sources of uncertainty represented by a d-dimensional Brownian motion

$\underset{\sim}{W}(t) = [W_1(t), W_2(t), \dots, W_n(t)]$. Let $\{\mathcal{W}(t)\}_{t\geq 0}$ be the filtration generated by the Brownian motion $\underset{\sim}{W}(t)$. Suppose further that we have a derivative security with underlying assets with price process $\{S_i(t)\}_{t\geq 0}$ $(i=1,\dots,m)$ which has payoff process $\{V(t)\}_{t\geq 0}$. It is naturally assumed that $V(T)$ is $\mathcal{W}(T)$ measurable where as always T is the expiration time of the derivative security.

We assume the existence of an equivalent risk neutral probability measure $\tilde{\mathbb{P}}$ under which the discounted payoff price process $\{D(t)V(t)\}_{t\geq 0}$ is a martingale and with the use of the Girsanov's multidimensional version theorem we found the market risk prices $\Theta_1(t), \Theta_2(t), \dots, \Theta_n(t)$ and defined the Radon-Nikodým derivative process $Z(t)$ and found $\underset{\sim}{\tilde{W}}(t) = [\tilde{W}_1(t), \tilde{W}_2(t), \dots, \tilde{W}(t)]$.

Then according to the Martingale Representation Theorem the multidimensional version, there are processes $\tilde{\Gamma}_1(u), \tilde{\Gamma}_2(u), \dots, \tilde{\Gamma}_n(u)$ such that

$$D(t)V(t) = V(0) + \sum_{j=1}^{d} \int_0^t \tilde{\Gamma}_j(s)\, d\tilde{W}_j(s)$$

Now, consider a portofolio with price process $\{X(t)\}_{t\geq 0}$ which is constructed by having $\Delta_i(t)$ $(i=1,2,\dots,m)$ unit of asset i and the money market account.

Then as we have seen in page 6.25

$$d\big(D(t)X(t)\big) = \sum_{i=1}^{m} \Delta_i(t)\, d\,(D(t)S_i(t))$$
$$= \sum_{j=1}^{n} \sum_{i=1}^{m} \Delta_i(t) D(t) S_i(t) \sigma_{ij}(t)\, d\tilde{W}_j(t).$$

Integrating both sides of the above equation we get

$$D(t)X(t) = X(0) + \sum_{j=1}^{n} \int_0^t \sum_{i=1}^{m} \Delta_i(s) D(s) S_i(s) \sigma_{ij}(s)\, d\tilde{W}_j(s)$$

We would like to be able to hedge a short position in the derivative security and thus we need to have $X(0)=V(0)$ and $X(t)=V(t)$ for every t. In order to do so we must choose the hedging strategy $\Delta_1(t), \Delta_2(t), \dots, \Delta_m(t)$ that satisfies the following equations

$$\frac{\tilde{\Gamma}_j(t)}{D(t)} = \sum_{i=1}^{m} \Delta_i(t) S_i(t) \sigma_{ij}(t)$$

called the hedging equations.

Theorem 6.6 (Second fundamental theorem of asset pricing).

Let a market where m assets are traded and let a market model for which an equivalent risk-neutral probability measure exists. The model is complete if and only if the risk-neutral probability measure is unique

Proof: a) Let us first assume that the market is complete. We will show that there can only be one risk neutral or equivalent martingale measure

Suppose that there are two equivalent real world probability measure $\mathbb{P}$ risk neutral probability measures $\tilde{\mathbb{P}}_1$ and $\tilde{\mathbb{P}}_2$.

Since the market is complete then any derivative security can be hedged. This implies that there exist for any derivative security with payoff process $V(t)$ a portofolio value process $X(t)$ with some initial condition $X(0) = V(0)$ and $X(T) = V(T)$. Thus the same will apply for the derivative security with payoff

$$V(T) = \mathbb{I}_A \frac{1}{D(T)}$$

where A be any set in $\mathcal{F}(T)$.

Since both $\tilde{\mathbb{P}}_1$ and $\tilde{\mathbb{P}}_2$ are risk-neutral probability measures the discounted portofolio value process $D(t)X(t)$ is a martingale under both $\tilde{\mathbb{P}}_1$ and $\tilde{\mathbb{P}}_2$. Thus we have for any $A \in \mathcal{F}(T)$

$$\tilde{\mathbb{P}}_1(A) = E_{\tilde{\mathbb{P}}_1}(\mathbb{I}_A) = E_{\tilde{\mathbb{P}}_1}\left[D(T)V(T)\right] = E_{\tilde{\mathbb{P}}_1}\left[D(T)X(T)\right] = X(0) =$$

$$= E_{\tilde{\mathbb{P}}_2}\left[D(T)X(T)\right] = E_{\tilde{\mathbb{P}}_2}\left[D(T)V(T)\right] = E_{\tilde{\mathbb{P}}_2}\left[\mathbb{I}_A\right] = \tilde{\mathbb{P}}_2(A),$$

consequently,

$$\tilde{\mathbb{P}}_1(A) = \tilde{\mathbb{P}}_2(A) \quad \text{for any } A \in \mathcal{F}(T)$$

and the two risk-neutral probability measures coincide.

b) Assume that there a exists a unique risk-neutral probability measure. We will prove that the market is complete i.e. any derivative security can be hedged.

Uniqueness of the risk-neutral probability measure implies that the market price of risk equations

$$\mu_i(t) - r(t) = \sum_{j=1}^{n} \sigma_{ij}(t)\,\Theta_j(t) \quad , \quad i = 1, \dots, m$$

have only one solution

$$[\Theta_1(t), \Theta_2(t), \dots, \Theta_n(t)] .$$

In order to prove that every derivative security can be hedged, we must be able to solve the equations

$$\frac{\tilde{\Gamma}_j(t)}{D(t)} = \sum_{i=1}^{m} \Delta_i(t) S_i(t) \sigma_{ij}(t) \quad , \quad j = 1, 2, \dots, d .$$

for $\Delta_1(t), \Delta_2(t), \dots, \Delta_m(t)$.

Consider in matrix form the market price of risk equations.

$$\begin{bmatrix} \sigma_{11}(t) & \sigma_{12}(t) & \cdots & \sigma_{1n}(t) \\ \sigma_{21}(t) & \sigma_{22}(t) & \cdots & \sigma_{2n}(t) \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{m1}(t) & \sigma_{m2}(t) & \cdots & \sigma_{mn}(t) \end{bmatrix} \cdot \begin{bmatrix} \Theta_1(t) \\ \Theta_2(t) \\ \vdots \\ \Theta_n(t) \end{bmatrix} = \begin{bmatrix} \mu_1(t) - r(t) \\ \mu_2(t) - r(t) \\ \cdots \\ \mu_m(t) - r(t) \end{bmatrix}$$

Also consider in matrix form the hedging equations.

$$\begin{bmatrix} \sigma_{11}(t) & \sigma_{21}(t) & \cdots & \sigma_{m1}(t) \\ \sigma_{12}(t) & \sigma_{22}(t) & \cdots & \sigma_{m2}(t) \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{1n}(t) & \sigma_{2n}(t) & \cdots & \sigma_{mn}(t) \end{bmatrix} \begin{bmatrix} \Delta_1(t) S_1(t) \\ \Delta_2(t) S_2(t) \\ \cdots \\ \Delta_m(t) S_m(t) \end{bmatrix} = \begin{bmatrix} \tilde{\Gamma}_1(t)/D(t) \\ \tilde{\Gamma}_2(t)/D(t) \\ \cdots \\ \tilde{\Gamma}_n(t)/D(t) \end{bmatrix}$$

Note that if we denote by $\underset{\sim}{\sigma}(t) = \{\sigma_{ij}(t)\}_{i,j}$ the matrix in the first matrix form equation then in the second matrix form equation appears the transpose of $\underset{\sim}{\sigma}(t)$. It is a known result from Linear Algebra that the uniqueness of a solution for the first matrix equation for

$$[\Theta_1(t), \Theta_2(t), \dots, \Theta_n(t)]$$

implies the existence of a solution from the second matrix equation for

$$[\Delta_1(t) S_1(t), \Delta_2(t) S_2(t), \dots, \Delta_m(t) S_m(t)]$$

and consequently for

$$[\Delta_1(t), \Delta_2(t), \dots, \Delta_m(t)] .$$

▲

6. Forward and Futures.

Let us assume a market which consists of an asset, a zero-coupon bond and the market money account.

Assume the usual stage i.e. in this market there exists a risk-neutral measure $\tilde{\mathbb{P}}$ or an equivalent martingale probability measure and the market is complete. This implies that the risk-neutral probability measure is unique and that means that there exists a hedging strategy for any derivative security. The asset in our market is assumed to be non-paying dividend paying.

Let $\{R(t)\}_{t\geq 0}$ be the interest rate process in a large interval $[0, T^{(f)}]$. All non-zero bonds and derivative securities we consider will mature or expire within that interval.

As usual, we define the discount process

$$D(t) = \exp\left\{-\int_0^t R(s)\,ds\right\}$$

Let $S(t),\ 0\leq t\leq T^{(f)}$ be an asset price process. According to what we have so far established we have if we define as before by $V(t)$ the payoff price of any derivative security we have that

$$V(t) = \frac{1}{D(t)} E_{\tilde{\mathbb{P}}}\left[D(T)\,V(T)/\mathcal{F}(t)\right]$$

Let now $B(t,T)$ be the price at time t of a zero-coupon bond paying one unit of money at time T (maturity). Then according to the previous formula we arrive at

$$\begin{aligned} B(t,T) &= \frac{1}{D(t)} E_{\tilde{\mathbb{P}}}\left[D(T)\,B(T,T)/\mathcal{F}(t)\right] \\ &= \frac{1}{D(t)} E_{\tilde{\mathbb{P}}}\left[D(T)/\mathcal{F}(t)\right]. \end{aligned}$$

The assumption of the existence of the risk-neutral probability model guaranties that there is no arbitrage in the market of zero-coupon bond, the asset and the market money account

A forward contract is an agreement to pay a specified delivery price K at at a delivery date T, where $0 \leq T \leq T^{(f)}$, for the asset whose price at time t is $S(t)$.

Definition 6.3

The T-forward price $For_S(t,T)$ of the asset whose price process is $\{S(t)\}_{t\geq 0}$, $0 \leq t \leq T^{(f)}$ is the value of delivery price K that makes the forward contract have no-arbitrage price zero at time t.

We will now provide the following theorem for the pricing of a forward contract

Theorem 6.7.

Let that the assumptions of the present section are valid and let that the zero-coupon bonds of all maturities can be traded. Then

$$For_S(t,T) = \frac{S(t)}{B(t,T)} \quad , 0 \leq t \leq T \leq T^{(f)}$$

Proof:

We have assumed the existence of a risk-neutral probability model. For the forward contract the payoff price at the time of maturity T is $V(T) = S(T) - K$. The payoff price at time t, $V(t)$ will be given by

$$V(t) = \frac{1}{D(t)} E_{\tilde{P}}\left[D(T)(S(T)-K)/\mathcal{F}(t)\right]$$

$$= \frac{1}{D(t)} E_{\tilde{P}}\left[D(T)S(T)/\mathcal{F}(t)\right] - \frac{K}{D(t)} E_{\tilde{P}}\left[D(T)/\mathcal{F}(t)\right]$$

$$= S(t) - K B(t,T)$$

According to definition the T-forward price $For_S(t,T)$ is the value of K that makes the above value of $V(t)$ equals to zero, thus

$$S(t) - For_S(t,T)\, B(t,T) = 0$$

or

$$For_S(t,T) = \frac{S(t)}{B(t,T)} ,$$

▲

Now consider a time interval $[0,T]$, and let a thin partition of it i.e.

$$0=t_0<t_1<t_2<\dots<t_n=T.$$

For obvious reasons we shall refer to each subinterval $[t_k,t_{k+1})$ as a "day".

Assume that the interest rate is constant within each day then the discount process for $k=0,1,\dots,n-1$ is given by

$$D(t_{k+1})=\exp\left\{-\sum_{j=0}^{k} r(t_j)(t_{j+1}-t_j)\right\}.$$

which apparently is $\mathcal{F}(t_k)$-measurable.

At time t_k the risk-neutral pricing of the zero-coupon bond paying a unit of money at maturity T is

$$B(t_k,T)=\frac{1}{D(t_k)}\,E_{\tilde{\mathbb{P}}}\left[D(T)/\mathcal{F}(t_k)\right].$$

At time t_k the T-forward price of an asset whose price process is $S(t)$ is given by

$$For_S(t_k,T)=\frac{S(t_k)}{B(t_k,T)}, \quad \mathcal{F}(t_k)\text{-measurable}.$$

Assume now that we agree to receive a unit of the asset at time T whose price is $S(T)$ and pay $For_S(t_k,T)$ at time T, i.e. we take a long position in the forward contract at time t_k.

At any time $t_j\geq t_k$ the payoff process $V(t_k,t_j)$ will be given by

$$V(t_k,t_j)=\frac{1}{D(t_j)}\,E_{\tilde{\mathbb{P}}}\left[D(T)\left(S(T)-\frac{S(t_k)}{B(t_k,T)}\right)/\mathcal{F}(t_j)\right]$$

$$=\frac{1}{D(t_j)}\,E_{\tilde{\mathbb{P}}}\left[D(T)S(T)/\mathcal{F}(t_j)\right]-\frac{S(t_k)}{B(t_k,T)}\frac{1}{D(t_j)}\times$$

$$E_{\tilde{\mathbb{P}}}\left[D(T)/\mathcal{F}(t_j)\right]=$$

$$=S(t_j)-S(t_k)\frac{B(t_j,T)}{B(t_k,T)}$$

From the above formula it is evident that if the asset's price grows faster than the value of the money of $S(t_k)$ at the future time $t_j>t_k$ then the forward contract takes on a positive value. Otherwise, it takes on a negative value.

In order to deal with the concern for default by the other party the agent creates a margin account and acquires what is called a futures contract. If an agent holds a long futures position between times t_k and t_{k+1} then at time t_{k+1} he receives a payment

$$Fut_S(t_{k+1}, T) - Fut_S(t_k, T)$$

where $Fut_S(t, T)$ is the future price process to be found in what follows. The future price process $Fut_S(t, T)$ is constructed so that $Fut_S(t, T)$ is $\mathcal{F}(t)$-measurable for every t and finally

$$Fut_S(T, T) = S(T).$$

Therefore, the sum of payments received by an agent who purchases a futures contract at time zero and holds it until delivery date T is

$$\begin{aligned} &\big(Fut_S(t_1, T) - Fut_S(0, T)\big) + \\ &\big(Fut_S(t_2, T) - Fut_S(t_1, T)\big) + \\ &\quad \cdots\cdots\cdots \\ &\big(Fut_S(t_n, T) - Fut_S(t_{n-1}, T)\big) = \\ &= Fut_S(T, T) - Fut_S(0, T) = S(T) - Fut_S(0, T). \end{aligned}$$

This is called marking to margin. In contrast to the case of forward contract the payment from holding a futures contract is distributed over the life of the contract. If the agent takes delivery of the asset at time T, paying market price $S(T)$ for it, his total income from futures contract and the delivery payment is $-Fut_S(0, T)$. Ignoring the time value of money, he has effectively paid the price $Fut_S(0, T)$ for the asset, a price that was locked at time zero.

In addition to satisfying $Fut_S(T, T) = S(T)$, the future price process is chosen so that at each time t_k the value of the payment to be received at time t_{k+1}, and indeed at all future times $t_j > t_k$ is zero. This means that at any time one may enter or close out a position in the contract without incurring any cost other than payments already made. This condition in terms of an equation involving the relevant variables could written as follows

$$\frac{1}{D(t_k)} E_{\tilde{\mathbb{P}}}\left[D(t_{k+1})\left(Fut_S(t_{k+1},T) - Fut_S(t_k,T)\right) / \mathcal{F}(t_k)\right] = 0$$

$$= \frac{D(t_{k+1})}{D(t_k)} \left\{ E_{\tilde{\mathbb{P}}}\left[Fut_S(t_{k+1},T) / \mathcal{F}(t_k)\right] - Fut_S(t_k,T) \right\}$$

or equivalently,

$$E_{\tilde{\mathbb{P}}}\left[Fut_S(t_{k+1},T) / \mathcal{F}(t_k)\right] = Fut_S(t_k,T), \quad k = 0,1,\dots,n-1.$$

Thus from the above equation we conclude that $Fut_S(t,T)$ as constructed should be a martingale under $\tilde{\mathbb{P}}$. Since we have already required that $Fut_S(T,T) = S(T)$ we arrive at

$$Fut_S(t_k,T) = E_{\tilde{\mathbb{P}}}\left[S(T)/\mathcal{F}(t_k)\right], \quad k = 0,1,\dots,n$$

It remains to show that for $Fut_S(t_k,T)$ given from the above equation we have that the payment to be received at time t_j is zero for every $j \geq k+1$. In this respect we have

$$\frac{1}{D(t_k)} E_{\tilde{\mathbb{P}}}\left[D(t_j)\left(Fut_S(t_j,T) - Fut_S(t_{j-1},T)\right) / \mathcal{F}(t_k)\right] =$$

$$= \frac{1}{D(t_k)} E_{\tilde{\mathbb{P}}}\left[E_{\tilde{\mathbb{P}}}\left[D(t_j)\left(Fut_S(t_j,T) - Fut_S(t_{j-1},T)\right) / \mathcal{F}(t_{j-1})\right] / \mathcal{F}(t_k)\right]$$

$$= \frac{1}{D(t_k)} E_{\tilde{\mathbb{P}}}\left[D(t_j) E_{\tilde{\mathbb{P}}}\left[Fut_S(t_j,T)/\mathcal{F}(t_{j-1})\right] - D(t_j) Fut_S(t_{j-1},T) / \mathcal{F}(t_k)\right]$$

$$= \frac{1}{D(t_k)} E_{\tilde{\mathbb{P}}}\left[D(t_j) Fut_S(t_{j-1},T) - D(t_j) Fut_S(t_{j-1},T) / \mathcal{F}(t_k)\right] = 0.$$

The above results by analogy are trasformed in the continuous case as follows:

Definition 6.4. The future price process of an asset whose value at time T is $S(T)$ is given by the formula

$$Fut_S(t,T) = E_{\tilde{\mathbb{P}}}[S(T)/\mathcal{F}(t)] \quad , \; 0 \le t \le T$$

We will now provide without proof the following theorem

Theorem 6.8 The future price process $Fut_S(t,T)$ is a martingale under the risk neutral measure $\tilde{\mathbb{P}}$, it satisfies $Fut_S(T,T) = S(T)$, and the value of a long (or short) futures position to be held over an interval of time is always zero.

We define the forward - futures spread, to be

$$For_S(0,T) - Fut_S(0,T).$$

Taking into account that under the risk neutral measure $\tilde{\mathbb{P}}$ we have

$$B(0,T) = E_{\tilde{\mathbb{P}}}[D(T)]$$

then we arrive at

$$For_S(0,T) - Fut_S(0,T) = \frac{S(0)}{E_{\tilde{\mathbb{P}}}[D(T)]} - E_{\tilde{\mathbb{P}}}[S(T)]$$

$$= \frac{1}{E_{\tilde{\mathbb{P}}}[D(T)]}\left\{ E_{\tilde{\mathbb{P}}}[D(T)S(T)] - E_{\tilde{\mathbb{P}}}[D(T)]\, E_{\tilde{\mathbb{P}}}[S(T)] \right\}$$

$$= \frac{1}{B(0,T)}\, Cov_{\tilde{\mathbb{P}}}[D(T), S(T)].$$

~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~

REFERENCES.

1. Björk, T (1998) Arbitrage theory in Continuous time. Oxford University Press.
2. Duffie, D (1992). Dynamic Asset Pricing Theory. Princeton University Press.
3. Karatzas, I. and Shreeve, S. (1998). Methods of Mathematical Finance, Springer-Verlag, New York.
4. Musiela, M and Rutkowski, M (1997). Martingale methods in Financial modelling, Springer-Verlag. New York.
5. Bingham, N.H and R. Kiesel. (1998). Risk-Neutral Valuation Pricing and Hedging of Financial Derivatives. Springer.
6. Shiryaev, A.N. (1999). Essentials of Stochastic Finance. Facts, Models, Theory. World Scientific.