UNIVERSITY OF TARTU
INSTITUTE OF MATHEMATICAL STATISTICS

Martingales

Lecture notes

Fall 2016

Kalev Pärna

J. Liivi 2, 50409 Tartu,

Email: kalev.parna@ut.ee

These lecture notes give an overview of martingales and their use in financial mathematics. It assumes that the reader has passed a measure theoretic course in probability. The course starts with introducing the concept of conditional expectation, followed by a treatment of discrete time martingales. Then continuous time martingales are covered, including Brownian motion. The stochastic integral is defined and Itô formula is shown. The theory is applied for pricing of options by considering classical Black-Scholes model.

The notes are mainly based on the following books.

D. Williams. *Probability with Martingales.* Cambridge University Press, 1995. Chapters 9, 10, 11, 12

A. Etheridge.*A Course in Financial Calculus.* Cambridge University Press, 2004.

D. Lamperton, B. Lapeyre. *Introduction to Stochastic Calculus Applied to Finance.* Chapman & Hall, 1996.

English - Estonian dictionary

Probability space $(\Omega, \mathcal{F}, \mathbf{P})$ - tõenäosusruum

random variable - juhuslik suurus

conditional probability - tinglik tõenäosus

expectation (mean value) - keskväärtus

variance - dispersioon

conditional expectation - tinglik keskväärtus

$\mathcal{G}$- measurable - $\mathcal{G}$-mõõtuv

filtration - filtratsioon

adopted to filtration - filtratsioonile kohandatud

martingale - martingaal

submartingale - submartingaal

supermartingale - supermartingaal

fair game - õiglane mäng, aus mäng

unfair game - ebaõiglane mäng

previsible process - ennustatav protsess (ettenähtav protsess)

gamble - hasartmäng

stopping time - peatumisaeg

stopped martingale - peatatud martingaal

stake on game n - panus mängus n

martingale transform - martingaalteisendus

stochastic integral - stohhastiline integraal

convergence theorem - koondumisteoreem

upcrossing - ülestõus (lõigu läbimine altpoolt üles)

forward convergence theorem - ettepoole koondumise teoreem

bounded - tõkestatud

Doob decomposition - Doobi lahutus

quadratic variation - ruutvariatsioon

uniform integrability - ühtlane integreeruvus

Brownian motion - Browni liikumine (= Wieneri protsess)

random walk - juhuslik ekslemine

central limit theorem - tsentraalne piirteoreem

stationary increments - statsionaarsed juurdekasvud

drift - triiv, trend

trajectory (path) - trajektoor

martingales with continuous time - pideva ajaga martingaalid

variation - variatsioon

semimartingale - semimartingaal

cross-variation - ristvariatsioon

change of measure - mõõdu vahetus

option - optsioon

maturity - täitmisaeg

payoff function - maksefunktsioon

option pricing - optsiooni hindamine (hinnastamine)

volatiliy - volatiilsus

stock price - aktsia hind

discounted price - diskonteeritud hind

1 Introduction

Martingales in simple words

A martingale is a mathematical model for a 'fair' game. What is a fair game? Consider the following example. A die is thrown and you earn 1 eur if the result is 1, 2, or 3, and you lose the same amount if the result is 4, 5, or 6. Your expected win is therefore 0 ($= 1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{2}$), which means that you can not earn systematically any additional money (random gains, both positive or negative, still possible). Denote by Y_n your total win after the game n, $Y_n = X_1 + X_2 + ... + X_n$ where X_i is the win in game i (each $X_i = +1$ or -1). The random process Y_n obtained in this way is a martingale. (Such a process is also known under the name "simple random walk".)

What is typical for a martingale? As in the example above, the martingale makes steps whose average length is equal to 0. The consequence is that given the current value Y_n, the next value Y_{n+1} is in average the same as the current value Y_n. Mathematically, it is expressed by writing $\mathbf{E}(Y_{n+1}|Y_n) = Y_n$. The expression on the left-hand side of the last equality is called 'conditional expectation' of Y_{n+1} given Y_n - one of the main concepts in our course.

How can one benefit from martingales? Suppose the process we are interested in is a martingale. Then, if we know its today's value, we also know something very important about its tomorrow's value - namely, we know that tomorrow's mean value is exactly the same as today's value (which we know!). However, on financial markets, any information about the future is very useful. Therefore, in our course we will try to identify martingales in our market models, in order to solve important pricing problems (e.g. option pricing).

Martingales have been intensively studied since 1940's (Doob, Itô stochastic calculus). They have been used in financial economics since 1970's due to pioneering works by F. Black and M. Scholes (Noble Prize to R. Merton and M. Scholes for a new method to determine the value of derivatives).

Basic concepts of probability used in this course

Random experiment (trial) is an action whose consequence is not predetermined.

Space of elementary events (sample space) Ω is the set of all possible outcomes of the random trial. Each element ω in Ω is called an elementary event (sample point).

σ-algebra of events (event space) $\mathcal{F}$ is a collection of subsets of Ω satisfying:

1. $\emptyset, \Omega \in \mathcal{F}$
2. if $A_1, A_2, \ldots \in \mathcal{F}$, then also $\cup_i A_i \in \mathcal{F}$
3. if $A \in \mathcal{F}$, then also $\bar{A} \in \mathcal{F}$.

Remark: From 1.-3. it follows that $\mathcal{F}$ is also closed w.r.t. intersections of its elements: if $A_i \in \mathcal{F}$ then also $\cap_i A_i \in \mathcal{F}$.

All elements of the event space $\mathcal{F}$ are called **events**.

Ex.1. Let us throw a die. Then $\Omega = \{1, 2, \ldots, 6\}$ and we can define $\mathcal{F}$ as a collection of **all** possible subsets of Ω, i.e. $\mathcal{F} = 2^{\Omega}$, which contains $2^6 = 64$ elements.

Ex.2. In the previous example, suppose you are betting on the result of a die throw: you win or lose 1 euro depending on whether the outcome is an odd number $(1, 3, 5)$ or even number $(2, 4, 6)$. Then you can use a simpler event space, namely $\mathcal{F} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$.

In the latter example, we say that event space $\mathcal{F}$ is generated by the event $A = \{1, 3, 5\}$, and we write $\mathcal{F} = \sigma(A)$. This is the smallest σ- algebra which contains the subset A. σ-algebras can also be generated in a more sophisticated way (see below σ-algebras induced by random variables).

Probability (probability measure) $\mathbf{P}$ is a function on $\mathcal{F}$ satisfying:

1. $\mathbf{P}(A) \geq 0$ for each $A \in \mathcal{F}$
2. $\mathbf{P}(\emptyset) = 0, \ \mathbf{P}(\Omega) = 1$
3. if $A_1, A_2, \ldots$ do not intersect and each $A_i \in \mathcal{F}$, then $\mathbf{P}(\cup_i A_i) = \sum_i \mathbf{P}(A_i)$.

The triple $(\Omega, \mathcal{F}, \mathbf{P})$ is called a **probability space**.

Borel sets and Borel σ-algebra:

Consider a special case where Ω is the real line $\mathcal{R}$ (imagine that we are throwing random point to the real line, or we measure yearly profit or loss of an enterprise). Consider all possible intervals $[a, b], \ a < b$. The collection of all such intervals is not a σ-algebra by itself (why?), and we have to add other necessary subsets in order to get the requirements 1-3 fulfilled. For example, we have to add all unions of intervals, their complements etc. Finally, we see that also open intervals (a, b) must be included, together with their unions etc. The smallest σ-algebra which contains all the intervals above is called **Borel σ-algebra**, and it is denoted by $\mathcal{B}$. Each element B of $\mathcal{B}$ is called a **Borel set**. We can say that the Borel σ-algebra $\mathcal{B}$ is generated by the class K of all intervals, and we write $\mathcal{B} = \sigma(K)$.

Random variables

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a given probability space. Let X be a function $X : \Omega \to \mathcal{R}$. For a Borel set B we can consider its *inverse image* $X^{-1}(B) = \{\omega : X(\omega) \in B\}$. Depending on the function X such an inverse set can be an event (i.e. $X^{-1}(B) \in \mathcal{F}$), or not. The function X is called **measurable** (w.r.t. $\mathcal{F}$) if for *each* Borel set $B \in \mathcal{B}$ its inverse set $X^{-1}(B) \in \mathcal{F}$. Measurable functions are also called **random variables**.

Example. The function X defined by

$$X(\omega) = \begin{cases} +1, & \text{if } \omega = 1, 3, 5 \\ -1, & \text{if } \omega = 2, 4, 6 \end{cases}$$

is measurable w.r.t. the σ-algebra $\mathcal{F}_1 = \{\emptyset, \Omega, \{1,3,5\}, \{2,4,6\}\}$ but it is not measurable w.r.t. $\mathcal{F}_2 = \{\emptyset, \Omega, \{1,2,3\}, \{4,5,6\}\}$.
Remark: We see that X is measurable w.r.t. $\mathcal{F}$ if the value of X is changed only on the borders of subsets of $\mathcal{F}$.

Each random variable X generates its σ-algebra $\sigma(X)$ which is the collection of all subsets of the form $X^{-1}(B)$, where B is a Borel subset:

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}\}.$$

Note that always $\sigma(X) \subset \mathcal{F}$ (sub-σ–algebra).

Example: $\Omega = \{\text{all students in the classroom}\}$.
Let $X(\omega) = 0$, if ω =male student, and $X(\omega) = 1$, if ω =female student.
Then $\sigma(X)$ consists of 4 subsets (which ones?)

Distribution function of X is defined as $F(t) = \mathbf{P}\{X \le t\} \equiv \{\omega : X(\omega) \le t\}$

Expectation (or expected value)

There are two important special classes of RV's - discrete and continuous RV's.

A **discrete random variable** has at most countably many different values $x_1, x_2, \ldots$, with respective probabilities $p_1, p_2, \ldots$. The expectation of a discrete RV is calculated as $\mathbf{E}X = \sum_i x_i p_i$.

A **continuous random variable** has density function $f_X(x) = F'(x)$, its expectation is calculated as $\mathbf{E}X = \int_{-\infty}^{\infty} x f(x) dx$

These two formulas are special cases of the **general formula of expectation**:

$$\mathbf{E}X = \int_\Omega X(\omega)\mathbf{P}(dw) = \int_\Omega X d\mathbf{P}$$

which is called Lebesgue integral (here X is an arbitrary RV).
Alternative notation is $\mathbf{E}X = \int x dF(x)$ - Stiltjes integral.

2 Conditional expectation

2.1 Conditional expectation with respect to discrete random variable

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and X, Z two discrete random variables with values $x_1, x_2, \ldots, x_m$ ja $z_1, z_2, \ldots, z_n$, respectively. From elementary probability theory the conditional probability is known to be:

$$\mathbf{P}(X = x_i | Z = z_j) = \frac{\mathbf{P}(X = x_i, Z = z_j)}{\mathbf{P}(Z = z_j)},$$

and conditional expectation of X with respect to the random event $\{Z = z_j\}$ is

$$E(X|Z = z_j) = \sum x_i \mathbf{P}(X = x_i | Z = z_j).$$

Conditional expectation of a random variable X with respect to another random variable Z is defined as a (new) random variable $Y = E(X|Z)$, given by the equation

$$\text{if } \ Z(\omega) = z_j, \ \text{then } Y(\omega) = E(X|Z = z_j) := y_j.$$

It is useful to look at this in a new way. The random variable Z generates (creates) a σ-algebra $\mathcal{G} = \sigma(Z)$ consisting of sets of the form [1] $\{Z \in B\}, \ B \in \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-algebra. As in our case Z is discrete, each set $\{Z \in B\}$ is a union of some G_j's and $\mathcal{G} = \sigma(Z)$ consists of all possible unions of the sets G_j (the total number of such unions is 2^n, including the empty set $\emptyset$). Since the conditional expectation Y is (similarly to Z) is constant on subsets G_j, then for each $B \in \mathcal{B}$ also $Y^{-1}(B) \in \mathcal{G}$ or, in other words,

$$Y \ \text{ is } \ \mathcal{G} - \text{measurable.} \tag{1}$$

[1] By notation $\{Z \in B\}$ we mean the set $\{\omega : Z(\omega) \in B\} := Z^{-1}(B)$, i.e. inverse image of a Borel subset B.

Next, since Y takes constant value y_j on the subset G_j, we have

$$\begin{aligned}\int_{G_j} Y d\mathbf{P} &= y_j \mathbf{P}(Z = z_j) = \sum_i x_i \mathbf{P}(X = x_i | Z = z_j)\mathbf{P}(Z = z_j) \\ &= \sum_i x_i \mathbf{P}(X = x_i, Z = z_j) = \int_{G_j} X d\mathbf{P}.\end{aligned}$$

Since each $G \in \mathcal{G}$ is a union of certain G_j's, then, by summing up respective integrals, we get

$$\int_G Y d\mathbf{P} = \int_G X d\mathbf{P}, \quad \forall G \in \mathcal{G}. \tag{2}$$

Results (1) and (2) suggest the following central definition of modern probability.

2.2 Conditional expectation

Theorem 1. (Kolmogorov, 1933) *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and X a random variable with $\mathbf{E}(|X|) < \infty$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Then there exists a random variable Y such that*
(a) Y is $\mathcal{G}$-measurable,
(b) $\mathbf{E}(|Y|) < \infty$,
(c) $\int_G Y d\mathbf{P} = \int_G X d\mathbf{P}, \quad \forall\ G \in \mathcal{G}$.
If $\tilde{Y}$ is another random variable with these properties then $\tilde{Y} = Y$ a.s., that is $\mathbf{P}(\tilde{Y} = Y) = 1$. Each random variable Y with properties (a)–(c) is called the ***conditional expectation*** *of X w.r.t. $\mathcal{G}$ and we write $Y = \mathbf{E}(X|\mathcal{G})$ a.s.*

Remark. The random variables with properties (a)–(c) are also called the *versions* of conditional expectation $\mathbf{E}(X|\mathcal{G})$. It follows that all versions are pair-wise equal a.s.

Notation. If $\mathcal{G} = \sigma(Z)$, then we usually write $\mathbf{E}(X|Z)$ instead of $\mathbf{E}[X|\sigma(Z)]$.

Proof of Theorem1. We apply Radon-Nikodym theorem: If μ and ν are σ-finite measures and $\nu \ll \mu$, then there exists a non-negative function f (*density*) such that $\nu(A) = \int_A f d\mu$ for each $A \in \mathcal{F}$. If g is another such density, then $\mu\{f \neq g\} = 0$.

Suppose first that $X \geq 0$. Denote $\nu(G) := \int_G X d\mathbf{P}, G \in \mathcal{G}$. Then $\nu \ll \mathbf{P}$ and by R-N theorem there exists a ($\mathcal{G}$-measurable) function $Y \geq 0$ such that $\int_G Y d\mathbf{P} = \nu(G) \equiv \int_G X d\mathbf{P}$. If X can also take negative values, then we use the decomposition $X = X^+ - X^-$ and apply R-N separately for two parts, thus obtaining Y^+ and Y^- and define $Y = Y^+ - Y^-$. $\square$

Interpretation of conditional expectation. A random experiment is exercised. Suppose we do not know the outcome ω exactly but, instead, we know about each event $G \in \mathcal{G}$ whether it occurred or not (i.e. whether $\omega \in G$ or not). Then $Y(\omega) = \mathbf{E}(X|\mathcal{G})(\omega)$ is the expected value of X given such an information. If $\mathcal{G}$ is trivial σ-algebra $\{\emptyset, \Omega\}$, it does not give any additional information about ω (since $\emptyset$ never occurs and Ω always occurs), and then the conditional expectation reduces to the (ordinary) expected value, $\mathbf{E}(X|\mathcal{G})(\omega) = \mathbf{E}(X)$ for all ω.

Conditional expectation as the best $\mathcal{G}$-measurable prediction of X.
If $\mathbf{E}(X^2) < \infty$, then the conditional expectation $Y = \mathbf{E}(X|\mathcal{G})$ is orthogonal projection of X onto the space $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$. In other words, Y minimizes the expression $\mathbf{E}(X - Y)^2$ over all possible $\mathcal{G}$-measurable functions (i.e. amongst all predictors which can be computed from the available infornmation G).

2.3 Properties of conditional expectation

Assume that $\mathbf{E}(|X|) < \infty$ and that $\mathcal{G}$, $\mathcal{H}$ are sub-σ-algebras of $\mathcal{F}$, i.e. $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$.

(a) If Y is any version of $\mathbf{E}(X|\mathcal{G})$, then $\mathbf{E}(Y) = \mathbf{E}(X)$.

(b) If X is $\mathcal{G}$-measurable, then $\mathbf{E}(X|\mathcal{G}) = X$ a.s.

(c) **(Linearity)** $\mathbf{E}(a_1X_1 + a_2X_2|\mathcal{G}) = a_1\mathbf{E}(X_1|\mathcal{G}) + a_2\mathbf{E}(X_2|\mathcal{G})$ a.s.

(d) **(Positivity)** If $X \geq 0$, then $\mathbf{E}(X|\mathcal{G}) \geq 0$ a.s.

(e) **(cMon)** If $0 \leq X_n \uparrow X$, then $\mathbf{E}(X_n|\mathcal{G}) \uparrow \mathbf{E}(X|\mathcal{G})$ a.s.

(f) **(cFatou)** If $X_n \geq 0$, then $\mathbf{E}(\liminf X_n|\mathcal{G}) \leq \liminf E(X_n + |\mathcal{G})$ a.s.

(g) **(cDom)** If $|X_n(\omega)| \leq V(\omega), \forall n, \quad \mathbf{E}V < \infty$ and $X_n \to X$ a.s., then

$$E(X_n|\mathcal{G}) \to E(X|\mathcal{G}) \text{ a.s.}$$

(h) **(cJensen)** If $c : \mathbf{R} \to \mathbf{R}$ is convex and $\mathbf{E}|c(X)| < \infty$, then

$$\mathbf{E}[c(X)|\mathcal{G}] \geq c(\mathbf{E}[X|\mathcal{G}]) \text{ a.s.}$$

Corollary: $\|\mathbf{E}(X|\mathcal{G})\|_p \geq \|X\|_p$ p.k.

(i) **(Tower property)** If $\mathcal{H}$ is sub-σ-algebra of $\mathcal{G}$, then

$$\mathbf{E}[\mathbf{E}(X|\mathcal{G})|\mathcal{H}] = \mathbf{E}[X|\mathcal{H}] \text{ a.s.}$$

(j) **(Take out what is known)** If Z is $\mathcal{G}$-measurable and bounded, then

$$\mathbf{E}(ZX|\mathcal{G}) = Z\mathbf{E}(X|\mathcal{G}) \text{ a.s.}$$

It also holds if $p > 1, p^{-1} + q^{-1} = 1, X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbf{P}), Z \in \mathcal{L}^q(\Omega, \mathcal{G}, \mathbf{P})$, or if $X \in (m\mathcal{F})^+, Z \in (m\mathcal{G})^+, \mathbf{E}X < \infty$ ja $\mathbf{E}(ZX) < \infty$.

(k) **(Independence)** If $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$, then

$$\mathbf{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbf{E}[X|\mathcal{G}] \text{ a.s.}$$

In particular, if X is independent of $\mathcal{H}$-st, then $\mathbf{E}(X|\mathcal{H}) = \mathbf{E}(X)$ a.s.

2.4 Proofs of properties of conditional expectation

The proofs are based on the definition of conditional expectation and corresponding properties of the expected value (integral).

The proofs of properties (a)–(c) are left to the reader.

Property (d). Let $X \geq 0$ and $Y = \mathbf{E}(X|\mathcal{G})$. Suppose, contraversially, that $\mathbf{P}(Y < 0) > 0$. Then there exists an n such that the set $G = \{Y < -1/n\}$ has positive probability, and hence [2]

$$0 \leq \mathbf{E}(X;G) = \mathbf{E}(Y;G) < -\frac{1}{n}P(G) < 0.$$

We reached a contradiction. □

Property (e). Let $Y_n = \mathbf{E}(X_n|\mathcal{G})$. Then for positivity of conditional expectation we have that $0 \leq Y_1 \leq Y_2 \leq \ldots$. Define $Y = \limsup Y_n$. Then Y is $\mathcal{G}$-measurable and $Y_n \uparrow Y$ a.s.. Show that $Y = \mathbf{E}(X|\mathcal{G})$. For that, we apply monotone convergence theorem to both sides of the equality

$$\mathbf{E}(Y_n;G) = \mathbf{E}(X_n;G),\ G \in \mathcal{G}.$$

The result is $\mathbf{E}(Y;G) = \mathbf{E}(X;G),\ \forall G \in \mathcal{G}$. □

Property (f). Denote

$$Y_n = \inf_{m \geq n} X_n,\quad Y = \lim_{n\to\infty} Y_n \equiv \liminf X_n.$$

Then $Y_n \uparrow Y$ and by (e) we have $\mathbf{E}(Y_n|\mathcal{G}) \uparrow \mathbf{E}(Y|\mathcal{G})$. Therefore

$$\mathbf{E}(\liminf X_n|\mathcal{G}) = \mathbf{E}(Y|\mathcal{G}) = \liminf E(Y_n|\mathcal{G}) \leq \liminf \mathbf{E}(X_n|\mathcal{G}) \text{ a.s.}$$

where the inequality comes from the property (d). □

[2] $\mathbf{E}(X;G)$ denotes the integral $\int_G X d\mathbf{P}$

Property (g). Apply the property (f) to sequences $X_n + V$ and $V - X_n$.

Property (h).

Property (i) follows immediately from the definition.

Property (j).

3 Martingales

3.1 Filtered spaces. Adopted processes

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space.

Definition 1. *An increasing family of sub-σ-algebras $\{\mathcal{F}_n, n \geq 0\}$ satisfying*

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}$$

is called filtration.

Denote

$$\mathcal{F}_\infty := \sigma\left(\bigcup_n \mathcal{F}_n\right) \subseteq \mathcal{F}.$$

Intuitive idea. The concept of filtration is used to express the growth of information in time. At time n the information about ω consists precisely of the values of all $\mathcal{F}_n$-measurable functions $Z(\omega)$. Usually $\{\mathcal{F}_n\}$ is the **natural filtration**

$$\mathcal{F}_n = \sigma(W_0, W_1, \ldots, W_n)$$

of some random (stochastic) process $\{W_n\}$, and then the information about ω which we have at time n is the current history of the process i.e.

$$W_0(\omega), W_1(\omega), \ldots, W_n(\omega)$$

(from which ω can not be determined uniquely, as a rule).

Definition 2. *We say that the random process $X = (X_n : n \geq 0)$ is **adopted** to the filtration $\{\mathcal{F}_n\}$ if for each n, X_n is $\mathcal{F}_n$ -measurable.*

Intuitive idea. If X is adopted, then its value $X_n(\omega)$ is known to us at time n. Usually, $\mathcal{F}_n = \sigma(W_0, W_1, \ldots, W_n)$ and $X_n = f_n(W_0, W_1, \ldots, W_n)$ for some measurable function f_n.

3.2 Martingale, supermartingale, submartingale

Definition 3. *A process X is called **martingale** (relative to $\{\mathcal{F}_n\}$) if*

1. *X is adopted,*
2. *$\mathbf{E}|X_n| < \infty$,*
3. *$\mathbf{E}(X_n|\mathcal{F}_{n-1}) = X_{n-1}$ a.s. $(n \geq 1)$.*

Supermartingale is defined similarly, except that 3. is replaced by

$$\mathbf{E}(X_n|\mathcal{F}_{n-1}) \leq X_{n-1} \text{ a.s. } (n \geq 1)$$

and a **submartingale** is defined with 3. replaced by

$$\mathbf{E}(X_n|\mathcal{F}_{n-1}) \geq X_{n-1} \text{ a.s. } (n \geq 1).$$

A supermartingale decreases on average, a submartingale increases on average in time.

Note that X is martingale if it is both supermartingale and submartingale. Assuming that $X_0 \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$, X is martingale if and only if $X - X_0$ has the same property. The same is true for supermartingales and submartingales. So we can focus attention on processes with $X_0 = 0$.

If X is for example a supermartingale, then the Tower Property (i) shows that for $m < n$,

$$\mathbf{E}(X_n|\mathcal{F}_m) = \mathbf{E}(X_n|\mathcal{F}_{n-1}|\mathcal{F}_m) \leq \mathbf{E}(X_{n-1}|\mathcal{F}_m) \leq \ldots \leq X_m \text{ a.s.}$$

3.3 Examples of martingales

Example 1. Sums of independent zero-mean RV's.
Let $X_1, X_2, \ldots$ be a sequence of independent RVs with $\mathbf{E}|X_k| < \infty$, $\forall k$ and

$$\mathbf{E}X_k = 0, \ \forall k.$$

Define $S_0 = 0, \quad \mathcal{F}_0 = \{\emptyset, \Omega\}$ and

$$S_n = \sum_{i=1}^{n} X_i, \ n \geq 1, , \quad \mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n), \ n \geq 1.$$

Then $S = (S_n \, : \, n \geq 0)$ is martingale relative to filtration $\{\mathcal{F}_n\}$ (show that!).

Example 2. Products of non-negative independent RVs of mean 1.
Let $X_1, X_2, \ldots$ be a sequence independent non-negative RVs with $X_k \geq 0$ ja

$$\mathbf{E}X_k = 1, \ \forall k.$$

Define $M_0 = 1$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and

$$M_n = \prod_{i=1}^{n} X_i, \quad \mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n).$$

Then the process $X = (X_n \, : \, n \geq 0)$ is martingale relative to $\{\mathcal{F}_n\}$ (show that!).

Näide 3. Accumulating data about a random variable
Let $\{\mathcal{F}_n\}_{n \geq 0}$ be our filtration and let $\xi \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ (it means that ξ is integrable, $\mathbf{E}|\xi| < \infty$.). Define $M_n = \mathbf{E}(\xi \mid \mathcal{F}_n)$ ('some version of CE'). The random variable M_n is a 'coarse' ($\mathcal{F}_n$-measurable) version of ξ, i.e. the best prediction of ξ given the information available to us at time n, which becomes more and more precise as $n \to \infty$. Show that the process $M = (M_n : \ n \geq 0)$ is a martingale. By the Tower Property of CE, we have (a.s.)

$$\mathbf{E}(M_n | \mathcal{F}_{n-1}) = \mathbf{E}(\xi | \mathcal{F}_n | \mathcal{F}_{n-1}) = \mathbf{E}(\xi | \mathcal{F}_{n-1}) = M_{n-1}.$$

Hence M is a martingale.

Later we will be able to show that

$$M_n \to M_\infty := \mathbf{E}(\xi | \mathcal{F}_\infty), \ a.s.$$

3.4 Fair and unfair games

Let $X_n - X_{n-1}$ be your net winnings per unit stake in game n ($n \geq 1$) in a series of games, played at times $n = 1, 2, \ldots$. There is no game at time 0. A simple example is obtained by a series of coin tosses where the outcome of the toss at time k is

$$\Delta_k = \begin{cases} +1, \text{if head} \\ -1, \text{if tail} \end{cases}$$

and $X_n = \sum_{k=1}^{n} \Delta_k$.

In the martingale case
(a) $\mathbf{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) = 0$, (game series is *fair*)
and in the supermartingale case
(b) $\mathbf{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) \leq 0$, (game series is *unfavorable* to you).

Note that we have the case (a) if the coin is symmetric, and case (b) if -1 is more probable than $+1$.

3.5 Predictable process, gambling strategy

Definition 4. *A process $C = (C_n : n \geq 1)$ is called* **predictable**,[3] *if C_n is $\mathcal{F}_{n-1}$-measurable ($n \geq 1$).*

One can think about C_n as your stake on game n. You have to decide on the value of C_n based on the history up to (and including) time $n-1$. This is the intuitive meaning of the predictable character of C. Your winnings on game n are $C_n \cdot (X_n - X_{n-1})$ and your total winnings up to time n are

$$Y_n = \sum_{i=1}^{n} C_i(X_i - X_{i-1}) =: (C \bullet X)_n. \tag{3}$$

Note that $Y_0 = (C \bullet X)_0 := 0$ and $Y_n - Y_{n-1} = C_n(X_n - X_{n-1})$.

[3] also called *previsible*

The process $C \bullet X$ is called the **martingale transform** of X by C. This is the discrete analogue of the stochastic integral $\int C\, dX$. Stochastic-integral theory is one of the greatest achievements of the modern theory of probability.

Theorem 2. (You can't beat the system)
(i) If C is a bounded ($|C_n| \leq K < \infty, \forall n$) predictable process and X is martingale, then $C \bullet X$ is a martingale null at 0.
(ii) If, in addition, C is a non-negative and X is a supermartingale, then $C \bullet X$ is a supermartingale null at 0.

Remark. Assertions (i) and (ii) ramain valid when the boundedness of C is replaced by the condition that $\mathbf{E}|C_n(X_n - X_{n-1})| < \infty,\ \forall n \geq 1$.

<u>Proof.</u> Let C be bounded and previsible and let X be adopted. Let $Y = C \bullet X$. Then, by the property (j) ('Take out what is known') of conditional expectation, we have

$$\mathbf{E}[Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}] = \mathbf{E}[C_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}] = C_n \mathbf{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}]. \quad (4)$$

If X is a martingale, then $\mathbf{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) = 0$, and by the equality (4) $Y = C\bullet$ is also a martingale. If $C_n \geq 0$ and X is a supermartingale, then from (4) it follows that

$$\mathbf{E}[Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}] \leq 0,$$

and thus $C \bullet X$ is also a supermartingale. $\square$

4 Stopping times

Definition 5. *A map* $T : \Omega \to \{0, 1, \ldots; \infty\}$ *is called* **a stopping time**, *if*

$$\{T \leq n\} \in \mathcal{F}_n \quad \forall n \in \{0, 1, \ldots, \infty\}$$

or, eqvivalently,

$$\{T = n\} \in \mathcal{F}_n \quad \forall n \in \{0, 1, \ldots, \infty\}.$$

Exercise. Show that the two conditions above are equivalent.

Note that T can be ∞.

Intuitive idea. T is a time when you can decide to stop playing the game. Whether or not you stop immediately after the n^{th} game depends only on the history up to (and including) time n i.e. $\{T = n\} \in \mathcal{F}_n$.

Example. Let A be an adopted process and let B be a Borel set. Then

$$T = inf\{n \, : \, A_n \in B\}$$

– the time of the first entry into the set B – is a stopping time.

By convention, $inf\{\emptyset\} = \infty$, so that $T = \infty$ means the process A never enters set B.

Example. Let $\{X_n\}$ be a series of die throws, and let $S_n = X_1 + X_2 + \ldots + X_n$. Convince yourself that

$$L := sup\{n \, : \, S_n \leq 100\}$$

is NOT a stopping time.

4.1 Stopped martingales

Let X be a random process (a (super)martingale, for example) and let T be a stopping time. For simplicity, we denote

$$T \wedge n = min\{T, n\}.$$

The process $X^T := (X_{T\wedge n} : n \geq 0)$ is called **a stopped process.** The stopped process remains constant after the stopping time T. Show that stopping does not change the martingale (or supermartingale) property of X.

Theorem 3. *If X is a martingale (supermartingale) and T is a stopping time then the* ***stopped process*** *$X^T = (X_{T\wedge n} : n \geq 0)$ is a martingale (supermartingale) again.*

Proof. The proof is based on Theorem 2. For that, we present the stopped process as a total winnings process. Suppose you always bet 1 unit and quit playing at (immediately after) time T; it means your 'stake process' is C^T where

$$C_n^T(\omega) = \begin{cases} 1, & \text{if } n \leq T(\omega); \\ 0, & \text{otherwise.} \end{cases}$$

Now, let's consider the 'winnings process' $C^T \bullet X$ with value at time n equal to

$$(C^T \bullet X)_n = \sum_{i=1}^{n} C_i^T (X_i - X_{i-1}) = X_{T\wedge n} - X_0 = X_{T\wedge n}.$$

We see that, indeed, the stopped process $\{X_{T\wedge n}\}$ can be regarded as a winning process. Clearly C^T is bounded (by 1) and non-negative. Moreover, C^T is previsible because $\forall n \quad \{C_n^T = 0\} = \{T \leq n-1\} \in \mathcal{F}_{n-1}$. Now Theorem 2 part (i) applies and $X_{T\wedge n}$ is a martingale. If the initial process X is a supermartingale, part (ii) of the same theorem gives the result.

4.2 Doob's Optional Stopping Theorem

It is important to know whether the martingale property $\mathbf{E}X_n = \mathbf{E}X_0$ (or respective supermartingale property $\mathbf{E}X_n \leq \mathbf{E}X_0$) remains true when n is replaced by a stopping time T. The main difference between two cases is that n is constant but T is a random variable which depends on trajectory (on ω). The expected value $\mathbf{E}X_n$ is calculated over all trajectories at the same time n, while in case of $\mathbf{E}X_T$ each trajectory is taken into account at its individual (random) time moment $T = T(\omega)$.

Theorem 4. (Doob's Optional Stopping Theorem)
a) Let T be a stopping time and let X be supermartingale. Then each of the following conditions ensures that X_T is integrable and $\mathbf{E}X_T \leq \mathbf{E}X_0$:

(i) T is bounded ($T \leq K$),

(ii) X is bounded ($\exists K : |X_n| \leq K, \ \forall n \geq 0$) and $T < \infty$ a.s.

(iii) $\mathbf{E}T < \infty$ and for some $K > 0$ $|X_n - X_{n-1}| \leq K, \ \ \forall n \geq 1$.

(iv) $X \geq 0$ and $T < \infty$ a.s.

b) If T is a stopping time, X is a martingale and at least one of the conditions (i)-(iii) is satisfied , then $\mathbf{E}X_T = \mathbf{E}X_0$.

Proof
Part a). Under the condition (i) the proof is simplest. Consider the stopped process $X^T = (X_{T\wedge n}, n \geq 0)$. As trajectories of X^T remain constant after the stopping time T and as $T \leq K$, we have $X_T \equiv X_T^T = X_K^T$, from which $\mathbf{E}X_T = \mathbf{E}X_K^T$. At the same time, by Theorem 3, X^T is a supermartingale, and therefore $\mathbf{E}X_K^T \leq \mathbf{E}X_0^T \equiv \mathbf{E}X_0$. Hence also $\mathbf{E}X_T \leq \mathbf{E}X_0$, as needed.

Under each of conditions (ii) – (iv) first observe that $T < \infty$ a.s. However, the latter ensures that the stopped process X^T converges:

$$X_n^T = X_{T\wedge n} \to X_T \text{ a.s.}, \tag{5}$$

as $n \to \infty$. To proceed, in case (ii) use the Lebesgue theorem of dominated convergence (DOM) which gives

$$\lim_n \mathbf{E}X_n^T = \mathbf{E}X_T \text{ a.s.} \tag{6}$$

Again, as X^T is a supermartingale, for each n we have $\mathbf{E}X_n^T \le \mathbf{E}X_0$. Then, due to (6), we also have $\mathbf{E}X_T \le \mathbf{E}X_0$, as needed.

In case of (iii) we first use

$$|X_{T\wedge n} - X_0| = |\sum_{k=1}^{T\wedge n}(X_k - X_{k-1})| \le KT$$

showing that X^T has an integrable upper bound. Then, again the Lebesgue theorem applies and the rest is the same as in case (ii).

In the case (iv) apply Fatou Lemma giving

$$\mathbf{E}(\liminf_n X_n^T) \le \liminf_n(\mathbf{E}X_n^T)$$

Due to (5), the LHS is equal to $\mathbf{E}X_T$ and the RHS is less than $\mathbf{E}X_0$ (for X^T being a martingale). Therefore, we have $\mathbf{E}X_T \le \mathbf{E}X_0$ again.

The part b) follows from a) (applied twice), since if X is a martingale then both X and $-X$ are supermartingales. $\square$

Corollary 1. *Let M be a martingale such that for some $K_1 > 0$ we have $|M_n - M_{n-1}| \le K_1$ $\forall n \ge 1$. Let C be a predictable process such that $|C_n| \le K_2$ for some $K_2 > 0$. If T is a stopping time such that $\mathbf{E}T < \infty$, then*

$$\mathbf{E}(C \bullet M)_T = 0.$$

Comment. The theorem says that when you play with bounded stake and the win in a single game is also bounded, you quit the game with zero average.

Proof. By Theorem 2 the process

$$(C \bullet M)_n = \sum_{i=1}^{n} C_i(M_i - M_{i-1})$$

is a martingale. By Theorem 3, the stopped process $(C \bullet M)_n^T$ is also a martingale, and its increments are bounded:

$$C_i(M_i - M_{i-1}) \leq |C_i||M_i - M_{i-1}| \leq K_2 \cdot K_1.$$

Now Theorem 4 part b) case (iii) gives $\mathbf{E}(C \bullet M)_T = \mathbf{E}(C \bullet M)_0 = 0$. □

In order to use Optional Stopping Theorem one has to show that $T < \infty$ and sometimes $\mathbf{E}T < \infty$. The following Lemma can be used for such a purpose.

Lemma 1. *Let T be a stopping time such that for some $N \in \mathbf{N}$ and $\varepsilon > 0$ the following inequality holds:*

$$\mathbf{P}(T \leq n + N \mid \mathcal{F}_n) \equiv \mathbf{E}(I_{T \leq n+N} \mid \mathcal{F}_n) > \varepsilon \quad p.k.\ \forall n \geq 1,$$

Then $\mathbf{E}T < \infty$.

Sketch of the proof:

First we show, by means of mathematical induction, that $\mathbf{P}(T > kN) \leq (1-\varepsilon)^k$. Indeed, let us assume that it holds for $k-1$, i.e. $\mathbf{P}(T > (k-1)N) \leq (1-\varepsilon)^{k-1}$. Then

$$\begin{aligned} \mathbf{P}(T > kN) &= \mathbf{P}(T > kN | T > (k-1)N)) \cdot \mathbf{P}(T > (k-1)N)) \\ &= (1-\varepsilon)(1-\varepsilon)^{k-1} = (1-\varepsilon)^k. \end{aligned}$$

Now, using the fact that the probabilities $\mathbf{P}(T > k)$ decrease when k grows, and grouping them by N, the expected value of T can be estimated as

$$\mathbf{E}T = \sum_{k=0}^{\infty} \mathbf{P}(T > k) \leq \sum_{k=0}^{\infty} N\mathbf{P}(T > kN) \leq N \sum_{k=0}^{\infty} (1-\varepsilon)^k = \frac{N}{\varepsilon} < \infty. \ \square$$

5 The Convergence Theorem

Many applications of martingales are based on their convergence property. Let X be a random process. Let a and b, $a < b$ be some real numbers. We denote by $U_N[a,b](\omega)$ the largest k such that there exist $0 \le s_0 < t_0 < s_1 < t_1 < \ldots < s_k < t_k \le N$ for which

$$X_{s_i} < a, \quad X_{t_i} > b \quad \forall i \in \{0, 1, \ldots, N\}.$$

The number $U_N[a,b](\omega)$ is called the number of upcrossing of [a,b] made by the trajectory $\{X_n(\omega)\}$ by time N. The function $U_N[a,b]$ is $\mathcal{F}_N$-measurable (why?).

Lemma 2. (Doob's Upcrossing Lemma) *Let X be a supermartingale. Then*

$$(b-a)\mathbf{E}U_N[a,b] \le \mathbf{E}[(X_N - a)^-].$$

Proof

The idea is to use Theorem 2 ('You can't beat the system!'). For that, we define 0–1 stakes by $C_1 = I_{\{X_0 < a\}}$,

$$C_i = I_{\{C_{i-1}=0\}\cap\{X_{i-1}<a\}} + I_{\{C_{i-1}=1\}\cap\{X_{i-1}\le b\}}, \quad i \ge 2.$$

The process C is previsible, bounded and non-negative, thus by the Theorem 2 the process $Y = C \bullet X$ with $Y_N = \sum_{i=1}^N C_i(X_i - X_{i-1})$ is also a supermartingale. Therefore we have $\mathbf{E}(Y_N) \le \mathbf{E}(Y_0) = 0$.

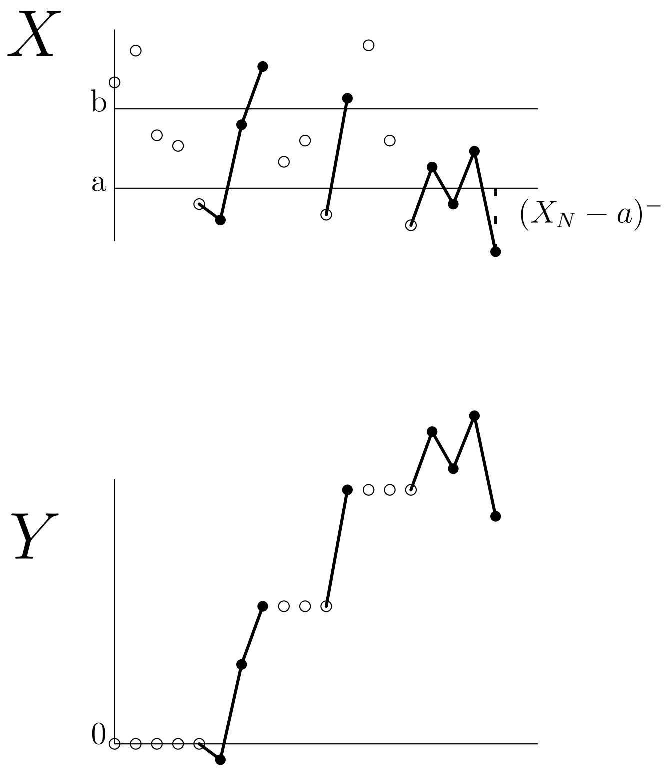

Figure 1: Above is the initial process X, below is $Y = C \bullet X$. White circles correspond to stakes $C_i = 0$, black circles correspond to $C_i = 1$.

However, Y_N satisfies

$$Y_N \geq (b-a)U_N[a,b] - (X_N - a)^-$$

(look at the picture!) Taking expectations gives

$$(b-a)\mathbf{E}(U_N[a,b]) - E[(X_N - a)^-] \leq \mathbf{E}(Y_N) \leq \mathbf{E}(Y_0) = 0.$$

Thus we have

$$(b-a)\mathbf{E}(U_N[a,b]) \leq E[(X_N - a)^-].$$

□

Define $U_\infty[a,b] = \lim_{N\to\infty} U_N[a,b]$.

We say that a process X is *bounded in* L^1 if there exists a constant K such that $\mathbf{E}|X_n| \le K$ for each $n = 1, 2, ...$ or, equivalently, $K := \sup_n \mathbf{E}(|X_n|) < \infty$.

Lemma 3. *Let X be a supermartingale bounded in L^1 (i.e. $\sup_n \mathbf{E}(|X_n|) < \infty$). Then for any $a, b \in \mathbf{R}$, $a < b$, we have*

$$\mathbf{P}(U_\infty[a,b] = \infty) = 0.$$

<u>Proof</u>

By Lemma 2 we have

$$(b-a)\mathbf{E}U_N[a,b] \le \mathbf{E}[(X_N - a)^-] \le |a| + \mathbf{E}(|X_N|) \le |a| + K < \infty, \quad \forall N \ge 1.$$

Letting $N \to \infty$ and using Monotone Convergence Theorem (MON) we have $\mathbf{E}U_\infty[a,b] < \infty$. But then it must be that

$$\mathbf{P}(U_\infty[a,b] = \infty) = 0.$$

□

Theorem 5. (Doob's Forward Convergence theorem)

Let X be a supermartingale bounded in L^1 (i.e. $K = \sup_n E|X_n| < \infty$). Then, almost surely, the limit $X_\infty := \lim_{n\to\infty} X_n$ exists and is finite.

<u>Proof</u>

Note that

$$\begin{aligned} A &= \{\omega : X_n(\omega) \text{ does not converge to a limit in } [-\infty; +\infty]\} \\ &= \{\omega : \liminf_n X_n(\omega) < \limsup_n X_n(\omega)\} \\ &= \cup_{a,b\in\mathbf{Q}:a<b}\{\omega : \liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)\} \\ &= \cup_{a,b\in\mathbf{Q}:a<b}\{\omega : U_\infty[a,b](\omega) = \infty\}. \end{aligned}$$

Since A is a countable union of subsets $\{\omega : U_\infty[a,b](\omega) = \infty\}$ of zero-probability (previous Lemma applies!), we see that $\mathbf{P}(A) = 0$, whence

$$\lim_n X_n \textit{ exists a.s. in } [-\infty; +\infty].$$

Show now that $-\infty$ and ∞ are in fact excluded, i.e. $\lim_n X_n$ can only take finite values a.s. Indeed, by Fatou's Lemma we have

$$\mathbf{E}(|X_\infty|) = \mathbf{E}(\liminf_n |X_n|) \leq \liminf_n \mathbf{E}(|X_n|) \leq K < \infty,$$

so that $\mathbf{P}(|X_\infty| < \infty) = 1$. □

Corollary 2. *If X is a non-negative supermartingale, then, almost surely, the limit $X_\infty := \lim_{n\to\infty} X_n$ exists and is finite.*

Proof. Non-negative X is bounded in $\mathcal{L}^1$, since $\mathbf{E}|X_n| = \mathbf{E}X_n \leq \mathbf{E}_0$ for each n. Now Theorem 5 applies. □

6 Martingales bounded in L^2

One of the easiest ways of proving that a martingale M is bounded in L^1 (needed in Doob's Convergence Theorem!) is to prove that it is bounded in L^2 in the sense that $\sup_n \mathbf{E}(X_n^2) < \infty$. (This is based on the simple fact that $|X_n| \le X_n^2 + 1$, showing that $\mathbf{E}(X_n^2) < \infty$ implies $\mathbf{E}|X_n| < \infty$.)

Let $M = (M_n : n \ge 0)$ be a L^2-martingale. Then for any time moments $0 \le s \le t \le u \le v$ the increments $M_t - M_s$ and $M_v - M_u$ are uncorrelated (orthogonal in L^2):

$$\begin{aligned} \mathbf{E}[(M_t - M_s)(M_v - M_u)] &= \mathbf{E}[\mathbf{E}((M_t - M_s)(M_v - M_u) \mid \mathcal{F}_u)] \\ &= \mathbf{E}[(M_t - M_s)\mathbf{E}(M_v - M_u) \mid \mathcal{F}_u)] = 0, \end{aligned}$$

since $E(M_v - M_u \mid \mathcal{F}_u) = 0$ in the martingale case. Now the simple formula

$$M_n = M_0 + \sum_{i=1}^{n} (M_i - M_{i-1}) \tag{7}$$

expresses M_n as the sum of uncorrelated terms, and therefore we obtain

$$\mathbf{E}(M_n^2) = \mathbf{E}(M_0^2) + \sum_{i=1}^{n} \mathbf{E}(M_i - M_{i-1})^2. \tag{8}$$

Note that all terms on the right-hand side are non-negative, so that the second moment (and also the variance) of M_n grow monotonically together with n.

Theorem 6. *Let M be a L^2-martingale. Then*
(a) M is bounded in L^2 if and only if

$$\sum_{k=1}^{\infty} \mathbf{E}(M_k - M_{k-1})^2 < \infty. \tag{9}$$

(b) If M is bounded in L^2, then $M_n \to M_\infty$ a.s. and in L^2.

Proof. (a) For (8) the L^2-boundedness of M is equivalent to (9). (b) Suppose now (9) holds. Then M is bounded in L^2, hence also in L^1. By Doob's Convergence Theorem we have that $M_n \to M_\infty =: \liminf M_n$ a.s.. Show now the convergence in L^2, meaning that $\lim_n \mathbf{E}(M_\infty - M_n)^2 = 0$. For orthogonality of increments we have

$$\mathbf{E}(M_{n+r} - M_n)^2 = \sum_{k=n+1}^{n+r} \mathbf{E}(M_k - M_{k-1})^2.$$

Letting $r \to \infty$ and applying Fatou's Lemma, we obtain

$$\begin{aligned}\mathbf{E}(M_\infty - M_n)^2 &= \mathbf{E}(\liminf_r (M_{n+r} - M_n)^2) \le \liminf_r \mathbf{E}(M_{n+r} - M_n)^2 \\ &= \sum_{k=n+1}^{\infty} \mathbf{E}(M_k - M_{k-1})^2,\end{aligned}$$

which tends to 0 when $n \to \infty$ (the residual sum of a convergent series!). Hence

$$\lim_n \mathbf{E}(M_\infty - M_n)^2 = 0,$$

so that $M_n \to M_\infty$ in L^2. $\square$

7 Doob Decomposition. Quadratic Variation

Theorem 7. (Doob decomposition)
(a) Let X be an adopted process, with $X_n \in L^1$, $\forall n$. Then X can be represented in the form

$$X = X_0 + M + A,$$

where M is a martingale null at 0, and A is a previsible process null at 0. Moreover, this decomposition is unique up to the a.s. equivalence in the sense that if $X = X_0 + M' + A'$ is another such decomposition, then $P\{M = M', A = A', \forall n\} = 1$.
(b) The process X is a submartingale if and only if the process A is increasing a.s.

Proof. (a) Define

$$\begin{aligned} A_0 &= 0, \\ A_n &= \sum_{i=1}^{n} \mathbf{E}(X_i - X_{i-1} \mid \mathcal{F}_{i-1}), \quad n \geq 1, \end{aligned}$$

and

$$M_n = X_n - X_0 - A_n.$$

Show that A and M satisfy the conditions stated in the theorem. Obviously, A is predictable ($A_n \in \mathcal{F}_{n-1}$). Check that M is a martingale:

$$\begin{aligned} \mathbf{E}(M_n - M_{n-1} \mid \mathcal{F}_{n-1}) &= \mathbf{E}(X_n - X_{n-1} - (A_n - A_{n-1}) \mid \mathcal{F}_{n-1}) && (10) \\ &= \mathbf{E}(X_n - X_{n-1} - (X_n - X_{n-1}) \mid \mathcal{F}_{n-1}) && (11) \\ &= 0. && (12) \end{aligned}$$

Thus M is a martingale.
(b) If X is a submartingale, then $\mathbf{E}(X_i - X_{i-1} \mid \mathcal{F}_{i-1}) \geq 0, \forall i$, and we have

$$A_n = \sum_{i=1}^{n} \mathbf{E}(X_i - X_{i-1} \mid \mathcal{F}_{i-1}) \geq A_{n-1}, \tag{13}$$

so that A is increasing. □

Example. Consider asymmetric random walk $S_n = \sum_{i=1}^n X_i$ with $P(X_i = +1) = 0,6$ and $P(X_i = -1) = 0,4$. Assume that $S_0 = 0$. The process S is a submartingale. It can be decomposed as $S_n = S_0 + M_n + A_n$, where $A_n = 0,2 \cdot n$ and M_n is a martingale given by $M_n = S_n - A_n = \sum_{i=1}^n X_i - 0,2 \cdot n = \sum_{i=1}^n (X_i - 0,2) = \sum_{i=1}^n X_i'$, where the new 'steps' $X_i' = X_i - 0,2$ take values 0,8 and -1,2 with probabilities 0,6 and 0,4, as before. (Check that new steps have zero means!)

Let now M be a L^2-martingale (i.e. $\mathbf{E}(M_n^2) < \infty, \ \forall n$) starting from zero. Then, by Jensen inequality (see property (h) of CE) we have $\mathbf{E}(M_n^2 \mid \mathcal{F}_{n-1}) \geq$ $\geq [\mathbf{E}(M_n \mid \mathcal{F}_{n-1})]^2 = M_{n-1}^2$, and hence

M^2 is a submartingale.

Thus, by Theorem 7, M^2 has the Doob decomposition

$$M^2 = N + A, \tag{14}$$

where N is a martingale and A is a previsible increasing process, both N and A being null at 0.

The process A is often written $\langle M \rangle$ and is called *quadratic variation of M.*

Example. Consider a simple symmetric random walk $S_n = X_1 + \ldots + X_n$, where X_i are independent random variables taking values $+1$ or -1 with equal probabilities $\frac{1}{2}$. We know that the processes S_n and $S_n^2 - n$ are both martingales. Hence the Doob decomposition (14) of the submartingale S_n^2 is given by

$$S_n^2 = (S_n^2 - n) + n$$

with martingale part $N_n = S_n^2 - n$ and previsible increasing part (i.e. quadartic variation) $A_n = \langle M \rangle_n = n$.

Define the limit

$$\langle M \rangle_\infty = \lim_{n \to \infty} \langle M \rangle_n.$$

The following two observations are useful:

(i) M is bounded in L^2 if and only if $\mathbf{E}\langle M\rangle_\infty < \infty$.

Indeed, since $\mathbf{E}(M_n^2) = \mathbf{E}(A_n)$ is increasing in n, we have $\sup \mathbf{E}(M_n^2) = \mathbf{E}\langle M\rangle_\infty$.

(ii) $\langle M\rangle_n - \langle M\rangle_{n-1} = \mathbf{E}(M_n^2 - M_{n-1}^2 \mid \mathcal{F}_{n-1}) = \mathbf{E}[(M_n - M_{n-1})^2 \mid \mathcal{F}_{n-1}]$.

Indeed, since $\mathbf{E}(M_{n-1}M_n \mid \mathcal{F}_{n-1}) = M_{n-1}\mathbf{E}(M_n \mid \mathcal{F}_{n-1}) = M_{n-1}^2$, we can write

$$\begin{aligned}
\mathbf{E}(M_n^2 - M_{n-1}^2 \mid \mathcal{F}_{n-1}) &= \mathbf{E}(M_n^2 - 2M_{n-1}M_n + 2M_{n-1}M_n - M_{n-1}^2 \mid \mathcal{F}_{n-1}) \\
&= \mathbf{E}(M_n^2 - 2M_{n-1}M_n + M_{n-1}^2 \mid \mathcal{F}_{n-1}) \\
&= \mathbf{E}[(M_n - M_{n-1})^2 \mid \mathcal{F}_{n-1}].
\end{aligned}$$

- the conditional variance of the increment of M.

8 Brownian motion

8.1 Transition to continuous time. Brownian motion.

So far, we have studied martingales with discrete time, $t = 0, 1, 2, \ldots$. The simpliest example was Simple Random Walk (SRW). In finance, however, it is also useful to consider processes with continous time, where the time parameter $t \in \mathbf{R}^+ = [0, \infty)$. For example, stock prices can change at any time instant within a business day. Among random processes with continous time one of the most important is so-called Brownian motion. Brownian motion is a simple continuous stochastic process that is widely used in physics and finance for modeling random behavior that evolves over time. Examples of such behavior are the random movements of a molecule of gas or fluctuations in an asset's price.

Brownian motion gets its name from the botanist Robert Brown (1828) who observed in 1827 that tiny particles of pollen suspended in water moved erratically on a microscopic scale; but he was not able to determine the mechanisms that caused this motion. The physicist Albert Einstein published a paper in 1905 explaining that the motion was caused by water molecules randomly bombarding the particle of pollen, and thus helping to firmly establish the atomic theory of matter. Later on, starting in 1918, American mathematician Norbert Wiener created a precise mathematical model for this phenomena. This is what we will study now.

In order to define the Brownian motion, we start from Simple Random Walk (SRW) and let its step sizes (both time step and space step) to go to 0. Denote

$$X_t = \Delta x \cdot (X_1 + X_2 + \ldots + X_{[\frac{t}{\Delta t}]}),$$

where

$$X_i = \begin{cases} +1, & \text{with probability } \frac{1}{2}, \\ -1, & \text{with probability } \frac{1}{2}, \end{cases}$$

Δx - space step size,

Δt - time step size,

$\frac{t}{\Delta t}$ - the number of time steps in the time interval $[0, t]$.

Since $X_1, X_2, \ldots, X_{[\frac{t}{\Delta t}]}$ are independent identically distributed (IID) random variables with mean value $EX_i = 0$ and variance $DX_i = 1$, then for each Δx and Δt we have $EX_t = 0$ and $DX_t = (\Delta x)^2[\frac{t}{\Delta t}]$. When $\Delta t \to 0, \Delta x \to 0$, then the number of summands $[\frac{t}{\Delta t}]$ increases unboundedly and according to the Central Limit Theorem

$$\frac{X_t}{\Delta x\sqrt{[\frac{t}{\Delta t}]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

By choosing the relationship between Δx and Δt such that $\frac{(\Delta x)^2}{\Delta t} =$ const $=: C^2$ we get, on the limit, that $X_t \sim \mathcal{N}(0, C\sqrt{t})$. The limiting process X_t preserves some important features of SRW:

(i) The increments of X_t are independent i.e. for $0 \leq s \leq t \leq u \leq v$ the increments $X_t - X_s$ and $X_v - X_u$ are indpendent r.v. (the same is valid for any n time intervals).

(ii) The increments of X_t are stationary i.e. the distribution of $X_{s+t} - X_s$ only depends on t (and not on s).

These properties suggest the following definition.

Definition 6 (Brownian motion)**.** *The random process $\{W_t,\ t \geq 0\}$ is called Brownian motion (Wieneri process), if*

(i) $W(0) = 0$,

(ii) for all $t > 0$ the r.v. $W_t \sim \mathcal{N}(0, C\sqrt{t})$, where $C > 0$ is a constant,

(iii) increments of W_t are independent and stationary,

(iv) the paths of W_t are a.s. continuous (in t).

From the definition of BM it follows that also the increments of BM are normally distributed: by the stationarity of increments

$$W_t - W_s \overset{\mathcal{D}}{=} W_{t-s} - W_0 = W_{t-s} \sim \mathcal{N}(0, C\sqrt{t-s}),$$

where $\overset{\mathcal{D}}{=}$ is to be read as "has same distribution as" .

Note that, in fact, the property (iv) can be deduced from properties (i)-(iii).

BM is a mathematical model widely used in physics (diffusions), economics (price models) e.t.c. .

If $C = 1$, the BM is called Standard Brownian Motion (SBM). The process

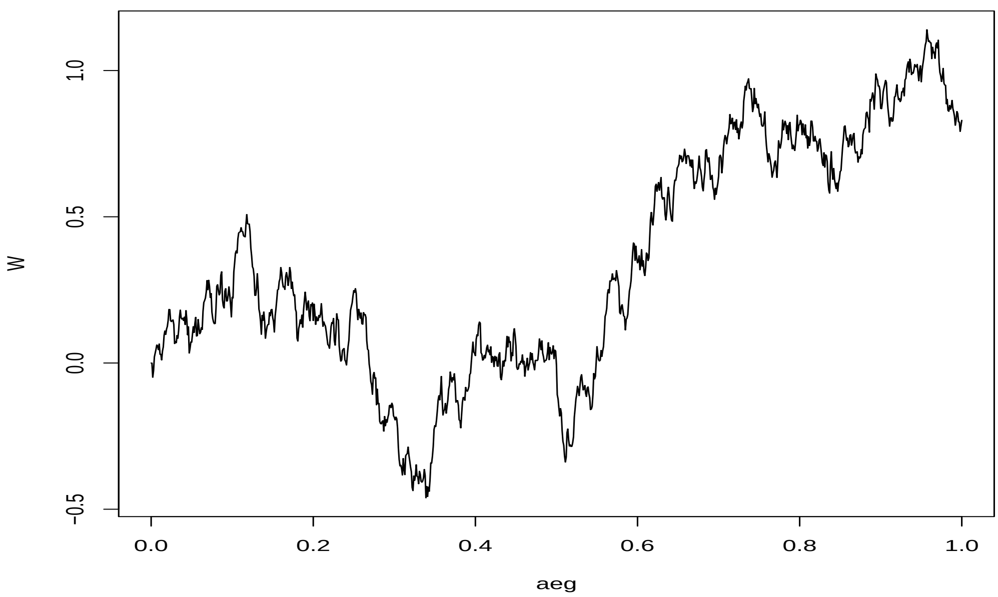

Figure 2: A trajectory of standard Brownian motion

$W_t + \mu t$, where μ is a real number, is called Brownian motion with *drift*(μ is called *drift coefficient*). The mean value of BM with positive (resp negative) drift increases (resp decreases) in time.

8.2 Some properties of Brownian motion

1) Finite-dimensional distributions of SBM

The joint distribution of $(W_{t_1}, W_{t_2}, \ldots, W_{t_n})$ where $0 < t_1 < t_2 < \ldots < t_n$ can easily be calculated.

For each t_i the density of W_{t_i} is

$$f_{W_{t_i}}(x) = \frac{1}{\sqrt{2\pi t_i}} e^{-\frac{x^2}{2t_i}},$$

provided that $C = 1$. Since the equalities

$$\begin{cases} W_{t_1} = x_1 \\ W_{t_2} = x_2 \\ \ldots \\ W_{t_n} = x_n \end{cases}$$

are equivalent to the equalities

$$\begin{cases} W_{t_1} = x_1 \\ W_{t_2} - W_{t_1} = x_2 - x_1 \\ \ldots \\ W_{t_n} - W_{t_{n-1}} = x_n - x_{n-1} \end{cases}$$

and since the increments $W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent, we have

$$\begin{aligned} & f_{W_{t_1},\ldots,W_{t_n}}(x_1, \ldots, x_n) = \\ & = f_{W_{t_1}, W_{t_2}-W_{t_1},\ldots,W_{t_n}-W_{t_{n-1}}}(x_1, x_2 - x_1, \ldots, x_n - x_{n-1}) = \\ & = f_{W_{t_1}}(x_1) \cdot f_{W_{t_2}-W_{t_1}}(x_2 - x_1) \cdot \ldots \cdot f_{W_{t_n}-W_{t_{n-1}}}(x_n - x_{n-1}) = \\ & = \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{{x_1}^2}{2t_1}} \cdot \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{(x_2-x_1)^2}{2(t_2-t_1)}} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi t_n - t_{n-1}}} e^{-\frac{(x_n-x_{n-1})^2}{2(t_n-t_{n-1})}}. \end{aligned}$$

The formula obtained can be used for many purposes.

2) Conditional distribution.

Let's use the formula above to solve one particular problem. Suppose we know that at time t BM has taken value $W_t = B$. Let s be an earlier time, $s < t$. What is the conditional distribution of W_s given the event $W_t = B$? It is known that the conditional density is the ratio of joint density and the density of the condition, we can calculate

$$f_{W_s|W_t}(x|B) = \frac{f_{W_s,W_t}(x,B)}{f_{W_t}(B)} =$$

$$= \frac{\frac{1}{\sqrt{2\pi s}} \cdot e^{\frac{x^2}{2s}} \cdot \frac{1}{\sqrt{2\pi(t-s)}} \cdot e^{\frac{(B-x)^2}{2(t-s)}}}{\frac{1}{\sqrt{2\pi t}} \cdot e^{\frac{B^2}{2t}}} = \ldots = \frac{1}{\sqrt{2\pi \frac{s}{t}(t-s)}} \cdot e^{-\frac{t(x-B\frac{s}{t})^2}{2s(t-s)}}.$$

Hence, the conditional distribution of W_s is normal distribution with mean $B \cdot \frac{s}{t}$ and variance $\frac{s(t-s)}{t}$.

3) First passage time.

Let $W_0 = 0$ and $a > 0$. We are interested in the time which elapses before BM attains the level a. We call it the *first passage time* to the point a and denote $T_a = \inf\{T : W_t = a\}$. T_a is a random variable since its value depends on the path of BM (on ω.) Let's find its distribution function $F_{T_a}(t) = P\{T_a \le t\}$. Using the formula of total probability, we have:

$$P\{W_t \ge a\} = P\{W_t \ge a|T_a \le t\} \cdot P\{T_a \le t\} + P\{W_t \ge a|T_a > t\} \cdot P\{T_a > t\}.$$

For symmetry of normal distribution $P\{W_t \ge a|T_a \le t\} = \frac{1}{2}$. At the same time obviously $P\{W_t \ge a|T_a > t\} = 0$. Therefore $P\{W_t \ge a\} = \frac{1}{2}P\{T_a \le t\}$, and

$$P\{T_a \le t\} = 2P\{W_t \ge a\}.$$

Since $W_t \sim N(0, \sqrt{t})$, we have

$$F_{T_a}(t) = P\{T_a \le t\} = 2\frac{1}{\sqrt{2\pi t}} \int_a^\infty e^{-\frac{x^2}{2t}} dx = 2[1 - \Phi\left(\frac{a}{\sqrt{t}}\right)].$$

If $a < 0$, then by symmetry $P\{T_a \le a\} = 2[1 - \Phi\left(\frac{|a|}{\sqrt{t}}\right)]$.

If $a = 0$, then $T_0 = 0$. Taking all together, we have that, for any a,

$$P\{T_a \le t\} = 2[1 - \Phi\left(\frac{|a|}{\sqrt{t}}\right)]. \tag{15}$$

By differentiating the distribution function above, one gets the density function of T_a. For $a > 0$ it calculates as

$$f_{T_A}(t) = F'_{T_a}(t) = -2\varphi(\frac{a}{\sqrt{t}}) \cdot (-\frac{a}{2}) \cdot t^{-\frac{3}{2}} = \frac{a}{\sqrt{2\pi t^3}}\ e^{-a^2/2t}.$$

This distribution is called *inverse Gaussian distribution* (also Wald distribution). From (15) it is also seen that if $t \to \infty$, then

$$\mathbf{P}(T_a < \infty) = \lim_{t\to\infty} \mathbf{P}(T_a \le t) = 1. \tag{16}$$

4) Maxima of Brownian motion .

If $a > 0$, then $P\{\max_{0\le s\le t} W_s \ge a\} = P\{T_a \le t\} = 2[1 - \Phi\left(\frac{|a|}{\sqrt{t}}\right)]$.

If $a < 0$, then $P\{\max_{0\le s\le t} W_s \ge a\} = 1$.

5) Brownian motion between two boundaries

Let $A > 0, B > 0$. Let us find the probability that, starting from 0, BM reaches level A before $-B$. Recall that in the case of SRW the answer to the same question is $\frac{B}{A+B}$. As the same answer remains true for any time and space steps sizes, we have

$$P\{W_t \text{ reaches } A \text{ before } -B\} = P\{T_A < T_B\} = \frac{B}{A+B}.$$

Exercise 1. *Let W_t be a standard Brownian motion. Assume that $W_1 = 2$. Find the probability that $W_5 < 0$.*

Without any given condition we would have $P\{W_5 < 0\} = \frac{1}{2}$, since $W_t \sim N(0, \sqrt{t})$.

However, under the condition $W_1 = 2$ the increment $W_5 - W_1 \sim N(0, \sqrt{4})$ – this distribution only depends on the length of the time interval and not on its location. Therefore we have

$$P\{W_5 < 0 | W_1 = 2\} = P\{W_5 - W_1 < -2 | W_1 = 2\} =$$

$$= P\{W_5 - W_1 < -2\} = P\{W_4 < -2\} =$$

$$= P\{N(0, \sqrt{4}) < -2\} = \Phi\left(\frac{-2}{2}\right) = \Phi(-1) = 0.16 \ .$$

9 Martingales with continous time

Many important concepts and results that are known for martingales with discrete time can be transferred to continuous time without any major change.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and let t be a real-valued parameter interpreted as time. Most often, t takes values from the half-line $\mathcal{R}^+$ or finite interval $[0, T]$.

Definition 7. *A family of σ-algebras $\{\mathcal{F}_t, t \geq 0\}$ is called a* filtration, *if 1) all its members $\mathcal{F}_t$ are sub-σ-algebras of $\mathcal{F}$ and, 2) for $s < t$ one has $\mathcal{F}_s \subseteq \mathcal{F}_t$.*

As in the case of dicsrete time, we are mainly interested in the *natural filtration* $\{\mathcal{F}_t^X, t \geq 0\}$, generated by a random process X. As before, $\mathcal{F}_t^X$ contains the information induced by the random process X within the time interval $[0, t]$. It means that an event $A \in \mathcal{F}_t^X$ if and only if one can decide whether A ocurred or not on the basis of the trajectory $\{X_s, 0 \leq s \leq t\}$ that the process X generates by time t.

Definition 8. *If $\{Y_t, t \geq 0\}$ is a random process such that for each t the random variable Y_t is $\mathcal{F}_t$-measurable, then it is said that the process Y is* adopted *to filtration $\{\mathcal{F}_t, t \geq 0\}$.*

Examples:

1. The random process $Z_t = \int_0^t X_s ds$ is adopted to the filtration $\{\mathcal{F}_t^X, t \geq 0\}$, since knowing the path of X within time interval $[0, t]$ is sufficient to determine Z_t.
2. The process $M_t = \max_{0 \leq s \leq t} W_s$ is adopted to the filtration $\{\mathcal{F}_t^W, t \geq 0\}$.
3. The process $Z_t = W_{t+1}^2 - W_t^2$ is not adopted to the filtration $\{\mathcal{F}_t^W, t \geq 0\}$.

Similarly to discrete time we define martingales with continous time.

Definition 9. *Let* $(\Omega, \mathcal{F}, \mathbf{P})$ *be a probability space endowed with a filtration* $\{\mathcal{F}_t, t \geq 0\}$. *A random process* $\{M_t, t \geq 0\}$ *is called a* martingale, *if*

1. M *is adopted to the filtration* $\{\mathcal{F}_t, t \geq 0\}$,
2. $\mathbf{E}|M_t| < \infty, \quad \forall t$
3. *for any* $s \leq t$ *we have* $\mathbf{E}(M_t|\mathcal{F}_s) = M_s \quad$ *a.s.*

If the equality in 3. is replaced by the inequality $\leq$ (or $\geq$), then we speak about a *supermartingale* (respectively *submartingale).*

Remark: Similarly, a martingale can be defined on a finite time interval $[0, T]$. In the following mostly the filtration $\{\mathcal{F}_t^W, t \geq 0\}$, generated by a standard Brownian motion W, is used.

Very important continous time martingales are related to Brownian motion.

Lemma 4. *Let* W_t *be a standard Brownian motion and let* $\{\mathcal{F}_t, t \geq 0\}$ *be the filtration induced by* W. *Then*

1. $W_t \quad$ *is a martingale,*
2. $W_t^2 - t \quad$ *is a martingale,*
3. $\exp(\sigma W_t - \frac{\sigma^2}{2} t) \quad$ *is a martingale (called* exponential martingale).

Proof: The proofs resemble each other. Consider, for example, the process $M_t = W_t^2 - t$. Obviously, $\mathbf{E}|M_t| < \infty$. Find now

$$\begin{aligned} \mathbf{E}(W_t^2 - W_s^2|\mathcal{F}_s) &= \mathbf{E}[(W_t - W_s)^2 + 2W_s(W_t - W_s)|\mathcal{F}_s)] \\ &= \mathbf{E}[(W_t - W_s)^2|\mathcal{F}_s] + 2W_s\mathbf{E}[(W_t - W_s)|\mathcal{F}_s] \\ &= t - s. \end{aligned}$$

So that

$$\begin{aligned} \mathbf{E}(W_t^2 - t|\mathcal{F}_s) &= \mathbf{E}(W_t^2 - W_s^2 + W_s^2 - (t - s) - s|\mathcal{F}_s) \\ &= (t - s) + W_s^2 - (t - s) - s = W_s^2 - s, \end{aligned}$$

and thus the process $W_t^2 - t$ is a martingale.

Definition 10. *A random variable T is called a* stopping time *with respect to the filtration $\{\mathcal{F}_t, t \geq 0\}$, if for each $t \geq 0$ the event $\{T \leq t\} \in \mathcal{F}_t$.*

Let's introduce a more general concept of local martingale.

Definition 11. *A process $\{X_t, t \geq 0\}$ is* local martingale, *if there exists a sequence of stopping times T_n, $n = 1, 2, \ldots$ such that the stopped process $\{X_{t \wedge T_n}, t \geq 0\}$ is a martingale for each n and*

$$\mathbf{P}[\lim_{t\to\infty} T_n = \infty] = 1.$$

Every martingale is a local martingale (how to choose T_n?), however, a local martingale need not be a martingale. This is the reason why we will assume some boundedness condition to be satisfied in the following.

It is very useful to have a 'continuous' version of Doob's convergence theorem. However, for that the trajectories of the martingale must be 'nice' enough. In all our examples below the trajectories are right continuous and have left limits or, in short, *càdlàg* (in French: *continues à droite, limites à gauche*).

For example, continuous functions (like trajectories of Brownian motion) are automatically *càdlàg.*

Theorem 8. (Stopping theorem in continous time)
If $\{M_t, t \geq 0\}$ is a càdlàg martingale w.r.t. the filtration $\{\mathcal{F}_t, t \geq 0\}$ and τ_1, τ_2 are bounded stopping times, $\tau_1 \leq \tau_2 \leq K$, then

$$\mathbf{E}|M_{\tau_2}| < \infty$$

and

$$\mathbf{E}(M_{\tau_2}|\mathcal{F}_{\tau_1}) = M_{\tau_1}, \quad \mathbf{P} - \text{a.s.} \tag{17}$$

Note: Let τ is a bounded stopping time. Then, by choosing $\tau_1 = 0$, $\tau_2 = \tau$, and by taking expectations from both sides in (17), we have $\mathbf{E}M_\tau = \mathbf{E}M_0$.

We next consider an interesting application of the stopping theorem.

Lemma 5. *Let W_t be a standard Brownian motion and let T_a be the first passage time of level a, $T_a = \inf\{t \geq 0 : W_t = a\}$. Then for any $\theta > 0$*

$$\mathbf{E}\left[e^{-\theta T_a}\right] = e^{-\sqrt{2\theta}|a|} \tag{18}$$

Proof. Assume that $a \geq 0$ (the case $a < 0$ can be handled by symmetry). Consider the martingale $M_t = \exp(\sigma W_t - \sigma^2 t/2)$. As the stopping time T_a is not bounded, Theorem 8 does not apply directly. Instead, let's consider bounded stopping times $\tau_1 = 0$ and $\tau_2 = T_a \wedge n$. Then, by Theorem 8, we have

$$\mathbf{E}(M_{T_a \wedge n}) = \mathbf{E}M_0 = 1, \ \forall n.$$

Show now that $\mathbf{E}M_{T_a} = \lim_n \mathbf{E}(M_{T_a \wedge n})$. First recall that $T_a < \infty$ a.s. (cf the formula (16)), hence starting from some value of n we have $T_a < n$ and therefore

$$M_{T_a \wedge n} \to M_{T_a} \ \ \text{a.s.}$$

At the same time the stopped martingale $M_{T_a \wedge n}$ is bounded from both sides: indeed, for $W_{T_a \wedge n} \leq a$ we have

$$0 \leq M_{T_a \wedge n} = exp\left(\sigma W_{T_a \wedge n} - \frac{\sigma^2}{2}(T_a \wedge n)\right) < e^{\sigma a}.$$

Therefore the bounded convergence theorem applies, giving us

$$\mathbf{E}M_{T_a} = \lim_n \mathbf{E}(M_{T_a \wedge n}) = 1.$$

From the other side, by the definition of M_t, the expectation $\mathbf{E}M_{T_a}$ can be written as

$$\mathbf{E}M_{T_a} = \mathbf{E}\left[e^{(\sigma a - \frac{1}{2}\sigma^2 T_a)}\right].$$

Hence we have

$$1 = \mathbf{E}\left[e^{(\sigma a - \frac{1}{2}\sigma^2 T_a)}\right],$$

and by choosing $\sigma^2 = 2\theta$, we obtain the required formula (18). $\square$

The Lemma above can be used, for example, for the calculation of $\mathbf{E}T_a$. (Expected value of T_a can be expressed via its moment generating function $m(\theta) = \mathbf{E}(e^{\theta T_a})$, namely, $\mathbf{E}T_a = m'(0)$.) Try to show that $\mathbf{E}T_a = \infty$!

10 Stochastic integral

Our aim here is to give a meaning to the integral $\int_0^t X_s\, dM_s$, where X is an adopted process and M is a martingale, e.g. Brownian motion. This new integral is rather different from the classical Stiltjes integral. In fact, we have already done a useful piece of work in this direction when considering discrete time martingales.

For discrete time martingales we have defined the martingale transform as the process

$$Y_n = \sum_{i=1}^{n} C_i(X_i - X_{i-1}) =: (C \bullet X)_n,$$

which was interpreted as the total winnings of a player after the game n (recall that C_i is the stake of the player on game i and $X_i - X_{i-1}$ is the net winnings per unit stake in game i.) Most importantly, we have shown (Theorem 2) that if the process C_i is predictable and X is a martingale, then the process Y is a martingale again. At the same time, the formula of the martingale transform above looks like a certain integral sum. Our next task is to define similar concept (stochastic integral) for continuous time martingales. Stochastic integral is an efficient tool to solve problems in various areas, including option pricing.

We first show that it is not possible to integrate with respect to continuous time martingales in a traditional manner i.e. path-wise (as Riemann- Stiltjes integral).

10.1 Variation. Quadratic variation of Brownian motion

Let us consider a function f of a real variable. Variability of the function f within the interval $[a, b]$ can be described by dividing the interval into smaller subintervals using cutting points $a = t_0 < t_1 < \ldots < t_n = b$ and by summing up the absolute values of the increments $|f(t_i) - f(t_{i-1}|$. Instead of absolute values, one could also use the squares, cubes etc. In order to avoid the dependence of the result on the interval partition method, we allow the number of cut-points n to approach infinity. We thus reach the following definition.

Definition 12. *The p-variation of a function f over the interval* $[a, b]$ *is defined as*

$$Var_p(f; a, b) = \limsup_{\|\pi_n\| \to 0} \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|^p, \}$$

where π_n *is a partition of* $[a, b]$ *by cutting points* $a = t_0 < t_1 < \ldots < t_n = b$*, and* $\|\pi_n\|$ *is the length of the longest subinterval of* π_n*.*

It is easy to see that when f is continuous and the partition is fine enough, so that the increments $|f(t_i) - f(t_{i-1}|$ are small numbers (smaller than 1), then the higher is the order p the less is the result. Thus for $p > q$ the p-variation is less than the q-variation.

At the same time, while studying martingales, we have used the term 'quadratic variation': the quadratic variation of a martingale M was a predictable process A such that the difference $M^2 - A$ is again martingale. For example, for a standard Brownian motion W the process $W_t^2 - t$ is a martingale, and hence t is the quadratic variation of standard Brownian motion. The question arises whether such a coincidence of terminology is justified. The positive answer is provided by the following lemma. Consider a partition of the interval $[0, t]$

$$\pi_n : 0 = t_0 < t_1 < \ldots < t_n = t$$

and let us denote

$$\begin{aligned}
\Delta_i &= t_i - t_{i-1},\ i = 1, 2, \ldots, n \\
\Delta W_i &= W_{t_i} - W_{t_{i-1}}, \\
Q_n(t) &= \sum_{i=1}^{n} (W_{t_i} - W_{t_{i-1}})^2 = \sum_{i=1}^{n} (\Delta W_i)^2.
\end{aligned}$$

Lemma 6. *The following (mean-square) convergence takes place:*

$$\mathbf{E}[Q_n(t) - t]^2 \to 0, \quad n \to \infty.$$

Proof. The increments of Brownian motion ΔW_i are independent and $\Delta W_i \sim N(0, \sqrt{\Delta_i})$. Therefore

$$\mathbf{E}Q_n(t) = \sum_{i=1}^{n} \mathbf{E}(\Delta W_i)^2 = \sum_{i=1}^{n} \Delta_i = t.$$

At the same time the variance

$$D(Q_n(t)) = \sum_{i=1}^{n} D((\Delta W_i)^2) = \sum_{i=1}^{n}[\mathbf{E}(\Delta W_i)^4 - \Delta_i^2].$$

It is well known that the 4-th order moment of a $N(0,1)$– distributed random variable is 3, thus $\mathbf{E}(W_1^4) = 3$. Hence, by taking into account the stationarity of increments, we have

$$\mathbf{E}(\Delta W_i)^4 = \mathbf{E}(W_{t_i} - W_{t_{i-1}})^4 = \mathbf{E}W^4_{t_i - t_{i-1}} = \mathbf{E}W^4_{\Delta_i} = \mathbf{E}\left(\sqrt{\Delta_i}W_1\right)^4 = 3\Delta_i^2,$$

from which

$$D(Q_n(t)) = 2\sum_{i=1}^{n}\Delta_i^2.$$

Therefore, if $\|\pi_n\| = \max \Delta_i \to 0$, then

$$D(Q_n(t)) \leq 2\|\pi_n\| \cdot \sum_{i=1}^{n}\Delta_i = 2t\ \|\pi_n\| \to 0.$$

Since

$$D(Q_n(t)) = \mathbf{E}(Q_n(t) - t)^2,$$

the proof is completed.□

Now it is not difficult to show that the variation (1-variation) of Brownian motion is unbounded.

Corollary. The trajectories of Brownian motion have a.s. unbounded variation, i.e. $Var_1(W;0,t) = \infty$.

Proof. Obviously, the following inequalities are valid:

$$\begin{aligned} Q_n(t) &= \sum_{i=1}^{n}(\Delta W_i)^2 \\ &\leq \max_i |\Delta W_i| \cdot \sum_{i=1}^{n}|\Delta W_i| \\ &\leq \max_i |\Delta W_i| \cdot Var_1(W;0,t). \end{aligned} \quad (19)$$

We now let $\|\pi_n\| = \max \Delta_i \to 0$. Since the trajectories of Brownian motion are uniformly continuous in the interval $[0,t]$, we have the convergence $\max_i |\Delta W_i| \to$

0. Suppose now, in contrary, that $Var_1(W;0,t) < \infty$. Then the product (19) also tends to zero, while a smaller quantity $Q_n(t)$ converges to t - a contradiction. Therefore, it must be that $Var_1(W;0,t) = \infty$. $\square$

The last fact makes it more complicated to integrate with respect to Brownian motion (and other martingales), as it becomes impossible to integrate along individual trajectories. Indeed, it is known that Riemann-Stiltjes integral $\int_0^1 f(t)dg(t)$ exists if both f and g have bounded variations and their discontinuity points do not overlap. [4] It is therefore necessary to introduce a new type of integrals (called stochastic integral) that we first do in a form of simple but important example.

10.2 First example of stochastic integral

In the next, we try to give meaning to the integral $\int_0^t W_s dW_s$. From the discussion above, it is clear that this integral can not exist in Riemann- Stiltjes sense, because of the unbounded variation of W_t. Thus for each separate trajectory the integral sums

$$S_n = \sum_{i=1}^{n} W_{t_{i-1}} \Delta W_i = \sum_{i=1}^{n} W_{t_{i-1}} (W_{t_i} - W_{t_{i-1}}) \tag{20}$$

do not converge as ordinary number sequence, when $n \to \infty$. However, by using $(W_{t_i} - W_{t_{i-1}})^2 = W_{t_i}^2 + W_{t_{i-1}}^2 - 2W_{t_i} W_{t_{i-1}}$, it is possible (after little algebra) to represent S_n in the form

$$S_n = \frac{1}{2} W_t^2 - \frac{1}{2} Q_n(t).$$

By Lemma 6, the mean-square convergence ($\mathcal{L}^2$- convergence) $S_n \to \frac{1}{2}W_t^2 - \frac{1}{2}t$ takes place. The limit $\frac{1}{2}W_t^2 - \frac{1}{2}t =: I_t$ is called *stochastic integral* and we write

$$\int_0^t W_s dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} t. \tag{21}$$

We see that, in addition to Riemann-Stiltjes integral, an additional term $-t/2$ has appeared. Let us take together the discussion above:

[4] Recent studies, however, have shown that for the existence of Riemann-Stiltjes integral it is sufficient that f has bounded p-variation and g has bounded q-variation, where p, $q > 0$ satisfy $1/p + 1/q > 1$.

- For each separate trajectory $W_t(\omega)$ the integral sum $S_n(\omega)$ does not converge.
- At the same time, the average of S_n (over all ω) is equal to $I_t = \frac{1}{2}W_t^2 - \frac{1}{2}t$, so that $\mathbf{E}(S_n - I_t) = 0$ for each n.
- S_n fluctuates around the process I_t and its deviation from I_t vanishes when n increases to infinity: for each t we have $\mathbf{E}(S_n - I_t)^2 \to 0$ (convergence in $\mathcal{L}^2$).

Thus the stochastic integral I_t is a random process which is not necessarily close to the integral sum S_n for each separate trajectory, but it works well (as much as possible) for all trajectories simultaneously.

It is important to note, that the integral sum $S_n = \sum_{i=1}^n W_{t_{i-1}}(W_{t_i} - W_{t_{i-1}})$ is a martingale transform and therefore (by Theorem 2) it is a martingale. We also know that the limit process $I_t = \frac{1}{2}W_t^2 - \frac{t}{2}$ is a martingale (see Lemma 4, assertion 2)).

Comment: If we define S_n slightly differently from (20), $S_n = \sum_{i=1}^n W_{\bar{t}_i}\Delta W_i$, using $\bar{t}_i = (t_{i-1} + t_i)/2$ (instead of t_{i-1}), the process S_n stops to be a martingale. The integral obtained by such an alternative construction is called *Stratonovich integral* and it has useful technical applications.

We now proceed to a more general definition of the stochastic integral. In doing that we do not restrict ourselves with integration merely w.r.t. the Brownian motion.

10.3 Definition of stochastic integral

In order to define a stochastic integral in general case, we follow the standard scheme:
1) the stochastic integral is first defined for "simple" processes (resembles a martingale transform),

2) a general process is represented as the limit of simple processes,

3) the stochastic integral of a general process is defined as the limit of integrals of respective simple processes.

10.3.1 Stochastic integral of simple process

Let $T = [0, \infty)$ and let $\{\mathcal{F}_t, t \in T\}$ be a filtration. Let M be a continuous square integrable martingale w.r.t. the filtration $\{\mathcal{F}_t, t \in T\}$.

Definition 13. *The process η is called a* simple process *if there exists a finite number of time instances $0 = t_0 < t_1 < \cdots < t_n = \infty$ and random variables $\xi_i,\ i = 1, 2, \ldots, n$, with finite variances such that ξ_i is $\mathcal{F}_{t_{i-1}}$-measurable for each i and*

$$\eta_t = \sum_{i=1}^{n} I_{[t_{i-1}, t_i)}(t)\xi_i.$$

Definition 14. *The stochastic integral of a simple process η with respect to a martingale M is defined as the process*

$$\begin{aligned} Int_t = \int_0^t \eta_s \, dM_s : &= \sum_{i=1}^{n_t} \xi_i (M_{t_i} - M_{t_{i-1}}) + \xi_{n_t+1}(M_t - M_{t_{n_t}}) \\ &\equiv \sum_{i=1}^{n} \xi_i (M_{t \wedge t_i} - M_{t \wedge t_{i-1}}), \end{aligned}$$

where n_t is an integer such that $t_{n_t} \le t < t_{n_t+1}$.

Notice that the last formula has the structure of a martingale transform (3). Therefore, by Theorem 2, the process Int_{t_i} is a martingale as well.

It is easy to see that by choosing $\eta \equiv 1$, we obtain an useful formula

$$\int_0^t dM_s = M_t - M_0,$$

as in the case of classical R-S integral.

10.3.2 Stochastic integral of continuous process

Let M be a martingale and let Z be an adopted and continuous process such that

$$\mathbf{E}(\int_0^t Z_s^2 \, d\langle M\rangle_s) < \infty \ \forall t > 0.$$

Definition 15. *The stochastic integral of a process Z with respect to a martingale M is defined as the limit (in $\mathcal{L}^2$)*

$$Int_t = \int_0^t Z_s \, dM_s = \lim_{\|\pi_n\| \to 0} \sum_{i=1}^{n} Z_{t_{i-1}} (M_{t_i} - M_{t_{i-1}}),$$

where π_n is a partition of $[0,t]$ into n subintervals:

$$0 = t_0 < t_1 < \cdots < t_n = t.$$

Theorem 9. *Stochastic integral has following properties:*

(i) The process $(Int_t)_{t\in[0,\infty)}$ is a square integrable martingale.

(ii) The quadratic variation of Int is the process

$$\langle Int\rangle_t = \int_0^t Z_s^2 \, d\langle M\rangle_s.$$

Proof: The properties are relatively easy to prove for simple processes. The second property is known as *isometry.*

Generally speaking, the definition above is not suitable for immediate calculations, except for simple processes. (The same is true for classical Riemann integral!). Therefore certain rules have been worked out to calculate stochastic integrals in practice. Here the central role is played by so called *Itô formula.*

11 Itô formula and its applications

Consider an stochastic process X_t, which can be decomposed as

$$X_t = X_0 + V_t + M_t, \tag{22}$$

where
– M_t is a martingale with continuous trajectories that have non-zero quadratic variation, $Var_2 > 0$ (such trajectories are called 'rough', e.g. Brownian motion),
– V_t is an adopted process with continuously differentiable (or 'smooth') trajectories that have finite variation, $Var_1 < \infty$, and zero quadratic variation, $Var_2 = 0$.

Such processes X_t are called *semimartingales.*

We are interested in an expression for the increment $f(X_t) - f(X_0)$ for some functions f.

In classical analysis, it is known that if X_t is a "common" process (e.g. with continuously differentiable trajectories) then the Newton-Leibnizi formula gives

$$f(X_t) - f(X_0) = \int_0^t f'(X_s) dX_s.$$

However, for more general processes (22) the answer is different, as shows the Itô formula below.

11.1 Itô formula

Let f be twice continuously differentiable function. Then

$$f(X_t) - f(X_0) = \int_0^t f'(X_s) dV_s + \int_0^t f'(X_s) dM_s + \frac{1}{2} \int_0^t f''(X_s) d\langle M \rangle_s \, . \tag{23}$$

The second integral at the right hand side of Itô formula is a stochastic integral, whereas the other two are traditional (Riemann-Stiltjes) integrals. Hence, the stochastic integral can be expressed in terms of usual integrals. This is why the

Itô formula is so important.

Example 1. Let W_t be standard Brownian motion and $X_t = W_t$ (i.e. the decomposition (22) contains only the martingale part while $X_0 = V_t = 0$. Using Itô formula, show that $\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t$ (that, in fact, we know already from before) (21).)
Hint: Take $f(x) = x^2$.

The proof of Itô formula. We present only the basic idea of the proof. For simplicity, let $V_t = 0$ and let $M_t = W_t$ - a standard Brownian motion. Since the quadratic variation is $\langle W \rangle_s = s$, we need to show that

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)dW_s + \frac{1}{2}\int_0^t f''(X_s)ds. \tag{24}$$

Let us consider a partition of the interval $[0, t]$ by points $0 = t_0 < t_1 < \ldots < t_n = t$. Then

$$f(X_t) - f(X_0) \equiv \sum_{i=1}^{n} \left[f(X_{t_i}) - f(X_{t_{i-1}})\right].$$

In each subinterval we apply usual Taylor's formula:

$$f(X_t) - f(X_0) = \sum_{i=1}^{n} f'(X_{t_{i-1}})(W_{t_i} - W_{t_{i-1}}) + \frac{1}{2}\sum_{i=1}^{n} f''(X_{\xi_i})(W_{t_i} - W_{t_{i-1}})^2,$$

where $\xi_i \in [t_i, t_{i-1}]$. When the partition gets finer, so that the maximum length of subintervals tends to zero, the first sum converges (in mean square) to stochastic integral $\int_0^t f'(X_s)dW_s$. At the same time, in the second term the square of the increment of the Brownian motion $(W_{t_i} - W_{t_{i-1}})^2$ can be approximated by its mean value $\Delta_i = t_i - t_{i-1}$ (related calculations were made in the proof of the lemma 6 where it came out that the variance $D((W_{t_i} - W_{t_{i-1}})^2) = 2\Delta_i^2 \to 0$). Therefore, the second term can be approximated by the expression

$\frac{1}{2}\sum_{i=1}^n f''(X_{\xi_i})\Delta_i$, which converges to usual integral $\frac{1}{2}\int_0^t f''(X_s)ds$. $\square$

Example 2. A widely used model in financial mathematics to describe the behaviour of a stock price S_t is the following:

$$dS_t = S_t\,(\mu dt + \sigma dW_t), \tag{25}$$

where
μ shows relative change of the price per time unit (a constant in this model),
dW_t is the random part of the price change, that within a short time interval Δt behaves like an increment of a Brownian motion ,
σ shows the importance of the random component in the development of the price (volatility parameter).
The relationship above is, in fact, a short notation of the following equation:

$$S_t - S_0 = \int_0^t S_s \mu ds + \int_0^t S_s \sigma dW_s. \tag{26}$$

The equation (25) is called *stochastic differential equation* (SDE) . Our aim is to solve this SDE for S_t. It turns out that the easiest way to do that is first to find $ln(S_t/S_0)$.
Hint: apply Itô formula for $f(x) = lnx$.

11.2 Itô formula in differential form

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)d\langle M\rangle_t \ . \tag{27}$$

The formula (27) comes immediately from the Itô formula (23), being simply its shorter (and more convenient) notation. Here the expression $d\langle M\rangle_t$ is the usual differential of the function $\langle M\rangle_t$, however dX_t is a symbol of (nn *stochastic differential*), whose precise meaning is given by the notion of stochastic integral $\int_0^t f'(X_s)dX_s$.

Example 3. Solve the problem in Example 2, using the differential form of Itô formula.

11.3 A generalization of Itô formula

Let the function f also depend on time, $f = f(X_t, t)$, where the process X_t is still of the form (22). Assuming that the function $f(x,t)$ is smooth enough (sufficiently times differentiable), the following formula is valid:

$$df(X_t,t) = \frac{\partial f}{\partial t}(X_t,t)\ dt + \frac{\partial f}{\partial x}(X_t,t)\ dX_t + \frac{1}{2}\,\frac{\partial^2 f}{\partial x^2}(X_t,t)\ d\langle M\rangle_t\ . \tag{28}$$

Example 4. Assume that a stock price S_t behaves in accordance with the following model:

$$dS_t = S_t\,(rdt + \sigma dW_t),$$

where r is the risk free interest rate and σ is the price volatility. Show that then the discounted price process $e^{-rt}S_t$ is a martingale.
Hint: use $f(x,t) = e^{-rt}x$.

Example 5. Find $d\left(e^{-rt}V(S_t,t)\right)$, where $V(s,t)$ is known twice continuously differentiable function and the price process S_t is driven by the equation $dS_t = S_t\left(rdt + \sigma dW_t\right)$.

Hint: Take $f(s,t) = e^{-rt}V(s,t)$.

The result is the following stochastic differential equation

$$d\left(e^{-rt}V(S_t,t)\right) = e^{-rt}\left[-rV(S_t,t) + \frac{\partial V}{\partial t} + rS_t\frac{\partial V}{\partial s} + \frac{1}{2}S^2\sigma^2\frac{\partial^2 V}{\partial s^2}\right]dt + e^{-rt}\sigma S_t\frac{\partial V}{\partial s}dW_t$$

Corrollary: If the function $V = V(s,t)$ satisfies the relationship

$$\frac{\partial V}{\partial t} + rs\frac{\partial V}{\partial s} + \frac{1}{2}s^2\sigma^2\frac{\partial^2 V}{\partial s^2} - rV = 0, \quad \forall s,t \tag{29}$$

then the coefficient of the term dt becomes zero and the result is the martingale

$$d\left(e^{-rt}V(S_t,t)\right) = e^{-rt}\sigma S_t\frac{\partial V}{\partial s}\,dW_t.$$

We have discovered a relationship between random processes (SDE) and differential equations (PDE). The formula (29) is called Black-Scholes equation.

The fact that, under some circumstances, $e^{-rt}V(S_t,t)$ is a martingale can be used to solve several problems in financial mathematics. This is based on exploiting the martingale's property to keep its mean value over the time. Let us consider an example of this type.

Example 6. Option pricing

11.4 Itô formula for functions of several variables

Consider a function of several variables in the form $f(x_1, x_2, \ldots, x_n, t)$, for example $\sum_{i=1}^{n} x_i + t$. Our aim is to find the stochastic differential $d\, f(X_{1t}, X_{2t}, \ldots, X_{nt}, t)$. We first introduce the concept of cross-variation.

Definition 16. The cross-variation *of martingales M and N is the process*

$$\langle M, N\rangle_t = \frac{1}{4}\left(\langle M+N\rangle_t - \langle M-N\rangle_t\right).$$

Some most important properties of cross-variation are:

1. $\langle M, M\rangle_t = \langle M\rangle_t$. (It suffices to take $M = N$ in the definition)
2. If $Int_{1t} = \int_0^t \xi_s \, dM_s$ ja $Int_{2t} = \int_0^t \eta_s \, dN_s$, then
$$\langle Int_1, Int_2\rangle_t = \int_0^t \xi_s \eta_s \; d\langle M, N\rangle_s.$$
3. For independent Brownian motions $\langle W_1, W_2\rangle_t = 0$.

Itô formula for functions of several variables

If f is twice continuously differentiable function, then

$$\begin{aligned} d\, f(X_{1t}, X_{2t}, \ldots, X_{nt}, t) &= \frac{\partial f}{\partial t}(\overrightarrow{X}_t, t)\, dt + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\overrightarrow{X}_t, t)\; dX_{it} + \\ &+ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\overrightarrow{X}_t, t)\; d\langle M_i, M_j\rangle_t \end{aligned}$$

where X_{it} is of the form $X_{it} = X_{i_0} + V_{it} + M_{it}$,
V is a term with 'nice' properties (having finite variation), and M is a martingale.

Näide 7.

12 Change of measure. Girsanov's theorem

12.1 Options

An *option* is a contract between a buyer and a seller that gives the buyer of the option the right, but not the obligation, to buy or to sell a specified asset (underlying) on or before the option's expiration time, at an agreed price, the strike price. In return for granting the option, the seller collects a payment (the premium) from the buyer. Granting the option is also referred to as "selling" or "writing" the option. A *call* option gives the buyer of the option the right but not the obligation to buy the underlying at the strike price. A *put* option gives the buyer of the option the right but not the obligation to sell the underlying at the strike price.

If the buyer chooses to exercise this right, the seller is obliged to sell or buy the asset at the agreed price. If the option may only be exercised at expiration time T then it is called European option. An American option, in contrary, may be exercised on any trading day on or before expiration.

The price of the underlying S_t can rise by time T higher than strike price K. In such case the owner of the option at time T executes his right to buy shares at price K and sells them at the market price, earning $S_T - K$ per share. At the same time, the seller of the option can lose the same amount of money, provided he does not undertake anything during the option period. However, as we will see below, the seller can take measures to avoid the loss.

Since the option never results in negative value, more exactly, at time T its value is $C(S_T) := \max(S_T - K, 0)$, then it is natural that in return for granting the option, the seller collects a payment (the premium) from the buyer. We see two problems related to options:

- What is the fair *price* of the option (at the time of writing it but also at any later time since we may want to sell the option on)?

- How can the seller (writer) of the option avoid possible loss caused by an unfavorable change of the price of the underlying?

12.2 Option pricing

The basis for pricing an option is so called *parity principle*. According to that principle, the fair price is the amount of money X_0 such that if the seller, starting trading with that amount at time 0 and using appropriate trading strategy, can reach the same value X_T at time T as the option itself , i.e. $X_T = C(S_T) = \max(S_T - K, 0)$.

It is not difficult to show that when the option is priced differently from X_0, there will be a possibility for *arbitrage* where one side of the agreement (the buyer or the seller) can earn a risk free profit.

In the following we assume that the stock price S_t is driven by the model

$$d\,S_t = S_t\,(\mu dt + \sigma\, dW_t), \tag{30}$$

where μ is a constant component of the return, called *drift*, and σ is a constant showing the importance of the random component of the return, called *volatility*.

Now consider an investor (= seller of the option) who invests the amount X_0 at time 0 partly in shares and puts the rest of the money onto a bank account with risk free interest rate r. Then he starts trading the same stock putting the money earned to the bank accunt and taking the money necessary for purchase of shares from the same account (self-financing portfolio).

Let $\eta(t)$ be the number of shares owned by the investor at time t. Then the value X_t of the portfolio at time t consists of two parts: the value of shares $\eta(t) \cdot S_t$ plus the money in the bank $X_t - \eta(t)\,S_t$. Let's find the increment of the value of such portfolio during a short time interval dt:

$$dX_t = r\,[X_t - \eta(t)\,S_t]\,dt + \eta(t) \cdot dS_t.$$

By using Itô formula and the market model (30) one can also calculate the change of the discounted wealth:

$$d(e^{-rt}X_t) = \eta(t)d(e^{-rt}S_t)$$

(show this!). From here it is seen that as soon as the discounted price process $e^{-rt}S_t$ is a martingale, so does the discounted wealth $e^{-rt}X_t$. However, any martingale keeps its mean value over the time. Hence, its (non-random) initial value X_0 is equal to the mean value at any later time epoch, including time T:

$$X_0 = e^{-r0}X_0 = \mathbf{E}(e^{-rT}X_T) = \mathbf{E}(e^{-rT}C(S_T)) \tag{31}$$

Only one question remains: whether the discounted price process $e^{-rt}S_t$ is a martingale? In order to decide that, we will find its stochastic differential, using Itô formula:

$$d\left(e^{-rt}S_t\right) = e^{-rt}S_t\,\sigma\left(\frac{\mu - r}{\sigma}dt + dW_t\right).$$

The result will definitely be a martingale if $\mu = r$, since then the right hand side reduces to the stochastic integral w.r.t. W_t, which is, as we know, a martingale. However, it turns out that even when $\mu \neq r$ the whole expression in the brackets in the last formula, $\widetilde{W}_t = \frac{\mu - r}{\sigma}\,t + W_t$, can be regarded as a martingale. For that, it is only necessary to redefine appropriately the probabilities of the trajectories of the initial Brownian motion W_t. In other words, the measure $\mathbf{P}$ must be changed in such a way that $\widetilde{W}_t$ is martingale with respect to the new measure $\mathbf{Q}$. How this can be achieved? This possibility is granted by Girsanov's Theorem (also called Cameron-Martin Theorem) which also indicates the relationship between the old and new meausres.

Theorem 10. (Girsanov's Theorem)
Let W_t be a standard Brownian motion w.r.t. probability space $(\Omega, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{P})$. Let $\widetilde{W}_t = \int_0^t \theta_s\,ds + W_t$, where the process θ is adopted to the filtration $\{\mathcal{F}_t\}$ and satisfies the requirement $\mathbf{E}\left(e^{\frac{1}{2}\int_0^T \theta_s^2 ds}\right) < \infty$.
Then there exists a probability measure $\mathbf{Q}$ such that $\widetilde{W}_t$ is a standard Brownian motion w.r.t. that measure. The measure $\mathbf{Q}$ is related to $\mathbf{P}$ via the formula

$\mathbf{Q}(A) = \int_A M_T(\omega)\mathbf{P}(d\omega)$, *where*

$$M_t = \exp\left(- \int_0^t \theta_s \, dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds \right).$$

The measure $\mathbf{Q}$ is called *risk neutral* or martingale measure. Let us apply Girsanov's Theorem in our case where we have $\theta_t \equiv \frac{\mu - r}{\sigma}$. The assumption $\mathbf{E}\left(e^{\frac{1}{2}\int_0^T \theta_s^2 ds} \right) < \infty$ is clearly fulfilled. Thus the existence of the martingale measure $\mathbf{Q}$ is guaranteed, and the option price is the expected value w.r.t. to that new measure $\mathbf{Q}$:

$$X_0 = e^{-r0} X_0 = \mathbf{E}^{\mathbf{Q}}(e^{-rT} C(S_T)). \tag{32}$$

Finally, let's explain the term "risk neutral measure". As

$$dW_t = d\widetilde{W}_t - \frac{\mu - r}{\sigma} dt,$$

the market model (30) reduces (after simple calculations) to

$$d\, S_t = S_t \,(rdt + \sigma \, d\widetilde{W}_t). \tag{33}$$

The formula differs from the model (30) only in that μ is replaced by the risk free interest rate r. We see that in the case of the measure $\mathbf{Q}$ the average return of the stock price is equal to the risk free interest rate r. Hence the term interest "risk neutral measure".

Some remarks on how to calculate the option price as the expected value $X_0 = \mathbf{E}^{\mathbf{Q}}(e^{-rT} C(S_T))$. Basically, there are two possibilities for that:

1. **Simulation** where a large number of price trajectories is generated, in accordance with the market model (33). For each trajectory $C(S_T)$ is found, the results are averaged and multiplied by the discounting factor e^{-rT}.

2. **Analytical** calculation of the expectation. This is only possible for some simple pay-off functions $C(S_t)$, like for example European type options. In the latter case the result is well known as *Black-Scholes* pricing formula. (a good exercise!)