1.9 Random vectors

Preliminary. - The Borel σ-field on $\mathbb{R}^n$ is defined as

$$\mathcal{B}(\mathbb{R}^n) = \sigma\left(\{]a_1, b_1[\times \ldots \times]a_n, b_n[: \; a_i < b_i, \, \forall i\}\right)$$

It contains nearly all possible subsets of $\mathbb{R}^n$ (not only rectangles!).

- The Lebesgue measure on $\mathbb{R}^n$ is defined as

$$|\,]a_1, b_1[\times \ldots \times]a_n, b_n[\, | = \prod_{i=1}^{n} (b_i - a_i)$$

and can be extended by σ-additivity to any Borel subset of $\mathbb{R}^n$.

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

Definition 1.31. A *random vector of dimension* $n \geq 1$ is a map $X : \begin{cases} \Omega \to \mathbb{R}^n \\ \omega \mapsto X(\omega) = (X_1(\omega), \ldots, X_n(\omega)) \end{cases}$ such that

$$\{\omega \in \Omega \; : \; X(\omega) \in B\} \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbb{R}^n).$$

Proposition 1.32. - $X = (X_1, \ldots, X_n)$ is a random vector if and only if

$$\{\omega \in \Omega \; : \; X_1(\omega) \leq t_1, \ldots, X_n(\omega) \leq t_n\} \in \mathcal{F}, \quad \forall t_1, \ldots, t_n \in \mathbb{R}.$$

- If X is a random vector, then each component X_i is a random variable, but the reciprocal statement is wrong.

Two important classes of random vectors.

Discrete random vectors.

Definition 1.33. X is a discrete random vector if it takes values in a countable subset C of $\mathbb{R}^n$.

If $g : \mathbb{R}^n \to \mathbb{R}$ is a Borel-measurable and bounded function, then

$$\mathbb{E}(g(X_1, \ldots, X_n)) = \sum_{x_1, \ldots, x_n \in C} g(x_1, \ldots, x_n)\, \mathbb{P}(\{X_1 = x_1, \ldots, X_n = x_n\}).$$

Proposition 1.34. X is a discrete random vector if and only if each X_i, $1 \leq i \leq n$, is a discrete random variable.

Continuous random vectors.

Definition 1.35. X is a *continuous random vector* if $\mathbb{P}(\{X \in B\}) = 0$ for every $B \in \mathcal{B}(\mathbb{R}^n)$ such that $|B| = 0$.

Fact. If X is a continuous random vector, then there exists a Borel-measurable function $f_X : \mathbb{R}^n \to \mathbb{R}$ such that

$$f_X(x_1, \ldots, x_n) \geq 0, \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n, \quad \int_{\mathbb{R}^n} f_X(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1$$

and

$$\mathbb{P}\left(\{X \in B\}\right) = \int_B f_X(x_1, \ldots, x_n)\, dx_1 \cdots dx_n, \quad \forall B \in \mathcal{B}(\mathbb{R}^n).$$

Terminology. f_X is called the joint probability density function of the random vector X.

If $g : \mathbb{R}^n \to \mathbb{R}$ is a Borel-measurable and bounded function then

$$\mathbb{E}(g(X_1, \ldots, X_n)) = \int_{\mathbb{R}^n} g(x_1, \ldots, x_n)\, f_X(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$

Proposition 1.36. If X is a continuous random vector, then each X_i, $1 \leq i \leq n$, is a continuous random variable, but the reciprocal statement is wrong.

Here is a counter-example. Let Y be a continuous random variable and $X = (Y, Y)$; then X is not a continuous random vector (indeed, let $\Delta = \{(x, y) \in \mathbb{R}^2 \ : \ x = y\}$: $|\Delta| = 0$, but $\mathbb{P}(\{X \in \Delta\}) = 1$).

Expectation and covariance of random vector.

Let X be an n-dimensional random vector such that each component X_i, $1 \leq i \leq n$, is square-integrable.

Expectation (or mean) of X : $\mathbb{E}(X) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_n))$, n-variate vector.

Covariance (matrix) of X : $\text{Cov}(X) = K$, $n \times n$ matrix, where

$$K_{ij} = \text{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\,\mathbb{E}(X_j).$$

Properties. - K is symmetric, i.e. $K_{ij} = K_{ji}$.

- K is positive semi-definite, i.e. $\forall c_1, \ldots c_n \in \mathbb{R}$, $\sum_{i,j=1}^n c_i\, c_j\, K_{ij} \geq 0$. Indeed,

$$\sum_{i,j=1}^n c_i\, c_j\, K_{ij} = \sum_{i,j=1}^n c_i\, c_j\, \text{Cov}(X_i, X_j) = \text{Cov}\left(\sum_{i=1}^n c_i\, X_i, \sum_{j=1}^n c_j\, X_j\right) = \text{Var}\left(\sum_{i=1}^n c_i\, X_i\right) \geq 0.$$

Remark. If K is a symmetric $n \times n$ matrix, then K is positive semi-definite if and only if all its eigenvalues are non-negative.

Proposition 1.37. If $X_1, \ldots, X_n$ are independent and square-integrable then $X = (X_1, \ldots, X_n)$ is a random vector and $\text{Cov}(X)$ is a diagonal matrix (i.e. $\text{Cov}(X_i, X_j) = 0$, $\forall i \neq j$).

Gaussian random vectors.

Convention. If $Y(\omega) = c$, $\forall \omega \in \Omega$, then Y is said to be a Gaussian random variable with mean c and variance 0 ($Y \sim \mathcal{N}(c, 0)$).

Definition 1.38. A random vector $X = (X_1, \ldots, X_n)$ is Gaussian if $\forall c_1, \ldots, c_n \in \mathbb{R}$, $c_1 X_1 + \ldots + c_n X_n$ is a Gaussian random variable (possibly with variance 0).

Remark. This is more than saying that every X_i, $1 \leq i \leq n$, is Gaussian (see below)!

Proposition 1.39. If $X_1, \ldots, X_n$ are independent Gaussian random variables, then $X = (X_1, \ldots X_n)$ is a Gaussian random vector.

Proposition 1.40. Let $X = (X_1, \ldots, X_n)$ be a Gaussian random vector. Then the random variables $X_1, \ldots, X_n$ are independent if and only if $\text{Cov}(X)$ is a diagonal matrix.

But. If $X_1, \ldots, X_n$ are Gaussian random variables, then it is not necessarily true that $X = (X_1, \ldots, X_n)$ is a Gaussian random vector. Also $\text{Cov}(X_i, X_j) = 0$, $\forall i \neq j$ does not imply in general that $X_1, \ldots, X_n$ are independent!

Remark. If X is a Gaussian random vector with mean m and covariance K, this is denoted as $X \sim \mathcal{N}(m, K)$. Moreover, X is entirely characterized by its mean m and its covariance K.

Let X be an n-dimensional Gaussian random vector with mean $m = \mathbb{E}(X)$ and covariance $K = \text{Cov}(X)$.

Definition 1.41. X is *non-degenerate* if $\text{rank}(K) = n$.

Reminder. Let K be an $n \times n$ symmetric matrix.

- $\text{rank}(K) = n$ if and only if K is invertible if and only if $\det(K) \neq 0$ if and only if all its eigenvalues are non-zero.

- More generally, $\text{rank}(K) =$ number of non-zero eigenvalues of K.

Proposition 1.42. Let X be a non-degenerate n-dimensional Gaussian random vector, with mean m and covariance K. Then X is a continuous random vector with joint pdf

$$f_X(x_1,\ldots,x_n) = \frac{1}{\sqrt{(2\pi)^n \det(K)}} \exp\left(-\frac{1}{2} \sum_{i,j=1}^n (x_i - m_i)\,(K^{-1})_{ij}\,(x_j - m_j) \right), \quad (x_1,\ldots,x_n) \in \mathbb{R}^n.$$

Proposition 1.43. Let X be an n-dimensional Gaussian random vector with mean 0 and covariance K. Let also $k = \text{rank}(K) \in \{0,\ldots,n\}$. Then there exist k i.i.d. random variables $U_1,\ldots,U_k \sim \mathcal{N}(0,1)$ and $\alpha_{ij} \in \mathbb{R}$ ($1 \le i \le n$, $1 \le j \le k$) such that $X_i = \sum_{j=1}^k \alpha_{ij}\,U_j$, $i = 1,\ldots,n$.

Remark. In matrix form, $X = AU$ and $AA^T = K$.

Example. Let $X = (Y,Y)$, where $Y \sim \mathcal{N}(0,1)$. In this simple case, we have

$$\text{Cov}(X) = K = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

so $k = 1$, and $U_1 = Y$, $\alpha_{11} = \alpha_{21} = 1$.

2 Discrete-time stochastic processes

A discrete-time stochastic process can be viewed

a) either as a collection of random variables $(X_n,\, n \in \mathbb{N})$

b) or as a random sequence $X : \begin{cases} \Omega & \to & \mathbb{R}^{\mathbb{N}} \\ \omega & \mapsto & (X_n(\omega),\, n \in \mathbb{N}) \end{cases}$

Canonical example: the random walk.

Let $(\xi_n,\, n \ge 1)$ be a collection of i.i.d. random variables such that $\mathbb{P}(\{\xi_1 = +1\}) = \mathbb{P}(\{\xi_1 = -1\}) = 1/2$. Let $S_0 = 0$, $S_n = \xi_1 + \ldots + \xi_n$, $n \ge 1$: the process $(S_n,\, n \in \mathbb{N})$ is called the simple symmetric random walk. As simple as it may be, this process exhibits already many fascinating properties.

Remarks. - $\mathbb{E}(\xi_1) = 0$, so $\mathbb{E}(S_n) = \mathbb{E}(\xi_1) + \ldots + \mathbb{E}(\xi_n) = 0$.

- $\text{Var}(\xi_1) = 1$, so by independence, $\text{Var}(S_n) = \text{Var}(\xi_1) + \ldots + \text{Var}(\xi_n) = n$.

- By the (strong) law of large numbers, we know that $\dfrac{S_n}{n} \underset{n\to\infty}{\to} 0$ a.s.

- By the central limit theorem, we also know that $\dfrac{S_n}{\sqrt{n}} \overset{d}{\to} Z \sim \mathcal{N}(0,1)$, i.e. $S_n \simeq \sqrt{n}\,Z$.

Variations.

- simple asymmetric random walk: $S_n = \xi_1 + \ldots + \xi_n$, with $\mathbb{P}(\{\xi_1 = +1\}) = p = 1 - \mathbb{P}(\{\xi_1 = -1\})$ and $p \neq 1/2$.

- random walk with values in $\mathbb{R}$: $S_n = \xi_1 + \ldots + \xi_n$ with $\xi_1 \sim \mathcal{N}(0,1)$ e.g.

- continuous-time random walk = Brownian motion (see next chapter).

2.1 Martingales

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

Definition 2.1. A *filtration* is a sequence $(\mathcal{F}_n,\, n \in \mathbb{N})$ of sub-σ-fields of $\mathcal{F}$ such that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, $\forall n \in \mathbb{N}$.

Example. Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $X_n(\omega) = n^{th}$ decimal of ω, for $n \geq 1$. Let also $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Then $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, $\forall n \in \mathbb{N}$.

Definitions 2.2. - A discrete-time process $(X_n,\, n \in \mathbb{N})$ is said to be *adapted* to the filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$ if X_n is $\mathcal{F}_n$-measurable $\forall n \in \mathbb{N}$.

- The *natural filtration* of a process $(X_n,\, n \in \mathbb{N})$ is defined as $\mathcal{F}_n^X = \sigma(X_0, \ldots, X_n)$, $n \in \mathbb{N}$. It represents the available amount of information about the process at time n.

Remark. A process is adapted to its natural filtration, by definition.

Let now $(\mathcal{F}_n,\, n \in \mathbb{N})$ be a given filtration.

Definition 2.3. A discrete-time process $(M_n,\, n \in \mathbb{N})$ is a *martingale* with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ if

(i) $\mathbb{E}(|M_n|) < \infty$, $\forall n \in \mathbb{N}$.

(ii) M_n is $\mathcal{F}_n$-measurable, $\forall n \in \mathbb{N}$ (i.e., $(M_n,\, n \in \mathbb{N})$ is adapted to $(\mathcal{F}_n,\, n \in \mathbb{N})$).

(iii) $\mathbb{E}(M_{n+1}|\mathcal{F}_n) = M_n$ a.s., $\forall n \in \mathbb{N}$.

A martingale is therefore a fair game: the expectation of the process at time $n+1$ given the information at time n is equal to the value of the process at time n.

Remark. Conditions (ii) and (iii) are actually redundant, as (iii) implies (ii).

Properties. If $(M_n,\, n \in \mathbb{N})$ is a martingale, then

- $\mathbb{E}(M_{n+1}) = \mathbb{E}(M_n)\, (= \ldots = \mathbb{E}(M_0))$, $\forall n \in \mathbb{N}$.

- $\mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = 0$ a.s.

- $\mathbb{E}(M_{n+m}|\mathcal{F}_n) = M_n$ a.s., $\forall n, m \in \mathbb{N}$.

This last property is important, as it says that the martingale property propagates over time.

Example: the simple symmetric random walk.

Let $(S_n,\, n \in \mathbb{N})$ be the simple symmetric random walk : $S_0 = 0$, $S_n = \xi_1 + \ldots + \xi_n$, where the ξ_n are i.i.d. and $\mathbb{P}(\{\xi_1 = +1\}) = \mathbb{P}(\{\xi_1 = -1\}) = 1/2$.

Let us define the following filtration: $\mathcal{F}_0 = \{\emptyset,\, \Omega\}$, $\mathcal{F}_n = \sigma(\{\xi_1,\, \ldots,\, \xi_n\})$, $n \geq 1$. Then $(S_n,\, n \in \mathbb{N})$ is a martingale with respect to $(\mathcal{F}_n, ,n \in \mathbb{N})$. Indeed:

(i) $\mathbb{E}(|S_n|) \leq \mathbb{E}(|\xi_1|) + \ldots + \mathbb{E}(|\xi_n|) = 1 + \ldots + 1 = n < \infty$, $\forall n \in \mathbb{N}$.

(ii) $S_n = \xi_1 + \ldots + \xi_n$ is a function of $(\xi_1, \ldots, \xi_n)$, i.e., is $\sigma(\xi_1, \ldots, \xi_n) = \mathcal{F}_n$-measurable.

(iii) We have

$$\begin{aligned} \mathbb{E}(S_{n+1}|\mathcal{F}_n) &= \mathbb{E}(S_n + \xi_{n+1}|\mathcal{F}_n) = \mathbb{E}(S_n|\mathcal{F}_n) + \mathbb{E}(\xi_{n+1}|\mathcal{F}_n) \\ &= S_n + \mathbb{E}(\xi_{n+1}) = S_n + 0 = S_n \quad \text{a.s.} \end{aligned}$$

The first equality on the second line follows from the fact that S_n is $\mathcal{F}_n$-measurable and that ξ_{n+1} is independent of $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$.

Generalization. If the random variables ξ_n are i.i.d. and such that $\mathbb{E}(|\xi_1|) < \infty$ and $\mathbb{E}(\xi_1) = 0$, then $(S_n,\, n \in \mathbb{N})$ is also a martingale (in particular, $\xi_1 \sim \mathcal{N}(0,1)$ works).

Definition 2.4. Let $(\mathcal{F}_n,\, n \in \mathbb{N})$ be a filtration. A process $(M_n,\, n \in \mathbb{N})$ is a *submartingale* (resp. a *supermartingale*) with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ if

(i) $\mathbb{E}(|M_n|) < \infty$, $\forall n \in \mathbb{N}$.

(ii) M_n is $\mathcal{F}_n$-measurable, $\forall n \in \mathbb{N}$.

(iii) $\mathbb{E}(M_{n+1}|\mathcal{F}_n) \geq M_n$ a.s., $\forall n \in \mathbb{N}$ (resp. $\mathbb{E}(M_{n+1}|\mathcal{F}_n) \leq M_n$ a.s., $\forall n \in \mathbb{N}$).

Remarks. - Not every process is either a sub- or a supermartingale!

- The appellations sub- and supermartingale are counter-intuitive. They are due to historical reasons.

- Condition (ii) is now necessary in itself, as (iii) does not imply it.

- If $(M_n,\, n \in \mathbb{N})$ is both a submartingale and a supermartingale, then it is a martingale.

Example: the simple asymmetric random walk.

- If $\mathbb{P}(\{\xi_1 = +1\}) = p = 1 - \mathbb{P}(\{\xi_1 = -1\})$ with $p \geq 1/2$, then $S_n = \xi_1 + \ldots + \xi_n$ is a submartingale.

- More generally, $S_n = \xi_1 + \ldots + \xi_n$ is a submartingale if $\mathbb{E}(\xi_1) \geq 0$.

Proposition 2.5. If $(M_n,\, n \in \mathbb{N})$ is a martingale with respect to a filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$ and $\varphi : \mathbb{R} \to \mathbb{R}$ is a Borel-measurable and convex function such that $\mathbb{E}(|\varphi(M_n)|) < \infty$, $\forall n \in \mathbb{N}$, then $(\varphi(M_n),\, n \in \mathbb{N})$ is a submartingale.

Proof. (i) $\mathbb{E}(|\varphi(M_n)|) < \infty$ by assumption.

(ii) $\varphi(M_n)$ is $\mathcal{F}_n$-measurable as M_n is (and φ is Borel-measurable).

(iii) $\mathbb{E}(\varphi(M_{n+1})|\mathcal{F}_n) \geq \varphi(\mathbb{E}(M_{n+1}|\mathcal{F}_n)) = \varphi(M_n)$ a.s.

In (iii), the first inequality follows from Jensen's inequality and the second follows from the fact that M is a martingale. □

Example. If $(M_n,\, n \in \mathbb{N})$ is a square-integrable martingale (i.e., $\mathbb{E}(M_n^2) < \infty$, $\forall n \in \mathbb{N}$), then the process $(M_n^2,\, n \in \mathbb{N})$ is a submartingale (as $x \mapsto x^2$ is convex).

Doob's Decomposition theorem.

Definition 2.6. A process $(A_n,\, n \in \mathbb{N})$ is said to be *predictable* with respect to a filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$ if $A_0 = 0$ and A_n is $\mathcal{F}_{n-1}$-measurable $\forall n \geq 1$.

Remark. If a process is predictable, then it is adapted.

Theorem 2.7. Let $(X_n,\, n \in \mathbb{N})$ be a submartingale with respect to a filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$. Then there exists a martingale $(M_n,\, n \in \mathbb{N})$ with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ and a process $(A_n,\, n \in \mathbb{N})$ predictable with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ and increasing (i.e, $A_n \leq A_{n+1}$ $\forall n \in \mathbb{N}$) such that $A_0 = 0$ and $X_n = M_n + A_n$, $\forall n \in \mathbb{N}$. Moreover, this decomposition of the process X is unique.

Proof. (main idea)
$\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$, so a natural candidate for the process A is to set $A_0 = 0$ and $A_{n+1} = A_n + \mathbb{E}(X_{n+1}|\mathcal{F}_n) - X_n \,(\geq A_n)$, which is a predictable and increasing process. Then, $M_0 = X_0$ and $M_{n+1} - M_n = X_{n+1} - X_n - (A_{n+1} - A_n) = X_{n+1} - \mathbb{E}(X_{n+1}|\mathcal{F}_n)$ is indeed a martingale, as $\mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = 0$. □

2.2 Stopping times

Definitions 2.8. - A *random time* is a random variable T with values in $\mathbb{N} \cup \{+\infty\}$.

- Given a process $(X_n,\, n \in \mathbb{N})$, one defines $X_T(\omega) = X_{T(\omega)}(\omega) = \sum_{n\in\mathbb{N}} X_n(\omega)\, 1_{\{T=n\}}(\omega)$.

- A *stopping time* with respect to a filtration $(\mathcal{F}_n, n \in \mathbb{N})$ is a random time T such that $\{T \leq n\} \in \mathcal{F}_n$, $\forall n \in \mathbb{N}$.

Proposition 2.9. T is a stopping time with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ if and only if $\{T = n\} \in \mathcal{F}_n$, $\forall n \in \mathbb{N}$.

Example. Let $(X_n,\, n \in \mathbb{N})$ be a process adapted to $(\mathcal{F}_n,\, n \in \mathbb{N})$ and $a > 0$. Then, $T_a = \inf\{n \in \mathbb{N} : |X_n| \geq a\}$ is a stopping time with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$. Indeed:

$$\begin{aligned} \{T_a = n\} &= \{|X_i| < a,\ \forall 0 \leq i \leq n-1 \text{ and } |X_n| \geq a\} \\ &= \bigcap_{i=0}^{n-1} \underbrace{\{|X_i| < a\}}_{\in \mathcal{F}_i \subset \mathcal{F}_n\ \forall i=0,\ldots,n-1} \quad \cap \quad \{|X_n| \geq a\} \in \mathcal{F}_n, \quad \forall n \in \mathbb{N}. \end{aligned}$$

Definition 2.10. Let T be a stopping time with respect to a filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$. One defines the information one possesses at time T as the following σ-field:

$$\mathcal{F}_T = \{A \in \mathcal{F} \,:\, A \cap \{T = n\} \in \mathcal{F}_n,\, \forall n \in \mathbb{N}\}.$$

Facts.

- If $T(\omega) = N$, $\forall \omega \in \Omega$, then $\mathcal{F}_T = \mathcal{F}_N$.

- If $T_1(\omega) \leq T_2(\omega)$, $\forall \omega \in \Omega$, then $\mathcal{F}_{T_1} \subset \mathcal{F}_{T_2}$.

[Here is an example of stopping times T_1, T_2 such that $T_1 \leq T_2$:
let $0 < a < b$ and consider $T_1 = \inf\{n \in \mathbb{N} \,:\, |X_n| \geq a\}$ and $T_2 = \inf\{n \in \mathbb{N} \,:\, |X_n| \geq b\}$.]

- A random variable Y is $\mathcal{F}_T$-measurable if and only if $Y\, 1_{\{T=n\}}$ is $\mathcal{F}_n$-measurable, $\forall n \in \mathbb{N}$.

As a consequence:

- If $(X_n,\, n \in \mathbb{N})$ is adapted to $(\mathcal{F}_n,\, n \in \mathbb{N})$, then X_T is $\mathcal{F}_T$-measurable.

Doob's optional sampling theorem.
Let $(M_n,\, n \in \mathbb{N})$ be a martingale with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ and T_1, T_2 be two stopping times such that $0 \leq T_1(\omega) \leq T_2(\omega) \leq N < \infty$, $\forall \omega \in \Omega$. Then

$$\mathbb{E}(M_{T_2} | \mathcal{F}_{T_1}) = M_{T_1} \text{ a.s.}$$

In particular, $\mathbb{E}(M_{T_2}) = \mathbb{E}(M_{T_1})$ (this consequence is referred to as the *optional stopping* theorem).

In particular, if T is a stopping time such that $0 \leq T(\omega) \leq N < \infty$, $\forall \omega \in \Omega$, then

$$\mathbb{E}(M_T) = \mathbb{E}(M_0).$$

Remarks. - The above theorem says that the martingale property holds even if one is given the option to stop at any (bounded) stopping time.

- The theorem also holds for sub- and supermartingales.

Proof. - We first show that if T is a stopping time such that $0 \leq T(\omega) \leq N$, then $\mathbb{E}(M_N | \mathcal{F}_T) = M_T$ $(*)$:

Indeed, let $Z = M_T = \sum_{n=0}^{N} M_n\, 1_{\{T=n\}}$. We check below that Z is the conditional expectation of M_N given $\mathcal{F}_T$:

(i) Z is $\mathcal{F}_T$-measurable: $Z\,1_{\{T=n\}} = M_n\,1_{\{T=n\}}$, so by the above mentioned fact, Z is $\mathcal{F}_T$-measurable.

(ii) $\mathbb{E}(ZU) = \mathbb{E}(M_N U)$, $\forall U$ $\mathcal{F}_T$-measurable and bounded:

$$\mathbb{E}(ZU) = \sum_{n=0}^{N} \mathbb{E}(M_n 1_{\{T=n\}} U) = \sum_{n=0}^{N} \mathbb{E}(\mathbb{E}(M_N|\mathcal{F}_n)\underbrace{1_{\{T=n\}}U}_{\mathcal{F}_n-\text{measurable}}) = \sum_{n=0}^{N} \mathbb{E}(M_N 1_{\{T=n\}} U) = \mathbb{E}(M_N U).$$

- Second, let us check that $\mathbb{E}(M_{T_2}|\mathcal{F}_{T_1}) = M_{T_1}$:

$$M_{T_1} \underset{(*)\text{ with } T=T_1}{=} \mathbb{E}(M_N|\mathcal{F}_{T_1}) \underset{\mathcal{F}_{T_1}\subset\mathcal{F}_{T_2}}{=} \mathbb{E}(\mathbb{E}(M_N|\mathcal{F}_{T_2})|\mathcal{F}_{T_1}) \underset{(*)\text{ with } T=T_2}{=} \mathbb{E}(M_{T_2}|\mathcal{F}_{T_1}).$$

This concludes the proof of the theorem. □

2.3 Martingale transforms

Let $(\mathcal{F}_n,\, n\in\mathbb{N})$ be a filtration, $(H_n,\, n\in\mathbb{N})$ be a predictable process with respect to $(\mathcal{F}_n,\, n\in\mathbb{N})$ and $(M_n,\, n\in\mathbb{N})$ be a martingale with respect to $(\mathcal{F}_n,\, n\in\mathbb{N})$.

Definition 2.11. The process G defined as

$$G_0 = 0, \quad G_n = (H\cdot M)_n = \sum_{i=1}^{n} H_i(M_i - M_{i-1}), \quad n\geq 1,$$

is called the *martingale transform* of M through H.

Remark. This process is the discrete version of the stochastic integral. It represents the gain obtained by applying the strategy H to the game M:

- H_i = amount bet on day i ($\mathcal{F}_{i-1}$-measurable)

- $M_i - M_{i-1}$ = increment of the process M on day i.

- G_n = gain on day n.

Proposition 2.12. If H_n is a bounded random variable for each n (i.e, $|H_n(\omega)|\leq K_n$ $\forall\omega\in\Omega$), then the process G is a martingale with respect to $(\mathcal{F}_n,\, n\in\mathbb{N})$.

In other words, one cannot win on a martingale!

Proof. (i) $\mathbb{E}(|G_n|) \leq \sum_{i=1}^n \mathbb{E}(|H_i|\,|M_i - M_{i-1}|) \leq \sum_{i=1}^n K_i\,(\mathbb{E}(|M_i|) + \mathbb{E}(|M_{i-1}|)) < \infty$.

(ii) G_n is $\mathcal{F}_n$-measurable by construction.

(iii) $\mathbb{E}(G_{n+1}|\mathcal{F}_n) = \mathbb{E}(G_n + H_{n+1}\,(M_{n+1}-M_n)|\mathcal{F}_n) = G_n + H_{n+1}\,\mathbb{E}(M_{n+1}-M_n|\mathcal{F}_n) = G_n + 0 = G_n$. □

Example: "the" martingale.

Let $(M_n,\, n\in\mathbb{N})$ be the simple symmetric random walk ($M_n = \xi_1 + \ldots + \xi_n$) and consider the following strategy:

$$H_0 = 0,\; H_1 = 1,\; H_{n+1} = \begin{cases} 2H_n, & \text{if } \xi_1 = \ldots = \xi_n = -1, \\ 0. & \text{otherwise.} \end{cases}$$

Notice that all the H_n are bounded random variables. Then by the above proposition, the process G defined as

$$G_0 = 0, \quad G_n = \sum_{i=1}^{n} H_i\,(M_i - M_{i-1}) = \sum_{i=1}^{n} H_i\,\xi_i, \quad n\geq 1,$$

is a martingale. So $\mathbb{E}(G_n) = \mathbb{E}(G_0) = 0$, $\forall n \in \mathbb{N}$. Let now

$$T = \inf\{n \geq 1 \,:\, \xi_n = +1\}.$$

T is a stopping time and it is easily seen that $G_T = +1$. But then $\mathbb{E}(G_T) = 1 \neq 0 = \mathbb{E}(G_0)$? Is there a contradiction? Actually no. The optional stopping theorem does not apply here, because the time T is unbounded: $\mathbb{P}(T = n) = 2^{-n}$, $\forall n \in \mathbb{N}$, i.e., there does not exist N fixed such that $T(\omega) \leq N$, $\forall \omega \in \Omega$.

2.4 Markov processes

Let $(X_n,\, n \in \mathbb{N})$ be a discrete-time process adapted to a filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$.

Definition 2.13. $(X_n,\, n \in \mathbb{N})$ is said to be a *Markov process* with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ if

$$\mathbb{P}(X_{n+1} \in B \,|\, \mathcal{F}_n) = \mathbb{P}(X_{n+1} \in B \,|\, X_n), \quad \forall n \in \mathbb{N},\, B \in \mathcal{B}(\mathbb{R}).$$

(remember that $\mathbb{P}(X_{n+1} \in B \,|\, X_n) = \mathbb{P}(X_{n+1} \in B \,|\, \sigma(X_n))$ by definition.)

Proposition 2.14. $(X_n,\, n \in \mathbb{N})$ is a Markov process with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ if and only if

$$\mathbb{E}(g(X_{n+1})|\mathcal{F}_n) = \mathbb{E}(g(X_{n+1})|X_n), \quad \forall n \in \mathbb{N}$$

and $\forall g : \mathbb{R} \to \mathbb{R}$ Borel-measurable and bounded.

Particular class of Markov process. Let $(\xi_n,\, n \geq 1)$ be a sequence of independent random variables (not necessarily i.i.d.) and

$$\mathcal{F}_0 = \{\emptyset,\, \Omega\}, \quad \mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n).$$

Then any process defined recursively as

$$X_0 = \text{cst}, \quad X_{n+1} = f(X_n, \xi_{n+1}), \quad n \geq 0,$$

with $f : \mathbb{R}^2 \to \mathbb{R}$ a Borel-measurable function, is a Markov process with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$.

Proof. As X_n is both and $\mathcal{F}_n$-measurable and $\sigma(X_n)$-measurable, and ξ_{n+1} is independent of $\mathcal{F}_n$, and therefore of X_n, we have by Proposition 1.28:

$$\mathbb{E}(g(X_{n+1})|\mathcal{F}_n) = \mathbb{E}(g(f(X_n, \xi_{n+1}))|\mathcal{F}_n) = \psi(X_n), \quad \text{where } \psi(y) = \mathbb{E}(g(f(y, \xi_{n+1})))$$

and also

$$\mathbb{E}(g(X_{n+1})|X_n) = \mathbb{E}(g(f(X_n, \xi_{n+1}))|X_n) = \psi(X_n).$$

□

Example. Let S be a "generalized" random walk with independent increments, i.e. $S_0 = 0$, $S_n = \xi_1 + \ldots + \xi_n$, $n \geq 1$, where the random variables ξ_i are independent. Then $S_{n+1} = S_n + \xi_{n+1}$ is indeed a Borel-measurable function of S_n and ξ_{n+1}; it falls therefore into the above-mentioned class of Markov processes.

Remarks.

- It is important not to mix martingale property ($\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$) with the Markov property ($\mathbb{E}(g(X_{n+1})|\mathcal{F}_n) = \mathbb{E}(g(X_{n+1})|X_n)$, $\forall g$). None of the two implies the other one.

- Markov chains form a particular class of Markov processes (those with discrete state space).

- Stationarity is not included in the above definition of a Markov process.