

# 1 Conditional Expection

We start with some preliminaries, things known or less known.

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

**Definition 1.1.** *We define the Lebesgue spaces*

$L^p(\Omega, \mathcal{F}, \mathbb{P}) = L^p = \{X : \Omega \rightarrow \mathbb{R} \mid X \text{ is } \mathcal{F} \text{ measurable and } \mathbb{E}(|X|^p) < \infty\}$   
where  $1 \leq p < \infty$ .

$L^\infty = \{X : \Omega \rightarrow \mathbb{R} \mid X \text{ is } \mathcal{F} \text{ measurable and } \exists C > 0 : \mathbb{P}(|X| \leq C) = 1\}$  is the space of almost surely bounded random variables.

**Theorem 1.1** (Hölder's inequality). *Let  $X \in L^p, Y \in L^q$  for conjugate  $p, q \geq 1$  (i.e.  $1/p + 1/q = 1$ ) then we have  $XY \in L^1$  moreover Hölder's inequality holds true*

$$\mathbb{E}(|XY|) \leq \mathbb{E}(|X|^p)^{\frac{1}{p}} \mathbb{E}(|X|^q)^{\frac{1}{q}}$$

**Remark 1.1.**

1. Hölder's inequality is the reason why probability theory behaves "nicer" in consideration of Lebesgue spaces, i.e. let  $0 < r < s$  and set  $p = \frac{s}{r}$  then for the conjugate  $q = \frac{p}{p-1}$  we can apply Hölder's inequality to  $|X|^r$  and the constant 1 function.

$$\mathbb{E}(|X|^r) \leq \mathbb{E}(|X|^{rp})^{1/p} = \mathbb{E}(|X|^s)^{r/s} \implies \mathbb{E}(|X|^r)^{\frac{1}{r}} \leq \mathbb{E}(|X|^s)^{\frac{1}{s}}$$

in particular if  $X$  is a random variable such that  $X \in L^s$  then it always follows that  $X \in L^r$  for any  $r < s$ . In general measure theory, where the spaces don't need to be of finite measure, this is not the case at all.

2. It is an important Theorem of Riesz that states that for all  $1 \leq p \leq \infty$  the  $L^p$  spaces are complete, in particular they are Banach spaces. Especially for  $p = 2 = q$  the space  $L^2$  is even a Hilbert space.

**Proposition 1.1** (Jensen's inequality). *Let  $X \in L^1$  and  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  a convex function, then we have*

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X))$$

**Proposition 1.2** (Existence of the conditional expectation). *Let  $\mathcal{G} \subset \mathcal{F}$  be a subsigma-algebra. We define  $Z = \mathbb{E}(X \mid \mathcal{G})$  to be the unique  $\mathcal{G}$ -measurable variable such that*

$$\mathbb{E}(Z1_A) = \mathbb{E}(X1_A) \text{ for all } A \in \mathcal{G}$$

## 2 Discrete time Martingales

**Definition 2.1.** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, an increasing sequence of  $\sigma$ -algebras  $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}$  is called a filtration on  $\Omega$ . We call the space  $(\Omega, \mathcal{F}, (\mathcal{F})_{n \in \mathbb{N}}, \mathbb{P})$  a filtered probability space.

**Definition 2.2.** A stochastic process  $X = (X_n)_{n \in \mathbb{N}}$  is called adapted if  $X_n$  is  $\mathcal{F}_n$ -measurable for all  $n \in \mathbb{N}$ . Moreover, an adapted process is called a martingale if  $X_n \in L^1$  for all  $n \in \mathbb{N}$  and

$$\mathbb{E}(X_n \mid \mathcal{F}_m) = X_{n \wedge m} \text{ for all } n, m \geq 0$$

**Remark 2.1.** Analogously we define sub-martingales if instead of an equality we have  $\geq$  above, or a super-martingale for  $\leq$  respectively.

**Definition 2.3.** A stopping time  $T$  is a random variable  $T : \Omega \rightarrow \mathbb{N}_0^\infty$  such that  $\{T \leq n\} \in \mathcal{F}_n$  for all  $n \in \mathbb{N}$ . The sigma-algebra at the stopping time  $T$  is then defined as

$$\mathcal{F}_T := \{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}\}$$

**Remark 2.2.** The sigma algebra at the stopping time  $T$   $\mathcal{F}_T$  encodes the information up to the random time  $T$ . In other words, if we interpret our filtered probability space as an random experiment, then the maximum information that can be found until the random time  $T$  is in  $\mathcal{F}_T$ .

**Theorem 2.1** (Optional stopping theorem - weak version). Let  $X$  be a martingale. Let  $X^T := (X_n^T)_{n \in \mathbb{N}}$  be given by  $X_n^T := X_{n \wedge T}$  then  $X^T$  is also a martingale. Moreover if  $S, T$  are almost surely finite stopping times with  $S \leq T$  almost surely, then

$$\mathbb{E}(X_T \mid \mathcal{F}_S) = X_S$$

**Theorem 2.2** (Martingale convergence Theorem). Let  $X$  be a super-martingale that is bounded in  $L^1$  (i.e.  $\sup_{n \in \mathbb{N}} \mathbb{E}(|X_n|) < \infty$ ). Then there exists a random variable  $X_\infty \in L^1$  such that  $X_n \rightarrow X_\infty$  almost surely, moreover

$$\mathbb{E}(|X_\infty|) \leq \sup_{n \in \mathbb{N}} \mathbb{E}(|X_n|)$$

**Corollary 2.1.** Let  $X$  be a super-martingale that is bounded from below (i.e.  $X_n \geq c$  a.s. for some  $c \in \mathbb{R}$  for all  $n \in \mathbb{N}$ ). Then  $X_n$  converges almost surely to  $X_\infty \in L^1$ .

**Theorem 2.3** (Behaviour of Martingales with bounded increments). Let  $X = (X_n)_{n \in \mathbb{N}}$  be a martingale with bounded increments (i.e.  $|X_{n+1} - X_n| \leq C$ ). Let  $\mathcal{C} := \{\limsup X_n = \liminf X_n \in \mathbb{R}\}$ ,  $\mathcal{O} := \{\limsup X_n = +\infty\} \cap \{\liminf X_n = -\infty\}$ , then  $\mathbb{P}(\mathcal{O} \cup \mathcal{C}) = 1$ .

**Lemma 2.1** (Doob's inequalities). *Let  $X$  be a non-negative submartingale and  $X_n^* := \max_{k \leq n} X_k$  for  $n \in \mathbb{N}$ , then we have for any  $\lambda > 0$*

$$\lambda \mathbb{P}(X_n^* > \lambda) \leq \mathbb{E}(X_n 1_{X_n^* > \lambda}) \leq \mathbb{E}(X_n)$$

*In addition, for  $p > 1$ , we have*

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

**Theorem 2.4** (Closed martingale convergence theorem). *Let  $X$  be a martingale and  $p > 1$ . Then the following statements are equivalent:*

1.  $\sup_{n \in \mathbb{N}} \|X_n\|_p < \infty$
2.  $X_n$  converges almost surely and in  $L^p$  to  $X_\infty \in L^p$
3. There exists a random variable  $X_\infty \in L^p$  such that

$$\mathbb{E}(X_\infty | \mathcal{F}_n) = X_n \text{ for all } n \in \mathbb{N}$$

*If any (and consequently all) of these conditions hold true, we say that  $X$  is a closed martingale in  $L^p$ .*

**Definition 2.4.** *A family  $\mathcal{H}$  of random variables is called uniformly integrable (UI) if*

$$\lim_{\lambda \rightarrow \infty} \sup_{X \in \mathcal{H}} \mathbb{E}(|X| 1_{|X| > \lambda}) = 0$$

**Exercise:** Prove that  $\mathcal{H}$  is UI if there exists  $G : [0, \infty) \rightarrow [0, \infty)$  non-decreasing such that

1.  $\lim_{t \rightarrow \infty} \frac{G(t)}{t} = \infty$
2.  $\sup_{X \in \mathcal{H}} \mathbb{E}(G(X)) < \infty$

**Theorem 2.5** (UI convergence Theorem). *Let  $(X_n)_{n \in \mathbb{N}}$  a stochastic process, then the following statements are equivalent:*

1.  $(X_n)_{n \in \mathbb{N}}$  is UI and  $X_n \rightarrow X_\infty$  in probability.
2.  $X_n \rightarrow X_\infty$  in  $L^1$ .

**Theorem 2.6** (Characterisation of UI martingales through closedness). *Let  $X$  be a martingale. The following assertions are equivalent:*

1.  $X$  is UI.
2.  $X$  converges almost surely and in  $L^1$  to  $X_\infty \in L^1$ .
3. There exists  $X_\infty \in L^1$  such that  $\mathbb{E}(X_\infty | \mathcal{F}_n) = X_n$ , i.e. the martingale is closed.

**Theorem 2.7** (Optional stopping - strong version). *Let  $X$  be a closed (and thus UI) martingale. Then for any stopping times  $S$  and  $T$ , one has*

$$\mathbb{E}(X_T | \mathcal{F}_S) = X_{T \wedge S}$$

**Definition 2.5.** Let  $\mathcal{F}$  be a sigma-algebra and consider a sequence  $\cdots \subset \mathcal{H}_2 \subset \mathcal{H}_1 \subset \mathcal{F}$  of sigma-algebras (i.e. we lose information over time). Let  $Y_1 \in L^p(\Omega, \mathcal{H}_1, \mathbb{P})$  for some  $p \in [1, \infty)$ . Then  $Y = (Y_n)_{n \in \mathbb{N}}$  is called a backward-martingale if  $Y_n = \mathbb{E}(Y_1 \mid \mathcal{H}_n)$  for all  $n \in \mathbb{N}$ .

**Theorem 2.8** (Convergence theorem for backward-martingales). Let  $Y = (Y_n)_{n \in \mathbb{N}}$  be a backward-martingale. Then  $Y$  converges almost surely and in  $L^p$  to  $Y_\infty = \mathbb{E}(Y_1 \mid \mathcal{H}_\infty)$  where  $\mathcal{H}_\infty = \bigcap_{n \in \mathbb{N}} \mathcal{H}_n$ .

**Theorem 2.9** (Kolmogorov's Law of Large Numbers). Let  $(\xi_n)_{n \in \mathbb{N}}$  be a sequence of i.i.d. random variables with  $\xi_1 \in L^1$ . Let  $S_n = \xi_1 + \cdots + \xi_n$  be the simple random walk, then we have

$$\frac{S_n}{n} \rightarrow \mathbb{E}(X_1) \text{ almost surely and in } L^1$$

*Proof. Exercise.* Kolmogorov's LLN can be shown in it's full generality (i.e. for  $L^1$  RV) by using backwards martingales and Kolmogorov's 0-1 Law.  $\square$

**Theorem 2.10** (Kolmogorov's 0-1 Law). Let  $(\xi_n)_{n \in \mathbb{N}}$  be a sequence of i.i.d. random variables and let  $\mathcal{H}_n = \sigma(\xi_j : j \geq n)$ . Then, the sigma algebra  $\mathcal{H}_\infty := \bigcup_{n \in \mathbb{N}} \mathcal{H}_n$  is trivial, i.e. for every event  $A \in \mathcal{H}_\infty$  we have  $\mathbb{P}(A) \in \{0, 1\}$ .

Next we will remind the very important Borel-Cantelli Lemmas. We recall the following definition. Let  $(A_n)_{n \in \mathbb{N}}$  be a sequence of events in a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  then we define

$$A = \limsup_{n \rightarrow \infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{i \geq n}^{\infty} A_i = \{A_n \text{ infinitely often}\}$$

$$A^c := \liminf_{n \rightarrow \infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{i \geq n}^{\infty} A_i = \{A_n \text{ eventually}\}$$

We remark that the following is true

$$\omega \in \bigcap_{n=1}^{\infty} \bigcup_{i \geq n}^{\infty} A_i \iff \forall n \in \mathbb{N}, \exists i \geq n : \omega \in A_i$$

$$\omega \in \bigcup_{n=1}^{\infty} \bigcap_{i \geq n}^{\infty} A_i^c \iff \exists n \in \mathbb{N}, \forall i \geq n : \omega \in A_i^c$$

**Theorem 2.11** (Borel-Cantelli Lemma 1). If  $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < \infty$ , then we have  $\mathbb{P}(A_n \text{ i.o.}) = 0$ .

**Theorem 2.12** (Borel-Cantelli Lemma 2). If the events  $A_n$  are independent and  $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = \infty$ , then we have  $\mathbb{P}(A_n \text{ i.o.}) = 1$

**Theorem 2.13** (Extension of Borel Cantelli Lemma). *Let  $\mathcal{F} = (\mathcal{F}_n)_{n=1}^\infty$  be a Filtration and suppose  $A_n \in \mathcal{F}_n$ , then we have*

$$\{A_n \text{ i.o.}\} = \left\{ \sum_{n \geq k} \mathbb{P}(A_n \mid \mathcal{F}_{n-1}) = \infty \right\}$$

for all  $k \in \mathbb{N}$  up to a set of measure zero.

We now give some results that are related to the Central Limit Theorem (CLT). We will first recall the classical result.

**Theorem 2.14** (CLT). *Let  $X_1, X_2, \dots$  be a sequence of i.i.d.  $L^2$  random variables with  $\mathbb{E}(X) = \mu$  and  $\text{Var}(X) = \sigma > 0$ . Let  $S_n = X_1 + \dots + X_n$ , then we have*

$$\frac{S_n - n\mu}{n\sigma} \implies \mathcal{N}(0, 1)$$

It is our goal to generalize this result, i.e. weaken the conditions and then give a central limit theorem for martingales.

The outline is the following: Let  $\{X_{n,k} : 1 \leq k \leq J(n), n \in \mathbb{N}_{\geq 1}\}$  be an array of centered random variables. Here  $K$  can be random, however it is almost surely finite with  $J(n) \rightarrow \infty$  as  $n \rightarrow \infty$ .

Let  $(\xi_j)_{j \in \mathbb{N}}$  be a sequence of (centered?) i.i.d. random variables, we then set

$$X_{n,k} = \frac{\xi_k}{\sqrt{n}}$$

Suppose that  $X_{n,k} \in L^2(\mathcal{F})$  and let

$$S_n := \sum_{k=1}^{J(n)} X_{n,k}$$

**Theorem 2.15** (Mc Leish). *In the situation as above, let us define*

$T_n = \prod_{k=1}^{J(n)} (1 + itX_{n,k})$  *and suppose that*

1.  $T_n$  is UI and  $\mathbb{E}(T_n) \rightarrow 1$  as  $n \rightarrow \infty$ .
2.  $\sum_{k=1}^{J(n)} X_{n,k}^2 \rightarrow 1$  in Probability.
3.  $\max_{k \leq n} |X_{n,k}| \rightarrow 0$  in Probability.

*Then  $S_n \implies \mathcal{N}(0, 1)$  as  $n \rightarrow \infty$ .*

**Theorem 2.16** (Martingale Central Limit Theorem). *Let  $X_{n,k}$  where  $1 \leq k \leq m_n$  ( $m_n$  is deterministic with  $m_n \rightarrow \infty$  as  $n \rightarrow \infty$ ) be a  $\mathcal{F}_{n,k}$ -measurable martingale difference array. Suppose that*

1.  $\mathbb{E}(X_{n,k} \mid \mathcal{F}_{n,k-1}) = 0$
2.  $V_n := \sum_{j=1}^{m_n} X_{n,j}^2 \rightarrow 1$  in Probability.
3. (Lindenberg Condition) For any  $\epsilon > 0$  we have that

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E}(X_{n,k}^2 1_{|X_{n,k}| > \epsilon}) = 0$$

Then we have as  $n \rightarrow \infty$

$$S_n = \sum_{k=1}^{m_n} X_{n,k} \implies \mathcal{N}(0, 1)$$

### 3 Stochastic processes

Let  $(H_i)_{i \in I}$  be a sequence of complete, separable metric spaces, where  $I$  is an arbitrary index set.

**Definition 3.1** (Product space). *We define the product topology for  $H = \prod_{i \in I} H_i$  to be the coarsest (that is the weakest/smallest) topology such that all the maps  $\pi_i : H \rightarrow H_i$  are continuous.*

**Remark 3.1.**

1. This construction is equivalent to the construction of the weak topology.
2. This guarantees that the Borel  $\sigma$ -Algebra  $\mathcal{B}(H)$  coincides with the product  $\sigma$ -algebra

$$\sigma \left( \left\{ \prod_{i \in I} A_i : A_i \in \mathcal{B}(H_i) \text{ for } i \in K, A_i = H_i \text{ for } i \notin K, K \subset I \text{ finite} \right\} \right)$$

that is, the product  $\sigma$ -algebra on  $\mathbb{R}^{\mathbb{N}}$  is generated by

$$\bigcup_{n=1}^{\infty} \{A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R} \times \mathbb{R} \times \cdots \mid A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})\}$$

**Definition 3.2** (Product measure). *Let  $\mu_i$  be a probability measure on  $H_i$ , we define the product measure  $\mu = \bigotimes_{i \in I} \mu_i$  on  $H$  to be the unique measure such that*

$$\mu \left( \prod_{i \in I} A_i \right) = \int 1_{\prod_{i \in I} A_i} d\mu = \prod_{i \in I} \mu_i(A_i)$$

*This implies that*

$$\int \prod_{j \in K} f_j(\pi_j(\omega)) \mu(d\omega) = \prod_{j \in K} \int f_j d\mu_j, \text{ for all } f_j \in L^1(H_j, \mathcal{B}(H_j), \mu_j)$$

*under the assumption that  $A_i \in \mathcal{B}(H_i)$  for  $i \in K$  and  $A_i = H_i$  for  $i \notin K$ .*

It is by no means clear that such a unique product measure exists at all, this is however part of the celebrated Kolmogorov Extension theorem which we will state next.

**Theorem 3.1** (Kolmogorov's Extension Theorem). *Let  $K \subset I$  be a finite subset and let  $\mu_K$  be a probability measure on  $\prod_{i \in K} H_i$ . Suppose that for any  $K \subset J$  where  $J$  is another finite subset of  $I$  the following consistency condition holds:*

$$\mu_J \left( \prod_{i \in J} A_i \right) = \mu_K \left( \prod_{i \in K} B_i \right), \text{ with } A_i = \begin{cases} B_i \in \mathcal{B}(H_i), & \text{if } i \in K \\ H_i, & \text{if } i \notin K \end{cases}$$

*Then there exists a unique measure on  $H = \prod_{i \in I} H_i$  such that for all finite subsets  $K \subset I$  one has*

$$\mu \left( \prod_{i \in I} A_i \right) = \mu_K \left( \prod_{i \in K} B_i \right), \text{ with } A_i = \begin{cases} B_i \in \mathcal{B}(H_i), & \text{if } i \in K \\ H_i, & \text{if } i \notin K \end{cases}$$

**Remark 3.2.**

1. Here  $(\mu_K)_{K \subset I, K \text{ finite}}$  is called the collection of finite dimensional marginal distributions of  $\mu$ .
2. Kolmogorov extension theorem says that, given the finite dimensional marginales  $(\mu_K)$  as above, if they satisfy the consistency condition as requested in the theorem, then there exists a unique extension  $\mu$  to the whole product space  $H$  with marginales  $(\mu_K)$
3. The Theorem is also known as Kolmogorov's existence theorem or Kolmogorov's consistency theorem. It guarantees that a suitably "consistent" (here consistency is meant as in the theorem) collection of finite-dimensional distributions will define a stochastic process.

**Definition 3.3.** *A stochastic process  $X = (X_i)_{i \in I}$ , where  $I$  is an arbitrary index set, is a collection of random variables defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , i.e.  $X_i : \Omega \rightarrow H_i$  is  $\mathcal{F}$ -measurable.*

*The law of  $X$  is a probability measure  $\mu$  on  $H = \prod_{i \in I} H_i$  such that for all finite  $K \subset I$ , the marginal law  $\mu_K$  of  $\mu$  is given by*

$$\mu_K \left( \prod_{j \in K} A_j \right) = \mathbb{P}(X_j \in A_j, j \in K) \text{ for all } A_j \in \mathcal{B}(H_j)$$

**Remark 3.3.** We can always choose  $(\Omega, \mathcal{F}, \mathbb{P}) = (H = \prod_{i \in I} H_i, \mathcal{B}(H), \mu)$ , then we have  $X_i = \pi_i$  where  $\pi_i$  are the coordinate maps, i.e.

$$\pi_i : \begin{cases} H & \longrightarrow H_i \\ \omega & \longmapsto \omega_i \end{cases}$$

This is sometimes called the *canonical process*.

There is a section about the Galton-Watson process, it is an important exercise and will be left open on this summary.



## 4 Markov Processes

**Definition 4.1.** A stochastic process  $(X_t)_{t \in I}$  is a Markov if for all  $f : \mathcal{H} \rightarrow \mathbb{R}_+$  measurable ( $\mathcal{H}$  is a metric space) and for all  $s < t$  we have

$$\mathbb{E}(f(X_t) \mid \mathcal{F}_s^X) = \mathbb{E}(f(X_t) \mid \sigma(X_s)), \text{ where } \mathcal{F}_s^X := \sigma(X_t : t \leq s)$$

**Definition 4.2.** A process  $(X_n)_{n \in \mathbb{N}}$  is a time-homogenous Markov chain if there exists a measurable map  $Q : \mathcal{H} \rightarrow \mathcal{M}(\mathcal{H}) := \{\text{probability measures on } \mathcal{H}\}$  such that  $\mathbb{P}(X_n \in A \mid X_{n-1} = x) = Q(x, A)$  for all Borel sets  $A \subset \mathcal{H}$ . We call  $Q$  the transition probability of the Markov chain.  $\mathcal{H}$  is called the state space of the Chain, moreover  $x \in \mathcal{H}$  is called a state of the chain.

**Proposition 4.1.**  $\mathcal{M}(\mathcal{H})$  equipped with the topology of weak convergence is a completely metrizable, separable space.

Recall that a sequence  $\mu_n \in \mathcal{M}(\mathcal{H})$  converges weakly to  $\mu \in \mathcal{M}(\mathcal{H})$  if and only if

$$\int f d\mu_n \rightarrow \int f d\mu, \text{ for all } f \in C_b(\mathcal{H} \rightarrow \mathbb{R})$$

Given  $Q$  as above, then there exists a Markov chain with  $\mathbb{P}_\nu(X_0 \in A) = \nu(A)$  and transition probability  $Q$ . We define for all  $n \in \mathbb{N}$

$$\mathbb{P}_\nu(x_0 \in A_0, x_1 \in A_1, \dots, x_n \in A_n) = \int_{A_0 \times \dots \times A_n} \nu(dx_0) Q(x_0, dx_1) \dots Q(dx_{n-1}, dx_n)$$

By Kolmogorov's extension theorem, there exists a unique probability measure  $\mathbb{P}_\nu$  on  $(\Omega, \mathcal{F}_\infty)$ , where  $\mathcal{F}_\infty = \bigvee_{n=1}^\infty \mathcal{F}_n$  is the product  $\sigma$ -algebra on  $\Omega$ , with marginal distribution as above.

**Definition 4.3.** We define the shift operator acting on  $\Omega$  by  $\Theta : \Omega \rightarrow \Omega$ ,  $\Theta(\omega) = (\omega_1, \omega_2, \omega_3, \dots)$  where  $\omega = (\omega_0, \omega_1, \dots)$ . We then have  $\Theta^n(\omega) = (\omega_{n+j})_{j=0}^\infty$  for all  $n \in \mathbb{N}$ .

**Theorem 4.1** (Markov Property). Under  $\mathbb{P}_\nu$ , conditionally on the canonical filtration  $\mathcal{F}_n = \sigma(\omega_0, \omega_1, \dots, \omega_n)$ ,  $\Theta^n(\omega)$  has law given by  $\mathbb{P}_{\omega_n}$  and is independent of  $\mathcal{F}_{n-1}$  for all  $n \in \mathbb{N}_{\geq 1}$ .

**Theorem 4.2** (Strong Markov Property). Let  $T$  be a stopping time such that  $\mathbb{P}_\nu(T < \infty) > 0$ . Conditionally on the event  $\{T < \infty\}$  and  $\mathcal{F}_T$  we have that  $\Theta^n$  is a Markov Chain with law given by  $\mathbb{P}_{X_T}$  and independent of  $\mathcal{F}_{T-1}$

We now work (hopefully) with  $\mathcal{H}$  a countable (or finite) state space, the transition probability  $Q$  is then represented as a matrix, i.e. we have

$$Q(x, A) = \sum_{y \in A} Q_{xy}, \text{ for all } x \in \mathcal{H}$$

A probability measure on  $\mathcal{H}$  is a vector such that we have

$$\nu(A) = \sum_{y \in A} \nu_y$$

We have

$$\mathcal{M}(\mathcal{H}) = \{\mu \in \mathbb{R}^{\mathcal{H}} : \mu_x \geq 0 \text{ for all } x \in \mathcal{H} \text{ and } \sum_{x \in \mathcal{H}} \mu_x = 1\}$$

**Definition 4.4.** We define for all  $x \in \mathcal{H}$

$R_x := \inf\{n \geq 1 : \omega_n = x\}$ , i.e. the first return to the state  $x$ .

$N_x := \#\{n \geq 0 : \omega_n = x\}$ , i.e. the number of visits of the state  $x$ .

A state  $x \in \mathcal{H}$  is called recurrent if  $\mathbb{P}_x(R_x < \infty) = 1$ , else  $x$  is called transient.

**Proposition 4.2.** If  $x$  is recurrent, then  $\mathbb{P}(N_x = \infty) = 1$ .  
If  $x$  is transient, then we have

$$\mathbb{E}_x(N_x) = \frac{1}{\mathbb{P}_x(R_x = \infty)} < \infty$$

**Definition 4.5.** The Green Kernel of the Markov Chain is defined as

$$G_{xy} := \mathbb{E}_x(N_y) = \sum_{n \in \mathbb{N}} Q_{xy}^n = \sum_{n \in \mathbb{N}} \mathbb{P}_x(\omega_n = y)$$

i.e. the expected number of visits of  $y$  starting from the state  $x$ .  
(Recall that  $Q_{xy}^n = \mathbb{P}_x(\omega_n = y)$ .)

**Proposition 4.3.** Let  $x, y \in \mathcal{H}$ ,  $x \neq y$ , then the following holds

1.  $G_{xx} = \infty$  if and only if  $x$  is recurrent.
2.  $G(x, y) = 0$  if and only if  $\mathbb{P}_x(R_y < \infty) = 0$ .  
(i.e. impossible to go from  $x$  to  $y$ .)
3. If  $\mathbb{P}_x(R_y < \infty) > 0$ , then  $G(x, y) = G(y, y)\mathbb{P}_x(R_y < \infty)$ .

*Proof.* See Exercises! □

**Corollary 4.1.** *Let  $R := \{x \in \mathcal{H} : \mathbb{P}_x(R_x < \infty) = 1\}$  the set of recurrent states. We have that if  $x \in R$  and  $G(x, y) > 0$ , then  $y \in R$  and also  $G(y, x) > 0$ ,  $\mathbb{P}_y(R_x < \infty) = 1$ .*

**Remark 4.1.** The above corollary yields that the being recurrent gives equivalent classes on  $R$ . We have  $R = \cup_{j \in I} R_j$  and the sets  $R_j$  are disjoint, the  $R_j$ 's are called the recurrence classes. We define the equivalence relation  $x \sim y$  if  $G(x, y) > 0$ . It follows that  $x \in R_j$ , then  $G(x, y) > 0$  if and only if  $y \in R_j$ .

**Theorem 4.3** (Classification of states). *The following holds true*

1. *If  $x \in R_j$  for some  $j \in I$ , then*

$$\begin{cases} \mathbb{P}_x(N_y = \infty) = 1, & \text{for all } y \in R_j \\ \mathbb{P}_x(N_y = 0) = 1, & \text{for all } y \notin R_j \end{cases}$$

2. *Let  $T = \inf\{n \geq 0 : \omega_n \in R\}$  be the first time, then some weird statement that should be reformulated.*

**Definition 4.6.** *The chain is called irreducible if  $G_{xy} > 0$  for all  $x, y \in \mathcal{H}$ . Irreducibility means that either we have  $R = \mathcal{H}$  or  $R = \emptyset$ .*

**Corollary 4.2.** *If the chain is irreducible, one of the following applies*

1. *All states are recurrent, in particular  $\mathbb{P}_x(N_y = \infty) = 1$  for all  $x \in \mathcal{H}$ .*
2. *All states are transient, in particular  $\mathbb{P}_x(N_y < \infty) = 1$  for all  $x \in \mathcal{H}$ .*

## 4.1 Stationary measures

We consider a Markov Chain with transition probability  $Q$ .

**Definition 4.7.** Let  $\mu$  be a non-trivial measure on  $\mathcal{H}$  such that  $\mu_x < \infty$  for all  $x \in \mathcal{H}$ . We say that  $\mu$  is stationary if  $\mu = \mu Q$ . That is

$$\mu(x) = \sum_{y \in \mathcal{H}} \mu(y) Q(y, x), \text{ for all } x \in \mathcal{H}$$

**Definition 4.8.** We say that  $\mu$  is reversible if it satisfies the detailed balance, i.e.

$$\mu_x Q_{xy} = \mu_y Q_{yx}, \text{ for all } x, y \in \mathcal{H}$$

**Exercise:** Check that if  $\mu$  is reversible, then it is also stationary.

**Lemma 4.1.** Let  $z \in R$  and define for all  $x \in \mathcal{H}$

$$\hat{\pi}(x) := \sum_{n=0}^{\infty} \mathbb{P}_z(\omega_n = x, R_z > n)$$

Then  $\hat{\pi}$  is a stationary measure and we have  $\hat{\pi}(x) > 0$  if and only if  $x \sim z$  i.e.  $G_{zx} > 0$ .

**Proposition 4.4.** If the chain is irreducible the solution  $\mu = \mu Q$  is unique up to a multiplicative constant.

*Proof.* Exercise! □

**Theorem 4.4.** Assume that the chain is irreducible and recurrent. Then we have the following dichotomy.

1. The chain is positive recurrent, i.e. there exists a stationary measure with finite mass. Moreover we then have, the probability measure  $\pi_x = 1/\mathbb{E}_x(R_x)$  is stationary,  $\pi$  is called the equilibrium measure.
2. The chain is null recurrent, i.e. all stationary measures have infinite mass. Moreover we then have,  $\mathbb{E}_x(R_x) = \infty$  for all  $x \in \mathcal{H}$ .

**Proposition 4.5.** Assume the chain is irreducible ( $G_{xy} > 0$ ) and there exists a equilibrium measure  $\pi$ . Then the chain is positive recurrent, in particular  $\pi$  is unique.

**Theorem 4.5** (Ergodic Theorem). *Let  $f, g : \mathcal{H} \rightarrow \mathbb{R}_+$ . and  $z \in R$ . Assume that  $\int f d\hat{\pi} < \infty$  and  $g(x) > 0$  for some  $x \sim z$ . Then we have  $\mathbb{P}_z$ -almost surely*

$$\frac{\sum_{j=0}^n f(\omega_j)}{\sum_{j=0}^n g(\omega_j)} \rightarrow \frac{\int f d\hat{\pi}}{\int g d\hat{\pi}} \text{ as } n \rightarrow \infty$$

**Corollary 4.3.** *The following holds*

1. *If the chain is positive recurrent, then for  $\mu \in \mathcal{M}(\mathcal{H})$  one has  $\mathbb{P}_\mu$  almost surely that*

$$\frac{1}{n} \sum_{j=0}^n f(\omega_j) \rightarrow \int f d\pi, \text{ as } n \rightarrow \infty$$

*where  $\pi$  is the equilibrium measure.*

2. *If the chain is null-recurrent, then for any  $\mu \in \mathcal{M}(\mathcal{H})$  one has  $\mathbb{P}_\mu$  almost surely that*

$$\frac{1}{n} \sum_{j=0}^n 1_{\omega_j=x} \rightarrow 0, \text{ as } n \rightarrow \infty \text{ for all } x \in \mathcal{H}.$$

**Definition 4.9.** *Let  $x \in R$  be a recurrent state and consider the set  $P_x := \{n \in \mathbb{N}_{\geq 1} : Q_{x,x}^n > 0\}$ . This set is then infinite (because  $x$  is recurrent) and the period of  $x$  is defined to be the greatest common divisor of the set  $P_x$ . That is, the period of  $x$  is the number  $d = \gcd P_x$ . A state  $x$  is called aperiodic if  $d = 1$ , else it is called periodic with period  $d$ . A Markov chain is called aperiodic if every state is aperiodic.*

**Proposition 4.6.** *A chain is aperiodic if and only if there exists a positive integer  $k > 0$  such that  $Q_{xy}^k > 0$  for all  $x, y \in \mathcal{H}$ .*

**Proposition 4.7.** *Let  $X$  be an irreducible markov chain, then all states  $x$  have the same period  $d$ . In particular for an irreducible chain it is enough to find one recurrent state to conclude that the chain is aperiodic.*

**Theorem 4.6.** *Suppose the chain is positive recurrent and aperiodic, then*

$$\sum_{x \in \mathcal{H}} |\mathbb{P}_z(\omega_n = x) - \pi(x)| \rightarrow 0, \text{ as } n \rightarrow \infty$$

*i.e. the law of  $\omega_n$  under  $\mathbb{P}_z$  converges in total variation to  $\pi$ .*

**Exercise:** Given a transition probability  $Q$ . Show that  $\tilde{Q} = (I + Q)/2$  defines an irreducible aperiodic chain, it is called the lazy chain.

**Theorem 4.7.** *Let  $Z = (Z_k)_{k \in \mathbb{N}}$  identically distributed random variables such that*

1.  $\mathbb{E}(Z_k^2) = 1$
2.  $\mathbb{E}(Z_k \mid \mathcal{F}_{k-1}) = 0$  where  $\mathcal{F}_k = \sigma(Z_1, \dots, Z_k)$
3.  $Z$  is erogidic, in this regime this means

$$\frac{1}{n} \sum_{j=1}^n f(Z_j) \rightarrow \mathbb{E}(f(Z)), \text{ almost surely, for all integrable } f.$$

*Then we have that*

$$\frac{Z_1 + \dots + Z_n}{\sqrt{n}} \implies \mathcal{N}(0, 1), \text{ as } n \rightarrow \infty$$

*Proof.* Follows by McLeish CLT. □

**Corollary 4.4.** *Let  $\mathcal{H}$  be a finite state space and  $Q$  be an irreducible transition Matrix with equilibrium measure  $\pi$ . If  $\int f d\pi = 0$  and  $\sigma = \int f^2 d\pi < \infty$  then we have that*

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n f(\omega_j) \implies \mathcal{N}(0, \sigma^2), \text{ as } n \rightarrow \infty$$

## 5 Ergodicity

In this section we consider  $H$  to be complete, separable metric space with Borel  $\sigma$ -algebra  $\mathcal{F}$ . For all  $t \in \mathbb{R}$  let  $T_t : H \rightarrow H$  be a family of bijective measurable transformations such that they satisfy the semigroup property, i.e.

$$\begin{cases} T_t \circ T_s = T_{t+s} \\ T_0(x) = x \end{cases}$$

**Definition 5.1.** Let  $\mu$  be a measure on  $H$ . We say that  $\mu$  is an invariant measure if  $\mu(T_t^{-1}(A)) = \mu(A)$  for all  $A \in \mathcal{F}$  and for all  $t \in \mathbb{R}$ .  $(H, \mathcal{F}, \mu, T_t)$  is called a dynamical system.

**Definition 5.2.**  $\mu$  is called ergodic if for every  $A \in \mathcal{F}$  with  $T_t^{-1}(A) = A$  for all  $t \in \mathbb{R}$  we have  $\mu(A) \in \{0, 1\}$ .

We consider the functional expression

$$R_t(f) = \frac{1}{t} \int_0^t f(T_s x) ds, \text{ for } f : H \rightarrow \mathbb{R} \text{ measurable}$$

and we wonder under what conditions we have that  $R_t(f) \rightarrow \int f d\mu$  as  $t \rightarrow \infty$  and for which  $\mu$ ? We start by discussing some functional analytic properties.

**Definition 5.3.** A unitary operator is a bounded linear operator  $U : H \rightarrow H$  on a Hilbert space  $H$  that satisfies the following:

1.  $U$  is surjective, and
2.  $U$  preserves the inner product of the Hilbert space  $H$ . That is we have

$$\langle Ux, Uy \rangle_H = \langle x, y \rangle_H, \text{ for all } x, y \in H.$$

**Lemma 5.1.** Let  $U_s : L^2(H, \mathcal{F}, \mu) \rightarrow L^2(H, \mathcal{F}, \mu)$  be defined as  $U_s f(x) = f(T_s x)$  for all  $s \in \mathbb{R}$ . Then  $U_s$  is a unitary operator on  $L^2(\mu)$ .

**Lemma 5.2.** If  $U$  is a unitary operator on a Hilbert space  $H$ , then we have  $\text{ran}(I - U)^\perp = \ker(I - U)$  where  $I : H \rightarrow H$  is the identity operator.

**Theorem 5.1** (Von Neumann Ergodic Theorem). Let  $U$  be a unitary operator on a Hilbert space  $H$ . Let  $P$  be the orthogonal projection on  $\ker(I - U)$ . Then we have for all  $x \in H$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n U^k f = Pf \text{ in } H$$

**Remark 5.1.** The convergence above is of course meant with respect to the norm in  $H$  which is induced by the scalar product that is given on  $H$ .

**Lemma 5.3.** *Let us define  $\mathcal{I} := \{A \in \mathcal{F} : T_t^{-1}(A) = A \text{ for all } t \in \mathbb{R}\}$ , then  $\mathcal{I}$  is a  $\sigma$ -algebra, it is called the  $\sigma$ -algebra of invariant sets.*

*Proof.* Exercise! □

**Corollary 5.1.** *Let  $\mu$  be an invariant measure under  $(T_t)_{t \in \mathbb{R}}$ ,  $\mathcal{I}$  the  $\sigma$ -algebra of invariant sets as above, then we have*

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(T_s \omega) ds = \mathbb{E}(f \mid \mathcal{I}) \text{ in } L^2(\mu).$$



## 6 Continuous time Martingales

**Definition 6.1.** Let  $X = (X_t)_{t \in T}$  be a stochastic process on some filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in T}, \mathbb{P})$ . Then we say that  $X$  is a martingale if

1.  $X$  is  $\mathcal{F}_t$ -adapted, i.e.  $X_t$  is  $\mathcal{F}_t$ -measurable for all  $t \in T$ .
2.  $X_t$  is  $\mathbb{P}$ -integrable for all  $t \in T$ .
3.  $\mathbb{E}(X_t | \mathcal{F}_s) = X_s$  for all  $s < t$ .

**Definition 6.2.** Let  $(X_t)_{t \geq 0}$  and  $(Y_t)_{t \geq 0}$  be stochastic processes. We say

1.  $X_t$  and  $Y_t$  are indistinguishable if  $\mathbb{P}(X_t = Y_t, \text{ for all } t \in \mathbb{R}_0^+) = 1$
2.  $X_t$  is a modification of  $Y_t$  (and vice versa) if  $\mathbb{P}(X_t = Y_t) = 1$  for all  $t \in \mathbb{R}_0^+$
3.  $X_t$  and  $Y_t$  have the same FDD (Finite Dimensional Distributions) if for all  $n \in \mathbb{N}$  and for all  $H \in \mathcal{B}(\mathbb{R}^n)$  we have

$$\mathbb{P}((X_{t_1}, \dots, X_{t_n}) \in H) = \mathbb{P}((Y_{t_1}, \dots, Y_{t_n}) \in H)$$

**Remark 6.1.** We do indeed have that  $1) \implies 2) \implies 3)$ .

**Definition 6.3.** Let  $(\mathcal{F}_t)_{t \geq 0}$  be a Filtration. We define the left respectively right continuous Filtrations as

$$\mathcal{F}_{t-} := \bigvee_{s < t} \mathcal{F}_s \text{ and } \mathcal{F}_{t+} := \bigcap_{s > t} \mathcal{F}_s$$

**Remark 6.2.** We have the inclusions  $\mathcal{F}_{t-} \subset \mathcal{F}_t \subset \mathcal{F}_{t+}$  for all  $t \in \mathbb{R}_0^+$  and it is very well possible that these inclusions are strict.

**Definition 6.4.** Let  $(\mathcal{F}_t)_{t \geq 0}$  be a Filtration. We define

$$\tilde{\mathcal{F}}_t := \sigma(\mathcal{F}_{t+}, \mathcal{N}), \text{ where } \mathcal{N} \text{ is the set of } \mathbb{P}\text{-negligible sets.}$$

We say that our Filtration  $(\mathcal{F}_t)_{t \geq 0}$  satisfies the usual conditions if  $\mathcal{F}_t = \tilde{\mathcal{F}}_t$  for all  $t \in \mathbb{R}_0^+$ .

We come now to our main result in this section. It states that we can regulate martingales, i.e. represent them as modifications through cadlag (continue a droite, limites a gauche) martingales.

**Theorem 6.1** (Martingale regularization theorem).

Let  $(X_t)_{t \geq 0}$  be a  $\mathcal{F}_t$ -martingale, then there exists  $(\widetilde{X}_t)_{t \geq 0}$  a  $\widetilde{\mathcal{F}}_t$ -martingale which is cadlag such that  $X_t = \mathbb{E}(\widetilde{X}_t | \mathcal{F}_t)$  almost surely for all  $t \geq 0$ .

Moreover, if  $(\mathcal{F}_t)_{t \geq 0}$  satisfies the usual conditions, then  $\widetilde{X}_t$  is a cadlag modification of  $X_t$ .

The Martingale regularization theorem is a powerful tool when it comes to extending statements we already know from discrete time martingales to continuous time martingales, because we can regularize them by cadlag representations and then approximate them (using for instance that  $\mathbb{Q}$  is dense in  $\mathbb{R}$ ).

**Theorem 6.2** (Convergence Theorem). *Let  $X$  be a cadlag martingale with respect to  $(\mathcal{F}_t)_{t \geq 0}$ . If  $X$  is bounded in  $L^1$ , i.e. if  $\sup_{t \geq 0} \|X_t\|_{L^1} < \infty$ , then  $X_t \rightarrow X_\infty$  as  $t \rightarrow \infty$  for some  $X_\infty \in L^1(\Omega, \mathcal{F}_\infty, \mathbb{P})$ . Moreover if  $X$  is UI, then  $X_t \rightarrow X_\infty$  in  $L^1(\Omega, \mathcal{F}, \mathbb{P})$ .*

*Proof.* Exercise! □

**Theorem 6.3** (Stopping Theorem). *Let  $X$  be a cadlag martingale and let  $\tau \leq \Lambda$  be stopping times with respect to  $(\mathcal{F}_t)_{t \geq 0}$ . If  $\Lambda \leq C$  for some constant  $C > 0$ , then*

$$\mathbb{E}(X_\Lambda \mid \mathcal{F}_\tau) = X_\tau \tag{*}$$

*Moreover, if  $X$  is UI, then (\*) above is true for all stopping times  $\tau \leq \Lambda$ .*

*Proof.* Exercise! □

## 7 Brownian Motion

**Definition 7.1.** A stochastic process  $X = (X_t)_{t \geq 0}$  is Gaussian if for every  $n \in \mathbb{N}_{\geq 1}$  and times  $t_1, \dots, t_n \in [0, \infty)$  the vector  $(X_{t_1}, \dots, X_{t_n})$  has a multivariate Gaussian distribution.

**Remark 7.1.**

1. A Gaussian process  $X$  is fully characterized by its mean  $\mathbb{E}(X_t)$  and by its covariance kernel  $K(t, t') := \text{Cov}(X_t, X_{t'})$
2. If  $\mathbb{E}(X_t) = 0$ , then we have  $K(t, t') = \text{Cov}(X_t, X_{t'}) = \mathbb{E}(X_t X_{t'})$

**Proposition 7.1** (Characterization). *The following two statements are equivalent:*

1.  $(X_t)_{t \geq 0}$  has stationary, independent increments and  $X_t \sim \mathcal{N}(0, t)$  for all  $t \geq 0$ .
2.  $(X_t)_{t \geq 0}$  is a Gaussian process with mean zero and covariance kernel  $\mathbb{E}(X_t X_{t'}) = t \wedge t'$  for all  $t, t' \in \mathbb{R}_0^+$ .

**Remark 7.2.** We recall the following

1. Independence of the increments means that for all times  $0 \leq t_1 \leq t_2 \leq \dots \leq t_n < \infty$  where  $n \in \mathbb{N}$ ,  $X_{t_n} - X_{t_{n-1}}, \dots, X_{t_3} - X_{t_2}, X_{t_2} - X_{t_1}$  are independent.
2. Stationarity of the increments means that for all  $s < t$  we have that  $X_t - X_s$  is equal in distribution as  $X_{t-s}$ .
3. We can also summarize 1) and 2) above as, if for all  $t \in \mathbb{R}_0^+$  we have

$$\begin{aligned} & (X_{t_n} - X_{t_{n-1}}, \dots, X_{t_3} - X_{t_2}, X_{t_2} - X_{t_1}) \\ & \sim (X_{t+t_n} - X_{t+t_{n-1}}, \dots, X_{t+t_3} - X_{t+t_2}, X_{t+t_2} - X_{t+t_1}) \end{aligned}$$

4. Since  $X_t \sim \mathcal{N}(0, t)$  we have by the stationarity  $X_t - X_s \sim X_{t-s} \sim \mathcal{N}(0, t-s)$ .

**Definition 7.2.** A Brownian motion is a stochastic process  $(B_t)_{t \geq 0}$  with continuous sample paths (i.e. a continuous stochastic process) which satisfies either property 1) or 2) (and consequently both of them) in Proposition 7.1.