

# Advanced Probability (M24)

James Norris and Wei Qian

The aim of the course is to introduce students to advanced topics in modern probability theory. The emphasis is on tools required in the rigorous analysis of stochastic processes, such as Brownian motion, and in applications where probability theory plays an important role.

**Review of measure and integration:** sigma-algebras, measures and filtrations; integrals and expectation; convergence theorems; product measures, independence and Fubini's theorem.

**Conditional expectation:** Discrete case, Gaussian case, conditional density functions; existence and uniqueness; basic properties.

**Martingales:** Martingales and submartingales in discrete time; optional stopping; Doob's inequalities, upcrossings, martingale convergence theorems; applications of martingale techniques.

**Stochastic processes in continuous time:** Kolmogorov's criterion, regularization of paths; martingales in continuous time.

**Weak convergence:** Definitions and characterizations; convergence in distribution, tightness, Prokhorov's theorem; characteristic functions, Lévy's continuity theorem.

**Sums of independent random variables:** Strong laws of large numbers; central limit theorem; Cramér's theory of large deviations.

**Brownian motion:** Wiener's existence theorem, scaling and symmetry properties; martingales associated with Brownian motion, the strong Markov property, hitting times; properties of sample paths, recurrence and transience; Brownian motion and the Dirichlet problem; Donsker's invariance principle.

**Poisson random measures:** Construction and properties; integrals.

**Lévy processes:** Lévy-Khinchin theorem.

## Pre-requisites

A basic familiarity with measure theory and the measure-theoretic formulation of probability theory is very helpful. These foundational topics will be reviewed at the beginning of the course, but students unfamiliar with them are expected to consult the literature (for instance, Williams' book) to strengthen their understanding.

## Literature

- Lecture notes online: [www.statslab.cam.ac.uk/~james/Lectures/ap.pdf](http://www.statslab.cam.ac.uk/~james/Lectures/ap.pdf)
- D. Applebaum, Lévy processes (2nd ed.), Cambridge University Press 2009.
- R. Durrett, Probability: Theory and Examples (4th ed.), CUP 2010.
- O. Kallenberg, Foundations of Modern Probability, Springer-Verlag, 1997.
- D. Williams, Probability with martingales, CUP 1991.

## Additional support

Four example sheets will be provided along with supervisions or examples classes. There will be a revision class in Easter term.

# Advanced Probability

Lectured by : Prof. James Norris(<http://www.statslab.cam.ac.uk/~james/>), Dr. Wei Qian  
Example Classes given by : Peter Taylor, Piet Lammers

Typed by : Jiwoon Park

*Some remarks :* This document was created during the lecture “*Advanced Probability*”, Michaelmas term of 2018, University of Cambridge, Mathematics Tripos Part III. This is a 24-hours lecture course, and comes with 4 example sheets that can be found on <http://www.statslab.cam.ac.uk/~james/Lectures/>. This course is intended for students in first year of Masters degree in mathematics or equivalent. Background on elementary measure theory and functional analysis would be very helpful in this course.

I tried to follow closely to the materials provided in the lecture, but minor modification also had been make. If there is any error in this notes, it is surely mine, and it has nothing to do with the lecturer or the university.

This document was uploaded on <http://www.github.com/ChocoUyou>, along with its source code in  $\text{\LaTeX}$ . I hold no right on this document, and you are free to redistribute this document with any of its details below modified, but please do not just redistribute with any of the detail above this line modified.

## 0. Review

### 0.1. Measure Spaces

Let  $E$  be a set.

Let  $\mathcal{E}$  be a set of subsets of  $E$ . We say that  $\mathcal{E}$  is a  **$\sigma$ -algebra** if  $\emptyset \in \mathcal{E}$  and for  $A \in \mathcal{E}$  and for any sequence  $(A_n : n \in \mathbb{N})$  in  $\mathcal{E}$ , we have  $E \setminus A \in \mathcal{E}$  and  $\cup_n A_n \in \mathcal{E}$ . Then  $(E, \mathcal{E})$  is called a **measurable space** and we call the elements of  $\mathcal{E}$  **measurable**.

Let  $\mu$  be a function  $\mathcal{E} \rightarrow [0, \infty]$ . We say that  $\mu$  is a **measure** if  $\mu(\emptyset) = 0$  and for all sequences  $(A_n : n \in \mathbb{N})$  of disjoint elements, we have  $\mu(\cup_n A_n) = \sum_n \mu(A_n)$ . (**countable additivity**). The triple  $(E, \mathcal{E}, \mu)$  is called the **measure space**.

Given a topological space  $E$  there is a smallest  $\sigma$ -algebra containing all the open sets. This is the **Borel  $\sigma$ -algebra**, denoted  $\mathcal{B}(E)$ .

E.g. For  $\mathbb{R}$ , has  $\mathcal{B}(\mathbb{R}) = \mathcal{B}$ , and  $[0, \infty]$  (the extended real positive line) has analogous Borel  $\sigma$ -algebra.

### 0.2. Integration of Measurable Functions

Let  $(E, \mathcal{E})$  and  $(E', \mathcal{E}')$  be measurable spaces. A function  $f : E \rightarrow E'$  is **measurable** if

$$f^{-1}(A) = \{x \in E : f(x) \in A\} \in \mathcal{E} \quad \forall A \in \mathcal{E}'$$

If we refer to a **measurable function**  $f$  without specifying range, the default is  $(\mathbb{R}, \mathcal{B})$ .

If we refer to  $f$  as a **non-negative measurable function**, then we mean  $E' = [0, \infty]$ ,  $\mathcal{E}' = \mathcal{B}([0, \infty])$ . Write  $m\mathcal{E}^+$  for the set of non-negative measurable functions.

**Theorem 0.2.1.)** Let  $(E, \mathcal{E}, \mu)$  be a measure space. Then there exists a *unique* map  $\tilde{m}u : m\mathcal{E}^+ \rightarrow [0, \infty]$  such that

- (a)  $\tilde{m}u(1_A) = \mu(A)$  for all  $A \in \mathcal{E}$ .
- (b)  $\tilde{m}u(\alpha f + \beta g) = \alpha \tilde{m}u(f) + \beta \tilde{m}u(g)$  for all  $\alpha, \beta \in [0, \infty)$  and all  $f, g \in m\mathcal{E}^+$ .
- (c) (**monotone convergence**)  $\tilde{m}u(f) = \lim_{n \rightarrow \infty} \tilde{m}u(f_n)$  for a non-decreasing sequence of functions  $(f_n)_{n \in \mathbb{N}}$  in  $m\mathcal{E}^+$  such that  $f_n(x) \rightarrow f(x)$  as  $n \rightarrow \infty$  for all  $x \in E$ .

(End of statement)  $\square$

For now on, write  $\mu$  for  $\tilde{m}u$ . Call  $\mu(f)$  the **integral of  $f$  w.r.t.  $\mu$** . Also write  $\int_E f d\mu = \int_E f(x) \mu(dx)$ .

A **simple function** is one which is a finite linear combination of indicator functions of measurable sets with positive coefficients. So  $f$  is simple if  $f = \sum_{k=1}^n \alpha_k 1_{A_k}$  for some  $n \geq 0$ ,  $\alpha_k \in (0, \infty)$  and  $A_k \in \mathcal{E}$  for  $k \in \{1, \dots, n\}$ .

Form (a) and (b) for  $f$  simple, we have

$$\mu(f) = \sum_{k=1}^n \alpha_k \mu(A_k)$$

Also, if  $f, g \in m\mathcal{E}^+$  with  $f \leq g$  then  $f + h = g$  where  $h = g - f1_{f < \infty} \in m\mathcal{E}^+$ . Since  $\mu(h) \geq 0$ , (b) implies  $\mu(f) \leq \mu(g)$ .

Take  $f \in m\mathcal{E}^+$ . Define for  $x \in E$ ,  $n \in \mathbb{N}$ ,

$$f_n(x) = 2^{-n} [2^n f(x)] \wedge n$$

Then  $(f_n : n \in \mathbb{N})$  is a non-decreasing sequence of simple functions such that  $f_n(x) \rightarrow f(x)$  for all  $x \in E$ . By monotone convergence(property (c)), we have

$$\mu(f) = \lim_{n \rightarrow \infty} \mu(f_n)$$

These procedure characterizes  $\mu(f)$  uniquely in terms of the measure  $\mu : \mathcal{E} \rightarrow [0, +\infty]$ , (a), (b) and (c), so we have shown the uniqueness in **Theorem 0.2.1**.

When is  $\mu(f) = 0$ ? (for  $f \in m\mathcal{E}^+$ )

For measurable functions  $f, g$  we say  $f = g$  **almost everywhere**(or a.e.) if

$$\mu(\{x \in E : f(x) \neq g(x)\}) = 0$$

Can show, for  $f \in m\mathcal{E}^+$ , we have  $\mu(f) = 0$  if and only if  $f = 0$  a.e.

Let  $f$  be a measurable function. We say that  $f$  is **integrable** if  $\mu(|f|) < \infty$ . Write  $L^1 = L^1(E, \mathcal{E}, \mu)$  for the set of all integrable functions. We extend the integral to  $L^1$  by setting  $\mu(f) = \mu(f^+) - \mu(f^-)$  where  $f^\pm(x) = 0 \vee (\pm f(x))$  so that  $f = f^+ - f^-$ . Then can show  $L^1$  is a vector space  $\mu : L^1 \rightarrow \mathbb{R}$  is linear.

**Lemma 0.2.2.) (Fatou's lemma)** Let  $(f_n : n \in \mathbb{N})$  be a sequence in  $m\mathcal{E}^+$ . Then

$$\mu(\liminf_{n \rightarrow \infty} f_n) \leq \liminf_{n \rightarrow \infty} \mu(f_n)$$

(End of statement)  $\square$

The proof is done by applying monotone convergence to the sequence  $(\inf_{m \geq n} f_m : n \in \mathbb{N})$ .

(8th October, Monday)

**Theorem 0.2.3)** (Dominated Convergence) Let  $(f_n : n \in \mathbb{N})$  be a sequence of measurable functions on  $(E, \mathcal{E})$ . Suppose  $f_n(x) \rightarrow f(x)$  as  $n \rightarrow \infty$  for some function  $f$  and for every  $x \in E$ . Suppose furthermore that  $|f_n| \leq g$  for all  $n$ , for some measurable function  $g$ . Then  $f_n$  is integrable for all  $n$  and so is  $f$ . Moreover, we have  $\mu(f_n) \rightarrow \mu(f)$  as  $n \rightarrow \infty$ .

We call a measure space  $(\Omega, \mathcal{F}, \mathbb{P})$  such that  $\mathbb{P}(\Omega) = 1$  a **probability space**, with the following terms replaced:

measurable functions( $f$ )  $\leftrightarrow$  random variable( $X$ )

measurable sets  $\leftrightarrow$  events

almost everywhere  $\leftrightarrow$  almost surely

integral  $\leftrightarrow$  expectation( $\mathbb{E}X = \int_{\Omega} X d\mathbb{P}$  in place of  $\mathbb{P}(X)$ )

### 0.3. Production measure and Fubini's theorem

Take finite (or  $\sigma$ -finite) measure spaces  $(E_1, \mathcal{E}_1, \mu_1)$  and  $(E_2, \mathcal{E}_2, \mu_2)$ . Write  $\mathcal{E}_1 \otimes \mathcal{E}_2$  for the  $\sigma$ -algebra on  $E_1 \times E_2$  generated by sets of the form  $A_1 \times A_2$  for  $A_1 \in \mathcal{E}_1$  and  $A_2 \in \mathcal{E}_2$  and call this **product  $\sigma$ -algebra**.

**Theorem 0.3.1.)** There exists a unique measure  $\mu = \mu_1 \otimes \mu_2$  on  $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$  such that

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \quad \forall A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2$$

**Theorem 0.3.2.)** (Fubini's theorem) Let  $f$  be a non-negative measurable function on  $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ .

- $x_1 \in E_1$ , set  $f_{x_1}(x_2) = f(x_1, x_2)$ , then  $f_{x_1}$  is  $\mathcal{E}_2$ -measurable for all  $x_1 \in E_1$ .
- Set  $f_1(x_1) = \mu_2(f_{x_1})$ . Then  $f_1$  is  $\mathcal{E}_1$ -measurable, and  $\mu_1(f_1) = \mu(f)$ .

(End of statement)  $\square$

Now define  $\hat{f}$  on  $E_2 \times E_1$  by  $\hat{f}(x_2, x_1) = f(x_1, x_2)$ . Then we can show that  $\hat{f}$  is  $\mathcal{E}_2 \otimes \mathcal{E}_1$ -measurable and  $\mu_2 \otimes \mu_1(\hat{f}) = (\mu_1 \otimes \mu_2)(f)$ . So by Fubini, we have  $\mu_2(f_2) = \hat{\mu}(\hat{f}) = \mu(f)\mu_1(f_1)$  where  $f_2$  is defined as for  $f_1$ .

$$\text{i.e. } \int_{E_2} \left( \int_{E_1} f(x_1, x_2) \mu_1(dx_1) \right) \mu_2(dx_2) = \int_{E_1} \left( \int_{E_2} f(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1).$$

Note,  $f$  being integrable is also sufficient to for Fubini's theorem.

## Chapter 1. Conditional Expectation

### 1.1. Discrete Case

Suppose  $(G_n : n \in \mathbb{N})$  is a sequence of disjoint sets in  $\mathcal{F}$  such that  $\cup_n G_n = \Omega$ . Let  $X$  be an integrable random variable.

- Set  $\mathcal{G} = \sigma(G_n : n \in \mathbb{N}) = \{\cup_{n \in I} G_n : I \subset \mathbb{N}\}$ .
- Define a new random variable  $Y = \sum_{n \in \mathbb{N}} \mathbb{E}(X|G_n) 1_{G_n}$  where  $\mathbb{E}(X|G_n) = \mathbb{E}(X 1_{G_n}) / \mathbb{P}(G_n)$  (used convention  $0/0=0$ ). Then,
  - (a)  $Y$  is  $\mathcal{G}$ -measurable.
  - (b)  $Y$  is integrable and  $\mathbb{E}(Y 1_A) = \mathbb{E}(X 1_A)$  for all  $A \in \mathcal{G}$ .

We denote  $Y = \mathbb{E}(X|\mathcal{G})$  and call (a version of) the **conditional expectation** of  $X$  given  $\mathcal{G}$ .

## 1.2. Gaussian Case

Let  $(W, X)$  be a Gaussian random variable in  $\mathbb{R}^2$  that has distribution  $N(\mu, \Sigma)$ . Take  $\mathcal{G} = \sigma(W) = \{W^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$ .

Consider, for  $a, b \in \mathbb{R}$ , the random variable  $Y = aW + b$ .

We can choose  $a, b$  so that  $\mathbb{E}(Y) = a\mathbb{E}(X) + b = \mathbb{E}(X)$  and  $\text{Cov}(Y - X, W) = a\text{Var}(W) - \text{Var}(X, W) = 0$ . Then

- (a)  $Y$  is  $\mathcal{G}$ -measurable.
- (b)  $Y$  is integrable and  $\mathbb{E}(Y1_A) = \mathbb{E}(X1_A)$  for all  $A \in \mathcal{G}$ .

To see this, note  $Y - X$  and  $W$  are independent as  $\text{Cov}(Y - X, W) = 0$  so  $\mathbb{E}((Y - X)1_A) = \mathbb{E}(Y - X)\mathbb{P}(A) = 0$  for all  $A \in \mathcal{G}$ .

## 1.3. Conditional density functions

Let  $(U, V)$  be a random variable in  $\mathbb{R}^2$  with density function  $f(u, v)$ . That is,  $\mathbb{P}((U, V) \in A) = \int_A f(u, v) du dv$ . Take  $\mathcal{G} = \sigma(U) = \{U^{-1}(B) : B \in \mathcal{B}\}$ . Take a Borel measurable function  $h$  on  $\mathbb{R}$  and set  $X = h(V)$ . Assume  $X \in L^1(\mathbb{P})$ . Note  $U$  has density function  $f(u) = \int f(u, v) dv$ . Define the **conditional density function**  $f(u|v) = f(u, v)/f(v)$  (again with convention  $0/0=0$ ). Now set  $Y = g(U)$  where  $g(u) = \int_{\mathbb{R}} h(v)f(v|u)dv$ . Then  $g$  is a Borel function on  $\mathbb{R}$  so

- (a)  $Y$  is a  $\mathcal{G}$ -measurable random variable.
- (b)  $Y$  is integrable and for all  $A = U^{-1}(B) \in \mathcal{G}$ , we have  $\mathbb{E}(Y1_A) = \mathbb{E}(X1_A)$ .

To see this,

$$\begin{aligned} \mathbb{E}(Y1_A) &= \int_{\mathbb{R}} g(u)1_B(u)f(u)du \\ &= \int \int h(v)f(v|u)dv1_B(u)f(u)du \\ &= \mathbb{E}(X1_A) \end{aligned}$$

using Fubini's theorem to go from second to the third line.

(10th October, Wednesday)

## 1.4. Existence and uniqueness of conditional expectation

$(\Omega, \mathcal{F}, \mathbb{P})$  is always the probability space behind.

**Theorem 1.4.1.)** Let  $X$  be integrable random variable and let  $\mathcal{G}$  be a sub- $\sigma$ -algebra of  $\mathcal{F}$ . There exist a random variable  $Y$  such that

- (a)  $Y$  is  $\mathcal{F}$ -measurable,
- (b)  $Y$  is integrable and for all  $A \in \mathcal{F}$ ,  $\mathbb{E}(Y1_A) = \mathbb{E}(X1_A)$ .

Moreover, if  $Y'$  is another random variable satisfying (a) and (b) then  $Y' = Y$  a.s.

We will write  $Y = \mathbb{E}(X|\mathcal{F})$  a.s. (because  $\mathbb{E}(X|\mathcal{F})$  is not a fully specified random variable) and we say  $Y$  is a **(version of) the conditional expectation of  $X$  given  $\mathcal{F}$** . In case  $X = 1_A$  write  $Y = \mathbb{P}(A|\mathcal{F})$  a.s.

An analogous statement holds with 'integrable' replaced by 'non-negative' throughout the statement.

**proof** (*Uniqueness*) Suppose  $Y$  satisfies (a) and (b), and  $Y'$  satisfies (a) and (b) for an integrable random variable  $X'$ , with  $X \leq X'$ . Consider the non-negative random variable  $Z = (Y - Y')1_A$  where  $A = \{Y \geq Y'\} \in \mathcal{F}$ . Then

$$\mathbb{E}(Y1_A) = \mathbb{E}(X1_A) \leq \mathbb{E}(X'1_A) = \mathbb{E}(Y'1_A)$$

so  $\mathbb{E}(Z) \leq 0$  so  $Z = 0$  a.s. and so  $Y \leq Y'$  a.s.

By symmetry, if  $X = X'$ , we see that  $Y = Y'$  a.s.

(*Existence*) Consider for now the case  $X \in L^2(\mathcal{F})$  (that is,  $\mathbb{E}(|X|^2) < \infty$ , which is stronger than being integrable). Since  $L^2(\mathcal{F})$  is a Hilbert space, and in fact  $L^2(\mathcal{F})$  is a closed subspace of  $L^2(\mathcal{F})$ , there exists an orthogonal projection  $Y$  of  $X$  on  $L^2(\mathcal{F})$ . That is,  $Y \in L^2(\mathcal{F})$  and for all  $Z \in L^2(\mathcal{F})$ , we have  $\mathbb{E}((Y - X)Z) = 0$ . (This is because  $(X - Y)$  should be orthogonal to  $L^2(\mathcal{F})$ , and the inner product is given by  $\langle X - Y, Z \rangle = \mathbb{E}((X - Y)Z) = 0$ .) So for any  $A \in \mathcal{F}$ , take  $Z = 1_A$  to see  $\mathbb{E}(Y1_A) = \mathbb{E}(X1_A)$ . So  $Y$  satisfies (a) and (b).

Suppose now that  $X \geq 0$ . Set  $X_n = X \wedge n$  then  $X_n \in L^2(\mathcal{F})$ . We have shown there exist  $Y_n \in L^2(\mathcal{F})$  such that for all  $A \in \mathcal{F}$ , we have  $\mathbb{E}(Y_n1_A) = \mathbb{E}(X_n1_A)$ ,  $0 \leq X_n \leq X_{n+1}$  and  $0 \leq Y_n \leq Y$  for all  $n$ . Set  $\Omega_0 = \{\omega \in \Omega : 0 \leq Y_n(\omega) \leq Y_{n+1}(\omega) \text{ for all } n\}$ . Then  $\mathbb{P}(\Omega_0) = 1$  (because it is a countable intersection of sets of measure 1). Define a non-negative  $\mathcal{F}$ -measurable random variable

$$Y_\infty = \lim_{n \rightarrow \infty} 1_{\Omega_0} Y_n$$

Then  $0 \leq 1_{\Omega_0} Y_n \nearrow Y_\infty$  as  $n \rightarrow \infty$  so by monotone convergence, we have for all  $A \in \mathcal{F}$

$$\mathbb{E}(Y_\infty 1_A) = \lim_{n \rightarrow \infty} \mathbb{E}(Y_n 1_A) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n 1_A) = \mathbb{E}(X 1_A)$$

In particular, taking  $A = \Omega$ , we see  $Y_\infty$  is integrable, so  $\mathbb{P}(Y_\infty) = 0$ . So if we set  $Y = Y_\infty 1_{Y_\infty < \infty}$ , then  $Y$  is a random variable and satisfies (a) and (b).

In the general case we can apply the proceeding to  $X^\pm = 0 \vee (\pm X)$  (so  $X = X^+ - X^-$ ) to obtain  $Y^\pm$  and then check  $Y = Y^+ - Y^-$  satisfies (a) and (b).

(*End of proof*)  $\square$

## 1.5. Properties of conditional expectation

Conditional expectations plays a central role in many core materials of the course. e.g. martingales and Brownian motion. So it would be useful to investigate some properties of conditional expectation.

Fix an integrable random variable  $X$  and sub- $\sigma$ -algebra  $\mathcal{F}$ . **Theorem 1.4.1** has useful consequences:

- (i)  $\mathbb{E}(\mathbb{E}(X|\mathcal{F})) = \mathbb{E}(X)$

- (ii) If  $X$  is  $\mathcal{F}$ -measurable then  $\mathbb{E}(X|\mathcal{F}) = X$  a.s.
- (iii) If  $X$  is independent of  $\mathcal{F}$  then  $\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(X)$  a.s.  
 : check for  $A \in \mathcal{F}$ ,  $\mathbb{E}(X1_A) = \mathbb{E}(X)\mathbb{P}(A) = \mathbb{E}(\mathbb{E}(X)1_A)$  so  $\mathbb{E}(X)$  satisfies condition (b).
- (iv) If  $X \geq 0$  a.s. then  $\mathbb{E}(X|\mathcal{F}) \geq 0$  a.s. (follows from the proof of **Theorem 1.4.1**)
- (v) For all  $\alpha, \beta \in \mathbb{R}$ , all integrable random variables  $X, Y$ , has

$$\mathbb{E}(\alpha X + \beta Y|\mathcal{F}) = \alpha \mathbb{E}(X|\mathcal{F}) + \beta \mathbb{E}(Y|\mathcal{F}) \quad \text{a.s.}$$

: check that the expression on the RHS satisfies condition (b) to show that it is a version of conditional expectation of  $\alpha X + \beta Y$  given  $\mathcal{F}$ .

Now consider a sequence of random variables  $(X_n)_n$ .

- (vi) (*conditional monotone convergence*) If  $0 \leq X_n \nearrow X$  pointwise then  $\mathbb{E}(X_n|\mathcal{F}) \rightarrow \mathbb{E}(X|\mathcal{F})$  a.s.  
 : To see this, we know  $\mathbb{E}(X_n|\mathcal{F}) \nearrow Y$  a.s. for some non-negative  $\mathcal{F}$ -measurable random variable  $Y$ . Take  $A \in \mathcal{F}$  then by monotone convergence, we have

$$\begin{aligned} \mathbb{E}(Y1_A) &= \lim_{n \rightarrow \infty} \mathbb{E}(\mathbb{E}(X_n|\mathcal{F})1_A) \\ &= \lim_{n \rightarrow \infty} \mathbb{E}(X_n1_A) = \mathbb{E}(X1_A) \end{aligned}$$

so  $Y$  satisfies (b) for  $X$ .

- (vii) (*Conditional Fatou's lemma*) For any non-negative random variables  $X_n$ ,

$$\mathbb{E}(\liminf_{n \rightarrow \infty} X_n|\mathcal{F}) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n|\mathcal{F})$$

: By conditional monotone convergence, we have

$$\begin{aligned} \mathbb{E}(\liminf_{n \rightarrow \infty} X_n|\mathcal{F}) &= \lim_{n \rightarrow \infty} \mathbb{E}(\inf_{m \geq n} X_m|\mathcal{F}) \leq \lim_{n \rightarrow \infty} \inf_{m \geq n} \mathbb{E}(X_m|\mathcal{F}) \\ &= \liminf_{n \rightarrow \infty} \mathbb{E}(X_n|\mathcal{F}) \end{aligned}$$

so we have the result.

- (viii) (*Conditional dominated convergence*) If  $X_n(\omega) \rightarrow X(\omega)$  for all  $\omega$  as  $n \rightarrow \infty$  and there exist  $Y$  integrable so  $|X_n| \leq Y$  for all  $n$ , then

$$\mathbb{E}(X_n|\mathcal{F}) \rightarrow \mathbb{E}(X|\mathcal{F}) \quad \text{a.s.}$$

: First, note that  $X_n$  and  $X$  are integrable, since they are bounded by  $Y$ . Consider the sequence of non-negative random variables  $Z_n = Y - X_n$ . By Fatou's lemma,

$$\mathbb{E}(\liminf_n Z_n|\mathcal{F}) = \mathbb{E}(Y - X|\mathcal{F}) \leq \liminf_n \mathbb{E}(Y - X_n|\mathcal{F}) = \mathbb{E}(Y|\mathcal{F}) - \limsup_n \mathbb{E}(X_n|\mathcal{F})$$

and hence

$$\limsup_n \mathbb{E}(X_n|\mathcal{F}) \leq \mathbb{E}(X|\mathcal{F})$$



On the other hand, if we let  $W_n = Y + X_n$ , applying Fatou's lemma gives

$$\mathbb{E}(\liminf_n W_n | \mathcal{F}) = \mathbb{E}(Y | \mathcal{F}) + \mathbb{E}(X | \mathcal{F}) \leq \mathbb{E}(Y | \mathcal{F}) + \liminf_n \mathbb{E}(X_n | \mathcal{F})$$

and so

$$\liminf_n \mathbb{E}(X_n | \mathcal{F}) \geq \mathbb{E}(X | \mathcal{F})$$

The two inequalities give the result.

(12th October, Friday)

(ix) (*Conditional Jensen*) Let  $c : \mathbb{R} \rightarrow (-\infty, \infty]$  be convex. Then  $c(\mathbb{E}(X | \mathcal{F})) \leq \mathbb{E}(c(X) | \mathcal{F})$  a.s.

: To see this, since  $c$  is convex there exist real sequences  $(a_n), (b_n)$  such that  $c(x) = \sup_n (a_n x + b_n)$ . Then  $c(X) \geq a_n X + b_n$  for all  $n$ . So

$$\mathbb{E}(c(X) | \mathcal{F}) \geq \mathbb{E}(a_n X + b_n | \mathcal{F}) = a_n \mathbb{E}(X | \mathcal{F}) + b_n \quad \text{a.s.}$$

so  $\mathbb{E}(c(X) | \mathcal{F}) \geq c(\mathbb{E}(X | \mathcal{F}))$  a.s.

(x)  $\|\mathbb{E}(X | \mathcal{F})\|_p \leq \|X\|_p$  for all  $p \in [0, \infty)$  (where  $\|X\|_p = \mathbb{E}(|X|^p)^{1/p}$ )

: To see this, take  $c(x) = |x|^p$  in conditional Jensen's inequality.

(xi) (*Tower property*) This is also important for martingales.

Suppose  $\mathcal{H} \subset \mathcal{F} \subset \mathcal{F}$  be sub- $\sigma$ -algebras. Then

$$\mathbb{E}(X | \mathcal{H}) = \mathbb{E}(\mathbb{E}(X | \mathcal{F}) | \mathcal{H}) \quad \text{a.s.}$$

: Proof is an exercise.

(xii) (*Taking out what is known*) This is related to the 'filtration' that is to be introduced in the theory of martingales.

Let  $Y$  be a bounded  $\mathcal{F}$ -measurable random variable. Then

$$\mathbb{E}(YX | \mathcal{F}) = Y\mathbb{E}(X | \mathcal{F}) \quad \text{a.s.} \quad \dots\dots\dots (\star)$$

: To see this, consider first the case when  $Y = 1_B$ ,  $B \in \mathcal{F}$ . Take  $A \in \mathcal{F}$ , then

$$\mathbb{E}(Y\mathbb{E}(X | \mathcal{F})1_A) = \mathbb{E}(\mathbb{E}(X | \mathcal{F})1_A) = \mathbb{E}(X1_{AB}) = \mathbb{E}(YX1_A)$$

so above identity  $(\star)$  holds. The general case follows by a monotone class argument. i.e.  $(\star)$  extends by linearity to case where  $Y$  is simple. Next consider  $X, Y \geq 0$  and set  $Y_n = (2^{-n} \lfloor 2^n Y \rfloor) \wedge n$ . Then  $Y_n$  is simple, so  $\mathbb{E}(Y_n X 1_A) = \mathbb{E}(Y_n \mathbb{E}(X | \mathcal{F}))$  for all  $A \in \mathcal{F}$ . Let  $n \rightarrow \infty$  using monotone convergence to obtain  $(\star)$  for  $Y$ . Finally extend to  $X, Y$  general using  $X = X^+ - X^-$  and  $Y = Y^+ - Y^-$ . (Or instead, use *monotone class theorem*, that abstracts this process)

(xiii) Let  $\mathcal{H}$  be a  $\sigma$ -algebra and suppose  $\mathcal{H}$  is independent of  $\sigma(X, \mathcal{F})$ . Then

$$\mathbb{E}(X | \sigma(\mathcal{F}, \mathcal{H})) = \mathbb{E}(X | \mathcal{H}) \quad \text{a.s.}$$

: To see this, put  $Z = \mathbb{E}(X|\sigma(\mathcal{F}, \mathcal{H}))$  and  $Y = \mathbb{E}(X|\mathcal{H})$ . Take  $A \in \mathcal{F}$ ,  $B \in \mathcal{H}$ .

$$\begin{aligned}\mathbb{E}(Y1_{A \cap B}) &= \mathbb{E}(X1_A1_B) = \mathbb{E}(X1_A)\mathbb{P}(B) \\ &= \mathbb{E}(Y1_A)\mathbb{P}(B) = \mathbb{E}(Y1_{A \cap B})\end{aligned}$$

Hence  $\mathbb{E}((Z - Y)1_{A \cap B}) = 0$  for all  $A \in \mathcal{F}$ ,  $B \in \mathcal{H}$ . Also,  $(Z - Y)$  is  $\sigma(\mathcal{F}, \mathcal{H})$ -measurable, and therefore  $Y = Z$  by standard argument, e.g. using the following lemma.

**Lemma)** Suppose  $X$  is integrable and  $\mathbb{E}(X1_A) = 0$  for all  $A \in \mathcal{A}$  where  $\mathcal{A}$  is a  $\pi$ -system generating  $\mathcal{F}$ . Then  $X = 0$  a.s.

**proof)** Consider  $\mathcal{D} = \{A \in \mathcal{F} : \mathbb{E}(X1_A) = 0\}$  then  $\mathcal{D}$  is a  $d$ -system, contains  $\mathcal{A}$  so  $\mathcal{D} = \mathcal{F}$  by Dynkin's lemma. Finally,  $\mathbb{E}(X1_{\{X > 0\}}) = 0$  so  $X = 0$  a.s.

We end the chapter with a lemma that is going to be useful.

**Lemma 1.5.1.)** Let  $X \in L^1(\mathbb{P})$ . Set  $\mathcal{Y} = \{\mathbb{E}(X|\mathcal{F}) : \mathcal{F} \subset \mathcal{F} \text{ a sub-}\sigma\text{-algebra}\}$ . Then  $\mathcal{Y}$  is UI(*uniformly integrable*). That is,

$$\sup_{Y \in \mathcal{Y}} \mathbb{E}(|Y|1_{|Y| \geq \lambda}) \rightarrow 0 \quad \text{as } \lambda \rightarrow \infty$$

**proof)** Since  $X$  is integrable, given  $\epsilon > 0$ , there exists  $\delta > 0$  s.t.  $\mathbb{E}(|X|1_A) \leq \epsilon$  whenever  $\mathbb{P}(A) \leq \delta$ . (Suppose not, then we should find  $\epsilon > 0$  and  $A_n \in \mathcal{F}$  s.t.  $\mathbb{P}(A_n) \leq 2^{-n}$  and  $\mathbb{E}(|X|1_{A_n}) > \epsilon$ . Set  $A = \bigcap_n \bigcup_{m \geq n} A_m$ . Then  $\mathbb{P}(A) = 0$  by Borel-Cantelli lemma. But  $\mathbb{E}(|X|1_{\bigcup_{m \geq n} A_m}) \geq \epsilon$ , while this converges to  $\mathbb{E}(|X|1_A) = 0$  as  $n \rightarrow \infty$ , which is a contradiction.)

Choose  $\lambda \in (0, \infty)$  so that  $\mathbb{E}(|X|) \leq \lambda\delta$ . Take  $Y = \mathbb{E}(X|\mathcal{F}) \in \mathcal{Y}$ . Since  $x \mapsto |x|$  is a convex function, by conditional Jensen's inequality,

$$Y \leq \mathbb{E}(|X||\mathcal{F}) \quad \text{so} \quad \mathbb{E}(|Y|) \leq \mathbb{E}(|X|)$$

(In fact, by decomposition  $X = X^+ + X^-$ , we see that  $|Y| = \mathbb{E}(X^+|\mathcal{F}) + \mathbb{E}(X^-|\mathcal{F}) = \mathbb{E}(|X||\mathcal{F})$  a.s.... but does not need this at this point)

By Chebyshev's inequality,  $\mathbb{P}(|Y| \geq \lambda) \leq \lambda^{-1}\mathbb{E}(|Y|) \leq \delta$  hence

$$\begin{aligned}\mathbb{E}(|Y|1_{|Y| \geq \lambda}) &\leq \mathbb{E}(\mathbb{E}(|X||\mathcal{F})1_{|Y| \geq \lambda}) \quad \text{as } 1_{|Y| \geq \lambda} \in \mathcal{F} \\ &= \mathbb{E}(|X|1_{|Y| \geq \lambda}) \leq \epsilon\end{aligned}$$

So  $\sup_{Y \in \mathcal{Y}} \mathbb{E}(|Y|1_{|Y| \geq \lambda}) \leq \epsilon$ . This is true for any  $\epsilon > 0$ , therefore we have the uniform integrability.

(End of proof)  $\square$

(15th October 2018, Monday)

## Chapter 2. Martingales in Discrete Time

### 2.1. Definitions.

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

- A **Filtration** for  $(\Omega, \mathcal{F}, \mathbb{P})$  is a sequence  $(\mathcal{F}_n)_{n \geq 0}$  of  $\sigma$ -algebras s.t. for all  $n \geq 0$ , we have

$$\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$$

Set  $\mathcal{F}_\infty = \sigma(\mathcal{F}_n : n \geq 0)$  then  $\mathcal{F}_\infty \subset \mathcal{F}$ . We allow  $\mathcal{F}_\infty \neq \mathcal{F}$ . We interpret  $n$  as times and  $\mathcal{F}_n$  as the extent of knowledge at time  $n$ .

- A **Random process(in discrete time)** is a sequence of random variables  $(X_n)_{n \geq 0}$ . It has a natural filtration  $(\mathcal{F}_n^X)_{n \geq 0}$  given by

$$\mathcal{F}_n^X = \sigma(X_0, \dots, X_n)$$

That is, the knowledge obtained from  $X_n$  by time  $n$ . We say  $(X_n)_{n \geq 0}$  is **adapted to**  $(\mathcal{F}_n)_{n \geq 0}$  if  $X_n$  is  $\mathcal{F}_n$ -measurable for all  $n \geq 0$ . This is equivalent to having  $\mathcal{F}_n^X \subset \mathcal{F}_n$ , for all  $n \geq 0$ . (Here,  $X_n$  are real-valued)

- We would say  $(X_n)_{n \geq 0}$  is **integrable** if  $X_n$  is integrable for all  $n \geq 0$ .
- A **martingale** is an *adapted, integrable random process*  $(X_n)_{n \geq 0}$  s.t. for all  $n \geq 0$ ,

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n \quad \text{a.s.}$$

In the case  $\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq X_n$  a.s.,  $(X_n)_n$  is called a **super-martingale** and in the case  $\mathbb{E}[X_{n+1} | \mathcal{F}_n] \geq X_n$  a.s.,  $(X_n)_n$  is called a **sub-martingale**.

## Optional Stopping

- A random variable  $T : \Omega \rightarrow \{0, 1, 2, \dots\} \cup \{\infty\}$  is a **stopping time** if  $\{T \leq n\} \in \mathcal{F}_n$  for all  $n \geq 0$ .
- For a stopping time  $T$ , we set  $\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq n\} \in \mathcal{F}_n \text{ for all } n \geq 0\}$ . It is easy to check  $\mathcal{F}_T$  is indeed a  $\sigma$ -algebra and that if  $T(\omega) = n$  for all  $\omega \in \Omega$ , then  $T$  is a stopping time and  $\mathcal{F}_T = \mathcal{F}_n$ .
- Given  $X$ , define  $X_T(\omega) = X_{T(\omega)}(\omega)$  whenever  $T(\omega) < \infty$  and define the **stopped process**  $X^T$  by

$$X_n^T(\omega) = X_{T(\omega) \wedge n}(\omega) \quad \text{for } n \geq 0$$

**Proposition 2.2.1.)** Let  $X$  be an adapted process. Let  $S, T$  be stopping times for  $X$ . Then

- $S \wedge T$  is a stopping time for  $X$ .
- $\mathcal{F}_T$  is a  $\sigma$ -algebra.
- If  $S \leq T$  then  $\mathcal{F}_S \subset \mathcal{F}_T$ .
- $X_T 1_{T < \infty}$  is an  $\mathcal{F}_T$ -measurable random variable.
- $X^T$  is adapted.
- If  $X$  is integrable, then  $X^T$  is also integrable.

**proof)**

- $\{S \wedge T \leq n\} = \{S \leq n\} \cup \{T \leq n\} \in \mathcal{F}_n$  for all  $n \geq 0$ , so  $S \wedge T$  is a stopping times

- (b) Directly from the definition, we see that  $\phi\mathcal{F}_T$ . Also, given  $A \in \mathcal{F}_T$  and a sequence  $(A_m)_m \subset \mathcal{F}_T$ , we have

$$\begin{aligned} A^c \cap \{T \leq n\} &= \{T \leq n\} - A \cap \{T \leq n\} \in \mathcal{F}_n \Rightarrow A^c \in \mathcal{F}_T \\ (\cup_m A_m) \cap \{T \leq n\} &= \cup_m (A_m \cap \{T \leq n\}) \in \mathcal{F}_n \Rightarrow \cup_m A_m \in \mathcal{F}_T \end{aligned}$$

hence  $\mathcal{F}_T$  is a  $\sigma$ -algebra.

- (c) Let  $A \in \mathcal{F}_S$ . Then  $A \cap \{T \leq n\} = A \cap \{S \leq n\} \cap \{T \leq n\} \in \mathcal{F}_n$ , hence  $A \in \mathcal{F}_T$ .  
(d) For each  $t \in \mathbb{R}$ , we have  $\{X_T 1_T > t\} = \cup_m \{X_m > t, T = n\}$  so for any  $n \geq 0$ ,

$$\{X_T 1_T > t\} \cap \{T \leq n\} = \cup_{m=1}^n \{X_m > t, T = n\} \in \mathcal{F}_n$$

and so  $X_T 1_T$  is  $\mathcal{F}_T$ -measurable.

- (e) By definition of being a stopping time, for any  $t \in \mathbb{R}$ ,

$$\{(X^T)_n > t\} = \{T > n, X_n > t\} \cup \left( \cup_{m=0}^n \{T = m, X_m > t\} \right) \in \mathcal{F}_n$$

so  $X^T$  is adapted.

- (f) First consider the case where  $X$  is non-negative integrable. Then

$$\mathbb{E}(X_n^T) = \mathbb{E}(\mathbb{E}(X_n^T | T)) = \sum_{m \geq n} \mathbb{P}(T = m) \mathbb{E}(X_m) + \mathbb{P}(T > n) \mathbb{E}(X_n) < \infty$$

for any  $n$ , so we have the result for non-negative  $X$ .

For the general case, divide  $X$  into a non-negative and a negative part.

(End of proof)  $\square$

**Theorem 2.2.2)** (*Optional stopping theorem*) Let  $X$  be a super-martingale and let  $S, T$  be bounded stopping times with  $S \leq T$  a.s. Then

$$\mathbb{E}[X_T] \leq \mathbb{E}[X_S]$$

**proof)** Fix  $n \geq 0$  such that  $T \leq n$  a.s. Then

$$\begin{aligned} X_T &= X_S + \sum_{S \leq k < T} X_{k+1} - X_k \\ &= X_S + \sum_{k=0}^n (X_{k+1} - X_k) 1_{S \leq k < T} \end{aligned}$$

Now  $\{S \leq k\}$  is in  $\mathcal{F}_k$  and  $\{T > k\}$  is in  $\mathcal{F}_k$ , so

$$\begin{aligned} \mathbb{E}[(X_{k+1} - X_k) 1_{S \leq k < T}] &= \mathbb{E}[\mathbb{E}[(X_{k+1} - X_k) 1_{S \leq k < T} | \mathcal{F}_k]] \\ &= \mathbb{E}[\mathbb{E}[X_{k+1} - X_k | \mathcal{F}_k] 1_{S \leq k < T}] \end{aligned}$$

but since  $(X_n)$  was a super-martingale,  $\mathbb{E}[X_{k+1} - X_k | \mathcal{F}_k] \leq 0$  a.s. and therefore  $\mathbb{E}[(X_{k+1} - X_k) 1_{S \leq k < T}] \leq 0$  a.s. Hence  $\mathbb{E}(X_T) \leq \mathbb{E}(X_S)$ .

(End of proof)  $\square$

★ Note that  $X$  is a sub-martingale *if and only if*  $(-X)$  is a super-martingale, and that  $X$  is a martingale *if and only if*  $X$  and  $(-X)$  are super-martingales. Hence, we obtain sub-martingale and martingale versions of the theorem :

$$\begin{aligned} \text{If } (X_n) \text{ is a sub-martingale, } \mathbb{E}[X_T] &\geq \mathbb{E}[X_S] \\ \text{If } (X_n) \text{ is a martingale, } \mathbb{E}[X_T] &= \mathbb{E}[X_S] \end{aligned}$$

**Theorem 2.2.3.)** Let  $X$  be an adapted integrable process. Then the followings are equivalent.

- (a)  $X$  is a super-martingale.
- (b) for all bounded stopping times  $T$  and stopping time  $S$ ,

$$\mathbb{E}(X_T | \mathcal{F}_S) \leq X_{S \wedge T} \quad \text{a.s.},$$

- (c) for all stopping times  $T$ , the stopped process  $X^T$  is a super-martingale,
- (d) for all bounded stopping times  $T$  and all stopping times  $S$  with  $S \leq T$  a.s,

$$\mathbb{E}(X_T) \leq \mathbb{E}(X_S)$$

★ The theorem gives an inverse statement of the optional stopping theorem.

**proof)**

- (a)  $\Rightarrow$  (b) Suppose  $X$  is a super-martingale and  $S, T$  are stopping times. Let  $T \leq n$ , for some  $n < \infty$ . Then

$$X_T = X_{S \wedge T} + \sum_{k=0}^T (X_{k+1} - X_k) 1_{S \leq k < T} \cdots \cdots (*)$$

Let  $A \in \mathcal{F}_S$ . Then  $A \cap \{S \leq k\} \in \mathcal{F}_k$  and  $\{T > k\} \in \mathcal{F}_k$  so

$$\mathbb{E}[(X_{k+1} - X_k) 1_{S \leq k < T} 1_A] = \mathbb{E}[\mathbb{E}[(X_{k+1} - X_k) 1_{S \leq k < T} 1_A | \mathcal{F}_k]] \leq 0$$

and

$$\begin{aligned} \mathbb{E}[(X_T - X_{S \wedge T}) 1_A] &= \mathbb{E}\left[\sum_{n=0}^T (X_{k+1} - X_k) 1_{S \leq k < T} 1_A\right] \leq 0 \\ \Rightarrow \mathbb{E}[X_T 1_A] &\leq \mathbb{E}[X_{S \wedge T} 1_A] \end{aligned}$$

But since this inequality is true for any  $A \in \mathcal{F}_S$  and noting that  $X_{S \wedge T} \in \mathcal{F}_S$ , we see

$$\mathbb{E}[X_T | \mathcal{F}_S] \leq X_{S \wedge T} \quad \text{a.s.}$$

The implications (b) $\Rightarrow$ (c) and (c) $\Rightarrow$ (d) are obvious.

(d)  $\Rightarrow$  (a) Let  $m \leq n$  and  $A \in \mathcal{F}_n$ . Set  $T = m1_A + n1_{A^c}$ . Then  $T$  is a stopping time with  $T \leq n$ . Then

$$\mathbb{E}(X_n 1_A - X_m 1_A) = \mathbb{E}(X_n) - \mathbb{E}(X_T) \leq 0$$

(note, if  $\omega \in A$  then  $(X_n 1_A - X_m 1_A)(\omega) = X_n(\omega) - X_m(\omega)$  and 0 otherwise) so

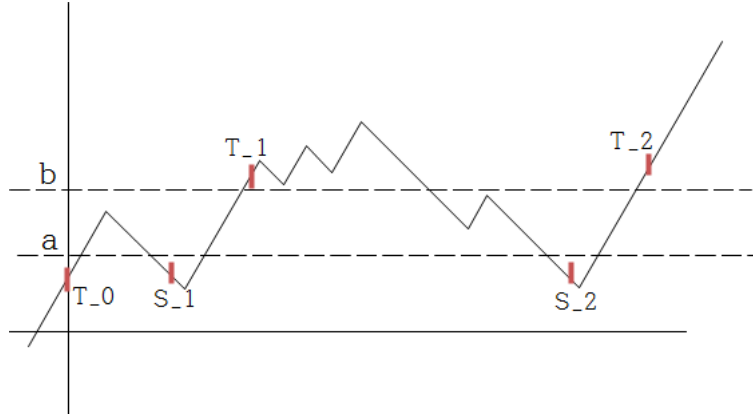
$$\mathbb{E}[X_n | \mathcal{F}_m] \leq X_m$$

(End of proof)  $\square$

(17th October, Wednesday)

### 2.3. Doob's upcrossing inequality

- Let  $X$  be a random process and let  $a, b \in \mathbb{R}$  s.t.  $a < b$ . Fix  $\omega \in \Omega$ . By an **upcrossing** of  $[a, b]$  by  $X(\omega)$ , we mean an interval of times  $\{j, j+1, \dots, k\}$  s.t.  $X_j(\omega) < a$ ,  $X_k(\omega) > b$ .
- Write  $U_n[a, b](\omega)$  for the number of disjoint upcrossings contained in  $\{0, 1, \dots, n\}$ , and  $U_n[a, b] \nearrow U[a, b]$  as  $n \rightarrow \infty$ .



**Theorem 2.3.1.)** (Doob's upcrossing inequality) Let  $X$  be a *super-martingale*. Then

$$(b - a)\mathbb{E}[U[a, b]] \leq \sup_{n \geq 0} \mathbb{E}[(X_n - a)^-]$$

(Recall,  $x^- = (-x) \vee 0$ )

In fact, in this theorem, we prove  $(b - a)\mathbb{E}[U_n[a, b]] \leq \mathbb{E}[(X_n - a)^-]$ .

**proof)** Set  $T_0 = 0$  and define recursively for  $k \geq 0$ ,

$$S_{k+1} = \inf\{m \geq T_k : X_m < a\}, \quad T_{k+1} = \sup\{m \geq S_{k+1} : X_m > b\}$$

Note that if  $T_k < \infty$ , then  $\{S_k, S_k + 1, \dots, T_k\}$  is an upcrossing of  $[a, b]$  by  $X$ , and  $T_k$  is the time of completion of the  $k$ -th upcrossing. Also note that  $U_n[a, b] \leq n$ . For  $m \leq n$ , we have

$$\{U_n[a, b] = m\} = \{T_m \leq n < T_{m+1}\}$$

On this event,

$$X_{T_k \wedge n} - X_{S_k \wedge n} = \begin{cases} X_{T_k} - X_{S_k} \geq b - a & \text{if } k \leq m \\ X_n - X_{S_k} \geq X_n - a & \text{if } k \geq m+1, S_{m+1} \leq n \\ 0 & \text{otherwise} \end{cases}$$

Hence

$$\begin{aligned} \sum_{k=1}^n (X_{T_k \wedge n} - X_{S_k \wedge n}) &\geq (b-a)U_n[a, b] + X_n - a \\ &\geq (b-a)U_n[a, b] - (X_n - a)^- \end{aligned}$$

Since  $X$  is a super-martingale and  $T_k \wedge n$  and  $S_k \wedge n$  are *bounded stopping times* with  $S_k \leq T_k$ , by optional stopping theorem, we have

$$\mathbb{E}(X_{T_k \wedge n}) \leq \mathbb{E}(X_{S_k \wedge n})$$

By  $\mathbb{E}(\sum_{k=1}^n (X_{T_k \wedge n} - X_{S_k \wedge n})) \leq 0$  we get

$$(b-a)\mathbb{E}(U_n[a, b]) \leq \mathbb{E}[(X_n - a)^-]$$

Apply monotone convergence, with  $n \rightarrow \infty$ , then we are done.

(End of proof)  $\square$

This theorem does not seem to have any significance at the moment, but it will turn out to be important later on.

## 2.4. Doob's maximal inequalities.

Define  $X_n^* = \sum_{k \geq n} |X_k|$

In the next two theorems, we see that the martingale(or sub-martingale) property allows us to obtain estimates on this  $X_n^*$  in terms of expectations for  $X_n$ .

**Theorem 2.4.1** (Doob's maximal inequality) Let  $X$  be a *martingale* or a *non-negative sub-martingale*. Then for all  $\lambda \geq 0$ ,

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

**proof** If  $X$  is a martingale, then  $|X|$  is a non-negative sub-martingale. It suffices to consider the case where  $X$  is a non-negative sub-martingale.

Set  $T = \inf\{k \geq 0 : X_k \geq \lambda\} \wedge n$ . Then  $T$  is a stopping time and  $T \leq n$ , so by optional stopping, has

$$\begin{aligned} \mathbb{E}(X_n) &\geq \mathbb{E}(X_T) = \mathbb{E}(X_T 1_{X_n^* \geq \lambda}) + \mathbb{E}(X_T 1_{X_n^* < \lambda}) \\ &= \mathbb{E}(\lambda 1_{X_n^* \geq \lambda}) + \mathbb{E}(X_n 1_{X_n^* < \lambda}) \end{aligned}$$

and

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

(End of proof)  $\square$

**Theorem 2.4.2)** (Doob's  $L^p$ -inequality) Let  $X$  be a *martingale* or a *non-negative sub-martingale*. Then, for all  $p > 1$  and  $q = p/(p-1)$ , we have

$$\|X_n^*\|_p \leq q \|X_n\|_q$$

**proof)** Again, it suffices to consider when  $X$  is a non-negative sub-martingale. Fix  $k < \infty$ . Then

$$\begin{aligned} \mathbb{E}[(X_n^* \wedge k)^p] &= \mathbb{E} \int_0^k p \lambda^{p-1} 1_{\{X_n^* \geq \lambda\}} d\lambda \quad (\text{integration by parts}) \\ &= \int_0^k p \lambda^{p-1} \mathbb{P}(X_n^* \geq \lambda) d\lambda \quad (\text{Fubini}) \\ &\leq \int_0^k p \lambda^{p-2} \mathbb{E}(X_n 1_{X_n^* \geq \lambda}) d\lambda \quad (\text{Doob's maximal inequality}) \\ &= \frac{p}{p-1} \mathbb{E}(X_n (X_n^* \wedge k)^{p-1}) \\ &\leq q \|X_n\|_p \|X_n^* \wedge k\|_p^{p-1} \quad (\text{Hölder's inequality}) \end{aligned}$$

Hence,  $\|X_n^* \wedge k\|_p \leq q \|X_n\|_p$ . Apply monotone convergence theorem with  $k \rightarrow \infty$ , then we have the desired result.

(End of proof)  $\square$

Doob's maximal and  $L^p$  inequalities have different versions which apply under the same hypothesis to

$$X^* = \sum_{n \geq 0} |X_n|$$

since  $X_n^* \nearrow X^*$ . Letting  $n \rightarrow \infty$  in Doob's maximal inequality gives

$$\lambda \mathbb{P}(X^* \geq \lambda) \lim_{n \rightarrow \infty} \lambda \mathbb{P}(X_n^* \geq \lambda) \leq \sup_{n \geq 0} \mathbb{E}(|X_n|)$$

We can then replace  $\lambda \mathbb{P}(X^* > \lambda)$  by  $\lambda \mathbb{P}(X^* \geq \lambda)$  by taking limits from the right in  $\lambda$ .

Similarly, for  $p \in (1, \infty)$  by monotone convergence,

$$\|X^*\|_p \leq q \sup_{n \geq 0} \|X_n\|_p$$

(19th October, Friday)

## 2.5. Doob's martingale convergence theorems

We are going to study three different martingale convergence theorems. They are all important.

- We say that a random process  $X$  is  $L^p$ -**bounded** if  $\sum_{n \geq 0} \|X_n\|_p < \infty$ .



- We say that  $X$  is **uniformly integrable** if

$$\sup_{n \geq 0} \mathbb{E}(|X_n| 1_{|X_n| > \lambda}) \rightarrow 0 \quad \text{as } \lambda \rightarrow \infty$$

- If  $X$  is  $L^p$  bounded for some  $p > 1$ , then this implies that  $X$  is uniformly integrable. This again implies that  $X$  is  $L^1$  bounded. The first implication follows from Hölder inequality. The second implication is true because  $\mathbb{E}(|X_n|) = \mathbb{E}(|X_n| 1_{|X_n| \leq \lambda}) + \mathbb{E}(|X_n| 1_{|X_n| > \lambda}) \leq \lambda + \mathbb{E}(|X_n| 1_{|X_n| > \lambda})$ .

$X$  is  $L^p$ -bounded,  $p > 1$

$\Downarrow$

$X$  is Uniformly integrable

$\Downarrow$

$X$  is  $L^1$ -bounded

**Theorem 2.5.1)** (*Almost sure martingale convergence theorem*) Let  $X$  be an  $L^1$ -bounded super-martingale. Then there exists an integrable and  $\mathcal{F}_\infty$ -measurable random variable  $X_\infty$  such that

$$X_n \rightarrow X \quad \text{a.s. as } n \rightarrow \infty$$

**proof)** For a sequence of real numbers  $(x_n)_{n \geq 0}$ , as  $n \rightarrow \infty$ ,  $(x_n)_n$  either converges or  $|x_n| \rightarrow \infty$ , or  $\liminf_n x_n < \limsup_n x_n$ . In the last case, since the rationals are dense in  $\mathbb{R}$ , there exist  $a, b \in \mathbb{Q}$  such that  $\liminf x_n < a < b < \limsup x_n$ .

Set  $\Omega_0 = \Omega_\infty \cap (\bigcap_{a,b \in \mathbb{Q}, a < b} \Omega_{a,b})$  where  $\Omega_\infty = \{\liminf |X_n| < \infty\}$ ,  $\Omega_{a,b} = \{U[a, b] < \infty\}$  (Recall that  $U[a, b]$  is the number of upcrossings). Then  $X_n(\omega)$  converges for all  $\omega \in \Omega_0$ . By Fatous' lemma,

$$\mathbb{E}(\liminf |X_n|) \leq \liminf \mathbb{E}|X_n| < \infty$$

so this implies  $\mathbb{P}(\Omega_\infty) = 1$ . By Doob's inequality, for  $a < b$ , has

$$(b - a)\mathbb{E}(U[a, b]) \leq |a| + \sup_{n \geq 0} \mathbb{E}|X_n| < \infty$$

and therefore  $\mathbb{P}(\Omega_{a,b}) = 1$ . Putting this together, we deduce that  $\mathbb{P}(\Omega_0) = 1$ , and we can find a random variable  $X_\infty$  defined by

$$X_\infty = \lim_{n \rightarrow \infty} X_n 1_{\Omega_0}$$

Then  $X_n \rightarrow X_\infty$  a.s. Also  $X_\infty$  is  $\mathcal{F}_\infty$ -measurable and  $|X_\infty| \leq \liminf |X_n|$  so  $\mathbb{E}(|X_\infty|) < \infty$ . Hence  $X_\infty$  is integrable.

(End of proof)  $\square$

**Remark :** Every non-negative integrable super-martingale is  $L^1$ -bounded, hence it converges a.s.

**Theorem 2.5.2)** ( *$L^1$  martingale convergence theorem*) Let  $(X_n)_{n \geq 0}$  be a uniformly integrable martingale. Then there exists a random variable  $X_\infty \in L^1(\mathcal{F}_\infty)$  such that

$$X_n \xrightarrow{n \rightarrow \infty} X_\infty \quad \text{a.s. and in } L^1$$

Moreover,  $X_n = \mathbb{E}(X_\infty | \mathcal{F}_n)$  a.s. for all  $n \geq 0$ .

Conversely, for all  $Y \in L^1(\mathcal{F}_\infty)$ , on choosing version  $X_n$  of  $\mathbb{E}(Y | \mathcal{F}_n)$  for all  $n$ , we obtain a uniformly integrable martingale  $(X_n)_{n \geq 0}$  such that

$$X_n \xrightarrow{n \rightarrow \infty} Y \quad \text{a.s. and in } L^1$$

We can think of this theorem as establishing the bijection

$$\text{unif. integrable martingale/a.s.} \leftrightarrow L^1(\mathcal{F}_\infty)$$

**proof)** Let  $(X_n)_{n \geq 0}$  be a uniformly integrable martingale. By the almost sure martingale convergence theorem, there exists  $X_\infty \in L^1(\mathcal{F}_\infty)$  s.t.  $X_n \rightarrow X_\infty$  a.s. Since  $X$  is uniformly integrable, it also follows that  $X_n \rightarrow X_\infty$  in  $L^1$ . (see PM, Thm 2.5.1. and 6.2.3.)

Next, for  $m \geq n$ ,

$$\begin{aligned} \|X_n - \mathbb{E}(X_\infty | \mathcal{F}_n)\|_1 &= \|\mathbb{E}(X_m - X_\infty | \mathcal{F}_n)\|_1 \\ &= \|X_m - X_\infty\|_1 \rightarrow 0 \quad \text{as } m \rightarrow \infty \end{aligned}$$

Hence  $X_n = \mathbb{E}(X_\infty | \mathcal{F}_n)$  a.s.

For the converse statement, suppose  $Y \in L^1(\mathcal{F}_\infty)$  and let  $X_n$  be a version of  $\mathbb{E}(Y | \mathcal{F}_n)$  for all  $n$ . Then  $(X_n)_{n \geq 0}$  is a martingale by the tower property, and is uniformly integrable by **Lemma 1.5.1**. Hence there exists  $X_\infty \in L^1(\mathcal{F}_\infty)$  such that  $X_n \rightarrow X_\infty$  a.s. and in  $L^1$ . For all  $n \geq 0$  and all  $A \in \mathcal{F}_n$ , we have

$$\mathbb{E}(X_\infty 1_A) = \lim_{m \rightarrow \infty} \mathbb{E}(X_m 1_A) = \lim_{n \leq m \rightarrow \infty} \mathbb{E}(\mathbb{E}(Y 1_A | \mathcal{F}_m)) = \mathbb{E}(Y 1_A)$$

where the second equality follows because  $\mathbb{E}(X_m | \mathcal{F}_n) = \mathbb{E}(Y | \mathcal{F}_n)$ . Now  $X_\infty, Y \in L^1(\mathcal{F}_\infty)$  and  $\cup_n \mathcal{F}_n$  is a  $\pi$ -system generating  $\mathcal{F}_\infty$ . Hence, by Dynkin's lemma,

$$X_\infty = Y \quad \text{a.s.}$$

(End of proof)  $\square$

**Theorem 2.5.3)** ( *$L^p$ -martingale convergence theorem*) Let  $p \in (1, \infty)$ . Let  $(X_n)_{n \geq 0}$  be an  $L^p$ -bounded martingale. Then there exists a random variable  $X_\infty \in L^p(\mathcal{F}_\infty)$  s.t.

$$X_n \rightarrow X_\infty \quad \text{a.s. and in } L^p$$

Moreover,  $X_n = \mathbb{E}(X_\infty | \mathcal{F}_n)$  a.s. for all  $n \geq 0$ .

Conversely, for all  $Y \in L^p(\mathcal{F}_\infty)$ , on choosing a version  $X_n$  of  $\mathbb{E}(Y | \mathcal{F}_n)$  for all  $n$ , we obtain an  $L^p$ -bounded martingale such that  $X_n \rightarrow Y$  a.s. and in  $L^p$ .

This is very similar to the statement of  $L^1$ -martingale convergence theorem. Indeed, the proof is also very similar.

**proof)** Let  $(X_n)$  be an  $L^p$ -bounded martingale. By *a.s. martingale convergence theorem*, there exists  $X_\infty \in L^1(\mathcal{F}_\infty)$ ,  $X_n \rightarrow X_\infty$  a.s.

By *Doob's  $L^p$ -inequality*,  $\|X^*\|_p \leq q \sup_{n \geq 0} \|X_n\|_p < \infty$ , where  $X^* = \sup_{n \geq 0} |X_n|$ . Also, since  $|X_n - X_\infty|^p \leq (2X^*)^p$  for all  $n$ , we may apply dominated convergence theorem to deduce that  $X_n \rightarrow X_\infty$  in  $L^p$ . Then  $X_n = \mathbb{E}(X_\infty | \mathcal{F}_n)$  a.s. for all  $n$ , as in the  $L^1$ -convergence.

For the converse statement, suppose  $Y \in L^p(\mathcal{F}_\infty)$  and let  $X_n$  be a version of  $\mathbb{E}(Y | \mathcal{F}_n)$ . Then  $(X_n)_{n \geq 0}$  is a martingale by the tower property and by Jensen inequality,

$$\|X_n\|_p = \|\mathbb{E}(Y | \mathcal{F}_n)\|_p \leq \|Y\|_p$$

Let  $X_n \rightarrow X_\infty$  a.s. and in  $L^p$  for  $X_\infty \in L^p(\mathcal{F}_\infty)$ , using the previous part. Then proceed as in the proof of  $L^1$ -convergence to prove that in fact  $Y = X_\infty$  a.s.

(End of proof)  $\square$

(22nd October, Monday)

Recall that, for a stopping time  $T$  and a random process  $X$ ,  $X_T$  has been defined only on  $\{T < \infty\}$ . Given an almost sure limit  $X_\infty$  for  $X$ , we define  $X_T = X_\infty$  on  $\{T = \infty\}$ . Then the optional stopping theorem extends to all stopping times for uniformly integrable martingales.

**Theorem 2.5.5.)** Let  $X$  be a uniformly integrable martingale and let  $T$  be any stopping time. Then  $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ . Moreover, for all stopping time  $S$  and  $T$ , we have

$$\mathbb{E}(X_T | \mathcal{F}_S) = X_{S \wedge T} \quad \text{a.s.}$$

This theorem is an extension of Optional stopping theorem, **Theorem 2.2.2** and **Theorem 2.2.3**.

**proof)** By the optional stopping time theorem and **2.2.3**, when applied to the bounded stopping time  $T \wedge n$ , we have

$$\begin{aligned} \mathbb{E}(X_{T \wedge n}) &= \mathbb{E}(X_0) \\ \mathbb{E}(X_{T \wedge n} | \mathcal{F}_S) &= X_{S \wedge T \wedge n} \end{aligned}$$

In order to get the claim by letting  $n \rightarrow \infty$ , we need to prove  $X_{T \wedge n} \rightarrow X_T$  a.s. and in  $L^1$ . This will imply that

$$\mathbb{E}(X_{T \wedge n} | \mathcal{F}_S) \rightarrow \mathbb{E}(X_T | \mathcal{F}_S) \quad \text{in } L^1$$

**Claim :**  $X_{T \wedge n} \rightarrow X_T$  a.s. and in  $L^1$

**proof)** By the  $L^1$  martingale convergence theorem, there exists  $X_\infty \in L^1(\mathcal{F}_\infty)$  s.t.  $X_n \rightarrow X_\infty$  a.s. and in  $L^1$  and  $X_n = \mathbb{E}(X_\infty | \mathcal{F}_n)$ . This implies  $X_{T \wedge n} \rightarrow X_T$  a.s. as  $n \rightarrow \infty$ . (if  $T < \infty$ , the convergence trivial, and in the case  $T = \infty$ , the convergence justified the previous statement). Since  $F_{T \wedge n} \subset F_n$ , by **Theorem 2.2.3**. and the tower property we have

$$X_{T \wedge n} = \mathbb{E}(X_n | \mathcal{F}_{T \wedge n}) = \mathbb{E}(X_\infty | \mathcal{F}_{T \wedge n})$$

By **Lemma 1.5.1**,  $(X_{T \wedge n})_{n \geq 0}$  is uniformly integrable. Hence

$$X_{T \wedge n} \rightarrow X_T \quad \text{in } L^1$$

(End of proof)  $\square$

### Backward martingale

- A **backward filtration**  $(\hat{\mathcal{F}}_n)_{n \geq 0}$  is a sequence of  $\sigma$ -algebras such that  $\mathcal{F} \supset \hat{\mathcal{F}}_n \supset \hat{\mathcal{F}}_{n+1}$ .
- This also defines  $\hat{\mathcal{F}}_\infty = \bigcap_{n \geq 0} \hat{\mathcal{F}}_n$

**Theorem 2.5.4.)** (*Backward martingale convergence theorem*) For all  $Y \in L^1(\mathcal{F})$ , we have

$$\mathbb{E}(Y|\hat{\mathcal{F}}_n) \rightarrow \mathbb{E}(Y|\hat{\mathcal{F}}_\infty) \quad \text{a.s. and in } L^1 \quad \text{as } n \rightarrow \infty$$

Note that we do not need a uniformly integrability condition, because our assumption of backward filtration already implies uniform convergences.

**proof)** Write  $X_n = \mathbb{E}(Y|\hat{\mathcal{F}}_n)$  for all  $n \geq 0$ . Fix  $n \geq 0$ , by the Tower property,  $(X_{n-k})_{0 \leq k \leq n}$  is a martingale for the filtration  $(\hat{\mathcal{F}}_{n-k})_{0 \leq k \leq n}$ . For  $a < b$ , the number  $U_n[0, \infty]$  of upcrossings of  $[a, b]$  by  $(X_k)_{0 \leq k \leq n}$  equals the number of upcrossings of  $[-b, -a]$  by the process  $(-X_{n-k})_{0 \leq k \leq n}$ . Hence by (the note on) **Theorem 2.3.1**,

$$(b - a)\mathbb{E}(U_n[a, b]) \leq \mathbb{E}((X_0 - b)^+)$$

and so by monotone convergence,

$$(b - a)\mathbb{E}(U[a, b]) \leq \mathbb{E}((X_0 - b)^+) \leq \mathbb{E}(|X|) + |b| \leq \mathbb{E}(|Y|) + |b| < \infty$$

where the third inequality follows because of Jensen's inequality. Also,

$$\mathbb{E}(\liminf |X_n|) \leq \liminf \mathbb{E}|X_n| \leq \mathbb{E}|Y| < \infty$$

With these properties in hand, we can apply the same proof used to prove almost sure martingale convergence theorem to show that  $\mathbb{P}(\hat{\Omega}_0) = 1$ , where  $\hat{\Omega}_0 = \{X_n \text{ converges as } n \rightarrow \infty\}$  - observe that  $\hat{\Omega}_0 = \{\liminf_n |X_n| < \infty\} \cap (\bigcap_{a,b \in \mathbb{Q}, a < b} \{U[a, b] < \infty\})$  and we see that each set in the intersection has measure 1, and therefore  $\mathbb{P}(\hat{\Omega}_0) = 1$ .

Set  $X_\infty = 1_{\hat{\Omega}_0} \lim_{n \rightarrow \infty} X_n$ . Then  $X_\infty \in L^1(\hat{\mathcal{F}}_\infty)$  and  $X_n \rightarrow X_\infty$  a.s. Now  $(X_n)_{n \geq 0}$  is uniformly integrable (by **Lemma 1.5.1**), so  $X_n \xrightarrow{L^1} X_\infty$ . Finally, for all  $A \in \hat{\mathcal{F}}_\infty$ , we have

$$\mathbb{E}((X_\infty - \mathbb{E}(Y|\hat{\mathcal{F}}_\infty))1_A) = \lim_{n \rightarrow \infty} \mathbb{E}((X_n - Y)1_A) = 0$$

This implies  $X_\infty = \mathbb{E}(Y|\hat{\mathcal{F}}_\infty)$  a.s.

(End of proof)  $\square$

### 3. Applications of martingale theory

#### Sums of independent random variables

Let  $S_n = X_1 + \cdots + X_n$ , where  $(X_n)_{n \geq 0}$  is a sequence of independent random variables.

**Theorem 3.1.1** (*Strong Law of Large Numbers*) Let  $(X_n)_{n \geq 0}$  be a sequence of independent identically distributed (*i.i.d*) integrable random variables. Set  $\mu = \mathbb{E}(X_1)$ . Then

$$S_n/n \rightarrow \mu \quad \text{a.s. and in } L^1$$

**proof** Define  $\hat{\mathcal{F}}_n = \sigma(S_m : m \geq n)$ ,  $\mathcal{T}_n = \sigma(X_m : m \geq n+1)$  and  $\mathcal{T} = \cap_{n \geq 1} \mathcal{T}_n$ . Then  $\hat{\mathcal{F}}_n = \sigma(S_n, \mathcal{T}_n)$  and  $(\hat{\mathcal{F}}_n)_{n \geq 1}$  is a backward filtration. Since  $\sigma(X_1, S_n)$  is independent of  $\mathcal{T}_n$ , we have

$$\mathbb{E}(X_1 | \hat{\mathcal{F}}_n) = \mathbb{E}(X_1 | S_n) \quad \text{a.s.}$$

For  $k \leq n$  and all Borel sets  $B$ , we have

$$\mathbb{E}(X_k 1_{\{S_n \in B\}}) = \mathbb{E}(X_1 1_{\{S_n \in B\}})$$

by symmetry  $(X_k, S_n) \stackrel{d}{=} (X_1, S_n)$  in distribution, so  $\mathbb{E}(X_k | S_n) = \mathbb{E}(X_1 | S_n)$  a.s. But

$$\mathbb{E}(X_1 | S_n) + \cdots + \mathbb{E}(X_n | S_n) = \mathbb{E}(S_n | S_n) = S_n \quad \text{a.s.}$$

so  $\mathbb{E}(X_1 | \hat{\mathcal{F}}_n) = S_n/n$  almost surely. Then by backward martingale convergence theorem, has  $S_n/n \rightarrow Y$  a.s. and in  $L^1$  for some random variable  $Y$ . Then  $Y \in \mathcal{T}$ . By Kolmogorov's 0-1 law [PM **Theorem 2.6.1**],  $Y$  is almost surely a constant. Hence

$$Y = \mathbb{E}(Y) = \lim \mathbb{E}(S_n/n) = \mu \quad \text{a.s.}$$

where the second equality follows from  $L^1$  convergence  $S_n/n \rightarrow Y$ .

(End of proof)  $\square$

Since a.s. convergence implies convergence in probability, we have the following corollary.

**Corollary 3.1.2** (*Weak law of large numbers*) Let  $(X_n)_{n \geq 1}$  be a sequence of i.i.d. integrable r.v.. Set  $\mu = \mathbb{E}(X_1)$ . Then

$$\mathbb{P}(|\frac{S_n}{n} - \mu| > \epsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad \forall \epsilon > 0$$

(24th October, Wednesday)

#### 3.2. Non-negative martingale and change of measure

- Given a random variable  $X$ ,  $\mathcal{F}$ -measurable with  $X \geq 0$  and  $\mathbb{E}(X) = 1$ , we can define a new probability measure for  $\tilde{\mathbb{P}}$  on  $\mathcal{F}$  by

$$\tilde{\mathbb{P}}(A) = \mathbb{E}(X 1_A) \quad \forall A \in \mathcal{F}$$

Moreover, by [PM, Prop 3.1.4], given  $\tilde{\mathbb{P}}$ , this equation determines  $X$  uniquely, up to a.s. modification. We say  $\tilde{\mathbb{P}}$  **has a density w.r.t.**  $\mathbb{P}$  and  $X$  is a version of the density.

- Let  $(\mathcal{F}_n)_{n \geq 0}$  be a filtration in  $\mathcal{F}$  and assume  $\mathcal{F} = \mathcal{F}_\infty$ . Let  $(X_n)_{n \geq 0}$  be an adapted random process, with  $X_n \geq 0$  and  $\mathbb{E}(X_n) = 1$  for all  $n$ . We can define, for each  $n$ , a new probability measure  $\tilde{\mathbb{P}}_n$  on  $\mathcal{F}_n$  by

$$\tilde{\mathbb{P}}_n(A) = \mathbb{E}(X_n 1_A) \quad \forall A \in \mathcal{F}_n$$

Since we require each  $X_n$  to be  $\mathcal{F}_n$ -measurable, this equation determines  $X_n$  uniquely up to a.s. modification.

**Proposition 3.2.1.)** The measures  $\tilde{\mathbb{P}}_n$  are consistent. That is

$$\tilde{\mathbb{P}}_{n+1}|_{\mathcal{F}_n} = \tilde{\mathbb{P}}_n \quad \forall n \quad \text{iff} \quad (X_n)_{n \geq 0} \quad \text{is a martingale}$$

Moreover, there is a measure  $\tilde{\mathbb{P}}$  on  $\mathcal{F}$ , which has a density w.r.t  $\mathbb{P}$  such that

$$\tilde{\mathbb{P}}|_{\mathcal{F}_n} = \tilde{\mathbb{P}}_n \quad \forall n \quad \text{iff} \quad (X_n)_n \quad \text{is a uniformly integrable martingale}$$

**proof)** (The proof was an exercise.) For the first point,

$$\begin{aligned} \tilde{\mathbb{P}}_n(A) &= \tilde{\mathbb{P}}_{n+1}(A|\mathcal{F}_n) = \mathbb{E}(X_{n+1} 1_A | \mathcal{F}_n) = \mathbb{E}(X_{n+1} | \mathcal{F}_n) 1_A \quad \forall A \in \mathcal{F}_n \\ &\Leftrightarrow \mathbb{E}(X_{n+1} | \mathcal{F}_n) = X_n \quad \text{a.s.} \quad \Leftrightarrow (X_n) \text{ is a martingale} \end{aligned}$$

For the second point, suppose  $\tilde{\mathbb{P}}|_{\mathcal{F}_n} = \tilde{\mathbb{P}}_n \quad \forall n$ . Then  $\tilde{\mathbb{P}}_n|_{\mathcal{F}_m} = (\tilde{\mathbb{P}}|_{\mathcal{F}_n})|_{\mathcal{F}_m} = \tilde{\mathbb{P}}_m$  whenever  $n \geq m$ , so we find that  $(X_n)_n$  is a martingale. Since we assumed that  $\mathbb{E}(X_n) = 1$  for all  $n$ , by almost everywhere martingale convergence theorem, we find a random variable  $X$  such that  $X_n \rightarrow X$  a.s. Now for any  $A \in \mathcal{F}$ , we may find  $N \geq 0$  such that  $A \in \mathcal{F}_k$  for all  $k \geq N$ , so

$$\tilde{\mathbb{P}}(A) = \tilde{\mathbb{P}}(A|\mathcal{F}_k) = \tilde{\mathbb{P}}_k(A) = \mathbb{E}(X_k 1_A) \xrightarrow{k \rightarrow \infty} \mathbb{E}(X 1_A)$$

and therefore  $\tilde{\mathbb{P}}(A) = \mathbb{E}(X 1_A)$ . Hence for all  $n \geq 0$ , we have

$$\mathbb{E}(X 1_A | \mathcal{F}_n) = \tilde{\mathbb{P}}(A|\mathcal{F}_n) = \mathbb{E}(X_n 1_A) \quad \forall A \in \mathcal{F}_n$$

and therefore  $X_n = \mathbb{E}(X | \mathcal{F}_n)$ . This shows that  $(X_n)_n$  is uniformly integrable martingale.

For the converse direction, assume that  $(X_n)_n$  is uniformly integrable. Then by  $L^1$ -martingale convergence theorem, we may find  $X \in \mathcal{F}_\infty$  such that  $X_n \rightarrow X$  in  $L^1$  and a.s. Define  $\tilde{\mathbb{P}}(A) = \mathbb{E}(X 1_A)$ . Then  $\tilde{\mathbb{P}}(A|\mathcal{F}_n) = \mathbb{E}(\mathbb{E}(X 1_A | \mathcal{F}_n)) = \mathbb{E}(X_n 1_A)$  for any  $n \geq 0$  and  $A \in \mathcal{F}_n$ , and therefore  $\tilde{\mathbb{P}}|_{\mathcal{F}_n} = \tilde{\mathbb{P}}_n$ .

(End of proof)  $\square$

**Theorem 3.2.3)** (Radon-Nikodym theorem) Let  $\mu$  and  $\nu$  be  $\sigma$ -finite measures on a measurable space  $(E, \mathcal{E})$ . Then the followings are equivalent :

- $\nu(A) = 0$  for all  $A \in \mathcal{E}$  such that  $\mu(A) = 0$ , i.e.  $\nu$  is **absolutely continuous** with respect to  $\mu$ .
- There exists a measurable function  $f$  on  $E$  such that  $f \geq 0$  and  $\nu(A) = \mu(f 1_A)$  for all  $A \in \mathcal{E}$ .

The function  $f$  which is unique up to modification  $\mu$ -a.e. is called (a version of) the **Radon-Nikodym derivative** of  $\nu$  with respect to  $\mu$ . We write  $f = d\nu/d\mu$  almost surely.

We will give a proof for the case *where  $\mathcal{E}$  is countably generated*. We assume there is a sequence  $(G_n : n \in \mathbb{N})$  of subsets of  $E$  which generates  $\mathcal{E}$ . This holds, for example, whenever  $\mathcal{E}$  is the Borel  $\sigma$ -algebra of a topology with countable basis. A further martingale argument is required to prove the general case, but we omit it.

**proof** The direction (b)  $\Rightarrow$  (a) is obvious. So we aim to prove (a)  $\Rightarrow$  (b)

By assumption, there is a countable partition of  $E$  by measurable sets on which both  $\mu$  and  $\nu$  are finite. (since  $\mu, \nu$  are  $\sigma$ -finite.) It suffices to show (b) holds on each of these sets, so we can reduce to the case where  $\mu, \nu$  are finite.

The case  $\nu(E) = 0$  is clear, as we can just take  $f \equiv 0$ . So assume  $\nu(E) > 0$ . Then  $\mu(E) > 0$  by (a). Write  $\Omega = E$  and  $\mathcal{F} = \mathcal{E}$  and consider the probability measures

$$\mathbb{P} = \mu/\mu(E) \quad \text{and} \quad \tilde{\mathbb{P}} = \nu/\nu(E) \quad \text{on } (\Omega, \mathcal{F})$$

It will suffice to show that there is a random variable  $X \geq 0$  such that  $\tilde{\mathbb{P}}(A) = \mathbb{E}(X1_A)$  for all  $A \in \mathcal{F}$ .

Set  $\mathcal{F}_n = \sigma(G_k : k \leq n)$ . There exists  $m \in \mathbb{N}$  and a partition of  $\Omega$  by events  $A_1, \dots, A_m$  such that  $\mathcal{F}_n = \sigma(A_1, \dots, A_m)$  (e.g. choose  $A_1 = G_1$ ,  $A_2 = G_2 \setminus G_1$ ,  $A_3 = G_3 \setminus (G_1 \cup G_2)$  and so on). Set

$$X_n = \sum_{j=1}^m a_j 1_{A_j}$$

where  $a_j = \tilde{\mathbb{P}}(A_j)/\mathbb{P}(A_j)$  if  $\mathbb{P}(A_j) > 0$  and  $a_j = 0$  otherwise. Then  $X_n \geq 0$ ,  $X_n \in \mathcal{F}_n$ .

Observe that  $(\mathcal{F}_n)_{n \geq 0}$  is a filtration and  $(X_n)_{n \geq 0}$  is a non-negative martingale adapted to  $(\mathcal{F}_n)_{n \geq 0}$  (has to check this). We will show that  $(X_n)_{n \geq 0}$  is *uniformly integrable*. Once shown this, by the  $L^1$ -martingale convergence theorem, there exists  $X \geq 0$  such that  $\mathbb{E}(X1_A) = \mathbb{E}(X_n1_A)$  for all  $A \in \mathcal{F}_n$ . Define a probability measure  $\mathbb{Q}$  on  $\mathcal{F}$  by

$$\mathbb{Q}(A) = \mathbb{E}(X1_A) \quad \forall A \in \mathcal{F}$$

Then  $\mathbb{Q} = \tilde{\mathbb{P}}$  on  $\cup_n \mathcal{F}_n$  which is a  $\pi$ -system generating  $\mathcal{F}$ . Hence  $\mathbb{Q} = \tilde{\mathbb{P}}$  on  $\mathcal{F}$ , by uniqueness of extension. [PM, Thm 1.7.1], which implies (b).

It remains to show that  $(X_n)_n$  is uniformly integrable. Given  $\epsilon > 0$ , we can find  $\delta > 0$  such that  $\tilde{\mathbb{P}}(B) < \epsilon$  for all  $B \in \mathcal{F}$  with  $\mathbb{P}(B) < \delta$ . (If not, there would be a sequence of sets  $(B_n)_n \subset \mathcal{F}$  with  $\mathbb{P}(B_n) < 2^{-n}$  and  $\tilde{\mathbb{P}}(B_n) \geq \epsilon$  for all  $n$ . Then by Borel-Cantelli lemma,  $\mathbb{P}(\limsup B_n) = 0$ , but  $\tilde{\mathbb{P}}(\limsup B_n) > \epsilon$ , which contradicts (a)). Set  $\lambda = 1/\delta$ . Then by Markov inequality,

$$\mathbb{P}(X_n > \lambda) \leq \frac{\mathbb{E}(X_n)}{\lambda} = \frac{1}{\lambda} = \delta \quad \forall n$$

so  $\mathbb{E}(X_n 1_{X_n > \lambda}) = \tilde{\mathbb{P}}(X_n > \lambda) < \epsilon$  for all  $n$ . Hence  $(X_n)_n$  is uniformly integrable by its definition

(End of proof)  $\square$

### 3.3. Markov Chains

- Let  $E$  be a *countable set*. We identify each measure  $\mu$  on  $E$  with  $(\mu_x : x \in E)$  where  $\mu_x = \mu(\{x\})$ . Then for each function  $f$  on  $E$  write

$$\mu(f) = \mu f = \sum_{x \in E} \mu_x f_x \quad (\text{vector product})$$

where  $f_x = f(x)$ .

- A **transition matrix** on  $E$  is a matrix  $P = (p_{xy} : x, y \in E)$  such that each row  $(p_{xy} : y \in E)$  is a probability measure.
- Given a filtration  $(\mathcal{F}_n)_{n \geq 0}$  and  $(X_n)_{n \geq 0}$ , and adapted process with values in  $E$ , we say that  $(X_n)_{n \geq 0}$  is a **Markov chain with transition matrix**  $P$  if, for all  $n \geq 0$ , all  $x, y \in E$  and all  $A \in \mathcal{F}_n$  with  $A \subset \{x_n = x\}$  and  $\mathbb{P}(A) > 0$ ,

$$\mathbb{P}(X_{n+1} = y | A) = p_{xy}$$

Our notion of Markov chain depends on the choice of  $(\mathcal{F}_n)_n$ . The following results show that our definition agrees with the usual one with the choice of the natural filtration of  $(X_n)_n$ .

**Proposition 3.3.1)** Let  $(X_n)_{n \geq 0}$  be a random process in  $E$  and take  $\mathcal{F}_n = \sigma(X_k : k \geq n)$ . Then the following are equivalent :

- $(X_n)_{n \geq 0}$  is a Markov chain with initial distribution  $\mu$  and transition matrix  $P$ .
- For all  $n$  and all  $x_0, x_1, \dots, x_n \in E$ ,

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mu_{x_0} p_{x_0 x_1} \cdots p_{x_{n-1} x_n}$$

**Proposition 3.3.2)** Let  $E^*$  denote the set of sequence  $x = (x_n : n \geq 0)$  taking values in  $E$  and define  $X_n : E^* \rightarrow E$  by  $X_n(x) = x_n$ . Set  $\mathcal{E} = \sigma(X_k : k \geq 0)$ . Let  $P$  be a transition matrix on  $E$ . Then, for each  $y \in E$ , there is a unique probability measure  $\mathbb{P}_y$  on  $(E^*, \mathcal{E}^*)$  such that  $(X_n)_{n \geq 0}$  is a Markov chain with transition matrix  $P$  and starting from  $y$ .

**proof)** The choice of probability measure should be obvious from the transition matrix  $P$ . To show uniqueness, use Dynkin's lemma.

An example of a Markov chain in  $\mathbb{Z}^d$  is the simple symmetric random walk with transition matrix

$$p_{xy} = \begin{cases} 1/2d & \text{if } |x - y| = 1 \\ 0 & \text{otherwise} \end{cases}$$

The following result shows a simple instance of a general relationship between Markov processes and martingale.

**Proposition 3.3.3)** Let  $(X_n)_{n \geq 0}$  be an adapted process in  $E$ . TFAE :

- $(X_n)_{n \geq 0}$  is a Markov chain with transition matrix  $P$ .
- For all bounded functions  $f$  on  $E$ , the following process is a *martingale*

$$M_n^f = f(X_n) - f(X_0) - \sum_{k=0}^{n-1} (P - I)f(X_k)$$



**proof)** (exercise) (Be careful that  $(P - I)f(X_n)$  is not  $P - I$  applied to  $f(X_n)$  but  $(P - I)f$  applied to  $X_n$ .) Suppose that  $(X_n)_n$  is a Markov chain. Then

$$\mathbb{E}(f(X_{n+1})|\mathcal{F}_n) = \mathbb{E}\left(\sum_{y \in E} f(X_{n+1})1_{X_{n+1}=y}|\mathcal{F}_n\right) = \sum_{y \in E} f(y)\mathbb{E}(1_{X_{n+1}=y}|\mathcal{F}_n)$$

**Claim :**  $\mathbb{E}(1_{X_{n+1}=y}|\mathcal{F}_n) = \sum_{x \in X} p_{xy}1_{X_n=x}$

**proof)** Observe that  $\mathbb{E}(1_{X_{n+1}=y}|\mathcal{F}_n) = \mathbb{E}(\mathbb{E}(1_{X_{n+1}=y}|X_n)|\mathcal{F}_n)$ , so it is sufficient to prove that  $\mathbb{E}(1_{X_{n+1}=y}|X_n) = \sum_{x \in X} p_{xy}1_{X_n=x}$ . The expression on the right hand side is clearly  $\sigma(X_n)$ -measurable. Also, for any  $A = \{X_n = w\} \in \sigma(X_n)$

$$\mathbb{E}(\mathbb{E}(1_{X_{n+1}=y}|X_n)1_A) = \mathbb{E}(1_{X_{n+1}=y}1_A) = \mathbb{P}(X_{n+1} = y, X_n = w) = p_{wy}\mathbb{P}(X_n = w)$$

and

$$\mathbb{E}\left(\sum_{x \in X} p_{xy}1_{X_n=x}1_A\right) = \sum_{x \in X} p_{xy}\mathbb{P}(X_n=x, 1_A) = p_{wy}\mathbb{P}(X_n=w)$$

Since  $\{\{X_n = w\} : w \in E\}$  generates  $\sigma(X_n)$ , we have the result.

Therefore,  $\mathbb{E}(f(X_{n+1})|\mathcal{F}_n) = \sum_{x,y \in E} f(y)p_{xy}1_{X_n=x} = P(f)(X_n)$  and therefore

$$\mathbb{E}(f(X_n + 1) - f(X_0) - \sum_{k=0}^n (P - I)f(X_k)|\mathcal{F}_n) = f(X_n) - f(X_0) - \sum_{k=0}^{n-1} (P - I)f(X_k)$$

Now if  $(M_n^f)_n$  is a martingale for any bounded function, then it follows that  $\mathbb{E}(f(X_n + 1)|X_n) = P(f)(X_n)$  for any bounded  $f$  and  $n$ , and therefore  $X_n$  is a Markov chain.

- A bounded function  $f$  on  $E$  is said to be **harmonic** (for the transition matrix  $P$ ) if

$$P(f) = f \quad \text{i.e.} \quad \sum_{y \in E} p_{xy}f_y = f_x \quad \forall x \in E$$

- If  $f$  is a *bounded harmonic function*, then  $(f(X_n))_{n \geq 0}$  is a *bounded martingale*. Then by Doob's convergence theorem,  $f(X_n)$  converges a.s. and in  $L^p$  for all  $p < \infty$ .
- More generally, for  $D \subset E$ , a *bounded function*  $f$  on  $E$  is **harmonic on  $D$**  if

$$\sum_{y \in E} p_{xy}f_y = f_x \quad \forall x \in D$$

- Let  $\partial D = E \setminus D$  and fix a bounded function  $f$  on  $\partial D$ . Set  $T = \inf\{n \geq 0 : X_n \in \partial D\}$  and define a function  $u$  on  $E$  by

$$u(x) = \mathbb{E}_x(f(X_T)1_{T < \infty})$$

where  $\mathbb{E}_x$  is the unique probability measure of a Markov chain starting at  $x \in E$ , as defined in **Prop 3.3.2**.

**Theorem 3.3.4)** The function  $u$  is bounded, harmonic in  $D$ , and  $u = f$  on  $\partial D$ . Moreover, if  $\mathbb{P}_x(T < \infty) = 1$  for all  $x \in D$ , then  $u$  is the unique bounded extension of  $f$  which is harmonic in  $D$ .

**proof)** It is clear that  $u$  is bounded and  $u = f$  on  $\partial D$ . For all  $x, y \in E$  with  $p_{xy} > 0$  under  $\mathbb{P}_x$ , conditional on  $\{X_1 = y\}$ ,  $(X_{n+1})_{n \geq 0}$  has distribution  $\mathbb{P}_y$ . So for  $x \in D$ ,  $u(x) = \sum_{y \in E} p_{xy} u(y)$  showing  $u$  is harmonic in  $D$ .

On the other hand, suppose that  $g$  is a bounded function harmonic in  $D$  such that  $g = f$  on  $\partial D$ . Then  $M = M^g$  (where  $M$  is as defined in **Prop 3.3.3**) is a martingale and  $T$  is a stopping time, so  $M^T$  is also a martingale by optional stopping theorem. But  $M_{T \wedge n} = g(X_{T \wedge n})$  so if  $\mathbb{P}_x(T < \infty) = 1$  for all  $x \in D$ , then

$$M_{T \wedge n} \rightarrow g(X_T) = f(X_T) \quad \text{a.s.}$$

So by bounded convergence, for all  $x \in D$ ,

$$g(x) = \mathbb{E}_x(M_0) = \mathbb{E}_x(M_{T \wedge n}) \rightarrow \mathbb{E}_x(f(X_T)) = u(x)$$

therefore  $g(x) = u(x)$  for all  $x \in E$ .

(End of proof)  $\square$

(29th October, Monday)

## 4. Random processes in continuous time

### 4.1. Definitions

- A **continuous random process** is a family of random variables  $(X_t)_{t \geq 0}$  such that for all  $\omega \in \Omega$ , the path

$$t \mapsto X_t(\omega) : [0, \infty) \rightarrow \mathbb{R}$$

is continuous.

- A function  $x : [0, \infty) \rightarrow \mathbb{R}$  is said to be **cadlag** if it is *right-continuous with left limits*, i.e. for all  $t \geq 0$

$$x_s \rightarrow x_t \quad \text{as } s \rightarrow t^+$$

and for all  $t > 0$ , there exists  $x_{t-} \in \mathbb{R}$  such that

$$x_s \rightarrow x_{t-} \quad \text{as } s \rightarrow t^-$$

- A **cadlag random process** is a family of random variables  $(X_t)_{t \geq 0}$  such that for all  $\omega \in \Omega$ , the path

$$t \mapsto X_t(\omega) : [0, \infty) \rightarrow \mathbb{R}$$

is cadlag.

- The space of continuous and cadlag functions on  $[0, \infty)$  are denoted  $C([0, \infty), \mathbb{R})$  and  $D([0, \infty), \mathbb{R})$  respectively. We equip these spaces with the  $\sigma$ -algebra generated by the coordinate functions  $\sigma(x \mapsto x_t : t \geq 0)$ . A continuous(/cadlag) random process  $(X_t)_{t \geq 0}$  can then be considered as a random variable  $X$  in  $C([0, \infty), \mathbb{R})/(D([0, \infty), \mathbb{R}))$ , given by

$$X(\omega) = (t \mapsto X_t(\omega) : t \geq 0)$$

- The **finite-dimensional distributions** of a continuous or cadlag process  $X$  are the laws  $\mu_{t_1, \dots, t_n}$  on  $\mathbb{R}^n$  given by  $\mu_{t_1, \dots, t_n}(A) = \mathbb{P}((X_{t_1}, \dots, X_{t_n}) \in A)$  for  $A \in \mathcal{B}(\mathbb{R}^n)$  where  $n \in \mathbb{N}$  and  $0 \leq t_1 < \dots < t_n < \infty$ . Since the cylinder sets  $\{(X_{t_1}, \dots, X_{t_n}) \in A\}$  form a generating  $\pi$ -system, they determine uniquely the law of  $X$ . We make analogous definition when  $\mathbb{R}$  is replaced by a general topological space.

## Kolmogorov's Criterion

**Theorem 4.2.1** (*Kolmogorov's criterion*) Let  $p \in (1, \infty)$  and  $\beta \in (1/p, 1]$ . Let  $I$  be a dense subset of  $[0, 1]$  and let  $(\xi_t)_{t \in I}$  a family of random variables such that for some constant  $C < \infty$ ,

$$\|\xi_s - \xi_t\|_p \leq C|s - t|^\beta \quad \forall s, t \in I \quad \dots\dots\dots (\dagger)$$

Then there exists a continuous random process  $(X_t)_{t \in [0, 1]}$  such that  $X_t = \xi_t$  a.s. for all  $t \in I$ . Moreover  $(X_t)_{t \in [0, 1]}$  may be chosen such that for all  $\alpha \in [0, \beta - \frac{1}{p})$  such that

$$|X_s - X_t| \leq K_\alpha |s - t|^\alpha \quad \forall s, t \in [0, 1]$$

This theorem indicates that  $L^p$ -Hölder continuity on a dense subset implies Hölder continuity of the random process. Later, this becomes important in construction of Brownian motion, and different stochastic processes.

**proof)** For  $n \geq 0$ , write

$$\begin{aligned} \mathbb{D}_n &= \{k2^{-n} : k \in \mathbb{Z}^+\} & \mathbb{D} &= \bigcup_{n \geq 0} \mathbb{D}_n \\ D_n &= \mathbb{D}_n \cap [0, 1] & D &= \mathbb{D} \cap [0, 1] \end{aligned}$$

By taking limits in  $L^p$ , we can extend  $(\xi_t)_{t \in I}$  to all  $t \in D$  and such that  $(\dagger)$  holds for all  $s, t \in D \cup I$ . (The limit exist, because each sequence converging to a point forms a Cauchy sequence).

For  $n \geq 0$  and  $\alpha \in [0, \beta - \frac{1}{p})$ , define non-negative random variable by

$$K_n = \sup_{t \in D_n} |\xi_{t+2^{-n}} - \xi_t|, \quad K_\alpha = 2 \sum_{n \geq 0} 2^{n\alpha} K_n$$

Then

$$\mathbb{E}((K_n)^p) \leq \mathbb{E}\left(\sum_{t \in D_n} |\xi_{t+2^{-n}} - \xi_t|^p\right) \leq 2^n C^p (2^{-n})^{\beta p} \quad (\text{by } (\dagger))$$

so

$$\|K_\alpha\|_p \leq 2 \sum_{n \geq 0} 2^{n\alpha} \|K_n\|_p \leq 2C \sum_{n \geq 0} 2^{-(\beta - \alpha - \frac{1}{p})n} < \infty$$

For  $s, t \in D$ , with  $s < t$ , choose  $m \geq 0$  such that  $2^{-m-1} < t - s \leq 2^{-m}$ . Then interval  $[s, t)$  can be expressed as a finite disjoint union of intervals of the form  $[r, r + 2^{-n})$  where  $r \in D_n$  and  $n \geq m + 1$  and no 3 intervals have the same length. Hence  $|\xi_t - \xi_s| \leq 2 \sum_{n \geq m+1} K_n$  and so

$$\frac{|\xi_t - \xi_s|}{/} (t - s)^\alpha \leq 2 \sum_{n \geq m+1} K_n 2^{(m+1)\alpha} \leq K_\alpha$$

Now define

$$X_t(\omega) = \begin{cases} \lim_{s \rightarrow t, s \in D} \xi_s(\omega) & \text{if } K_\alpha(\omega) < \infty \text{ for all } \alpha \in [0, \beta - \frac{1}{p}) \\ 0 & \text{otherwise} \end{cases}$$

but  $\mathbb{P}(K_\alpha(\omega) < \infty \text{ for all } \alpha \in [0, \beta - \frac{1}{p})) = 1$  (note that  $K_\alpha$  is an increasing function of  $\alpha$ , so it is enough to take countable sequence  $(\alpha_k)_k \rightarrow \beta - \frac{1}{p}$ ), so  $X_t(\omega) = \lim_{s \rightarrow t} \xi_s(\omega)$  a.s. Then  $(X_t)_{t \in [0,1]}$  is a continuous random process with the claimed properties.

(End of proof)  $\square$

### 4.3. Martingales in continuous time

We assume in this section that our probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  is equipped with a **continuous filtration**, i.e. a family of  $\sigma$ -algebras  $(\mathcal{F}_t)_{t \geq 0}$  such that  $\mathcal{F}_s \subset \mathcal{F}_t$  for all  $s \leq t$ .

- Define for  $t \geq 0$ ,  $\mathcal{F}_{t+} = \bigcap_{s > t} \mathcal{F}_s$ ,  $\mathcal{F}_\infty = \sigma(\mathcal{F}_t : t \geq 0)$  and

$$\mathcal{N} = \{A \in \mathcal{F}_\infty : \mathbb{P}(A) = 0\}$$

- The filtration  $(\mathcal{F}_t)_{t \geq 0}$  is said to satisfy the **usual conditions** if  $\mathcal{N} \subset \mathcal{F}_0$  and  $\mathcal{F}_t = \mathcal{F}_{t+}$  for all  $t$ .
- A continuous(/cadlag) adapted integrable random process is said to be a **martingale** if, for all  $s, t \geq 0$  with  $s \leq t$ ,

$$\mathbb{E}(X_t | \mathcal{F}_s) = X_s \quad \text{a.s.}$$

Define super-martingale and sub-martingale accordingly.

---

(31st October, Wednesday)

Recall  $\mathbb{D}_n = \{k2^{-n} : k \in \mathbb{Z}^+\}$ ,  $\mathbb{D} = \bigcup_{n \geq 0} \mathbb{D}_n$ . Define, for a cadlag random process  $X$ ,

$$X^* = \sup_{t \geq 0} |X_t| \quad \text{and} \quad X^{(n)*} = \sup_{t \in \mathbb{D}_n} |X_t|$$

The cadlag property implies  $X^{(n)*} \rightarrow X^*$  as  $n \rightarrow \infty$ .

If  $(X_t)_{t \geq 0}$  is a cadlag (super/sub) martingale, then  $(X_t)_{t \in \mathbb{D}_n}$  is a discrete-time (super/sub) martingale, for  $(\mathcal{F}_t)_{t \in \mathbb{D}_n}$ . Therefore, we can find an analogy of discrete Doob's inequality in a continuous time cadlag (super/sub) martingale by sending  $n \rightarrow \infty$ .

**Theorem 4.3.1)** (Doob's maximal inequality) Let  $X$  be a cadlag martingale or non-negative sub-martingale. Then for all  $\lambda \geq 0$ ,

$$\lambda \mathbb{P}(X^* \geq \lambda) \leq \sup_{t \geq 0} \mathbb{E}(|X_t|)$$

**Theorem 4.3.2)** (Doob's  $L^p$  inequality) Let  $X$  be a cadlag martingale or non-negative sub-martingale. Then for all  $p > 1$  and  $q = p/(p-1)$ ,

$$\|X^*\|_p \leq q \sup_{t \geq 0} \|X_t\|_p$$

**Theorem 4.3.3)** (Doob's upcrossing inequality) Let  $X$  be a cadlag super-martingale and let  $a, b \in \mathbb{R}$  with  $a < b$ . Then

$$(b-a)\mathbb{E}(U[a, b]) \leq \sup_{t \geq 0} \mathbb{E}((X_t - a)^-)$$

where  $U[a, b]$  is the number of disjoint upcrossing of  $[a, b]$  by  $X_t$ .

**proof)** The proof of above theorems is exercise.

We also have different versions of martingale convergence theorem.

**Theorem 4.3.4)** (a.s. martingale convergence theorem) Let  $X$  be an  $L^1$ -bounded cadlag super-martingale. Then there exists an integrable  $\mathcal{F}_\infty$ -measurable random variable  $X_\infty$  such that  $X_t \rightarrow X_\infty$  almost surely as  $t \rightarrow \infty$ .

**Theorem 4.3.5)** ( $L^1$  martingale convergence theorem) Let  $(X_t)_{t \geq 0}$  be a uniformly integrable cadlag martingale. Then there exists a random variable  $X_\infty \in L^1(\mathcal{F}_\infty)$  such that  $X_t \rightarrow X_\infty$  a.s. and in  $L^1$ . Moreover,  $X_t = \mathbb{E}(X_\infty | \mathcal{F}_t)$  a.s. for all  $t \geq 0$ .

Conversely, if  $(\mathcal{F}_t)_t$  satisfies the usual condition, then for all  $Y \in L^1(\mathcal{F}_\infty)$ , there exists a uniformly integrable cadlag martingale  $(X_t)_t$  such that  $X_t = \mathbb{E}(Y | \mathcal{F}_t)$  a.s. for all  $t$ , and  $X_t \rightarrow Y$  a.s. and in  $L^1$ .

**proof)** Choose a version  $\xi_t$  of  $\mathbb{E}(Y | \mathcal{F}_t)$  for all  $t \in \mathbb{D}$ . Then  $(\xi_t)_{t \in \mathbb{D}}$  is uniformly integrable and  $(\xi_t)_{t \in \mathbb{D}_n}$  is a discrete-time martingale for all  $n \geq 0$ . Set  $\xi^* = \sup_{t \in \mathbb{D}} |\xi_t|$  and write  $U[a, b]$  for the total number of disjoint upcrossings of  $[a, b]$  by  $(\xi_t)_{t \in \mathbb{D}}$ . Set

$$\Omega_0 = \Omega^* \cap \left( \bigcap_{a, b \in \mathbb{Q}, a < b} \Omega_{a, b} \right)$$

where  $\Omega^* = \{\xi^* < \infty\}$  and  $\Omega_{a, b} = \{U[a, b] < \infty\}$ . Then discrete Doob's inequality implies  $\mathbb{P}(\Omega_0) = 1$  (this involves some technical problems - we may first show that  $\mathbb{P}(\Omega_{0, n}) = 0$  for all  $n$  where  $\Omega_{0, n} = \Omega_n^* \cap \bigcap_{a, b \in \mathbb{Q}, a < b} \Omega_{a, b, n}$ ,  $\Omega_n^* = \{\sup_{t \in \mathbb{D}_n} |\xi_t| < \infty\}$ ,  $\Omega_{a, b, n} = \{U_{(n)}[a, b] < \infty\}$ , and  $U_{(n)}[a, b]$  is the number of upcrossings of  $[a, b]$  by  $(\xi_t)_{t \in \mathbb{D}_n}$ . This result is almost immediate from the case of discrete martingales). Define for  $t \geq 0$

$$X_t = \lim_{s \rightarrow t^+, s > t, s \in \mathbb{D}} \xi_s 1_{\Omega_0}$$

The usual conditions ensure that  $(X_t)_{t \geq 0}$  is adapted to  $(\mathcal{F}_t)_{t \geq 0}$ . It is clear that  $(X_t)_{t > 0}$  is cadlag and  $X_t = \mathbb{E}(Y | \mathcal{F}_t)$  a.s. for all  $t \geq 0$  so  $(X_t)_{t \geq 0}$  is a uniformly integrable cadlag martingale.

If  $X_t \rightarrow X_\infty$  a.s. and in  $L^1$  then  $X_\infty = Y$  a.s. by the same argument as in the discrete time.

(End of proof)  $\square$

**Theorem 4.3.6)** ( $L^p$  martingale convergence theorem) Let  $(X_t)_{t \geq 0}$  be a  $L^p$ -bounded cadlag martingale. Then there exists a random variable  $X_\infty \in L^p(\mathcal{F}_\infty)$  such that  $X_t \rightarrow X_\infty$  a.s. and in  $L^p$ . Moreover,  $X_t = \mathbb{E}(X_\infty | \mathcal{F}_t)$  a.s. for all  $t \geq 0$ .

Moreover, if  $(\mathcal{F}_t)_t$  satisfies the usual condition, then for all  $Y \in L^1(\mathcal{F}_\infty)$ , there exists an  $L^p$ -bounded cadlag martingale  $(X_t)_t$  such that  $X_t = \mathbb{E}(Y | \mathcal{F}_t)$  a.s. for all  $t$ , and  $X_t \rightarrow Y$  a.s. and in  $L^p$ .

**proof)** The proof is almost the same as the  $L^1$  case.

(End of proof)  $\square$

- We say that a random variable  $T : \Omega \rightarrow [0, \infty]$  is a **stopping time** if  $\{T \leq t\} \in \mathcal{F}_t$  for all  $t \geq 0$ .
- For a stopping time  $T$ , we set

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\}$$

This indeed is a  $\sigma$ -algebra. (stated in the next proposition)

- Given a cadlag random process  $X$ , we define  $X_T$  and  $X^T$  as

$$\begin{aligned} X_T(\omega) &= X_{T(\omega)}(\omega) \\ X^T(\omega) &= X_{T(\omega) \wedge t}(\omega) \end{aligned}$$

where we leave  $X_{T(\omega)}$  undefined if  $T(\omega) = \infty$  and  $X_t(\omega)$  fails to converge as  $t \rightarrow \infty$ .

**Proposition 4.3.7)** Let  $S$  and  $T$  be stopping times and let  $X$  be a cadlag adapted process. Then

- (a)  $S \wedge T$  is a stopping time.
- (b)  $\mathcal{F}_T$  is a  $\sigma$ -algebra.
- (c) If  $S \leq T$ , then  $\mathcal{F}_S \subset \mathcal{F}_T$ .
- (d)  $X_T 1_{T < \infty}$  is  $\mathcal{F}_T$ -measurable.
- (e)  $X^T$  is adapted.

**Theorem 4.3.8)** Let  $X$  be a cadlag adapted integrable process. TFAE

- (a)  $X$  is a martingale.
- (b) for all bounded stopping times  $T$  and all stopping times  $S$ ,  $X_T$  is integrable and

$$\mathbb{E}(X_T | \mathcal{F}_S) = X_{S \wedge T} \quad \text{a.s.}$$

- (c) for all stopping times  $T$ ,  $X^T$  is a martingale.
- (d) for all bounded stopping times  $T$ ,  $X_T$  is integrable and  $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ .

Moreover, if  $X$  is uniformly integrable, then (b) and (d) hold for all stopping times.

**proof)** We are only going to prove (a) $\Rightarrow$ (b) here.

(a) $\Rightarrow$ (b) Let  $S$  and  $T$  be stopping times and  $T \leq t$ . Let  $A \in \mathcal{F}_S \subset \mathcal{F}_{S_n}$ . For  $n \geq 0$ , set  $S_n = 2^{-n} \lceil 2^n S \rceil$ ,  $T_n = 2^{-n} \lceil 2^n T \rceil$ . (note that this is not what we usually do - people usually take  $\lfloor \cdot \rfloor$  instead of  $\lceil \cdot \rceil$ . However, in this case, taking ceiling is more useful, because  $X$  is right-continuous so want to make  $T_n \searrow T$ ) Then  $S_n$  and  $T_n$  are stopping times and  $S_n \searrow S$  and  $T_n \searrow T$  as  $n \rightarrow \infty$ . Since  $(X_t)_{t \geq 0}$  is right-continuous,

$$X_{T_n} \xrightarrow{\text{a.s.}} X_T \quad \text{as } n \rightarrow \infty$$

By **Theorem 2.2.3**,  $X_{T_n} = \mathbb{E}(X_{t+1} | \mathcal{F}_{T_n})$  so  $(X_{T_n} : n \geq 0)$  is uniformly integrable and so  $X_{T_n} \rightarrow X_T$  in  $L^1$  by backward martingale convergence. In particular,  $X_T$  is integrable. Similarly,  $X_{S_n \wedge T_n} \rightarrow X_{S \wedge T}$  in  $L^1$ . By **Theorem 2.2.3**, as  $\mathbb{E}(X_{T_n} | \mathcal{F}_{S_n}) = X_{S_n \wedge T_n}$ ,

$$\mathbb{E}(X_{T_n} 1_A) = \mathbb{E}(X_{S_n \wedge T_n} 1_A)$$

Letting  $n \rightarrow \infty$ ,

$$\mathbb{E}(X_T 1_A) = \mathbb{E}(X_{S \wedge T} 1_A)$$

by dominated convergence. So we deduce (b).

(End of proof)  $\square$

(2nd November, Friday)

## Chapter 5. Weak Convergence

### 5.1. Definitions

Let  $E$  be a metric space. Whenever we are talking about a metric space, the  $\sigma$ -algebra is given by the Borel  $\sigma$ -algebra. Write  $C_b(E)$  for the set of bounded continuous functions on  $E$ .

- Let  $(\mu_n : n \in \mathbb{N})$  be a sequence of probability measures and let  $\mu$  be another probability measure on  $E$ . We say that  $\mu_n \rightarrow \mu$  **weakly** (as  $n \rightarrow \infty$ ) if  $\mu_n(f) \rightarrow \mu(f)$  for all  $f \in C_b(\mathbb{R})$ .

**Theorem 5.1.1)** The following are equivalent.

- (a)  $\mu_n \rightarrow \mu$  weakly on  $E$
- (b)  $\liminf_{n \rightarrow \infty} \mu_n(U) \geq \mu(U)$  for all  $U$  open
- (c)  $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$  for all  $F$  closed.
- (d)  $\mu_n(B) \rightarrow \mu(B)$  for all  $B \in \mathcal{B}$  such that  $\mu(\partial B) = 0$ . (Boundary is the set of limit points of  $B$  that are not contained in  $B$ .)

**proof)** Exercise.

For an example, consider a sequence  $(x_n)_n \subset \mathbb{R}$  such that  $x_n \rightarrow 0$  as  $n \rightarrow \infty$ . We want to have  $\delta_{x_n} \rightarrow \delta_0$ . Indeed, this is true in the weak sense. However, the sequence has  $\delta_{x_n}(\{0\}) = 0$  for all  $n$ , hence we should have inequality in condition (c).

We have a similar version of the theorem for the real line.

**Proposition 5.1.2)** Consider the case  $E = \mathbb{R}$ . TFAE

- (a)  $\mu_n \rightarrow \mu$  weakly for some probability measure  $\mu$ .
- (b)  $F_n(x) \rightarrow F(x)$  for all  $x \in \mathbb{R}$  such that  $F(x^-) = F(x)$ . (Here,  $F(x) = \mu((-\infty, x])$  is the **distribution function** of  $\mu$ .) (Sometimes called convergence of distributions)
- (c) There exists a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and random variables  $X_n, X$  on  $\Omega$  such that  $X_n \sim \mu_n$ ,  $X \sim \mu$  and  $X_n \rightarrow X$  almost surely.

**proof)** See probability and measure notes.

## 5.2. Prohorov's Theorem

When does a sequence of probability measures has a converging subsequence?

Let  $E$  be a metric space and  $(\mu_n : n \in \mathbb{N})$  be a sequence of probability measures on  $E$ .

- We say that  $(\mu_n)_n$  is **tight** if for all  $\epsilon > 0$ , there is a compact set  $K \subset E$  such that

$$\mu_n(E \setminus K) \leq \epsilon \quad \forall n \in \mathbb{N}$$

For example, the sequence  $(\delta_n)_n$  is *not* tight.

**Theorem 5.2.1)** Let  $(\mu_n : n \in \mathbb{N})$  be a sequence of probability measures on a metric space  $E$  and suppose that  $(\mu_n : n \in \mathbb{N})$  is tight. Then there exists a subsequence  $(n_k)_k \subset \mathbb{N}$  and probability measure  $\mu$  on  $E$  such that  $\mu_{n_k} \rightarrow \mu$  weakly as  $k \rightarrow \infty$ .

This gives a version of weakly sequential compactness of probability measures. We are only going to prove this for  $\mathbb{R}$ . This theorem is hard to prove in general.(e.g. there is a method using Monge-Kantorovich metric defined for Polish spaces. For this method, see "Topics in Optimal Transport", C.Villani, Ame.Soc.Math. For the general version, see the attached note)

**proof for  $E = \mathbb{R}$ )** By a diagonal argument and by passing to a subsequence, it suffices to consider the case where  $F_n(x) \rightarrow g(x)$  as  $n \rightarrow \infty$  for all  $x \in \mathbb{Q}$  for some  $g(x) \in [0, 1]$ , where  $F_n$  is the distribution function of  $\mu_n$ . Now  $g : \mathbb{Q} \rightarrow [0, 1]$  is non-decreasing so  $g$  has a non-decreasing extension  $G : \mathbb{R} \rightarrow [0, 1]$ , i.e.

$$G(x) = \lim_{q \searrow x, q \in \mathbb{Q}} g(q)$$

which has only countably many discontinuities.(because there should be a rational number in each discontinuity). Now we must have

$$F_n(x) \rightarrow G(x) \quad \forall x \text{ s.t. } G \text{ is continuous at } x$$



Set  $F(x) = G(x^+)$ , then  $F$  and  $G$  have same points of continuity, so  $F_n(x) \rightarrow F(x)$  for all  $x \in \mathbb{R}$ .

We are only left to check that  $G(x) \rightarrow 1$  as  $x \rightarrow \infty$  using tightness condition.

Since  $(\mu_n : n \in \mathbb{N})$  is tight, given  $\epsilon > 0$ , there exists  $R < \infty$  such that  $\mu_n(\mathbb{R} \setminus (-R, R)) \leq \epsilon$  for all  $n$  so  $F_n(-R) \leq \epsilon$ ,  $F_n(R) \geq 1 - \epsilon$ . So

$$\begin{aligned} F(x) &\rightarrow 0 \quad \text{as } x \rightarrow -\infty \\ F(x) &\rightarrow 1 \quad \text{as } x \rightarrow \infty \end{aligned}$$

So  $F$  is distribution function. So there exists a probability measure  $\mu$  such that  $\mu((-\infty, x]) = F(x)$ . Then  $\mu_n \rightarrow \mu$  by **Prop 5.1.2**.

(End of proof)  $\square$

### 5.3. Weak Convergence and Characteristic Functions

Take  $E = \mathbb{R}^d$ .

- For a probability measures  $\mu$  on  $\mathbb{R}^d$ , define its **characteristic function**  $\phi : \mathbb{R}^d \rightarrow \mathbb{C}$  by

$$\phi(u) = \int_{\mathbb{R}^d} e^{i\langle u, x \rangle} \mu(dx)$$

**Lemma 5.3.1)** Fix  $d = 1$ . For all  $\lambda \in (0, \infty)$ ,

$$\mu(\mathbb{R} \setminus (-\lambda, \lambda)) \leq C\lambda \int_0^\lambda (1 - \operatorname{Re}(\phi(u))) du$$

where  $C = (1 - \sin(1))^{-1} < \infty$ .

**proof)** Consider for  $t \geq 1$ . Let  $A(t) = t^{-1} \int_0^t (1 - \cos v) dv$ . Then

$$A(t) \geq A(0) = 1 - \sin(t)$$

(to see this, observe that  $A(t)$  is the average of  $(1 - \cos(v))$  on interval  $(0, t)$  and divide the cases  $|t| \leq \pi/2$  and  $|t| \geq \pi/2$ )

So  $Ct^{-1} \int_0^t (1 - \cos(v)) dv \geq 1$ . Substitute  $v = uy$ ,  $u = v/y$ ,

$$Ct^{-1} \int_0^{t/y} (1 - \cos(uy)) y du \geq 1$$

Put  $t/y = 1/\lambda$ ,  $\lambda = y/t$ ,  $t = y/\lambda \geq 1$  to see

$$C\lambda \int_0^{1/\lambda} (1 - \cos(uy)) du \geq 1$$

whenever  $t = y/\lambda \geq 1$  (this was the assumption we started with). Now for general  $y \in \mathbb{R}$ , has

$$C\lambda \int_0^{1/\lambda} (1 - \cos(uy)) du \geq 1_{|y| \geq \lambda}$$

Now integrate with respect to  $\mu$  and use Fubini.

$$\begin{aligned}\mu(\mathbb{R} \setminus (-\lambda, \lambda)) &\leq C\lambda \int_{\mathbb{R}} \int_0^{1/\lambda} (1 - \cos(uy)) du \mu(dy) \\ &= C\lambda \int_0^{1/\lambda} \int_{\mathbb{R}} (1 - \cos(uy)) du \mu(dy)\end{aligned}$$

(End of proof)  $\square$

(5th November, Monday)

**Theorem 5.3.2)** Let  $\mu_n, \mu$  be probability measures on  $\mathbb{R}^d$  with characteristic functions  $\phi_n, \phi$ . Then the following are equivalent

- (a)  $\mu_n \rightarrow \mu$  weakly on  $\mathbb{R}^d$ .
- (b)  $\phi_n(u) \rightarrow \phi(u)$  for all  $u \in \mathbb{R}^d$ .

We will prove only for the case  $d = 1$ .

**proof)** It is clear that (a) implies (b). Suppose (b) holds. We prove via a 'compactness argument'. We aim to show that the sequence  $(\mu_n)_n$  is tight, and therefore has a converging subsequence, and show that the converging point is in fact  $\mu$ .

Note that  $\phi(0) = 1$  and  $\phi$  is continuous. Given  $\epsilon > 0$ , there exists  $\lambda < \infty$  such that

$$C\lambda \int_0^{1/\lambda} (1 - \operatorname{Re}(\phi(u))) du \leq \epsilon/2$$

with  $C = (1 - \sin(1))^{-1} < \infty$ . By dominated convergence,

$$\int_0^{1/\lambda} (1 - \operatorname{Re}(\phi_n(u))) du \xrightarrow{n \rightarrow \infty} \int_0^{1/\lambda} (1 - \operatorname{Re}(\phi(u))) du$$

so for sufficiently large  $n$ , by **Lemma 5.3.1**,

$$\mu_n(\mathbb{R} \setminus (-\lambda, \lambda)) \leq C\lambda \int_0^{1/\lambda} (1 - \operatorname{Re}(\phi_n(u))) du \leq \epsilon$$

Since  $\epsilon$  was arbitrary, we see that  $(\mu_n : n \in \mathbb{N})$  is tight. By Prohorov's theorem, we have a converging subsequence  $\mu_{n_k} \rightarrow \nu$  for some probability measure  $\nu$ .

Suppose for a contradiction that  $\nu \neq \mu$ . Therefore, there exists  $\epsilon > 0$ , and  $f \in C_b(\mathbb{R}^n)$  such that

$$|\mu_{n_k}(f) - \mu(f)| \geq \epsilon \quad \forall k$$

By above argument, we have  $\mu_{n_k} \rightarrow \nu$ . But then, since  $e^{inx}$  is a bounded continuous function,

$$\int_{\mathbb{R}} e^{inx} \nu(dx) = \lim_{k \rightarrow \infty} \phi_{n_k}(n) = \phi(n)$$

which indicates  $\mu = \nu$  by uniqueness of characteristic functions (see PM notes), a contradiction.

(End of proof)  $\square$

In fact, the proof of the theorem implies a slightly stronger statement, which is less useful.

**Theorem 5.3.3)** (*Lévy's continuity theorem for characteristic functions*) Let  $(\mu_n : n \in \mathbb{N})$  be a sequence of probability measures on  $\mathbb{R}^n$  with characteristic functions  $\phi_n$ . Suppose  $\phi_n(u) \rightarrow \phi(u)$  for all  $u$  for some function  $\phi$  (not necessarily a characteristic function) such that  $\phi$  is continuous at 0. Then  $\phi$  is the characteristic function of some probability measure  $\mu$  on  $\mathbb{R}^d$  and  $\mu_n \rightarrow \mu$  weakly on  $\mathbb{R}^d$ .

## 6. Large Deviations

### 6.1. Cramér's theorem

**Theorem 6.1.1)** Let  $(X_n : n \in \mathbb{N})$  be a sequence of integrable *i.i.d.* random variables in  $\mathbb{R}$ . Set  $m = \mathbb{E}(X_1)$ ,  $S_n = X_1 + \cdots + X_n$ . We know  $S_n/n \rightarrow \delta_m$  in probability, so if  $(m - \epsilon, m + \epsilon) \cap B = \emptyset$  then  $\mathbb{P}(S_n/n \in B) \rightarrow 0$  as  $n \rightarrow \infty$ . Then in fact the convergence rate is given by  $\sim \exp(-n\alpha(B))$  for some  $\alpha$ . To be precise, for all  $a \geq m = \mathbb{E}(X_1)$ , as  $n \rightarrow \infty$ ,

$$\frac{1}{n} \log \mathbb{P}(S_n \geq na) \rightarrow -\psi^*(a)$$

where  $\psi^*$  is the *Legendre transform* of the *cumulant generating function*  $\psi(\lambda) = \log(\mathbb{E}(e^{\lambda X_1}))$ , where Legendre transform is given by

$$\psi^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \psi(\lambda)\}$$

In particular, for  $n$  sufficiently large and in case  $\psi^*(a) < \infty$ , we get

$$-\psi^*(a) - \epsilon \leq \frac{1}{n} \log(\mathbb{P}(S_n \geq a)) \leq -\psi^*(a) + \epsilon$$

and therefore

$$e^{-n(\psi^*(a)+\epsilon)} \leq \mathbb{P}(S_n \geq na) \leq e^{-n(\psi^*(a)-\epsilon)}.$$

**Note :**  $\psi$  is always a convex function, so  $\psi^*$  is also a convex function.

**Examples :**

(i)  $X_1 \sim N(0, 1)$ , then  $\mathbb{E}(e^{\lambda X_1}) = e^{\lambda^2/2}$ ,  $\psi(\lambda) = \lambda^2/2$  and  $\psi^*(x) = x^2/2$ . Hence

$$\frac{1}{n} \log(\mathbb{P}(S_n \geq a)) \rightarrow -\frac{a^2}{2} \quad \forall a \geq 0$$

Can check this directly, using the fact that  $S_n \sim N(0, n)$  in this case.

(ii)  $X_1 \sim \text{Exp}(1)$ , then

$$\mathbb{E}(e^{\lambda X_1}) = \int_0^\infty e^{\lambda x} e^{-x} dx = \begin{cases} \infty & \text{if } \lambda \geq 1 \\ \frac{1}{1-\lambda} & \text{if } \lambda < 1 \end{cases}$$

so  $\psi(\lambda) = -\log(1 - \lambda)$  if  $\lambda < 1$  and  $\infty$  otherwise, and  $\psi^*(x) = x - 1 - \log(x)$  for  $x > 0$ . Cramér's theorem implies that

$$\frac{1}{n} \log \mathbb{P}(S_n \geq na) \rightarrow -(a - 1 - \log(a)) \quad \forall a \geq 1$$

On the other hand,  $\text{Var}(X_1) = 1 < \infty$ , so  $\frac{S_n - n}{\sqrt{n}} \rightarrow N(0, 1)$  by CLT. So

$$\mathbb{P}(S_n \geq n + a\sqrt{n}) \rightarrow \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

so Cramér's theorem gives a result of a different flavour from CLT for distributions with bounded variation : while CLT provides a description for distribution near the average, Cramér gives an explanation of tail distribution of  $S_n$ .

**preparation for proof of Cramér's theorem)** Let  $\mu(B) = \mathbb{P}(X_1 \in B)$ . Exclude the easy case where  $\mu = \delta_m$ . Define for  $\lambda \geq 0$  with  $\psi(\lambda) < \infty$ , the **tilted distribution**  $\mu_\lambda$  by

$$\mu_\lambda(dx) \propto e^{\lambda x} \mu(dx)$$

For  $K \geq m = \mathbb{E}(X_1)$ , define the conditional distribution by

$$\mu_K(dx|x \leq K) \propto 1_{\{x \leq K\}} \mu(dx)$$

The CGF(cumulant generating function) of  $\mu_K$  is then given by

$$\psi_K(\lambda) = \log(\mathbb{E}(e^{\lambda X_1} | X_1 \leq K))$$

(7th November, Wednesday)

We now start proving the following theorem.

**Theorem 6.1.1)** Let  $(X_n : n \in \mathbb{N})$  be a sequence of integrable *i.i.d.* random variables in  $\mathbb{R}$ . Set  $m = \mathbb{E}(X_1)$ ,  $S_n = X_1 + \dots + X_n$ . Then for all  $a \geq m = \mathbb{E}(X_1)$ , as  $n \rightarrow \infty$ ,

$$\frac{1}{n} \log \mathbb{P}(S_n \geq na) \rightarrow -\psi^*(a)$$

where  $\psi(\lambda) = \log(\mathbb{E}(e^{\lambda X_1}))$ , and  $\psi^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \psi(\lambda)\}$ .

**proof) (Upper bound)** For all  $\lambda \geq 0$  and  $n \geq 1$

$$\mathbb{P}(S_n \geq na) = \mathbb{P}(e^{\lambda S_n} \geq e^{\lambda na}) \leq e^{-\lambda na} \mathbb{E}(e^{\lambda S_n}) = e^{-(\lambda a - \psi(\lambda))n}$$

so  $\frac{1}{n} \log \mathbb{P}(S_n \geq na) \leq -(\lambda a - \psi(\lambda))$  and

$$\frac{1}{n} \log \mathbb{P}(S_n \geq na) \leq -\psi^*(a)$$

(Lower bound) It remains to show the lower bound. That is, we aim to prove

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) \geq -\psi^*(a)$$

Consider first the case where  $\mathbb{P}(X_1 \leq a) = 1$ . Then

$$\mathbb{E}(e^{\lambda(X_1 - a)}) \xrightarrow{\lambda \rightarrow \infty} \mathbb{P}(X_1 = a)$$

Call  $p = \mathbb{P}(X_1 = a)$ , so  $\lambda a - \psi(\lambda) \rightarrow -\log(p)$ . So in particular,

$$\psi^*(a) \geq -\log(p)$$

Now  $\mathbb{P}(S_n \geq na) = p^n$  so

$$\frac{1}{n} \log \mathbb{P}(S_n \geq na) = \log(p) \geq -\psi^*(a)$$

hence we can eliminate the case  $\mathbb{P}(X_1 \leq a) = 1$ .

Next consider the case  $\psi(\lambda) < \infty$  for all  $\lambda \geq 0$  and  $\mathbb{P}(X_1 > a) > 0$ . Fix  $\epsilon > 0$  and set  $b = a + \epsilon$ ,  $c = a + 2\epsilon$ , choosing  $\epsilon$  small enough so  $\mathbb{P}(X_1 > b) > 0$ . Then there exists  $\lambda$  such that  $\psi'(\lambda) = b$  - where the differentiability and the existence is justified in the following proposition :

**Proposition 6.1.2)** Suppose  $X$  is integrable and not a.s. constant. Then

$$\begin{aligned} \psi_K(\lambda) &= \log \mathbb{E}(e^{\lambda X_1} | X_1 \leq K) < \infty \quad \forall K < \infty \\ \text{and } \psi_K(\lambda) &\nearrow \psi(\lambda) \quad \text{as } K \rightarrow \infty \end{aligned}$$

Moreover in the case  $\psi(\lambda) < \infty$  for all  $\lambda \geq 0$ ,  $\psi$  has a continuous derivative on  $[0, \infty)$  and is  $C^2$  on  $(0, \infty)$  with

$$\begin{aligned} \psi'(\lambda) &= \int_{\mathbb{R}} x \mu_\lambda(dx) \\ \psi''(\lambda) &= \text{Var}(\mu_\lambda) > 0 \end{aligned}$$

and  $\psi'$  is a homeomorphism from  $[0, \infty)$  to  $[m, \sup(\text{supp}(\mu))]$ .

**proof)** (Exercise)

Now we use the idea of tilting the probability measure. Define a new probability measure  $\mathbb{P}_\lambda$  by  $d\mathbb{P}_\lambda = e^{\lambda S_n - n\psi(\lambda)} d\mathbb{P}$ . Then observe that under  $\mathbb{P}_\lambda$  the random variables  $X_1, \dots, X_n$  are independent with distributions  $\mu_\lambda$  and that  $\mathbb{E}_\lambda(X_1) = b$ . Consider the event

$$A_n = \left\{ \left| \frac{S_n}{n} - b \right| \leq \epsilon \right\} = \{(b - \epsilon)n = an \leq S_n \leq (b + \epsilon)n = cn\}$$

By the weak law of large numbers,  $\mathbb{P}_\lambda(A_n) \rightarrow 1$ . So

$$\begin{aligned} \mathbb{P}(S_n \geq na) &\geq \mathbb{P}(A_n) = \mathbb{E}_\lambda \left( 1_{A_n} e^{-\lambda S_n + \psi(\lambda)n} \right) \\ &\geq e^{-\lambda cn + \psi(\lambda)n} \mathbb{P}_\lambda(A_n) \end{aligned}$$

So

$$\frac{1}{n} \log \mathbb{P}(S_n \geq na) \geq -\lambda c + \psi(\lambda) + \frac{\log(\mathbb{P}_\lambda(A_n))}{n}$$

and

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) \geq -(\lambda c - \psi(\lambda)) \geq -\psi^*(c)$$

Now  $\psi^*$  is continuous at  $a$  (recall,  $\psi^*$  is a Legendre transform of a convex function so is convex, and therefore continuous. Or, see **Lemma 6.1.3**) and  $\epsilon > 0$  is arbitrary so the desired lower bound follows on letting  $\epsilon \rightarrow 0$ .

Finally, consider the general case  $\mathbb{P}(X_1 > a) > 0$  but allowing  $\psi(\lambda) = \infty$  for some  $\lambda \geq 0$ . For  $K > a$ , we have  $\mathbb{P}(X_1 > a | X_1 \leq K) > 0$  and  $\psi_K(\lambda) < \infty$  for all  $\lambda \geq 0$ . So preceding argument shows

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_K(S_n > na) \geq -\psi_K^*(a)$$

where  $\mathbb{P}_K$  is the probability measure given by

$$d\mathbb{P}_K^{(n)} \propto 1_{\{X_1 \leq K, \dots, X_n \leq K\}} d\mathbb{P}$$

(To see this, note, under  $\mathbb{P}_K$ , random variables  $X_1, \dots, X_n$  are independent with distribution  $\mu(\cdot | x \leq K)$ ). But

$$\mathbb{P}(S_n \geq na) \geq \mathbb{P}(S_n \geq na | X_1 \leq K, \dots, X_n \leq K) = \mathbb{P}_K(S_n \geq na)$$

and  $\psi_K^*(a) \searrow \psi^*(a)$  as  $K \rightarrow \infty$  (by **Lemma 6.1.3**) so we see

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) \geq -\psi_K^*(a) \nearrow -\psi^*(a)$$

(End of proof)  $\square$

One different way to see that  $\psi^*$  is continuous at  $a$  is presented in the following lemma.

**Lemma 6.1.3)** For all  $a \geq m$ , with  $\mathbb{P}(X_1 > 0) > 0$  we have  $\psi_K^*(a) \searrow \psi^*(a)$  as  $K \rightarrow \infty$ . Moreover in the case  $\psi(\lambda) < \infty$  for all  $\lambda \geq 0$ ,  $\psi^*$  is continuous at  $a$  and we have  $\psi^*(a) = \lambda^* a - \psi(\lambda^*)$  where  $\lambda^*$  is uniquely determined by  $\psi'(\lambda^*) = a$ .

**proof)** Consider first the later case where  $\psi(\lambda) < \infty$  for  $\lambda \geq 0$ . Then by **Proposition 6.1.2** we see that

$$\psi^*(a) = \lambda^* a - \psi(\lambda^*)$$

where  $a = \psi'(\lambda^*)$  and  $\psi^*$  is continuous at  $a$  with  $\lambda^* = (\psi')^{-1}(a)$ .

For the first part, note that  $\psi_K^*$  is non-increasing in  $K$ . For  $K$  sufficiently large, we have

$$\mathbb{P}(X_1 > a | X_1 \leq K) > 0$$

and  $a \geq m \geq m_K$  (where  $m_K = \mathbb{E}(X_1 | \leq X_1 \leq K)$ ) and  $\psi_K(\lambda) < \infty$  for all  $\lambda \geq 0$ , so we may apply the preceding argument to  $\mu_K$  to see that

$$\psi_K^*(a) = \lambda_K^* a - \psi_K(\lambda_K^*)$$

where  $\lambda_K^* \geq 0$  is determined by  $\psi'_K(\lambda_K^*) = a$ . Now  $\psi'_K(\lambda)$  is non-decreasing in  $K$  and  $\lambda$ , so  $\lambda_K^* \searrow \lambda^*$  for some  $\lambda^* \geq 0$ . Also  $\psi'_K(\lambda) \geq m_K$  for all  $\lambda \geq 0$  so

$$\psi_K(\lambda_K^*) \geq \psi_K(\lambda^*) + m_K(\lambda_K^* - \lambda^*)$$

Then

$$\psi_K^*(a) = \lambda_K^* a - \psi_K(\lambda_K^*) \leq \lambda_K^* a - \psi_K(\lambda^*) - m_K(\lambda_K^* - \lambda^*) \rightarrow \lambda^* a - \psi(\lambda^*) \leq \psi^*(a)$$

So  $\psi_K^*(a) \searrow \psi^*(a)$  as  $K \rightarrow \infty$  as claimed.

(End of proof)  $\square$

(9th November, Friday)

## 7. Brownian Motion

### 7.1. Definition

Let  $(X_t)_{t \geq 0}$  is a random process in  $\mathbb{R}^d$ . We say  $(X_t)_{t \geq 0}$  is a **Brownian motion** if :

- (i) For all  $s, t \geq 0$ , the random variable  $X_{s+t} - X_s$  is Gaussian, of mean 0 and variance  $tI$  and is independent of  $\mathcal{F}_s^X = \sigma(X_r : r \leq s)$
- (ii) for all  $\omega \in \Omega$  the map  $t \mapsto X_t(\omega) : [0, \infty) \rightarrow \mathbb{R}^d$  is *continuous*.

Condition (i) means that, for all  $s \geq 0, t > 0$ , all Borel sets  $B \subset \mathbb{R}^d$  and all  $A \in \mathcal{F}_s^X$ ,

$$\mathbb{P}(\{X_{s+t} - X_s \in B\} \cap A) = \mathbb{P}(A) \int_B (2\pi t)^{-\frac{d}{2}} e^{-|y|^2/2t} dy$$

Or, in terms of conditional expectation, (i) is equivalent to : for all  $s, t \geq 0$  and all  $f \in C_b(\mathbb{R}^d)$ ,

$$\mathbb{E}(f(X)_{s+t} | \mathcal{F}_s^X) = P_t f(X_s) \quad \text{a.s.}$$

where  $P_t$  is the **heat semigroup**, i.e.

$$P_0 f(x) = \int_{\mathbb{R}^d} p(t, x, y) f(y) dy, \quad p(t, x, y) = (2\pi t)^{-\frac{d}{2}} e^{-|y-x|^2/2t}$$

If  $X_0 = x$  then we call  $(X_t)_{t \geq 0}$  a **Brownian motion starting from  $x$** . In this case, condition (i) is equivalent following property : for all  $t_1, \dots, t_n \geq 0$  with  $t_1 < \dots < t_n$  and all  $B \in \mathcal{B}(\mathbb{R}^{dn})$

$$\mathbb{P}((X_{t_1}, \dots, X_{t_n}) \in B) = \int_B \prod_{i=1}^n p(s_i, x_{i-1}, x_i) dx_i$$

where  $t_0 = 0, x_0 = x, s_i = t_i - t_{i-1}$ .

Given independent Brownian motions  $(X_t^1)_{t \geq 0}, \dots, (X_t^d)_{t \geq 0}$  in  $\mathbb{R}$  starting from 0 and given  $x = (x^1, \dots, x^d) \in \mathbb{R}^d$ , the process  $(x + (X_t^1, \dots, X_t^d))_{t \geq 0}$  is a Brownian motion in  $\mathbb{R}^d$  starting from  $x$  and we obtain all Brownian motion starting from  $x$  in  $\mathbb{R}^d$  in this way.

## 7.2. Wiener's theorem

Brownian motion was established as a mathematical object only after 1920's.

Let  $W_d = C([0, \infty), \mathbb{R}^d)$ , and  $x_t : W_d \rightarrow \mathbb{R}^d$ ,  $x_t(w) = w(t)$  be the coordinate functions. We may endow  $W_d$  with  $\sigma$ -algebra  $\mathcal{W}_d = \sigma(x_t : t \geq 0)$ .

Given a continuous process  $(X_t)_{t \geq 0}$  in  $\mathbb{R}^d$  on  $\Omega$ , we can define

$$X : \Omega \rightarrow W_d, \quad X(\omega)(t) = X_t(\omega)$$

then  $X$  is  $\mathcal{W}_d$ -measurable so  $X$  has a law on  $(W_d, \mathcal{W}_d)$ .

**Theorem 7.2.1.)** (*Wiener*) For all  $d \geq 1$  and  $x \in \mathbb{R}^d$ , there exist a unique probability measure  $\mu_x$  on  $(W_d, \mathcal{W}_d)$  such that  $(x_t)_{t \geq 0}$  is a Brownian motion in  $\mathbb{R}^d$  starting from  $x$ . In particular, Brownian motion exists.

**proof)** Conditions (i) and (ii) determine the finite dimensional distributions of a Brownian motion and hence determine the law of any BM on  $(W_d, \mathcal{W}_d)$  (with given starting point - hence such probability measure is unique.

Suppose we have a measure  $\mu_0$  on  $(W_1, \mathcal{W}_1)$  such that  $(x_t)_{t \geq 0} \sim \text{BM}_0$  in  $\mathbb{R}$ . For  $x \in \mathbb{R}$ ,  $(x + x_t)_{t \geq 0} \sim \text{BM}_x$  so could take  $\mu_x$  as law of this process. Then for  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ , the measure  $\mu_{x_1} \otimes \dots \otimes \mu_{x_d}$  has required properties. So we only have to work in 1 dimension, starting at 0.

Define  $\mathbb{D}_N = \{k2^{-N} : k \in \mathbb{Z}^+\}$  and  $\mathbb{D} = \bigcup_{N \geq 0} \mathbb{D}_N$ . There exists some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and a family  $(Y_t : t \in \mathbb{D})$  of independent  $N(0, 1)$  random variables. First define for  $t \in \mathbb{D}_0 = \mathbb{Z}^+$ ,

$$\xi_t = Y_1 + \dots + Y_t$$

Then  $(\xi_n)_{n \in \mathbb{D}_0}$  is Gaussian and  $(\xi_{t+1} - \xi_t : t \in \mathbb{D}_0)$  are independent and has distribution  $\sim N(0, 1)$ . We define recursively  $(\xi_t)_{t \in \mathbb{D}_N}$  as follows for  $t \in \mathbb{D}_{N+1} \setminus \mathbb{D}_N$  :

: set  $r = t - 2^{-N-1}$ ,  $s = t + 2^{-N-1} \in \mathbb{D}_N$ , set  $Z_t = 2^{-\frac{N+2}{2}} Y_t$  and define  $\xi_t = \frac{\xi_r + \xi_s}{2} + Z_t$ .

We will show by induction that for all  $N \geq 0$ ,  $(\xi_{t+2^{-N}} - \xi_t : t \in \mathbb{D}_N)$  are independent,  $\sim N(0, 2^{-N})$  random variables

: Suppose true for  $N$ . Take  $t \in \mathbb{D}_{N+1} - \mathbb{D}_N$  and  $r, s$  as above. Then

$$\begin{aligned} \xi_t - \xi_r &= \frac{\xi_s - \xi_r}{2} + Z_t, & \xi_s - \xi_t &= \frac{\xi_s - \xi_r}{2} - Z_t \\ \text{Var}\left(\frac{\xi_s - \xi_r}{2}\right) &= \frac{1}{4}2^{-N}, & \text{Var}(Z_t) &= 2^{-N-2} \end{aligned}$$

so

$$\begin{aligned} \text{Var}(\xi_t - \xi_r) &= \frac{1}{4}2^{-N} + 2^{-N-2} = 2^{-N-1} = \text{Var}(\xi_s - \xi_r) \\ \text{cov}(\xi_t - \xi_r, \xi_s - \xi_t) &= 0 \end{aligned}$$

Also for any interval  $(u, v]$  disjoint from  $(r, s]$  with  $u, v \in \mathbb{D}_{N+1}$ ,

$$\text{cov}(\xi_s - \xi_r, \xi_v - \xi_u) = \text{cov}(\xi_s - \xi_t, \xi_v - \xi_u) = 0$$

So the induction proceeds.



**proof of Wiener's theorem continues)** We constructed  $(\xi_t : t \in \mathbb{D})$  such that for all  $s, t \geq 0$ ,  $\xi_{s+t} - \xi_s \sim N(0, t)$  and is independent of  $\sigma(\xi_r : r \leq s, r \in \mathbb{D})$ . Choose  $p > 2$  and set  $C_p = \mathbb{E}(|\xi_1|^p)$ . Then  $C_p < \infty$  and  $\mathbb{E}(|\xi_t - \xi_s|^p) = C_p |t - s|^{p/2}$  so by Kolmogorov's lemma (**Theorem 4.2.1**), there exists a continuous process  $(X_t)_{t \geq 0}$  such that  $X_t = \xi_t$  a.s. for all  $t \in \mathbb{D}$ .

For  $s \geq 0, t > 0$  and for any  $A \in \mathcal{F}_s^X$  there exist sequence  $(s_n)$  in  $[0, \infty)$ ,  $(t_n)$  in  $(0, \infty)$  and  $A_0 \in \sigma(\xi_r : r \leq s, r \in \mathbb{D})$  s.t.  $s_n \rightarrow s, t_n \rightarrow t$  as  $n \rightarrow \infty$  and  $1_A = 1_{A_0}$  a.s. (to check this, consider  $\{A \in \mathcal{F}_s : \text{this holds}\}$ ). Then for any  $f \in C_b(\mathbb{R})$ ,

$$\begin{aligned} \mathbb{E}(f(X_{s_n+t_n} - X_{s_n})1_A) &= \mathbb{E}(f(\xi_{s_n+t_n} - \xi_{s_n})1_{A_0}) \\ &= \mathbb{P}(A_0) \int_{\mathbb{R}} p(t_0, 0, y) f(y) dy \end{aligned}$$

so letting  $n \rightarrow \infty$  using bounded convergence theorem gives

$$\mathbb{E}(f(X_{t+s} - X_s)1_A) = \mathbb{P}(A) \int_{\mathbb{R}} p(t, 0, y) f(y) dy$$

so  $X$  is a Brownian motion as required.

(End of proof)  $\square$

### 7.3. Symmetries of Brownian Motion

**Proposition 7.3.1** Let  $(X_t)_{t \geq 0}$  be a  $\text{BM}_0(\mathbb{R}^d)$  and let  $\sigma \in (0, \infty)$  and  $U \in O(d)$ . Then the following processes are also  $\text{BM}_0(\mathbb{R}^d)$ .

- (i) **(Scaling property)**  $(\sigma X_{\sigma^{-2}t})_{t \geq 0}$ ,
- (ii) **(Rotation invariance)**  $(UX_t)_{t \geq 0}$ .

In fact  $\text{BM}_0(\mathbb{R}^d)$  is characterized among continuous Gaussian processes by its means and covariances,

$$\mathbb{E}(X_t) = 0, \quad \text{cov}(X_s^i, X_t^j) = \mathbb{E}(X_s^i X_t^j) = \delta_{ij}(s \wedge t)$$

**proof)** Exercise.

### 7.4. Brownian Motion in a Given Filtration

Suppose given a filtration  $(\mathcal{F}_t)_{t \geq 0}$  on  $(\Omega, \mathcal{F}, \mathbb{P})$ .

- We say that  $(X_t)_{t \geq 0}$  is a  $(\mathcal{F}_t)_{t \geq 0}$ -**BM** if
  - (a)  $X_t$  is  $\mathcal{F}_t$ -measurable,
  - (b) for all  $s, t \geq 0$ , the random variable  $X_{s+t} - X_s \sim N(0, tI)$  and is independent of  $\mathcal{F}_s$ ,
  - (c) for all  $\omega \in \Omega$ , the map  $t \mapsto (X_t(\omega) : [0, \infty) \rightarrow \mathbb{R}^d)$  is continuous.

This implies that  $X$  is a BM in the old sense and any process  $X$  which is a BM in the old sense is an  $(\mathcal{F}_t^X)_{t \geq 0}$ -BM.

**Proposition 7.4.1)** Let  $X = (X_t)_{t \geq 0}$  be a  $\text{BM}(\mathbb{R}^d)$  and let  $F$  be a bounded measurable function on  $W_d$ . Define

$$f(x) = \int F(\omega) \mu_x(d\omega), \quad x \in \mathbb{R}^d$$

Then  $f$  is measurable on  $\mathbb{R}^d$  and  $\mathbb{E}(F(X)|\mathcal{F}_0) = f(X_0)$  a.s. (recall,  $\mu_x$  is the law of  $\text{BM}_x$ )

**proof)** (Was an exercise.) First consider  $F$  of form  $1_{A^{(a,\eta)}}$  where  $A^{(a,\eta)} = \{g \in W_d : g(a) \geq (\eta_1, \dots, \eta_d)\} \subset W_d$ . Then

$$\int A^{(a,\eta)}(\omega) \mu_x(d\omega) = \mathbb{P}(B^{(x)}(a) \geq \eta) = \int_{\eta_1}^{\infty} \cdots \int_{\eta_d}^{\infty} \frac{1}{(2\pi)^{d/2}} e^{-y^2/2} d^d y$$

where  $B^{(x)}$  is a Brownian motion starting at  $x$ , so this is measurable. Also, we easily see that any finite intersection of sets  $A^{(a,\eta)}$  if of form

$$B_{(\eta_1, \dots, \eta_m)}^{(a_1, \dots, a_m)} = \{g \in W_d : g(a_1) \geq \eta_1, \dots, g(a_1 + \dots + a_m) \geq \eta_m\}$$

with  $a_1, \dots, a_m \geq 0$  so

$$\int B_{(\eta_1, \dots, \eta_m)}^{(a_1, \dots, a_m)} \mu_x(d\omega) = \int A^{(a_1, \eta_1)}(\omega) \mu_x(d\omega) \prod_{k=2}^m \int A^{(a_k, \eta_k - \eta_{k-1})}(\omega) \mu_0(d\omega)$$

is measurable. We also see that  $\mathbb{E}(F(X)|\mathcal{F}_0) = f(X_0)$  a.s. for each case  $F = 1_{A^{(a,\eta)}}$  or  $F = 1_{B_{(\eta_1, \dots, \eta_m)}^{(a_1, \dots, a_m)}}$ .

Now notice that the sets  $(B_{(\eta_1, \dots, \eta_m)}^{(a_1, \dots, a_m)} : m \geq 1, a_j \geq 0, \eta_j \in \mathbb{R})$  forms a  $\pi$ -system generating the  $\sigma$ -algebra  $\mathcal{W}_d$ . Since any limit of measurable functions is measurable, and  $\mathbb{E}(F_n(X)|\mathcal{F}_0) = f_n(X_0) \forall n$ ,  $F_n \nearrow F$  implies  $\mathbb{E}(F(X)|\mathcal{F}_0) = f(X_0)$ , we see from monotone class theorem that the statement also holds for general bounded measurable functions  $F$ .

(End of proof)  $\square$

## 7.5. Martingales of BM

**Theorem 7.5.1)** Let  $(X_t)_{t \geq 0}$  be an  $(\mathcal{F}_t)_{t \geq 0}$ -BM in  $\mathbb{R}^d$  and let  $f \in C_b^2(\mathbb{R}^d)$ . Define

$$M_t = f(X_t) - f(X_0) - \int_0^t \frac{1}{2} \Delta f(X_s) ds$$

Then  $(M_t)_{t \geq 0}$  is an  $(\mathcal{F}_t)_{t \geq 0}$ -martingale.

There are two ways for doing this. One is to use stochastic calculus, and the other is to do Markovian-way(which would be made clear soon). You will see the first method in Stochastic Calculus course next term, and we are going to prove this in the second way.

**proof)** It is clear that  $(M_t)_{t \geq 0}$  is adapted and integrable. Consider for now the case  $X_0 = x \in \mathbb{R}^d$ . Set

$$m(x) = \mathbb{E}(M_t) = \int_{W_d} \left[ f(\omega(t)) - f(\omega(0)) - \int_0^t \frac{1}{2} \Delta f(\omega(s)) ds \right] \mu_x(d\omega)$$

Fix  $s \in (0, t]$ . We have

$$M_t - M_s = f(X_t) - f(X_s) - \int_s^t \frac{1}{2} \Delta f(X_r) dr$$

so using Fubini,

$$\mathbb{E}(M_t - M_s) = \int_{\mathbb{R}^d} p(t, x, y) f(y) dy - \int p(s, x, y) f(y) dy - \int_s^t \int_{\mathbb{R}^d} p(r, x, y) \times \frac{1}{2} \Delta f(y) dy$$

But since  $p$  is a heat kernel, has  $\frac{d}{dt}p = \frac{1}{2} \Delta p$  so

$$\begin{aligned} \int_{\mathbb{R}^d} p(r, x, y) \times \frac{1}{2} \Delta f(y) dy &= \int_{\mathbb{R}} \frac{1}{2} \Delta_y p(r, x, y) f(y) dy \\ &= \int_{\mathbb{R}} \frac{d}{dt} p(r, x, y) f(y) dy \end{aligned}$$

and

$$\begin{aligned} \int_s^t \int_{\mathbb{R}^d} p(r, x, y) \times \frac{1}{2} \Delta f(y) dy dr &= \int_s^t \int_{\mathbb{R}} \frac{d}{dt} p(r, x, y) f(y) dy dr \\ &= \int_{\mathbb{R}^d} (p(t, x, y) - p(s, x, y)) dy \end{aligned}$$

so  $\mathbb{E}(M_t - M_s) = 0$ . But  $M_s \rightarrow 0$  a.s. as  $s \rightarrow 0$  so by bounded convergence,  $\mathbb{E}(M_s) \rightarrow 0$ . Hence  $m(x) = \mathbb{E}(M_t) = 0$  for all  $x$ .

Return to the case of general  $X_0$ . By **Prop 7.4.1**,

$$\mathbb{E}(M_t | \mathcal{F}_0) = m(X_0) = 0 \quad \text{a.s.}$$

Now for any  $s, t \geq 0$ ,

$$M_{s+t} - M_s = f(X_{s+t}) - f(X_s) - \frac{1}{2} \int_0^s \Delta f(X_{s+r}) dr$$

and  $(X_{s+t})_{t \geq 0}$  is an  $(\mathcal{F}_{s+t})_{t \geq 0}$ -BM. So we see that

$$\mathbb{E}(M_{s+t} - M_s | \mathcal{F}_s) = 0 \quad \text{a.s.}$$

as required.

(End of proof)  $\square$

We are in fact apply the theorem in the case where the assumption  $f \in C_b^2(\mathbb{R}^d)$  does not hold. We can actually relax this condition using almost the same proof, e.g. any  $f$  of polynomial growth rate would work.

**Exercise :** see how much we can relax  $C_b^2(\mathbb{R}^d)$ .

-Every step other than the integration by parts  $\int_{\mathbb{R}^d} p(r, x, y) \times \frac{1}{2} \Delta f(y) dy = \int_{\mathbb{R}} \frac{1}{2} \Delta_y p(r, x, y) f(y) dy$  works without assumption that  $f \in C_b^2(\mathbb{R}^d)$ . Hence, it is enough to have  $f \in W^{2,1}(\mathbb{R}^d)$  (the Sobolev space) with sub-exponential growth rate.

---

(14th November, Wednesday)

**Examples :** Let  $(X_t^i)_{t \geq 0}$ ,  $i = 1, \dots, n$  be a BM.

- $f(x) = x^i$  then  $\Delta f = 0$ , so is a martingale.
- $f(x) = x^i x^j$  then  $\Delta f = 0$ , so is a martingale.
- $(X_t^i X_t^j - \delta_{ij} t)_{t \geq 0}$  is a martingale.
- $\exp(\lambda X_t - \frac{1}{2} \lambda^2 t)_{t \geq 0}$  is a martingale.
- For all harmonic functions  $f$  bounded,  $(f(X_t))_{t \geq 0}$  is a martingale.

## 7.6. Strong Markov Property

Strong Markov Property of Brownian Motion would be particularly useful.

**Theorem 7.6.1** Let  $(X_t)_{t \geq 0}$  be an  $(\mathcal{F}_t)_{t \geq 0}$ -BM and let  $T$  be a stopping time. Then conditioned on  $\{T < \infty\}$ ,  $(X_{T+t})_{t \geq 0}$  is an  $(\mathcal{F}_{T+t})_{t \geq 0}$ -BM.

**proof** Clearly on  $\{T < \infty\}$ ,  $(X_{T+t})_{t \geq 0}$  is continuous on  $[0, \infty)$ . Also on  $\{T < \infty\}$ ,  $X_{T+t}$  is  $\mathcal{F}_{T+t}$ -measurable. Fix  $s \geq 0$ ,  $t > 0$ ,  $f \in C_b(\mathbb{R}^d)$ ,  $m, n \in \mathbb{N}$ ,  $A \in \mathcal{F}_{T+s}$  with  $A \subset \{T \leq m\}$ . For  $k = 0, 1, \dots, 2^n m$ , set  $t_k = k2^{-n}$  and  $A_k = A \cap \{t_k - 2^{-n} < T \leq t_k\}$ . Note  $A_k \in \mathcal{F}_{t_k+s}$  and  $A_k = A \cap \{T_n = t_k\}$  where  $T_n = 2^{-n} \lceil 2^n T \rceil$ , so

$$\begin{aligned} \mathbb{E}(f(X_{T_n+s+t})1_{A_k}) &= \mathbb{E}(f(X_{t_k+s+t})1_{A_k}) \\ &= \mathbb{E}(P_t f(X_{t_k+s})1_{A_k}) = \mathbb{E}(P_t f(X_{T_n+s})1_{A_k}) \end{aligned}$$

Now sum over  $k$  and let  $n \rightarrow \infty$  using bounded convergence to obtain

$$\mathbb{E}(f(X_{T+s+t})1_A) = \mathbb{E}(P_t f(X_{T+s})1_A)$$

But  $m$  and  $A$  were arbitrary so this implies, on  $\{T < \infty\}$ ,

$$\mathbb{E}(f(X_{T+t+s})|\mathcal{F}_{T+s}) = P_t f(X_{T+s}) \quad \text{a.s.}$$

as required.

(End of proof)  $\square$

## 7.7. Properties of 1-d BM

**Proposition 7.7.1** Let  $(X_t)_{t \geq 0}$  be a  $\text{BM}_0(\mathbb{R})$ . Set  $T_a = \inf\{t \geq 0 : X_t = a\}$ . Then

$$\begin{aligned} \mathbb{P}(T_a < \infty) &= 1 \quad \text{for all } a \in \mathbb{R} \quad \text{and} \\ \mathbb{P}(T_{-a} \leq T_b) &= \frac{b}{a+b} \quad \forall a, b \geq 0 \quad \text{and} \\ \mathbb{E}(T_a \wedge T_b) &= ab \end{aligned}$$

Moreover  $T_a$  has a density  $f_a$  on  $[0, \infty)$  given by

$$f_a(t) = \frac{1}{\sqrt{2\pi t^3}} e^{-a^2/2t} \quad t \geq 0$$

Moreover the following holds almost surely.

- (a)  $X_t/t \rightarrow 0$  as  $t \rightarrow \infty$ .
- (b)  $\inf_{t \geq 0} X_t = -\infty$ ,  $\sup_{t \geq 0} X_t = \infty$ .
- (c) for all  $s \geq 0$ , there exist  $t, n \geq s$  such that  $X_t < 0 < X_n$ .
- (d) for all  $s > 0$  there exist  $t, n \in [0, s)$  such that  $X_t < 0 < X_n$ .

**proof)** Exercise.

**Theorem 7.7.2)** Let  $X \sim \text{BM}_0(\mathbb{R})$ . Then the following properties hold almost surely :

- (a) for all  $\alpha < 1/2$ ,  $(X_t)_{t \geq 0}$  is locally Hölder continuous of exponent  $\alpha$ .
- (b) for all  $\alpha > 1/2$  there is no non-trivial interval on which  $X$  is Hölder continuous of exponent  $\alpha$ .

**proof)**

- (a) For  $s, t \geq 0$  with  $s < t$ ,  $X_t - X_s \sim (t - s)^{1/2} X_1$  so for all  $p < \infty$ ,

$$\mathbb{E}(|X_t - X_s|^p) = (t - s)^{p/2} C_p$$

where  $C_p = \mathbb{E}(|X_1|^p) < \infty$ . Given  $\alpha < 1/2$  we can find  $p < \infty$  so that  $\alpha + \frac{1}{p} < \frac{1}{2}$ . Then by Kolmogorov's lemma there exist  $K \in L^p(\mathbb{P})$  such that

$$|X_t - X_s| \leq K|t - s|^\alpha \quad \text{for all } s, t \in [0, 1]$$

So by scaling of BM,  $(X_t)_{t \geq 0}$  is locally Hölder continuous of exponent  $\alpha$  a.s.

Finally consider  $\alpha_n \nearrow 1/2$ ,  $\alpha_n < 1/2$  to see  $(X_t)_{t \geq 0}$  is Hölder continuous of exponent  $\alpha$  a.s. for all  $\alpha < 1/2$ .

- (b) Let  $m, n \in \mathbb{N}$  with  $m \geq n$ . For  $s, t \in \mathbb{D}_n$ , define

$$[X]_{s,t}^m = \sum_{\tau} (X_{\tau+2^{-m}} - X_{\tau})^2$$

where we sum over all  $\tau \in \mathbb{D}_n$  with  $\tau \in (s, t]$ . The random variables  $(X_{\tau+2^{-m}} - X_{\tau})^2$  are independent of mean  $2^{-m}$  and variance  $2^{-2m+1}$  (Here we need  $\mathbb{E}(X_1^4) = 3$ ,  $\text{Var}(X_1^2) = 2$ ). So  $\mathbb{E}([X]_{s,t}^m) = 2^m(t - s)2^{-m} = t - s$  and  $\text{Var}([X]_{s,t}^m) = 2^m(t - s)2^{-2m+1}$ . So  $[X]_{s,t}^m \rightarrow t - s$  a.s.

Suppose  $X$  is Hölder  $\alpha$  on  $[s, t]$  then for some constant  $K$ ,

$$|X_{\tau+2^{-m}} - X_{\tau}|^2 \leq K^2(2^{-m})^{2\alpha}$$

so  $[X]_{s,t}^m \leq 2^m(t - s)K^22^{-2m\alpha} \rightarrow 0$  as  $m \rightarrow \infty$ . So  $(X_t)_{t \geq 0}$  is not Hölder  $\alpha$  on  $[s, t]$ , which contradicts with our earlier result.

(End of proof)  $\square$

## 7.8. Recurrence and Transience of Brownian Motion

The following theorem tells about recurrence and transience in different dimensions.

**Theorem 7.8.1)** Let  $(X_t)_{t \geq 0}$  is a  $\mathbf{BM}_0(\mathbb{R}^d)$ .

- (a) For  $d = 1$ ,  $\mathbb{P}(\{t \geq 0 : X_t = 0\} \text{ is unbounded}) = 1$ . (**Point recurrent**)
- (b) For  $d = 2$ , and all  $\epsilon > 0$ ,  $\mathbb{P}(X_t = 0 \text{ for some } t > 0) = 0$  but  $\mathbb{P}(\{t \geq 0 : |X_t| \leq \epsilon\} \text{ is unbounded}) = 1$ . (**Neighbourhood recurrent**, but not recurrent)
- (c) For  $d \geq 3$ ,  $\mathbb{P}(|X_t| \rightarrow \infty \text{ as } t \rightarrow \infty) = 1$ . (**Transient**)

**proof)**

(a) Is direct from **Prop 7.7.1 (b)**

(b) Note that  $x \mapsto \log |x|$  is harmonic in  $\mathbb{R}^2 \setminus \{0\}$ . Let  $(X_t)_{t \geq 0} \sim \mathbf{BM}_x(\mathbb{R}^2)$ . For  $\epsilon > 0$ , set

$$p_0(x) = \mathbb{P}(X_t = 0 \text{ for some } t > 0), \quad p_\epsilon(x) = \mathbb{P}(|X_t| \leq \epsilon \text{ for some } t > 0)$$

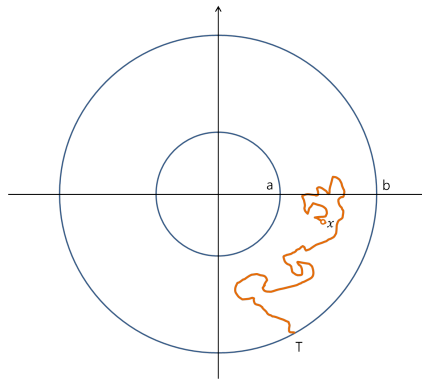
Consider the case  $|x| = 1$ . For  $a < 1 < b$ , there is a function  $f \in C_b^2(\mathbb{R}^2)$  such that  $f(x) = \log |x|$  for  $|x| \in [a, b]$ . Also consider the martingale (a martingale by **Theorem 7.5.1**)

$$M_t = f(X_t) - f(X_0) - \int_0^t \frac{1}{2} \Delta f(X_s) ds$$

and the stopping time

$$T = \inf\{t \geq 0 : |X_t| \in \{a, b\}\}$$

Note that  $\mathbb{P}(T < \infty) = 1$ . Also for  $t \leq T$ ,  $M_t = \log |X_t|$  so  $M$  is bounded up to



$T$ . So optional stopping theorem applies to show

$$0 = \mathbb{E}(M_0) = \mathbb{E}(M_T) = \mathbb{E}(\log |X_T|) = p \log a + (1 - p) \log b$$

where  $p = p(a, b) = \mathbb{P}(|X_T| = a)$ . Consider the limit  $a \rightarrow 0$ , then  $\log a \rightarrow -\infty$ , so  $p \rightarrow 0$ . Hence for  $|x| = 1$ ,  $p_0(x) = \lim_{b \rightarrow \infty} \lim_{a \rightarrow 0} p(a, b) = 0$ . By scaling, we see  $p_0(x) = 0$  for all  $x \neq 0$ .

In the case  $x = 0$ , for  $n \in \mathbb{N}$ , by Markov property,

$$\mathbb{P}(|X_t| = 0 \text{ for some } t \geq \frac{1}{n}) = \int_{\mathbb{R}^2} p(\frac{1}{n}, 0, y) p_0(y) dy = 0$$

Let  $n \rightarrow \infty$  to see  $p_0(0) = 0$ .

For  $|x| = 1$ , let  $b \rightarrow \infty$  with  $a = \epsilon$  fixed. Since  $\log b \rightarrow \infty$ ,  $p(a, b) \rightarrow 1$  so

$$p_\epsilon(x) = \lim_{b \rightarrow \infty} p(\epsilon, b) = 1.$$

By scaling  $p_\epsilon(x) = 1$  for all  $x \neq 0$ .

If  $x = 0$ , for  $n \in \mathbb{N}$ , by Markov property

$$\mathbb{P}(|X_t| \leq \epsilon \text{ for some } t \geq n) = \int_{\mathbb{R}^2} p(n, x, y) p_\epsilon(y) dy = 1$$

Hence  $\mathbb{P}(\{t \geq 0, |X_t| \leq \epsilon\} \text{ is unbounded}) = 1$ .

(c) It will suffice to consider case  $d = 3$  and to show for all  $N \in \mathbb{N}$  that

$$\mathbb{P}(\{t \geq 0 : |X_t| \leq N\} \text{ is unbounded}) = 0$$

We repeat the argument from (b) with  $1/|x|$  in place of  $\log(|x|)$  - in  $d = 3$ ,  $1/|x|$  is the harmonic function on  $\mathbb{R}^3 \setminus \{0\}$  - to obtain

$$\frac{p}{a} + \frac{1-p}{b} = 1$$

for  $p = \mathbb{P}(|X_T| = a)$  and  $(X_T)_{t \geq 0} \sim \text{BM}_x(\mathbb{R}^3)$ ,  $|x| = 1$ . Let  $b \rightarrow \infty$  to see

$$\lim_{b \rightarrow \infty} p(a, b) = \mathbb{P}(|X_t| \leq a \text{ for some } t \geq 0) = a$$

By scaling, for  $|x| = N + 1$ ,

$$\mathbb{P}(|X_t| \leq N \text{ for some } t \geq 0) = \frac{N}{N+1}$$

Set  $T_0 = 0$  and for  $k \geq 1$ ,

$$S_k = \inf\{t \geq T_{k-1} : |X_t| = N + 1\}, \quad T_k = \inf\{t \geq S_k : |X_t| = N\}$$

Then

$$\mathbb{P}(T_k < \infty) \leq \frac{N}{N+1} \mathbb{P}(T_{k-1} < \infty) \leq \left(\frac{N}{N+1}\right)^k \rightarrow 0$$

by strong Markov property and induction. Set  $K = \sup\{k : T_k < \infty\}$ . Then  $K < \infty$  a.s. Now for all  $k \geq 1$ , on  $\{T_{k-1} < \infty\}$ , has  $S_k < \infty$  a.s. Hence  $S_{K+1} < \infty$  a.s. But on  $\{S_{K+1} \leq \infty\}$ , has  $|X_t| > N$  for all  $t \geq S_{K+1}$  hence  $\mathbb{P}(\{t \geq 0 : |X_t| \leq N\} \text{ is unbounded}) = 0$ .

(End of proof)  $\square$

## 7.9. Brownian Motion and the Dirichlet Problem

Let  $D$  be a connected open set in  $\mathbb{R}^d$ . Write  $\partial D$  for the boundary. Let  $c : D \rightarrow [0, \infty)$ ,  $f : \partial D \rightarrow [0, \infty)$  be measurable functions. For  $x \in \overline{D}$ , let  $(X_t)_{t \geq 0}$  be a Brownian motion starting at  $x$ .

Define the **expected total cost** by

$$\phi(x) = \mathbb{E} \left( \int_0^T c(X_t) dt + f(X_T) 1_{T < \infty} \right)$$

where  $T = \inf\{t \geq 0 : X_t \notin D\}$ . By a

By a **solution of the Dirichlet problem (in  $D$  with data  $c$  and  $f$ )**, we mean a function  $\psi \in C(\overline{D}) \cap C^2(D)$  such that

$$\begin{cases} -\frac{1}{2}\Delta\psi = c & \text{in } D \\ \psi = f & \text{on } \partial D \end{cases}$$

Accordingly, say  $\psi$  is a **super-solution** if  $-\frac{1}{2}\Delta\psi \geq c$ ,  $\psi \geq f$ , and call  $\psi$  a **sub-solution** with inverse inequalities.

Suppose  $\psi \in C_b^2(\mathbb{R}^d)$  restrict to a solution of Dirichlet problem on  $D$ . Consider the martingale (as in **Theorem 7.5.1**)

$$M_t = \psi(X_t) - \psi(X_0) - \int_0^t \frac{1}{2} \Delta\psi(X_s) ds$$

Suppose we could justify optional stopping theorem at  $T$  and suppose  $T < \infty$  a.s. Then

$$0 = \mathbb{E}(M_0) = \mathbb{E}(M_T) = \mathbb{E}(\phi(X_T)) - \psi(x) - \mathbb{E} \left( \int_0^T \frac{1}{2} \Delta\psi(X_s) ds \right)$$

so

$$\psi(x) = \mathbb{E}(f(X_T)) + \mathbb{E} \left( \int_0^T c(X_s) ds \right) = \phi(x)$$

We will spend time in exploring the condition in which this holds.

(19th November, Monday)

Settings from the last lecture :  $D$  a connected open set  $\subset \mathbb{R}^d$ ,  $c : D \rightarrow [0, \infty)$ ,  $f : \partial D \rightarrow [0, \infty)$  measurable functions,

$$\phi(x) = \mathbb{E} \left[ \int_0^T c(X_t) dt + f(X_T) 1_{T < \infty} \right]$$

$(X_t)_{t \geq 0} \sim \text{BM}_x(\mathbb{R}^d)$ ,  $T = \inf\{t \geq 0 : X_t \notin D\}$ , times spent in  $B \subset D$  before leaving.

**Theorem 7.9.2 (a)** Let  $\psi$  be a *non-negative super-solution* of the Dirichlet Problem(DP),  $\psi \in C(\overline{D}) \cap C^2(D)$ ,

$$\begin{cases} -\frac{1}{2}\Delta\psi \geq c & \text{in } D \\ \psi \geq f & \text{on } \partial D \end{cases}$$

Then  $\psi \geq \phi$ .



**proof)** Clear that  $\psi \geq \phi$  on  $\partial D$ . Fix  $x \in D$  and  $N \geq 1$  such that  $x \in D_N = \{y \in D : |y| < N, \text{dist}(y, \partial D) > 1/N\}$ . There exists  $g \in C_b^2(\mathbb{R}^d)$  such that  $g = \psi$  on  $D_N$ . Consider

$$M_t = g(X_t) - g(X_0) - \int_0^t \frac{1}{2} \Delta g(X_s) ds$$

and  $T_N = \inf\{t \geq 0 : X_t \notin D_N\}$ . Then  $M$  is a martingale. By the optional stopping theorem, has

$$0 = \mathbb{E}(M_0) = \mathbb{E}(M_{T_N \wedge t}) = \mathbb{E}(\psi(X_{T_N \wedge t})) - \psi(x) + \mathbb{E}\left[\int_0^{T_N \wedge t} \left(-\frac{1}{2} \Delta \psi\right)(X_s) ds\right] \quad \dots\dots\dots (\star)$$

Consider the limit  $N \rightarrow \infty$  and then  $t \rightarrow \infty$ . Then  $T_N \wedge t \nearrow T$ , hence

$$\mathbb{E} \int_0^{T_N \wedge t} \left(-\frac{1}{2} \Delta \psi\right)(X_s) ds \geq \mathbb{E} \int_0^{T_N \wedge t} c(X_s) ds \rightarrow \mathbb{E} \int_0^T c(X_s) ds$$

by monotone convergence. On  $\{T < \infty\}$ ,

$$\psi(X_{T_N \wedge t}) \rightarrow \psi(X_T) \quad \text{a.s.}$$

so since  $\psi \geq 0$ , by Fatou's lemma,

$$\liminf \mathbb{E}(\psi(X_{T_N \wedge t})) \geq \mathbb{E}(\psi(X_T) 1_{T < \infty}) \geq \mathbb{E}(f(X_T) 1_{T < \infty})$$

So taking  $\liminf$  in  $(\star)$ , we see  $\psi \geq \phi$ .

(End of proof)  $\square$

**Theorem 7.9.2 (b))** Let  $\psi$  be a *non-negative bounded solution* of the Dirichlet Problem,  $\psi \in C(\overline{D}) \cap C^2(D)$

$$\begin{cases} -\frac{1}{2} \Delta \psi = c & \text{in } D \\ \psi = f & \text{on } \partial D \end{cases}$$

and suppose that  $\mathbb{E}(\psi(X_t) 1_{t < T}) \rightarrow 0$  as  $t \rightarrow \infty (\dots (\dagger))$ . Then  $\psi = \phi$ .

**Remark about the condition  $(\dagger)$  :** Since we assume  $\psi$  bounded and

$$|\mathbb{E}(\psi | X_t) 1_{t < T}| \leq \|\psi\|_\infty \mathbb{P}(t < T)$$

We see condition  $(\dagger)$  holds whenever  $T < \infty$  a.s. for all starting points  $x$ . e.g. we're OK if ( $d = 1$  and  $\partial D \neq \emptyset$ ) or ( $d = 2$  and  $\mathbb{R}^2 \setminus D$  contains an open ball) or ( $d \geq 3$  and  $D$  bounded).

**proof)** Clear that  $\psi = \phi$  on  $\partial D$ . Fix  $x \in D$  and  $N \geq 1$  such that  $x \in D_N = \{y \in D : |y| < N, \text{dist}(y, \partial D) > 1/N\}$ . There exists  $g \in C_b^2(\mathbb{R}^d)$  such that  $g = \psi$  on  $D_N$ . Consider

$$M_t = g(X_t) - g(X_0) - \int_0^t \frac{1}{2} \Delta g(X_s) ds$$

and  $T_N = \inf\{t \geq 0 : X_t \notin D_N\}$ . Then  $M$  is a martingale. By the optional stopping theorem, has

$$0 = \mathbb{E}(M_0) = \mathbb{E}(M_{T_N \wedge t}) = \mathbb{E}(\psi(X_{T_N \wedge t})) - \psi(x) + \mathbb{E}\left[\int_0^{T_N \wedge t} \left(-\frac{1}{2} \Delta \psi\right)(X_s) ds\right] \quad \dots\dots\dots (\star)$$

Consider the limit  $N \rightarrow \infty$  and then  $t \rightarrow \infty$ . Then  $T_N \wedge t \nearrow T$ , hence

$$\mathbb{E} \int_0^{T_N \wedge t} \left( -\frac{1}{2} \Delta \psi \right) (X_s) ds = \mathbb{E} \int_0^{T_N \wedge t} c(X_s) ds \rightarrow \mathbb{E} \int_0^T c(X_s) ds$$

by monotone convergence. Now

$$\mathbb{E}(\psi(X_{T_N \wedge t})) = \mathbb{E}(\psi(X_{T_N})1_{T_N \leq t}) + \mathbb{E}(\psi(X_t)1_{t < T_N})$$

We have, since  $\psi \in C(\overline{D})$ ,

$$\psi(X_{T_N})1_{T_N \leq t} \rightarrow \psi(X_T)1_{T < \infty} \quad \text{a.s.}$$

and since  $\psi$  is bounded, by dominated convergence,

$$\mathbb{E} \left[ \psi(X_{T_N})1_{T_N \leq t} \right] \rightarrow \mathbb{E}(\psi(X_T)1_{T < \infty}) = \mathbb{E}(f(X_T)1_{T < \infty})$$

Also by the assumption

$$\lim_{N \rightarrow \infty} \mathbb{E}(\psi(X_t)1_{t < T_N}) = \mathbb{E}(\psi(X_t)1_{t < T}) \rightarrow 0 \quad \text{as } t \rightarrow \infty$$

Take limit in  $(\star)$  to see  $\psi = \phi$ .

(End of proof)  $\square$

This theorem tells us about uniqueness of solution of the Dirichlet problem, but it does not say anything about existence of the solution. We are next going to examine this problem.

**Theorem 7.9.2 (c).(I)** Assume  $d \geq 3$  and  $D = \mathbb{R}^d$  and  $c$  has compact support and  $c \in C^2(D)$ . Then  $\phi \in C^2(\mathbb{R}^d)$  and  $-\frac{1}{2}\Delta\phi = c$ .

**proof)** Suppose  $g$  is continuous and of compact support on  $\mathbb{R}^d$ . Note

$$P_t g(x) = \mathbb{E} g(x + X_t) = \int_{\mathbb{R}^d} (2\pi t)^{-d/2} e^{-|x-y|^2/2t} g(y) dy$$

where  $(X_t)_{t \geq 0} \sim \text{BM}(\mathbb{R}^d)$  so

$$\begin{cases} \|P_t g\|_\infty \leq \|g\|_\infty & \text{and also} \\ \|P_t g\|_\infty \leq (2\pi t)^{-d/2} \text{vol}(\text{supp}(g)) \|g\|_\infty \end{cases} \dots\dots\dots (\diamond)$$

So splitting the integral at 1 gives

$$\mathbb{E} \int_0^t g(x + X_t) dt = \int_0^\infty P_t g(x) dt \leq (1 + \text{vol}(\text{supp}(g))) \|g\|_\infty$$

Hence also for  $\epsilon > 0$ ,

$$\mathbb{E} \int_0^\infty \sup_{|x-y| < \epsilon} |g(y + X_t)| dt < \infty$$

(note,  $g_\epsilon(x) = \sup_{|y-x| < \epsilon} |g(y)|$  is also continuous of compact support.) Now we can apply this to  $g$  and its derivative to justify differentiating

$$\phi(x) = \mathbb{E} \int_0^\infty c(x + X_t) dt$$

under the integrals to obtain  $\phi \in C^2(D)$  and

$$\Delta\phi(x) = \mathbb{E} \int_0^\infty \Delta c(x + X_t) dt = \left( \int_0^s + \int_s^t + \int_t^\infty \right) \mathbb{E}(\Delta c(x + X_r)) dr$$

for  $0 < s < t$ . Consider the limits  $s \rightarrow 0$  and  $t \rightarrow \infty$ . First and the third integrals tend to 0 using the estimates above. Now

$$\begin{aligned} \frac{1}{2} \int_s^t \mathbb{E}(\Delta c(x + X_r)) dr &= \frac{1}{2} \int_s^t \int_{\mathbb{R}^d} \Delta c(y) p(r, x, y) dy dr \\ &= \frac{1}{2} \int_s^t \int_{\mathbb{R}^d} c(y) \Delta p(t, x, y) dy dr \\ &= \int_s^t \int_{\mathbb{R}^d} c(y) \dot{p}(r, x, y) dy dr \\ &= \int_{\mathbb{R}^d} c(y) (p(t, x, y) - p(s, x, y)) dy \quad (\text{FTC}) \end{aligned}$$

But  $\int c(y) p(t, x, y) dy \rightarrow 0$  as  $t \rightarrow \infty$  and  $\int c(y) p(s, x, y) dy = \mathbb{E}(c(x + X_s)) \rightarrow c(x)$  as  $s \rightarrow 0$ . Hence  $-\frac{1}{2} \Delta\psi = c$  as required.

(End of proof)  $\square$

(21st November, Wednesday)

(part iii drop-in hours tomorrow 10:30 - 12:30).

(Bring a few different colored pens to class on Friday, if possible)

(Office hours on Monday, time and place TBD)

**Theorem 7.9.4)** Let  $D$  be a connected open set  $\subset \mathbb{R}^d$  and let  $\phi$  be a *non-negative* measurable function on  $D$ . Suppose

$$\phi(x) = \int_{S(x,p)} \phi(y) \sigma_{x,p}(dy)$$

whenever  $B(x, p) \subset D$  where  $\sigma_{x,p}$  is the uniform distribution on  $S(x, p) = \{y : |x - y| = p\}$ . Then *either*  $\phi(x) = \infty$  for all  $x$  *or*  $\phi \in C^\infty(D)$  with  $\Delta\phi = 0$ .

This would not be examinable, and will not be proved in the lecture - see online notes for proof.

**Theorem 7.9.5)** (*Blumenthal's zero-one law*) Let  $(X_t)_{t \geq 0} \sim \text{BM}_x(\mathbb{R}^d)$ . Then for all  $A \in \mathcal{F}_{0+}^X = \bigcap_{t > 0} \mathcal{F}_t^X$ , has  $\mathbb{P}(A) \in \{0, 1\}$ .

**proof)** Consider the  $\pi$ -system  $\mathcal{A} = \bigcup_{s > 0} \sigma(X_t - X_s : t \geq s)$ . Then for all  $A_0 \in \mathcal{F}_{0+}^X$  and all  $A \in \mathcal{A}$ , we have  $\mathbb{P}(A_0 \cap A) = \mathbb{P}(A_0) \mathbb{P}(A)$ . By Dynkin's lemma, this extend to all  $A \in \sigma(\mathcal{A})$  (check details).

Now for all  $s > 0, t \geq s$ ,  $X_t - X_s$  is  $\sigma(\mathcal{A})$ -measurable. But  $X_s \rightarrow 0$  as  $s \rightarrow 0$ , so  $X_t$  is  $\sigma(\mathcal{A})$ -measurable. So  $\mathcal{F}_\infty^X = \sigma(\mathcal{A}) \supset \mathcal{F}_{0+}^X$ . So for  $A \in \mathcal{F}_{0+}^X$ ,

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

and therefore  $\mathbb{P}(A) \in \{0, 1\}$ .

(End of proof)  $\square$

**Proposition 7.9.6)** (*Brownina motion enters all cones immediately*) Let  $A$  be an open subset of  $S^{d-1}$  and let  $\epsilon > 0$ . Set  $C$  a cone,

$$C = \{ty : y \in A, t \in (0, \epsilon)\}$$

Let  $X \sim \text{BM}_0(\mathbb{R}^d)$ , and define  $T_C = \inf\{t \geq 0 : X_t \in C\}$ . Then  $\mathbb{P}(T_C = 0) = 1$ .

See example sheet.

Some definition (**External cone condition, ECC**) : Let  $D \subset \mathbb{R}^d$  be open and connected. Then  $D$  satisfies ECC if for all  $x \in \partial D$ , there exist  $A$  open  $\subset S^{d-1}$ ,  $\epsilon > 0$  such that

$$\{x + ty : t \in (0, \epsilon), y \in A\} \cap D = \emptyset$$

This is weaker than the Lipschitz boundary condition.

**Theorem 7.9.2(c))** Let  $D$  be a connected open set  $\subset \mathbb{R}^d$ . Let  $c \in C^2(\mathbb{R}^d)$ ,  $f \in C(\partial D)$ . Set

$$\phi(x) = \mathbb{E}\left(\int_0^T c(X_t)dt + f(X_T)1_{T<\infty}\right) \quad \forall x \in \overline{D}$$

where  $X \sim \text{BM}_0(\mathbb{R}^d)$  and  $T = \inf\{t \geq 0 : X_t \notin D\}$ . Assume  $D$  satisfies the *external cone condition (ECC)* and that  $\phi$  is locally bounded. Then  $\phi \in C^2(D) \cap C(\overline{D})$  and  $-\frac{1}{2}\Delta\phi = c$  in  $D$  and  $\phi = f$  on  $\partial D$ . i.e.  $\phi$  solves the Dirichlet condition.

This is significant - the proof using ECC is hard to achieve just using usual theories of partial differential equations.(recall, we usually prove theorems under Lipschitz boundary condition in PDEs). Probability wins!

**Lemma 7.9.3)** Let  $D_0$  be a *bounded subset* of  $D$ . Define  $T_0 = \inf\{t \geq 0 : X_t \notin D_0\}$ ,  $x \in \overline{D}$ . Then

$$\phi(x) = \mathbb{E}\left(\int_0^{T_0} c(X_t)dt + \phi(X_{T_0})\right)$$

**proof)** Set  $\tilde{X}_t = X_{T_0+t}$ ,  $\tilde{\mathcal{F}}_t = \mathcal{F}_{T_0+t}$ , and  $\tilde{T} = \inf\{t \geq 0 : \tilde{X}_t \notin D\}$ . By *Strong Markov property*,  $(\tilde{X}_t)$  is an  $(\tilde{\mathcal{F}}_t)$ -BM. Now

$$\int_0^T c(X_t)dt + f(X_T)1_{T<\infty} = \int_0^{T_0} c(X_t)dt + \int_0^{\tilde{T}} c(\tilde{X}_t)dt + f(\tilde{X}_{\tilde{T}})1_{\tilde{T}<\infty} \quad \dots\dots\dots (\dagger)$$

Here we used that  $\tilde{T} < \infty$  iff  $T < \infty$  and  $X_T = \tilde{X}_{\tilde{T}}$ . By **Prop 7.4.1**,

$$\mathbb{E}\left(\int_0^{\tilde{T}} c(\tilde{X}_t)dt + f(\tilde{X}_{\tilde{T}})1_{\tilde{T}<\infty} | \mathcal{F}_{T_0}\right) = \phi(X_{T_0}) \quad \text{a.s.}$$

So we obtain result on taking  $\mathbb{E}$  in  $(\dagger)$

(End of proof)  $\square$

Recall, we proved **Theorem 7.9.2(c)** in a weaker setting earlier. Here, we prove in a slightly more general setting.

**proof of Theorem 7.9.2(c) II.** We show in the case  $c = 0$  that  $\phi \in C^\infty(D)$  with  $\Delta\phi = 0$  -then follows from a simple argument using symmetry of a sphere and **Theorem 7.9.4**.

See online notes.

**proof of Theorem 7.9.2(c) III.** We show for all  $y \in \partial D$  that

$$\lim_{x \rightarrow y, x \in \bar{D}} \phi(x) = \phi(y)$$

Fix  $y \in \partial D$  and choose  $U$  bounded open set  $\subset \mathbb{R}^d$  with  $u \in U$ . Set  $D_0 = D \cap U$ . By *exterior cone condition*, there exist  $A \in S^{d-1}$ ,  $\epsilon > 0$  such that  $\{y + tz : t \in (0, \epsilon), z \in A\} \cap D = \emptyset$ . Let  $(X_t)_{t \geq 0}$  be a  $\text{BM}_0(\mathbb{R}^d)$ . Set  $T_0(x) = \inf\{t \geq 0 : x + X_t \notin D\}$ . By **Proposition 7.9.6**, for  $T_C = \inf\{t \geq 0 : y + X_t \in C\}$  we have  $T_C = 0$  a.s. Now on the event  $\{T_C = 0\}$ , we have  $T_0(x) \rightarrow 0$  as  $x \rightarrow y$  ( $x$  staying in  $\bar{D}$ ) so  $X_{T_0(x)} + x \in \partial D$  eventually as  $x \rightarrow y$  and  $X_{T_0(x)} + x \rightarrow y$  as  $x \rightarrow y$ . Hence

$$\phi(X_{T_0(x)} + x) = f(X_{T_0(x)} + x) \rightarrow f(y) \quad \text{as } x \rightarrow y$$

Also by **Prop 7.7.1**,  $\mathbb{E}(\sup_{x \in D_0} T_0(x)) < \infty$ . Now

$$\phi(x) = \mathbb{E}\left(\int_0^{T_0(x)} c(x + X_t) dt + \phi(X_{T_0(x)} + x)\right)$$

Let  $x \rightarrow y$ , then using dominated convergence to see  $\phi(x) \rightarrow f(y)$  as  $x \rightarrow y$ .

(End of proof)  $\square$

(23rd November, Friday)

The following is the last part of proof of **Theorem 7.9.2(c)**

**proof of Theorem 7.9.2(c) IV.** It will suffice to consider the case where  $d \geq 3$ , or set  $\tilde{D} = D \times \mathbb{R}^2$  and extend  $c$  as a constant in new directions.

Consider for now the case  $D$  bounded. There exists  $\tilde{c} \in C^2(\mathbb{R}^d)$  of compact support with  $\tilde{c} = c$  in  $D$ . Set

$$\phi_0(x) = \mathbb{E} \int_0^\infty \tilde{c}(X_t) dt$$

By **part I**,  $\phi_0 \in C^2(\mathbb{R}^d)$  with  $-\frac{1}{2}\Delta\phi_0 = \tilde{c}$ . Now

$$\phi_0(x) = \mathbb{E}\left(\int_0^T \tilde{c}(X_t) dt + \phi_0(X_T)\right) = \phi(x) + \phi_1(x)$$

(by **Lemma 7.9.3**) where  $\phi_1(x) = \mathbb{E}(\phi_0(X_T))$ . By **part II**,  $\phi_1 \in C^\infty(D)$  and  $\Delta\phi_1 = 0$ . So  $\phi \in C^2(D)$  and  $-\frac{1}{2}\Delta\phi = c$ .

Turn to the case  $D$  unbounded. Choose any bounded open  $D_0 \subset D$ . Then

$$\phi(x) = \phi_0(x) + \phi_1(x)$$

(new  $\phi_0$  and  $\phi_1$ ) where  $\phi_0$  and  $\phi_1$  are defined by

$$\phi_0(x) = \mathbb{E}\left(\int_0^{T_0} c(X_t)dt\right), \quad \phi_1(x) = \mathbb{E}(\phi(X_{T_0}))$$

where  $T_0 = \inf\{t \geq 0 : X_t \notin D_0\}$ . We know  $\phi_0 \in C^2(D)$  and  $-\frac{1}{2}\Delta\phi_0 = c$  and since  $\phi$  is locally bounded,  $\phi_1 \in C^\infty(D_0)$  with  $\Delta\phi_1 = 0$ . Hence  $\phi \in C^2(D_0)$  and  $-\frac{1}{2}\Delta\phi = c$ .

(End of proof)  $\square$

Plan for the rest of the course : we will not cover all the materials(e.g. Donsker's invariance principle or Skorohod embedding for random walks) - do not have enough time. But will discuss about Poisson random measures.

## 8. Poisson Random Measures

Our ambition is to classify all stationary independent increments - this is what Lévy-Khinchin theorem tells us. To prove this, we need some works on Poisson random measures.

### 8.1. Construction

For  $\lambda \in (0, \infty)$ , a random variable  $X$  with values in  $\mathbb{Z}^+ \cup \{\infty\}$  has distribution  $P(\lambda)$  if

$$\mathbb{P}(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x \in \mathbb{Z}^+$$

Also conventionally write  $X \sim P(0)$  if  $X \equiv 0$ , and  $P \sim P(\infty)$  if  $X \equiv \infty$ .

**Proposition 8.1.1.)** (*Addition property*) Let  $(N_k : k \in \mathbb{N})$  be a sequence of independent random variables with  $N_k \sim P(\lambda_k)$  for all  $k$ . Then

$$\sum_k N_k \sim P\left(\sum_k \lambda_k\right)$$

**Proposition 8.1.2.)** (*Splitting property*) Suppose  $N \sim P(\lambda)$ . Let  $(Y_n : n \in \mathbb{N})$  be a sequence of i.i.d. random variables in  $\mathbb{N}$ . Set  $N_k = \sum_{n=1}^N 1_{Y_n=k}$  for  $k \geq 1$ . Then  $(N_k : k \in \mathbb{N})$  is a sequence of independent  $P(\lambda_k)$  random variables with  $\lambda_k = \lambda \mathbb{P}(Y_1 = k)$ .

The proofs are just elementary calculations, but please do them!!!

Let  $(E, \mathcal{E}, \mu)$  be a  $\sigma$ -finite measure space.

- A **Poisson random measure** of intensity  $\mu$  is a map  $M : \Omega \times \mathcal{E} \rightarrow \mathbb{Z}^+ \cup \{\infty\}$  (for  $A \in \mathcal{E}$ , think of  $M(A)$  as  $M(\{\omega\} \times A)$ , a random variable) such that for any disjoint sequence  $(A_k : k \in \mathbb{N})$  in  $\mathcal{E}$ ,

$$(i) \quad M(\cup_k A_k) = \sum_k M(A_k).$$

$$(ii) \quad M(A_k) \text{ is a } P(\mu(A_k)) \text{ random variable for all } k.$$

$$(iii) \quad \text{The random variables } M(A_k), k \in \mathbb{N} \text{ are independent.}$$

- We can construct a  $\text{PRM}(\mu)$  as follows.

First in case  $\mu(E) < \infty$ , take  $N \sim P(\lambda)$ ,  $\lambda = \mu(E)$  and take independent random variables  $(Y_n : n \in \mathbb{N})$  independent of  $N$  with distribution  $\frac{1}{\lambda}\mu$ . Define  $M(A) = \sum_{n=1}^N 1_{Y_n \in A}$ . Then  $M \sim \text{PRM}(\mu)$ , i.e.  $M(A) \sim P(\mu(A))$ . (**an exercise**)

In general, if  $\mu(E) = \infty$ , as  $E$  is  $\sigma$ -finite, there exist disjoint  $E_n \in \mathcal{E}$  with  $\cup_n E_n = E$  such that  $\mu(E_n) < \infty$  for all  $n$ . Using preceding construction, we obtain  $(M_n)_n$  independent  $\text{PRM}(\mu 1_{E_n})$ 's. Then set  $M = \sum_n M_n$ . Then  $M \sim \text{PRM}(\mu)$ . (**also an exercise**)

- The uniqueness of such map also can be shown - refer to the online lecture notes.

## 8.2. Integrals with respect to Poisson Random Measures

**Theorem 8.2.1)** Let  $M$  be a  $\text{PRM}(\mu)$  and let  $g$  be a measurable function on  $(E, \mathcal{E})$ . Assume  $\mu(E) < \infty$ . Define

$$M(g) = \begin{cases} \int_E g(y) M(dy) & \text{if } M(E) < \infty \\ 0 & \text{o/w} \end{cases}$$

Then  $M(g)$  is a well-defined random variable and

$$\mathbb{E}(e^{inM(g)}) = \exp \left[ \int_E (e^{ing(y)} - 1) \mu(dy) \right]$$

and if  $g \in L^1(\mu)$  then  $M(g) \in L^1(\mathbb{P})$  with  $\mathbb{E}(M(g)) = \mu(g)$  and  $\text{Var}(M(g)) = \mu(g^2)$ .

**proof)** Let's see these all true in the case  $M(A) = \sum_{n=1}^N 1_{Y_n \in A}$ . Then  $M(g) = \sum_{n=1}^N g(Y_n)$  and this is clearly a well-defined measurable function. We have

$$\mathbb{E}(e^{inM(g)} | N = n) = \mathbb{E}(e^{in \sum_{k=1}^n g(Y_k)}) = \prod_{k=1}^n \mathbb{E}(e^{ing(Y_k)}) = \left( \int_E e^{ing(y)} \frac{\mu(dy)}{\lambda} \right)^n$$

So

$$\begin{aligned} \mathbb{E}(e^{inM(g)}) &= \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \left( \int_E e^{ing(y)} \frac{\mu(dy)}{\lambda} \right)^n \\ &= e^{-\lambda} \exp \left( \int_E e^{ing(y)} \mu(dy) \right) \end{aligned}$$

so we obtain the average and variance of  $M(g)$  from the moment generating function.

For the general case, since  $(E, \mathcal{E}, \mu)$  is  $\sigma$ -finite, let  $E = \cup_n E_n$  where  $E_n \in \mathcal{E}$ , disjoint and  $\mu(E_n) < \infty$ . Let  $M_n$  be the Poisson random measure on each  $E_n$ . Then  $1_{M < \infty} \sum_{n=1}^{\infty} M_n \rightarrow M 1_{M < \infty}$ . The statement holds for each  $M_n$  by above, so by dominated convergence (dominated by 1),

$$\begin{aligned} \mathbb{E}(e^{inM(g)}) &= \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \left( \int_E e^{ing(y)} \frac{\mu(dy)}{\lambda} \right)^n \\ &= e^{-\lambda} \exp \left( \int_E e^{ing(y)} \mu(dy) \right) \end{aligned}$$

also holds.

(End of proof)  $\square$

---

(26th November, Monday)

Last time :  $(E, \mathcal{E}, \mu)$ , then  $M : \Omega \times \mathcal{E} \rightarrow \mathbb{Z}^+ \cup \{\infty\}$ . Then  $M \sim \text{PRM}(\mu)$  if for any  $(A_k : k \in \mathbb{N})$  in  $\mathcal{E}$  *disjoint*,

- (i)  $M(\cup_k A_k) = \sum_k M(A_k)$ .
- (ii)  $M(A_k)$  is a  $P(\mu(A_k))$  random variable for all  $k$ .
- (iii) The random variables  $M(A_k)$ ,  $k \in \mathbb{N}$  are independent.

If  $\mu(E) = \lambda < \infty$ ,  $M(A) = \sum_{k=1}^N 1_{Y_k \in A}$  with  $N \sim P(\lambda)$ ,  $(Y_n : n \in \mathbb{N})$  i.i.d  $\sim \frac{1}{\lambda} \mu$ . Also,  $M(g) = \sum_{n=1}^N g(Y_n)$ , for  $g : E \rightarrow \mathbb{R}$ ,  $\mathbb{E}(M(g)) = \mu(g)$ ,  $\text{Var}(M(g)) = \mu(g^2)$ , and  $\mathbb{E}(e^{inM(g)}) = \exp \int_E (e^{ing(y)} - 1) \mu(dy)$ .

Now we want to introduce time to this construction. Let  $(E, \mathcal{E}, K)$  be a  $\sigma$ -finite measure space. Then let  $\mu = dt \otimes K$  on  $(0, \infty) \times E$ . Let  $M$  be a PRM on  $(0, \infty) \times E$  with intensity  $\mu$ . Write  $\tilde{M} = M - \mu$ . Call  $\tilde{M}$  a **compensated Poisson Random Measure**. Use filtration  $(\mathcal{F}_t)_{t \geq 0}$ ,  $\mathcal{F}_t = \sigma(\mathcal{F}_t^M, \mathcal{N})$  where  $\mathcal{F}_t^M = \sigma(M([0, s] \times A) : s \leq t, A \in \mathcal{E})$  and  $\mathcal{N} = \{B \in \mathcal{F}_\infty^M : P(B) = 0\}$  is the set of null sets.

Similarly as above, can use, for case  $K(E) < \infty$ , that  $M_t(g) = \sum_{n=1}^{N_t} g(Y_n)$  for  $N \sim \text{PP}(\lambda)$ ,  $g : E \rightarrow \mathbb{R}$ .

**Proposition 8.2.2)** Assume  $K(E) < \infty$  and  $g \in L^1(K)$ . Set

$$\tilde{M}_t(g) = \begin{cases} \int_{(0,t] \times E} g(y) \tilde{M}(ds, dy) & \text{if } \mu((0, t] \times E) < \infty \forall t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Then  $(\tilde{M}_t(g))_{t \geq 0}$  is a *cadlag martingale* with *stationary independent increments*. Moreover,

$$\begin{aligned} \mathbb{E}(\tilde{M}_t(g))^2 &= t \int_E g(y)^2 K(dy) & \dots\dots\dots (\heartsuit) \\ \mathbb{E}(e^{iu\tilde{M}_t(g)}) &= \exp \left[ t \int_E (e^{iug(y)} - 1) K(dy) \right] & \dots\dots\dots (\oplus) \end{aligned}$$

**proof)** Exercise.

This can be thought of a sort of pre-view of what we do in stochastic calculus course, in which we use Gaussian random variables instead of Poissons.

**Theorem 8.2.3)** Let  $g \in L^2(K)$ . Take  $E_n \in \mathcal{E}$  with  $E_n \nearrow E$  and  $K(E_n) < \infty$ . Set  $X_t^N = \tilde{M}_t(g1_{E_n})$ . Then there exists a *cadlag martingale*  $(X_t)_{t \geq 0}$  such that  $\mathbb{E}(\sup_{s \leq t} |X_s^n - X_s|^2) \rightarrow 0$  as  $n \rightarrow \infty$  for all  $t \geq 0$ .

Define  $\tilde{M}_t(g) = X_t$ . Then  $(\tilde{M}_t(g))_{t \geq 0}$  has *stationary independent increments* and  $(\heartsuit)$  and  $(\oplus)$  still hold - so we may write, in convention,

$$\tilde{M}_t(g) = \int_{(0,t] \times E} g(y) \tilde{M}(ds, dy)$$

and call  $(\tilde{M}_t(g))_{t \geq 0}$  (a version of the stochastic) integral of  $g$  with respect to  $\tilde{M}$ . (Although the integral does not converge absolutely!!!)



**proof)** For  $n \leq m$ ,

$$\mathbb{E}(|X_t^n - X_t^m|^2) = \mathbb{E}(\tilde{M}_t(g1_{E_m \setminus E_n})^2) = t \cdot \int_{E_m \setminus E_n} g(y)^2 K(dy) \rightarrow 0 \quad (\text{dominated convergence})$$

So by Doob's  $L^2$ -inequality,

$$\mathbb{E}(\sup_{s \leq t} |X_s^n - X_s^m|^2) \leq 4\mathbb{E}(|X_t^n - X_t^m|^2) \rightarrow 0 \quad \text{as } n, m \rightarrow \infty \quad \dots\dots\dots (\otimes)$$

So there exists  $(n_k)$  such that

$$\sup_{s \leq t} |X_s^{n_k} - X_s^{n_j}| \rightarrow 0 \quad \text{a.s. as } j, k \rightarrow \infty$$

Since the uniform limit of cadlag function is cadlag, there exists a cadlag process  $(X_t)_{t \geq 0}$  such that

$$\sup_{s \leq t} |X_s^{n_k} - X_s| \rightarrow 0 \quad \text{a.s. as } k \rightarrow \infty$$

Note that  $X_s \in L^2(\mathbb{P})$ , because  $X_t \leq X^0 + \sup_n (\sup_{s \leq t} (X_s^n - X^0))$ . By Fatou's lemma with  $(\otimes)$ , has

$$\mathbb{E}(\sup_{s \leq t} |X_s^n - X_s|^2) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

Now for  $A \in \mathcal{F}_s$ , has

$$\begin{aligned} \mathbb{E}((X_t^n - X_s^n)1_A) &= 0 \\ \mathbb{E}(e^{iu(X_t^n - X_s^n)}1_A) &= \mathbb{P}(A) \times \exp \left[ t \int_{E_n} (e^{iug(y)-1-ug(y)})K(dy) \right] \end{aligned}$$

But  $X_t^n \rightarrow X_t$  in  $L^2(\mathbb{P})$  for all  $t$  so on letting  $n \rightarrow \infty$  in these equations we see (using  $|e^{iug} - 1 - iug| \leq \frac{1}{2}u^2g^2$ )

$$\begin{aligned} \mathbb{E}((X_t - X_s)1_A) &= 0 \\ \mathbb{E}(e^{in(X_t - X_s)}1_A) &= \mathbb{P}(A) \times \exp \left[ t \int_n (e^{iug(y)-1-ug(y)})K(dy) \right] \end{aligned}$$

so  $(X_t)_{t \geq 0}$  is a martingale and both  $(\heartsuit)$  and  $(\oplus)$  hold and  $(X_t)_{t \geq 0}$  has stationary independent increments.

(End of proof)  $\square$

## 9. Lévy Processes

### 0.0.1 9.1. Definitions

A **Lévy process** is a cadlag process with stationary independent increments starting at 0. Say that  $(a, b, K)$  is a **Lévy triple** if  $a \in [0, \infty)$ ,  $b \in \mathbb{R}$  and  $K$  is a Borel measure on  $\mathbb{R}$  with  $K(\{0\}) = 0$  and

$$\int_{\mathbb{R}} (|y|^2 \wedge 1) K(dy) < \infty$$

We call  $a$  the **diffusivity**,  $b$  the **drift** and  $K$  the **Lévy measure**. These notions generalize naturally to processes with values in  $\mathbb{R}^d$  but we will consider only the case  $d = 1$ . Take a Brownian motion  $(B_t)$  and a Poisson random measure  $M$  on  $(0, \infty) \times \mathbb{R}$  of intensity  $\mu = dt \otimes K$ . Set

$$X_t = bt + \sqrt{a}B_t \int_{(0,t] \times \mathbb{R}} y 1_{|y| \leq 1} \tilde{M}(ds, dy) + \int_{(0,t] \times \mathbb{R}} y 1_{|y| > 1} M(ds, dy)$$

Note that the first integral exists because  $|y| \wedge 1 \in L^2(K)$  and the second integral exists because  $\mu((0, \infty] \times \{y : |y| \geq 1\}) < \infty$ . Now putting the results from **Section 8** together, we see that,  $(X_t)$  is *cadlag with stationary independent increments* and

$$\mathbb{E}(e^{iuX_t}) = e^{t\psi_{a,b,K}(u)}$$

where  $\psi_{a,b,K}(u) = iub - \frac{1}{2}u^2a + \int_{\mathbb{R}} (e^{iuy} - 1 - iuy1_{|y| \leq 1})K(dy)$ . Thus, to every Lévy triple there corresponds a Lévy process.

(28th November, Wednesday)

## 9.2. Lévy-Khinchine Theorem

**Theorem 9.2.1** Let  $(X_t)_{t \geq 0}$  be a Lévy process (i.e. cadlag, stationary independent increments with  $X_0 = 0$ .) Then there exists a Lévy triple  $(a, b, K)$  such that  $\mathbb{E}(e^{iuX_t}) = e^{t\psi(u)}$  for all  $t \geq 0$ ,  $u \in \mathbb{R}$  where

$$\psi(u) \equiv \psi_{a,b,K}(u) = ibu - \frac{1}{2}au^2 + \int_{\mathbb{R}} (e^{iuy} - 1 - iuy1_{|y| \leq 1})K(dy)$$

and  $\int_{\mathbb{R}} (1 \wedge |y|^2)K(dy) < \infty$ .

**proof)**

I. Set  $\phi_t(y) = \mathbb{E}(e^{iuX_t})$ . We'll show  $\phi_t(y) = e^{t\psi(u)}$  for some  $\psi : \mathbb{R} \rightarrow \mathbb{C}$  continuous.

First note  $\phi_t : \mathbb{R} \rightarrow \mathbb{C}$  is continuous. Since  $X$  has stationary independent increments and

$$X_{nt} = X_t + (X_{2t} - X_t) + \cdots + (X_{nt} - X_{(n-1)t})$$

so we have  $\phi_{nt}(u) = (\phi_t(u))^n$ . Since  $X$  is cadlag, for all  $s \geq 0$ , for  $t > s$ , has  $X_t \rightarrow X_s$  as  $t \rightarrow s$  so

$$|\phi_t(u) - \phi_s(u)| \leq \mathbb{E}(|e^{iu(X_t - X_s)} - 1|) \leq \mathbb{E}((u|X_t - X_s| \wedge 2))$$

So  $\phi_t(u) \rightarrow \phi_s(u)$  uniformly in  $u$  on compact domains. Since  $X_0 = 0$ ,  $\phi_0(u) = 1$ , now  $|\phi_t(u)|^{1/n} = |\phi_{t/n}(u)| \rightarrow 1$  so  $\phi_t(y) \neq 0$  for all  $t \geq 0$ ,  $u \in \mathbb{R}$ .

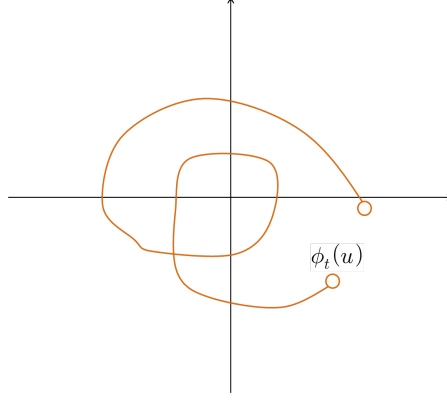
Consider path  $(\phi_t(r) : r \in [0, n])$ . Define

$$\psi_t(u) = \int_1^{\phi_t(u)} \frac{dz}{z}$$

where we integrate along contour homotopic to the path  $(\phi_t(r) : r \in [0, u])$  in  $\mathbb{C} \setminus \{0\}$ . Then  $\psi_t : \mathbb{R} \rightarrow \mathbb{C}$  is continuous and

$$\phi_t(u) = e^{\psi_t(u)}$$

Now  $\psi_{nt}(u) = n\psi_t(u)$ ,  $\psi_0(u) = 0$ , and  $t \mapsto \psi_t(y)$  is right-continuous. Hence  $\psi_t(u) = t\psi_1(u)$  for all  $t \geq 0$  - so define  $\psi$  to be  $\psi = \psi_1$ , then  $\psi_t(u) = t\psi(u)$ .



II. We produce a limit object.

Set  $\nu_n = \text{law of } X_{1/n}$ . Then

$$\psi(u) = \lim_{n \rightarrow \infty} n(\phi_{1/n}(u) - 1)$$

uniformly in  $u$  on compact domains and  $n(\phi_{1/n}(u) - 1) = \int_{\mathbb{R}} (e^{iuy} - 1)n\nu_n(dy)$  so

$$\int_{\mathbb{R}} (1 - \cos(uy))n\nu_n(dy) \rightarrow -\text{Re}\psi(u)$$

uniformly in  $u$  on compact domains. There exists  $C < \infty$  such that

$$y^2 1_{|y| \leq 1} \leq C(1 - \cos(y)) \quad \forall y \in \mathbb{R}$$

and also for all  $\lambda > 0$ , has

$$1_{|y| \geq \lambda} \leq C\lambda \int_0^{1/\lambda} (1 - \cos(uy))du$$

Set  $\eta_n(dy) = n\nu_n(dy)(1 \wedge y^2)$ . Then integration gives

$$\eta_n([-1, 1]) \leq \int_{\mathbb{R}} C(1 - \cos(y))n\nu_n(dy) \rightarrow -C \cdot \text{Re}(\psi(1))$$

and for  $\lambda > 1$

$$\begin{aligned} \eta_n(\mathbb{R} \setminus (-\lambda, \lambda)) &\leq C\lambda \int_{\mathbb{R}} \int_0^{1/\lambda} (1 - \cos(uy))du \cdot n\nu_n(dy) \\ &= C\lambda \int_0^{1/\lambda} \int_{\mathbb{R}} (1 - \cos(uy))n\nu_n(dy)du \\ &\xrightarrow{n \rightarrow \infty} C\lambda \int_0^{1/\lambda} (-\text{Re}(\psi(u)))du \end{aligned}$$

Since  $\psi(0) = 0$  and  $\psi$  continuous, this converges to 0 as  $\lambda \rightarrow \infty$ . Hence  $(\eta_n : n \in \mathbb{N})$  is bounded in total mass and tight, so by Prohorov's theorem, there exists a finite Borel measure  $\eta$  and a subsequence  $(n_k)_k \subset \mathbb{N}$  such that  $\eta_{n_k} \rightarrow \eta$  weakly.

III. Now divide  $\psi$  into two parts by

$$\begin{aligned} \psi(u) &= \lim_{n \rightarrow \infty} \int_{\mathbb{R}} \frac{(e^{iuy} - 1)}{1 \wedge |y|^2} \eta_n(dy) \\ &= \lim_{k \rightarrow \infty} \left( \int_{\mathbb{R}} \mathcal{O}(u, y) \eta_{n_k}(dy) + iu\beta_{n_k} \right) \end{aligned}$$

where

$$\mathcal{O}(u, y) = \begin{cases} (e^{iuy} - 1 - iuy\chi(y))/(1 \wedge y^2) & \text{if } y \neq 0 \\ -u^2/2 & \text{if } y = 0 \end{cases}$$

$\chi : \mathbb{R} \rightarrow [0, 1]$  continuous,  $1_{|y| \leq 1} \leq \chi \leq 1_{|y| \leq 2}$  and

$$\beta_n = \int_{\mathbb{R}} \frac{y\chi(y)}{1 \wedge y^2} \cdot \eta_n(dy)$$

Now  $\mathcal{O}(u, \cdot) \in C_b(\mathbb{R})$  so  $\int_{\mathbb{R}} \mathcal{O}(u, y) \eta_{n_k}(dy)$  converges as  $k \rightarrow \infty$  and in turn  $\beta_{n_k}$  converges as  $k \rightarrow \infty$ , say  $\beta_{n_k} \rightarrow \beta$  for some  $\beta \in \mathbb{R}$ . Hence  $\psi(u) = \psi_{a,b,K}(u)$  where

$$\begin{aligned} a &= \eta(\{0\}) \\ K(dy) &= 1_{y \neq 0} \eta(dy) / (1 \wedge y^2) \\ b &= \beta - \int_{\mathbb{R}} y \left( \chi(y) - 1_{|y| \leq 1} \right) K(dy) \end{aligned}$$

That is,

$$\psi(u) = iub - \frac{1}{2}u^2a + \int_E (e^{iuy} - 1 - iuy1_{|y| \leq 1}) K(dy)$$

(End of proof)  $\square$