

# INTRODUCTION TO PROBABILITY

CARSON JAMES

## CONTENTS

|   |    |
|---|----|
| 1. Basic Probability                                | 2  |
| 2. Kolomogorov Extension Theorem                    | 2  |
| 3. Probability                                      | 3  |
| 3.1. Distributions                                  | 3  |
| 3.2. Independence                                   | 6  |
| 3.3. $L^p$ Spaces for Probability                   | 9  |
| 3.4. Borel Cantelli Lemma                           | 14 |
| 4. Probability on locally compact Groups            | 16 |
| 4.1. Action on Probability Measures                 | 16 |
| 5. Weak Convergence of Measures                     | 17 |
| 6. Concentration Inequalities                       | 18 |
| 6.1. Introduction                                   | 18 |
| 6.2. Sub $\alpha$ -Exponential Random Variables     | 18 |
| 7. Conditional Expectation and Probability          | 22 |
| 7.1. Conditional Expectation                        | 22 |
| 7.2. Conditional Probability                        | 23 |
| 8. Markov Chains                                    | 26 |
| 9. Probabilities Induced by Isometric Group Actions | 27 |
| 9.1. Applications to Bayesian Statistics            | 27 |
| 10. Stochastic Integration                          | 28 |
| 11. TODO  | 29 |
| References  | 30 |

## 1. BASIC PROBABILITY

## 2. KOLOMOGOROV EXTENSION THEOREM

Show that for each  $x$ ,  $\pi_{st}(x) = \pi_s(x)\pi_t(x)$  and this implies that for each feasible  $(s, A), (t, B)$  so that  $\pi_{st}^{-1}(A \times B) = \pi_s^{-1}(A) \cap \pi_t^{-1}(B)$ .

For  $u \subset \mathbb{C}^{\mathbb{N}}$ , set  $\pi_u \mathbb{C}^{\mathbb{N}} = \{x \in \mathbb{C}^{\mathbb{N}} : \text{for each } j \in u^c, x_j = 0\}$ . Consider lattice of subspaces of  $\pi_u \mathbb{C}^{\mathbb{N}}$ , ordered by inclusion indexed by finite subsets  $u \subset \mathbb{N}$ .

show that if  $u \subset v$ , then  $\mathcal{B}(\pi_u \mathbb{C}^{\mathbb{N}}) = \iota_{u,v}^*(\mathcal{B}(\pi_v \mathbb{C}^{\mathbb{N}}))$  where  $\iota_{u,v} : \pi_u \mathbb{C}^{\mathbb{N}} \rightarrow \pi_v \mathbb{C}^{\mathbb{N}}$  is inclusion.

Equip  $(\pi_u \mathbb{C}^{\mathbb{N}}, \mathcal{B}(\pi_u \mathbb{C}^{\mathbb{N}}))$  with measure  $\mu_u$  such that  $\pi_u$

with pullback sigma algebra  $\iota_{u,v}^*(\mathcal{B}(\pi_v \mathbb{C}^{\mathbb{N}}))$  where  $u \subset v$  and  $\iota_{u,v} : \pi_u \mathbb{C}^{\mathbb{N}} \rightarrow \pi_v \mathbb{C}^{\mathbb{N}}$ . Equip each object  $(\pi_u \mathbb{C}^{\mathbb{N}}, \mathcal{B}(\pi_u \mathbb{C}^{\mathbb{N}}))$  with measures  $\mu_u$  satisfying  $\mu_u$  of in stoch has maximum

**Definition 2.0.1.** Let

## 3. PROBABILITY

## 3.1. Distributions.

**Definition 3.1.1.** Let  $\Omega$  be a set and  $\mathcal{P} \subset \mathcal{P}(X)$ . Then  $\mathcal{P}$  is said to be a  $\pi$ -system on  $\Omega$  if for each  $A, B \in \mathcal{P}$ ,  $A \cap B \in \mathcal{P}$ .

**Definition 3.1.2.** Let  $\Omega$  be a set and  $\mathcal{L} \subset \mathcal{P}(\Omega)$ . Then  $\mathcal{L}$  is said to be a  $\lambda$ -system on  $\Omega$  if

- (1)  $\mathcal{L} \neq \emptyset$
- (2) for each  $A \in \mathcal{L}$ ,  $A^c \in \mathcal{L}$
- (3) for each  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{L}$ , if  $(A_n)_{n \in \mathbb{N}}$  is disjoint, then  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{L}$

**Exercise 3.1.3.** Let  $\Omega$  be a set and  $\mathcal{L}$  a  $\lambda$ -system on  $\Omega$ . Then

- (1)  $\Omega, \emptyset \in \mathcal{L}$

*Proof.* Straightforward. □

**Definition 3.1.4.** Let  $\Omega$  be a set and  $\mathcal{C} \subset \mathcal{P}(\Omega)$ . Put

$$\mathcal{S} = \{\mathcal{L} \subset \mathcal{P}(\Omega) : \mathcal{L} \text{ is a } \lambda\text{-system on } \Omega \text{ and } \mathcal{C} \subset \mathcal{L}\}$$

We define the  $\lambda$ -system on  $\Omega$  generated by  $\mathcal{C}$ ,  $\lambda(\mathcal{C})$ , to be

$$\lambda(\mathcal{C}) = \bigcap_{\mathcal{L} \in \mathcal{S}} \mathcal{L}$$

**Exercise 3.1.5.** Let  $\Omega$  be a set and  $\mathcal{C} \subset \mathcal{P}(\Omega)$ . If  $\mathcal{C}$  is a  $\lambda$ -system and  $\mathcal{C}$  is a  $\pi$ -system, then  $\mathcal{C}$  is a  $\sigma$ -algebra.

*Proof.* Suppose that  $\mathcal{C}$  is a  $\lambda$ -system and  $\mathcal{C}$  is a  $\pi$ -system. Then we need only verify the third axiom in the definition of a  $\sigma$ -algebra. Let  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{C}$ . Define  $B_1 = A_1$  and for  $n \geq 2$ , define  $B_n = A_n \cap \left( \bigcup_{k=1}^{n-1} A_k \right)^c = A_n \cap \left( \bigcap_{k=1}^{n-1} A_k^c \right) \in \mathcal{C}$ . Then  $(B_n)_{n \in \mathbb{N}}$  is disjoint and therefore  $\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} B_n \in \mathcal{C}$ . □

**Theorem 3.1.6.** (Dynkin's Theorem)

Let  $\Omega$  be a set.

- (1) Let  $\mathcal{P}$  be a  $\pi$ -system on  $\Omega$  and  $\mathcal{L}$  a  $\lambda$ -system on  $\Omega$ . If  $\mathcal{P} \subset \mathcal{L}$ , then  $\sigma(\mathcal{P}) \subset \mathcal{L}$ .
- (2) Let  $\mathcal{P}$  be a  $\pi$ -system on  $\Omega$ . Then  $\sigma(\mathcal{P}) = \lambda(\mathcal{P})$

**Exercise 3.1.7.** Let  $(\Omega, \mathcal{F})$  be a measurable space and  $\mu, \nu$  probability measures on  $(\Omega, \mathcal{F})$ . Put  $\mathcal{L}_{\mu, \nu} = \{A \in \mathcal{F} : \mu(A) = \nu(A)\}$ . Then  $\mathcal{L}_{\mu, \nu}$  is a  $\lambda$ -system on  $\Omega$ .

*Proof.*

- (1)  $\emptyset \in \mathcal{L}_{\mu, \nu}$ .
- (2) Let  $A \in \mathcal{L}_{\mu, \nu}$ . Then  $\mu(A) = \nu(A)$ . Thus

$$\begin{aligned} \mu(A^c) &= 1 - \mu(A) \\ &= 1 - \nu(A) \\ &= \nu(A^c) \end{aligned}$$

So  $A^c \in \mathcal{L}_{\mu, \nu}$ .

- (3) Let  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{L}_{\mu, \nu}$ . So for each  $n \in \mathbb{N}$ ,  $\mu(A_n) = \nu(A_n)$ . Suppose that  $(A_n)_{n \in \mathbb{N}}$  is disjoint. Then

$$\begin{aligned} \mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= \sum_{n \in \mathbb{N}} \mu(A_n) \\ &= \sum_{n \in \mathbb{N}} \nu(A_n) \\ &= \nu\left(\bigcup_{n \in \mathbb{N}} A_n\right) \end{aligned}$$

Hence  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{L}_{\mu, \nu}$ .

□

**Exercise 3.1.8.** Let  $(\Omega, \mathcal{F})$  be a measurable space,  $\mu, \nu$  probability measures on  $(\Omega, \mathcal{F})$  and  $\mathcal{P} \subset \mathcal{F}$  a  $\pi$ -system on  $\Omega$ . Suppose that for each  $A \in \mathcal{P}$ ,  $\mu(A) = \nu(A)$ . Then for each  $A \in \sigma(\mathcal{P})$ ,  $\mu(A) = \nu(A)$ .

*Proof.* Using the previous exercise, we see that  $\mathcal{P} \subset \mathcal{L}_{\mu, \nu}$ . Dynkin's theorem implies that  $\sigma(\mathcal{P}) \subset \mathcal{L}_{\mu, \nu}$ . So for each  $A \in \sigma(\mathcal{P})$ ,  $\mu(A) = \nu(A)$ . □

**Definition 3.1.9.** Let  $F : \mathbb{R} \rightarrow \mathbb{R}$ . Then  $F$  is said to be a **probability distribution function** if

- (1)  $F$  is right continuous
- (2)  $F$  is increasing
- (3)  $F(-\infty) = 0$  and  $F(\infty) = 1$

**Definition 3.1.10.** Let  $P$  be a probability measure on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . We define  $F_P : \mathbb{R} \rightarrow \mathbb{R}$ , by

$$F_P(x) = P((-\infty, x])$$

We call  $F_P$  the **probability distribution function of  $P$** .

**Exercise 3.1.11.** Let  $(\Omega, \mathcal{F}, P)$  be a probability measure. Then  $F_P$  is a probability distribution function.

*Proof.* (1) Let  $x \in \mathbb{R}$  and  $(x_n)_{n \in \mathbb{N}} \subset [x, \infty)$ . Suppose that  $x_n \rightarrow x$ . Then  $(x, x_n] \rightarrow \emptyset$  because  $\limsup_{n \rightarrow \infty} (x, x_n] = \emptyset$ . Thus

$$F(x_n) - F(x) = P((x, x_n]) \rightarrow P(\emptyset) = 0$$

This implies that

$$F(x_n) \rightarrow F(x)$$

So  $F$  is right continuous.

- (2) Clearly  $F_P$  is increasing.
- (3) Continuity from below tells us that

$$F(-\infty) = \lim_{n \rightarrow -\infty} F(n) = \lim_{n \rightarrow -\infty} P((-\infty, n]) = 0$$

and continuity from above tell us that

$$F(\infty) = \lim_{n \rightarrow \infty} F(n) = \lim_{n \rightarrow \infty} P((-\infty, n]) = 1$$

□

**Exercise 3.1.12.** Let  $\mu, \nu$  be probability measures on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Then  $F_\mu = F_\nu$  iff  $\mu = \nu$ .

*Proof.* Clearly if  $\mu = \nu$ , then  $F_\mu = F_\nu$ . Conversely, suppose that  $F_\mu = F_\nu$ . Then for each  $x \in \mathbb{R}$ ,

$$\begin{aligned}\mu((-\infty, x]) &= F_\mu(x) \\ &= F_\nu(x) \\ &= \nu((-\infty, x])\end{aligned}$$

Put  $\mathcal{C} = \{(-\infty, x] : x \in \mathbb{R}\}$ . Then  $\mathcal{C}$  is a  $\pi$ -system and for each  $A \in \mathcal{C}$ ,  $\mu(A) = \nu(A)$ . Hence for each  $A \in \sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$ ,  $\mu(A) = \nu(A)$ . So  $\mu = \nu$ .  $\square$

**Definition 3.1.13.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X : \Omega \rightarrow \mathbb{R}^n$ . Then  $X$  is said to be a **random vector** on  $(\Omega, \mathcal{F})$  if  $X$  is  $\mathcal{F}$ - $\mathcal{B}(\mathbb{R}^n)$  measurable. If  $n = 1$ , then  $X$  is said to be a **random variable**. We define

$$L_n^0(\Omega, \mathcal{F}, P) = \{X : \Omega \rightarrow \mathbb{R}^n : X \text{ is a random vector}\}$$

and

$$L_n^p(\Omega, \mathcal{F}, P) = \left\{ X \in L_n^0 : \int \|X\|^p dP < \infty \right\}$$

**Definition 3.1.14.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X$  a random variable on  $(\Omega, \mathcal{F})$ . We define the **probability distribution** of  $X$ ,  $P_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ , to be the measure

$$P_X = X_*P$$

That is, for each  $A \in \mathcal{B}(\mathbb{R})$ ,

$$P_X(A) = P(X^{-1}(A))$$

We define the **probability distribution function** of  $X$ ,  $F_X : \mathbb{R} \rightarrow [0, 1]$ , to be

$$F_X = F_{P_X}$$

**Definition 3.1.15.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X$  a random variable on  $(\Omega, \mathcal{F})$ . If  $P_X \ll m$ , we define the **probability density** of  $X$ ,  $f_X : \mathbb{R} \rightarrow \mathbb{R}$ , by

$$f_X = \frac{dP_X}{dm}$$

**Exercise 3.1.16.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $(X_n)_{n \in \mathbb{N}}$  be a sequence of random variables on  $(\Omega, \mathcal{F})$ . Then for each  $x \in \mathbb{R}$ ,

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} X_n > x\right) \leq \liminf_{n \rightarrow \infty} P(X_n > x)$$

*Proof.* Let  $\omega \in \left\{ \liminf_{n \rightarrow \infty} X_n > x \right\}$ . Then  $x < \liminf_{n \rightarrow \infty} X_n(\omega) = \sup_{n \in \mathbb{N}} \left( \inf_{k \geq n} X_k(\omega) \right)$ . So there exists  $n^* \in \mathbb{N}$  such that  $x < \inf_{k \geq n^*} X_k(\omega)$ . Then for each  $k \in \mathbb{N}$ ,  $k \geq n^*$  implies that  $x < X_k(\omega)$ .

So there exists  $n^* \in \mathbb{N}$  such that for each  $k \in \mathbb{N}$ ,  $k \geq n^*$  implies that  $\mathbf{1}_{\{X_k > x\}}(\omega) = 1$ . Hence

$\inf_{k \geq n^*} \mathbf{1}_{\{X_k > x\}}(\omega) = 1$ . Thus  $\liminf_{n \rightarrow \infty} \mathbf{1}_{\{X_n > x\}}(\omega) = \sup_{n \in \mathbb{N}} \left( \inf_{k \geq n} \mathbf{1}_{\{X_k > x\}}(\omega) \right) = 1$ . Therefore

$\omega \in \liminf_{n \rightarrow \infty} \{X_k > x\}$  and we have shown that

$$\left\{ \liminf_{n \rightarrow \infty} X_n > x \right\} \subset \liminf_{n \rightarrow \infty} \{X_k > x\}$$

Then

$$\begin{aligned} P\left(\liminf_{n \rightarrow \infty} X_n > x\right) &\leq P\left(\liminf_{n \rightarrow \infty} \{X_k > x\}\right) \\ &\leq \liminf_{n \rightarrow \infty} P(\{X_k > x\}) \end{aligned}$$

□

**Definition 3.1.17.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X \in L^+(\Omega) \cup L^1$ . Define the **expectation of X**,  $E(X)$ , to be

$$E(X) = \int X dP$$

### 3.2. Independence.

**Definition 3.2.1.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\mathcal{C} \subset \mathcal{F}$ . Then  $\mathcal{C}$  is said to be **independent** if for each  $(A_i)_{i=1}^n \subset \mathcal{C}$ ,

$$P\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n P(A_k)$$

**Definition 3.2.2.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\mathcal{C}_1, \dots, \mathcal{C}_n \subset \mathcal{F}$ . Then  $\mathcal{C}_1, \dots, \mathcal{C}_n$  are said to be **independent** if for each  $A_1 \in \mathcal{C}_1, \dots, A_n \in \mathcal{C}_n$ ,  $A_1, \dots, A_n$  are independent.

**Note 3.2.3.** We will explicitly say that for each  $i = 1, \dots, n$ ,  $\mathcal{C}_i$  is independent when talking about the independence of the elements of  $\mathcal{C}_i$  to avoid ambiguity.

**Definition 3.2.4.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X_1, \dots, X_n$  random variables on  $(\Omega, \mathcal{F})$ . Then  $X_1, \dots, X_n$  are said to be **independent** if for each  $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$ ,  $X_1^{-1}B_1, \dots, X_n^{-1}B_n$  are independent.

**Exercise 3.2.5.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X_1, \dots, X_n$  random variables on  $(\Omega, \mathcal{F})$ . Then  $X_1, \dots, X_n$  are independent iff  $\sigma(X_1), \dots, \sigma(X_n)$  are independent.

*Proof.* Suppose that  $X_1, \dots, X_n$  are independent. Let  $A_1 \in \sigma(X_1), \dots, A_n \in \sigma(X_n)$ . Then for each  $i = 1, \dots, n$ , there exists  $B_i \in \mathcal{B}(\mathbb{R})$  such that  $A_i = X_i^{-1}(B_i)$ . Then  $A_1, \dots, A_n$  are independent. Hence  $\sigma(X_1), \dots, \sigma(X_n)$  are independent. Conversely, suppose that  $\sigma(X_1), \dots, \sigma(X_n)$  are independent. Let  $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$ . Then for each  $i = 1, \dots, n$ ,  $X_i^{-1}B_i \in \sigma(X_i)$ . Then  $X_1^{-1}B_1, \dots, X_n^{-1}B_n$  are independent. Hence  $X_1, \dots, X_n$  are independent. □

**Exercise 3.2.6.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X_1, \dots, X_n$  random variables on  $(\Omega, \mathcal{F})$  and  $\mathcal{F}_1, \dots, \mathcal{F}_n \subset \mathcal{F}$  a collection of  $\sigma$ -algebras on  $\Omega$ . Suppose that for each  $i = 1, \dots, n$ ,  $X_i$  is  $\mathcal{F}_i$ -measurable. If  $\mathcal{F}_1, \dots, \mathcal{F}_n$  are independent, then  $X_1, \dots, X_n$  are independent.

*Proof.* For each  $i = 1, \dots, n$ ,  $\sigma(X_i) \subset \mathcal{F}_i$ . So  $\sigma(X_1), \dots, \sigma(X_n)$  are independent. Hence  $X_1, \dots, X_n$  are independent. □

**Exercise 3.2.7.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\mathcal{C}_1, \dots, \mathcal{C}_n \subset \mathcal{F}$ . Suppose that for each  $i = 1, \dots, n$ ,  $\mathcal{C}_i$  is a  $\pi$ -system and  $\mathcal{C}_1, \dots, \mathcal{C}_n$  are independent, then  $\sigma(\mathcal{C}_1), \dots, \sigma(\mathcal{C}_n)$  are independent.

*Proof.* Let  $A_2 \in \mathcal{C}_2$ . Define  $\mathcal{L} = \{A \in \mathcal{F} : P(A \cap A_2) = P(A)P(A_2)\}$ . Then

- (1)  $\Omega \in \mathcal{L}$
- (2) If  $A \in \mathcal{L}$ , then

$$\begin{aligned} P(A^c \cap A_2) &= P(A_2) - P(A_2 \cap A) \\ &= P(A_2) - P(A_2)P(A) \\ &= (1 - P(A))P(A_2) \\ &= P(A^c)P(A_2) \end{aligned}$$

So  $A^c \in \mathcal{L}$

- (3) If  $(B_n)_{n \in \mathbb{N}} \subset \mathcal{L}$  is disjoint, then

$$\begin{aligned} P\left(\left[\bigcup_{n \in \mathbb{N}} B_n\right] \cap A_2\right) &= P\left(\bigcup_{n \in \mathbb{N}} B_n \cap A_2\right) \\ &= \sum_{n \in \mathbb{N}} P(B_n \cap A_2) \\ &= \sum_{n \in \mathbb{N}} P(B_n)P(A_2) \\ &= \left[\sum_{n \in \mathbb{N}} P(B_n)\right]P(A_2) \\ &= P\left(\bigcup_{n \in \mathbb{N}} B_n\right)P(A_2) \end{aligned}$$

So  $\bigcup_{n \in \mathbb{N}} B_n \in \mathcal{L}$ .

Thus  $\mathcal{L}$  is a  $\lambda$ -system. Since  $\mathcal{C}_1 \subset \mathcal{L}$  is a  $\pi$ -system, Dynkin's theorem tells us that  $\sigma(\mathcal{C}_1) \subset \mathcal{L}$ . Since  $A_2 \in \mathcal{C}_2$  is arbitrary  $\sigma(\mathcal{C}_1)$  and  $\mathcal{C}_2$  are independent. The same reasoning implies that  $\sigma(\mathcal{C}_1)$  and  $\sigma(\mathcal{C}_2)$  are independent. Let  $A_2 \in \mathcal{C}_1, \dots, A_n \in \mathcal{C}_n$ . We may do the same process with

$$\mathcal{L} = \left\{A \in \mathcal{F} : P\left(A \cap \left(\bigcap_{i=2}^n A_i\right)\right) = P(A) \prod_{i=2}^n P(A_i)\right\}$$

and conclude that  $\sigma(\mathcal{C}_1), \mathcal{C}_2, \dots, \mathcal{C}_n$  are independent. Which, using the same reasoning would imply that  $\sigma(\mathcal{C}_1), \dots, \sigma(\mathcal{C}_n)$  are independent.  $\square$

**Exercise 3.2.8.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X_1, \dots, X_n$  random variables on  $(\Omega, \mathcal{F})$ . Then  $X_1, \dots, X_n$  are independent iff for each  $x_1, \dots, x_n \in \mathbb{R}$ ,

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i)$$

*Proof.* Suppose that  $X_1, \dots, X_n$  are independent. Then  $\sigma(X_1), \dots, \sigma(X_n)$  are independent. Let  $x_1, \dots, x_n \in \mathbb{R}$ . Then for each  $i = 1, \dots, n$ ,  $\{X_i \leq x_i\} \in \sigma(X_i)$ . Hence

$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i)$ . Conversely, suppose that for each  $x_1, \dots, x_n \in \mathbb{R}$ ,  $P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i)$ . Define  $\mathcal{C} = \{(-\infty, x] : x \in \mathbb{R}\}$ . Then  $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$ . For each  $i = 1, \dots, n$ , define  $\mathcal{C}_i = X_i^{-1}\mathcal{C}$ . Then for each  $i = 1, \dots, n$ ,  $\mathcal{C}_i$  is a  $\pi$ -system and

$$\begin{aligned}\sigma(\mathcal{C}_i) &= \sigma(X_i^{-1}(\mathcal{C})) \\ &= X_i^{-1}(\sigma(\mathcal{C})) \\ &= X_i^{-1}(\mathcal{B}(\mathbb{R})) \\ &= \sigma(X_i)\end{aligned}$$

By assumption,  $\mathcal{C}_1, \dots, \mathcal{C}_n$  are independent. The previous exercise tells us that  $\sigma(X_1), \dots, \sigma(X_n)$  are independent. Then  $X_1, \dots, X_n$  are independent.  $\square$

**Exercise 3.2.9.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X_1, \dots, X_n$  random variables on  $(\Omega, \mathcal{F})$ . Define  $X = (X_1, \dots, X_n)$ . If  $X_1, \dots, X_n$  are independent, then

$$P_X = \prod_{i=1}^n P_{X_i}$$

.

*Proof.* Let  $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ . Then

$$\begin{aligned}P_X(A_1 \times \dots \times A_n) &= P(X \in A_1 \times \dots \times A_n) \\ &= P(X_1 \in A_1, \dots, X_n \in A_n) \\ &= P(X_1 \in A_1) \dots P(X_n \in A_n) \\ &= P_{X_1}(A_1) \dots P_{X_n}(A_n) \\ &= \prod_{i=1}^n P_{X_i}(A_1 \times \dots \times A_n)\end{aligned}$$

Put

$$\mathcal{P} = \{A_1 \times \dots \times A_n : A_1 \in \mathcal{B}(\mathbb{R}), \dots, A_n \in \mathcal{B}(\mathbb{R})\}$$

Then  $\mathcal{P}$  is a  $\pi$ -system and

$$\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^n)$$

A previous exercise then tells us that  $P_X = \prod_{i=1}^n P_{X_i}$   $\square$

**Exercise 3.2.10.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X_1, \dots, X_n$  random variables on  $(\Omega, \mathcal{F})$  and  $f_1, \dots, f_n : \mathbb{R} \rightarrow \mathbb{R} \in L^0$ . Suppose that  $f_1 \circ X_1, \dots, f_n \circ X_n \in L^+(\Omega)$  or  $f_1 \circ X_1, \dots, f_n \circ X_n \in L^1(\Omega)$ . If  $X_1, \dots, X_n$  are independent, then

$$E(f_1(X_1) \dots f_n(X_n)) = \prod_{i=1}^n E(f_i(X_i))$$



*Proof.* Define the random vector  $X : \Omega \rightarrow \mathbb{R}^n$  by  $X = (X_1, \dots, X_n)$  and  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  by  $g(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$ . Suppose that for each  $i = 1, \dots, n$ ,  $f_i \in L^+(\mathbb{R})$ . Then  $g \in L^+(\mathbb{R}^n)$  and by change of variables,

$$\begin{aligned}
 E(f_1(X_1) \cdots f_n(X_n)) &= E(g(X)) \\
 &= \int_{\Omega} g \circ X \, dP \\
 &= \int_{\mathbb{R}^n} g(x) \, dP_X(x) \\
 &= \int_{\mathbb{R}^n} g(x) \, d \prod_{i=1}^n P_{X_i}(x) \\
 &= \prod_{i=1}^n \int_{\mathbb{R}} f_i(x) \, dP_{X_i}(x) \\
 &= \prod_{i=1}^n \int_{\Omega} f_i \circ X \, dP \\
 &= \prod_{i=1}^n E(f_i(X_i))
 \end{aligned}$$

If for each  $i = 1, \dots, n$ ,  $f_i \in L^1(\mathbb{R}, P_{X_i})$ , then following the above reasoning with  $|g|$  tells us that  $g \in L^1(\mathbb{R}^n, P_X)$  and we use change of variables and Fubini's theorem to get the same result.  $\square$

### 3.3. $L^p$ Spaces for Probability.

**Note 3.3.1.** Recall that for a probability space  $(\Omega, \mathcal{F}, P)$  and  $1 \leq p \leq q \leq \infty$  we have  $L^q \subset L^p$  and for each  $X \in L^q$ ,  $\|X\|_p \leq \|X\|_q$ . Also recall that for  $X, Y \in L^2$ , we have that  $\|XY\|_1 \leq \|X\|_2 \|Y\|_2$ .

**Definition 3.3.2.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X \in L^2$ . Define the **variance of X**,  $Var(X)$ , to be

$$Var(X) = E([(X - E(X))^2])$$

.

**Definition 3.3.3.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X, Y \in L^2$ . Define the

**Definition 3.3.4.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X, Y \in L^2$ . Define the **covariance of X and Y**,  $Cov(X, Y)$ , to be

$$Cov(X, Y) = E([X - E(X)][Y - E(Y)])$$

**Exercise 3.3.5.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X, Y \in L^2$ . Then the covariance is well defined and  $Cov(X, Y)^2 \leq Var(X)Var(Y)$

*Proof.* By Holder's inequality,

$$\begin{aligned}
|Cov(X, Y)| &= \left| \int (X - E(X))(Y - E(Y)) dP \right| \\
&\leq \int |(X - E(X))(Y - E(Y))| dP \\
&= \|(X - E(X))(Y - E(Y))\|_1 \\
&\leq \|X - E(X)\|_2 \|Y - E(Y)\|_2 \\
&= \left( \int |X - E(X)|^2 dP \right)^{\frac{1}{2}} \left( \int |Y - E(Y)|^2 dP \right)^{\frac{1}{2}} \\
&= Var(X)^{\frac{1}{2}} Var(Y)^{\frac{1}{2}}
\end{aligned}$$

So  $Cov(X, Y)^2 \leq Var(X)Var(Y)$ . □

**Exercise 3.3.6.** Let  $(\Omega, \mathcal{F}, P)$  be a measure space and  $X, Y \in L^2$ . Then

- (1)  $Cov(X, Y) = E(XY) - E(X)E(Y)$
- (2) If  $X, Y$  are independent, then  $Cov(X, Y) = 0$
- (3)  $Var(X) = E(X^2) - E(X)^2$
- (4) for each  $a, b \in \mathbb{R}$ ,  $Var(aX + b) = a^2 Var(X)$ .
- (5)  $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

*Proof.*

- (1) We have that

$$\begin{aligned}
Cov(X, Y) &= E \left[ (X - E(X))(Y - E(Y)) \right] \\
&= E(XY - E(Y)X - E(X)Y + E(X)E(Y)) \\
&= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\
&= E(XY) - E(X)E(Y)
\end{aligned}$$

- (2) Suppose that  $X, Y$  are independent. Then  $E(XY) = E(X)E(Y)$ . Hence

$$\begin{aligned}
Cov(X, Y) &= E(XY) - E(X)E(Y) \\
&= E(X)E(Y) - E(X)E(Y) \\
&= 0
\end{aligned}$$

- (3) Part (1) implies that

$$\begin{aligned}
Var(X) &= Cov(X, X) \\
&= E(X^2) - E(X)^2
\end{aligned}$$

(4) Let  $a, b \in \mathbb{R}$ . Then

$$\begin{aligned}
 \text{Var}(aX + b) &= E[(aX + b)^2] - E(aX + b)^2 \\
 &= E[a^2X^2 + 2abX + b^2] - (aE(X) + b)^2 \\
 &= a^2E(X^2) + 2abE(X) + b^2 - (a^2E(X)^2 + 2abE(X) + b^2) \\
 &= a^2(E(X^2) - E(X)^2) \\
 &= a^2\text{Var}(X)
 \end{aligned}$$

(5) We have that

$$\begin{aligned}
 \text{Var}(X + Y) &= E[(X + Y)^2] - E[X + Y]^2 \\
 &= E[X^2 + 2XY + Y^2] - (E(X) + E(Y))^2 \\
 &= E(X^2) + 2E[XY] + E(Y^2) - (E(X)^2 + 2E(X)E(Y) + E(Y)^2) \\
 &= (E(X^2) - E(X)^2) + (E(Y^2) - E(Y)^2) + 2(E[XY] - E(X)E(Y)) \\
 &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)
 \end{aligned}$$

□

**Definition 3.3.7.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X, Y \in L^2$ . The **correlation of X and Y**,  $\text{Cor}(X, Y)$ , is defined to be

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

**Exercise 3.3.8.**

**Exercise 3.3.9. Jensen's Inequality:**

Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X \in L^1$  and  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ . If  $\phi$  is convex, then

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

*Proof.* Put  $x_0 = E(X)$ . Since  $\phi$  is convex, there exist  $a, b \in \mathbb{R}$  such that  $\phi(x_0) = ax_0 + b$  and for each  $x \in \mathbb{R}$ ,  $\phi(x) \geq ax + b$ . Then

$$\begin{aligned}
 E[\phi(X)] &= \int \phi(X) dP \\
 &\geq \int [aX + b] dP \\
 &= a \int X dP + b \\
 &= aE(X) + b \\
 &= ax_0 + b \\
 &= \phi(x_0) \\
 &= \phi(E(X))
 \end{aligned}$$

□

**Exercise 3.3.10.** Markov's Inequality: Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X \in L^+$ . Then for each  $a \in (0, \infty)$ ,

$$P(X \geq a) \leq \frac{E(X)}{a}$$

*Proof.* Let  $a \in (0, \infty)$ . Then  $a\mathbf{1}_{\{X \geq a\}} \leq X\mathbf{1}_{\{X \geq a\}}$ . Thus

$$\begin{aligned} aP(X \geq a) &= \int a\mathbf{1}_{\{X \geq a\}} dP \\ &= \int X\mathbf{1}_{\{X \geq a\}} dP \\ &\leq \int X dP \\ &= E(X) \end{aligned}$$

Therefore

$$P(X \geq a) \leq \frac{E(X)}{a}$$

□

**Exercise 3.3.11.** Chebychev's Inequality: Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X \in L^2$ . Then for each  $a \in (0, \infty)$ ,

$$P(|X - E(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}$$

*Proof.* Let  $a \in (0, \infty)$ . Then

$$\begin{aligned} P(|X - E(X)| \geq a) &= P((X - E(X))^2 \geq a^2) \\ &\leq \frac{E[(X - E(X))^2]}{a^2} \\ &= \frac{\text{Var}(X)}{a^2} \end{aligned}$$

□

**Exercise 3.3.12.** Chernoff's Bound: Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X \in L^2$ . Then for each  $a, t \in (0, \infty)$ ,

$$P(X \geq a) \leq e^{-ta} E[e^{tX}]$$

*Proof.* Let  $a, t \in (0, \infty)$ . Then

$$\begin{aligned} P(X \geq a) &= P(tX \geq ta) \\ &= P(e^{tX} \geq e^{ta}) \\ &\leq e^{-ta} E[e^{tX}] \end{aligned}$$

□

**Exercise 3.3.13.** Weak Law of Large Numbers: Let  $(\Omega, \mathcal{F}, P)$  be a probability space  $(X_i)_{i \in \mathbb{N}} \subset L^2$ . Suppose that  $(X_i)_{i \in \mathbb{N}}$  are iid. Then

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} E[X_1]$$

*Proof.* Put  $\mu = E[X_1]$  and  $\sigma^2 = \text{Var}(X_1)$ . Then

$$\begin{aligned} E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] &= \frac{1}{n} \sum_{i=1}^n E[X_i] \\ &= \frac{1}{n} \sum_{i=1}^n \mu \\ &= \mu \end{aligned}$$

and

$$\begin{aligned} \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) &= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Let  $\epsilon > 0$ . Then

$$\begin{aligned} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - E[X_1]\right| \geq \epsilon\right) &= P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \epsilon\right) \\ &= P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - E\left[\frac{1}{n} \sum_{i=1}^n X_i\right]\right| \geq \epsilon\right) \\ &\leq \frac{\text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right)}{\epsilon^2} \\ &= \frac{\sigma^2/n}{\epsilon^2} \\ &= \frac{\sigma^2}{n\epsilon^2} \rightarrow 0 \end{aligned}$$

So

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} E[X_1]$$

□

### 3.4. Borel Cantelli Lemma.

#### Exercise 3.4.1. Borel Cantelli Lemma:

Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ .

- (1) If  $\sum_{n \in \mathbb{N}} P(A_n) < \infty$ , then  $P(\limsup_{n \rightarrow \infty} A_n) = 0$ .
- (2) If  $(A_n)_{n \in \mathbb{N}}$  are independent and  $\sum_{n \in \mathbb{N}} P(A_n) = \infty$ , then  $P(\limsup_{n \rightarrow \infty} A_n) = 1$ .

*Proof.*

- (1) Suppose that  $\sum_{n \in \mathbb{N}} P(A_n) < \infty$ . Recall that

$$\limsup_{n \rightarrow \infty} A_n = \left\{ \omega \in \Omega : \sum_{n \in \mathbb{N}} 1_{A_n}(\omega) = \infty \right\}$$

Then

$$\begin{aligned} \infty &> \sum_{n \in \mathbb{N}} P(A_n) \\ &= \sum_{n \in \mathbb{N}} \int 1_{A_n} dP \\ &= \int \sum_{n \in \mathbb{N}} 1_{A_n} dP \end{aligned}$$

Thus  $\sum_{n \in \mathbb{N}} 1_{A_n} < \infty$  a.e. and  $P(\limsup_{n \rightarrow \infty} A_n) = 0$ .

- (2) Suppose that  $(A_n)_{n \in \mathbb{N}}$  are independent and  $\sum_{n \in \mathbb{N}} P(A_n) = \infty$ .

□

**Exercise 3.4.2.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $(X_n)_{n \in \mathbb{N}} \subset L^0$  and  $X \in L^0$ .

- (1) If for each  $\epsilon > 0$ ,  $\sum_{n \in \mathbb{N}} P(|X_n - X| \geq \epsilon) < \infty$ , then  $X_n \rightarrow X$  a.e.
- (2) If  $(X_n)_{n \in \mathbb{N}}$  are independent and there exists  $\epsilon > 0$  such that  $\sum_{n \in \mathbb{N}} P(|X_n - X| \geq \epsilon) = \infty$ , then  $X_n \not\rightarrow X$  a.e.

*Proof.*

- (1) For  $\epsilon > 0$  and  $n \in \mathbb{N}$ , set  $A_n(\epsilon) = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}$ . Suppose that for each  $\epsilon > 0$ ,  $\sum_{n \in \mathbb{N}} P(|X_n - X| \geq \epsilon) < \infty$ . The Borel-Cantelli lemma implies that for each  $m \in \mathbb{N}$ ,

$$P(\limsup_{n \rightarrow \infty} A_n(1/m)) = 0$$

Let  $\omega \in \Omega$ . Then  $X_n(\omega) \not\rightarrow X(\omega)$  iff

$$\omega \in \bigcup_{m \in \mathbb{N}} \limsup_{n \rightarrow \infty} A_n(1/m)$$

So

$$\begin{aligned} P(X_n \not\rightarrow X) &= P\left(\bigcup_{m \in \mathbb{N}} \limsup_{n \rightarrow \infty} A_n(1/m)\right) \\ &\leq \sum_{m \in \mathbb{N}} P(\limsup_{n \rightarrow \infty} A_n(1/m)) \\ &= 0 \end{aligned}$$

(2) Hence  $X_n \rightarrow X$  a.e.

□

## 4. PROBABILITY ON LOCALLY COMPACT GROUPS

**Note 4.0.1.** In this section, familiarity with Haar measure will be assumed. This section is intended as a continuation of section 7 of [3].

## 4.1. Action on Probability Measures.

**Note 4.1.1.** We recall some notation from section 7.1 of [3].

- $l_g \in \text{Homeo}(G)$ ,  $l_g(x) = gx$
- $L_g \in \text{Sym}(L_0(G))$ ,  $L_g f = f \circ l_g^{-1}$  We continue from section 7

**Note 4.1.2.** The next exercise generalizes the notion of a scale-family.

**Exercise 4.1.3.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $G$  a locally compact group,  $\mu$  a left Haar measure on  $G$ ,  $X \in L_G^0$  and  $g \in G$ . If  $P_X \ll \mu$ , then  $f_{gX} = L_g f_X$ .

*Proof.* Suppose that  $P_X \ll \mu$ . Let  $A \in \mathcal{B}(G)$ . Then

$$\begin{aligned}
 P_{gX}(A) &= P(gX \in A) \\
 &= P(X \in g^{-1}A) \\
 &= P_X(g^{-1}A) \\
 &= P_X(l_g^{-1}(A)) \\
 &= l_{g*}P_X(A) \\
 &= g \cdot P_X(A)
 \end{aligned}$$

The previous exercise tells us that  $f_{gX} = L_g f_X$ . □



## 5. WEAK CONVERGENCE OF MEASURES

## 6. CONCENTRATION INEQUALITIES

## 6.1. Introduction.

**Exercise 6.1.1.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X, Y \in L^0_{\mathbb{R}}(\Omega, \mathcal{F}, P)$ . Then for each  $s, t \in \mathbb{R}$ ,

$$P(X + Y \geq s + t) \leq P(X \geq s) + P(Y \geq t)$$

*Proof.* For  $Z \in L^0_{\mathbb{R}}(\Omega, \mathcal{F}, P)$  and  $t \in \mathbb{R}$ , define  $A_Z^t \in \mathcal{F}$  by

$$A_Z^t = \{\omega \in \Omega : Z(\omega) \geq t\}$$

Let  $s, t \in \mathbb{R}$ . Since  $(A_X^s)^c \cap (A_Y^t)^c \subset (A_{X+Y}^{s+t})^c$ , we have that  $A_{X+Y}^{s+t} \subset A_X^s \cup A_Y^t$ . Then

$$\begin{aligned} P(X + Y \geq s + t) &= P(A_{X+Y}^{s+t}) \\ &\leq P(A_X^s) + P(A_Y^t) \\ &= P(X \geq s) + P(Y \geq t) \end{aligned}$$

□

**Exercise 6.1.2.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X, Y \in L^+(\Omega, \mathcal{F}, P)$ . Then for each  $s, t \geq 0$ ,

$$P(XY \geq st) \leq P(X \geq s) + P(Y \geq t)$$

*Proof.* For  $Z \in L^0_{\mathbb{R}}(\Omega, \mathcal{F}, P)$  and  $t \in \mathbb{R}$ , define  $A_Z^t \in \mathcal{F}$  by

$$A_Z^t = \{\omega \in \Omega : Z(\omega) \geq t\}$$

Let  $s, t \in \mathbb{R}$ . Since  $(A_X^s)^c \cap (A_Y^t)^c \subset (A_{XY}^{st})^c$ , we have that  $A_{XY}^{st} \subset A_X^s \cup A_Y^t$ . Then

$$\begin{aligned} P(XY \geq st) &= P(A_{XY}^{st}) \\ &\leq P(A_X^s) + P(A_Y^t) \\ &= P(X \geq s) + P(Y \geq t) \end{aligned}$$

□

6.2. Sub  $\alpha$ -Exponential Random Variables.

**Definition 6.2.1.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X \in L^0(\Omega, \mathcal{F}, P)$  and  $\alpha > 0$ . Then  $X$  is said to be **sub  $\alpha$ -exponential** if there exist  $M, K > 0$  such that for each  $t \geq 0$ ,

$$P(|X| \geq t) \leq Me^{-Kt^\alpha}$$

**Exercise 6.2.2.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X \in L^0(\Omega, \mathcal{F}, P)$  and  $\alpha > 0$ . Then the following are equivalent:

- (1)  $X$  is sub  $\alpha$ -exponential
- (2) there exists  $K > 0$  such that for each  $p \geq 1$ ,  $\|X\|_p \leq Kp^{1/\alpha}$
- (3)
- (4)

*Proof.*

- (1)  $\implies$  (2):

Choose  $C_\alpha > 0$  such that for each  $x \geq \alpha^{-1}$ ,  $\Gamma(x) \leq C_\alpha x^\alpha$ . Since  $X$  is sub  $\alpha$ -exponential, there exist  $M, K_0 > 0$  such that for each  $t \geq 0$ ,  $P(|X| \geq t) \leq Me^{-Kt^\alpha}$ . Set  $K = \max(M\alpha^{-1}C_\alpha, 1)2K_0^{-1/\alpha}\alpha^{-1/\alpha}$ . Let  $p \geq 1$ . Then  $p\alpha^{-1} \geq \alpha^{-1}$

$$\begin{aligned}
\|X\|_p^p &= E(|X|^p) \\
&= \int_0^\infty P(|X|^p \geq t) dt \\
&= \int_0^\infty P(|X| \geq t^{1/p}) dt \\
&\leq \int_0^\infty Me^{-K_0 t^{\alpha/p}} dt \\
&= Mp\alpha^{-1} \int_0^\infty u^{p/\alpha-1} e^{-K_0 u} du \\
&= Mp\alpha^{-1} \Gamma(p/\alpha) K_0^{-p/\alpha} \\
&\leq Mp\alpha^{-1} C_\alpha (p\alpha^{-1})^{p/\alpha} K_0^{-p/\alpha}
\end{aligned}$$

Therefore

$$\begin{aligned}
\|X\|_p &\leq (M\alpha^{-1}C_\alpha)^{1/p} p^{1/p} K_0^{-1/\alpha} \alpha^{-1/\alpha} p^{1/\alpha} \\
&\leq \max(M\alpha^{-1}C_\alpha, 1) 2K_0^{-1/\alpha} \alpha^{-1/\alpha} p^{1/\alpha} \\
&= K p^{1/\alpha}
\end{aligned}$$

- (2)  $\implies$  (3):

□

**Definition 6.2.3.** Let  $\psi : [0, \infty) \rightarrow [0, \infty)$ . Then  $\psi$  is said to be an **Orlicz function** if

- (1)  $\psi$  is convex
- (2)  $\psi$  is increasing
- (3)  $\psi(x) \rightarrow \infty$  as  $x \rightarrow \infty$
- (4)  $\psi(0) = 0$

**Definition 6.2.4.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\psi : [0, \infty) \rightarrow [0, \infty)$  and Orlicz function. We define the **Orlicz  $\psi$ -norm**, denoted  $\|\cdot\|_\psi : L^0(\Omega, \mathcal{F}, P) \rightarrow [0, \infty]$ , by

$$\|X\|_\psi = \inf\{t > 0 : E[\psi(|X|/t)] \leq 1\}$$

We define the **Orlicz  $\psi$ -space**, denoted  $L^\psi(\Omega, \mathcal{F}, P)$ , by

$$L^\psi(\Omega, \mathcal{F}, P) = \{X \in L^1(\Omega, \mathcal{F}, P) : \|X\|_\psi < \infty\}$$

**Exercise 6.2.5.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\psi : [0, \infty) \rightarrow [0, \infty)$  an Orlicz function. Then  $L^\psi$  is a vector space and  $\|\cdot\|_\psi : L^\psi \rightarrow [0, \infty)$  is a norm.

**Hint:** note that

- for  $s, t > 0$ ,

$$\frac{|X|}{s+t} + \frac{|Y|}{s+t} = \frac{s}{s+t} \frac{|X|}{s} + \frac{t}{s+t} \frac{|Y|}{t}$$

- $\psi$  is star-shaped, i.e. for each  $x, t \geq 0$  and  $f(tx) \leq tf(x)$ .

*Proof.* For  $X \in L^0(\Omega, \mathcal{F}, P)$ , define  $A_X \in \mathcal{F}$  by

$$A_X = \{t > 0 : E[\psi(|X|/t)] \leq 1\}$$

Let  $X, Y \in L^\psi(\Omega, \mathcal{F}, P)$  and  $\lambda \in \mathbb{C}$ . Since  $\|X\|_\psi < \infty$  and  $\|Y\|_\psi < \infty$ , we have that  $A_X \neq \emptyset$  and  $A_Y \neq \emptyset$ ,

- (1) **subadditivity:** Let  $\epsilon > 0$ . Then there exists  $s \in A_X$  and  $t \in A_Y$  such that  $s < \inf A_X + \epsilon/2$  and  $t < \inf A_Y + \epsilon/2$ . Since  $\psi$  is convex and increasing, we have that

$$\begin{aligned} \psi\left(\frac{|X+Y|}{s+t}\right) &\leq \psi\left(\frac{|X|}{s+t} + \frac{|Y|}{s+t}\right) \\ &= \psi\left(\frac{s}{s+t} \frac{|X|}{s} + \frac{t}{s+t} \frac{|Y|}{t}\right) \\ &\leq \frac{s}{s+t} \psi\left(\frac{|X|}{s}\right) + \frac{t}{s+t} \psi\left(\frac{|Y|}{t}\right) \end{aligned}$$

Therefore,

$$\begin{aligned} E\left[\psi\left(\frac{|X+Y|}{s+t}\right)\right] &\leq \frac{s}{s+t} E\left[\psi\left(\frac{|X|}{s}\right)\right] + \frac{t}{s+t} E\left[\psi\left(\frac{|Y|}{t}\right)\right] \\ &\leq \frac{s}{s+t} + \frac{t}{s+t} \\ &\leq 1 \end{aligned}$$

Hence  $s+t \in A_{X+Y}$ . Thus  $A_{X+Y} \neq \emptyset$ . Since  $s < \inf A_X + \epsilon/2$  and  $t < \inf A_Y + \epsilon/2$ , we have that

$$\begin{aligned} \|X+Y\|_\psi &= \inf A_{X+Y} \\ &\leq s+t \\ &< \inf A_X + \inf A_Y + \epsilon \\ &= \|X\|_\psi + \|Y\|_\psi + \epsilon \end{aligned}$$

Since  $\epsilon > 0$  is arbitrary,  $\|X+Y\|_\psi \leq \|X\|_\psi + \|Y\|_\psi$ . So  $X+Y \in L^\psi(\Omega, \mathcal{F}, P)$

- (2) **absolute homogeneity:** Suppose that  $\lambda = 0$ . Then for each  $t > 0$ ,

$$\begin{aligned} E[\psi(|\lambda X|/t)] &= E[\psi(0)] \\ &= E[0] \\ &= 0 \\ &\leq 1 \end{aligned}$$

Thus

$$\begin{aligned} \|\lambda X\|_\psi &= 0 \\ &= |\lambda| \|X\|_\psi \end{aligned}$$

Suppose that  $\lambda \neq 0$ . Since for each  $t > 0$ ,

$$\psi\left(\frac{|X|}{t}\right) = \psi\left(\frac{|\lambda X|}{|\lambda|t}\right)$$

we have that for each  $t > 0$ ,  $t \in A_X$  iff  $|\lambda|t \in A_{\lambda X}$ . Therefore  $A_{\lambda X} = |\lambda|A_X$  and

$$\begin{aligned}\|\lambda X\|_\psi &= \inf A_{\lambda X} \\ &= |\lambda| \inf A_X \\ &= |\lambda| \|X\|_\psi\end{aligned}$$

So  $\|\lambda X\| \in L^\psi(\Omega, \mathcal{F}, P)$ .

- (3) **positive definiteness:** Note that since  $\psi$  is increasing, for each  $t_0 > 0$ ,  $t_0 \in A_X$  implies that  $[t_0, \infty) \subset A_X$ . Suppose that  $\|X\|_\psi = 0$ . Then  $(0, \infty) \subset A_X$ . Jensen's inequality implies that for each  $t > 0$ ,

$$\begin{aligned}\psi\left(\frac{E|X|}{t}\right) &\leq E\left[\psi\left(\frac{|X|}{t}\right)\right] \\ &\leq 1\end{aligned}$$

For the sake of contradiction, suppose that  $E|X| > 0$ . Since  $\psi(x) \rightarrow \infty$  as  $x \rightarrow \infty$ , Choose  $x > 0$  such that  $\psi(x) > 1$ . Set  $t = E|X|/x$ . Then  $t > 0$  and

$$\begin{aligned}1 &< \psi(x) \\ &= \psi\left(\frac{E|X|}{t}\right) \\ &\leq 1\end{aligned}$$

which is a contradiction. Hence  $E|X| = 0$ . Thus  $X = 0$  a.s. □

**Exercise 6.2.6.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\psi : [0, \infty) \rightarrow [0, \infty)$  an Orlicz function. Then  $L^\psi(\Omega, \mathcal{F}, P)$  is a Banach space.

*Proof.* □

## 7. CONDITIONAL EXPECTATION AND PROBABILITY

## 7.1. Conditional Expectation.

**Exercise 7.1.1.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $\mathcal{G}$  a sub  $\sigma$ -algebra of  $\mathcal{F}$  and  $X \in L^1(\Omega, \mathcal{F}, P)$ . Define  $P_{\mathcal{G}} = P|_{\mathcal{G}}$  and  $Q : \mathcal{G} \rightarrow [0, \infty)$  by  $Q(G) = \int_G X dP$ . Then  $Q \ll P_{\mathcal{G}}$ .

*Proof.* Let  $G \in \mathcal{G}$ . Suppose that  $P_{\mathcal{G}}(G) = 0$ . By definition,  $P(G) = 0$ . So  $Q(G) = 0$  and  $Q \ll P_{\mathcal{G}}$ .  $\square$

**Definition 7.1.2.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $\mathcal{G}$  a sub  $\sigma$ -algebra of  $\mathcal{F}$  and  $X, Y \in L^1(\Omega, \mathcal{F}, P)$ . Then  $Y$  is said to be a **conditional expectation of  $X$  given  $\mathcal{G}$**  if

- (1)  $Y$  is  $\mathcal{G}$ -measurable
- (2) for each  $G \in \mathcal{G}$ ,

$$\int_G Y dP = \int_G X dP$$

Since (2) implies that conditional expectations of  $X$  given  $\mathcal{G}$  are equal  $P_{\mathcal{G}}$ -a.e., we write  $Y = E(X|\mathcal{G})$ .

**Note 7.1.3.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $(S, \mathcal{S})$  a measurable space,  $X \in L^1(\Omega, \mathcal{F}, P)$  and  $Y \in L^0_S(\Omega, \mathcal{F})$ . We typically write  $E(X|Y)$  instead of  $E(X|Y^*\mathcal{S})$ .

**Exercise 7.1.4. Existence of Conditional Expectation:**

Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $\mathcal{G}$  a sub  $\sigma$ -algebra of  $\mathcal{F}$  and  $X \in L^1(\Omega, \mathcal{F}, P)$ . Define  $Q$  and  $P_{\mathcal{G}}$  as in the previous exercise. Define  $Y = dQ/dP_{\mathcal{G}}$ . Then  $Y$  is a conditional expectation of  $X$  given  $\mathcal{G}$ .

*Proof.* The Radon-Nikodym theorem implies that  $Y$  is  $\mathcal{G}$ -measurable. Since  $Q$  is finite, so is  $|Q|$ . Since  $d|Q| = |Y| dP_{\mathcal{G}}$ , we have that  $Y \in L^1(\Omega, \mathcal{G}, P_{\mathcal{G}})$ . An exercise in section 3.3 of [3], implies that for each  $G \in \mathcal{G}$

$$\begin{aligned} \int_G Y dP &= \int_G Y dP_{\mathcal{G}} \\ &= Q(G) \\ &= \int_G X dP \end{aligned}$$

$\square$

**Definition 7.1.5.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $(S, \mathcal{S})$  a measurable space,  $X \in L^1(\Omega, \mathcal{F}, P)$  and  $Y \in L^0_S(\Omega, \mathcal{F})$ . Let  $\phi \in L^0(Y(\Omega), \mathcal{S} \cap Y(\Omega))$ . Then  $\phi$  is said to be a **conditional expectation function of  $X$  given  $Y$**  if for each  $B \in \mathcal{S} \cap Y(\Omega)$ ,

$$\int_{Y^{-1}(B)} X dP = \int_B \phi dP_Y$$

To denote this, we write  $\phi(y) = E[X|Y = y]$ .

**Exercise 7.1.6. Existence of Conditional Expectation Function:**

Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $(S, \mathcal{S})$  a measurable space,  $X \in L^1(\Omega, \mathcal{F}, P)$  and  $Y \in L^0_S(\Omega, \mathcal{F})$ . Suppose that for each  $y \in S$ ,  $\{y\} \in \mathcal{S}$ . Then there exists  $\phi \in L^0(Y(S), \mathcal{S} \cap Y(\Omega))$  such that  $\phi$  is a conditional expectation function of  $X$  given  $Y$ .

**Hint:** Doob-Dynkin lemma

*Proof.* Since  $E[X|Y] \in L^0(\Omega, Y^*\mathcal{S})$ , the Doob-Dynkin lemma implies that there exists  $\phi \in L^0(Y(\Omega), \mathcal{S} \cap Y(\Omega))$  such that  $\phi \circ Y = E(X|Y)$ . Let  $B \in \mathcal{S} \cap Y(\Omega)$ . Then

$$\begin{aligned} \int_B \phi dP_Y &= \int_{Y^{-1}(B)} \phi \circ Y dP \\ &= \int_{Y^{-1}(B)} E(X|Y) dP \\ &= \int_{Y^{-1}(B)} X dP \end{aligned}$$

□

## 7.2. Conditional Probability.

**Definition 7.2.1.** Let  $(A, \mathcal{A})$  be a measurable space,  $(B, \mathcal{B}, P_Y)$  a probability space and  $Q : B \times \mathcal{A} \rightarrow [0, 1]$ . Then  $Q$  is said to be a **stochastic transition kernel from  $(B, \mathcal{B}, P)$  to  $(A, \mathcal{A})$**  if

- (1) for each  $E \in \mathcal{A}$ ,  $Q(\cdot, E)$  is  $\mathcal{B}$ -measurable
- (2) for  $P$ -a.e.  $b \in B$ ,  $Q(b, \cdot)$  is a probability measure on  $(A, \mathcal{A})$

**Definition 7.2.2.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X, Y \in L_n^0(\Omega, \mathcal{F}, P)$  and  $Q : \mathbb{R}^n \times \mathcal{F} \rightarrow [0, 1]$ . Then  $Q$  is said to be a **conditional probability distribution of  $X$  given  $Y$**  if

- (1)  $Q$  is a stochastic transition kernel from  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_Y)$  to  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$
- (2) for each  $A, B \in \mathcal{F}$ ,

$$\int_B Q(y, A) dP_Y(y) = P(X \in A, Y \in B)$$

**Note 7.2.3.** It is helpful to connect this notion of conditional probability with the elementary one by writing  $Q(y, A) = P(X \in A | Y = y)$ . If  $P_Y \ll \mu$ , then property (2) in the definition becomes

$$\begin{aligned} P(X \in A, Y \in B) &= \int_B Q(y, A) dP_Y(y) \\ &= \int_B P(X \in A | Y = y) f_Y(y) d\mu(y) \end{aligned}$$

as in a first course on probability.

**Exercise 7.2.4.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X, Y \in L_n^0$  and  $Q : \mathbb{R}^n \times \mathcal{F} \rightarrow [0, 1]$ . Suppose that for each  $A \in \mathcal{F}$ ,  $Q(\cdot, A)$  is  $\mathcal{B}(\mathbb{R}^n)$ -measurable, for  $P_Y$ -a.e.  $y \in \mathbb{R}^n$ ,  $P_{X|Y}(y, \cdot)$  is a probability measure on  $(\Omega, \mathcal{F})$  and  $Q(Y, A) = P(X \in A | Y)$  a.e. Then  $Q$  is a conditional probability of  $X$  given  $Y$ .

*Proof.* By assumption, for each  $A \in \mathcal{F}$ ,  $Q(\cdot, A)$  is  $\mathcal{B}(\mathbb{R}^n)$ -measurable and for  $P_Y$ -a.e.  $y \in \mathbb{R}^n$ ,  $Q(y, \cdot)$  is a probability measure on  $(\Omega, \mathcal{F})$ . Let  $A, B \in \mathcal{F}$ . Then

$$\begin{aligned}
 \int_B Q(y, A) dP_Y(y) &= \int_{Y^{-1}(B)} Q(Y(\omega), A) dP(\omega) \\
 &= \int_{Y^{-1}(B)} P(X \in A | Y) dP \\
 &= \int_{Y^{-1}(B)} E[1_{X^{-1}(A)} | Y] dP \\
 &= \int_{Y^{-1}(B)} 1_{X^{-1}(A)} dP \\
 &= \int 1_{X^{-1}(A)} 1_{Y^{-1}(B)} dP \\
 &= \int 1_{X^{-1}(A) \cap Y^{-1}(B)} dP \\
 &= P(X \in A, Y \in B)
 \end{aligned}$$

So  $Q$  is a conditional probability distribution of  $X$  given  $Y$ . □

**Definition 7.2.5.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X, Y \in L_n^0$  and  $\mu$  a  $\sigma$ -finite measure on  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ . Suppose that  $P_X, P_Y \ll \mu$ . Then  $P_{X,Y} \ll \mu^2$ . Let  $f_X = dP_X/d\mu$ ,  $f_Y = dP_Y/d\mu$  and  $f_{X,Y} = dP_{X,Y}/d\mu^2$ . Define  $f_{X|Y} : \mathbb{R}^n \times \mathbb{R}^n$  by

$$f_{X|Y}(x, y) = \begin{cases} \frac{f_{X,Y}(x, y)}{f_Y(y)}, & y \in \text{supp } Y \\ 0, & y \notin \text{supp } Y \end{cases}$$

Then  $f_{X|Y}$  is called the **conditional probability density of  $X$  given  $Y$** .

**Exercise 7.2.6.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X, Y \in L_n^0$  and  $\mu$  a  $\sigma$ -finite measure on  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ . Suppose that  $P_X, P_Y \ll \mu$ . Define  $Q : \mathbb{R}^n \times \mathcal{F} \rightarrow [0, 1]$  by

$$Q(y, A) = \int_A f_{X|Y}(x, y) d\mu(x)$$

Then  $Q$  is a conditional probability distribution of  $X$  given  $Y$ .

*Proof.* By the Fubini-Tonelli Theorem, for each  $A \in \mathcal{F}$ ,  $Q(\cdot, A)$  is  $\mathcal{B}(\mathbb{R}^n)$ -measurable and for  $P_Y$ -a.e.  $y \in \mathbb{R}^n$ ,  $Q(y, \cdot)$  is a probability measure on  $(\Omega, \mathcal{F})$ . Let  $A, B \in \mathcal{F}$ . Then

$$\begin{aligned}
 \int_B Q(y, A) dP_Y(y) &= \int_B \left[ \int_A f_{X|Y}(x, y) d\mu(x) \right] dP_Y(y) \\
 &= \int_{B \cap \text{supp } Y} \left[ \int_{A \cap \text{supp } Y} \frac{f_{X,Y}(x, y)}{f_Y(y)} d\mu(x) f_Y(y) \right] d\mu(y) \\
 &= \int_{B \cap \text{supp } Y} \left[ \int_A f_{X,Y}(x, y) d\mu(x) \right] d\mu(y) \\
 &= P(X \in A, Y \in B \cap \text{supp } Y) \\
 &= P(X \in A, Y \in B)
 \end{aligned}$$



□

**Theorem 7.2.7.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X, Y \in L_n^1(\Omega, \mathcal{F}, P)$ . Suppose that  $\text{Im } X \in \mathcal{B}(\mathbb{R}^n)$ . Then there exists a conditional probability distribution of  $Y$  given  $X$ .

## 8. MARKOV CHAINS

**Definition 8.0.1.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $(X_n)_{n \in \mathbb{N}_0} \in L_n^0$ . Then  $(X_n)_{n \in \mathbb{N}_0}$  is said to be a **homogeneous Markov chain** if for each  $A \in \mathcal{F}$  and  $n \in \mathbb{N}$ ,  $P(X_n \in A | X_1, \dots, X_{n-1}) = P(X_1 \in A | X_0)$  a.e.

## 9. PROBABILITIES INDUCED BY ISOMETRIC GROUP ACTIONS

## 9.1. Applications to Bayesian Statistics.

**Exercise 9.1.1.** Let  $(\mathcal{X}, \mathcal{A})$  be a measurable space  $(\Theta, d)$  a metric space,  $G$  a group,  $\phi : G \times \Theta \rightarrow \Theta$  an isometric group action. Suppose that  $\bar{d}$  is a metric on  $\Theta/G$ . Let

- $H_p^\Theta$  be the Hausdorff measure on  $\Theta$ ,  $\mu_{\mathcal{X}}$  a measure on  $\mathcal{X}$ ,
- $p$  a density on  $\Theta$  and for each  $\theta \in \Theta$ ,  $p(\cdot|\theta)$  a density on  $\mathcal{X}$ .
- $\theta_0 \in \Theta$  and for  $j \in \mathbb{N}$ ,  $X_j \sim p(x|\theta_0)$

Suppose that  $\mu_\Theta$  is  $G$ -invariant,  $p$  is  $G$ -invariant and continuous on  $\Theta$  and for each  $x \in \mathcal{X}$ ,  $p(x|\cdot)$  is  $G$ -invariant and continuous on  $\Theta$ . For  $n \in \mathbb{N}$ , set  $p(\cdot|X^{(n)}) \propto f(X_1, \dots, X_n|\cdot)p(\cdot)$ . Define the posterior measure  $P_{\Theta|X^{(n)}} : \mathcal{B}(\Theta) \rightarrow [0, 1]$  by

$$dP_{\Theta|X^{(n)}}(\theta) = p(\theta|X^{(n)}) dH_p^\Theta(\theta)$$

Then there exists a density  $\bar{p}(\cdot|X^{(n)})$  on  $\Theta/G$  such that

$$d\bar{P}_{\Theta|X^{(n)}}(\theta) = \bar{p}(\theta|X^{(n)}) d\bar{H}^\Theta(\theta)$$

*Proof.* Clear from previous work. □

**Exercise 9.1.2.** Let  $(\mathcal{X}, \mathcal{A})$  be a measurable space  $(\Theta, d)$  a metric space,  $G$  a group,  $\phi : G \times \Theta \rightarrow \Theta$  an isometric group action. Suppose that  $\bar{d}$  is a metric on  $\Theta/G$ . Let

- $H_p^\Theta$  be the Hausdorff measure on  $\Theta$ ,  $\mu_{\mathcal{X}}$  a measure on  $\mathcal{X}$ ,
- $p$  a density on  $\Theta$  and for each  $\theta \in \Theta$ ,  $p(\cdot|\theta)$  a density on  $\mathcal{X}$ .
- $\theta_0 \in \Theta$  and for  $j \in \mathbb{N}$ ,  $X_j \sim p(x|\theta_0)$

Suppose  $p$  is  $G$ -invariant and continuous on  $\Theta$  and for each  $x \in \mathcal{X}$ ,  $p(x|\cdot)$  is  $G$ -invariant and continuous on  $\Theta$ . For  $n \in \mathbb{N}$ , set  $p(\cdot|X^{(n)}) \propto f(X_1, \dots, X_n|\cdot)p(\cdot)$ . Define the posterior measure  $P_{\Theta|X^{(n)}} : \mathcal{B}(\Theta) \rightarrow [0, 1]$  by

$$dP_{\Theta|X^{(n)}}(\theta) = p(\theta|X^{(n)}) dH_p^\Theta(\theta)$$

Suppose that  $(P_{\Theta|X^{(n)}})_{n \in \mathbb{N}}$  concentrates on  $\bar{\theta}_0 \subset \Theta$  a.s. or in probability. Then  $(\bar{P}_{\Theta|X^{(n)}})_{n \in \mathbb{N}}$  concentrates a.s. or in probability on  $\{\bar{\theta}_0\} \subset \Theta/G$  (i.e. is consistent a.s. or in probability)

*Proof.* Let  $V \in \mathcal{N}_{\bar{\theta}_0}$ . Then  $\pi^{-1}(V) \in \mathcal{N}_{\bar{\theta}_0}$ . By definition,

$$\begin{aligned} \bar{P}_{\Theta|X^{(n)}}(V^c) &= P_{\Theta|X^{(n)}}(\pi^{-1}(V^c)) \\ &= P_{\Theta|X^{(n)}}(\pi^{-1}(V)^c) \\ &\xrightarrow{\text{a.s./}P} 0 \end{aligned}$$

□

**Note 9.1.3.** Some examples of  $G$ -invariant priors would be the uniform distribution, or  $N_n(0, \sigma^2 I)$  on  $\mathbb{R}^n$  when acted on by  $O(n)$ . An example of a  $G$ -invariant likelihood would be  $f(A|Z) \sim \text{Ber}(ZZ^T)$  as in a latent position random graph model where  $Z \in \mathbb{R}^{n \times d}$  is the parameter is invariant under right multiplication by  $U \in O_d$ .

## 10. STOCHASTIC INTEGRATION

**Exercise 10.0.1.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X$  a set  $\mathcal{A}_0$  an algebra,  $\mu_0 : \mathcal{A}_0 \rightarrow \mathbb{C}$  and  $B : \mathcal{A}_0 \rightarrow L^2(\Omega, \mathcal{F}, P)$ . Suppose that

- (1)  $B(\emptyset) = 0$
- (2) for each  $E, F \in \mathcal{A}_0$ , if  $E \cap F = \emptyset$ , then  $B(E \cup F) = B(E) + B(F)$
- (3)  $E(B(E)B(F)^*) = \mu_0(E \cap F)$

Then

- (1) for each  $E \in \mathcal{A}_0$ ,  $\mu_0(E) = E(|B(E)|^2)$ .
- (2) for each  $E \in \mathcal{A}_0$ ,  $0 \leq \mu_0(E) < \infty$
- (3) for each  $E, F \in \mathcal{A}_0$ , if  $E \cap F = \emptyset$ , then  $\mu_0(E \cup F) = \mu_0(E) + \mu_0(F)$

*Proof.*

- (1) Clear
- (2) Clear
- (3) Let  $E, F \in \mathcal{A}_0$ . Suppose that  $E \cap F = \emptyset$ . Then

$$\begin{aligned}
 E(B(E)B(F)^*) &= \mu_0(E \cap F) \\
 &= \mu_0(\emptyset) \\
 &= E(|B(\emptyset)|^2) \\
 &= E(0) \\
 &= 0
 \end{aligned}$$

This implies that

$$\begin{aligned}
 \mu_0(E \cup F) &= E(|B(E \cup F)|^2) \\
 &= E(|B(E) + B(F)|^2) \\
 &= E(|B(E)|^2) + E(|B(F)|^2) + 2\operatorname{Re}E(B(E)B(F)^*) \\
 &= \mu_0(E) + \mu_0(F) + 0 \\
 &= \mu_0(E) + \mu_0(F)
 \end{aligned}$$

□

**Definition 10.0.2.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X$  a set  $\mathcal{A}_0$  an algebra,  $\mu_0 : \mathcal{A}_0 \rightarrow [0, \infty)$  a premeasure and  $B : \mathcal{A}_0 \rightarrow L^2(\Omega, \mathcal{F}, P)$ . Suppose that

- (1)  $B(\emptyset) = 0$
- (2) for each  $E, F \in \mathcal{A}_0$ , if  $E \cap F = \emptyset$ , then  $B(E \cup F) = B(E) + B(F)$
- (3)  $E(B(E)B(F)^*) = \mu_0(E \cap F)$

Then  $B$  is said to be a **stochastic premeasure with sturcture**  $\mu_0$

## 11. TODO

Incorporate vector measures to have conditional expectations of Banach space valued functions given a  $\sigma$ -algebra

## REFERENCES

- [1] [Introduction to Analysis](#)
- [2] [Introduction to Group Theory](#)
- [3] [Introduction to Measure and Integration](#)