

# Notes on Measure-Theoretic Probability

Paulo Eduardo Rauber

2022

## 1 Measure spaces

A set  $S$  contains  $s$  if  $s \in S$ . A set  $S$  includes  $F$  if  $F \subseteq S$ .

An algebra  $\Sigma_0$  on a set  $S$  is a set of subsets of  $S$  such that

- $S \in \Sigma_0$ ,
- If  $F \in \Sigma_0$ , then  $F^c \in \Sigma_0$ , where  $F^c = S \setminus F$ ,
- If  $F, G \in \Sigma_0$ , then  $F \cup G \in \Sigma_0$ .

Consequently, if  $\Sigma_0$  is an algebra on  $S$ ,

- $\emptyset \in \Sigma_0$ ,
- If  $F, G \in \Sigma_0$ , then  $F \cap G \in \Sigma_0$ .

A trivial algebra on  $S$  is given by  $\{\emptyset, S\}$ .

A  $\sigma$ -algebra  $\Sigma$  on  $S$  is an algebra on  $S$  such that

$$\bigcup_{n \in \mathbb{N}} F_n \in \Sigma$$

for any sequence  $(F_n \in \Sigma \mid n \in \mathbb{N})$ , which also implies

$$\bigcap_{n \in \mathbb{N}} F_n \in \Sigma.$$

A measurable space  $(S, \Sigma)$  is a pair composed of a set  $S$  and a  $\sigma$ -algebra  $\Sigma$  on  $S$ . An element of  $\Sigma$  is called a  $\Sigma$ -measurable subset of  $S$ .

Let  $\mathcal{C}$  be a set of subsets of  $S$ . The  $\sigma$ -algebra  $\sigma(\mathcal{C})$  generated by  $\mathcal{C}$  is the smallest  $\sigma$ -algebra  $\Sigma$  on  $S$  such that  $\mathcal{C} \subseteq \Sigma$ . The  $\sigma$ -algebra  $\sigma(\mathcal{C})$  is the intersection of all the  $\sigma$ -algebras on  $S$  that include  $\mathcal{C}$ . Note that the set  $\mathcal{P}(S)$  of all subsets of  $S$  is a  $\sigma$ -algebra on  $S$  that includes any set of subsets  $\mathcal{C}$ .

The Borel  $\mathcal{B}(\mathbb{R})$   $\sigma$ -algebra is the  $\sigma$ -algebra on  $\mathbb{R}$  generated by the set of open sets of real numbers.

Let  $\pi(\mathbb{R}) = \{(-\infty, x] \mid x \in \mathbb{R}\}$  be the set that contains every interval that contains every real number smaller or equal to every real number  $x \in \mathbb{R}$ . We will now show that the  $\sigma$ -algebra generated by  $\pi(\mathbb{R})$  is  $\sigma(\pi(\mathbb{R})) = \mathcal{B}(\mathbb{R})$ .

First, recall that  $(-\infty, x] = \bigcap_{n \in \mathbb{N}^+} (-\infty, x + n^{-1})$ . Because  $\mathcal{B}(\mathbb{R})$  is a  $\sigma$ -algebra on  $\mathbb{R}$  that contains every  $(-\infty, x + n^{-1})$ , we have  $(-\infty, x] \in \mathcal{B}(\mathbb{R})$ . Because  $\mathcal{B}(\mathbb{R})$  is a  $\sigma$ -algebra on  $\mathbb{R}$  that includes  $\pi(\mathbb{R})$  and  $\sigma(\pi(\mathbb{R}))$  is the smallest  $\sigma$ -algebra on  $\mathbb{R}$  that includes  $\pi(\mathbb{R})$ , we have  $\sigma(\pi(\mathbb{R})) \subseteq \mathcal{B}(\mathbb{R})$ .

Second, recall that every open set of real numbers is a countable union of open intervals. Because  $\sigma(\pi(\mathbb{R}))$  is a  $\sigma$ -algebra on  $\mathbb{R}$ , if it contains every open interval, then it contains every open set of real numbers. This would also imply that  $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\pi(\mathbb{R}))$ , since  $\sigma(\pi(\mathbb{R}))$  is a  $\sigma$ -algebra on  $\mathbb{R}$  and  $\mathcal{B}(\mathbb{R})$  is the smallest  $\sigma$ -algebra on  $\mathbb{R}$  that contains every open set of real numbers. In order to show that  $\sigma(\pi(\mathbb{R}))$  contains every open interval, first note that  $(a, u] = (-\infty, u] \cap (-\infty, a]^c \in \sigma(\pi(\mathbb{R}))$  for any  $u > a$  and then note that  $(a, b) = \bigcup_{n \in \mathbb{N}^+} (a, b - \epsilon n^{-1}]$  for  $\epsilon = (b - a)/2$ .

Consider an algebra  $\Sigma_0$  on a set  $S$ . A function  $\mu_0 : \Sigma_0 \rightarrow [0, \infty]$  is called additive if  $\mu_0(\emptyset) = 0$  and, for any  $F, G \in \Sigma_0$  such that  $F \cap G = \emptyset$ ,

$$\mu_0(F \cup G) = \mu_0(F) + \mu_0(G).$$

A function  $\mu_0 : \Sigma_0 \rightarrow [0, \infty]$  is called countably additive if  $\mu_0(\emptyset) = 0$  and, for any sequence  $(F_n \in \Sigma_0 \mid n \in \mathbb{N})$  such that  $F_n \cap F_m = \emptyset$  for  $n \neq m$ ,

$$\mu_0\left(\bigcup_{n \in \mathbb{N}} F_n\right) = \sum_{n \in \mathbb{N}} \mu_0(F_n)$$

whenever  $\bigcup_{n \in \mathbb{N}} F_n \in \Sigma_0$ . This last requirement is always met when  $\Sigma_0$  is a  $\sigma$ -algebra.

Let  $(S, \Sigma)$  be a measurable space. A countably additive function  $\mu : \Sigma \rightarrow [0, \infty]$  is called a measure on  $(S, \Sigma)$ . The triple  $(S, \Sigma, \mu)$  is called a measure space, which has the following properties:

- If  $\mu(S) < \infty$  and  $A, B \in \Sigma$ , then  $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$ ,
- If  $A, B \in \Sigma$ , then  $\mu(A \cup B) \leq \mu(A) + \mu(B)$ ,
- $\mu(\bigcup_{n \in \mathbb{N}} F_n) \leq \sum_{n \in \mathbb{N}} \mu(F_n)$  for any sequence  $(F_n \in \Sigma \mid n \in \mathbb{N})$ .

A measure  $\mu$  on the measurable space  $(S, \Sigma)$  is called finite if  $\mu(S) < \infty$ . A measure  $\mu$  on the measurable space  $(S, \Sigma)$  is called  $\sigma$ -finite if there is a sequence  $(S_n \in \Sigma \mid n \in \mathbb{N})$  such that  $\mu(S_n) < \infty$  and  $\bigcup_{n \in \mathbb{N}} S_n = S$ .

A measure  $\mu$  on the measurable space  $(S, \Sigma)$  is called a probability measure if  $\mu(S) = 1$ . The triple  $(S, \Sigma, \mu)$  is then called a probability triple. A set  $F \in \Sigma$  is called  $\mu$ -null if  $\mu(F) = 0$ . If a statement is false only for elements of a  $\mu$ -null set  $F \in \Sigma$ , then the statement is said to be true almost everywhere.

A  $\pi$ -system  $\mathcal{I}$  on  $S$  is a set of subsets of  $S$  such that if  $I_1, I_2 \in \mathcal{I}$ , then  $I_1 \cap I_2 \in \mathcal{I}$ . Let  $\Sigma = \sigma(\mathcal{I})$  be the  $\sigma$ -algebra generated by a  $\pi$ -system  $\mathcal{I}$ . If  $\mu_1$  and  $\mu_2$  are measures on the measurable space  $(S, \Sigma)$  such that  $\mu_1(S) = \mu_2(S) < \infty$  and  $\mu_1(I) = \mu_2(I)$  for any  $I \in \mathcal{I}$ , then  $\mu_1(F) = \mu_2(F)$  for any  $F \in \Sigma$ . Therefore, if two probability measures agree on a  $\pi$ -system, then they agree on the  $\sigma$ -algebra generated by that  $\pi$ -system.

Carathéodory's extension theorem states that if  $\Sigma_0$  is an algebra on  $S$  and  $\Sigma = \sigma(\Sigma_0)$  is the  $\sigma$ -algebra generated by  $\Sigma_0$  and  $\mu_0 : \Sigma_0 \rightarrow [0, \infty]$  is a countably additive function, then there exists a measure  $\mu$  on the measurable space  $(S, \Sigma)$  such that  $\mu(F) = \mu_0(F)$  for any  $F \in \Sigma_0$ . If  $\mu_0(S) < \infty$ , then  $\mu$  is unique, since an algebra is a  $\pi$ -system.

Let  $\Sigma_0$  be the algebra on the set  $S = (0, 1]$  that contains every  $F$  such that

$$F = \bigcup_{k=1}^r (a_k, b_k],$$

where  $r \in \mathbb{N}$  and  $0 \leq a_1 \leq b_1 \leq \dots \leq a_r \leq b_r \leq 1$ .

Let  $\mu_0 : \Sigma_0 \rightarrow [0, 1]$  denote the countably additive function given by

$$\mu_0(F) = \sum_{k=1}^r (b_k - a_k).$$

Let  $\mathcal{B}((0, 1]) = \sigma(\Sigma_0)$  be the  $\sigma$ -algebra generated by  $\Sigma_0$ . The unique measure  $\mu : \mathcal{B}((0, 1]) \rightarrow [0, 1]$  on the measurable space  $((0, 1], \mathcal{B}((0, 1]))$  that agrees with  $\mu_0$  on the algebra  $\Sigma_0$  is called the Lebesgue measure  $\text{Leb}$  on  $((0, 1], \mathcal{B}((0, 1]))$ . The  $\sigma$ -finite Lebesgue measure  $\text{Leb}$  on the measurable space  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  is similarly defined. Intuitively, a Lebesgue measure  $\text{Leb}$  assigns *lengths*.

Let  $a_n \uparrow a$  denote that a sequence of real numbers  $(a_n \mid n \in \mathbb{N})$  is such that  $a_n \leq a_{n+1}$  and  $a = \lim_{n \rightarrow \infty} a_n$ . Similarly, let  $a_n \downarrow a$  denote that a sequence of real numbers  $(a_n \mid n \in \mathbb{N})$  is such that  $a_{n+1} \leq a_n$  and  $a = \lim_{n \rightarrow \infty} a_n$ .

Let  $A_n \uparrow A$  denote that a sequence of sets  $(A_n \mid n \in \mathbb{N})$  is such that  $A_n \subseteq A_{n+1}$  and  $A = \bigcup_{n \in \mathbb{N}} A_n$ . Similarly, let  $A_n \downarrow A$  denote that a sequence of sets  $(A_n \mid n \in \mathbb{N})$  is such that  $A_{n+1} \subseteq A_n$  and  $A = \bigcap_{n \in \mathbb{N}} A_n$ .

Consider the measure space  $(S, \Sigma, \mu)$ . For a sequence  $(F_n \in \Sigma \mid n \in \mathbb{N})$ , the monotone-convergence property of measure guarantees that if  $F_n \uparrow F$ , then  $\mu(F_n) \uparrow \mu(F)$ . Similarly, for a sequence  $(G_n \in \Sigma \mid n \in \mathbb{N})$ , if  $G_n \downarrow G$  and  $\mu(G_k) < \infty$  for some  $k$ , then  $\mu(G_n) \downarrow \mu(G)$ .

## 2 Events

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ . An element  $\omega \in \Omega$  is called an outcome. The set  $\Omega$  is called an outcome space. A set of outcomes  $F \in \mathcal{F}$  is called an event. The probability measure  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  is defined on a  $\sigma$ -algebra  $\mathcal{F}$  on the outcome space  $\Omega$ .

A probability  $\mathbb{P}(F)$  assigns a degree of belief to the statement that the outcome  $\omega \in \Omega$  of an experiment belongs to the event  $F \in \mathcal{F}$ . For instance, a probability  $\mathbb{P}(F) = 1$  indicates that  $\omega \in F$  almost surely, while a probability  $\mathbb{P}(F) = 0$  indicates that  $\omega \notin F$  almost surely. In general, a statement about an outcome is said to be true almost surely if  $\mathbb{P}(F) = 1$ , where  $F \in \mathcal{F}$  is the event that contains every outcome  $\omega \in \Omega$  for which the statement is true.

As an example, consider an experiment where a coin is tossed twice. Let  $H = 0$  represent heads and  $T = 1$  represent tails. The outcome space  $\Omega$  may be defined as  $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$ . The  $\sigma$ -algebra  $\mathcal{F}$  on the outcome space  $\Omega$  may be defined as the set of all subsets of  $\Omega$ , which is denoted by  $\mathcal{F} = \mathcal{P}(\Omega)$ . The event  $F$  where at least one head is observed is then given by  $F = \{(H, H), (H, T), (T, H)\}$ .

More interestingly, consider an experiment where a coin is tossed infinitely often. The outcome space  $\Omega$  may be defined as the set of infinite binary sequences  $\Omega = \{H, T\}^{\mathbb{N}}$ . In order to at least assign probabilities to every event  $F = \{\omega \in \Omega \mid \omega_n = W\}$  where  $n \in \mathbb{N}$  and  $W \in \{H, T\}$ , the  $\sigma$ -algebra  $\mathcal{F}$  on the outcome space  $\Omega$  may be generated as  $\mathcal{F} = \sigma(\{\{\omega \in \Omega \mid \omega_n = W\} \mid n \in \mathbb{N}, W \in \{H, T\}\})$ .

Consider a sequence of events  $(F_n \in \mathcal{F} \mid n \in \mathbb{N})$ . If  $\mathbb{P}(F_n) = 1$  for every  $n \in \mathbb{N}$ , then  $\mathbb{P}(\cap_{n \in \mathbb{N}} F_n) = 1$ .

The infimum  $\inf_n x_n$  of a sequence of real numbers  $(x_n \in \mathbb{R} \mid n \in \mathbb{N})$  is the largest  $r \in [-\infty, \infty]$  such that  $r \leq x_n$  for every  $n \in \mathbb{N}$ . The supremum  $\sup_n x_n$  of a sequence of real numbers  $(x_n \in \mathbb{R} \mid n \in \mathbb{N})$  is the smallest  $r \in [-\infty, \infty]$  such that  $r \geq x_n$  for every  $n \in \mathbb{N}$ .

The limit inferior of a sequence of real numbers  $(x_n \in \mathbb{R} \mid n \in \mathbb{N})$  is defined by

$$\liminf_{n \rightarrow \infty} x_n = \sup_m \inf_{n \geq m} x_n = \lim_{m \rightarrow \infty} \inf_{n \geq m} x_n.$$

Note that the sequence  $(\inf_{n \geq m} x_n \mid m \in \mathbb{N})$  is non-decreasing. Let  $z \in [-\infty, \infty]$ . If  $z < \liminf_{n \rightarrow \infty} x_n$ , then  $z < x_n$  for all sufficiently large  $n \in \mathbb{N}$ . If  $z > \liminf_{n \rightarrow \infty} x_n$ , then  $z > x_n$  for infinitely many  $n \in \mathbb{N}$ .

The limit superior of a sequence of real numbers  $(x_n \in \mathbb{R} \mid n \in \mathbb{N})$  is defined by

$$\limsup_{n \rightarrow \infty} x_n = \inf_m \sup_{n \geq m} x_n = \lim_{m \rightarrow \infty} \sup_{n \geq m} x_n.$$

Note that the sequence  $(\sup_{n \geq m} x_n \mid m \in \mathbb{N})$  is non-increasing. Let  $z \in [-\infty, \infty]$ . If  $z > \limsup_{n \rightarrow \infty} x_n$ , then  $z > x_n$  for all sufficiently large  $n \in \mathbb{N}$ . If  $z < \limsup_{n \rightarrow \infty} x_n$ , then  $z < x_n$  for infinitely many  $n \in \mathbb{N}$ .

For any sequence  $(x_n \in \mathbb{R} \mid n \in \mathbb{N})$ , the limit inferior and the limit superior are related by the fact that

$$-\liminf_{n \rightarrow \infty} x_n = \lim_{m \rightarrow \infty} -\inf_{n \geq m} x_n = \lim_{m \rightarrow \infty} \sup_{n \geq m} -x_n = \limsup_{n \rightarrow \infty} -x_n.$$

A sequence of real numbers  $(x_n \in \mathbb{R} \mid n \in \mathbb{N})$  is said to converge in  $[-\infty, \infty]$  if and only if

$$\liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} x_n.$$

The limit inferior of a sequence of sets  $(E_n \mid n \in \mathbb{N})$  is defined by

$$\liminf_{n \rightarrow \infty} E_n = \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} E_n.$$

Let  $F_m = \bigcap_{n \geq m} E_n$ . Note that  $F_m \subseteq F_{m+1}$ . Furthermore,  $\omega \in \liminf_{n \rightarrow \infty} E_n$  if and only if  $\omega \in E_n$  for all sufficiently large  $n \in \mathbb{N}$ .

The limit superior of a sequence of sets  $(E_n \mid n \in \mathbb{N})$  is defined by

$$\limsup_{n \rightarrow \infty} E_n = \bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} E_n.$$

Let  $F_m = \bigcup_{n \geq m} E_n$ . Note that  $F_m \supseteq F_{m+1}$ . Furthermore,  $\omega \in \limsup_{n \rightarrow \infty} E_n$  if and only if  $\omega \in E_n$  for infinitely many  $n \in \mathbb{N}$ .

For any sequence of sets  $(E_n \subseteq \Omega \mid n \in \mathbb{N})$ , the limit inferior and the limit superior are related by the fact that

$$\left( \liminf_{n \rightarrow \infty} E_n \right)^C = \limsup_{n \rightarrow \infty} E_n^C.$$

Consider a measurable space  $(\Omega, \mathcal{F})$ . The indicator function  $\mathbb{I}_F : \Omega \rightarrow \{0, 1\}$  of an event  $F \in \mathcal{F}$  is defined by

$$\mathbb{I}_F(\omega) = \begin{cases} 1, & \text{if } \omega \in F, \\ 0, & \text{if } \omega \notin F. \end{cases}$$

For any outcome  $\omega \in \Omega$  and sequence of events  $(E_n \in \mathcal{F} \mid n \in \mathbb{N})$ ,

$$\begin{aligned} \mathbb{I}_{\liminf_{n \rightarrow \infty} E_n}(\omega) &= \liminf_{n \rightarrow \infty} \mathbb{I}_{E_n}(\omega), \\ \mathbb{I}_{\limsup_{n \rightarrow \infty} E_n}(\omega) &= \limsup_{n \rightarrow \infty} \mathbb{I}_{E_n}(\omega). \end{aligned}$$

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a sequence  $(E_n \in \mathcal{F} \mid n \in \mathbb{N})$ . The reverse Fatou Lemma states that

$$\mathbb{P} \left( \limsup_{n \rightarrow \infty} E_n \right) \geq \limsup_{n \rightarrow \infty} \mathbb{P}(E_n).$$

We will now show this lemma. Let  $F_m = \bigcup_{n \geq m} E_n$  such that  $F_m \supseteq F_{m+1}$ . By definition,  $F_m \downarrow \limsup_{n \rightarrow \infty} E_n$ , which implies  $\mathbb{P}(F_m) \downarrow \mathbb{P}(\limsup_{n \rightarrow \infty} E_n)$ . Because  $A \subseteq (B \cup A)$  implies  $\mathbb{P}(A) \leq \mathbb{P}(B \cup A)$  for any events  $A, B \in \mathcal{F}$ ,

$$\mathbb{P}(F_m) = \mathbb{P}\left(\bigcup_{n \geq m} E_n\right) \geq \sup_{n \geq m} \mathbb{P}(E_n).$$

By taking the limit of both sides of the equation above when  $m \rightarrow \infty$ ,

$$\lim_{m \rightarrow \infty} \mathbb{P}(F_m) = \mathbb{P}\left(\limsup_{n \rightarrow \infty} E_n\right) \geq \lim_{m \rightarrow \infty} \sup_{n \geq m} \mathbb{P}(E_n) = \limsup_{n \rightarrow \infty} \mathbb{P}(E_n).$$

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a sequence  $(E_n \in \mathcal{F} \mid n \in \mathbb{N})$ . The Fatou Lemma for sets states that

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} E_n\right) \leq \liminf_{n \rightarrow \infty} \mathbb{P}(E_n).$$

We will now show this lemma. Let  $F_m = \bigcap_{n \geq m} E_n$  such that  $F_m \subseteq F_{m+1}$ . By definition,  $F_m \uparrow \liminf_{n \rightarrow \infty} E_n$ , which implies  $\mathbb{P}(F_m) \uparrow \mathbb{P}(\liminf_{n \rightarrow \infty} E_n)$ . Because  $(A \cap B) \subseteq B$  implies  $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$  for any events  $A, B \in \mathcal{F}$ ,

$$\mathbb{P}(F_m) = \mathbb{P}\left(\bigcap_{n \geq m} E_n\right) \leq \inf_{n \geq m} \mathbb{P}(E_n).$$

By taking the limit of both sides of the equation above when  $m \rightarrow \infty$ ,

$$\lim_{m \rightarrow \infty} \mathbb{P}(F_m) = \mathbb{P}\left(\liminf_{n \rightarrow \infty} E_n\right) \leq \lim_{m \rightarrow \infty} \inf_{n \geq m} \mathbb{P}(E_n) = \liminf_{n \rightarrow \infty} \mathbb{P}(E_n).$$

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a sequence of events  $(E_n \in \mathcal{F} \mid n \in \mathbb{N})$  such that  $\sum_{n=0}^{\infty} \mathbb{P}(E_n) < \infty$ . The first Borel-Cantelli Lemma states that

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} E_n\right) = 0.$$

We will now show this lemma. Let  $F_m = \bigcup_{n \geq m} E_n$  such that  $F_m \supseteq F_{m+1}$ . By definition,  $F_m \downarrow \limsup_{n \rightarrow \infty} E_n$ , which implies  $\mathbb{P}(F_m) \downarrow \mathbb{P}(\limsup_{n \rightarrow \infty} E_n)$ . Because  $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$  for any events  $A, B \in \mathcal{F}$ ,

$$\mathbb{P}(F_m) = \mathbb{P}\left(\bigcup_{n \geq m} E_n\right) \leq \sum_{n \geq m} \mathbb{P}(E_n).$$

By taking the limit of both sides of the equation above when  $m \rightarrow \infty$ ,

$$\lim_{m \rightarrow \infty} \mathbb{P}(F_m) = \mathbb{P}\left(\limsup_{n \rightarrow \infty} E_n\right) \leq \lim_{m \rightarrow \infty} \sum_{n \geq m} \mathbb{P}(E_n) = 0,$$

where the last equality comes from the fact that, for any  $\epsilon > 0$ , there is an  $N \in \mathbb{N}$  such that, for all  $m - 1 \geq N$ ,

$$\epsilon > \left| \sum_{n=0}^{\infty} \mathbb{P}(E_n) - \sum_{n=0}^{m-1} \mathbb{P}(E_n) \right| = \sum_{n \geq m} \mathbb{P}(E_n).$$

### 3 Random variables

Consider a measurable space  $(S, \Sigma)$  and a function  $h : S \rightarrow \mathbb{R}$ . The function  $h^{-1}$  is defined as

$$h^{-1}(A) = \{s \in S \mid h(s) \in A\}$$

for any  $A \subseteq \mathbb{R}$ . The function  $h$  is called  $\Sigma$ -measurable if  $h^{-1}(A) \in \Sigma$  for every  $A \in \mathcal{B}(\mathbb{R})$ . In an extended definition, a function  $h : S \rightarrow [-\infty, \infty]$  is called  $\Sigma$ -measurable if  $h^{-1}(A) \in \Sigma$  for every  $A \in \mathcal{B}([-\infty, \infty])$ . A  $\mathcal{B}(\mathbb{R})$ -measurable function  $h : \mathbb{R} \rightarrow \mathbb{R}$  is said to be Borel.

The set of  $\Sigma$ -measurable functions on  $S$  is denoted by  $m\Sigma$ . The set of non-negative  $\Sigma$ -measurable functions on  $S$  is denoted by  $(m\Sigma)^+$ . The set of bounded  $\Sigma$ -measurable functions on  $S$  is denoted by  $b\Sigma$ .

Consider a function  $h : S \rightarrow \mathbb{R}$ . For any set  $A \subseteq \mathbb{R}$ ,

$$h^{-1}(A^c) = \{s \in S \mid h(s) \in A^c\} = \{s \in S \mid h(s) \in A\}^c = (h^{-1}(A))^c.$$

Consider a function  $h : S \rightarrow \mathbb{R}$ . For any sequence of sets  $(A_n \subseteq \mathbb{R} \mid n \in \mathbb{N})$ ,

$$h^{-1}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \{s \in S \mid h(s) \in \bigcup_{n \in \mathbb{N}} A_n\} = \bigcup_{n \in \mathbb{N}} \{s \in S \mid h(s) \in A_n\} = \bigcup_{n \in \mathbb{N}} h^{-1}(A_n).$$

Similarly,

$$h^{-1}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \{s \in S \mid h(s) \in \bigcap_{n \in \mathbb{N}} A_n\} = \bigcap_{n \in \mathbb{N}} \{s \in S \mid h(s) \in A_n\} = \bigcap_{n \in \mathbb{N}} h^{-1}(A_n).$$

Consider a measurable space  $(S, \Sigma)$  and a function  $h : S \rightarrow \mathbb{R}$ . The set  $\mathcal{E} = \{B \in \mathcal{B}(\mathbb{R}) \mid h^{-1}(B) \in \Sigma\}$  is a  $\sigma$ -algebra on  $\mathbb{R}$ . First, note that  $h^{-1}(\mathbb{R}) = \{s \in S \mid h(s) \in \mathbb{R}\} = S$  and  $S \in \Sigma$ . Therefore,  $\mathbb{R} \in \mathcal{E}$ . Consider an element  $B \in \mathcal{E}$ . In that case,  $h^{-1}(B) \in \Sigma$ , which implies  $(h^{-1}(B))^c = h^{-1}(B^c) \in \Sigma$ . Therefore,  $B^c \in \mathcal{E}$ . Finally, consider a sequence  $(B_n \in \mathcal{E} \mid n \in \mathbb{N})$ . In that case,  $h^{-1}(B_n) \in \Sigma$  for every  $n \in \mathbb{N}$ , which implies  $\bigcup_n h^{-1}(B_n) \in \Sigma$ . Therefore,  $h^{-1}(\bigcup_n B_n) \in \Sigma$  and  $\bigcup_n B_n \in \mathcal{E}$ .

Consider a measurable space  $(S, \Sigma)$ , a function  $h : S \rightarrow \mathbb{R}$ , and a set  $\mathcal{C}$  of subsets of  $\mathbb{R}$ . If  $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$  and  $h^{-1}(C) \in \Sigma$  for every  $C \in \mathcal{C}$ , then  $h$  is  $\Sigma$ -measurable. First, note that the set  $\mathcal{E} = \{B \in \mathcal{B}(\mathbb{R}) \mid h^{-1}(B) \in \Sigma\}$  is a  $\sigma$ -algebra on  $\mathbb{R}$ . Because  $\mathcal{C} \subseteq \mathcal{E}$ ,  $\mathcal{E} \subseteq \mathcal{B}(\mathbb{R})$ , and  $\mathcal{B}(\mathbb{R})$  is the smallest  $\sigma$ -algebra that includes  $\mathcal{C}$ , we know that  $\mathcal{E} = \mathcal{B}(\mathbb{R})$ , which implies that  $h^{-1}(B) \in \Sigma$  for every  $B \in \mathcal{B}(\mathbb{R})$ .

If a function  $h : \mathbb{R} \rightarrow \mathbb{R}$  is continuous, then it is Borel. First, consider the measurable space  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and let  $\mathcal{C}$  be the set of open sets of real numbers. Recall that  $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$ . Second, recall that a function  $h$  is continuous if  $h^{-1}(A) \in \mathcal{C}$  is an open set for every open set  $A \in \mathcal{C}$ . Using the previous result,  $h^{-1}(B) \in \mathcal{B}(\mathbb{R})$  for every  $B \in \mathcal{B}(\mathbb{R})$ .

Consider a measurable space  $(S, \Sigma)$  and a function  $h : S \rightarrow \mathbb{R}$ . For any  $c \in \mathbb{R}$ , define

$$\{h \leq c\} = h^{-1}((-\infty, c]) = \{s \in S \mid h(s) \leq c\}.$$

If  $\{h \leq c\} \in \Sigma$  for every  $c \in \mathbb{R}$ , then  $h$  is  $\Sigma$ -measurable. First, let  $\mathcal{C} = \{(-\infty, x] \mid x \in \mathbb{R}\}$  be the set that contains every interval that contains every real number smaller or equal to every real number  $x \in \mathbb{R}$ . Recall that  $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$ . By assumption,  $h^{-1}(C) \in \Sigma$  for every  $C \in \mathcal{C}$ , and so  $h^{-1}$  is  $\Sigma$ -measurable. Note that analogous results apply for  $\{h \geq c\}$ ,  $\{h < c\}$ , and  $\{h > c\}$ .

Consider a measurable space  $(S, \Sigma)$ . Let  $h : S \rightarrow \mathbb{R}$ ,  $h_1 : S \rightarrow \mathbb{R}$ , and  $h_2 : S \rightarrow \mathbb{R}$  be  $\Sigma$ -measurable functions and let  $\lambda \in \mathbb{R}$  be a constant. In that case,  $h_1 + h_2$  is a  $\Sigma$ -measurable function,  $h_1 h_2$  is a  $\Sigma$ -measurable function, and  $\lambda h$  is a  $\Sigma$ -measurable function. We will now show the first of these statements. Based on the previous result, if  $\{h_1 + h_2 > c\} = \{s \in S \mid h_1(s) + h_2(s) > c\} \in \Sigma$  for every  $c \in \mathbb{R}$ , then  $h_1 + h_2$  is  $\Sigma$ -measurable. Recall that  $h_1(s) + h_2(s) > c$  if and only if there is a rational  $q \in \mathcal{Q}$  such that  $h_1(s) > q > c - h_2(s)$ . Therefore,

$$\{h_1 + h_2 > c\} = \{s \in S \mid h_1(s) > q \text{ and } q > c - h_2(s) \text{ for some } q \in \mathcal{Q}\} = \bigcup_{q \in \mathcal{Q}} \{s \in S \mid h_1(s) > q \text{ and } q > c - h_2(s)\},$$

which is a countable union of elements of  $\Sigma$  given by

$$\{h_1 + h_2 > c\} = \bigcup_{q \in \mathcal{Q}} \{s \in S \mid h_1(s) > q\} \cap \{s \in S \mid q > c - h_2(s)\} = \bigcup_{q \in \mathcal{Q}} \{h_1 > q\} \cap \{h_2 > c - q\}.$$

Consider a measurable space  $(S, \Sigma)$  and a  $\Sigma$ -measurable function  $h : S \rightarrow \mathbb{R}$ . Consider also the measurable space  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and a  $\mathcal{B}(\mathbb{R})$ -measurable function  $f : \mathbb{R} \rightarrow \mathbb{R}$ . For all  $s \in S$ , let  $(f \circ h)(s) = f(h(s))$ . For any  $A \subseteq \mathbb{R}$ ,

$$(f \circ h)^{-1}(A) = \{s \in S \mid (f \circ h)(s) \in A\} = \{s \in S \mid f(h(s)) \in A\}.$$

Note that  $f^{-1}(A) \subseteq \mathbb{R}$  for any  $A \subseteq \mathbb{R}$ , since  $f^{-1}(A) = \{r \in \mathbb{R} \mid f(r) \in A\}$ . Therefore,

$$(h^{-1} \circ f^{-1})(A) = h^{-1}(f^{-1}(A)) = \{s \in S \mid h(s) \in f^{-1}(A)\} = \{s \in S \mid f(h(s)) \in A\} = (f \circ h)^{-1}(A),$$

where we used the fact that  $f(h(s)) \in A$  if and only if  $h(s) \in f^{-1}(A)$ , for all  $s \in S$  and  $A \subseteq \mathbb{R}$ . Furthermore, since  $f^{-1}(A) \in \mathcal{B}(\mathbb{R})$  for any  $A \in \mathcal{B}(\mathbb{R})$  and  $h^{-1}(f^{-1}(A)) \in \Sigma$  for any  $f^{-1}(A) \in \mathcal{B}(\mathbb{R})$ , the function  $f \circ h$  is  $\Sigma$ -measurable.

Consider the measurable spaces  $(S_1, \Sigma_1)$  and  $(S_2, \Sigma_2)$ . A function  $h : S_1 \rightarrow S_2$  is called  $\Sigma_1/\Sigma_2$ -measurable if  $h^{-1}(A) \in \Sigma_1$  for every  $A \in \Sigma_2$ . Therefore, a function on a measurable space  $(S, \Sigma)$  is  $\Sigma$ -measurable if it is  $\Sigma/\mathcal{B}(\mathbb{R})$ -measurable.

Consider a measurable space  $(S, \Sigma)$  and a sequence of  $\Sigma/\mathcal{B}([-\infty, \infty])$ -measurable functions  $(h_n \mid n \in \mathbb{N})$ .

For any  $s \in S$ , the function  $\inf_n h_n : S \rightarrow [-\infty, \infty]$  is given by

$$\left( \inf_n h_n \right) (s) = \inf_n h_n(s).$$

We will now show that  $\inf_n h_n$  is  $\Sigma/\mathcal{B}([-\infty, \infty])$ -measurable. Note that if  $\{\inf_n h_n \geq c\} \in \Sigma$  for every  $c \in \mathbb{R}$ , then  $\inf_n h_n$  is  $\Sigma/\mathcal{B}([-\infty, \infty])$ -measurable. For every  $c \in \mathbb{R}$ ,

$$\left\{ \inf_n h_n \geq c \right\} = \left\{ s \in S \mid \inf_n h_n(s) \geq c \right\} = \left\{ s \in S \mid h_n(s) \geq c \text{ for all } n \in \mathbb{N} \right\},$$

where we used the fact that  $\inf_n h_n(s) \geq c$  if and only if  $h_n(s) \geq c$  for all  $n \in \mathbb{N}$ , for all  $s \in S$  and  $c \in \mathbb{R}$ . Therefore,

$$\left\{ \inf_n h_n \geq c \right\} = \bigcap_{n \in \mathbb{N}} \left\{ s \in S \mid h_n(s) \geq c \right\} = \bigcap_{n \in \mathbb{N}} \{h_n \geq c\},$$

which is a countable intersection of elements of  $\Sigma$ .

For any  $s \in S$ , the function  $\sup_n h_n : S \rightarrow [-\infty, \infty]$  is given by

$$\left( \sup_n h_n \right) (s) = \sup_n h_n(s).$$

We will now show that  $\sup_n h_n$  is  $\Sigma/\mathcal{B}([-\infty, \infty])$ -measurable. Note that if  $\{\sup_n h_n \leq c\} \in \Sigma$  for every  $c \in \mathbb{R}$ , then  $\sup_n h_n$  is  $\Sigma/\mathcal{B}([-\infty, \infty])$ -measurable. For every  $c \in \mathbb{R}$ ,

$$\left\{ \sup_n h_n \leq c \right\} = \left\{ s \in S \mid \sup_n h_n(s) \leq c \right\} = \left\{ s \in S \mid h_n(s) \leq c \text{ for all } n \in \mathbb{N} \right\},$$

where we used the fact that  $\sup_n h_n(s) \leq c$  if and only if  $h_n(s) \leq c$  for all  $n \in \mathbb{N}$ , for all  $s \in S$  and  $c \in \mathbb{R}$ . Therefore,

$$\left\{ \sup_n h_n \leq c \right\} = \bigcap_{n \in \mathbb{N}} \left\{ s \in S \mid h_n(s) \leq c \right\} = \bigcap_{n \in \mathbb{N}} \{h_n \leq c\},$$

which is a countable intersection of elements of  $\Sigma$ .

For any  $s \in S$ , the function  $\liminf_{n \rightarrow \infty} h_n : S \rightarrow [-\infty, \infty]$  is given by

$$\left( \liminf_{n \rightarrow \infty} h_n \right) (s) = \liminf_{n \rightarrow \infty} h_n(s).$$

We will now show that  $\liminf_{n \rightarrow \infty} h_n$  is  $\Sigma/\mathcal{B}([-\infty, \infty])$ -measurable. Each function in the sequence  $(L_n = \inf_{r \geq n} h_r \mid n \in \mathbb{N})$  is  $\Sigma/\mathcal{B}([-\infty, \infty])$ -measurable, which implies that  $\sup_n L_n$  is  $\Sigma/\mathcal{B}([-\infty, \infty])$ -measurable. Also,

$$\left( \liminf_{n \rightarrow \infty} h_n \right) (s) = \liminf_{n \rightarrow \infty} h_n(s) = \sup_n \inf_{r \geq n} h_r(s) = \sup_n \left( \inf_{r \geq n} h_r \right) (s) = \sup_n L_n(s) = \left( \sup_n L_n \right) (s).$$

For any  $s \in S$ , the function  $\limsup_{n \rightarrow \infty} h_n : S \rightarrow [-\infty, \infty]$  is given by

$$\left( \limsup_{n \rightarrow \infty} h_n \right) (s) = \limsup_{n \rightarrow \infty} h_n(s).$$

We will now show that  $\limsup_{n \rightarrow \infty} h_n$  is  $\Sigma/\mathcal{B}([-\infty, \infty])$ -measurable. Each function in the sequence  $(L_n = \sup_{r \geq n} h_r \mid n \in \mathbb{N})$  is  $\Sigma/\mathcal{B}([-\infty, \infty])$ -measurable, which implies that  $\inf_n L_n$  is  $\Sigma/\mathcal{B}([-\infty, \infty])$ -measurable. Also,

$$\left( \limsup_{n \rightarrow \infty} h_n \right) (s) = \limsup_{n \rightarrow \infty} h_n(s) = \inf_n \sup_{r \geq n} h_r(s) = \inf_n \left( \sup_{r \geq n} h_r \right) (s) = \inf_n L_n(s) = \left( \inf_n L_n \right) (s).$$

Consider the set  $F = \{s \in S \mid \lim_{n \rightarrow \infty} h_n(s) \text{ exists in } \mathbb{R}\}$ . Recall that  $\lim_{n \rightarrow \infty} h_n(s)$  exists in  $\mathbb{R}$  if and only if

$$-\infty < \liminf_{n \rightarrow \infty} h_n(s) = \limsup_{n \rightarrow \infty} h_n(s) < \infty.$$

Therefore,  $F \in \Sigma$ , since  $F$  is an intersection of elements of  $\Sigma$ :

$$F = \{s \in S \mid \liminf_{n \rightarrow \infty} h_n(s) > -\infty\} \cap \{s \in S \mid \limsup_{n \rightarrow \infty} h_n(s) < \infty\} \cap \{s \in S \mid \left(\limsup_{n \rightarrow \infty} h_n - \liminf_{n \rightarrow \infty} h_n\right)(s) = 0\}.$$

Consider a measurable space  $(\Omega, \mathcal{F})$ . An  $\mathcal{F}$ -measurable function  $X : \Omega \rightarrow \mathbb{R}$  is a random variable. By definition, for any  $B \in \mathcal{B}(\mathbb{R})$ ,  $X^{-1}(B) \in \mathcal{F}$ .

The indicator function  $\mathbb{I}_F : \Omega \rightarrow \{0, 1\}$  of any event  $F \in \mathcal{F}$  is a random variable. The function  $\mathbb{I}_F$  is defined by

$$\mathbb{I}_F(\omega) = \begin{cases} 1, & \text{if } \omega \in F, \\ 0, & \text{if } \omega \notin F. \end{cases}$$

Recall that if  $\{\omega \in \Omega \mid \mathbb{I}_F(\omega) \leq c\} \in \mathcal{F}$  for every  $c \in \mathbb{R}$ , then  $\mathbb{I}_F$  is  $\mathcal{F}$ -measurable. For every  $c < 1$ , we have  $\{\omega \in \Omega \mid \mathbb{I}_F(\omega) \leq c\} = \{\omega \in \Omega \mid \omega \notin F\} = F^c$ . For every  $c \geq 1$ , we have  $\{\omega \in \Omega \mid \mathbb{I}_F(\omega) \leq c\} = \Omega$ .

More interestingly, once again consider an experiment where a coin is tossed infinitely often. Let  $H = 0$  represent heads and  $T = 1$  represent tails. The outcome space  $\Omega$  may be defined as the set of infinite binary sequences  $\Omega = \{H, T\}^{\mathbb{N}^+}$ . Let  $F_{n,W} = \{\omega \in \Omega \mid \omega_n = W\}$  be the set of infinite binary sequences whose  $n$ -th element is  $W$ . The  $\sigma$ -algebra  $\mathcal{F}$  on the outcome space  $\Omega$  may be generated as  $\mathcal{F} = \sigma(\{F_{n,W} \mid n \in \mathbb{N}^+, W \in \{H, T\}\})$ . Note that  $\mathbb{I}_{F_{n,W}}$  is a random variable, since  $F_{n,W} \in \mathcal{F}$ . Therefore, for any  $n \in \mathbb{N}^+$ , the function  $A_{n,W}$  given by

$$A_{n,W}(\omega) = \left(n^{-1} \sum_{i=1}^n \mathbb{I}_{F_{i,W}}\right)(\omega) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{F_{i,W}}(\omega)$$

is also a random variable. For a given sequence  $\omega \in \Omega$ ,  $A_{n,W}(\omega)$  is the fraction of the first  $n$  tosses resulting in  $W$ .

For a given  $p \in [0, 1]$ , consider the set  $\Lambda_W = \{\omega \in \Omega \mid \lim_{n \rightarrow \infty} A_{n,W}(\omega) = p\}$ . Clearly,

$$\Lambda_W = \{\omega \in \Omega \mid \liminf_{n \rightarrow \infty} A_{n,W}(\omega) = p\} \cap \{\omega \in \Omega \mid \limsup_{n \rightarrow \infty} A_{n,W}(\omega) = p\},$$

which can be rewritten as

$$\Lambda_W = \left(\liminf_{n \rightarrow \infty} A_{n,W}\right)^{-1}(\{p\}) \cap \left(\limsup_{n \rightarrow \infty} A_{n,W}\right)^{-1}(\{p\}).$$

Note that  $\Lambda_W \in \mathcal{F}$ , since both the limit inferior and the limit superior of the sequence of  $\mathcal{F}$ -measurable functions  $(A_{n,W} \mid n \in \mathbb{N}^+)$  are  $\mathcal{F}$ -measurable functions. Therefore, a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  would define the probability  $\mathbb{P}(\Lambda_W)$  that the fraction of tosses with result  $W$  tends to a given  $p \in [0, 1]$ .

Consider a function  $X : \Omega \rightarrow \mathbb{R}$ . The  $\sigma$ -algebra  $\sigma(X)$  on  $\Omega$  is defined as  $\sigma(X) = \sigma(\{X^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R})\})$ . Note that if  $X$  is a random variable on a measurable space  $(\Omega, \mathcal{F})$ , then  $\sigma(X) \subseteq \mathcal{F}$ .

Consider a set of functions  $\{Y_\gamma \mid \gamma \in \mathcal{C}\}$  where  $Y_\gamma : \Omega \rightarrow \mathbb{R}$ . The  $\sigma$ -algebra  $\sigma(\{Y_\gamma \mid \gamma \in \mathcal{C}\})$  is defined by

$$\sigma(\{Y_\gamma \mid \gamma \in \mathcal{C}\}) = \sigma(\{Y_\gamma^{-1}(B) \mid \gamma \in \mathcal{C}, B \in \mathcal{B}(\mathbb{R})\}).$$

Note that if  $Y_\gamma : \Omega \rightarrow \mathbb{R}$  is a random variable on a measurable space  $(\Omega, \mathcal{F})$  for every  $\gamma$ , then  $\sigma(\{Y_\gamma \mid \gamma \in \mathcal{C}\}) \subseteq \mathcal{F}$ .

Consider a measurable space  $(\Omega, \mathcal{F})$  and a random variable  $Y : \Omega \rightarrow \mathbb{R}$ . For a set  $\mathcal{E}$  of subsets of  $\mathbb{R}$ , let  $Y^{-1}(\mathcal{E}) = \{Y^{-1}(E) \mid E \in \mathcal{E}\}$ . By definition,  $\sigma(Y) = \sigma(Y^{-1}(\mathcal{B}(\mathbb{R})))$ . We will now show that  $\sigma(Y) = Y^{-1}(\mathcal{B}(\mathbb{R}))$ .

By definition,  $Y^{-1}(\mathcal{B}(\mathbb{R})) = \{Y^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R})\}$ . Because  $\mathbb{R} \in \mathcal{B}(\mathbb{R})$ ,  $Y^{-1}(\mathbb{R}) \in Y^{-1}(\mathcal{B}(\mathbb{R}))$ , where  $Y^{-1}(\mathbb{R}) = \Omega$ . Consider an element  $Y^{-1}(B) \in Y^{-1}(\mathcal{B}(\mathbb{R}))$ . Because  $B^c \in \mathcal{B}(\mathbb{R})$ ,  $Y^{-1}(B^c) \in Y^{-1}(\mathcal{B}(\mathbb{R}))$ , where  $Y^{-1}(B^c) = (Y^{-1}(B))^c$ . Finally, consider a sequence  $(Y^{-1}(B_n) \in Y^{-1}(\mathcal{B}(\mathbb{R})) \mid n \in \mathbb{N})$ . Because  $\cup_n B_n \in \mathcal{B}(\mathbb{R})$ ,  $Y^{-1}(\cup_n B_n) \in Y^{-1}(\mathcal{B}(\mathbb{R}))$ , where  $Y^{-1}(\cup_n B_n) = \cup_n Y^{-1}(B_n)$ . Therefore,  $Y^{-1}(\mathcal{B}(\mathbb{R}))$  is a  $\sigma$ -algebra on  $\Omega$ . Because  $\sigma(Y)$  is the smallest  $\sigma$ -algebra on  $\Omega$  that includes  $Y^{-1}(\mathcal{B}(\mathbb{R}))$ , we know that  $\sigma(Y) = Y^{-1}(\mathcal{B}(\mathbb{R}))$ .

Furthermore, consider the  $\pi$ -system  $\pi(\mathbb{R}) = \{(-\infty, x] \mid x \in \mathbb{R}\}$  and let  $\pi(Y) = Y^{-1}(\pi(\mathbb{R}))$ . We will now show that  $\sigma(Y) = \sigma(\pi(Y))$ .

By definition,  $\sigma(\pi(Y)) = \sigma(\{Y^{-1}((-\infty, x]) \mid (-\infty, x] \in \pi(\mathbb{R})\})$ . Clearly,  $\pi(\mathbb{R}) \subseteq \mathcal{B}(\mathbb{R})$  implies  $\sigma(\pi(Y)) \subseteq \sigma(Y)$ , since  $\sigma(Y) = \sigma(\{Y^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R})\})$ . Because  $\{Y \leq x\} \in \sigma(\pi(Y))$  for every  $x \in \mathbb{R}$ ,  $Y$  is  $\sigma(\pi(Y))$ -measurable. Therefore,  $\sigma(Y) \subseteq \sigma(\pi(Y))$ .

If  $Y : \Omega \rightarrow \mathbb{R}$ , then  $Z : \Omega \rightarrow \mathbb{R}$  is a  $\sigma(Y)$ -measurable function if and only if there is a Borel function  $f : \mathbb{R} \rightarrow \mathbb{R}$  such that  $Z = f \circ Y$ .

If  $Y_1, Y_2, \dots, Y_n$  are functions from  $\Omega$  to  $\mathbb{R}$ , then a function  $Z : \Omega \rightarrow \mathbb{R}$  is  $\sigma(\{Y_1, Y_2, \dots, Y_n\})$ -measurable if and only if there is a Borel function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $Z(\omega) = f(Y_1(\omega), Y_2(\omega), \dots, Y_n(\omega))$  for every  $\omega \in \Omega$ .

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a random variable  $X : \Omega \rightarrow \mathbb{R}$ . For any  $B \in \mathcal{B}(\mathbb{R})$ ,  $X^{-1}(B) \in \sigma(X)$ ,  $\sigma(X) \subseteq \mathcal{F}$ , and  $\mathbb{P}(X^{-1}(B)) \in [0, 1]$ . For any  $B \in \mathcal{B}(\mathbb{R})$ , this allows defining the law  $\mathcal{L}_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  of  $X$  as

$$\mathcal{L}_X(B) = \mathbb{P}(X^{-1}(B)).$$

The law  $\mathcal{L}_X$  is a probability measure on the measurable space  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . First, note that

$$\begin{aligned}\mathcal{L}_X(\mathbb{R}) &= \mathbb{P}(X^{-1}(\mathbb{R})) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in \mathbb{R}\}) = \mathbb{P}(\Omega) = 1, \\ \mathcal{L}_X(\emptyset) &= \mathbb{P}(X^{-1}(\emptyset)) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in \emptyset\}) = \mathbb{P}(\emptyset) = 0.\end{aligned}$$

Second, consider a sequence of sets  $(B_n \in \mathcal{B}(\mathbb{R}) \mid n \in \mathbb{N})$  such that  $B_n \cap B_m = \emptyset$  for  $n \neq m$  and note that

$$\mathcal{L}_X\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{n \in \mathbb{N}} B_n\right)\right) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} X^{-1}(B_n)\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(X^{-1}(B_n)) = \sum_{n \in \mathbb{N}} \mathcal{L}_X(B_n),$$

where we used the fact that  $X^{-1}(B_n) \cap X^{-1}(B_m) = X^{-1}(B_n \cap B_m) = X^{-1}(\emptyset) = \emptyset$  for  $n \neq m$ .

The (cumulative) distribution function  $F_X : \mathbb{R} \rightarrow [0, 1]$  of the random variable  $X$  is defined by

$$F_X(c) = \mathcal{L}_X((-\infty, c]) = \mathbb{P}(X^{-1}((-\infty, c])) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \leq c\}) = \mathbb{P}(\{X \leq c\}).$$

Recall that the  $\sigma$ -algebra generated by  $\pi(\mathbb{R}) = \{(-\infty, x] \mid x \in \mathbb{R}\}$  is  $\sigma(\pi(\mathbb{R})) = \mathcal{B}(\mathbb{R})$ . Consider a probability measure  $\mu$  on the measurable space  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  such that  $\mu((-\infty, c]) = F_X(c) = \mathcal{L}_X((-\infty, c])$  for every  $c \in \mathbb{R}$ . Because  $\mu$  and  $\mathcal{L}_X$  agree on the  $\pi$ -system  $\pi(\mathbb{R})$ , we have  $\mu = \mathcal{L}_X$ . Therefore,  $F_X$  fully determines the law  $\mathcal{L}_X$  of  $X$ .

Consider a random variable  $X : \Omega \rightarrow \mathbb{R}$  carried by a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and the distribution function  $F_X : \mathbb{R} \rightarrow [0, 1]$ .

If  $a \leq b$ , then  $F_X(a) \leq F_X(b)$ . Clearly,  $\{X \leq a\} \subseteq \{X \leq b\}$ , which implies  $\mathbb{P}(\{X \leq a\}) \leq \mathbb{P}(\{X \leq b\})$ .

We will now show that  $\lim_{x \rightarrow -\infty} F_X(x) = 0$ . Recall that  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a function such that  $\lim_{x \rightarrow -\infty} f(x) = L$  for some  $L \in \mathbb{R}$  if and only if  $\lim_{n \rightarrow \infty} f(x_n) = L$  for all non-increasing sequences  $(x_n \in \mathbb{R} \mid n \in \mathbb{N})$  such that  $\lim_{n \rightarrow \infty} x_n = -\infty$ .

Consider a non-increasing sequence  $(x_n \in \mathbb{R} \mid n \in \mathbb{N})$  such that  $\lim_{n \rightarrow \infty} x_n = -\infty$  and the sequence of sets  $(A_n = (-\infty, x_n] \mid n \in \mathbb{N})$ . Because  $A_n \downarrow \emptyset$ ,  $\mathcal{L}_X(A_n) \downarrow 0$ . Therefore,  $\lim_{n \rightarrow \infty} \mathcal{L}_X((-\infty, x_n]) = 0$ , which implies

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{x \rightarrow -\infty} \mathcal{L}_X((-\infty, x]) = 0.$$

We will now show that  $\lim_{x \rightarrow \infty} F_X(x) = 1$ . Recall that  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a function such that  $\lim_{x \rightarrow \infty} f(x) = L$  for some  $L \in \mathbb{R}$  if and only if  $\lim_{n \rightarrow \infty} f(x_n) = L$  for all non-decreasing sequences  $(x_n \in \mathbb{R} \mid n \in \mathbb{N})$  such that  $\lim_{n \rightarrow \infty} x_n = +\infty$ .

Consider a non-decreasing sequence  $(x_n \in \mathbb{R} \mid n \in \mathbb{N})$  such that  $\lim_{n \rightarrow \infty} x_n = +\infty$  and the sequence of sets  $(A_n = (-\infty, x_n] \mid n \in \mathbb{N})$ . Because  $A_n \uparrow \mathbb{R}$ ,  $\mathcal{L}_X(A_n) \uparrow 1$ . Therefore,  $\lim_{n \rightarrow \infty} \mathcal{L}_X((-\infty, x_n]) = 1$ , which implies

$$\lim_{x \rightarrow \infty} F_X(x) = \lim_{x \rightarrow \infty} \mathcal{L}_X((-\infty, x]) = 1.$$

We will now show that  $F_X$  is right-continuous. Recall that  $f : \mathbb{R} \rightarrow \mathbb{R}$  is right continuous if and only if  $\lim_{n \rightarrow \infty} f(x_n) = f(x)$  for every  $x \in \mathbb{R}$  and every non-increasing sequence  $(x_n \in \mathbb{R} \mid n \in \mathbb{N})$  such that  $\lim_{n \rightarrow \infty} x_n = x$  and  $x_n > x$  for every  $n \in \mathbb{N}$ .

Consider  $x \in \mathbb{R}$  and a non-increasing sequence  $(x_n \in \mathbb{R} \mid n \in \mathbb{N})$  such that  $\lim_{n \rightarrow \infty} x_n = x$  and  $x_n > x$  for every  $n \in \mathbb{N}$ . Consider also the sequence of sets  $(A_n = (-\infty, x_n] \mid n \in \mathbb{N})$ . Because  $A_n \downarrow (-\infty, x]$ ,  $\mathcal{L}_X((-\infty, x_n]) \downarrow \mathcal{L}_X((-\infty, x])$ . Therefore,  $\lim_{n \rightarrow \infty} \mathcal{L}_X((-\infty, x_n]) = \mathcal{L}_X((-\infty, x])$ , which implies

$$\lim_{n \rightarrow \infty} F_X(x_n) = \lim_{n \rightarrow \infty} \mathcal{L}_X((-\infty, x_n]) = \mathcal{L}_X((-\infty, x]) = F_X(x).$$

Consider a right-continuous function  $F : \mathbb{R} \rightarrow [0, 1]$  such that if  $a \leq b$ , then  $F(a) \leq F(b)$ ;  $\lim_{x \rightarrow -\infty} F(x) = 0$ ; and  $\lim_{x \rightarrow \infty} F(x) = 1$ . We will show that there is a unique probability measure  $\mathcal{L}$  on the measurable space  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  such that  $\mathcal{L}((-\infty, x]) = F(x)$  for every  $x \in \mathbb{R}$ .



Consider the probability triple  $((0, 1), \mathcal{B}((0, 1)), \text{Leb})$  and a function  $X^- : (0, 1) \rightarrow \mathbb{R}$  given by

$$X^-(\omega) = \inf\{z \in \mathbb{R} \mid F(z) \geq \omega\}.$$

In words,  $X^-(\omega)$  is the infimum  $z \in \mathbb{R}$  such that  $F(z)$  reaches  $\omega \in (0, 1)$ .

First, note that  $\omega \leq F(c)$  if and only if  $X^-(\omega) \leq c$  for every  $c \in \mathbb{R}$ . Clearly, if  $\omega \leq F(c)$ , then  $X^-(\omega) \leq c$ . Now suppose  $X^-(\omega) \leq c$ . Because  $F$  is non-decreasing,  $F(X^-(\omega)) \leq F(c)$ . Because  $F$  is also right-continuous,  $F(X^-(\omega)) \geq \omega$ . Therefore,  $\omega \leq F(c)$ . This also implies that  $X^-$  is a random variable since, for every  $c \in \mathbb{R}$ ,

$$\{X^- \leq c\} = \{\omega \in (0, 1) \mid X^-(\omega) \leq c\} = \{\omega \in (0, 1) \mid \omega \leq F(c)\} = (0, F(c)].$$

For every  $c \in \mathbb{R}$ , the distribution function  $F_{X^-}$  on the probability triple  $((0, 1), \mathcal{B}((0, 1)), \text{Leb})$  is given by

$$F_{X^-}(c) = \mathcal{L}_{X^-}((-\infty, c]) = \text{Leb}(\{X^- \leq c\}) = \text{Leb}((0, F(c)]) = F(c).$$

Finally, recall that the distribution function  $F_{X^-}$  fully determines the law  $\mathcal{L}_{X^-}$  of  $X^-$ , which is the desired unique probability measure on the measurable space  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  such that  $\mathcal{L}_{X^-}((-\infty, x]) = F(x)$  for every  $x \in \mathbb{R}$ .

The monotone-class theorem states that if

- $\mathcal{H}$  is a set of bounded functions from a set  $S$  into  $\mathbb{R}$ ,
- $\mathcal{H}$  is a vector space over  $\mathbb{R}$ ,
- The constant function 1 is an element of  $\mathcal{H}$ ,
- If  $(f_n \in \mathcal{H} \mid n \in \mathbb{N})$  is a sequence of non-negative functions in  $\mathcal{H}$  such that  $f_n \uparrow f$ , where  $f$  is a bounded function on  $S$ , then  $f \in \mathcal{H}$ ,
- $\mathcal{H}$  contains the indicator function of every set in some  $\pi$ -system  $\mathcal{I}$ ,

then  $\mathcal{H}$  contains every bounded  $\sigma(\mathcal{I})$ -measurable function on  $S$ .

## 4 Independence

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ .

The sub- $\sigma$ -algebras  $\mathcal{G}_1, \mathcal{G}_2, \dots$  of  $\mathcal{F}$  are called independent if, for every choice of distinct indices  $i_1, i_2, \dots, i_n$  and events  $G_{i_1}, G_{i_2}, \dots, G_{i_n}$  such that  $G_{i_k} \in \mathcal{G}_{i_k}$  for every  $i_k$ ,

$$\mathbb{P}\left(\bigcap_{k=1}^n G_{i_k}\right) = \prod_{k=1}^n \mathbb{P}(G_{i_k}).$$

The random variables  $X_1, X_2, \dots$  are called independent if the  $\sigma$ -algebras  $\sigma(X_1), \sigma(X_2), \dots$  are independent.

The events  $E_1, E_2, \dots$  are called independent if the  $\sigma$ -algebras  $\mathcal{E}_1, \mathcal{E}_2, \dots$  are independent, where  $\mathcal{E}_k = \{\emptyset, E_k, E_k^c, \Omega\}$ . We have already shown that each indicator function  $\mathbb{I}_{E_k}$  is  $\mathcal{E}_k$ -measurable. Since  $\mathbb{I}_{E_k}^{-1}(\{1\}) = E_k$ , we know that  $E_k \in \sigma(\mathbb{I}_{E_k})$ , which implies  $\mathcal{E}_k = \sigma(\mathbb{I}_{E_k})$ . Therefore, the events  $E_1, E_2, \dots$  are called independent if and only if the random variables  $\mathbb{I}_{E_1}, \mathbb{I}_{E_2}, \dots$  are independent.

The events  $E_1, E_2, \dots$  are independent if and only if, for every choice of distinct indices  $i_1, i_2, \dots, i_n$ ,

$$\mathbb{P}\left(\bigcap_{k=1}^n E_{i_k}\right) = \prod_{k=1}^n \mathbb{P}(E_{i_k}).$$

If  $X_1, X_2, \dots$  are independent random variables, then the events  $\{X_1 \leq x_1\}, \{X_2 \leq x_2\}, \dots$  are independent for every  $x_1, x_2, \dots \in \mathbb{R}$ , since  $X_n^{-1}((-\infty, x_n]) \in \sigma(X_n)$  for every  $n \in \mathbb{N}^+$ .

Suppose that  $\mathcal{G}$  and  $\mathcal{H}$  are sub- $\sigma$ -algebras of  $\mathcal{F}$ . Furthermore, let  $\mathcal{I}$  and  $\mathcal{J}$  be  $\pi$ -systems such that  $\sigma(\mathcal{I}) = \mathcal{G}$  and  $\sigma(\mathcal{J}) = \mathcal{H}$ . If  $\mathbb{P}(I \cap J) = \mathbb{P}(I)\mathbb{P}(J)$  for every  $I \in \mathcal{I}$  and  $J \in \mathcal{J}$ , we say that  $\mathcal{I}$  and  $\mathcal{J}$  are independent. We will show that  $\mathcal{G}$  and  $\mathcal{H}$  are independent if and only if  $\mathcal{I}$  and  $\mathcal{J}$  are independent.

Suppose that  $\mathcal{G}$  and  $\mathcal{H}$  are independent. In that case,  $\mathbb{P}(G \cap H) = \mathbb{P}(G)\mathbb{P}(H)$  for every  $G \in \mathcal{G}$  and  $H \in \mathcal{H}$ . Since  $\mathcal{I} \subseteq \mathcal{G}$  and  $\mathcal{J} \subseteq \mathcal{H}$ ,  $\mathcal{I}$  and  $\mathcal{J}$  are independent.

Suppose that  $\mathcal{I}$  and  $\mathcal{J}$  are independent. For every  $I \in \mathcal{I}$  and  $H \in \mathcal{H}$ , let  $\mu_I(H) = \mathbb{P}(I \cap H)$  and  $\eta_I(H) = \mathbb{P}(I)\mathbb{P}(H)$ . Clearly,  $\mu_I(\emptyset) = 0 = \eta_I(\emptyset)$ . Also,  $\mu_I(\Omega) = \mathbb{P}(I) = \eta_I(\Omega)$ . Finally, if  $(H_n \in \mathcal{H} \mid n \in \mathbb{N})$  is a sequence of events such that  $H_n \cap H_m = \emptyset$  for  $n \neq m$ ,

$$\begin{aligned}\mu_I\left(\bigcup_n H_n\right) &= \mathbb{P}\left(I \cap \left(\bigcup_n H_n\right)\right) = \mathbb{P}\left(\bigcup_n (I \cap H_n)\right) = \sum_n \mathbb{P}(I \cap H_n) = \sum_n \mu_I(H_n), \\ \eta_I\left(\bigcup_n H_n\right) &= \mathbb{P}(I)\mathbb{P}\left(\bigcup_n H_n\right) = \mathbb{P}(I) \sum_n \mathbb{P}(H_n) = \sum_n \mathbb{P}(I)\mathbb{P}(H_n) = \sum_n \eta_I(H_n).\end{aligned}$$

Considered together, these results imply that  $\mu_I$  and  $\eta_I$  are finite measures on  $(\Omega, \mathcal{H})$ . By assumption,  $\mu_I(J) = \mathbb{P}(I \cap J) = \mathbb{P}(I)\mathbb{P}(J) = \eta_I(J)$  for every  $I \in \mathcal{I}$  and  $J \in \mathcal{J}$ . Therefore,  $\mu_I$  and  $\eta_I$  agree on the  $\pi$ -system  $\mathcal{J}$ , which implies that they agree on the  $\sigma$ -algebra  $\sigma(\mathcal{J}) = \mathcal{H}$ . In other words, for every  $I \in \mathcal{I}$  and  $H \in \mathcal{H}$ , we have  $\mathbb{P}(I \cap H) = \mu_I(H) = \eta_I(H) = \mathbb{P}(I)\mathbb{P}(H)$ .

For every  $H \in \mathcal{H}$  and  $G \in \mathcal{G}$ , let  $\mu'_H(G) = \mathbb{P}(H \cap G)$  and  $\eta'_H(G) = \mathbb{P}(H)\mathbb{P}(G)$ . Analogously,  $\mu'_H$  and  $\eta'_H$  are finite measures on  $(\Omega, \mathcal{G})$ . From our previous result, for every  $I \in \mathcal{I}$  and  $H \in \mathcal{H}$ , we have  $\mathbb{P}(I \cap H) = \mu'_H(I) = \eta'_H(I) = \mathbb{P}(I)\mathbb{P}(H)$ . Therefore,  $\mu'_H$  and  $\eta'_H$  agree on the  $\pi$ -system  $\mathcal{I}$ , which implies that they agree on the  $\sigma$ -algebra  $\sigma(\mathcal{I}) = \mathcal{G}$ . In other words, for every  $G \in \mathcal{G}$  and  $H \in \mathcal{H}$ , we have  $\mathbb{P}(G \cap H) = \mu'_H(G) = \eta'_H(G) = \mathbb{P}(G)\mathbb{P}(H)$ .

Consider the random variables  $X$  and  $Y$  on the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ . For every  $A \in \mathcal{B}(\mathbb{R})$  and  $B \in \mathcal{B}(\mathbb{R})$  such that  $\mathbb{P}(Y^{-1}(B)) > 0$ , let  $\mathbb{P}(X^{-1}(A) \mid Y^{-1}(B)) = \mathbb{P}(X^{-1}(A) \cap Y^{-1}(B)) / \mathbb{P}(Y^{-1}(B))$ . If  $X$  and  $Y$  are independent, then  $\mathbb{P}(X^{-1}(A) \mid Y^{-1}(B)) = \mathbb{P}(X^{-1}(A))$ , since  $X^{-1}(A) \in \sigma(X)$  and  $Y^{-1}(B) \in \sigma(Y)$ .

In what follows, we will employ a common abuse of notation. Consider the random variables  $X$  and  $Y$  on the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ . For every  $x \in \mathbb{R}$ , we will let  $\mathbb{P}(X \leq x)$  denote  $\mathbb{P}(\{X \leq x\})$ . Furthermore, for every  $x, y \in \mathbb{R}$ , we will let  $\mathbb{P}(X \leq x, Y \leq y)$  denote  $\mathbb{P}(\{X \leq x\} \cap \{Y \leq y\})$ . We will employ analogous notation when there are more random variables and different predicates.

Consider the random variables  $X$  and  $Y$  on the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ . Suppose that, for every  $x, y \in \mathbb{R}$ ,  $\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y)$ . We will now show that  $X$  and  $Y$  are independent.

Recall that  $\pi(\mathbb{R}) = \{(-\infty, x] \mid x \in \mathbb{R}\}$  and  $\pi(X) = \{X^{-1}((-\infty, x]) \mid (-\infty, x] \in \pi(\mathbb{R})\} = \{\{X \leq x\} \mid x \in \mathbb{R}\}$ . Note that  $\pi(X)$  is a  $\pi$ -system on  $\Omega$ : for any  $x_1, x_2 \in \mathbb{R}$ , if  $\{X \leq x_1\} \in \pi(X)$  and  $\{X \leq x_2\} \in \pi(X)$ , then  $\{X \leq x_1\} \cap \{X \leq x_2\} = \{\omega \in \Omega \mid X(\omega) \leq x_1 \text{ and } X(\omega) \leq x_2\} = \{\omega \in \Omega \mid X(\omega) \leq \min(x_1, x_2)\} = \{X \leq \min(x_1, x_2)\}$ . By assumption,  $\mathbb{P}(\{X \leq x\} \cap \{Y \leq y\}) = \mathbb{P}(\{X \leq x\})\mathbb{P}(\{Y \leq y\})$  for any  $\{X \leq x\} \in \pi(X)$  and  $\{Y \leq y\} \in \pi(Y)$ . By definition, the  $\pi$ -systems  $\pi(X)$  and  $\pi(Y)$  are independent. Therefore,  $\sigma(\pi(X))$  and  $\sigma(\pi(Y))$  are independent. Based on a previous result, we know that  $\sigma(\pi(X)) = \sigma(X)$  and  $\sigma(\pi(Y)) = \sigma(Y)$ .

In general, the random variables  $X_1, X_2, \dots, X_n$  are independent if and only if, for every  $x_1, x_2, \dots, x_n \in \mathbb{R}$ ,

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = \mathbb{P}\left(\bigcap_{k=1}^n \{X_k \leq x_k\}\right) = \prod_{k=1}^n \mathbb{P}(X_k \leq x_k).$$

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a sequence of independent events  $(E_n \in \mathcal{F} \mid n \in \mathbb{N})$  such that  $\sum_{n=0}^{\infty} \mathbb{P}(E_n) = \infty$ . The second Borel-Cantelli Lemma states that

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} E_n\right) = 1.$$

Because the events are independent, for any  $m, r \in \mathbb{N}$  such that  $m \leq r$ ,

$$\mathbb{P}\left(\bigcap_{m \leq n \leq r} E_n^c\right) = \prod_{m \leq n \leq r} \mathbb{P}(E_n^c) = \prod_{m \leq n \leq r} (1 - \mathbb{P}(E_n)).$$

Let  $e$  denote Euler's number. For any  $x \geq 0$ , recall that  $1 - x \leq e^{-x}$ . Therefore,

$$\mathbb{P}\left(\bigcap_{m \leq n \leq r} E_n^c\right) \leq \prod_{m \leq n \leq r} e^{-\mathbb{P}(E_n)} = e^{-\sum_{m \leq n \leq r} \mathbb{P}(E_n)}.$$

Because both sides of the inequation above are non-increasing with respect to  $r$ , we may take the limit of both sides when  $r \rightarrow \infty$  and use the fact that  $\sum_{n=0}^{\infty} \mathbb{P}(E_n) = \infty$  to conclude that

$$\lim_{r \rightarrow \infty} \mathbb{P}\left(\bigcap_{m \leq n \leq r} E_n^c\right) = \mathbb{P}\left(\bigcap_{n \geq m} E_n^c\right) \leq \lim_{r \rightarrow \infty} e^{-\sum_{m \leq n \leq r} \mathbb{P}(E_n)} = 0.$$

Using the relationship between the limit superior and the limit inferior,

$$\mathbb{P} \left( \left( \limsup_{n \rightarrow \infty} E_n \right)^c \right) = \mathbb{P} \left( \liminf_{n \rightarrow \infty} E_n^c \right) = \mathbb{P} \left( \bigcup_m \bigcap_{n \geq m} E_n^c \right) \leq \sum_m \mathbb{P} \left( \bigcap_{n \geq m} E_n^c \right) = 0.$$

A valid distribution function  $F : \mathbb{R} \rightarrow [0, 1]$  is a right-continuous function such that if  $a \leq b$ , then  $F(a) \leq F(b)$ ;  $\lim_{x \rightarrow -\infty} F(x) = 0$ ; and  $\lim_{x \rightarrow \infty} F(x) = 1$ . For any sequence of valid distribution functions  $(F_n \mid n \in \mathbb{N})$ , it is possible to show that there is a sequence of independent random variables  $(X_n \mid n \in \mathbb{N})$  on the probability triple  $([0, 1], \mathcal{B}([0, 1]), \text{Leb})$  such that  $F_n$  is the distribution function of  $X_n$ .

Let  $(X_n \mid n \in \mathbb{N})$  be a sequence of independent random variables on the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ . If  $\mathbb{P}(X_n \leq x) = F(x)$  for every  $n \in \mathbb{N}$ ,  $x \in \mathbb{R}$ , and a distribution function  $F : \mathbb{R} \rightarrow [0, 1]$ , then the random variables are considered independent and identically distributed.

As an application of the Borel-Cantelli lemmas, consider a sequence  $(X_n \mid n \in \mathbb{N}^+)$  of independent random variables on the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ . Suppose that each random variable  $X_n$  is exponentially distributed with rate 1 such that  $\mathbb{P}(X_n > x_n) = 1 - \mathbb{P}(X_n \leq x_n) = e^{-x_n}$  for every  $x_n \geq 0$ . If  $x_n = \alpha \log n$  for some  $\alpha > 0$ , then

$$\mathbb{P}(X_n > \alpha \log n) = e^{-\alpha \log n} = (e^{\log n})^{-\alpha} = \frac{1}{n^\alpha}.$$

For some  $\alpha > 0$ , consider the sequence of independent events  $(\{X_n > \alpha \log n\} \in \mathcal{F} \mid n \in \mathbb{N}^+)$  and recall that

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n > \alpha \log n) = \sum_{n=1}^{\infty} \frac{1}{n^\alpha} < \infty$$

if and only if  $\alpha > 1$ . Using the Borel-Cantelli lemmas,

$$\mathbb{P} \left( \limsup_{n \rightarrow \infty} \{X_n > \alpha \log n\} \right) = \begin{cases} 0, & \text{if } \alpha > 1, \\ 1, & \text{if } \alpha \leq 1. \end{cases}$$

Recall that  $\omega \in \limsup_{n \rightarrow \infty} \{X_n > \alpha \log n\}$  if and only if  $X_n(\omega) > \alpha \log n$  for infinitely many  $n \in \mathbb{N}$ . Furthermore, consider the random variable  $\limsup_{n \rightarrow \infty} X_n / \log n$ . It is also possible to show that

$$\mathbb{P} \left( \limsup_{n \rightarrow \infty} \frac{X_n}{\log n} = 1 \right) = \mathbb{P} \left( \left\{ \omega \in \Omega \mid \limsup_{n \rightarrow \infty} \frac{X_n(\omega)}{\log n} = 1 \right\} \right) = 1.$$

For any set  $\mathcal{C}$ , a set (or sequence) of random variables  $Y = (Y_\gamma \mid \gamma \in \mathcal{C})$  on a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  is called a stochastic process parameterized by  $\mathcal{C}$ .

Consider a measurable space  $(\Omega, \mathcal{F})$  and a function  $X : \Omega \rightarrow C$ , where  $C \subseteq \mathbb{N}$ . We will show that if  $\{X = c\} \in \mathcal{F}$  for every  $c \in C$ , then  $X$  is  $\mathcal{F}$ -measurable. For any  $B \in \mathcal{B}(\mathbb{R})$ , let  $A = B \cap C$  and note that

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} = \{\omega \in \Omega \mid X(\omega) \in B \text{ and } X(\omega) \in C\} = X^{-1}(B \cap C) = X^{-1}(A).$$

Furthermore, note that

$$X^{-1}(A) = X^{-1} \left( \bigcup_{a \in A} \{a\} \right) = \bigcup_{a \in A} X^{-1}(\{a\}) = \bigcup_{a \in A} \{X = a\}.$$

Because  $A \subseteq C$ , we have  $\{X = a\} \in \mathcal{F}$  for every  $a \in A$ . Because  $\mathcal{F}$  is a  $\sigma$ -algebra, we have  $X^{-1}(A) \in \mathcal{F}$ . Therefore, for every  $B \in \mathcal{B}(\mathbb{R})$ , we have  $X^{-1}(B) \in \mathcal{F}$ .

Consider a set  $E \subseteq \mathbb{N}$ . For every  $i, j \in E$ , let  $P$  be a stochastic matrix whose  $(i, j)$ -th element is given by  $p_{i,j} \geq 0$  and suppose that  $\sum_k p_{i,k} = 1$ . Let  $\mu$  be a probability measure on the measurable space  $(E, \mathcal{P}(E))$ , where  $\mathcal{P}(E)$  is the set of all subsets of  $E$ , and let  $\mu_i$  denote  $\mu(\{i\})$  for every  $i \in E$ . A time-homogeneous Markov chain  $Z = (Z_n \mid n \in \mathbb{N})$  on  $E$  with initial distribution  $\mu$  and 1-step transition matrix  $P$  is a stochastic process parameterized by  $\mathbb{N}$  such that, for every  $n \in \mathbb{N}$  and  $i_0, i_1, \dots, i_n \in E$ ,

$$\mathbb{P}(Z_0 = i_0, Z_1 = i_1, \dots, Z_n = i_n) = \mu_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, i_n} = \mu_{i_0} \prod_{k=1}^n p_{i_{k-1}, i_k}.$$

We will now show that a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  carrying the aforementioned stochastic process  $Z$  exists.

First, for any set of valid distribution functions  $\{F_n \mid n \in \mathbb{N}\}$ , recall that there is a set of independent random variables  $\{X_n \mid n \in \mathbb{N}\}$  on a certain probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  such that  $F_n$  is the distribution function of  $X_n$ . Using this result, for every  $i, j \in E$  and  $n \in \mathbb{N}^+$ , let  $Z_0 : \Omega \rightarrow E$  and  $Y_{i,n} : \Omega \rightarrow E$  be independent random variables on a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  such that  $\mathbb{P}(Z_0 = i) = \mu_i$  and  $\mathbb{P}(Y_{i,n} = j) = p_{i,j}$ .

For every  $\omega \in \Omega$  and  $n \in \mathbb{N}^+$ , let  $Z_n(\omega) = Y_{Z_{n-1}(\omega), n}(\omega)$ . Using induction, we will show that the function  $Z_n : \Omega \rightarrow E$  is a random variable for every  $n \in \mathbb{N}$ . We already know that  $Z_0$  is a random variable. Suppose that  $Z_{n-1}$  is a random variable. We will show that  $\{Z_n = i_n\} \in \mathcal{F}$  for every  $i_n \in E$ . By definition,

$$\{Z_n = i_n\} = \{\omega \in \Omega \mid Z_n(\omega) = i_n\} = \{\omega \in \Omega \mid Y_{Z_{n-1}(\omega), n}(\omega) = i_n\} = \bigcup_{i \in E} \{\omega \in \Omega \mid Z_{n-1}(\omega) = i \text{ and } Y_{i,n}(\omega) = i_n\},$$

which implies

$$\{Z_n = i_n\} = \bigcup_{i \in E} \{Z_{n-1} = i\} \cap \{Y_{i,n} = i_n\}.$$

Because  $Z_{n-1}$  and  $Y_{i,n}$  are random variables for every  $i \in E$ ,  $\{Z_n = i_n\} \in \mathcal{F}$ , as we wanted to show.

Using induction, we will now show that, for every  $n \in \mathbb{N}$  and  $i_0, i_1, \dots, i_n \in E$ ,

$$\bigcap_{k=0}^n \{Z_k = i_k\} = \{Z_0 = i_0\} \cap \bigcap_{k=1}^n \{Y_{i_{k-1}, k} = i_k\}.$$

The statement above is true when  $n = 0$ , so suppose it is true for some  $n - 1 \in \mathbb{N}$ . Using a previous result,

$$\bigcap_{k=0}^n \{Z_k = i_k\} = \left( \bigcap_{k=0}^{n-1} \{Z_k = i_k\} \right) \cap \{Z_n = i_n\} = \left( \bigcap_{k=0}^{n-1} \{Z_k = i_k\} \right) \cap \left( \bigcup_{i \in E} \{Z_{n-1} = i\} \cap \{Y_{i,n} = i_n\} \right).$$

By distributing the intersection over the union,

$$\bigcap_{k=0}^n \{Z_k = i_k\} = \bigcup_{i \in E} \left( \bigcap_{k=0}^{n-1} \{Z_k = i_k\} \right) \cap \{Z_{n-1} = i\} \cap \{Y_{i,n} = i_n\}.$$

Because  $\{Z_{n-1} = i_{n-1}\} \cap \{Z_{n-1} = i\} = \emptyset$  whenever  $i \neq i_{n-1}$ ,

$$\bigcap_{k=0}^n \{Z_k = i_k\} = \left( \bigcap_{k=0}^{n-1} \{Z_k = i_k\} \right) \cap \{Y_{i_{n-1}, n} = i_n\} = \{Z_0 = i_0\} \cap \bigcap_{k=1}^n \{Y_{i_{k-1}, k} = i_k\},$$

where the last equation follows from the inductive hypothesis.

The event above is the intersection of events from the  $\sigma$ -algebras of independent random variables, which implies

$$\mathbb{P}(Z_0 = i_0, Z_1 = i_1, \dots, Z_n = i_n) = \mathbb{P}\left(\bigcap_{k=0}^n \{Z_k = i_k\}\right) = \mathbb{P}(Z_0 = i_0) \prod_{k=1}^n \mathbb{P}(Y_{i_{k-1}, k} = i_k) = \mu_{i_0} \prod_{k=1}^n p_{i_{k-1}, i_k}.$$

Consider a time-homogeneous Markov chain  $Z = (Z_n \mid n \in \mathbb{N})$  on  $E$  with initial distribution  $\mu$  and 1-step transition matrix  $P$ . Consider also a finite sequence of elements of  $E$  given by  $I = i_0, i_1, \dots, i_n$ . We say that the sequence  $I$  appears in outcome  $\omega \in \Omega$  at time  $t$  if  $Z_{t+k}(\omega) = i_k$  for every  $k \leq n$ . We will now show how several interesting events related to the appearance of the sequence  $I$  may be defined.

The event  $M_t$  composed of outcomes where the sequence  $I$  appears at time  $t$  is given by

$$M_t = \bigcap_{k=0}^n \{Z_{t+k} = i_k\} = \bigcap_{k=0}^n \{\omega \in \Omega \mid Z_{t+k}(\omega) = i_k\}.$$

The event  $S_t$  composed of outcomes where the sequence  $I$  appears at least once at or after time  $t$  is given by

$$S_t = \bigcup_{t' \geq t} M_{t'}.$$

The event  $L_{t,m}$  composed of outcomes where the sequence  $I$  appears at least  $m$  times up to time  $t$  is given by

$$L_{t,m} = \bigcup_{l_1, \dots, l_m} \bigcap_{k=1}^m M_{l_k},$$

where  $l_1, \dots, l_m$  is a finite sequence of distinct elements of  $E$  such that  $l_k \leq t$  for every  $k \leq m$ .

The event  $L_m$  composed of outcomes where  $I$  appears at least  $m$  times is given by  $L_{t,m}$  when  $t = \infty$ .

The event  $E$  composed of outcomes where the sequence  $I$  appears infinitely many times is given by

$$E = \limsup_{t \rightarrow \infty} M_t.$$

## 5 Integration

Consider a measure space  $(S, \Sigma, \mu)$ . The integral with respect to  $\mu$  of a  $\Sigma$ -measurable function  $f : S \rightarrow \mathbb{R}$  is denoted by  $\mu(f)$ .

For any set  $A \in \Sigma$ , the integral with respect to  $\mu$  of the indicator function  $\mathbb{I}_A : S \rightarrow \{0, 1\}$  is defined as

$$\mu(\mathbb{I}_A) = \mu(A).$$

A simple function is a  $\Sigma$ -measurable function  $f : S \rightarrow [0, \infty]$  that can be written as

$$f(s) = \sum_{k=1}^m a_k \mathbb{I}_{A_k}(s)$$

for every  $s \in S$ , for some fixed  $a_1, a_2, \dots, a_m \in [0, \infty]$  and  $A_1, A_2, \dots, A_m \in \Sigma$ . Intuitively, when  $A_1, A_2, \dots, A_m$  partition  $S$ , each set  $A_k$  is assigned a value  $a_k$ .

The integral with respect to  $\mu$  of the simple function  $f : S \rightarrow [0, \infty]$  as written above is defined as

$$\mu(f) = \sum_{k=1}^m a_k \mu(A_k).$$

It is possible to show that the right side of the equation above is equivalent for every choice of sets and constants used to write the simple function  $f$ . Therefore, the integral  $\mu(f)$  with respect to  $\mu$  of a simple function  $f$  is well-defined. Intuitively, when  $A_1, A_2, \dots, A_m$  partition  $S$ , the integral with respect to  $\mu$  accumulates the measure  $\mu(A_k)$  of each set  $A_k$  multiplied by the value  $a_k$  assigned to it.

If  $f : S \rightarrow [0, \infty]$  and  $g : S \rightarrow [0, \infty]$  are simple functions, then

- $f + g$  is a simple function and  $\mu(f + g) = \mu(f) + \mu(g)$ ,
- if  $c \geq 0$ , then  $cf$  is a simple function and  $\mu(cf) = c\mu(f)$ ,
- if  $\mu(f \neq g) = \mu(\{s \in S \mid f(s) \neq g(s)\}) = 0$ , then  $\mu(f) = \mu(g)$ ,
- if  $f \leq g$  such that  $f(s) \leq g(s)$  for every  $s \in S$ , then  $\mu(f) \leq \mu(g)$ ,
- if  $h = \min(f, g)$  such that  $h(s) = \min(f(s), g(s))$  for every  $s \in S$ , then  $h$  is a simple function,
- if  $h = \max(f, g)$  such that  $h(s) = \max(f(s), g(s))$  for every  $s \in S$ , then  $h$  is a simple function.

The integral with respect to  $\mu$  of a  $\Sigma$ -measurable function  $f : S \rightarrow [0, \infty]$  is defined as

$$\mu(f) = \sup\{\mu(h) \mid h \text{ is simple and } h \leq f\}.$$

Consider a  $\Sigma$ -measurable function  $f : S \rightarrow [0, \infty]$ . We will now show that if  $\mu(f) = 0$ , then  $\mu(\{f > 0\}) = 0$ . Because the measure  $\mu$  is non-negative, this is equivalent to showing that if  $\mu(\{f > 0\}) > 0$ , then  $\mu(f) > 0$ .

For every  $n \in \mathbb{N}^+$ , let  $A_n = \{f > n^{-1}\} = \{s \in S \mid f(s) > n^{-1}\}$  and note that

$$\{f > 0\} = \{s \in S \mid f(s) > 0\} = \bigcup_{n \in \mathbb{N}^+} \{s \in S \mid f(s) > n^{-1}\} = \bigcup_{n \in \mathbb{N}^+} A_n.$$

For every  $s \in S$  and  $n \in \mathbb{N}^+$ , if  $f(s) > n^{-1}$ , then  $f(s) > (n+1)^{-1}$ . Therefore,  $A_n \subseteq A_{n+1}$  and  $A_n \uparrow \{f > 0\}$ . Furthermore, the monotone-convergence property of measure guarantees that  $\mu(A_n) \uparrow \mu(\{f > 0\})$ .

Suppose that  $\mu(\{f > 0\}) > 0$ . In that case, there is an  $n \in \mathbb{N}^+$  such that

$$\mu(\mathbb{I}_{\{f > n^{-1}\}}) = \mu(\{f > n^{-1}\}) = \mu(A_n) > 0.$$

For such an  $n \in \mathbb{N}^+$ , consider now the simple function  $g = n^{-1}\mathbb{I}_{\{f > n^{-1}\}}$  given by

$$g(s) = n^{-1}\mathbb{I}_{\{f > n^{-1}\}}(s) = \begin{cases} n^{-1} & f(s) > n^{-1}, \\ 0 & f(s) \leq n^{-1}. \end{cases}$$

The fact that  $f \geq g$  implies that  $\mu(f) \geq \mu(g)$  even if  $f$  is not simple. Therefore,

$$\mu(f) \geq \mu(g) = \mu(n^{-1}\mathbb{I}_{\{f > n^{-1}\}}) = n^{-1}\mu(\mathbb{I}_{\{f > n^{-1}\}}) > 0,$$

where the last inequality follows from the fact that  $n^{-1} > 0$ .

Let  $f_n \uparrow f$  denote that a sequence of functions  $(f_n : S \rightarrow \mathbb{R} \mid n \in \mathbb{N})$  is such that  $f_n(s) \uparrow f(s)$  for every  $s \in S$ . Similarly, let  $f_n \downarrow f$  denote that a sequence of functions  $(f_n : S \rightarrow \mathbb{R} \mid n \in \mathbb{N})$  is such that  $f_n(s) \downarrow f(s)$  for every  $s \in S$ .

The monotone-convergence theorem states that if  $(f_n : S \rightarrow [0, \infty] \mid n \in \mathbb{N})$  is a sequence of  $\Sigma$ -measurable functions such that  $f_n \uparrow f$ , then  $\mu(f_n) \uparrow \mu(f)$ .

Before showing how the integral with respect to  $\mu$  of a given  $\Sigma$ -measurable function is the limit of a sequence of integrals with respect to  $\mu$  of simple functions, it is convenient to introduce staircase functions.

Let  $\alpha_n : [0, \infty] \rightarrow [0, n]$  denote the  $n$ -th staircase function given by  $\alpha_n(x) = \min(n, \lfloor 2^n x \rfloor / 2^n)$  for every  $n \in \mathbb{N}$  and  $x \in [0, \infty]$ . Intuitively, the  $n$ -th staircase function partitions its domain into a sequence of intervals of length  $1/2^n$ . The value assigned to the first interval is zero, and the value of each following interval is  $1/2^n$  plus the value of the previous interval, with values truncated at  $n$ . Furthermore, let  $h : [0, \infty] \rightarrow [0, \infty]$  denote the identity function given by  $h(x) = x$  for every  $x \in [0, \infty]$ . We will now show that  $\alpha_n \uparrow h$ .

We will start by showing that  $\min(n, \lfloor 2^n x \rfloor / 2^n) = \alpha_n(x) \leq \alpha_{n+1}(x) = \min(n+1, \lfloor 2^{n+1} x \rfloor / 2^{n+1})$ , for every  $n \in \mathbb{N}$  and  $x \in [0, \infty]$ . When  $x = \infty$ , we have  $\alpha_n(x) = n \leq n+1 = \alpha_{n+1}(x)$ . When  $x < \infty$ , the fact that  $n \leq n+1$  implies that we only need to show that  $\lfloor 2^n x \rfloor / 2^n \leq \lfloor 2^{n+1} x \rfloor / 2^{n+1}$ . Note that  $\lfloor 2^n x \rfloor \leq 2^n x$ , which implies  $2\lfloor 2^n x \rfloor \leq 2^{n+1} x$ . By the monotonicity of the floor function,  $2\lfloor 2^n x \rfloor \leq \lfloor 2^{n+1} x \rfloor$ . Because the floor of an integer is itself an integer,  $2\lfloor 2^n x \rfloor \leq \lfloor 2^{n+1} x \rfloor$ . Dividing both sides of the previous inequation by  $2^{n+1}$  completes the proof.

In order to show that  $\alpha_n \uparrow h$ , it remains to show that, for every  $x \in [0, \infty]$ ,

$$\lim_{n \rightarrow \infty} \alpha_n(x) = x.$$

The case where  $x = \infty$  is trivial, since  $\alpha_n(x) = n$ . When  $x < \infty$ , note that  $2^n x \geq \lfloor 2^n x \rfloor$  implies  $x \geq \lfloor 2^n x \rfloor / 2^n$ , and so  $n > x$  implies  $n > \lfloor 2^n x \rfloor / 2^n$ . Therefore, for every sufficiently large  $n \in \mathbb{N}$ , we know that  $\alpha_n(x) = \lfloor 2^n x \rfloor / 2^n$  when  $x < \infty$ . It remains to show that  $\lim_{n \rightarrow \infty} \lfloor 2^n x \rfloor / 2^n = x$ . By noting that  $2^n x - 1 \leq \lfloor 2^n x \rfloor \leq 2^n x$  and dividing each term by  $2^n$ ,

$$x - \frac{1}{2^n} = \frac{2^n x - 1}{2^n} \leq \frac{\lfloor 2^n x \rfloor}{2^n} \leq \frac{2^n x}{2^n} = x.$$

Using the squeeze theorem with  $n \rightarrow \infty$  completes the proof that  $\alpha_n \uparrow h$ .

Consider a  $\Sigma$ -measurable function  $f : S \rightarrow [0, \infty]$ . For every  $n \in \mathbb{N}$ , consider  $f_n : S \rightarrow [0, n]$  such that

$$f_n(s) = \alpha_n(f(s)) = \sum_{k=1}^m a_k \mathbb{I}_{\{f_n = a_k\}}(s),$$

where  $a_1, \dots, a_m \in [0, n]$  are the (distinct) elements of the (finite) image of the function  $f_n$ . Because  $f$  is  $\Sigma$ -measurable and  $\alpha_n$  is  $\mathcal{B}([0, \infty])$ -measurable, we know that  $f_n = \alpha_n \circ f$  is  $\Sigma$ -measurable, which implies that  $f_n$  is also simple. For every  $s \in S$ , we have  $f(s) \in [0, \infty]$  and  $(\alpha_n \circ f)(s) \uparrow f(s)$ . Therefore,  $f_n \uparrow f$ . From the monotone-convergence theorem,  $\mu(f_n) \uparrow \mu(f)$ . Therefore, the integral with respect to  $\mu$  of a given  $\Sigma$ -measurable function  $f$  is the limit of a sequence of integrals with respect to  $\mu$  of simple functions  $(f_n : S \rightarrow [0, n] \mid n \in \mathbb{N})$ .

Let  $f : S \rightarrow [0, \infty]$  and  $g : S \rightarrow [0, \infty]$  be  $\Sigma$ -measurable functions. We will show that if  $\mu(\{f \neq g\}) = 0$ , then  $\mu(f) = \mu(g)$ . Recall that we already have the analogous result for simple functions.

For any  $n \in \mathbb{N}$ , let  $f_n = \alpha_n \circ f$  and  $g_n = \alpha_n \circ g$ , where  $\alpha_n$  is the  $n$ -th staircase function. Note that

$$\{f_n \neq g_n\} = \{s \in S \mid f_n(s) \neq g_n(s)\} \subseteq \{s \in S \mid f(s) \neq g(s)\} = \{f \neq g\},$$

which implies  $\mu(\{f_n \neq g_n\}) \leq \mu(\{f \neq g\}) = 0$ . Because  $f_n$  and  $g_n$  are simple functions such that  $\mu(\{f_n \neq g_n\}) = 0$ , we know that  $\mu(f_n) = \mu(g_n)$ . From the monotone-convergence theorem,  $\mu(f_n) \uparrow \mu(f)$  and  $\mu(g_n) \uparrow \mu(g)$ , so

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

Consider a  $\Sigma$ -measurable function  $f : S \rightarrow [0, \infty]$  and a sequence of  $\Sigma$ -measurable functions  $(f_n : S \rightarrow [0, \infty] \mid n \in \mathbb{N})$  such that  $f_n(s) \uparrow f(s)$  for every  $s \in S \setminus N$  for some  $\mu$ -null set  $N \subseteq S$ . We will show that  $\mu(f_n) \uparrow \mu(f)$ .

Consider the  $\Sigma$ -measurable function  $f \mathbb{I}_{S \setminus N}$  such that  $(f \mathbb{I}_{S \setminus N})(s) = f(s) \mathbb{I}_{S \setminus N}(s)$  for every  $s \in S$ . Clearly,  $\{f \mathbb{I}_{S \setminus N} \neq f\} \subseteq N$ . Therefore,  $\mu(\{f \mathbb{I}_{S \setminus N} \neq f\}) \leq \mu(N) = 0$  and  $\mu(f \mathbb{I}_{S \setminus N}) = \mu(f)$ .

Analogously, consider the  $\Sigma$ -measurable function  $f_n \mathbb{I}_{S \setminus N}$  such that  $(f_n \mathbb{I}_{S \setminus N})(s) = f_n(s) \mathbb{I}_{S \setminus N}(s)$  for every  $s \in S$  and  $n \in \mathbb{N}$ . Clearly,  $\{f_n \mathbb{I}_{S \setminus N} \neq f_n\} \subseteq N$ . Therefore,  $\mu(\{f_n \mathbb{I}_{S \setminus N} \neq f_n\}) \leq \mu(N) = 0$  and  $\mu(f_n \mathbb{I}_{S \setminus N}) = \mu(f_n)$ .

Note that  $(f_n \mathbb{I}_{S \setminus N})(s) \uparrow (f \mathbb{I}_{S \setminus N})(s)$ , whether  $s \in N$  or  $s \in S \setminus N$ . Therefore,  $\mu(f_n \mathbb{I}_{S \setminus N}) \uparrow \mu(f \mathbb{I}_{S \setminus N})$ , which implies  $\mu(f_n) \uparrow \mu(f)$ .

Consider a sequence of  $\Sigma$ -measurable functions  $(f_n : S \rightarrow [0, \infty] \mid n \in \mathbb{N})$ . The Fatou lemma states that

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

We will now show this lemma. For any  $m \in \mathbb{N}$ , consider the function  $g_m = \inf_{n \geq m} f_n$  such that

$$\liminf_{n \rightarrow \infty} f_n = \lim_{m \rightarrow \infty} \inf_{n \geq m} f_n = \lim_{m \rightarrow \infty} g_m.$$

Because  $g_{m+1} \geq g_m$  for every  $m \in \mathbb{N}$ , we have that  $g_m \uparrow \liminf_{n \rightarrow \infty} f_n$ . Because  $g_m : S \rightarrow [0, \infty]$  is also  $\Sigma$ -measurable for every  $m \in \mathbb{N}$ , the monotone-convergence theorem guarantees that  $\mu(g_m) \uparrow \mu(\liminf_{n \rightarrow \infty} f_n)$ .

For any  $n \geq m$ , note that  $g_m \leq f_n$  and  $\mu(g_m) \leq \mu(f_n)$ , which also implies  $\mu(g_m) \leq \inf_{n \geq m} \mu(f_n)$ . By taking the limit of both sides of the previous inequation when  $m \rightarrow \infty$ ,

$$\mu\left(\liminf_{n \rightarrow \infty} f_n\right) = \lim_{m \rightarrow \infty} \mu(g_m) \leq \lim_{m \rightarrow \infty} \inf_{n \geq m} \mu(f_n) = \liminf_{n \rightarrow \infty} \mu(f_n).$$

Consider a  $\Sigma$ -measurable function  $f : S \rightarrow [0, \infty]$  and a constant  $c \geq 0$ . We will now show that  $\mu(cf) = c\mu(f)$ . Recall that we already have the analogous result for simple functions.

For any  $n \in \mathbb{N}$ , let  $f_n = \alpha_n \circ f$ , where  $\alpha_n$  is the  $n$ -th staircase function. Because  $f_n \uparrow f$ , we know that  $cf_n \uparrow cf$ . Because  $cf_n$  is  $\Sigma$ -measurable for every  $n \in \mathbb{N}$ , the monotone-convergence theorem guarantees that  $\mu(cf_n) \uparrow \mu(cf)$ . Because  $\mu(cf_n) = c\mu(f_n)$ , we have  $c\mu(f_n) \uparrow \mu(cf)$ . Because  $c\mu(f_n) \uparrow c\mu(f)$ , we have  $\mu(cf) = c\mu(f)$ .

Consider a  $\Sigma$ -measurable function  $f : S \rightarrow [0, \infty]$  and a  $\Sigma$ -measurable function  $g : S \rightarrow [0, \infty]$ . We will now show that  $\mu(f + g) = \mu(f) + \mu(g)$ . Recall that we already have the analogous result for simple functions.

For any  $n \in \mathbb{N}$ , let  $f_n = \alpha_n \circ f$  and  $g_n = \alpha_n \circ g$ , where  $\alpha_n$  is the  $n$ -th staircase function. Because  $f_n \uparrow f$  and  $g_n \uparrow g$ , we know that  $f_n + g_n \uparrow f + g$ . Because  $f_n + g_n$  is  $\Sigma$ -measurable for every  $n \in \mathbb{N}$ , the monotone-convergence theorem guarantees that  $\mu(f_n + g_n) \uparrow \mu(f + g)$ . Because  $\mu(f_n + g_n) \uparrow \mu(f) + \mu(g)$ , we have  $\mu(f + g) = \mu(f) + \mu(g)$ .

Consider a sequence of  $\Sigma$ -measurable functions  $(f_n : S \rightarrow [0, \infty] \mid n \in \mathbb{N})$  such that  $f_n \leq g$  for every  $n \in \mathbb{N}$  and some  $\Sigma$ -measurable function  $g : S \rightarrow [0, \infty]$  such that  $\mu(g) < \infty$ . The reverse Fatou lemma states that

$$\mu\left(\limsup_{n \rightarrow \infty} f_n\right) \geq \limsup_{n \rightarrow \infty} \mu(f_n).$$

We will now show this lemma. For every  $n \in \mathbb{N}$ , consider the function  $h_n = g - f_n$ . Because  $g$  and  $f_n$  are  $\Sigma$ -measurable and  $f_n \leq g$ , we know that  $h_n : S \rightarrow [0, \infty]$  is  $\Sigma$ -measurable. From the Fatou lemma,

$$\mu\left(\liminf_{n \rightarrow \infty} (g - f_n)\right) \leq \liminf_{n \rightarrow \infty} \mu(g - f_n).$$

By using the fact that  $\mu(g) = \mu(g - f_n) + \mu(f_n)$  and moving  $g$  and  $\mu(g)$  outside the corresponding limits,

$$\mu\left(g + \liminf_{n \rightarrow \infty} -f_n\right) \leq \mu(g) + \liminf_{n \rightarrow \infty} -\mu(f_n).$$

By using the relationship between limit inferior and limit superior,

$$\mu\left(g - \limsup_{n \rightarrow \infty} f_n\right) \leq \mu(g) - \limsup_{n \rightarrow \infty} \mu(f_n).$$

By using the fact that  $\mu(g) = \mu(g - \limsup_{n \rightarrow \infty} f_n) + \mu(\limsup_{n \rightarrow \infty} f_n)$ ,

$$\mu(g) - \mu\left(\limsup_{n \rightarrow \infty} f_n\right) \leq \mu(g) - \limsup_{n \rightarrow \infty} \mu(f_n).$$

The proof is completed by reorganizing terms in the inequation above.

For a  $\Sigma$ -measurable function  $f : S \rightarrow \mathbb{R}$ , the  $\Sigma$ -measurable function  $f^+ : S \rightarrow [0, \infty]$  is given by

$$f^+(s) = \max(f(s), 0) = \begin{cases} f(s), & \text{if } f(s) > 0, \\ 0, & \text{if } f(s) \leq 0. \end{cases}$$

For a  $\Sigma$ -measurable function  $f : S \rightarrow \mathbb{R}$ , the  $\Sigma$ -measurable function  $f^- : S \rightarrow [0, \infty]$  is given by

$$f^-(s) = \max(-f(s), 0) = \begin{cases} 0, & \text{if } f(s) > 0, \\ -f(s), & \text{if } f(s) \leq 0. \end{cases}$$

Therefore, for a  $\Sigma$ -measurable function  $f : S \rightarrow \mathbb{R}$ , whether  $f(s) > 0$  or  $f(s) \leq 0$ ,

$$f(s) = f^+(s) - f^-(s).$$

Furthermore, whether  $f(s) > 0$  or  $f(s) \leq 0$ ,

$$|f(s)| = f^+(s) + f^-(s).$$

In other words,  $f = f^+ - f^-$  and  $|f| = f^+ + f^-$ .

A function  $f : S \rightarrow \mathbb{R}$  is  $\mu$ -integrable if it is  $\Sigma$ -measurable and  $\mu(|f|) = \mu(f^+ + f^-) = \mu(f^+) + \mu(f^-) < \infty$ .

The set of all  $\mu$ -integrable functions in the measure space  $(S, \Sigma, \mu)$  is denoted by  $\mathcal{L}^1(S, \Sigma, \mu)$ . The set of all non-negative  $\mu$ -integrable functions in the measure space  $(S, \Sigma, \mu)$  is denoted by  $\mathcal{L}^1(S, \Sigma, \mu)^+$ .

The integral  $\mu(f)$  with respect to  $\mu$  of a  $\mu$ -integrable function  $f : S \rightarrow \mathbb{R}$  is defined as

$$\mu(f) = \mu(f^+) - \mu(f^-).$$

Alternatively, the integral  $\mu(f)$  with respect to  $\mu$  of a  $\mu$ -integrable function  $f : S \rightarrow \mathbb{R}$  is denoted by

$$\int_S f d\mu = \int_S f(s) \mu(ds) = \mu(f).$$

If a function  $f : S \rightarrow \mathbb{R}$  is  $\mu$ -integrable, then  $\mu(f^+) < \infty$  and  $\mu(f^-) < \infty$ . By the triangle inequality,

$$|\mu(f)| = |\mu(f^+) + (-\mu(f^-))| \leq |\mu(f^+)| + |-\mu(f^-)| = \mu(f^+) + \mu(f^-) = \mu(|f|).$$

Consider a  $\mu$ -integrable function  $f : S \rightarrow \mathbb{R}$ . Because  $-f : S \rightarrow \mathbb{R}$  is  $\Sigma$ -measurable and  $\mu(|-f|) = \mu(|f|) < \infty$ , we know that  $-f$  is  $\mu$ -integrable. We will now show that  $\mu(-f) = -\mu(f)$ . For every  $s \in S$ ,  $(-f)^+(s) = \max(-f(s), 0) = f^-(s)$  and  $(-f)^-(s) = \max(f(s), 0) = f^+(s)$ . Therefore,

$$\mu(-f) = \mu((-f)^+) - \mu((-f)^-) = -(\mu((-f)^-) - \mu((-f)^+)) = -(\mu(f^+) - \mu(f^-)) = -\mu(f).$$

Consider a  $\mu$ -integrable function  $f : S \rightarrow \mathbb{R}$  and a constant  $c \in \mathbb{R}$ . Because  $cf : S \rightarrow \mathbb{R}$  is  $\Sigma$ -measurable and  $\mu(|cf|) = \mu(|c||f|) = |c|\mu(|f|) < \infty$ , we know that  $cf$  is  $\mu$ -integrable. We will now show that  $\mu(cf) = c\mu(f)$ .

Because  $f = f^+ - f^-$ , we know that  $cf = cf^+ - cf^-$ . Furthermore,  $(cf)^+ = (cf^+)^+ - (cf^-)^+$ . Therefore,

$$(cf)^+ - (cf)^- = cf^+ - cf^-.$$

By rearranging negative terms,

$$(cf)^+ + cf^- = (cf)^- + cf^+.$$



We will now consider the case where  $c \geq 0$ . By the linearity of the integral of non-negative functions,

$$\mu((cf)^+) + \mu(cf^-) = \mu((cf)^-) + \mu(cf^+).$$

By rearranging terms,

$$\mu((cf)^+) - \mu((cf)^-) = \mu(cf^+) - \mu(cf^-).$$

Because  $cf$  is  $\mu$ -integrable and by the linearity of the integral of non-negative functions,

$$\mu(cf) = c\mu(f^+) - c\mu(f^-) = c(\mu(f^+) - \mu(f^-)) = c\mu(f).$$

When  $c < 0$ , note that  $\mu(cf) = \mu(-|c|f) = |c|\mu(-f) = -|c|\mu(f) = c\mu(f)$ .

Consider a  $\mu$ -integrable function  $f : S \rightarrow \mathbb{R}$  and a  $\mu$ -integrable function  $g : S \rightarrow \mathbb{R}$ . Because  $f + g : S \rightarrow \mathbb{R}$  is  $\Sigma$ -measurable and  $|f + g| \leq |f| + |g|$  implies  $\mu(|f + g|) \leq \mu(|f|) + \mu(|g|) < \infty$ , we know that  $f + g$  is  $\mu$ -integrable. We will now show that  $\mu(f + g) = \mu(f) + \mu(g)$ .

We know that  $f + g = (f^+ - f^-) + (g^+ - g^-)$ . Furthermore,  $(f + g) = (f + g)^+ - (f + g)^-$ . Therefore,

$$(f + g)^+ - (f + g)^- = (f^+ - f^-) + (g^+ - g^-).$$

By rearranging negative terms,

$$(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+.$$

By the linearity of the integral of non-negative functions,

$$\mu((f + g)^+) + \mu(f^-) + \mu(g^-) = \mu((f + g)^-) + \mu(f^+) + \mu(g^+).$$

By rearranging terms,

$$\mu((f + g)^+) - \mu((f + g)^-) = (\mu(f^+) - \mu(f^-)) + (\mu(g^+) - \mu(g^-))$$

Because  $f + g$  is  $\mu$ -integrable,

$$\mu(f + g) = \mu(f) + \mu(g).$$

Let  $f : S \rightarrow \mathbb{R}$  and  $g : S \rightarrow \mathbb{R}$  be  $\mu$ -integrable functions. We will now show that if  $\mu(\{f \neq g\}) = 0$ , then  $\mu(f) = \mu(g)$ . Recall that we already have the analogous result for non-negative  $\Sigma$ -measurable functions.

First, note that if  $f^+(s) \neq g^+(s)$  or  $f^-(s) \neq g^-(s)$  for some  $s \in S$ , then  $f(s) \neq g(s)$ . Therefore,

$$\{s \in S \mid f^+(s) \neq g^+(s)\} \cup \{s \in S \mid f^-(s) \neq g^-(s)\} \subseteq \{s \in S \mid f(s) \neq g(s)\},$$

so that  $\mu(\{f^+ \neq g^+\}) + \mu(\{f^- \neq g^-\}) \leq \mu(\{f \neq g\})$ . Because  $\mu(\{f \neq g\}) = 0$ , we know that  $\mu(\{f^+ \neq g^+\}) = 0$  and  $\mu(\{f^- \neq g^-\}) = 0$ . Because  $f^+, f^-, g^+$ , and  $g^-$  are non-negative  $\Sigma$ -measurable functions, we know that  $\mu(f^+) = \mu(g^+)$  and  $\mu(f^-) = \mu(g^-)$ . Therefore,

$$\mu(f) = \mu(f^+) - \mu(f^-) = \mu(g^+) - \mu(g^-) = \mu(g).$$

The integral with respect to  $\mu$  of a  $\mu$ -integrable function  $f : S \rightarrow \mathbb{R}$  over the set  $A \in \Sigma$  is defined as

$$\mu(f; A) = \mu(f\mathbb{I}_A).$$

Because  $f\mathbb{I}_A$  is  $\Sigma$ -measurable and  $|f\mathbb{I}_A| \leq |f|$  implies  $\mu(|f\mathbb{I}_A|) \leq \mu(|f|) < \infty$ , we know that  $f\mathbb{I}_A$  is  $\mu$ -integrable. Alternatively, the integral  $\mu(f; A)$  with respect to  $\mu$  of  $f$  over the set  $A \in \Sigma$  is denoted by

$$\int_A f d\mu = \int_A f(s) \mu(ds) = \mu(f; A).$$

Consider a sequence of real numbers  $(x_n \mid n \in \mathbb{N})$  and the measure space  $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ , where  $\mu(\{n\}) = 1$  for every  $n \in \mathbb{N}$ . Furthermore, consider a function  $f : \mathbb{N} \rightarrow \mathbb{R}$  such that  $f(n) = x_n$ . We will now show that  $f$  is  $\mu$ -integrable if and only if  $\sum_n |x_n| < \infty$ . Also, if  $f$  is  $\mu$ -integrable, then  $\mu(f) = \sum_n x_n$ .

Suppose that  $f(n) \geq 0$  for every  $n \in \mathbb{N}$ . For every  $k \in \mathbb{N}$ , consider the function  $f_k : \mathbb{N} \rightarrow [0, \infty]$  given by

$$f_k(n) = \sum_{i=0}^k f(i) \mathbb{I}_{\{i\}}(n) = \begin{cases} f(n), & \text{if } n \leq k, \\ 0, & \text{if } n > k. \end{cases}$$

Clearly, if  $k \rightarrow \infty$ , then  $f_k \rightarrow f$ . Because  $f_k$  is a simple function,

$$\mu(f_k) = \sum_{i=0}^k f(i) \mu(\{i\}) = \sum_{i=0}^k f(i) = \sum_{i=0}^k x_i.$$

Because  $f_k \leq f_{k+1}$ , we have  $f_k \uparrow f$ . By the monotone-convergence theorem,  $\mu(f_k) \uparrow \mu(f)$ . Therefore,

$$\mu(f) = \lim_{k \rightarrow \infty} \sum_{i=0}^k x_i = \sum_n x_n.$$

Now suppose  $f(n) \in \mathbb{R}$  for every  $n \in \mathbb{N}$ . Based on our previous result,

$$\mu(|f|) = \mu(f^+) + \mu(f^-) = \sum_n \max(x_n, 0) + \max(-x_n, 0) = \sum_n |x_n|.$$

By definition,  $f$  is integrable if and only if  $\mu(|f|) = \sum_n |x_n| < \infty$ , in which case

$$\mu(f) = \mu(f^+) - \mu(f^-) = \sum_n \max(x_n, 0) - \max(-x_n, 0) = \sum_n x_n.$$

Consider a sequence of  $\Sigma$ -measurable functions  $(f_n : S \rightarrow \mathbb{R} \mid n \in \mathbb{N})$  and a  $\Sigma$ -measurable function  $f : S \rightarrow \mathbb{R}$  such that  $\lim_{n \rightarrow \infty} f_n = f$ . Furthermore, suppose there is a  $\mu$ -integrable non-negative function  $g \in \mathcal{L}^1(S, \Sigma, \mu)^+$  that dominates this sequence of functions such that  $|f_n| \leq g$  for every  $n \in \mathbb{N}$ . The dominated convergence theorem states that  $f$  is  $\mu$ -integrable and  $\lim_{n \rightarrow \infty} \mu(f_n) = \mu(f)$ . We will now show this theorem.

Because  $g$  is  $\mu$ -integrable and non-negative,  $\mu(g) = \mu(|g|) < \infty$ . Because  $|f_n| \leq g$  for every  $n \in \mathbb{N}$ , we know that  $\mu(|f_n|) \leq \mu(g) < \infty$ , which implies that  $f_n$  is  $\mu$ -integrable. Because the function  $|\cdot|$  is continuous, we know that  $\lim_{n \rightarrow \infty} |f_n| = |f|$ , which implies  $|f| \leq g$ . Because  $\mu(|f|) \leq \mu(g) < \infty$ , we know that  $f$  is  $\mu$ -integrable.

Because  $|f_n| \leq g$  and  $|f| \leq g$ , we know that  $|f_n| + |f| \leq 2g$ . By the triangle inequality,

$$|f_n - f| = |f_n + (-f)| \leq |f_n| + |f| \leq 2g.$$

Because  $|f_n - f| : S \rightarrow [0, \infty]$  is a  $\Sigma$ -measurable function and  $|f_n - f| \leq 2g$  for every  $n \in \mathbb{N}$ , where  $2g : S \rightarrow [0, \infty]$  is a  $\Sigma$ -measurable function such that  $\mu(2g) = 2\mu(g) < \infty$ , the reverse Fatou lemma states that

$$\mu \left( \limsup_{n \rightarrow \infty} |f_n - f| \right) \geq \limsup_{n \rightarrow \infty} \mu(|f_n - f|).$$

Since the function  $|\cdot|$  is continuous, we know that  $\lim_{n \rightarrow \infty} |f_n - f| = 0$ , where 0 is the zero function. Therefore,

$$\limsup_{n \rightarrow \infty} |f_n - f| = \liminf_{n \rightarrow \infty} |f_n - f| = \lim_{n \rightarrow \infty} |f_n - f| = 0.$$

By taking the integral with respect to  $\mu$  of these non-negative functions,

$$\mu \left( \limsup_{n \rightarrow \infty} |f_n - f| \right) = \mu \left( \liminf_{n \rightarrow \infty} |f_n - f| \right) = \mu \left( \lim_{n \rightarrow \infty} |f_n - f| \right) = \mu(0) = 0.$$

Because  $f_n - f$  is  $\mu$ -integrable for every  $n \in \mathbb{N}$  and  $|\mu(f_n - f)| \leq \mu(|f_n - f|)$ ,

$$0 \geq \limsup_{n \rightarrow \infty} \mu(|f_n - f|) \geq \limsup_{n \rightarrow \infty} |\mu(f_n - f)| \geq \liminf_{n \rightarrow \infty} |\mu(f_n - f)| \geq 0.$$

Because the limit superior and limit inferior in the inequation above must be equal to zero, we know that  $\lim_{n \rightarrow \infty} |\mu(f_n - f)| = 0$ , which implies  $\lim_{n \rightarrow \infty} \mu(f_n - f) = 0$ . By the linearity of the integral with respect to  $\mu$ ,

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

Consider a sequence of  $\mu$ -integrable non-negative functions  $(f_n : S \rightarrow [0, \infty] \mid n \in \mathbb{N})$  and a  $\mu$ -integrable non-negative function  $f : S \rightarrow [0, \infty]$  such that  $\lim_{n \rightarrow \infty} f_n = f$  (almost everywhere). Scheffé's lemma for non-negative functions states that

$$\lim_{n \rightarrow \infty} \mu(|f_n - f|) = 0 \text{ if and only if } \lim_{n \rightarrow \infty} \mu(f_n) = \mu(f).$$

We will now show this lemma. First, suppose  $\lim_{n \rightarrow \infty} \mu(|f_n - f|) = 0$ . Since  $0 \leq |\mu(f_n - f)| \leq \mu(|f_n - f|)$ , the squeeze theorem implies that  $\lim_{n \rightarrow \infty} |\mu(f_n - f)| = 0$ , which also implies that  $\lim_{n \rightarrow \infty} \mu(f_n - f) = 0$ . By the linearity of the integral with respect to  $\mu$ , we conclude that  $\lim_{n \rightarrow \infty} \mu(f_n) = \mu(f)$ .

Now suppose  $\lim_{n \rightarrow \infty} \mu(f_n) = \mu(f)$  and consider the function  $(f_n - f)^- : S \rightarrow [0, \infty]$  given by

$$(f_n - f)^-(s) = \max(-(f_n - f)(s), 0) = \max((f - f_n)(s), 0) = (f - f_n)^+(s) = \begin{cases} f(s) - f_n(s), & \text{if } f(s) > f_n(s), \\ 0, & \text{if } f(s) \leq f_n(s). \end{cases}$$

Note that  $(f_n - f)^- \leq f$  for every  $n \in \mathbb{N}$ . Because  $\lim_{n \rightarrow \infty} f_n = f$ , we know that for every  $s \in S$  and  $\epsilon > 0$  there is an  $N \in \mathbb{N}$  such that  $n > N$  guarantees that  $|f(s) - f_n(s)| < \epsilon$ . Note that, for every  $n > N$ , if  $f(s) > f_n(s)$ , then  $|(f_n - f)^-(s)| = |f(s) - f_n(s)| < \epsilon$ . If  $f(s) \leq f_n(s)$ , then  $|(f_n - f)^-(s)| = 0 < \epsilon$ . Therefore, for every  $s \in S$  and  $\epsilon > 0$ , there is an  $N \in \mathbb{N}$  such that  $n > N$  guarantees that  $|(f_n - f)^-(s)| < \epsilon$ . By definition,  $\lim_{n \rightarrow \infty} (f_n - f)^- = 0$ , where 0 denotes the zero function.

Consider the sequence of  $\Sigma$ -measurable functions  $((f_n - f)^- : S \rightarrow \mathbb{R} \mid n \in \mathbb{N})$  and the  $\Sigma$ -measurable function  $0 : S \rightarrow \mathbb{R}$  such that  $\lim_{n \rightarrow \infty} (f_n - f)^- = 0$ . Furthermore, consider the  $\mu$ -integrable non-negative function  $f \in \mathcal{L}^1(S, \Sigma, \mu)^+$  such that  $|(f_n - f)^-| = (f_n - f)^- \leq f$  for every  $n \in \mathbb{N}$ . By the dominated convergence theorem, we know that  $\lim_{n \rightarrow \infty} \mu((f_n - f)^-) = \mu(0) = 0$ .

For every  $n \in \mathbb{N}$ , recall that  $(f_n - f)^+ = (f_n - f) + (f_n - f)^-$ . By the linearity of the integral with respect to  $\mu$ ,

$$\lim_{n \rightarrow \infty} \mu((f_n - f)^+) = \lim_{n \rightarrow \infty} \mu(f_n) - \mu(f) + \mu((f_n - f)^-) = \mu(f) - \mu(f) + \lim_{n \rightarrow \infty} \mu((f_n - f)^-) = 0.$$

For every  $n \in \mathbb{N}$ , recall that  $|f_n - f| = (f_n - f)^+ + (f_n - f)^-$ . By the linearity of the integral with respect to  $\mu$ ,

$$\lim_{n \rightarrow \infty} \mu(|f_n - f|) = \lim_{n \rightarrow \infty} \mu((f_n - f)^+) + \mu((f_n - f)^-) = 0.$$

Consider a sequence of  $\mu$ -integrable functions  $(f_n : S \rightarrow \mathbb{R} \mid n \in \mathbb{N})$  and a  $\mu$ -integrable function  $f : S \rightarrow \mathbb{R}$  such that  $\lim_{n \rightarrow \infty} f_n = f$  (almost everywhere). Scheffé's lemma states that

$$\lim_{n \rightarrow \infty} \mu(|f_n - f|) = 0 \text{ if and only if } \lim_{n \rightarrow \infty} \mu(|f_n|) = \mu(|f|).$$

We will now show this lemma. First, suppose  $\lim_{n \rightarrow \infty} \mu(|f_n - f|) = 0$ . By the triangle inequality,

$$\begin{aligned} |f_n| &= |(f_n - f) + f| \leq |f_n - f| + |f|, \\ |f| &= |(f - f_n) + f_n| \leq |f_n - f| + |f_n|. \end{aligned}$$

Because the integral with respect to  $\mu$  is non-decreasing and linear,

$$\begin{aligned} \mu(|f_n - f|) &\geq \mu(|f_n|) - \mu(|f|), \\ \mu(|f_n - f|) &\geq \mu(|f|) - \mu(|f_n|). \end{aligned}$$

Because  $\mu(|f_n - f|) \geq a$  and  $\mu(|f_n - f|) \geq -a$  for  $a = \mu(|f_n|) - \mu(|f|)$ ,

$$\mu(|f_n - f|) \geq |\mu(|f_n|) - \mu(|f|)| \geq 0. \quad (1)$$

By the squeeze theorem,  $\lim_{n \rightarrow \infty} |\mu(|f_n|) - \mu(|f|)| = 0$ , which implies  $\lim_{n \rightarrow \infty} \mu(|f_n|) - \mu(|f|) = 0$ . By the linearity of the integral with respect to  $\mu$ , we conclude that  $\lim_{n \rightarrow \infty} \mu(|f_n|) = \mu(|f|)$ .

Now suppose  $\lim_{n \rightarrow \infty} \mu(|f_n|) = \mu(|f|)$ . Because the function  $g : \mathbb{R} \rightarrow \mathbb{R}$  given by  $g(x) = \max(x, 0)$  is continuous,

$$\begin{aligned} \lim_{n \rightarrow \infty} f_n^+(s) &= \lim_{n \rightarrow \infty} \max(f_n(s), 0) = \max(f(s), 0) = f^+(s), \\ \lim_{n \rightarrow \infty} f_n^-(s) &= \lim_{n \rightarrow \infty} \max(-f_n(s), 0) = \max(-f(s), 0) = f^-(s). \end{aligned}$$

Because  $(f_n^+ : S \rightarrow [0, \infty] \mid n \in \mathbb{N})$  and  $(f_n^- : S \rightarrow [0, \infty] \mid n \in \mathbb{N})$  are sequences of  $\Sigma$ -measurable functions, the Fatou lemma guarantees that

$$\begin{aligned}\mu(f^+) &= \mu\left(\lim_{n \rightarrow \infty} f_n^+\right) = \mu\left(\liminf_{n \rightarrow \infty} f_n^+\right) \leq \liminf_{n \rightarrow \infty} \mu(f_n^+), \\ \mu(f^-) &= \mu\left(\lim_{n \rightarrow \infty} f_n^-\right) = \mu\left(\liminf_{n \rightarrow \infty} f_n^-\right) \leq \liminf_{n \rightarrow \infty} \mu(f_n^-).\end{aligned}$$

Consider the integrals  $\mu(f_n^+)$  and  $\mu(f_n^-)$  written as

$$\begin{aligned}\mu(f_n^+) &= \mu(f_n^+) + \mu(f_n^-) - \mu(f_n^-), \\ \mu(f_n^-) &= \mu(f_n^-) + \mu(f_n^+) - \mu(f_n^+).\end{aligned}$$

By taking the limit superior of both sides,

$$\begin{aligned}\limsup_{n \rightarrow \infty} \mu(f_n^+) &= \limsup_{n \rightarrow \infty} (\mu(f_n^+) + \mu(f_n^-) - \mu(f_n^-)), \\ \limsup_{n \rightarrow \infty} \mu(f_n^-) &= \limsup_{n \rightarrow \infty} (\mu(f_n^-) + \mu(f_n^+) - \mu(f_n^+)).\end{aligned}$$

By the subadditivity of the limit superior,

$$\begin{aligned}\limsup_{n \rightarrow \infty} \mu(f_n^+) &\leq \limsup_{n \rightarrow \infty} (\mu(f_n^+) + \mu(f_n^-)) + \limsup_{n \rightarrow \infty} -\mu(f_n^-) \\ \limsup_{n \rightarrow \infty} \mu(f_n^-) &\leq \limsup_{n \rightarrow \infty} (\mu(f_n^-) + \mu(f_n^+)) + \limsup_{n \rightarrow \infty} -\mu(f_n^+).\end{aligned}$$

From our assumption that  $\lim_{n \rightarrow \infty} \mu(|f_n|) = \mu(|f|)$ ,

$$\limsup_{n \rightarrow \infty} (\mu(f_n^+) + \mu(f_n^-)) = \limsup_{n \rightarrow \infty} (\mu(f_n^-) + \mu(f_n^+)) = \limsup_{n \rightarrow \infty} \mu(|f_n|) = \lim_{n \rightarrow \infty} \mu(|f_n|) = \mu(|f|).$$

Therefore, by the relationship between the limit inferior and the limit superior,

$$\begin{aligned}\limsup_{n \rightarrow \infty} \mu(f_n^+) &\leq \mu(|f|) - \liminf_{n \rightarrow \infty} \mu(f_n^-), \\ \limsup_{n \rightarrow \infty} \mu(f_n^-) &\leq \mu(|f|) - \liminf_{n \rightarrow \infty} \mu(f_n^+).\end{aligned}$$

By non-decreasing the right sides of the previous inequations using our previous result,

$$\begin{aligned}\limsup_{n \rightarrow \infty} \mu(f_n^+) &\leq \mu(|f|) - \mu(f^-) = \mu(f^+) + \mu(f^-) - \mu(f^-) = \mu(f^+), \\ \limsup_{n \rightarrow \infty} \mu(f_n^-) &\leq \mu(|f|) - \mu(f^+) = \mu(f^+) + \mu(f^-) - \mu(f^+) = \mu(f^-).\end{aligned}$$

By noting that the limit superior is at least as large as the limit inferior and combining the previous results,

$$\begin{aligned}\mu(f^+) &\leq \liminf_{n \rightarrow \infty} \mu(f_n^+) \leq \limsup_{n \rightarrow \infty} \mu(f_n^+) \leq \mu(f^+), \\ \mu(f^-) &\leq \liminf_{n \rightarrow \infty} \mu(f_n^-) \leq \limsup_{n \rightarrow \infty} \mu(f_n^-) \leq \mu(f^-).\end{aligned}$$

Because the previous inequations imply that the limits must match,

$$\begin{aligned}\lim_{n \rightarrow \infty} \mu(f_n^+) &= \mu(f^+), \\ \lim_{n \rightarrow \infty} \mu(f_n^-) &= \mu(f^-).\end{aligned}$$

Because  $(f_n^+ : S \rightarrow [0, \infty] \mid n \in \mathbb{N})$  and  $(f_n^- : S \rightarrow [0, \infty] \mid n \in \mathbb{N})$  are sequences of  $\mu$ -integrable non-negative functions and  $f^+ : S \rightarrow [0, \infty]$  and  $f^- : S \rightarrow [0, \infty]$  are  $\mu$ -integrable non-negative functions such that  $\lim_{n \rightarrow \infty} f_n^+ = f^+$  and  $\lim_{n \rightarrow \infty} f_n^- = f^-$ , Scheffé's lemma for non-negative functions guarantees that

$$\begin{aligned}\lim_{n \rightarrow \infty} \mu(|f_n^+ - f^+|) &= 0, \\ \lim_{n \rightarrow \infty} \mu(|f_n^- - f^-|) &= 0.\end{aligned}$$

By the triangle inequality,

$$|f_n - f| = |(f_n^+ - f_n^-) - (f^+ - f^-)| = |(f_n^+ - f^+) + (f^- - f_n^-)| \leq |f_n^+ - f^+| + |f_n^- - f^-|.$$

Because the integral with respect to  $\mu$  is non-negative for non-negative functions, non-decreasing, and linear,

$$0 \leq \mu(|f_n - f|) \leq \mu(|f_n^+ - f^+|) + \mu(|f_n^- - f^-|).$$

By the squeeze theorem, and as we wanted to show,

$$\lim_{n \rightarrow \infty} \mu(|f_n - f|) = 0.$$

Consider the measure space  $(S, \Sigma, \mu)$ . For a set  $A \in \Sigma$ , consider the triple  $(A, \Sigma_A, \mu_A)$  such that  $\Sigma_A = \{B \in \Sigma \mid B \subseteq A\}$  and  $\mu_A(B) = \mu(B)$  for every  $B \in \Sigma_A$ . We will now show that  $(A, \Sigma_A, \mu_A)$  is a measure space restricted to  $A$ .

First, we will show that  $\Sigma_A$  is a  $\sigma$ -algebra on  $A$ . Because  $A \in \Sigma$  and  $A \subseteq A$ , we have  $A \in \Sigma_A$ . If  $B \in \Sigma_A$ , then  $B \in \Sigma$  and  $A \cap B^c \in \Sigma$ . Because  $A \cap B^c \subseteq A$ , we have  $A \setminus B \in \Sigma_A$ . For any sequence  $(B_n \in \Sigma_A \mid n \in \mathbb{N})$ , the fact that  $B_n \in \Sigma$  guarantees that  $\cup_n B_n \in \Sigma$ . Because  $B_n \subseteq A$  for every  $n \in \mathbb{N}$ , we know that  $\cup_n B_n \subseteq A$ , which implies  $\cup_n B_n \in \Sigma_A$ .

Second, we will show that the non-negative function  $\mu_A : \Sigma_A \rightarrow [0, \infty]$  is a measure on the measurable space  $(A, \Sigma_A)$ . Because  $\emptyset \in \Sigma$  and  $\emptyset \in \Sigma_A$ , we know that  $\mu_A(\emptyset) = \mu(\emptyset) = 0$ . For any sequence  $(B_n \in \Sigma_A \mid n \in \mathbb{N})$  such that  $B_n \cap B_m = \emptyset$  for every  $n \neq m$ , we know that  $\cup_n B_n \in \Sigma$  and  $\cup_n B_n \in \Sigma_A$  and

$$\mu_A\left(\bigcup_n B_n\right) = \mu\left(\bigcup_n B_n\right) = \sum_n \mu(B_n) = \sum_n \mu_A(B_n).$$

Consider the measure space  $(S, \Sigma, \mu)$  and a  $\Sigma$ -measurable function  $f : S \rightarrow \mathbb{R}$ . Consider also the measure space  $(A, \Sigma_A, \mu_A)$  restricted to  $A \in \Sigma$  and the function  $f|_A : A \rightarrow \mathbb{R}$  restricted to  $A$  given by  $f|_A(a) = f(a)$  for every  $a \in A$ . The function  $f|_A$  is  $\Sigma_A$ -measurable because, for every  $B \in \mathcal{B}(\mathbb{R})$ ,

$$(f|_A)^{-1}(B) = \{a \in A \mid f(a) \in B\} = \{s \in S \mid f(s) \in B\} \cap A = f^{-1}(B) \cap A.$$

Consider the measure space  $(S, \Sigma, \mu)$ , a  $\Sigma$ -measurable function  $f : S \rightarrow \mathbb{R}$ , and a set  $A \in \Sigma$ . We will now show that  $f|_A$  is  $\mu_A$ -integrable if and only if  $f\mathbb{I}_A$  is  $\mu$ -integrable, in which case  $\mu_A(f|_A) = \mu(f\mathbb{I}_A) = \mu(f; A)$ .

First, suppose  $f = \mathbb{I}_B$  for some set  $B \in \Sigma$ . Clearly,  $\mu(f\mathbb{I}_A) = \mu(\mathbb{I}_B\mathbb{I}_A) = \mu(\mathbb{I}_{B \cap A}) = \mu(B \cap A)$  and  $\mu_A(f|_A) = \mu_A(\mathbb{I}_B|_A) = \mu_A(\mathbb{I}_{B \cap A}) = \mu_A(B \cap A)$ . Because  $B \cap A \subseteq A$ , we have  $\mu_A(B \cap A) = \mu(B \cap A)$ , which implies  $\mu_A(f|_A) = \mu(f\mathbb{I}_A)$ . Because  $\mu(|f\mathbb{I}_A|) = \mu(f\mathbb{I}_A) = \mu_A(f|_A) = \mu_A(|f|_A)$ , we know that  $f|_A$  is  $\mu_A$ -integrable if and only if  $f\mathbb{I}_A$  is  $\mu$ -integrable.

Next, suppose  $f$  is a simple function that can be written as  $f = \sum_{k=1}^m a_k \mathbb{I}_{A_k}$  for some fixed  $a_1, a_2, \dots, a_m \in [0, \infty]$  and  $A_1, A_2, \dots, A_m \in \Sigma$ . In that case, the integral with respect to  $\mu$  of the function  $f\mathbb{I}_A$  is given by

$$\mu(f\mathbb{I}_A) = \mu\left(\sum_{k=1}^m a_k \mathbb{I}_{A_k} \mathbb{I}_A\right) = \mu\left(\sum_{k=1}^m a_k \mathbb{I}_{A_k \cap A}\right) = \sum_{k=1}^m a_k \mu(A_k \cap A).$$

Furthermore, the integral of the function  $f|_A$  with respect to  $\mu_A$  is given by

$$\mu_A(f|_A) = \mu_A\left(\left(\sum_{k=1}^m a_k \mathbb{I}_{A_k}\right)\Big|_A\right) = \mu_A\left(\sum_{k=1}^m a_k \mathbb{I}_{A_k \cap A}\right) = \sum_{k=1}^m a_k \mu_A(\mathbb{I}_{A_k \cap A}) = \sum_{k=1}^m a_k \mu_A(A_k \cap A).$$

Because  $A_k \cap A \subseteq A$  for every  $k \leq m$ , we have  $\mu_A(A_k \cap A) = \mu(A_k \cap A)$ , which implies  $\mu_A(f|_A) = \mu(f\mathbb{I}_A)$ . Because  $\mu(|f\mathbb{I}_A|) = \mu(f\mathbb{I}_A) = \mu_A(f|_A) = \mu_A(|f|_A)$ , we know that  $f|_A$  is  $\mu_A$ -integrable if and only if  $f\mathbb{I}_A$  is  $\mu$ -integrable.

Next, suppose  $f$  is non-negative. For any  $n \in \mathbb{N}$ , let  $f_n = \alpha_n \circ f$ , where  $\alpha_n$  is the  $n$ -th staircase function. Because  $(f_n\mathbb{I}_A \mid n \in \mathbb{N})$  is a sequence of  $\Sigma$ -measurable functions such that  $f_n\mathbb{I}_A \uparrow f\mathbb{I}_A$ , we know that  $\mu(f_n\mathbb{I}_A) \uparrow \mu(f\mathbb{I}_A)$ . Because  $(f_n|_A \mid n \in \mathbb{N})$  is a sequence of  $\Sigma_A$ -measurable functions such that  $f_n|_A \uparrow f|_A$ , we know that  $\mu_A(f_n|_A) \uparrow \mu_A(f|_A)$ . For every  $n \in \mathbb{N}$ , the fact that  $f_n$  is a simple function implies  $\mu(f_n\mathbb{I}_A) = \mu_A(f_n|_A)$ . Therefore,  $\mu_A(f_n|_A) \uparrow \mu(f\mathbb{I}_A)$ , and  $\mu(f_n\mathbb{I}_A) \uparrow \mu_A(f|_A)$ , and  $\mu(f\mathbb{I}_A) = \mu_A(f|_A)$ . Because  $\mu(|f\mathbb{I}_A|) = \mu(f\mathbb{I}_A) = \mu_A(f|_A) = \mu_A(|f|_A)$ , we know that  $f|_A$  is  $\mu_A$ -integrable if and only if  $f\mathbb{I}_A$  is  $\mu$ -integrable.

Finally, suppose  $f : S \rightarrow \mathbb{R}$ . By definition,

$$\mu(|f\mathbb{I}_A|) = \mu((f\mathbb{I}_A)^+) + \mu((f\mathbb{I}_A)^-) = \mu(f^+\mathbb{I}_A) + \mu(f^-\mathbb{I}_A) = \mu_A(f^+|_A) + \mu_A(f^-|_A) = \mu_A((f|_A)^+) + \mu_A((f|_A)^-) = \mu(|f|_A).$$

Therefore,  $f|_A$  is  $\mu_A$ -integrable if and only if  $f\mathbb{I}_A$  is  $\mu$ -integrable. In that case,

$$\mu(f\mathbb{I}_A) = \mu((f\mathbb{I}_A)^+) - \mu((f\mathbb{I}_A)^-) = \mu(f^+\mathbb{I}_A) - \mu(f^-\mathbb{I}_A) = \mu_A(f^+|_A) - \mu_A(f^-|_A) = \mu_A((f|_A)^+) - \mu_A((f|_A)^-) = \mu(f|_A).$$

Consider a  $\Sigma$ -measurable function  $f : S \rightarrow [0, \infty]$  and the function  $(f\mu) : \Sigma \rightarrow [0, \infty]$  defined by

$$(f\mu)(A) = \mu(f; A) = \mu(f\mathbb{I}_A) = \mu_A(f|_A).$$

We will now show that  $(f\mu)$  is a measure on  $(S, \Sigma)$ . Clearly,  $(f\mu)(\emptyset) = \mu(f\mathbb{I}_\emptyset) = \mu(0) = 0$ .

Consider a sequence  $(B_n \in \Sigma \mid n \in \mathbb{N})$  such that  $B_n \cap B_m = \emptyset$  for  $n \neq m$ . First, suppose  $f$  is a simple function that can be written as  $f = \sum_{k=1}^m a_k \mathbb{I}_{A_k}$  for some fixed  $a_1, a_2, \dots, a_m \in [0, \infty]$  and  $A_1, A_2, \dots, A_m \in \Sigma$ . In that case,

$$(f\mu)(\cup_n B_n) = \mu(f\mathbb{I}_{\cup_n B_n}) = \mu\left(\sum_{k=1}^m a_k \mathbb{I}_{A_k} \mathbb{I}_{\cup_n B_n}\right) = \mu\left(\sum_{k=1}^m a_k \mathbb{I}_{A_k \cap (\cup_n B_n)}\right) = \mu\left(\sum_{k=1}^m a_k \mathbb{I}_{\cup_n (A_k \cap B_n)}\right).$$

By the definition of integral with respect to  $\mu$  of a simple function and countable additivity,

$$(f\mu)(\cup_n B_n) = \sum_{k=1}^m a_k \mu(\cup_n (A_k \cap B_n)) = \sum_{k=1}^m a_k \sum_n \mu(A_k \cap B_n) = \sum_n \sum_{k=1}^m a_k \mu(A_k \cap B_n).$$

By the definition of integral with respect to  $\mu$  of a simple function,

$$(f\mu)(\cup_n B_n) = \sum_n \mu\left(\sum_{k=1}^m a_k \mathbb{I}_{A_k \cap B_n}\right) = \sum_n \mu\left(\sum_{k=1}^m a_k \mathbb{I}_{A_k} \mathbb{I}_{B_n}\right) = \sum_n \mu(f\mathbb{I}_{B_n}) = \sum_n (f\mu)(B_n).$$

Now suppose  $f$  is non-negative. For any  $n \in \mathbb{N}$ , let  $f_n = \alpha_n \circ f$ , where  $\alpha_n$  is the  $n$ -th staircase function. For every set  $B \in \Sigma$ , we know that  $(f_n \mathbb{I}_B : S \rightarrow [0, \infty] \mid n \in \mathbb{N})$  is a sequence of  $\Sigma$ -measurable functions such that  $f_n \mathbb{I}_B \uparrow f \mathbb{I}_B$ , which implies that  $\mu(f_n \mathbb{I}_B) \uparrow \mu(f \mathbb{I}_B)$ . Therefore,

$$(f\mu)(\cup_j B_j) = \mu(f\mathbb{I}_{\cup_j B_j}) = \lim_{n \rightarrow \infty} \mu(f_n \mathbb{I}_{\cup_j B_j}) = \lim_{n \rightarrow \infty} \sum_j \mu(f_n \mathbb{I}_{B_j}) = \sum_j \lim_{n \rightarrow \infty} \mu(f_n \mathbb{I}_{B_j}) = \sum_j \mu(f \mathbb{I}_{B_j}) = \sum_j (f\mu)(B_j).$$

Consider a  $\Sigma$ -measurable function  $f : S \rightarrow [0, \infty]$  and the measure space  $(S, \Sigma, (f\mu))$ . By definition, the integral with respect to  $(f\mu)$  of a  $\Sigma$ -measurable function  $h : S \rightarrow \mathbb{R}$  over the set  $A$  is given by

$$(f\mu)(h\mathbb{I}_A) = (f\mu)(h; A) = (h(f\mu))(A).$$

We will now show that  $(f\mu)(h\mathbb{I}_A) = \mu(fh\mathbb{I}_A)$ .

First, suppose  $h = \mathbb{I}_B$  for some set  $B \in \Sigma$ . In that case, the integral with respect to  $(f\mu)$  of  $h$  over the set  $A$  is given by

$$(f\mu)(h\mathbb{I}_A) = (f\mu)(\mathbb{I}_B \mathbb{I}_A) = (f\mu)(\mathbb{I}_{B \cap A}) = (f\mu)(B \cap A) = \mu(f\mathbb{I}_{B \cap A}) = \mu(f\mathbb{I}_B \mathbb{I}_A) = \mu(fh\mathbb{I}_A).$$

Next, suppose  $h$  is a simple function that can be written as  $h = \sum_{k=1}^m a_k \mathbb{I}_{A_k}$  for some fixed  $a_1, a_2, \dots, a_m \in [0, \infty]$  and  $A_1, A_2, \dots, A_m \in \Sigma$ . In that case, the integral with respect to  $(f\mu)$  of  $h$  over the set  $A$  is given by

$$(f\mu)(h\mathbb{I}_A) = (f\mu)\left(\sum_{k=1}^m a_k \mathbb{I}_{A_k} \mathbb{I}_A\right) = (f\mu)\left(\sum_{k=1}^m a_k \mathbb{I}_{A_k \cap A}\right) = \sum_{k=1}^m a_k (f\mu)(A_k \cap A) = \sum_{k=1}^m a_k \mu(f\mathbb{I}_{A_k \cap A}).$$

By the linearity of the integral with respect to  $\mu$ ,

$$(f\mu)(h\mathbb{I}_A) = \mu\left(\sum_{k=1}^m a_k f\mathbb{I}_{A_k \cap A}\right) = \mu\left(f\mathbb{I}_A \sum_{k=1}^m a_k \mathbb{I}_{A_k}\right) = \mu(fh\mathbb{I}_A).$$

Next, suppose  $h$  is non-negative. For any  $n \in \mathbb{N}$ , let  $h_n = \alpha_n \circ h$ , where  $\alpha_n$  is the  $n$ -th staircase function. Because  $(h_n \mathbb{I}_A : S \rightarrow [0, \infty] \mid n \in \mathbb{N})$  is a sequence of  $\Sigma$ -measurable functions such that  $h_n \mathbb{I}_A \uparrow h \mathbb{I}_A$ , we know

that  $(f\mu)(h_n\mathbb{I}_A) \uparrow (f\mu)(h\mathbb{I}_A)$ . Furthermore, because  $(fh_n\mathbb{I}_A : S \rightarrow [0, \infty] \mid n \in \mathbb{N})$  is a sequence of  $\Sigma$ -measurable functions such that  $fh_n\mathbb{I}_A \uparrow fh\mathbb{I}_A$ , we know that  $\mu(fh_n\mathbb{I}_A) \uparrow \mu(fh\mathbb{I}_A)$ . Therefore, the integral with respect to  $(f\mu)$  of  $h$  over the set  $A$  is given by

$$(f\mu)(h\mathbb{I}_A) = \lim_{n \rightarrow \infty} (f\mu)(h_n\mathbb{I}_A) = \lim_{n \rightarrow \infty} \mu(fh_n\mathbb{I}_A) = \mu(fh\mathbb{I}_A).$$

Finally, suppose  $h : S \rightarrow \mathbb{R}$ . By definition,

$$(f\mu)(|h\mathbb{I}_A|) = (f\mu)((h\mathbb{I}_A)^+) + (f\mu)((h\mathbb{I}_A)^-) = (f\mu)(h^+\mathbb{I}_A) + (f\mu)(h^-\mathbb{I}_A) = \mu(fh^+\mathbb{I}_A) + \mu(fh^-\mathbb{I}_A)$$

By the linearity of the integral with respect to  $\mu$ ,

$$(f\mu)(|h\mathbb{I}_A|) = \mu(fh^+\mathbb{I}_A + fh^-\mathbb{I}_A) = \mu(f\mathbb{I}_A(h^+ + h^-)) = \mu(f|h\mathbb{I}_A|) = \mu(|fh\mathbb{I}_A|).$$

Therefore,  $h\mathbb{I}_A$  is  $(f\mu)$ -integrable if and only if  $fh\mathbb{I}_A$  is  $\mu$ -integrable. In that case,

$$(f\mu)(h\mathbb{I}_A) = (f\mu)((h\mathbb{I}_A)^+) - (f\mu)((h\mathbb{I}_A)^-) = (f\mu)(h^+\mathbb{I}_A) - (f\mu)(h^-\mathbb{I}_A) = \mu(fh^+\mathbb{I}_A) - \mu(fh^-\mathbb{I}_A)$$

By the linearity of the integral with respect to  $\mu$ ,

$$(f\mu)(h\mathbb{I}_A) = \mu(fh^+\mathbb{I}_A - fh^-\mathbb{I}_A) = \mu(f\mathbb{I}_A(h^+ - h^-)) = \mu(fh\mathbb{I}_A).$$

Therefore, by considering integrals over the set  $S$ , if  $f : S \rightarrow [0, \infty]$  and  $h : S \rightarrow \mathbb{R}$  are  $\Sigma$ -measurable functions, then  $h$  is  $(f\mu)$ -measurable if and only if  $fh$  is  $\mu$ -measurable, in which case  $(f\mu)(h) = \mu(fh)$ .

Consider a measure space  $(S, \Sigma, \mu)$ , a  $\Sigma$ -measurable function  $f : S \rightarrow [0, \infty]$ , and the measure  $\lambda = (f\mu)$  on  $(S, \Sigma)$ . We say that  $\lambda$  has density  $f$  relative to  $\mu$ , which is denoted by  $d\lambda/d\mu = f$ .

For every  $A \in \Sigma$ , if  $\mu(A) = 0$ , we will now show that  $\lambda(A) = (f\mu)(A) = \mu(f\mathbb{I}_A) = 0$ . The fact that  $\{f\mathbb{I}_A \neq 0\} \subseteq A$  implies  $\mu(\{f\mathbb{I}_A \neq 0\}) \leq \mu(A) = 0$ . Because  $f\mathbb{I}_A$  and 0 are  $\Sigma$ -measurable functions such that  $\mu(\{f\mathbb{I}_A \neq 0\}) = 0$ , we know that  $\mu(f\mathbb{I}_A) = \mu(0) = 0$ .

If  $\mu$  and  $\lambda$  are  $\sigma$ -finite measures on  $(S, \Sigma)$  such that if  $\mu(A) = 0$  then  $\lambda(A) = 0$  for every  $A \in \Sigma$ , the Radon-Nykodým theorem states that  $\lambda = (f\mu)$  for some  $\Sigma$ -measurable function  $f : S \rightarrow [0, \infty]$ .

## 6 Expectation

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ . The expectation  $\mathbb{E}(X)$  of a  $\mathbb{P}$ -integrable random variable  $X : \Omega \rightarrow \mathbb{R}$  is defined as the integral of  $X$  with respect to the probability measure  $\mathbb{P}$ . Therefore,

$$\mathbb{E}(X) = \mathbb{P}(X) = \int_{\Omega} X d\mathbb{P} = \int_{\Omega} X(\omega) \mathbb{P}(d\omega).$$

The expectation  $\mathbb{E}(X)$  of a non-negative random variable  $X : \Omega \rightarrow [0, \infty]$  is also defined as the integral of  $X$  with respect to the probability measure  $\mathbb{P}$ .

Consider a sequence of random variables  $(X_n : \Omega \rightarrow \mathbb{R} \mid n \in \mathbb{N})$  and a random variable  $X : \Omega \rightarrow \mathbb{R}$  such that

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_n = X\right) = \mathbb{P}\left(\{\omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}\right) = 1.$$

The integration results discussed in the previous section can be restated as follows:

- By the monotone-convergence theorem, if  $X_n \geq 0$  and  $X_n \leq X_{n+1}$  for every  $n \in \mathbb{N}$ , then  $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$ .
- By the Fatou lemma, if  $X_n \geq 0$  for every  $n \in \mathbb{N}$ , then  $\mathbb{E}(X) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n)$ .
- By the dominated convergence theorem, if there is a  $\mathbb{P}$ -integrable non-negative function  $Y : \Omega \rightarrow [0, \infty]$  such that  $|X_n| \leq Y$  for every  $n \in \mathbb{N}$ , then  $X$  is  $\mathbb{P}$ -integrable and  $\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(X)$ .
- By Scheffé's lemma, if  $X$  and  $X_n$  are  $\mathbb{P}$ -integrable for every  $n \in \mathbb{N}$ , then  $\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|) = 0$  if and only if  $\lim_{n \rightarrow \infty} \mathbb{E}(|X_n|) = \mathbb{E}(|X|)$ .

As a special case of the dominated convergence theorem, the bounded convergence theorem guarantees that if there is a  $K \in [0, \infty)$  such that  $|X_n| \leq K$  for every  $n \in \mathbb{N}$ , then  $\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|) = 0$ . Note that the simple function  $Y = K$  is  $\mathbb{P}$ -integrable, since  $\mathbb{P}(|Y|) = \mathbb{P}(Y) = \mathbb{P}(K\mathbb{I}_\Omega) = K\mathbb{P}(\Omega) = K$ . Therefore,  $X$  is  $\mathbb{P}$ -integrable and  $\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(X)$ . The dominated convergence theorem also guarantees that  $\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|) = 0$ .

The expectation  $\mathbb{E}(X; F)$  of the  $\mathbb{P}$ -integrable random variable  $X : \Omega \rightarrow \mathbb{R}$  over the set  $F \in \mathcal{F}$  is defined as

$$\mathbb{E}(X; F) = \mathbb{E}(X\mathbb{I}_F) = \mathbb{P}(X; F) = \mathbb{P}(X\mathbb{I}_F) = \int_F X d\mathbb{P} = \int_F X(\omega) \mathbb{P}(d\omega).$$

Consider a random variable  $Z : \Omega \rightarrow \mathbb{R}$  and a  $\mathcal{B}(\mathbb{R})$ -measurable non-negative function  $g : \mathbb{R} \rightarrow [0, \infty]$  such that  $a \leq b$  implies  $g(a) \leq g(b)$ . Recall that the function  $g(Z) : \Omega \rightarrow [0, \infty]$  defined by  $g(Z) = g \circ Z$  is also a random variable. For every  $c \in \mathbb{R}$ , Markov's inequality states that

$$\mathbb{E}(g(Z)) \geq g(c)\mathbb{P}(Z \geq c),$$

since  $g(Z) \geq g(Z)\mathbb{I}_{\{Z \geq c\}} \geq g(c)\mathbb{I}_{\{Z \geq c\}}$  implies  $\mathbb{E}(g(Z)) \geq \mathbb{E}(g(c)\mathbb{I}_{\{Z \geq c\}}) = g(c)\mathbb{P}(Z \geq c)$ .

Consider a non-negative random variable  $Z : \Omega \rightarrow [0, \infty]$  and let  $g(c) = \max(c, 0)$ . For  $c \geq 0$ , Markov's inequality implies that  $\mathbb{E}(Z) \geq c\mathbb{P}(Z \geq c)$ .

Consider a random variable  $Z : \Omega \rightarrow \mathbb{R}$  and let  $g(c) = e^{\theta c}$  for some  $\theta > 0$ . Markov's inequality implies that  $\mathbb{E}(e^{\theta Z}) \geq e^{\theta c}\mathbb{P}(Z \geq c)$ .

Consider a non-negative random variable  $X : \Omega \rightarrow [0, \infty]$ . If  $\mathbb{E}(X) < \infty$ , then  $\mathbb{P}(X < \infty) = 1$ . Note that  $\infty\mathbb{I}_{\{X=\infty\}} \leq X$ , such that  $\infty\mathbb{P}(X = \infty) \leq \mathbb{E}(X)$ . Therefore,  $\mathbb{P}(X = \infty) > 0$  implies  $\mathbb{E}[X] = \infty$ .

Consider a sequence  $(Z_n : \Omega \rightarrow [0, \infty] \mid n \in \mathbb{N})$  of non-negative random variables. We will now show that

$$\mathbb{E}\left(\sum_k Z_k\right) = \sum_k \mathbb{E}(Z_k).$$

For any  $n \in \mathbb{N}$ , let  $Y_n = \sum_{k=0}^n Z_k$ , such that  $\mathbb{E}(Y_n) = \sum_{k=0}^n \mathbb{E}(Z_k)$ . Clearly,  $Y_n \geq 0$ ,  $Y_n \leq Y_{n+1}$ , and  $\lim_{n \rightarrow \infty} Y_n = \sum_k Z_k$ . Therefore,  $Y_n \uparrow \sum_k Z_k$ . By the monotone-convergence theorem,  $\mathbb{E}(Y_n) \uparrow \mathbb{E}(\sum_k Z_k)$ .

Consider a sequence  $(Z_n : \Omega \rightarrow [0, \infty] \mid n \in \mathbb{N})$  of non-negative random variables such that  $\sum_k \mathbb{E}(Z_k) < \infty$ . We will now show that  $\sum_k Z_k < \infty$  almost surely and  $\lim_{n \rightarrow \infty} Z_n = 0$  almost surely, where 0 denotes the zero function. Because  $\mathbb{E}(\sum_k Z_k) < \infty$ , we know that  $\mathbb{P}(\sum_k Z_k < \infty) = 1$ . Because the  $n$ -th term test implies that  $\{\sum_k Z_k < \infty\} \subseteq \{\lim_{n \rightarrow \infty} Z_n = 0\}$ , we know that  $1 = \mathbb{P}(\sum_k Z_k < \infty) \leq \mathbb{P}(\lim_{n \rightarrow \infty} Z_n = 0)$ .

Consider a sequence of events  $(F_n \in \mathcal{F} \mid n \in \mathbb{N})$  such that  $\sum_n \mathbb{P}(F_n) < \infty$ . Let  $(\mathbb{I}_{F_n} \mid n \in \mathbb{N})$  be the corresponding sequence of indicator functions. Because  $\mathbb{E}(\mathbb{I}_{F_k}) = \mathbb{P}(F_k)$ , we know that  $\sum_n \mathbb{E}(\mathbb{I}_{F_n}) < \infty$ , which implies  $\sum_n \mathbb{I}_{F_n} < \infty$  almost surely. Because  $\sum_n \mathbb{I}_{F_n}(\omega)$  is the number of times that the outcome  $\omega \in \Omega$  belongs to an event in the sequence, we know that the outcome  $\omega$  almost surely belongs to a finite number of events in the sequence, which implies that  $\mathbb{P}(\limsup_{n \rightarrow \infty} F_n) = 0$ . This is the Borel-Cantelli lemma.

A function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is convex if  $\lambda\phi(x) + (1-\lambda)\phi(y) \geq \phi(\lambda x + (1-\lambda)y)$ , for every  $x \in \mathbb{R}$ ,  $y \in \mathbb{R}$ , and  $\lambda \in [0, 1]$ . If  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is convex, it is also continuous and therefore  $\mathcal{B}(\mathbb{R})$ -measurable. Important examples of convex functions include  $x \mapsto |x|$ ,  $x \mapsto x^2$ , and  $x \mapsto e^{\theta x}$  for  $\theta \in \mathbb{R}$ .

If  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is a convex function, for every  $z \in \mathbb{R}$  there is a function  $g : \mathbb{R} \rightarrow \mathbb{R}$  given by  $g(x) = ax + b$  for every  $x \in \mathbb{R}$  and some  $a \in \mathbb{R}$  and  $b \in \mathbb{R}$  such that  $g(z) = \phi(z)$  and  $g(x) \leq \phi(x)$  for every  $x \in \mathbb{R}$ . In other words, for every point in the domain of a convex function, there is a linear function that never surpasses the convex function such that the value of the linear function at that point matches the value of the convex function at that point.

Consider a random variable  $X : \Omega \rightarrow \mathbb{R}$  such that  $\mathbb{E}(X) < \infty$  and a convex function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ . Jensen's inequality states that  $\mathbb{E}(\phi(X)) \geq \phi(\mathbb{E}(X))$ . We will now show this inequality.

Consider a function  $g : \mathbb{R} \rightarrow \mathbb{R}$  such that  $g(\mathbb{E}(X)) = \phi(\mathbb{E}(X))$  and  $g(x) = ax + b \leq \phi(x)$  for every  $x \in \mathbb{R}$  and some  $a, b \in \mathbb{R}$ . Clearly  $g(X) = g \circ X \leq \phi \circ X = \phi(X)$ . Therefore,

$$\mathbb{E}(\phi(X)) \geq \mathbb{E}(g(X)) = \mathbb{E}[aX + b] = a\mathbb{E}(X) + b = g(\mathbb{E}(X)) = \phi(\mathbb{E}(X)).$$

For every  $p \in [1, \infty)$ , the set  $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$  contains exactly each random variable  $X : \Omega \rightarrow \mathbb{R}$  such that  $\mathbb{E}(|X|^p) < \infty$ . If  $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ , the  $p$ -norm  $\|X\|_p$  of the random variable  $X$  is given by  $\|X\|_p = \mathbb{E}(|X|^p)^{1/p}$ .

For every  $p \in [1, \infty)$  and  $r \in [1, \infty)$  such that  $p \leq r$ , we will now show that if  $Y \in \mathcal{L}^r(\Omega, \mathcal{F}, \mathbb{P})$  then  $Y \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$  and  $\|Y\|_p \leq \|Y\|_r$ . For every  $n \in \mathbb{N}$ , consider the function  $X_n = \min(|Y|, n)^p$ . Clearly,  $0 \leq X_n \leq n^p$ , so  $0 \leq \mathbb{E}(|X_n|) = \mathbb{E}(X_n) \leq n^p$ . Consider also the convex function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  given by  $\phi(x) = |x|^{r/p}$  such that



$\phi(X_n) = |X_n|^{r/p} = X_n^{r/p}$ . Clearly,  $0 \leq X_n^{r/p} = \min(|Y|, n)^r \leq n^r$ , so  $0 \leq \mathbb{E}(|X_n^{r/p}|) = \mathbb{E}(X_n^{r/p}) \leq n^r$ . Using Jensen's inequality,

$$\mathbb{E}(X_n^{r/p}) = \mathbb{E}(\phi(X_n)) \geq \phi(\mathbb{E}(X_n)) = |\mathbb{E}(X_n)|^{r/p} = \mathbb{E}(X_n)^{r/p}.$$

Because  $X_n^{r/p} \geq 0$  and  $X_n^{r/p} \uparrow |Y|^r$ , the monotone-convergence theorem guarantees that  $\mathbb{E}(X_n^{r/p}) \uparrow \mathbb{E}(|Y|^r)$ . Because  $X_n \geq 0$  and  $X_n \uparrow |Y|^p$ , the monotone-convergence theorem guarantees that  $\mathbb{E}(X_n) \uparrow \mathbb{E}(|Y|^p)$ . By taking the limit of both sides of the previous inequation,

$$\mathbb{E}(|Y|^r) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n^{r/p}) \geq \lim_{n \rightarrow \infty} \mathbb{E}(X_n)^{r/p} = \left( \lim_{n \rightarrow \infty} \mathbb{E}(X_n) \right)^{r/p} = \mathbb{E}(|Y|^p)^{r/p}.$$

By taking the  $r$ -th root of both sides of the previous inequation,

$$\infty > \mathbb{E}(|Y|^r)^{1/r} \geq \mathbb{E}(|Y|^p)^{1/p}.$$

For every  $p \in [1, \infty)$ , we will now show that  $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$  is a vector space over the field  $\mathbb{R}$ . First, recall that the set of all functions from  $\Omega$  to  $\mathbb{R}$  is a vector space over the field  $\mathbb{R}$  when scalar multiplication and addition are performed pointwise. Because such set includes  $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ , it is sufficient to show that  $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$  is non-empty and closed under scalar multiplication and addition. Because  $0 : \Omega \rightarrow \mathbb{R}$  is a random variable and  $\mathbb{E}(|0|^p) = \mathbb{E}(0) = 0$ , we know that  $0 \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ . If  $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$  and  $c \in \mathbb{R}$ , then  $cX : \Omega \rightarrow \mathbb{R}$  is a random variable and  $\mathbb{E}(|cX|^p) = \mathbb{E}(|c|^p |X|^p) = |c|^p \mathbb{E}(|X|^p)$ , we know that  $cX \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ . Finally, if  $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$  and  $Y \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ , then

$$|X + Y|^p \leq (|X| + |Y|)^p \leq (2 \max(|X|, |Y|))^p \leq 2^p(|X|^p + |Y|^p),$$

which implies  $X + Y \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$  since

$$\mathbb{E}(|X + Y|^p) \leq \mathbb{E}(2^p(|X|^p + |Y|^p)) = 2^p \mathbb{E}(|X|^p) + 2^p \mathbb{E}(|Y|^p) < \infty.$$

Consider the random variables  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . The Schwarz inequality states that  $XY \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and  $\mathbb{E}(|XY|) \leq \|X\|_2 \|Y\|_2$ . We will now show this inequality.

First, consider the case where  $\|X\|_2 \neq 0$  and  $\|Y\|_2 \neq 0$ . Let  $Z = |X|/\|X\|_2$  and  $W = |Y|/\|Y\|_2$ . Clearly,  $\mathbb{E}(Z^2) = \mathbb{E}(|X|^2)/\|X\|_2^2 = 1$ . Analogously,  $\mathbb{E}(W^2) = 1$ . Because  $(Z - W)^2 \geq 0$ , we know that

$$0 \leq \mathbb{E}((Z - W)^2) = \mathbb{E}(Z^2) + \mathbb{E}(W^2) - \mathbb{E}(2ZW) = 2 - \mathbb{E}(2ZW).$$

Because the previous inequation implies that  $\mathbb{E}(ZW) \leq 1$ ,

$$1 \geq \mathbb{E}(ZW) = \mathbb{E}(|X||Y|/\|X\|_2\|Y\|_2) = \mathbb{E}(|XY|)/\|X\|_2\|Y\|_2.$$

Using the fact that  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ ,

$$\mathbb{E}(|XY|) \leq \|X\|_2 \|Y\|_2 < \infty.$$

Finally, consider the case where  $\|X\|_2 = \mathbb{E}(X^2)^{1/2} = 0$ , which will prove analogous to the case where  $\|Y\|_2 = 0$ . Because  $X^2$  is a non-negative random variable, the fact that  $\mathbb{E}(X^2) = 0$  implies that  $\mathbb{P}(X^2 > 0) = \mathbb{P}(X \neq 0) = 0$ . Therefore,  $\mathbb{P}(X = 0) = 1$ . Because  $\{X = 0\} \subseteq \{XY = 0\}$ , we know that  $\mathbb{P}(X = 0) \leq \mathbb{P}(XY = 0)$ , which implies  $\mathbb{P}(XY = 0) = \mathbb{P}(|XY| = 0) = 1$ . Because  $\{|XY| = 0\}$  happens almost surely, we know that  $\mathbb{E}(|XY|) = \mathbb{E}(0) = 0$ . Therefore,  $XY \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and  $0 = \mathbb{E}(|XY|) \leq \|X\|_2 \|Y\|_2 = 0$ .

Consider the random variables  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . Because  $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  is a vector space over  $\mathbb{R}$ , we know that  $X + Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . We will now show that  $\|X + Y\|_2 \leq \|X\|_2 + \|Y\|_2$ .

Since  $|X + Y| \leq |X| + |Y|$ , we know that  $|X + Y|^2 \leq (|X| + |Y|)^2 = |X|^2 + 2|X||Y| + |Y|^2$ . Therefore,

$$\mathbb{E}(|X + Y|^2) \leq \mathbb{E}(|X|^2) + 2\mathbb{E}(|X||Y|) + \mathbb{E}(|Y|^2) = \mathbb{E}(|X|^2) + 2\mathbb{E}(|XY|) + \mathbb{E}(|Y|^2).$$

Using the Schwarz inequality,

$$\mathbb{E}(|X + Y|^2) \leq \mathbb{E}(|X|^2) + 2\|X\|_2\|Y\|_2 + \mathbb{E}(|Y|^2) = (\|X\|_2 + \|Y\|_2)^2$$

By taking the square root of both sides,

$$\|X + Y\|_2 = \mathbb{E}(|X + Y|^2)^{1/2} \leq \|X\|_2 + \|Y\|_2.$$

Consider the random variables  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . Recall that  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and  $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $\mu_X = \mathbb{E}(X)$  and  $\mu_Y = \mathbb{E}(Y)$ . Because  $(X - \mu_X) \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $(Y - \mu_Y) \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ , we know that  $(X - \mu_X)(Y - \mu_Y) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . The covariance  $\text{Cov}(X, Y)$  between  $X$  and  $Y$  is defined by

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mu_X)(Y - \mu_Y)) = \mathbb{E}(XY) - \mathbb{E}(X\mu_Y) - \mathbb{E}(Y\mu_X) + \mathbb{E}(\mu_X\mu_Y) = \mathbb{E}(XY) - \mu_X\mu_Y.$$

Consider the random variable  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . The variance  $\text{Var}(X)$  of  $X$  is defined by

$$\text{Var}(X) = \text{Cov}(X, X) = \mathbb{E}((X - \mu_X)^2) = \mathbb{E}(X^2) - \mu_X^2.$$

Consider the random variables  $U \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $V \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . The inner product  $\langle U, V \rangle$  between  $U$  and  $V$  is given by  $\langle U, V \rangle = \mathbb{E}(UV)$ . If  $\|U\|_2 \neq 0$  and  $\|V\|_2 \neq 0$ , the cosine of the angle  $\theta$  between  $U$  and  $V$  is defined as

$$\cos \theta = \frac{\langle U, V \rangle}{\|U\|_2 \|V\|_2}.$$

Because  $|\langle U, V \rangle| = |\mathbb{E}(UV)| \leq \mathbb{E}(|UV|) \leq \|U\|_2 \|V\|_2$ , we know that  $|\cos \theta| \leq 1$ .

Consider the random variables  $U, V, W, Z \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . Note the following properties of the inner product:

- $\langle U, U \rangle = \mathbb{E}(U^2) = \|U\|_2^2$ .
- $\langle U, V \rangle = \mathbb{E}(UV) = \mathbb{E}(VU) = \langle V, U \rangle$ .
- $\langle aU, V \rangle = \mathbb{E}(aUV) = a\mathbb{E}(UV) = a\langle U, V \rangle$ , for any  $a \in \mathbb{R}$ .
- $\langle U, aV \rangle = \mathbb{E}(UaV) = a\mathbb{E}(UV) = a\langle U, V \rangle$ , for any  $a \in \mathbb{R}$ .
- $\langle U + V, W \rangle = \mathbb{E}((U + V)W) = \mathbb{E}(UW + VW) = \langle U, W \rangle + \langle V, W \rangle$ .
- $\langle U, V + W \rangle = \mathbb{E}(U(V + W)) = \mathbb{E}(UV + UW) = \langle U, V \rangle + \langle U, W \rangle$ .
- $\langle U + V, W + Z \rangle = \langle U, W + Z \rangle + \langle V, W + Z \rangle = \langle U, W \rangle + \langle U, Z \rangle + \langle V, W \rangle + \langle V, Z \rangle$ .

Consider the random variables  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $\mu_X = \mathbb{E}(X)$  and  $\mu_Y = \mathbb{E}(Y)$ . The correlation  $\rho$  between  $X$  and  $Y$  is defined as the cosine of the angle between  $X - \mu_X$  and  $Y - \mu_Y$ , which is given by

$$\rho = \frac{\langle X - \mu_X, Y - \mu_Y \rangle}{\|X - \mu_X\|_2 \|Y - \mu_Y\|_2} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$

Consider the random variables  $U \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $V \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . Because  $U + V \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ ,

$$\|U + V\|_2^2 = \mathbb{E}(|U + V|^2) = \mathbb{E}((U + V)^2) = \mathbb{E}(U^2) + 2\mathbb{E}(UV) + \mathbb{E}(V^2) = \|U\|_2^2 + \|V\|_2^2 + 2\langle U, V \rangle.$$

When  $\langle U, V \rangle = 0$ , we say that  $U$  and  $V$  are orthogonal, which is denoted by  $U \perp V$ . In that case,

$$\|U + V\|_2^2 = \|U\|_2^2 + \|V\|_2^2.$$

Consider the random variables  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . Note that  $X + Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and

$$\text{Var}(X + Y) = \mathbb{E}((X + Y)^2) - \mathbb{E}(X + Y)^2 = \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}(X)^2 + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y)^2).$$

By the linearity of expectation and reorganizing terms,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

Therefore, if  $\text{Cov}(X, Y) = 0$ , then  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ .

More generally, if  $X_1, \dots, X_n \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ , then

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \text{Var}(X_k) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Cov}(X_i, X_j).$$

Consider the random variables  $U \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $V \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . The parallelogram law states that

$$\|U + V\|_2^2 + \|U - V\|_2^2 = 2\|U\|_2^2 + 2\|V\|_2^2.$$

We will now show this law. Using the relationship between the inner product and the 2-norm,

$$\|U + V\|_2^2 + \|U - V\|_2^2 = \langle U + V, U + V \rangle + \langle U - V, U - V \rangle.$$

By the bilinearity of the inner product,

$$\|U + V\|_2^2 + \|U - V\|_2^2 = \langle U, U \rangle + \langle U, V \rangle + \langle V, U \rangle + \langle V, V \rangle + \langle U, U \rangle + \langle U, -V \rangle + \langle -V, U \rangle + \langle -V, -V \rangle.$$

By cancelling terms,

$$\|U + V\|_2^2 + \|U - V\|_2^2 = 2\langle U, U \rangle + 2\langle V, V \rangle = 2\|U\|_2^2 + 2\|V\|_2^2.$$

For some  $p \in [1, \infty)$ , consider a sequence of random variables  $(X_n \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) \mid n \in \mathbb{N})$  such that

$$\lim_{k \rightarrow \infty} \sup_{r, s \geq k} \|X_r - X_s\|_p = 0.$$

We will now show that there is a random variable  $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$  such that

$$\lim_{n \rightarrow \infty} \|X_n - X\|_p = 0.$$

By definition, for every  $\epsilon > 0$  there is an  $N \in \mathbb{N}$  such that  $k \geq N$  implies  $\sup_{r, s \geq k} \|X_r - X_s\|_p < \epsilon$ . Therefore, there is a sequence  $(k_n \in \mathbb{N} \mid n \in \mathbb{N})$  such that  $k_{n+1} \geq k_n$  and  $\sup_{r, s \geq k_n} \|X_r - X_s\|_p < 1/2^n$  for every  $n \in \mathbb{N}$ .

For every  $n \in \mathbb{N}$ , the monotonicity of the norm implies that

$$\mathbb{E}(|X_{k_{n+1}} - X_{k_n}|) = \|X_{k_{n+1}} - X_{k_n}\|_1 \leq \|X_{k_{n+1}} - X_{k_n}\|_p < \frac{1}{2^n}.$$

Because  $|X_{k_{n+1}} - X_{k_n}|$  is a non-negative random variable for every  $n \in \mathbb{N}$ ,

$$\sum_n \mathbb{E}(|X_{k_{n+1}} - X_{k_n}|) = \mathbb{E} \left( \sum_n |X_{k_{n+1}} - X_{k_n}| \right) \leq \sum_n \frac{1}{2^n} < \infty.$$

Because the expectation above is finite,

$$\mathbb{P} \left( \sum_n |X_{k_{n+1}} - X_{k_n}| < \infty \right) = 1.$$

Suppose  $\sum_n |X_{k_{n+1}}(\omega) - X_{k_n}(\omega)| < \infty$  for some  $\omega \in \Omega$ . For every  $\epsilon > 0$ , the Cauchy test guarantees that there is an  $N \in \mathbb{N}$  such that  $j > i > N$  implies

$$\left| \sum_{n=i}^j |X_{k_{n+1}}(\omega) - X_{k_n}(\omega)| \right| = \sum_{n=i}^j |X_{k_{n+1}}(\omega) - X_{k_n}(\omega)| < \epsilon.$$

Furthermore, for every  $j > i$ ,

$$|X_{k_j}(\omega) - X_{k_i}(\omega)| = \left| X_{k_j}(\omega) - X_{k_i}(\omega) + \sum_{n=i+1}^{j-1} X_{k_n}(\omega) - \sum_{n=i+1}^{j-1} X_{k_n}(\omega) \right| = \left| \sum_{n=i+1}^j X_{k_n}(\omega) - \sum_{n=i}^{j-1} X_{k_n}(\omega) \right|.$$

By shifting indices and using the triangle inequality, for  $j > i > N$ ,

$$|X_{k_j}(\omega) - X_{k_i}(\omega)| = \left| \sum_{n=i}^{j-1} X_{k_{n+1}}(\omega) - X_{k_n}(\omega) \right| \leq \sum_{n=i}^{j-1} |X_{k_{n+1}}(\omega) - X_{k_n}(\omega)| < \epsilon.$$

For  $j = i > N$ , note that  $|X_{k_j}(\omega) - X_{k_i}(\omega)| = 0 < \epsilon$ . Therefore, for every  $\epsilon > 0$  there is an  $N \in \mathbb{N}$  such that  $j > N$  and  $i > N$  implies  $|X_{k_j}(\omega) - X_{k_i}(\omega)| < \epsilon$ , such that  $(X_{k_n}(\omega) \mid n \in \mathbb{N})$  is a Cauchy sequence of real numbers.

Because every Cauchy sequence of real numbers converges to a real number, consider the random variable  $X = \limsup_{n \rightarrow \infty} X_{k_n}$  such that  $\lim_{n \rightarrow \infty} X_{k_n}(\omega) = \limsup_{n \rightarrow \infty} X_{k_n}(\omega) = X(\omega)$ .

Since  $\{\sum_n |X_{k_{n+1}} - X_{k_n}| < \infty\} \subseteq \{\lim_{n \rightarrow \infty} X_{k_n} = X\}$ ,

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_{k_n} = X\right) \geq \mathbb{P}\left(\sum_n |X_{k_{n+1}} - X_{k_n}| < \infty\right) = 1.$$

Suppose  $\lim_{n \rightarrow \infty} X_{k_n}(\omega) = X(\omega)$  for some  $\omega \in \Omega$ . For every  $r \in \mathbb{N}$ ,

$$\left|\lim_{n \rightarrow \infty} X_{k_n}(\omega) - X_r(\omega)\right|^p = \lim_{n \rightarrow \infty} |X_{k_n}(\omega) - X_r(\omega)|^p = |X(\omega) - X_r(\omega)|^p.$$

Because  $\{\lim_{n \rightarrow \infty} X_{k_n} = X\} \subseteq \{\lim_{n \rightarrow \infty} |X_{k_n} - X_r|^p = |X - X_r|^p\}$  for every  $r \in \mathbb{N}$ ,

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} |X_{k_n} - X_r|^p = |X - X_r|^p\right) \geq \mathbb{P}\left(\lim_{n \rightarrow \infty} X_{k_n} = X\right) = 1.$$

Because  $|X_{k_n} - X_r|^p \geq 0$  for every  $n \in \mathbb{N}$ , by the Fatou lemma,

$$\mathbb{E}(|X - X_r|^p) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(|X_{k_n} - X_r|^p).$$

For any  $t \in \mathbb{N}$ , suppose  $r \geq k_t$  and recall that  $k_n \geq k_t$  whenever  $n \geq t$ . In that case,

$$\mathbb{E}(|X_{k_n} - X_r|^p) = \|X_{k_n} - X_r\|_p^p < \frac{1}{2^{tp}}.$$

For any  $\epsilon > 0$ , choose  $t \in \mathbb{N}$  such that  $1/2^{tp} < \epsilon$ . In that case, for any  $r \geq k_t$ ,

$$\mathbb{E}(|X - X_r|^p) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(|X_{k_n} - X_r|^p) \leq \frac{1}{2^{tp}} < \epsilon.$$

Because  $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$  is a vector space over the field  $\mathbb{R}$ , the fact that  $(X - X_r) \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$  and  $X_r \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$  implies that  $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ . The previous inequality also shows that

$$\lim_{r \rightarrow \infty} \mathbb{E}(|X - X_r|^p) = \lim_{r \rightarrow \infty} \|X - X_r\|_p^p = 0.$$

A vector space  $\mathcal{K} \subseteq \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$  is said to be complete if for every sequence  $(V_n \in \mathcal{K} \mid n \in \mathbb{N})$  such that

$$\lim_{k \rightarrow \infty} \sup_{r, s \geq k} \|V_r - V_s\|_p = 0$$

there is a  $V \in \mathcal{K}$  such that

$$\lim_{n \rightarrow \infty} \|V_n - V\|_p = 0.$$

We will now show that if the vector space  $\mathcal{K} \subseteq \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  is complete, then for every  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  there is a so-called version  $Y \in \mathcal{K}$  of the orthogonal projection of  $X$  onto  $\mathcal{K}$  such that  $\|X - Y\|_2 = \inf\{\|X - W\|_2 \mid W \in \mathcal{K}\}$  and  $X - Y \perp Z$  for every  $Z \in \mathcal{K}$ . Furthermore, if  $Y$  and  $\tilde{Y}$  are versions of the orthogonal projection of  $X$  onto  $\mathcal{K}$ , then  $\mathbb{P}(Y = \tilde{Y}) = 1$ .

For some  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ , let  $\Delta = \inf\{\|X - W\|_2 \mid W \in \mathcal{K}\}$ . First, we will show that it is possible to choose a sequence  $(Y_n \in \mathcal{K} \mid n \in \mathbb{N})$  such that  $\lim_{n \rightarrow \infty} \|X - Y_n\|_2 = \Delta$ . Recall that for every  $\epsilon > 0$  there is a  $W \in \mathcal{K}$  such that  $\|X - W\|_2 < \Delta + \epsilon$ . Choose  $Y_n$  such that  $\|X - Y_n\|_2 < \Delta + \frac{1}{n+1}$ . For every  $\epsilon > 0$ , there is an  $N \in \mathbb{N}$  such that  $n \geq N$  implies that  $\|X - Y_n\|_2 < \Delta + \epsilon$ , which is equivalent to  $|\|X - Y_n\|_2 - \Delta| < \epsilon$  since  $\Delta \leq \|X - Y_n\|_2$ .

Let  $U = X - \frac{1}{2}(Y_r + Y_s)$  and  $V = \frac{1}{2}(Y_r - Y_s)$  such that  $U + V = X - Y_s$  and  $U - V = X - Y_r$ . Because  $U \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $V \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ , the parallelogram law guarantees that

$$\|X - Y_s\|_2^2 + \|X - Y_r\|_2^2 = 2 \left\| X - \frac{1}{2}(Y_r + Y_s) \right\|_2^2 + 2 \left\| \frac{1}{2}(Y_r - Y_s) \right\|_2^2.$$

Therefore,

$$2 \left\| \frac{1}{2}(Y_r - Y_s) \right\|_2^2 = 2 \left\langle \frac{1}{2}(Y_r - Y_s), \frac{1}{2}(Y_r - Y_s) \right\rangle = \|X - Y_s\|_2^2 + \|X - Y_r\|_2^2 - 2 \left\| X - \frac{1}{2}(Y_r + Y_s) \right\|_2^2.$$

Using properties of the inner product and reorganizing terms,

$$\|Y_r - Y_s\|_2^2 = 2\|X - Y_s\|_2^2 + 2\|X - Y_r\|_2^2 - 4\left\|X - \frac{1}{2}(Y_r + Y_s)\right\|_2^2.$$

Because  $(Y_r + Y_s)/2 \in \mathcal{K}$ , we know that  $\|X - (Y_r + Y_s)/2\|_2^2 \geq \Delta^2$ . Therefore,

$$\|Y_r - Y_s\|_2^2 \leq 2\|X - Y_s\|_2^2 + 2\|X - Y_r\|_2^2 - 4\Delta^2.$$

For every  $\epsilon > 0$ , since  $\lim_{n \rightarrow \infty} \|X - Y_n\|_2^2 = \Delta^2$ , there is a  $k$  such that  $n \geq k$  implies  $|\|X - Y_n\|_2^2 - \Delta^2| < \frac{\epsilon}{4}$ , which is equivalent to  $\|X - Y_n\|_2^2 < \frac{\epsilon}{4} + \Delta^2$ . Therefore, whenever  $r, s \geq k$ ,

$$\|Y_r - Y_s\|_2^2 \leq 2\|X - Y_s\|_2^2 + 2\|X - Y_r\|_2^2 - 4\Delta^2 < 2\left(\frac{\epsilon}{4} + \Delta^2\right) + 2\left(\frac{\epsilon}{4} + \Delta^2\right) - 4\Delta^2 = \epsilon,$$

which implies

$$\lim_{k \rightarrow \infty} \sup_{r, s \geq k} \|Y_r - Y_s\|_2 = 0.$$

Because  $\mathcal{K}$  is complete, there is an  $Y \in \mathcal{K}$  such that

$$\lim_{n \rightarrow \infty} \|Y_n - Y\|_2 = 0.$$

Let  $U = X - Y_n$  and  $V = Y_n - Y$  such that  $U + V = X - Y$ . Because  $U \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $V \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ ,

$$\Delta \leq \|X - Y\|_2 \leq \|X - Y_n\|_2 + \|Y_n - Y\|_2.$$

Using the squeeze theorem when  $n \rightarrow \infty$  shows that  $\|X - Y\|_2 = \Delta = \inf\{\|X - W\|_2 \mid W \in \mathcal{K}\}$ .

For some  $Z \in \mathcal{K}$  and  $t \in \mathbb{R}$ , let  $U = X - Y$  and  $V = -tZ$  such that  $U + V = X - Y - tZ$ . Because  $U \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $V \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and considering the bilinearity of the inner product,

$$\|X - Y - tZ\|_2^2 = \|X - Y\|_2^2 + \|-tZ\|_2^2 + 2\langle X - Y, -tZ \rangle = \|X - Y\|_2^2 + t^2\|Z\|_2^2 - 2t\langle X - Y, Z \rangle.$$

Because  $(Y + tZ) \in \mathcal{K}$ , we know that  $\|X - Y\|_2^2 \leq \|X - (Y + tZ)\|_2^2$ . Therefore, for every  $Z \in \mathcal{K}$  and  $t \in \mathbb{R}$ ,

$$t^2\|Z\|_2^2 \geq 2t\langle X - Y, Z \rangle.$$

We will now show that the previous inequation can only be true for every  $Z \in \mathcal{K}$  and  $t \in \mathbb{R}$  if  $\langle X - Y, Z \rangle = 0$  for every  $Z \in \mathcal{K}$ , which implies  $X - Y \perp Z$  for every  $Z \in \mathcal{K}$ .

In order to find a contradiction, suppose that  $\langle X - Y, Z \rangle \neq 0$  for some  $Z \in \mathcal{K}$ . Because  $(X - Y) \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $Z \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ , the Schwarz inequality implies that

$$\|X - Y\|_2\|Z\|_2 \geq \mathbb{E}(|(X - Y)Z|) \geq |\mathbb{E}((X - Y)Z)| \geq 0.$$

Clearly,  $|\mathbb{E}((X - Y)Z)| = 0$  when  $\|Z\|_2 = 0$ , which implies  $\mathbb{E}((X - Y)Z) = \langle X - Y, Z \rangle = 0$ . Therefore, we can suppose that  $\|Z\|_2 > 0$ . If  $\langle X - Y, Z \rangle > 0$ , then choose a  $t \in \mathbb{R}$  such that  $0 < t < 2\langle X - Y, Z \rangle / \|Z\|_2^2$ . If  $\langle X - Y, Z \rangle < 0$ , then choose a  $t \in \mathbb{R}$  such that  $2\langle X - Y, Z \rangle / \|Z\|_2^2 < t < 0$ . In either case,  $t^2\|Z\|_2^2 < 2t\langle X - Y, Z \rangle$ , which is a contradiction.

Suppose that  $Y$  and  $\tilde{Y}$  are versions of the orthogonal projection of  $X$  onto  $\mathcal{K}$ . Because  $(\tilde{Y} - Y) \in \mathcal{K}$ ,

$$\langle X - Y, \tilde{Y} - Y \rangle = \langle X - \tilde{Y}, \tilde{Y} - Y \rangle = 0.$$

By the bilinearity of the inner product,

$$\langle X, \tilde{Y} - Y \rangle + \langle -Y, \tilde{Y} - Y \rangle - \langle X, \tilde{Y} - Y \rangle - \langle -\tilde{Y}, \tilde{Y} - Y \rangle = \langle -Y, \tilde{Y} - Y \rangle - \langle -\tilde{Y}, \tilde{Y} - Y \rangle = \langle \tilde{Y} - Y, \tilde{Y} - Y \rangle = 0$$

Because  $\langle \tilde{Y} - Y, \tilde{Y} - Y \rangle = \mathbb{E}((\tilde{Y} - Y)^2) = 0$  and  $(\tilde{Y} - Y)^2$  is a non-negative random variable, we know that  $\mathbb{P}((\tilde{Y} - Y)^2 \neq 0) = 0$ , which implies that  $\mathbb{P}(\tilde{Y} = Y) = 1$ .

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a random variable  $X : \Omega \rightarrow \mathbb{R}$ . Recall that  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \Lambda_X)$  is also a probability triple, where  $\Lambda_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  is the law of  $X$  given by  $\Lambda_X(B) = \mathbb{P}(X^{-1}(B))$  for every  $B \in \mathcal{B}(\mathbb{R})$ . We

will now show that if  $h : \mathbb{R} \rightarrow \mathbb{R}$  is a Borel function, then  $(h \circ X) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  if and only if  $h \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \Lambda_X)$ . Furthermore, in that case,

$$\int_{\Omega} (h \circ X) d\mathbb{P} = \mathbb{P}(h \circ X) = \Lambda_X(h) = \int_{\mathbb{R}} h d\Lambda_X.$$

First, suppose  $h = \mathbb{I}_B$  for some  $B \in \mathcal{B}(\mathbb{R})$ . For every  $\omega \in \Omega$ ,

$$(h \circ X)(\omega) = \mathbb{I}_B(X(\omega)) = \mathbb{I}_{X^{-1}(B)}(\omega) = \begin{cases} 1, & \text{if } X(\omega) \in B, \\ 0, & \text{if } X(\omega) \notin B. \end{cases}$$

Therefore,  $\mathbb{P}(h \circ X) = \mathbb{P}(\mathbb{I}_{X^{-1}(B)}) = \mathbb{P}(X^{-1}(B)) = \Lambda_X(B) = \Lambda_X(\mathbb{I}_B) = \Lambda_X(h) < \infty$ . Because  $h$  is  $\mathcal{B}(\mathbb{R})$ -measurable and  $(h \circ X)$  is  $\mathcal{F}$ -measurable, this step is complete.

Next, suppose  $h$  is a simple function that can be written as  $h = \sum_{k=1}^m a_k \mathbb{I}_{A_k}$  for some fixed  $a_1, a_2, \dots, a_m \in [0, \infty]$  and  $A_1, A_2, \dots, A_m \in \mathcal{B}(\mathbb{R})$ . For every  $\omega \in \Omega$ ,

$$(h \circ X)(\omega) = \sum_{k=1}^m a_k \mathbb{I}_{A_k}(X(\omega)) = \sum_{k=1}^m a_k \mathbb{I}_{X^{-1}(A_k)}(\omega).$$

Therefore,  $\mathbb{P}(h \circ X) = \sum_{k=1}^m a_k \mathbb{P}(X^{-1}(A_k)) = \sum_{k=1}^m a_k \Lambda_X(A_k) = \Lambda_X(\sum_{k=1}^m a_k \mathbb{I}_{A_k}) = \Lambda_X(h)$ . Because  $h$  is  $\mathcal{B}(\mathbb{R})$ -measurable and  $(h \circ X)$  is  $\mathcal{F}$ -measurable, this step is complete since  $\Lambda_X(h) < \infty$  if and only if  $\mathbb{P}(h \circ X) < \infty$ .

Next, suppose  $h$  is a non-negative Borel function. For any  $n \in \mathbb{N}$ , consider the simple function  $h_n = \alpha_n \circ h$ , where  $\alpha_n$  is the  $n$ -th staircase function. Because  $h_n \uparrow h$ , the monotone-convergence theorem implies that  $\Lambda_X(h_n) \uparrow \Lambda_X(h)$ . Similarly, consider the simple function  $\alpha_n \circ (h \circ X) = (\alpha_n \circ h) \circ X = h_n \circ X$ . Because  $(h_n \circ X) \uparrow (h \circ X)$ , the monotone-convergence theorem implies that  $\mathbb{P}(h_n \circ X) \uparrow \mathbb{P}(h \circ X)$ . Because our previous result implies that  $\mathbb{P}(h_n \circ X) = \Lambda_X(h_n)$ , the limit when  $n \rightarrow \infty$  shows that  $\mathbb{P}(h \circ X) = \Lambda_X(h)$ . Because  $h$  is  $\mathcal{B}(\mathbb{R})$ -measurable and  $(h \circ X)$  is  $\mathcal{F}$ -measurable, this step is complete since  $\Lambda_X(h) < \infty$  if and only if  $\mathbb{P}(h \circ X) < \infty$ .

Finally, suppose  $h$  is a Borel function. Recall that  $h = h^+ - h^-$ , where  $h^+$  and  $h^-$  are non-negative Borel functions. Therefore, if either  $\mathbb{P}(|h \circ X|) < \infty$  or  $\Lambda_X(|h|) < \infty$ , then

$$\mathbb{P}(h \circ X) = \mathbb{P}((h \circ X)^+) - \mathbb{P}((h \circ X)^-) = \mathbb{P}(h^+ \circ X) - \mathbb{P}(h^- \circ X) = \Lambda_X(h^+) - \Lambda_X(h^-) = \Lambda_X(h) < \infty,$$

where the second equality follows from associativity. Because  $h$  is  $\mathcal{B}(\mathbb{R})$ -measurable and  $(h \circ X)$  is  $\mathcal{F}$ -measurable, this completes the proof, since  $\mathbb{P}(|h \circ X|) = \Lambda_X(|h|) = \infty$  implies  $(h \circ X) \notin \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and  $h \notin \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \Lambda_X)$ .

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ . A random variable  $X : \Omega \rightarrow \mathbb{R}$  has a probability density function  $f_X$  if  $f_X : \mathbb{R} \rightarrow [0, \infty]$  is a Borel function such that the law  $\Lambda_X$  of  $X$  is given by

$$\Lambda_X(B) = \mathbb{P}(X^{-1}(B)) = \text{Leb}(f_X; B) = \text{Leb}(f_X \mathbb{I}_B) = \int_B f_X d\text{Leb},$$

for every  $B \in \mathcal{B}(\mathbb{R})$ , where  $\text{Leb}$  is the Lebesgue measure on the measurable space  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

In that case, since  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \text{Leb})$  is a measure space and  $f_X : \mathbb{R} \rightarrow [0, \infty]$  is  $\mathcal{B}(\mathbb{R})$ -measurable, recall that the measure  $(f_X \text{Leb})$  on the measurable space  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  is given by  $(f_X \text{Leb})(B) = \text{Leb}(f_X; B)$  for every  $B \in \mathcal{B}(\mathbb{R})$ , so that  $\Lambda_X = (f_X \text{Leb})$ . Therefore, using the terminology introduced in the previous section, the law  $\Lambda_X$  of  $X$  has density  $f_X$  relative to the Lebesgue measure  $\text{Leb}$ , which is denoted by

$$\frac{d\Lambda_X}{d\text{Leb}} = f_X.$$

Consider a random variable  $X : \Omega \rightarrow \mathbb{R}$  that has a probability density function  $f_X : \mathbb{R} \rightarrow [0, \infty]$ . Furthermore, consider a Borel function  $g_X : \mathbb{R} \rightarrow [0, \infty]$  such that  $\text{Leb}(\{f_X \neq g_X\}) = 0$ . Because these two functions are non-negative and  $\text{Leb}(\{f_X \mathbb{I}_B \neq g_X \mathbb{I}_B\}) = 0$ , we know that  $\text{Leb}(f_X \mathbb{I}_B) = \text{Leb}(g_X \mathbb{I}_B)$ , which implies that the random variable  $X$  also has a probability density function  $g_X$ .

Consider a measure space  $(S, \Sigma, \mu)$ , a  $\Sigma$ -measurable function  $f : S \rightarrow [0, \infty]$ , and the measure  $\lambda = (f\mu)$  on  $(S, \Sigma)$ . Recall that we say that  $\lambda$  has density  $f$  relative to  $\mu$ , which is denoted by  $d\lambda/d\mu = f$ . We will now show that if  $h : S \rightarrow \mathbb{R}$  is a  $\Sigma$ -measurable function, then  $h \in \mathcal{L}^1(S, \Sigma, \lambda)$  if and only if  $hf \in \mathcal{L}^1(S, \Sigma, \mu)$ . Furthermore, in that case,

$$\int_S h d\lambda = \lambda(h) = \mu(hf) = \int_S hf d\mu.$$

First, note that if  $h$  is  $\Sigma$ -measurable then  $hf$  is also  $\Sigma$ -measurable.

Next, let  $h = \mathbb{I}_A$  for some  $A \in \Sigma$ . In that case,  $\mu(hf) = \mu(\mathbb{I}_A f) = \mu(f; A) = \lambda(A) = \lambda(\mathbb{I}_A) = \lambda(h)$ . This step is complete since  $\mu(|hf|) < \infty$  if and only if  $\lambda(|h|) < \infty$ .

Next, suppose  $h$  is a simple function that can be written as  $h = \sum_{k=1}^m a_k \mathbb{I}_{A_k}$  for some fixed  $a_1, a_2, \dots, a_m \in [0, \infty]$  and  $A_1, A_2, \dots, A_m \in \Sigma$ . By the linearity of the integral and considering the previous step,

$$\mu(hf) = \mu\left(\sum_{k=1}^m a_k \mathbb{I}_{A_k} f\right) = \sum_{k=1}^m a_k \mu(\mathbb{I}_{A_k} f) = \sum_{k=1}^m a_k \lambda(\mathbb{I}_{A_k}) = \lambda\left(\sum_{k=1}^m a_k \mathbb{I}_{A_k}\right) = \lambda(h).$$

This step is complete since  $\mu(|hf|) < \infty$  if and only if  $\lambda(|h|) < \infty$ .

Next, suppose  $h$  is a non-negative  $\Sigma$ -measurable function. For any  $n \in \mathbb{N}$ , consider the simple function  $h_n = \alpha_n \circ h$ , where  $\alpha_n$  is the  $n$ -th staircase function. Because  $h_n \uparrow h$ , the monotone-convergence theorem implies that  $\lambda(h_n) \uparrow \lambda(h)$ . Similarly, because  $h_n f \uparrow hf$ , the monotone-convergence theorem implies that  $\mu(h_n f) \uparrow \mu(hf)$ . Because our previous result implies that  $\lambda(h_n) = \mu(h_n f)$ , the limit when  $n \rightarrow \infty$  shows that  $\mu(hf) = \lambda(h)$ . This step is complete since  $\mu(|hf|) < \infty$  if and only if  $\lambda(|h|) < \infty$ .

Finally, suppose  $h : S \rightarrow \mathbb{R}$  is a  $\Sigma$ -measurable function. Recall that  $h = h^+ - h^-$ , where  $h^+$  and  $h^-$  are non-negative  $\Sigma$ -measurable functions. If either  $\lambda(|h|) < \infty$  or  $\mu(|hf|) < \infty$ , then

$$\mu(hf) = \mu((h^+ - h^-)f) = \mu(h^+ f) - \mu(h^- f) = \lambda(h^+) - \lambda(h^-) = \lambda(h) < \infty.$$

Since  $\lambda(|h|) = \mu(|hf|) = \infty$  implies  $h \notin \mathcal{L}^1(S, \Sigma, \lambda)$  and  $hf \notin \mathcal{L}^1(S, \Sigma, \mu)$ , the proof is complete.

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a random variable  $X : \Omega \rightarrow \mathbb{R}$  with a probability density function  $f_X : \mathbb{R} \rightarrow [0, \infty]$ . Recall that the law  $\Lambda_X = (f_X \text{ Leb})$  of  $X$  has density  $f_X$  relative to  $\text{Leb}$ , which is denoted by  $d\Lambda_X/d\text{Leb} = f_X$ . If  $h : \mathbb{R} \rightarrow \mathbb{R}$  is a Borel function, the fact that  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \text{Leb})$  is a measure space implies that  $h \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \Lambda_X)$  if and only if  $hf_X \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \text{Leb})$ . Furthermore, in that case,

$$\int_{\mathbb{R}} h d\Lambda_X = \Lambda_X(h) = \text{Leb}(hf_X) = \int_{\mathbb{R}} hf_X d\text{Leb}.$$

Consider a measure space  $(S, \Sigma, \mu)$ . For every  $p \in [1, \infty)$ , the set  $\mathcal{L}^p(S, \Sigma, \mu)$  contains exactly each  $\Sigma$ -measurable function  $f : S \rightarrow \mathbb{R}$  such that  $\mu(|f|^p) < \infty$ . If  $f \in \mathcal{L}^p(S, \Sigma, \mu)$ , the  $p$ -norm  $\|f\|_p$  of the function  $f$  is given by  $\|f\|_p = \mu(|f|^p)^{1/p}$ .

Suppose that  $p > 1$  and  $p^{-1} + q^{-1} = 1$ . Furthermore, suppose  $f, g \in \mathcal{L}^p(S, \Sigma, \mu)$  and  $h \in \mathcal{L}^q(S, \Sigma, \mu)$ . Hölder's inequality states that  $fh \in \mathcal{L}^1(S, \Sigma, \mu)$  and  $\mu(|fh|) \leq \|f\|_p \|h\|_q$ . Minkowski's inequality states that  $\|f + g\|_p \leq \|f\|_p + \|g\|_p$ . We will now show these inequalities.

First, note that  $fh \in \mathcal{L}^1(S, \Sigma, \mu)$  and  $\mu(|fh|) \leq \|f\|_p \|h\|_q$  if and only if  $|f||h| \in \mathcal{L}^1(S, \Sigma, \mu)$  and  $\mu(|f||h|) \leq \|f\|_p \|h\|_q$ . Therefore, we only need to consider the case where  $f$  and  $h$  are non-negative. In that case, if  $\mu(f^p) = 0$ , then  $0 = \mu(\{f^p > 0\}) = \mu(\{f \neq 0\}) \geq \mu(\{fh \neq 0\})$  and  $\mu(fh) = 0$ , so that Hölder's inequality is trivial.

Consider the case where  $f$  and  $h$  are non-negative and  $0 < \mu(f^p) < \infty$ . Let  $\mathbb{P} : \Sigma \rightarrow [0, 1]$  be given by

$$\mathbb{P}(A) = \frac{(f^p \mu)(A)}{\mu(f^p)} = \frac{\mu(f^p; A)}{\mu(f^p)} = \frac{\mu(f^p \mathbb{I}_A)}{\mu(f^p)} = \mu\left(\frac{f^p}{\mu(f^p)} \mathbb{I}_A\right) = \mu\left(\frac{f^p}{\mu(f^p)}; A\right).$$

The function  $\mathbb{P}$  is a probability measure on  $(S, \Sigma)$ . Clearly,  $\mathbb{P}(S) = 1$  and  $\mathbb{P}(\emptyset) = 0$ . Because  $(f^p \mu)$  is a measure on  $(S, \Sigma)$ , for any sequence  $(A_n \in \Sigma \mid n \in \mathbb{N})$  such that  $A_n \cap A_m = \emptyset$  for  $n \neq m$ ,

$$\mathbb{P}\left(\bigcup_n A_n\right) = \frac{(f^p \mu)(\bigcup_n A_n)}{\mu(f^p)} = \frac{\sum_n (f^p \mu)(A_n)}{\mu(f^p)} = \sum_n \frac{(f^p \mu)(A_n)}{\mu(f^p)} = \sum_n \mathbb{P}(A_n).$$

Note that the probability measure  $\mathbb{P}$  has density  $f^p/\mu(f^p)$  relative to  $\mu$ , so that  $d\mathbb{P}/d\mu = f^p/\mu(f^p)$ . Therefore, if  $v : S \rightarrow \mathbb{R}$  is a  $\Sigma$ -measurable function, then  $v \in \mathcal{L}^1(S, \Sigma, \mathbb{P})$  if and only if  $vf^p/\mu(f^p) \in \mathcal{L}^1(S, \Sigma, \mu)$ . In that case,

$$\int_S v d\mathbb{P} = \mathbb{P}(v) = \mu\left(\frac{vf^p}{\mu(f^p)}\right) = \int_S \frac{vf^p}{\mu(f^p)} d\mu.$$

Consider the  $\Sigma$ -measurable function  $u : S \rightarrow [0, \infty]$  given by

$$u(s) = \begin{cases} \frac{h(s)}{f(s)^{p-1}}, & \text{if } f(s) > 0, \\ 0, & \text{if } f(s) = 0. \end{cases}$$

By inspecting the pointwise definition of  $uf^p$ ,

$$\mathbb{P}(u) = \mu \left( \frac{uf^p}{\mu(f^p)} \right) = \frac{\mu(uf^p)}{\mu(f^p)} = \frac{\mu(hf)}{\mu(f^p)}.$$

Similarly, by inspecting the pointwise definition of  $u^q f^p$  and using the fact that  $q(p-1) = p$ ,

$$\mathbb{P}(u^q) = \mu \left( \frac{u^q f^p}{\mu(f^p)} \right) = \frac{\mu(u^q f^p)}{\mu(f^p)} = \frac{\mu(h^q \mathbb{I}_{\{f>0\}})}{\mu(f^p)}.$$

Suppose  $\mathbb{P}(u) = \infty$ . In that case,  $\mathbb{P}(u) = \mathbb{P}(u \mathbb{I}_{\{u<1\}}) + \mathbb{P}(u \mathbb{I}_{\{u \geq 1\}}) = \infty$ . The fact that  $\mathbb{P}(u \mathbb{I}_{\{u<1\}}) \leq \mathbb{P}(\mathbb{I}_{\{u<1\}}) = \mathbb{P}(\{u < 1\}) \leq 1$  implies that  $\mathbb{P}(u \mathbb{I}_{\{u \geq 1\}}) = \infty$ . Consequently,  $\mathbb{P}(u^q) \geq \mathbb{P}(u^q \mathbb{I}_{\{u \geq 1\}}) \geq \mathbb{P}(u \mathbb{I}_{\{u \geq 1\}}) = \infty$ , so that  $\mathbb{P}(u^q) \geq \mathbb{P}(u)^q$ . In contrast, suppose  $\mathbb{P}(u) < \infty$ . Consider the convex function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  given by  $\phi(x) = |x|^q$ . Jensen's inequality also guarantees that  $\mathbb{P}(u^q) \geq \mathbb{P}(u)^q$ . Therefore,

$$\frac{\mu(h^q \mathbb{I}_{\{f>0\}})}{\mu(f^p)} \geq \frac{\mu(hf)^q}{\mu(f^p)^q}.$$

By multiplying both sides of the previous inequality by  $\mu(f^p)^q$ ,

$$\mu(h^q \mathbb{I}_{\{f>0\}}) \frac{\mu(f^p)^q}{\mu(f^p)} = \mu(h^q \mathbb{I}_{\{f>0\}}) \mu(f^p)^{q-1} \geq \mu(hf)^q.$$

Because  $\mu(h^q) \geq \mu(h^q \mathbb{I}_{\{f>0\}})$ ,

$$\mu(h^q) \mu(f^p)^{q-1} \geq \mu(hf)^q.$$

From the definition of norm and using the fact that  $p(q-1) = q$ ,

$$\|h\|_q^q \|f\|_p^q \geq \mu(hf)^q,$$

which completes the proof of Hölder's inequality.

In order to show Minkowski's inequality, recall that  $|f+g| \leq |f| + |g|$ . Therefore,

$$|f+g|^p = |f+g||f+g|^{p-1} \leq |f||f+g|^{p-1} + |g||f+g|^{p-1}.$$

By integrating both sides of the previous inequality with respect to  $\mu$  and employing Hölder's inequality,

$$\mu(|f+g|^p) \leq \mu(|f||f+g|^{p-1}) + \mu(|g||f+g|^{p-1}) \leq \|f\|_p \|f+g|^{p-1}\|_q + \|g\|_p \|f+g|^{p-1}\|_q.$$

Note that  $\|f+g|^{p-1}\|_q = \mu(|f+g|^{(p-1)q})^{1/q} = \mu(|f+g|^p)^{1/q} < \infty$  because  $q(p-1) = p$ . Therefore,

$$\mu(|f+g|^p) \leq (\|f\|_p + \|g\|_p) \mu(|f+g|^p)^{1/q}.$$

By dividing both sides of the previous inequality by  $\mu(|f+g|^p)^{1/q}$  and using the fact that  $p^{-1} = 1 - q^{-1}$ ,

$$\|f+g\|_p = \mu(|f+g|^p)^{1/p} \leq \|f\|_p + \|g\|_p.$$

## 7 Strong law

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ , a random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and a random variable  $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . We will now show that if  $X$  and  $Y$  are independent, then  $XY \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ .

First, suppose that  $X$  and  $Y$  are non-negative and let  $\alpha_n$  denote the  $n$ -th staircase function. For any  $n \in \mathbb{N}$ , consider the simple function  $X_n = \alpha_n \circ X = \sum_{k_x=1}^{m_x} a_{k_x} \mathbb{I}_{A_{k_x}}$ , where  $a_1, \dots, a_{m_x} \in [0, n]$  are distinct and  $A_1, \dots, A_{m_x} \in \mathcal{F}$  partition  $\Omega$ . Similarly, consider the simple function  $Y_n = \alpha_n \circ Y = \sum_{k_y=1}^{m_y} b_{k_y} \mathbb{I}_{B_{k_y}}$ , where  $b_1, \dots, b_{m_y} \in [0, n]$  are distinct and  $B_1, \dots, B_{m_y} \in \mathcal{F}$  partition  $\Omega$ . In that case,

$$\begin{aligned} \mathbb{E}(X_n) &= \mathbb{E} \left( \sum_{k_x=1}^{m_x} a_{k_x} \mathbb{I}_{A_{k_x}} \right) = \sum_{k_x=1}^{m_x} a_{k_x} \mathbb{P}(A_{k_x}), \\ \mathbb{E}(Y_n) &= \mathbb{E} \left( \sum_{k_y=1}^{m_y} b_{k_y} \mathbb{I}_{B_{k_y}} \right) = \sum_{k_y=1}^{m_y} b_{k_y} \mathbb{P}(B_{k_y}). \end{aligned}$$



Because  $X_n \uparrow X$ , the monotone-convergence theorem guarantees that  $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$ . Similarly, because  $Y_n \uparrow Y$ , the monotone-convergence theorem guarantees that  $\mathbb{E}(Y_n) \uparrow \mathbb{E}(Y)$ . Because  $\mathbb{E}(X) < \infty$  and  $\mathbb{E}(Y) < \infty$ , we also know that  $\mathbb{E}(X_n)\mathbb{E}(Y_n) \uparrow \mathbb{E}(X)\mathbb{E}(Y)$ . By distributing terms and using the fact that  $\mathbb{I}_{A_{k_x}}\mathbb{I}_{B_{k_y}} = \mathbb{I}_{A_{k_x} \cap B_{k_y}}$ ,

$$\mathbb{E}(X_n Y_n) = \mathbb{E} \left[ \left( \sum_{k_x=1}^{m_x} a_{k_x} \mathbb{I}_{A_{k_x}} \right) \left( \sum_{k_y=1}^{m_y} b_{k_y} \mathbb{I}_{B_{k_y}} \right) \right] = \mathbb{E} \left( \sum_{k_x=1}^{m_x} \sum_{k_y=1}^{m_y} a_{k_x} b_{k_y} \mathbb{I}_{A_{k_x} \cap B_{k_y}} \right) = \sum_{k_x=1}^{m_x} \sum_{k_y=1}^{m_y} a_{k_x} b_{k_y} \mathbb{P}(A_{k_x} \cap B_{k_y}).$$

Recall that if  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a Borel function and  $Z : \Omega \rightarrow \mathbb{R}$  is a random variable, then

$$\sigma(f \circ Z) = \{(f \circ Z)^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R})\} = \{Z^{-1}(f^{-1}(B)) \mid B \in \mathcal{B}(\mathbb{R})\} \subseteq \{Z^{-1}(A) \mid A \in \mathcal{B}(\mathbb{R})\} = \sigma(Z).$$

Recall that  $X$  and  $Y$  are independent if and only if  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$  for every  $A \in \sigma(X)$  and  $B \in \sigma(Y)$ . Therefore,  $X_n$  and  $Y_n$  are also independent. Because  $A_{k_x} = X_n^{-1}(\{a_{k_x}\})$ , we know that  $A_{k_x} \in \sigma(X_n)$ . Because  $B_{k_y} = Y_n^{-1}(\{b_{k_y}\})$ , we know that  $B_{k_y} \in \sigma(Y_n)$ . Therefore,

$$\mathbb{E}(X_n Y_n) = \sum_{k_x=1}^{m_x} \sum_{k_y=1}^{m_y} a_{k_x} b_{k_y} \mathbb{P}(A_{k_x}) \mathbb{P}(B_{k_y}) = \left( \sum_{k_x=1}^{m_x} a_{k_x} \mathbb{P}(A_{k_x}) \right) \left( \sum_{k_y=1}^{m_y} b_{k_y} \mathbb{P}(B_{k_y}) \right) = \mathbb{E}(X_n) \mathbb{E}(Y_n).$$

Since  $X_n \uparrow X$  and  $Y_n \uparrow Y$  imply  $X_n Y_n \uparrow XY$ , the monotone-convergence theorem guarantees that  $\mathbb{E}(X_n Y_n) \uparrow \mathbb{E}(XY)$ . Since  $\mathbb{E}(X_n Y_n) = \mathbb{E}(X_n) \mathbb{E}(Y_n)$ , taking the limit when  $n \rightarrow \infty$  shows that  $\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y) < \infty$ , which completes the proof when  $X$  and  $Y$  are non-negative.

Finally, let  $X = X^+ - X^-$ , where  $X^+ \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and  $X^- \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  are non-negative. Analogously, let  $Y = Y^+ - Y^-$ . Because the absolute value function is Borel, we know that  $XY \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Therefore,

$$\mathbb{E}(XY) = \mathbb{E}((X^+ - X^-)(Y^+ - Y^-)) = \mathbb{E}(X^+ Y^+) - \mathbb{E}(X^+ Y^-) - \mathbb{E}(X^- Y^+) + \mathbb{E}(X^- Y^-).$$

Since  $X$  and  $Y$  are independent, each pair of variables inside an expectation above is independent. Therefore,

$$\mathbb{E}(XY) = \mathbb{E}(X^+) \mathbb{E}(Y^+) - \mathbb{E}(X^+) \mathbb{E}(Y^-) - \mathbb{E}(X^-) \mathbb{E}(Y^+) + \mathbb{E}(X^-) \mathbb{E}(Y^-) = (\mathbb{E}(X^+) - \mathbb{E}(X^-))(\mathbb{E}(Y^+) - \mathbb{E}(Y^-)),$$

which completes the proof.

Consider the random variables  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . If  $X$  and  $Y$  are independent, the previous result guarantees that  $\text{Cov}(X, Y) = 0$  and  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ .

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ , a random variable  $X : \Omega \rightarrow \mathbb{R}$ , and the random variables  $Y_1, \dots, Y_n$ , where  $n \in \mathbb{N}^+$ . Suppose that  $X, Y_1, \dots, Y_n$  are independent. We will now show that if  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a Borel function and  $Z : \Omega \rightarrow \mathbb{R}$  is a random variable given by  $Z(\omega) = f(Y_1(\omega), \dots, Y_n(\omega))$ , then  $X$  and  $Z$  are independent.

First, recall that a previous result establishes that  $Z$  is  $\sigma(\{Y_1, \dots, Y_n\})$ -measurable, so that

$$\sigma(Z) \subseteq \sigma(\{Y_1, \dots, Y_n\}) = \sigma(\{Y_i^{-1}(B) \mid i \in \{1, \dots, n\}, B \in \mathcal{B}(\mathbb{R})\}) = \sigma\left(\bigcup_{i=1}^n \sigma(Y_i)\right).$$

Therefore, if  $\sigma(X)$  and  $\sigma(\{Y_1, \dots, Y_n\})$  are independent, then  $X$  and  $Z$  are independent.

Consider the set  $\mathcal{I} = \{\cap_{i=1}^n A_i \mid (A_1, \dots, A_n) \in \sigma(Y_1) \times \dots \times \sigma(Y_n)\}$ . If  $B \in \mathcal{I}$  and  $C \in \mathcal{I}$ , then  $B = \cap_{i=1}^n A_i$  and  $C = \cap_{i=1}^n A'_i$ , where  $A_i \in \sigma(Y_i)$  and  $A'_i \in \sigma(Y_i)$  for every  $i \in \{1, \dots, n\}$ . Because

$$B \cap C = \left( \bigcap_{i=1}^n A_i \right) \cap \left( \bigcap_{i=1}^n A'_i \right) = \bigcap_{i=1}^n (A_i \cap A'_i)$$

and  $(A_i \cap A'_i) \in \sigma(Y_i)$  for every  $i \in \{1, \dots, n\}$ , we know that  $(B \cap C) \in \mathcal{I}$ . Therefore,  $\mathcal{I}$  is a  $\pi$ -system on  $\Omega$ .

Let  $\mathcal{J} = \sigma(X)$  and note that  $\mathcal{J}$  is also a  $\pi$ -system on  $\Omega$ . Consider a set  $(\cap_{i=1}^n A_i) \in \mathcal{I}$ , where  $A_i \in \sigma(Y_i)$  for every  $i \in \{1, \dots, n\}$ , and a set  $B \in \mathcal{J}$ . Since  $X, Y_1, \dots, Y_n$  are independent,

$$\mathbb{P}\left(\left(\bigcap_{i=1}^n A_i\right) \cap B\right) = \left(\prod_{i=1}^n \mathbb{P}(A_i)\right) \mathbb{P}(B) = \mathbb{P}\left(\bigcap_{i=1}^n A_i\right) \mathbb{P}(B),$$

which implies that  $\mathcal{I}$  and  $\mathcal{J}$  are independent. Because  $\sigma(\mathcal{I})$  and  $\sigma(\mathcal{J})$  are then independent from a previous result and  $\sigma(\mathcal{J}) = \sigma(X)$ , the proof will be complete if  $\sigma(\mathcal{I}) = \sigma(\{Y_1, \dots, Y_n\})$ , which we will now show.

Note that  $\Omega \in \sigma(Y_i)$  for every  $i \in \{1, \dots, n\}$ , which implies  $\sigma(Y_i) \subseteq \mathcal{I}$  for every  $i \in \{1, \dots, n\}$ . Therefore,  $\cup_{i=1}^n \sigma(Y_i) \subseteq \mathcal{I}$  and  $\sigma(\cup_{i=1}^n \sigma(Y_i)) = \sigma(\{Y_1, \dots, Y_n\}) \subseteq \sigma(\mathcal{I})$ .

Consider a set  $(\cap_{i=1}^n A_i) \in \mathcal{I}$ , where  $A_i \in \sigma(Y_i)$  for every  $i \in \{1, \dots, n\}$ . Clearly,  $A_i \in \cup_{j=1}^n \sigma(Y_j)$ . Because  $\sigma(\cup_{j=1}^n \sigma(Y_j)) = \sigma(\{Y_1, \dots, Y_n\})$  is a  $\sigma$ -algebra, we know that  $(\cap_{i=1}^n A_i) \in \sigma(\{Y_1, \dots, Y_n\})$ , which implies  $\mathcal{I} \subseteq \sigma(\{Y_1, \dots, Y_n\})$  and  $\sigma(\mathcal{I}) \subseteq \sigma(\{Y_1, \dots, Y_n\})$ .

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a sequence of independent random variables  $(X_k : \Omega \rightarrow \mathbb{R} \mid k \in \mathbb{N}^+)$ . Furthermore, suppose  $\mathbb{E}(X_k) = 0$  and  $\mathbb{E}(X_k^4) \leq K$  for some  $K \in [0, \infty)$ , for every  $k \in \mathbb{N}^+$ . The strong law of large numbers for a finite fourth moment guarantees that

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = 0\right) = 1.$$

We will now prove this law. Consider the random variable  $S_n = \sum_{k=1}^n X_k$ . From the multinomial theorem,

$$S_n^4 = \left(\sum_{k=1}^n X_k\right)^4 = \sum_{(k_1, \dots, k_n) \in I_4^{(n)}} \frac{4!}{k_1! \dots k_n!} \prod_{t=1}^n X_t^{k_t},$$

where  $I_4^{(n)} = \{(k_1, \dots, k_n) \mid k_i \in \{0, \dots, p\} \text{ for every } i \in \{1, \dots, n\} \text{ and } \sum_i k_i = p\}$ . By the linearity of expectation,

$$\mathbb{E}(S_n^4) = \sum_{(k_1, \dots, k_n) \in I_4^{(n)}} \frac{4!}{k_1! \dots k_n!} \mathbb{E}\left(\prod_{t=1}^n X_t^{k_t}\right).$$

From the restrictions imposed on  $(k_1, \dots, k_n) \in I_4^{(n)}$ , the expectation  $\mathbb{E}\left(\prod_{t=1}^n X_t^{k_t}\right)$  can be written as either  $\mathbb{E}(X_i^4)$ ,  $\mathbb{E}(X_i^3 X_j)$ ,  $\mathbb{E}(X_i^2 X_j^2)$ ,  $\mathbb{E}(X_i^2 X_j X_k)$ , or  $\mathbb{E}(X_i X_j X_k X_l)$ , where  $i, j, k, l \in \{1, \dots, n\}$  are distinct indices.

Consider the expectation  $\mathbb{E}(X_i^3 X_j)$ . Because  $X_i$  and  $X_j$  are independent,  $X_i^3$  and  $X_j$  are independent. By the monotonicity of the norm,  $X_i^3 \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and  $X_j \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Therefore,  $\mathbb{E}(X_i^3 X_j) = \mathbb{E}(X_i^3) \mathbb{E}(X_j) = 0$ .

Consider the expectation  $\mathbb{E}(X_i^2 X_j X_k)$ . Because  $X_i^2, X_j, X_k$  are independent,  $X_i^2 X_j$  and  $X_k$  are independent. By the monotonicity of the norm,  $X_i^2 \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ ,  $X_j \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ , and  $X_k \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Due to independence,  $X_i^2 X_j \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Therefore,  $\mathbb{E}(X_i^2 X_j X_k) = \mathbb{E}(X_i^2 X_j) \mathbb{E}(X_k) = 0$ .

Consider the expectation  $\mathbb{E}(X_i X_j X_k X_l)$ . Because  $X_i, X_j, X_k, X_l$  are independent,  $X_i X_j X_k$  and  $X_l$  are independent. By the monotonicity of the norm,  $X_i, X_j, X_k, X_l \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Because  $X_i$  and  $X_j$  are independent,  $X_i X_j \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Because  $X_i X_j$  and  $X_k$  are independent,  $X_i X_j X_k \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Therefore,  $\mathbb{E}(X_i X_j X_k X_l) = \mathbb{E}(X_i X_j X_k) \mathbb{E}(X_l) = 0$ .

These observations allow rewriting the expectation  $\mathbb{E}(S_n^4)$  as

$$\mathbb{E}(S_n^4) = \sum_{i=1}^n \mathbb{E}(X_i^4) + 6 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{E}(X_i^2 X_j^2).$$

For every  $k \in \mathbb{N}^+$ , recall that  $\|X_k\|_2 = \mathbb{E}(X_k^2)^{1/2} \leq \mathbb{E}(X_k^4)^{1/4} = \|X_k\|_4$ . Therefore,  $\mathbb{E}(X_k^2) \leq \mathbb{E}(X_k^4)^{1/2} \leq K^{1/2}$ . For every  $i \neq j$ ,  $X_i^2$  and  $X_j^2$  are independent and  $X_i^2, X_j^2 \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  by the monotonicity of the norm. Therefore,

$$\mathbb{E}(X_i^2 X_j^2) = \mathbb{E}(X_i^2) \mathbb{E}(X_j^2) \leq \mathbb{E}(X_i^4)^{1/2} \mathbb{E}(X_j^4)^{1/2} \leq K.$$

Consequently,

$$\mathbb{E}(S_n^4) \leq \sum_{i=1}^n K + 6 \sum_{i=1}^{n-1} \sum_{j=i+1}^n K = nK + 3n(n-1)K = K(3n^2 - 2n) \leq 3Kn^2.$$

Because  $\mathbb{E}(S_n^4/n^4) \leq 3K/n^2$  for every  $n \in \mathbb{N}^+$ ,

$$\sum_{n=1}^k \mathbb{E}\left(\frac{S_n^4}{n^4}\right) \leq 3K \sum_{n=1}^k \frac{1}{n^2}.$$

Because the summation on the right of the inequality above converges to a real number when  $k \rightarrow \infty$ ,

$$\sum_n \mathbb{E} \left( \frac{S_n^4}{n^4} \right) < \infty.$$

Since  $S_n^4/n^4$  is a non-negative random variable for every  $n \in \mathbb{N}^+$ , a previous result guarantees that

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} \frac{S_n^4}{n^4} = 0 \right) = \mathbb{P} \left( \lim_{n \rightarrow \infty} \frac{S_n}{n} = 0 \right) = \mathbb{P} \left( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = 0 \right) = 1.$$

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a sequence of independent random variables  $(X_k : \Omega \rightarrow \mathbb{R} \mid k \in \mathbb{N}^+)$ . Furthermore, suppose  $\mathbb{E}(X_k) = \mu$  and  $\mathbb{E}(X_k^4) \leq K$  for some  $\mu \in \mathbb{R}$  and  $K \in [0, \infty)$ , for every  $k \in \mathbb{N}^+$ . As a corollary, the strong law of large numbers for a finite fourth moment guarantees that

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = \mu \right) = 1.$$

We will now show this corollary. For every  $k \in \mathbb{N}^+$ , let  $Y_k = X_k - \mu$ . By the monotonicity of the norm,  $X_k \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ , so that  $\mathbb{E}(Y_k) = \mathbb{E}(X_k) - \mu = 0$ . Furthermore,  $(Y_k : \Omega \rightarrow \mathbb{R} \mid k \in \mathbb{N}^+)$  is a sequence of independent random variables, since  $\sigma(Y_k) \subseteq \sigma(X_k)$ . Using Minkowski's inequality and the fact that  $X_k \in \mathcal{L}^4(\Omega, \mathcal{F}, \mathbb{P})$ ,

$$\infty > \|X_k\|_4 + |\mu| = \|X_k\|_4 + \|-\mu \mathbb{I}_\Omega\|_4 \geq \|X_k - \mu \mathbb{I}_\Omega\|_4 = \|X_k - \mu\|_4 = \|Y_k\|_4.$$

Therefore,  $\mathbb{E}(Y_k^4) \leq K'$  for some  $K' \in [0, \infty)$ . Using the strong law of large numbers for a finite fourth moment,

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Y_k = 0 \right) = \mathbb{P} \left( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = \mu \right) = 1$$

Consider a random variable  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and let  $\mu = \mathbb{E}(X)$ . For  $c \geq 0$ , Chebyshev's inequality states that

$$\text{Var}(X) = \mathbb{E}(|X - \mu|^2) \geq c^2 \mathbb{P}(|X - \mu| \geq c),$$

where the inequality above is a consequence of Markov's inequality.

As an application of Chebyshev's inequality, consider the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a sequence of independent and identically distributed random variables  $(X_k : \Omega \rightarrow \{0, 1\} \mid k \in \mathbb{N}^+)$ . Let  $p = \mathbb{E}(X_k) = \mathbb{E}(\mathbb{I}_{\{X_k=1\}}) = \mathbb{P}(X_k = 1)$ . Since  $X_k^2 = X_k$ ,  $X_k \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $\text{Var}(X_k) = \mathbb{E}(X_k^2) - \mathbb{E}(X_k)^2 = p - p^2$ , so that  $\text{Var}(X_k) \leq 1/4$ .

Let  $S_n = \sum_{k=1}^n X_k$ , so that  $\mathbb{E}(S_n) = \sum_{k=1}^n \mathbb{E}(X_k) = np$ . Due to independence,

$$\text{Var}(S_n) = \text{Var} \left( \sum_{k=1}^n X_k \right) = \sum_{k=1}^n \text{Var}(X_k) = \sum_{k=1}^n p - p^2 = n(p - p^2) \leq \frac{n}{4}.$$

For any  $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $a \in \mathbb{R}$ ,  $\text{Var}(aY) = \mathbb{E}((aY)^2) - \mathbb{E}(aY)^2 = a^2 \mathbb{E}(Y^2) - a^2 \mathbb{E}(Y)^2 = a^2 \text{Var}(Y)$ . Therefore,  $\mathbb{E}(S_n/n) = p$  and  $\text{Var}(S_n/n) \leq 1/4n$ . Using Chebyshev's inequality, for any  $\delta > 0$ ,

$$\mathbb{P} \left( \left| \left( \frac{1}{n} \sum_{k=1}^n X_k \right) - p \right| \geq \delta \right) \leq \frac{1}{4n\delta^2}.$$

## 8 Product measure

Consider a measurable space  $(S_1, \Sigma_1)$  and a measurable space  $(S_2, \Sigma_2)$ . Let  $S = S_1 \times S_2$ . Consider also the functions  $\rho_1 : S \rightarrow S_1$  and  $\rho_2 : S \rightarrow S_2$  given by  $\rho_1(s_1, s_2) = s_1$  and  $\rho_2(s_1, s_2) = s_2$ . For  $B_1 \in \Sigma_1$  and  $B_2 \in \Sigma_2$ , let

$$\begin{aligned} \rho_1^{-1}(B_1) &= \{(s_1, s_2) \in S \mid \rho_1(s_1, s_2) \in B_1\} = \{(s_1, s_2) \in S \mid s_1 \in B_1\} = B_1 \times S_2, \\ \rho_2^{-1}(B_2) &= \{(s_1, s_2) \in S \mid \rho_2(s_1, s_2) \in B_2\} = \{(s_1, s_2) \in S \mid s_2 \in B_2\} = S_1 \times B_2. \end{aligned}$$

For  $i \in \{1, 2\}$ , let  $\mathcal{A}_i = \{\rho_i^{-1}(B_i) \mid B_i \in \Sigma_i\}$ . We will now show that  $\mathcal{A}_i$  is a  $\sigma$ -algebra on  $S$ . First, note that  $\rho_i^{-1}(S_i) = S$  and  $S_i \in \Sigma_i$ . Therefore,  $S \in \mathcal{A}_i$ . Consider an element  $\rho_i^{-1}(B_i) \in \mathcal{A}_i$ . Note that  $B_i^c \in \Sigma_i$  and

$\rho_i^{-1}(B_i^c) = \rho_i^{-1}(B_i)^c$ . Therefore,  $\rho_i^{-1}(B_i)^c \in \mathcal{A}_i$ . Finally, consider a sequence of sets  $(\rho_i^{-1}(B_{i,j}) \in \mathcal{A}_i \mid j \in \mathbb{N})$ . Note that  $\cup_j B_{i,j} \in \Sigma_i$  and  $\rho_i^{-1}(\cup_j B_{i,j}) = \cup_j \rho_i^{-1}(B_{i,j})$ . Therefore,  $\cup_j \rho_i^{-1}(B_{i,j}) \in \mathcal{A}_i$ .

Considering the previous result, let  $\sigma(\rho_1)$  and  $\sigma(\rho_2)$  denote the  $\sigma$ -algebras on  $S$  given by

$$\begin{aligned}\sigma(\rho_1) &= \mathcal{A}_1 = \{\rho_1^{-1}(B_1) \mid B_1 \in \Sigma_1\} = \{B_1 \times S_2 \mid B_1 \in \Sigma_1\}, \\ \sigma(\rho_2) &= \mathcal{A}_2 = \{\rho_2^{-1}(B_2) \mid B_2 \in \Sigma_2\} = \{S_1 \times B_2 \mid B_2 \in \Sigma_2\}.\end{aligned}$$

The product  $\Sigma$  between the  $\sigma$ -algebras  $\Sigma_1$  and  $\Sigma_2$  is a  $\sigma$ -algebra on  $S$  denoted by  $\Sigma_1 \times \Sigma_2$  but defined by

$$\Sigma = \Sigma_1 \times \Sigma_2 = \sigma(\{\rho_1, \rho_2\}) = \sigma(\sigma(\rho_1) \cup \sigma(\rho_2)),$$

which should not be confused with the Cartesian product between  $\Sigma_1$  and  $\Sigma_2$ .

Consider the set  $\mathcal{I} = \{B_1 \times B_2 \mid B_1 \in \Sigma_1 \text{ and } B_2 \in \Sigma_2\}$ . For any  $B_1 \in \Sigma_1$  and  $B_2 \in \Sigma_2$ , note that

$$B_1 \times B_2 = (B_1 \cap S_1) \times (S_2 \cap B_2) = (B_1 \times S_2) \cap (S_1 \times B_2).$$

Suppose  $B_1 \times B_2 \in \mathcal{I}$  and  $B'_1 \times B'_2 \in \mathcal{I}$ . In that case,  $(B_1 \times B_2) \cap (B'_1 \times B'_2) = (B_1 \cap B'_1) \times (B_2 \cap B'_2)$ . Because  $(B_1 \cap B'_1) \in \Sigma_1$  and  $(B_2 \cap B'_2) \in \Sigma_2$ , this implies that  $\mathcal{I}$  is a  $\pi$ -system on  $S$ .

We will now show that  $\sigma(\mathcal{I}) = \Sigma$ . For any  $B_1 \times B_2 \in \mathcal{I}$ , we know that  $B_1 \times B_2 \in \Sigma$  because  $(B_1 \times S_2) \in \sigma(\rho_1)$  and  $(S_1 \times B_2) \in \sigma(\rho_2)$ . Since  $\Sigma$  is a  $\sigma$ -algebra,  $\sigma(\mathcal{I}) \subseteq \Sigma$ . For any  $B_1 \in \Sigma_1$  and  $B_2 \in \Sigma_2$ , we know that  $B_1 \times S_2 \in \mathcal{I}$  and  $S_1 \times B_2 \in \mathcal{I}$ . Therefore,  $\sigma(\rho_1) \cup \sigma(\rho_2) \subseteq \mathcal{I}$ . Because  $\sigma(\mathcal{I})$  is a  $\sigma$ -algebra,  $\Sigma \subseteq \sigma(\mathcal{I})$ .

Consider a measurable space  $(S_1, \Sigma_1)$  and a measurable space  $(S_2, \Sigma_2)$ . Furthermore, consider the measurable space  $(S, \Sigma)$ , where  $S = S_1 \times S_2$  and  $\Sigma = \Sigma_1 \times \Sigma_2$ . Let  $\mathcal{H}$  denote a set that contains exactly each bounded  $\Sigma$ -measurable function  $f : S \rightarrow \mathbb{R}$  for which there is a  $\Sigma_2$ -measurable function  $f_{s_1} : S_2 \rightarrow \mathbb{R}$  and a  $\Sigma_1$ -measurable function  $f_{s_2} : S_1 \rightarrow \mathbb{R}$  such that  $f(s_1, s_2) = f_{s_1}(s_2) = f_{s_2}(s_1)$  for every  $s_1 \in S_1$  and  $s_2 \in S_2$ . We will now show that  $\mathcal{H}$  contains every bounded  $\Sigma$ -measurable function on  $S$ , so that  $\mathcal{H} = \text{b}\Sigma$ .

Note that the set of bounded  $\Sigma$ -measurable functions  $\text{b}\Sigma$  is a vector space over the field  $\mathbb{R}$  when scalar multiplication and addition are performed pointwise. Because  $\mathcal{H} \subseteq \text{b}\Sigma$ , showing that  $\mathcal{H}$  is a vector space only requires showing that  $\mathcal{H}$  is non-empty and closed under scalar multiplication and addition. For every  $s_1 \in S_1$  and  $s_2 \in S_2$ , let  $f = \mathbb{I}_S$ ,  $f_{s_1} = \mathbb{I}_{S_2}$ , and  $f_{s_2} = \mathbb{I}_{S_1}$ , so that  $\mathbb{I}_S(s_1, s_2) = \mathbb{I}_{S_2}(s_2) = \mathbb{I}_{S_1}(s_1) = 1$ . Clearly,  $f \in \mathcal{H}$ . Now suppose  $f \in \mathcal{H}$  and  $a \in \mathbb{R}$ . Note that  $af \in \text{b}\Sigma$ . For every  $s_1 \in S_1$  and  $s_2 \in S_2$ , also note that  $af_{s_1}$  is  $\Sigma_2$ -measurable,  $af_{s_2}$  is  $\Sigma_1$ -measurable, and  $(af)(s_1, s_2) = (af_{s_1})(s_2) = (af_{s_2})(s_1)$ . Therefore,  $af \in \mathcal{H}$ . Finally, suppose that  $g, h \in \mathcal{H}$ . Note that  $g + h \in \text{b}\Sigma$ . For every  $s_1 \in S_1$  and  $s_2 \in S_2$ , note that  $g_{s_1} + h_{s_1}$  is  $\Sigma_2$ -measurable,  $g_{s_2} + h_{s_2}$  is  $\Sigma_1$ -measurable, and  $(g + h)(s_1, s_2) = (g_{s_1} + h_{s_1})(s_2) = (g_{s_2} + h_{s_2})(s_1)$ . Therefore,  $g + h \in \mathcal{H}$ .

Suppose  $(f_n \in \mathcal{H} \mid n \in \mathbb{N})$  is a sequence of non-negative functions in  $\mathcal{H}$  such that  $f_n \uparrow f$ , where  $f : S \rightarrow [0, \infty)$  is a bounded function. Note that  $f \in \text{b}\Sigma$ , since  $f$  is the limit of a sequence of (bounded)  $\Sigma$ -measurable functions. For every  $s_1 \in S_1$  and  $s_2 \in S_2$ , note that  $f_{s_1} = \lim_{n \rightarrow \infty} f_{n, s_1}$  is  $\Sigma_2$ -measurable,  $f_{s_2} = \lim_{n \rightarrow \infty} f_{n, s_2}$  is  $\Sigma_1$ -measurable, and  $f(s_1, s_2) = f_{s_1}(s_2) = f_{s_2}(s_1)$ . Therefore,  $f \in \mathcal{H}$ .

Consider the  $\pi$ -system  $\mathcal{I} = \{B_1 \times B_2 \mid B_1 \in \Sigma_1 \text{ and } B_2 \in \Sigma_2\}$  and the indicator function  $f = \mathbb{I}_{B_1 \times B_2}$  of a set  $B_1 \times B_2 \in \mathcal{I}$ . Note that  $f$  is a bounded  $\Sigma$ -measurable function, since  $B_1 \times B_2 \in \Sigma$ . For every  $s_1 \in S_1$  and  $s_2 \in S_2$ , note that  $f_{s_1} = \mathbb{I}_{B_1}(s_1)\mathbb{I}_{B_2}$  is  $\Sigma_2$ -measurable,  $f_{s_2} = \mathbb{I}_{B_2}(s_2)\mathbb{I}_{B_1}$  is  $\Sigma_1$ -measurable, and  $f(s_1, s_2) = f_{s_1}(s_2) = f_{s_2}(s_1)$ . Therefore,  $f \in \mathcal{H}$ . Since  $\sigma(\mathcal{I}) = \Sigma$ , the monotone-class theorem completes the proof.

Consider a measure space  $(S_1, \Sigma_1, \mu_1)$ , a measure space  $(S_2, \Sigma_2, \mu_2)$ , and the measurable space  $(S, \Sigma)$ , where  $S = S_1 \times S_2$  and  $\Sigma = \Sigma_1 \times \Sigma_2$ . Furthermore, suppose  $\mu_1$  and  $\mu_2$  are finite measures.

For any bounded  $\Sigma$ -measurable function  $f : S \rightarrow \mathbb{R}$ , let  $I_1^f : S_1 \rightarrow \mathbb{R}$  and  $I_2^f : S_2 \rightarrow \mathbb{R}$  be given by

$$\begin{aligned}I_1^f(s_1) &= \int_{S_2} f(s_1, s_2) \mu_2(ds_2) = \int_{S_2} f_{s_1}(s_2) \mu_2(ds_2) = \mu_2(f_{s_1}), \\ I_2^f(s_2) &= \int_{S_1} f(s_1, s_2) \mu_1(ds_1) = \int_{S_1} f_{s_2}(s_1) \mu_1(ds_1) = \mu_1(f_{s_2}),\end{aligned}$$

where  $f_{s_1} : S_2 \rightarrow \mathbb{R}$  is a  $\Sigma_2$ -measurable function,  $f_{s_2} : S_1 \rightarrow \mathbb{R}$  is a  $\Sigma_1$ -measurable function, and  $f(s_1, s_2) = f_{s_1}(s_2) = f_{s_2}(s_1)$ , for every  $s_1 \in S_1$  and  $s_2 \in S_2$ . Note that  $\mu_2(|f_{s_1}|) < \infty$  because  $\mu_2$  is finite and  $|f_{s_1}| \in \text{b}\Sigma_2$ . Similarly,  $\mu_1(|f_{s_2}|) < \infty$  because  $\mu_1$  is finite and  $|f_{s_2}| \in \text{b}\Sigma_1$ . Therefore,  $I_1^f$  and  $I_2^f$  are bounded.

Let  $\mathcal{H}$  denote a set that contains exactly each function  $f \in \text{b}\Sigma$  such that  $I_1^f \in \text{b}\Sigma_1$  and  $I_2^f \in \text{b}\Sigma_2$  and

$$\mu_1(I_1^f) = \int_{S_1} I_1^f(s_1) \mu_1(ds_1) = \int_{S_2} I_2^f(s_2) \mu_2(ds_2) = \mu_2(I_2^f).$$

We will now show that  $\mathcal{H}$  contains every bounded  $\Sigma$ -measurable function on  $S$ , so that  $\mathcal{H} = \text{b}\Sigma$ .

Because  $\mathcal{H} \subseteq \text{b}\Sigma$ , showing that  $\mathcal{H}$  is a vector space only requires showing that  $\mathcal{H}$  is non-empty and closed under scalar multiplication and addition. For every  $s_1 \in S_1$  and  $s_2 \in S_2$ , let  $f = \mathbb{I}_S$ ,  $f_{s_1} = \mathbb{I}_{S_2}$ , and  $f_{s_2} = \mathbb{I}_{S_1}$ , so that  $I_1^f(s_1) = \mu_2(\mathbb{I}_{S_2}) = \mu_2(S_2)\mathbb{I}_{S_1}(s_1)$  and  $I_2^f(s_2) = \mu_1(\mathbb{I}_{S_1}) = \mu_1(S_1)\mathbb{I}_{S_2}(s_2)$ . Because  $S_1 \in \Sigma_1$ , we have  $I_1^f \in \text{b}\Sigma_1$ . Similarly, because  $S_2 \in \Sigma_2$ , we have  $I_2^f \in \text{b}\Sigma_2$ . In that case,  $f \in \mathcal{H}$ , since

$$\mu_1(I_1^f) = \int_{S_1} \mu_2(S_2)\mathbb{I}_{S_1}(s_1)\mu_1(ds_1) = \mu_1(S_1)\mu_2(S_2) = \int_{S_2} \mu_1(S_1)\mathbb{I}_{S_2}(s_2)\mu_2(ds_2) = \mu_2(I_2^f).$$

Now suppose that  $f \in \mathcal{H}$  and  $a \in \mathbb{R}$ . Note that  $af \in \text{b}\Sigma$ . For every  $s_1 \in S_1$  and  $s_2 \in S_2$ , note that  $I_1^{af}(s_1) = \mu_2(af_{s_1}) = a\mu_2(f_{s_1}) = aI_1^f(s_1)$  and  $I_2^{af}(s_2) = \mu_1(af_{s_2}) = a\mu_1(f_{s_2}) = aI_2^f(s_2)$ . Clearly,  $I_1^{af} \in \text{b}\Sigma_1$  and  $I_2^{af} \in \text{b}\Sigma_2$ . Therefore,  $af \in \mathcal{H}$ , since the fact that  $\mu_1(I_1^f) = \mu_2(I_2^f)$  implies

$$\mu_1(I_1^{af}) = \int_{S_1} aI_1^f(s_1)\mu_1(ds_1) = a\mu_1(I_1^f) = a\mu_2(I_2^f) = \int_{S_2} aI_2^f(s_2)\mu_2(ds_2) = \mu_2(I_2^{af}).$$

Finally, suppose that  $g, h \in \mathcal{H}$ . Note that  $g + h \in \text{b}\Sigma$ . For every  $s_1 \in S_1$  and  $s_2 \in S_2$ , note that  $I_1^{g+h}(s_1) = \mu_2(g_{s_1} + h_{s_1}) = \mu_2(g_{s_1}) + \mu_2(h_{s_1}) = I_1^g(s_1) + I_1^h(s_1)$  and  $I_2^{g+h}(s_2) = \mu_1(g_{s_2} + h_{s_2}) = \mu_1(g_{s_2}) + \mu_1(h_{s_2}) = I_2^g(s_2) + I_2^h(s_2)$ . Clearly,  $I_1^{g+h} \in \text{b}\Sigma_1$  and  $I_2^{g+h} \in \text{b}\Sigma_2$ . Therefore,  $g + h \in \mathcal{H}$ , since  $\mu_1(I_1^g) = \mu_2(I_2^g)$  and  $\mu_1(I_1^h) = \mu_2(I_2^h)$  imply

$$\int_{S_1} [I_1^g(s_1) + I_1^h(s_1)] \mu_1(ds_1) = \mu_1(I_1^g) + \mu_1(I_1^h) = \mu_2(I_2^g) + \mu_2(I_2^h) = \int_{S_2} [I_2^g(s_2) + I_2^h(s_2)] \mu_2(ds_2).$$

Suppose  $(f_n \in \mathcal{H} \mid n \in \mathbb{N})$  is a sequence of non-negative functions in  $\mathcal{H}$  such that  $f_n \uparrow f$ , where  $f : S \rightarrow [0, \infty)$  is a bounded function. Note that  $f \in \text{b}\Sigma$ , since  $f$  is the limit of a sequence of (bounded)  $\Sigma$ -measurable functions.

For every  $s_1 \in S_1$  and  $s_2 \in S_2$ , note that  $f_{n,s_1} \uparrow f_{s_1}$  and  $f_{n,s_2} \uparrow f_{s_2}$ , so that the monotone-convergence theorem implies that  $\mu_2(f_{n,s_1}) \uparrow \mu_2(f_{s_1})$  and  $\mu_1(f_{n,s_2}) \uparrow \mu_1(f_{s_2})$ . Therefore,

$$\begin{aligned} I_1^f(s_1) &= \mu_2(f_{s_1}) = \lim_{n \rightarrow \infty} \mu_2(f_{n,s_1}) = \lim_{n \rightarrow \infty} I_1^{f_n}(s_1), \\ I_2^f(s_2) &= \mu_1(f_{s_2}) = \lim_{n \rightarrow \infty} \mu_1(f_{n,s_2}) = \lim_{n \rightarrow \infty} I_2^{f_n}(s_2). \end{aligned}$$

Because  $I_1^f$  is the limit of (bounded)  $\Sigma_1$ -measurable functions,  $I_1^f \in \text{b}\Sigma_1$ . Similarly, because  $I_2^f$  is the limit of (bounded)  $\Sigma_2$ -measurable functions,  $I_2^f \in \text{b}\Sigma_2$ . Furthermore,  $I_1^{f_n} \uparrow I_1^f$  and  $I_2^{f_n} \uparrow I_2^f$ , since  $f_{n+1} \geq f_n$  implies

$$\begin{aligned} I_1^{f_{n+1}}(s_1) &= \mu_2(f_{n+1,s_1}) \geq \mu_2(f_{n,s_1}) = I_1^{f_n}(s_1), \\ I_2^{f_{n+1}}(s_2) &= \mu_1(f_{n+1,s_2}) \geq \mu_1(f_{n,s_2}) = I_2^{f_n}(s_2). \end{aligned}$$

Therefore,  $f \in \mathcal{H}$ , since the monotone-convergence theorem implies that

$$\mu_1(I_1^f) = \lim_{n \rightarrow \infty} \mu_1(I_1^{f_n}) = \lim_{n \rightarrow \infty} \mu_2(I_2^{f_n}) = \mu_2(I_2^f).$$

Consider the  $\pi$ -system  $\mathcal{I} = \{B_1 \times B_2 \mid B_1 \in \Sigma_1 \text{ and } B_2 \in \Sigma_2\}$  and the indicator function  $f = \mathbb{I}_{B_1 \times B_2}$  of a set  $B_1 \times B_2 \in \mathcal{I}$ . Note that  $f$  is a bounded  $\Sigma$ -measurable function, since  $B_1 \times B_2 \in \Sigma$ . For every  $s_1 \in S_1$  and  $s_2 \in S_2$ , note that  $I_1^f(s_1) = \mu_2(\mathbb{I}_{B_1}(s_1)\mathbb{I}_{B_2}) = \mathbb{I}_{B_1}(s_1)\mu_2(B_2)$  and  $I_2^f(s_2) = \mu_1(\mathbb{I}_{B_2}(s_2)\mathbb{I}_{B_1}) = \mathbb{I}_{B_2}(s_2)\mu_1(B_1)$ . Clearly,  $I_1^f \in \text{b}\Sigma_1$  and  $I_2^f \in \text{b}\Sigma_2$ . Therefore,  $f \in \mathcal{H}$ , since

$$\mu_1(I_1^f) = \mu_1(\mu_2(B_2)\mathbb{I}_{B_1}) = \mu_1(B_1)\mu_2(B_2) = \mu_2(\mu_1(B_1)\mathbb{I}_{B_2}) = \mu_2(I_2^f).$$

Because  $\sigma(\mathcal{I}) = \Sigma$ , the monotone-class theorem completes the proof.

Consider a measure space  $(S_1, \Sigma_1, \mu_1)$ , a measure space  $(S_2, \Sigma_2, \mu_2)$ , and the measurable space  $(S, \Sigma)$ , where  $S = S_1 \times S_2$  and  $\Sigma = \Sigma_1 \times \Sigma_2$ . Furthermore, suppose  $\mu_1$  and  $\mu_2$  are finite measures. For any  $F \in \Sigma$ , define  $\mu(F)$  by

$$\mu(F) = \mu_1(I_1^{\mathbb{I}_F}) = \int_{S_1} I_1^{\mathbb{I}_F}(s_1)\mu_1(ds_1) = \int_{S_2} I_2^{\mathbb{I}_F}(s_2)\mu_2(ds_2) = \mu_2(I_2^{\mathbb{I}_F}).$$

We will now show that  $\mu$  is a measure on  $(S, \Sigma)$ . The measure  $\mu$  is called the product measure of  $\mu_1$  and  $\mu_2$  and denoted by  $\mu = \mu_1 \times \mu_2$ . The measure space  $(S, \Sigma, \mu)$  is denoted by  $(S, \Sigma, \mu) = (S_1, \Sigma_1, \mu_1) \times (S_2, \Sigma_2, \mu_2)$ .

Consider the  $\pi$ -system  $\mathcal{I} = \{B_1 \times B_2 \mid B_1 \in \Sigma_1 \text{ and } B_2 \in \Sigma_2\}$ , the indicator function  $f = \mathbb{I}_{B_1 \times B_2}$  of a set  $B_1 \times B_2 \in \mathcal{I}$ , and recall that  $\mu_1(I_1^f) = \mu_1(B_1)\mu_2(B_2) = \mu_2(I_2^f)$ . Therefore,  $\mu(\emptyset) = \mu_1(\emptyset)\mu_2(\emptyset) = 0$ .

Consider a sequence  $(F_n \in \Sigma \mid n \in \mathbb{N})$  such that  $F_n \cap F_m = \emptyset$  for  $n \neq m$ . Furthermore, consider the sequence of non-negative (bounded)  $\Sigma$ -measurable functions  $(f_n : S \rightarrow \{0, 1\} \mid n \in \mathbb{N})$  given by

$$f_n = \mathbb{I}_{\bigcup_{k=0}^n F_k} = \sum_{k=0}^n \mathbb{I}_{F_k}.$$

Let  $f = \mathbb{I}_{\bigcup_k F_k}$  so that  $f_n \uparrow f$ . Because  $f$  is a bounded function,

$$\mu\left(\bigcup_k F_k\right) = \mu_1(I_1^f) = \lim_{n \rightarrow \infty} \mu_1(I_1^{f_n}) = \lim_{n \rightarrow \infty} \mu_2(I_2^{f_n}) = \mu_2(I_2^f).$$

By the linearity of the integral with respect to  $\mu_2$ ,

$$I_1^{f_n}(s_1) = \int_{S_2} \sum_{k=0}^n \mathbb{I}_{F_k}(s_1, s_2) \mu_2(ds_2) = \sum_{k=0}^n \int_{S_2} \mathbb{I}_{F_k}(s_1, s_2) \mu_2(ds_2) = \sum_{k=0}^n I_1^{\mathbb{I}_{F_k}}(s_1).$$

By the linearity of the integral with respect to  $\mu_1$ ,

$$\mu\left(\bigcup_k F_k\right) = \lim_{n \rightarrow \infty} \mu_1(I_1^{f_n}) = \lim_{n \rightarrow \infty} \int_{S_1} \sum_{k=0}^n I_1^{\mathbb{I}_{F_k}}(s_1) \mu_1(ds_1) = \lim_{n \rightarrow \infty} \sum_{k=0}^n \int_{S_1} I_1^{\mathbb{I}_{F_k}}(s_1) \mu_1(ds_1) = \sum_k \mu(F_k),$$

which completes the proof that  $\mu$  is a measure on  $(S, \Sigma)$ . The measure  $\mu$  is also finite since  $\mu(S_1 \times S_2) = \mu_1(S_1)\mu_2(S_2)$ .

Notably,  $\mu$  is the unique measure on  $(S, \Sigma)$  such that  $\mu(B_1 \times B_2) = \mu_1(B_1)\mu_2(B_2)$  for every  $B_1 \in \Sigma_1$  and  $B_2 \in \Sigma_2$ , since  $\mathcal{I}$  is a  $\pi$ -system on  $S$  such that  $\sigma(\mathcal{I}) = \Sigma$  and  $\mu$  is a finite measure on  $(S, \Sigma)$ .

We will now show that if  $f : S \rightarrow \mathbb{R}$  is a bounded  $\Sigma$ -measurable function, then

$$\mu(f) = \mu_1(I_1^f) = \int_{S_1} I_1^f(s_1) \mu_1(ds_1) = \int_{S_2} I_2^f(s_2) \mu_2(ds_2) = \mu_2(I_2^f).$$

Let  $\mathcal{H}$  denote a set that contains exactly each function  $f \in b\Sigma$  such that  $\mu(f) = \mu_1(I_1^f) = \mu_2(I_2^f)$ .

Consider the  $\pi$ -system  $\mathcal{I} = \{B_1 \times B_2 \mid B_1 \in \Sigma_1 \text{ and } B_2 \in \Sigma_2\}$ . Suppose that  $f = \mathbb{I}_{B_1 \times B_2}$  is the indicator function of a set  $B_1 \times B_2 \in \mathcal{I}$ . In that case,  $\mu(f) = \mu(B_1 \times B_2) = \mu_1(I_1^f) = \mu_2(I_2^f)$ , so that  $f \in \mathcal{H}$ . In particular,  $\mathbb{I}_S \in \mathcal{H}$ , since  $S_1 \times S_2 \in \mathcal{I}$ .

Because  $\mathcal{H} \subseteq b\Sigma$  and  $\mathcal{H}$  is non-empty, showing that  $\mathcal{H}$  is a vector space only requires showing that  $\mathcal{H}$  is closed under scalar multiplication and addition.

Suppose that  $f \in \mathcal{H}$  and  $a \in \mathbb{R}$ . Note that  $af \in b\Sigma$  and  $af \in \mathcal{L}^1(S, \Sigma, \mu)$ , so that  $\mu(af) = a\mu(f)$ . Because  $f \in \mathcal{H}$ , we have  $\mu(af) = \mu_1(aI_1^f) = \mu_1(I_1^{af})$  and  $\mu(af) = \mu_2(aI_2^f) = \mu_2(I_2^{af})$ , so that  $af \in \mathcal{H}$ .

Now suppose that  $g, h \in \mathcal{H}$ . Note that  $g + h \in b\Sigma$  and  $g + h \in \mathcal{L}^1(S, \Sigma, \mu)$ , so that  $\mu(g + h) = \mu(g) + \mu(h)$ . Because  $g, h \in \mathcal{H}$ , we have  $\mu(g + h) = \mu_1(I_1^g + I_1^h) = \mu_1(I_1^{g+h})$  and  $\mu(g + h) = \mu_2(I_2^g + I_2^h) = \mu_2(I_2^{g+h})$ , so that  $g + h \in \mathcal{H}$ .

Finally, suppose  $(f_n \in \mathcal{H} \mid n \in \mathbb{N})$  is a sequence of non-negative functions in  $\mathcal{H}$  such that  $f_n \uparrow f$ , where  $f : S \rightarrow [0, \infty)$  is a bounded function. By the monotone-convergence theorem,  $\mu(f_n) \uparrow \mu(f)$ . Since  $f_n \in \mathcal{H}$ ,

$$\mu(f) = \lim_{n \rightarrow \infty} \mu(f_n) = \lim_{n \rightarrow \infty} \mu_1(I_1^{f_n}) = \lim_{n \rightarrow \infty} \mu_2(I_2^{f_n}) = \mu_1(I_1^f) = \mu_2(I_2^f),$$

which implies  $f \in \mathcal{H}$ . Because  $\sigma(\mathcal{I}) = \Sigma$ , the monotone-class theorem completes the proof.

We will now show that if  $f : S \rightarrow [0, \infty]$  is a  $\Sigma$ -measurable function, then

$$\mu(f) = \mu_1(I_1^f) = \int_{S_1} I_1^f(s_1) \mu_1(ds_1) = \int_{S_2} I_2^f(s_2) \mu_2(ds_2) = \mu_2(I_2^f),$$

where the  $\Sigma_1$ -measurable function  $I_1^f : S_1 \rightarrow [0, \infty]$  and the  $\Sigma_2$ -measurable function  $I_2^f : S_2 \rightarrow [0, \infty]$  are given by

$$\begin{aligned} I_1^f(s_1) &= \int_{S_2} f(s_1, s_2) \mu_2(ds_2) = \int_{S_2} f_{s_1}(s_2) \mu_2(ds_2) = \mu_2(f_{s_1}), \\ I_2^f(s_2) &= \int_{S_1} f(s_1, s_2) \mu_1(ds_1) = \int_{S_1} f_{s_2}(s_1) \mu_1(ds_1) = \mu_1(f_{s_2}), \end{aligned}$$

where  $f_{s_1} : S_2 \rightarrow [0, \infty]$  is a  $\Sigma_2$ -measurable function,  $f_{s_2} : S_1 \rightarrow [0, \infty]$  is a  $\Sigma_1$ -measurable function, and  $f(s_1, s_2) = f_{s_1}(s_2) = f_{s_2}(s_1)$ , for every  $s_1 \in S_1$  and  $s_2 \in S_2$ .

For any  $n \in \mathbb{N}$ , let  $f_n = \alpha_n \circ f$ , where  $\alpha_n$  is the  $n$ -th staircase function. Because  $f_n : S \rightarrow [0, n]$  is bounded and  $\Sigma$ -measurable, there is a bounded  $\Sigma_2$ -measurable function  $f_{n,s_1} : S_2 \rightarrow [0, n]$  and a bounded  $\Sigma_1$ -measurable function  $f_{n,s_2} : S_1 \rightarrow [0, n]$  such that  $f_n(s_1, s_2) = f_{n,s_1}(s_2) = f_{n,s_2}(s_1)$  for every  $s_1 \in S_1$  and  $s_2 \in S_2$ . Since  $f_n \uparrow f$ , consider the  $\Sigma_2$ -measurable function  $f_{s_1} = \lim_{n \rightarrow \infty} f_{n,s_1}$  and the  $\Sigma_1$ -measurable function  $f_{s_2} = \lim_{n \rightarrow \infty} f_{n,s_2}$ . Note that  $f(s_1, s_2) = f_{s_1}(s_2) = f_{s_2}(s_1)$ .

For every  $s_1 \in S_1$  and  $s_2 \in S_2$ , note that  $f_{n,s_1} \uparrow f_{s_1}$  and  $f_{n,s_2} \uparrow f_{s_2}$ , so that the monotone-convergence theorem implies that  $\mu_2(f_{n,s_1}) \uparrow \mu_2(f_{s_1})$  and  $\mu_1(f_{n,s_2}) \uparrow \mu_1(f_{s_2})$ . Therefore,

$$\begin{aligned} I_1^f(s_1) &= \mu_2(f_{s_1}) = \lim_{n \rightarrow \infty} \mu_2(f_{n,s_1}) = \lim_{n \rightarrow \infty} I_1^{f_n}(s_1), \\ I_2^f(s_2) &= \mu_1(f_{s_2}) = \lim_{n \rightarrow \infty} \mu_1(f_{n,s_2}) = \lim_{n \rightarrow \infty} I_2^{f_n}(s_2). \end{aligned}$$

Since  $f_n \in \text{b}\Sigma$ , recall that  $I_1^{f_n} \in \text{b}\Sigma_1$  and  $I_2^{f_n} \in \text{b}\Sigma_2$ . Because  $I_1^f$  is the limit of  $\Sigma_1$ -measurable functions,  $I_1^f \in \text{m}\Sigma_1$ . Similarly, because  $I_2^f$  is the limit of  $\Sigma_2$ -measurable functions,  $I_2^f \in \text{m}\Sigma_2$ . Furthermore,  $I_1^{f_n} \uparrow I_1^f$  and  $I_2^{f_n} \uparrow I_2^f$ , since  $f_{n+1} \geq f_n$  implies

$$\begin{aligned} I_1^{f_{n+1}}(s_1) &= \mu_2(f_{n+1,s_1}) \geq \mu_2(f_{n,s_1}) = I_1^{f_n}(s_1), \\ I_2^{f_{n+1}}(s_2) &= \mu_1(f_{n+1,s_2}) \geq \mu_1(f_{n,s_2}) = I_2^{f_n}(s_2). \end{aligned}$$

Because  $f_n \uparrow f$ , the monotone-convergence theorem implies that  $\mu(f_n) \uparrow \mu(f)$ . Because  $I_1^{f_n} \uparrow I_1^f$  and  $I_2^{f_n} \uparrow I_2^f$ , the monotone-convergence theorem implies that  $\mu_1(I_1^{f_n}) \uparrow \mu_1(I_1^f)$  and  $\mu_2(I_2^{f_n}) \uparrow \mu_2(I_2^f)$ . Because  $f_n \in \text{b}\Sigma$ ,

$$\mu(f) = \lim_{n \rightarrow \infty} \mu(f_n) = \lim_{n \rightarrow \infty} \mu_1(I_1^{f_n}) = \mu_1(I_1^f) = \lim_{n \rightarrow \infty} \mu_2(I_2^{f_n}) = \mu_2(I_2^f).$$

Consider the measure space  $(S, \Sigma, \mu) = (S_1, \Sigma_1, \mu_1) \times (S_2, \Sigma_2, \mu_2)$ , where  $\mu_1$  and  $\mu_2$  are finite measures. Consider also a function  $f \in \mathcal{L}^1(S, \Sigma, \mu)$ , and recall that  $f = f^+ - f^-$  and  $|f| = f^+ + f^-$ , where  $f^+ : S \rightarrow [0, \infty]$  and  $f^- : S \rightarrow [0, \infty]$  are non-negative  $\Sigma$ -measurable functions. Therefore, for every  $s_1 \in S_1$  and  $s_2 \in S_2$ ,

$$\begin{aligned} f(s_1, s_2) &= f^+(s_1, s_2) - f^-(s_1, s_2) = f_{s_1}^+(s_2) - f_{s_1}^-(s_2) = f_{s_2}^+(s_1) - f_{s_2}^-(s_1), \\ |f(s_1, s_2)| &= f^+(s_1, s_2) + f^-(s_1, s_2) = f_{s_1}^+(s_2) + f_{s_1}^-(s_2) = f_{s_2}^+(s_1) + f_{s_2}^-(s_1), \end{aligned}$$

where  $f_{s_1}^+ : S_2 \rightarrow [0, \infty]$  and  $f_{s_1}^- : S_2 \rightarrow [0, \infty]$  are non-negative  $\Sigma_2$ -measurable functions and  $f_{s_2}^+ : S_1 \rightarrow [0, \infty]$  and  $f_{s_2}^- : S_1 \rightarrow [0, \infty]$  are non-negative  $\Sigma_1$ -measurable functions.

For every  $s_1 \in S_1$  and  $s_2 \in S_2$ , let  $f_{s_1} = f_{s_1}^+ - f_{s_1}^-$  and  $f_{s_2} = f_{s_2}^+ - f_{s_2}^-$ , so that  $f(s_1, s_2) = f_{s_1}(s_2) = f_{s_2}(s_1)$ . Note that  $f_{s_1}$  is  $\Sigma_2$ -measurable and  $f_{s_2}$  is  $\Sigma_1$ -measurable. Furthermore,  $|f_{s_1}| = f_{s_1}^+ + f_{s_1}^-$  and  $|f_{s_2}| = f_{s_2}^+ + f_{s_2}^-$ .

Finally, let  $F_1^f = \{s_1 \in S_1 \mid \mu_2(|f_{s_1}|) < \infty\}$  and  $F_2^f = \{s_2 \in S_2 \mid \mu_1(|f_{s_2}|) < \infty\}$ . We will now show that

$$\mu(f) = \mu_1(I_1^f; F_1^f) = \int_{F_1^f} I_1^f(s_1) \mu_1(ds_1) = \int_{F_2^f} I_2^f(s_2) \mu_2(ds_2) = \mu_2(I_2^f; F_2^f),$$

where  $I_1^f : S_1 \rightarrow \mathbb{R}$  and  $I_2^f : S_2 \rightarrow \mathbb{R}$  are given by

$$\begin{aligned} I_1^f(s_1) &= \int_{S_2} f(s_1, s_2) \mu_2(ds_2) = \int_{S_2} f_{s_1}(s_2) \mu_2(ds_2) = \mu_2(f_{s_1}), \\ I_2^f(s_2) &= \int_{S_1} f(s_1, s_2) \mu_1(ds_1) = \int_{S_1} f_{s_2}(s_1) \mu_1(ds_1) = \mu_1(f_{s_2}), \end{aligned}$$

for every  $s_1 \in F_1^f$  and  $s_2 \in F_2^f$ .

Because  $|f| : S \rightarrow [0, \infty]$  is a non-negative  $\Sigma$ -measurable function such that  $\mu(|f|) < \infty$ ,

$$\begin{aligned} \mu(|f|) &= \mu_1(I_1^{|f|}) = \mu_1(I_1^{f^+ + f^-}) = \mu_1(I_1^{f^+} + I_1^{f^-}) < \infty, \\ \mu(|f|) &= \mu_2(I_2^{|f|}) = \mu_2(I_2^{f^+ + f^-}) = \mu_2(I_2^{f^+} + I_2^{f^-}) < \infty. \end{aligned}$$

For every  $s_1 \in S_1$ , note that  $I_1^{f^+}(s_1) + I_1^{f^-}(s_1) = \mu_2(f_{s_1}^+) + \mu_2(f_{s_1}^-) = \mu_2(|f_{s_1}|)$ . Because  $\mu_1(I_1^{f^+} + I_1^{f^-}) < \infty$ , we know that  $\mu_1(S_1 \setminus F_1^f) = \mu_1(\{s_1 \in S_1 \mid \mu_2(|f_{s_1}|) = \infty\}) = 0$ . Similarly, for every  $s_2 \in S_2$ , note that  $I_2^{f^+}(s_2) + I_2^{f^-}(s_2) = \mu_1(f_{s_2}^+) + \mu_1(f_{s_2}^-) = \mu_1(|f_{s_2}|)$ . Because  $\mu_2(I_2^{f^+} + I_2^{f^-}) < \infty$ , we know that  $\mu_2(S_2 \setminus F_2^f) = \mu_2(\{s_2 \in S_2 \mid \mu_1(|f_{s_2}|) = \infty\}) = 0$ . Therefore, by the linearity of the integral,

$$\begin{aligned}\mu(f) &= \mu(f^+) - \mu(f^-) = \mu_1(I_1^{f^+}) - \mu_1(I_1^{f^-}) = \mu_1(I_1^{f^+} \mathbb{I}_{F_1^f}) - \mu_1(I_1^{f^-} \mathbb{I}_{F_1^f}) = \mu_1((I_1^{f^+} - I_1^{f^-}) \mathbb{I}_{F_1^f}) = \mu_1(I_1^f; F_1^f), \\ \mu(f) &= \mu(f^+) - \mu(f^-) = \mu_2(I_2^{f^+}) - \mu_2(I_2^{f^-}) = \mu_2(I_2^{f^+} \mathbb{I}_{F_2^f}) - \mu_2(I_2^{f^-} \mathbb{I}_{F_2^f}) = \mu_2((I_2^{f^+} - I_2^{f^-}) \mathbb{I}_{F_2^f}) = \mu_2(I_2^f; F_2^f).\end{aligned}$$

The previous result is also valid when  $\mu_1$  and  $\mu_2$  are  $\sigma$ -finite measures.

Consider the measure space  $(S, \Sigma, \mu) = (\Omega, \mathcal{F}, \mathbb{P}) \times ([0, \infty), \mathcal{B}([0, \infty)), \text{Leb})$ , where  $(\Omega, \mathcal{F}, \mathbb{P})$  is a probability triple. Furthermore, consider a random variable  $X : \Omega \rightarrow [0, \infty]$ . We will now show that

$$\mathbb{E}(X) = \int_{[0, \infty)} \mathbb{P}(X \geq x) \text{Leb}(dx).$$

First, let  $A = \{(\omega, x) \in S \mid x \leq X(\omega)\}$  and  $f(\omega, x) = x - X(\omega) = \rho_2(\omega, x) - X(\rho_1(\omega, x))$ . Because  $f$  is  $\Sigma$ -measurable and  $f^{-1}((-\infty, 0]) = A$ , we know that  $A \in \Sigma$ . For every  $(\omega, x) \in S$ , note that

$$\mathbb{I}_A(\omega, x) = \mathbb{I}_{\{\omega \in \Omega \mid x \leq X(\omega)\}}(\omega) = \mathbb{I}_{\{x \in [0, \infty) \mid x \leq X(\omega)\}}(x).$$

Because  $\mathbb{I}_A$  is a bounded  $\Sigma$ -measurable function,

$$\begin{aligned}I_1^{\mathbb{I}_A}(\omega) &= \text{Leb}(\{x \in [0, \infty) \mid x \leq X(\omega)\}) = X(\omega), \\ I_2^{\mathbb{I}_A}(x) &= \mathbb{P}(\{\omega \in \Omega \mid x \leq X(\omega)\}) = \mathbb{P}(X \geq x).\end{aligned}$$

By the definition of the product measure  $\mu$ ,

$$\mu(A) = \mathbb{P}(I_1^{\mathbb{I}_A}) = \mathbb{E}(X) = \text{Leb}(I_2^{\mathbb{I}_A}) = \int_{[0, \infty)} P(X \geq x) \text{Leb}(dx).$$

Let  $\mathcal{C}$  denote the set of open subsets of  $\mathbb{R}^2$ . The Borel  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R}^2)$  on  $\mathbb{R}^2$  is defined as  $\mathcal{B}(\mathbb{R}^2) = \sigma(\mathcal{C})$ . We will now show that  $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R})^2$ , where  $\mathcal{B}(\mathbb{R})^2$  is the product between the Borel  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R})$  on  $\mathbb{R}$  and itself.

Because the functions  $\rho_1 : \mathbb{R}^2 \rightarrow \mathbb{R}$  and  $\rho_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$  given by  $\rho_1(x, y) = x$  and  $\rho_2(x, y) = y$  for every  $(x, y) \in \mathbb{R}^2$  are continuous, recall that  $\rho_1^{-1}(A) \in \mathcal{C}$  and  $\rho_2^{-1}(A) \in \mathcal{C}$  for every open set  $A \subseteq \mathbb{R}$ , so that a previous result guarantees that  $\rho_1$  and  $\rho_2$  are  $\mathcal{B}(\mathbb{R}^2)$ -measurable. Therefore,  $\sigma(\rho_1) \cup \sigma(\rho_2) \subseteq \mathcal{B}(\mathbb{R}^2)$ . Because  $\mathcal{B}(\mathbb{R})^2 = \sigma(\sigma(\rho_1) \cup \sigma(\rho_2))$ , we know that  $\mathcal{B}(\mathbb{R})^2 \subseteq \mathcal{B}(\mathbb{R}^2)$ .

Recall that every open subset  $C \subseteq \mathbb{R}^2$  can be written as  $C = \bigcup_n (a_n, b_n) \times (c_n, d_n)$ , where  $a_n \leq b_n$  and  $c_n \leq d_n$  for every  $n \in \mathbb{N}$ . Because  $\mathcal{B}(\mathbb{R})$  contains every open interval and  $\mathcal{B}(\mathbb{R})^2 = \sigma(\{B_1 \times B_2 \mid B_1, B_2 \in \mathcal{B}(\mathbb{R})\})$ , we know that  $\mathcal{C} \subseteq \mathcal{B}(\mathbb{R})^2$ , so that  $\mathcal{B}(\mathbb{R}^2) \subseteq \mathcal{B}(\mathbb{R})^2$ . Therefore,  $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R})^2$ .

Consider the set  $\mathcal{I} = \{(-\infty, x] \times (-\infty, y] \mid x, y \in \mathbb{R}\}$ . We will now show that  $\mathcal{I}$  is a  $\pi$ -system on  $\mathbb{R}^2$  such that  $\sigma(\mathcal{I}) = \mathcal{B}(\mathbb{R})^2$ , where  $\mathcal{B}(\mathbb{R})^2$  is the product between the Borel  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R})$  on  $\mathbb{R}$  and itself.

Let  $A_1 = (-\infty, x_1] \times (-\infty, y_1]$  and  $A_2 = (-\infty, x_2] \times (-\infty, y_2]$  be elements of  $\mathcal{I}$ . In that case,

$$A_1 \cap A_2 = ((-\infty, x_1] \cap (-\infty, x_2]) \times ((-\infty, y_1] \cap (-\infty, y_2]) = (-\infty, \min(x_1, x_2)] \times (-\infty, \min(y_1, y_2)],$$

so that  $A_1 \cap A_2 \in \mathcal{I}$ . Therefore,  $\mathcal{I}$  is a  $\pi$ -system.

Because  $(-\infty, x] \in \mathcal{B}(\mathbb{R})$  and  $(-\infty, y] \in \mathcal{B}(\mathbb{R})$  for every  $x, y \in \mathbb{R}$  and  $\mathcal{B}(\mathbb{R})^2 = \sigma(\{B_1 \times B_2 \mid B_1, B_2 \in \mathcal{B}(\mathbb{R})\})$ , we know that  $\mathcal{I} \subseteq \mathcal{B}(\mathbb{R})^2$ , so that  $\sigma(\mathcal{I}) \subseteq \mathcal{B}(\mathbb{R})^2$ .

Note that  $(a, b] \times (c, d] \in \sigma(\mathcal{I})$  for every  $a \leq b$  and  $c \leq d$ , since

$$(a, b] \times (c, d] = ((-\infty, b] \times (-\infty, d]) \cap (((-\infty, b] \times (-\infty, c]) \cup ((-\infty, a] \times (-\infty, d]))^c.$$

Also note that  $(a, b) \times (c, d] \in \sigma(\mathcal{I})$  for every  $a \leq b$  and  $c \leq d$ , since

$$(a, b) \times (c, d] = \left( \bigcup_{n \in \mathbb{N}^+} (a, b - \epsilon_1 n^{-1}] \right) \times (c, d] = \bigcup_{n \in \mathbb{N}^+} (a, b - \epsilon_1 n^{-1}] \times (c, d],$$

where  $\epsilon_1 = (b - a)/2$ .



Finally, note that  $(a, b) \times (c, d) \in \sigma(\mathcal{I})$  for every  $a \leq b$  and  $c \leq d$ , since

$$(a, b) \times (c, d) = (a, b) \times \bigcup_{n \in \mathbb{N}^+} (c, d - \epsilon_2 n^{-1}] = \bigcup_{n \in \mathbb{N}^+} (a, b) \times (c, d - \epsilon_2 n^{-1}],$$

where  $\epsilon_2 = (d - c)/2$ .

Because every open set  $C \in \mathcal{C}$  can be written as  $C = \bigcup_n (a_n, b_n) \times (c_n, d_n)$ , where  $a_n \leq b_n$  and  $c_n \leq d_n$  for every  $n \in \mathbb{N}$ , we know that  $\mathcal{C} \subseteq \sigma(\mathcal{I})$ . Since  $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R})^2$ , we know that  $\mathcal{B}(\mathbb{R})^2 \subseteq \sigma(\mathcal{I})$ .

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and the random variables  $X : \Omega \rightarrow \mathbb{R}$  and  $Y : \Omega \rightarrow \mathbb{R}$ . Let  $Z : \Omega \rightarrow \mathbb{R}^2$  be given by  $Z(\omega) = (X(\omega), Y(\omega))$ . We will now show that  $Z$  is  $\mathcal{F}/\mathcal{B}(\mathbb{R})^2$ -measurable.

Let  $\rho_1 : \mathbb{R}^2 \rightarrow \mathbb{R}$  be given by  $\rho_1(x, y) = x$  and  $\rho_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$  be given by  $\rho_2(x, y) = y$ . Note that  $X = \rho_1 \circ Z$  and  $Y = \rho_2 \circ Z$ , so that  $X^{-1}(B) = (\rho_1 \circ Z)^{-1}(B) = Z^{-1}(\rho_1^{-1}(B))$  and  $Y^{-1}(B) = (\rho_2 \circ Z)^{-1}(B) = Z^{-1}(\rho_2^{-1}(B))$  for every  $B \in \mathcal{B}(\mathbb{R})$ . Because  $X$  and  $Y$  are  $\mathcal{F}$ -measurable,  $Z^{-1}(C) \in \mathcal{F}$  for every  $C \in (\sigma(\rho_1) \cup \sigma(\rho_2))^2$ .

Note that  $\mathcal{E} = \{\Gamma \in \mathcal{B}(\mathbb{R})^2 \mid Z^{-1}(\Gamma) \in \mathcal{F}\}$  is a  $\sigma$ -algebra on  $\mathbb{R}^2$ . Because  $(\sigma(\rho_1) \cup \sigma(\rho_2))^2 \subseteq \mathcal{B}(\mathbb{R})^2$ , we know that  $\sigma(\sigma(\rho_1) \cup \sigma(\rho_2))^2 \subseteq \mathcal{E}$ , so that  $\mathcal{E} = \mathcal{B}(\mathbb{R})^2$ . Therefore,  $Z$  is  $\mathcal{F}/\mathcal{B}(\mathbb{R})^2$ -measurable.

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and the random variables  $X : \Omega \rightarrow \mathbb{R}$  and  $Y : \Omega \rightarrow \mathbb{R}$ . For any  $\Gamma \in \mathcal{B}(\mathbb{R})^2$ , the joint law  $\mathcal{L}_{X,Y} : \mathcal{B}(\mathbb{R})^2 \rightarrow [0, 1]$  of  $X$  and  $Y$  is defined by

$$\mathcal{L}_{X,Y}(\Gamma) = \mathbb{P}(\{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in \Gamma\}) = \mathbb{P}((X, Y) \in \Gamma).$$

Note that  $\mathcal{L}_{X,Y}$  is a probability measure on  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2)$ . Clearly,  $\mathcal{L}_{X,Y}(\mathbb{R}^2) = \mathbb{P}(\Omega) = 1$  and  $\mathcal{L}_{X,Y}(\emptyset) = \mathbb{P}(\emptyset) = 0$ . Furthermore, for any sequence of sets  $(\Gamma_n \in \mathcal{B}(\mathbb{R})^2 \mid n \in \mathbb{N})$  such that  $\Gamma_n \cap \Gamma_m = \emptyset$  for  $n \neq m$ ,

$$\mathcal{L}_{X,Y} \left( \bigcup_n \Gamma_n \right) = \mathbb{P} \left( \left\{ \omega \in \Omega \mid (X(\omega), Y(\omega)) \in \bigcup_n \Gamma_n \right\} \right) = \mathbb{P} \left( \bigcup_n \{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in \Gamma_n\} \right) = \sum_n \mathcal{L}_{X,Y}(\Gamma_n).$$

The joint distribution  $F_{X,Y} : \mathbb{R}^2 \rightarrow [0, 1]$  of  $X$  and  $Y$  is defined by

$$F_{X,Y}(x, y) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \leq x \text{ and } Y(\omega) \leq y\}) = \mathbb{P}(X \leq x, Y \leq y) = \mathcal{L}_{X,Y}((-\infty, x] \times (-\infty, y]).$$

Because the  $\pi$ -system  $\mathcal{I} = \{(-\infty, x] \times (-\infty, y] \mid x, y \in \mathbb{R}\}$  generates  $\mathcal{B}(\mathbb{R})^2$ , the joint law  $\mathcal{L}_{X,Y}$  of  $X$  and  $Y$  is the unique measure on the measurable space  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2)$  such that  $\mathcal{L}_{X,Y}((-\infty, x] \times (-\infty, y]) = F_{X,Y}(x, y)$  for every  $(x, y) \in \mathbb{R}^2$ . Therefore, the joint distribution  $F_{X,Y}$  completely determines the joint law  $\mathcal{L}_{X,Y}$ .

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and the random variables  $X : \Omega \rightarrow \mathbb{R}$  and  $Y : \Omega \rightarrow \mathbb{R}$ . Consider also the measure space  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2, \text{Leb}^2) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \text{Leb})^2$ . The random variables  $X$  and  $Y$  have a joint probability density function  $f_{X,Y}$  if  $f_{X,Y} : \mathbb{R}^2 \rightarrow [0, \infty]$  is a  $\mathcal{B}(\mathbb{R})^2$ -measurable function such that the joint law  $\mathcal{L}_{X,Y}$  is given by

$$\mathcal{L}_{X,Y}(\Gamma) = \int_{\Gamma} f_{X,Y}(z) \text{Leb}^2(dz) = \int_{\mathbb{R}^2} \mathbb{I}_{\Gamma}(z) f_{X,Y}(z) \text{Leb}^2(dz).$$

In that case, the joint law  $\mathcal{L}_{X,Y}$  has density  $f_{X,Y}$  relative to  $\text{Leb}^2$ , which is denoted by  $d\mathcal{L}_{X,Y}/d\text{Leb}^2 = f_{X,Y}$ . Furthermore, because  $\mathbb{I}_{\Gamma} f_{X,Y}$  is a non-negative  $\mathcal{B}(\mathbb{R})^2$ -measurable function,

$$\mathcal{L}_{X,Y}(\Gamma) = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \mathbb{I}_{\Gamma}(x, y) f_{X,Y}(x, y) \text{Leb}(dy) \right] \text{Leb}(dx) = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \mathbb{I}_{\Gamma}(x, y) f_{X,Y}(x, y) \text{Leb}(dx) \right] \text{Leb}(dy).$$

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and the random variables  $X : \Omega \rightarrow \mathbb{R}$  and  $Y : \Omega \rightarrow \mathbb{R}$ . Note that

$$\begin{aligned} \mathcal{L}_X(B) &= \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in B\}) = \mathbb{P}(\{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in (B \times \mathbb{R})\}) = \mathcal{L}_{X,Y}(B \times \mathbb{R}), \\ \mathcal{L}_Y(B) &= \mathbb{P}(Y^{-1}(B)) = \mathbb{P}(\{\omega \in \Omega \mid Y(\omega) \in B\}) = \mathbb{P}(\{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in (\mathbb{R} \times B)\}) = \mathcal{L}_{X,Y}(\mathbb{R} \times B), \end{aligned}$$

for every  $B \in \mathcal{B}(\mathbb{R})$ , where  $\mathcal{L}_X$  is the law of  $X$  and  $\mathcal{L}_Y$  is the law of  $Y$ . Therefore,

$$\begin{aligned} \mathcal{L}_X(B) &= \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \mathbb{I}_{B \times \mathbb{R}}(x, y) f_{X,Y}(x, y) \text{Leb}(dy) \right] \text{Leb}(dx) = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \mathbb{I}_B(x) f_{X,Y}(x, y) \text{Leb}(dy) \right] \text{Leb}(dx), \\ \mathcal{L}_Y(B) &= \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \mathbb{I}_{\mathbb{R} \times B}(x, y) f_{X,Y}(x, y) \text{Leb}(dx) \right] \text{Leb}(dy) = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \mathbb{I}_B(y) f_{X,Y}(x, y) \text{Leb}(dx) \right] \text{Leb}(dy), \end{aligned}$$

for every  $B \in \mathcal{B}(\mathbb{R})$ . By the linearity of the integral with respect to Leb,

$$\begin{aligned}\mathcal{L}_X(B) &= \int_{\mathbb{R}} \mathbb{I}_B(x) \left[ \int_{\mathbb{R}} f_{X,Y}(x,y) \text{Leb}(dy) \right] \text{Leb}(dx) = \int_{\mathbb{R}} \mathbb{I}_B(x) f_X(x) \text{Leb}(dx) = \int_B f_X(x) \text{Leb}(dx), \\ \mathcal{L}_Y(B) &= \int_{\mathbb{R}} \mathbb{I}_B(y) \left[ \int_{\mathbb{R}} f_{X,Y}(x,y) \text{Leb}(dx) \right] \text{Leb}(dy) = \int_{\mathbb{R}} \mathbb{I}_B(y) f_Y(y) \text{Leb}(dy) = \int_B f_Y(y) \text{Leb}(dy),\end{aligned}$$

where  $f_X : \mathbb{R} \rightarrow [0, \infty]$  and  $f_Y : \mathbb{R} \rightarrow [0, \infty]$  are Borel functions given by

$$\begin{aligned}f_X(x) &= \int_{\mathbb{R}} f_{X,Y}(x,y) \text{Leb}(dy), \\ f_Y(y) &= \int_{\mathbb{R}} f_{X,Y}(x,y) \text{Leb}(dx).\end{aligned}$$

By definition,  $f_X$  is a probability density function for  $X$  and  $f_Y$  is a probability density function for  $Y$ .

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and the random variables  $X : \Omega \rightarrow \mathbb{R}$  and  $Y : \Omega \rightarrow \mathbb{R}$ . Let  $\mathcal{L}_{X,Y}$  denote the joint law of  $X$  and  $Y$ ,  $\mathcal{L}_X$  denote the law of  $X$ ,  $\mathcal{L}_Y$  denote the law of  $Y$ ,  $F_{X,Y}$  denote the joint distribution function of  $X$  and  $Y$ ,  $F_X$  denote the distribution function of  $X$ , and  $F_Y$  denote the distribution function of  $Y$ . We will now show that the following are equivalent:  $X$  and  $Y$  are independent;  $\mathcal{L}_{X,Y} = \mathcal{L}_X \times \mathcal{L}_Y$ ; and  $F_{X,Y} = F_X F_Y$ .

Suppose  $X$  and  $Y$  are independent. In that case, for every  $B_1, B_2 \in \mathcal{B}(\mathbb{R})$ ,

$$\mathcal{L}_{X,Y}(B_1 \times B_2) = \mathbb{P}(\{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in (B_1 \times B_2)\}) = \mathbb{P}(X^{-1}(B_1) \cap Y^{-1}(B_2)) = \mathcal{L}_X(B_1) \mathcal{L}_Y(B_2).$$

Because  $\mathcal{L}_X \times \mathcal{L}_Y$  is the unique measure on  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2)$  such that  $(\mathcal{L}_X \times \mathcal{L}_Y)(B_1 \times B_2) = \mathcal{L}_X(B_1) \mathcal{L}_Y(B_2)$  for every  $B_1, B_2 \in \mathcal{B}(\mathbb{R})$  and  $\mathcal{L}_{X,Y}$  is a measure on  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2)$ , we know that  $\mathcal{L}_{X,Y} = \mathcal{L}_X \times \mathcal{L}_Y$ .

Suppose  $\mathcal{L}_{X,Y} = \mathcal{L}_X \times \mathcal{L}_Y$ . In that case, for every  $x, y \in \mathbb{R}$ ,

$$F_{X,Y}(x, y) = (\mathcal{L}_X \times \mathcal{L}_Y)((-\infty, x] \times (-\infty, y]) = \mathcal{L}_X((-\infty, x]) \mathcal{L}_Y((-\infty, y]) = F_X(x) F_Y(y).$$

Finally, suppose that  $F_{X,Y} = F_X F_Y$ . In that case, for every  $x, y \in \mathbb{R}$ ,

$$\mathbb{P}(X \leq x, Y \leq y) = F_{X,Y}(x, y) = F_X(x) F_Y(y) = \mathbb{P}(X \leq x) \mathbb{P}(Y \leq y),$$

so that a previous result implies that  $X$  and  $Y$  are independent, which completes the proof.

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and the random variables  $X : \Omega \rightarrow \mathbb{R}$  and  $Y : \Omega \rightarrow \mathbb{R}$ . Suppose  $f_{X,Y}$  is a joint probability density function for  $X$  and  $Y$ ,  $f_X$  is a probability density function for  $X$ , and  $f_Y$  is a probability density function for  $Y$ . Furthermore, let  $F = \{(x, y) \in \mathbb{R}^2 \mid f_X(x) f_Y(y) \neq f_{X,Y}(x, y)\}$ . We will now show that  $\text{Leb}^2(F) = 0$  if and only if  $X$  and  $Y$  are independent random variables.

Suppose  $\text{Leb}^2(F) = 0$ . For every  $\Gamma \in \mathcal{B}(\mathbb{R})^2$ , let  $F_\Gamma = \{z \in \mathbb{R}^2 \mid \mathbb{I}_\Gamma(z) f_X(\rho_1(z)) f_Y(\rho_2(z)) \neq \mathbb{I}_\Gamma(z) f_{X,Y}(z)\}$ , so that  $F_\Gamma \subseteq F$ . Because  $F_\Gamma \subseteq F_{\mathbb{R}^2} = F$ , we know that  $\text{Leb}^2(F_\Gamma) = 0$ . Therefore, because  $\mathbb{I}_\Gamma(f_X \circ \rho_1)(f_Y \circ \rho_2)$  and  $\mathbb{I}_\Gamma f_{X,Y}$  are non-negative  $\mathcal{B}(\mathbb{R})^2$ -measurable functions,

$$\mathcal{L}_{X,Y}(\Gamma) = \int_{\mathbb{R}^2} \mathbb{I}_\Gamma(z) f_{X,Y}(z) \text{Leb}^2(dz) = \int_{\mathbb{R}^2} \mathbb{I}_\Gamma(z) f_X(\rho_1(z)) f_Y(\rho_2(z)) \text{Leb}^2(dz).$$

For every  $B_1, B_2 \in \mathcal{B}(\mathbb{R})$ , since  $\mathbb{I}_\Gamma(f_X \circ \rho_1)(f_Y \circ \rho_2)$  is a non-negative  $\mathcal{B}(\mathbb{R})^2$ -measurable function,

$$\mathcal{L}_{X,Y}(B_1 \times B_2) = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \mathbb{I}_{B_1 \times B_2}(x, y) f_X(x) f_Y(y) \text{Leb}(dy) \right] \text{Leb}(dx).$$

Using the fact that  $\mathbb{I}_{B_1 \times B_2}(x, y) = \mathbb{I}_{B_1}(x) \mathbb{I}_{B_2}(y)$  and the linearity of the integral with respect to Leb,

$$\mathcal{L}_{X,Y}(B_1 \times B_2) = \left[ \int_{\mathbb{R}} \mathbb{I}_{B_1}(x) f_X(x) \text{Leb}(dx) \right] \left[ \int_{\mathbb{R}} \mathbb{I}_{B_2}(y) f_Y(y) \text{Leb}(dy) \right] = \mathcal{L}_X(B_1) \mathcal{L}_Y(B_2).$$

Because  $\mathcal{L}_X \times \mathcal{L}_Y$  is the unique measure on  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2)$  such that  $(\mathcal{L}_X \times \mathcal{L}_Y)(B_1 \times B_2) = \mathcal{L}_X(B_1) \mathcal{L}_Y(B_2)$  for every  $B_1, B_2 \in \mathcal{B}(\mathbb{R})$  and  $\mathcal{L}_{X,Y}$  is a measure on  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2)$ , we know that  $X$  and  $Y$  are independent.

Suppose  $X$  and  $Y$  are independent. Let  $f = (f_X \circ \rho_1)(f_Y \circ \rho_2)$ . Because  $f$  is a  $\mathcal{B}(\mathbb{R})^2$ -measurable non-negative function, recall that  $(f \text{Leb}^2)$  is a measure on  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2)$  given by

$$(f \text{Leb}^2)(\Gamma) = \int_{\Gamma} f d\text{Leb}^2 = \int_{\mathbb{R}^2} \mathbb{I}_\Gamma(z) f_X(\rho_1(z)) f_Y(\rho_2(z)) \text{Leb}^2(dz) = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \mathbb{I}_\Gamma(x, y) f_X(x) f_Y(y) \text{Leb}(dy) \right] \text{Leb}(dx).$$

By the linearity of the integral with respect to  $\text{Leb}$ , for every  $B_1, B_2 \in \mathcal{B}(\mathbb{R})$ ,

$$\mathcal{L}_X(B_1)\mathcal{L}_Y(B_2) = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \mathbb{I}_{B_1 \times B_2}(x, y) f_X(x) f_Y(y) \text{Leb}(dy) \right] \text{Leb}(dx) = (f \text{Leb}^2)(B_1 \times B_2).$$

Because  $\mathcal{L}_X \times \mathcal{L}_Y$  is the unique measure on  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2)$  such that  $(\mathcal{L}_X \times \mathcal{L}_Y)(B_1 \times B_2) = \mathcal{L}_X(B_1)\mathcal{L}_Y(B_2)$  for every  $B_1, B_2 \in \mathcal{B}(\mathbb{R})$  and  $(f \text{Leb}^2)$  is a measure on  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2)$ , we know that  $\mathcal{L}_X \times \mathcal{L}_Y = (f \text{Leb}^2)$ . Since  $X$  and  $Y$  are independent,  $\mathcal{L}_{X,Y} = (f \text{Leb}^2)$ . Therefore,  $f$  is a joint probability density function for  $X$  and  $Y$ .

Let  $F_1 = \{z \in \mathbb{R}^2 \mid f(z) - f_{X,Y}(z) > 0\}$  and  $F_2 = \{z \in \mathbb{R}^2 \mid f_{X,Y}(z) - f(z) > 0\}$ , so that  $F = F_1 \cup F_2$ . Since  $F_1 \cap F_2 = \emptyset$ , we have  $\text{Leb}^2(F) = \text{Leb}^2(F_1) + \text{Leb}^2(F_2)$ . In order to find a contradiction, suppose  $\text{Leb}^2(F) > 0$ , so that  $\text{Leb}^2(F_1) > 0$  or  $\text{Leb}^2(F_2) > 0$ . Because  $(f - f_{X,Y})\mathbb{I}_{F_1}$  and  $(f_{X,Y} - f)\mathbb{I}_{F_2}$  are non-negative  $\mathcal{B}(\mathbb{R})^2$ -measurable functions, a previous result then implies that  $\text{Leb}^2((f - f_{X,Y})\mathbb{I}_{F_1}) > 0$  or  $\text{Leb}^2((f_{X,Y} - f)\mathbb{I}_{F_2}) > 0$ . The linearity of the integral with respect to  $\text{Leb}^2$  then implies that  $\mathcal{L}_{X,Y}(F_1) = \text{Leb}^2(f\mathbb{I}_{F_1}) > \text{Leb}^2(f_{X,Y}\mathbb{I}_{F_1}) = \mathcal{L}_{X,Y}(F_1)$  or  $\mathcal{L}_{X,Y}(F_2) = \text{Leb}^2(f_{X,Y}\mathbb{I}_{F_2}) > \text{Leb}^2(f\mathbb{I}_{F_2}) = \mathcal{L}_{X,Y}(F_2)$ , which is a contradiction. Therefore,  $\text{Leb}^2(F) = 0$ .

The results in this section can be generalized to products between any number of measure spaces.

Consider the measurable space  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and a sequence of probability measures  $(\Lambda_n \mid n \in \mathbb{N})$ . Let  $\Omega = \prod_n \mathbb{R}$ , so that each  $\omega \in \Omega$  corresponds to a sequence  $(\omega_n \in \mathbb{R} \mid n \in \mathbb{N})$ . For every  $n \in \mathbb{N}$ , let  $X_n : \Omega \rightarrow \mathbb{R}$  be given by  $X_n(\omega) = \omega_n$ . Furthermore, consider the  $\sigma$ -algebra  $\mathcal{F}$  on  $\Omega$  given by  $\mathcal{F} = \sigma(\cup_n \sigma(X_n))$ . Kolmogorov's extension theorem guarantees that there is a unique probability measure  $\mathbb{P}$  on the measurable space  $(\Omega, \mathcal{F})$  such that, for every sequence  $(B_n \in \mathcal{B}(\mathbb{R}) \mid n \in \mathbb{N})$ ,

$$\mathbb{P} \left( \prod_n B_n \right) = \prod_n \Lambda_n(B_n).$$

The measure space  $(\Omega, \mathcal{F}, \mathbb{P})$  is denoted by  $(\Omega, \mathcal{F}, \mathbb{P}) = \prod_n (\mathbb{R}, \mathcal{B}(\mathbb{R}), \Lambda_n)$ . The sequence  $(X_n : \Omega \rightarrow \mathbb{R} \mid n \in \mathbb{N})$  is composed of independent random variables on  $(\Omega, \mathcal{F}, \mathbb{P})$  so that  $\Lambda_n$  is the law of  $X_n$ .

## 9 Conditional expectation

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a random variable  $X : \Omega \rightarrow \mathbb{R}$ . For every  $\omega \in \Omega$ , note that knowing  $\mathbb{I}_{\{X=x\}}(\omega)$  for every  $x \in \mathbb{R}$  is equivalent to knowing  $X(\omega)$ . Furthermore, from a previous result,

$$\sigma(X) = \left\{ X^{-1} \left( \bigcup_{x \in B} \{x\} \right) \mid B \in \mathcal{B}(\mathbb{R}) \right\} = \left\{ \bigcup_{x \in B} X^{-1}(\{x\}) \mid B \in \mathcal{B}(\mathbb{R}) \right\} = \left\{ \bigcup_{x \in B} \{X = x\} \mid B \in \mathcal{B}(\mathbb{R}) \right\}.$$

Let  $F = \cup_{x \in B} \{X = x\}$  for some  $B \in \mathcal{B}(\mathbb{R})$ . For every  $\omega \in \Omega$ , note that  $\mathbb{I}_F(\omega) = \sum_{x \in B} \mathbb{I}_{\{X=x\}}(\omega)$ , since  $F$  is a union of disjoint sets. Finally, note that  $\{X = x\} \in \sigma(X)$  for every  $x \in \mathbb{R}$ . Therefore, for every  $\omega \in \Omega$ , knowing  $\mathbb{I}_{\{X=x\}}(\omega)$  for every  $x \in \mathbb{R}$  is also equivalent to knowing  $\mathbb{I}_F(\omega)$  for every  $F \in \sigma(X)$ .

In conclusion, for every  $\omega \in \Omega$ , knowing  $X(\omega)$  is equivalent to knowing  $\mathbb{I}_F(\omega)$  for every  $F \in \sigma(X)$ .

More generally, consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a set of random variables  $\{Y_\gamma \mid \gamma \in \mathcal{C}\}$  where  $Y_\gamma : \Omega \rightarrow \mathbb{R}$  for every  $\gamma \in \mathcal{C}$ . Suppose that an unknown outcome  $\omega \in \Omega$  results in a known value  $Y_\gamma(\omega) \in \mathbb{R}$  for every  $\gamma \in \mathcal{C}$ . The  $\sigma$ -algebra  $\sigma(\{Y_\gamma \mid \gamma \in \mathcal{C}\})$  contains exactly each event  $F \in \mathcal{F}$  such that it is possible to state whether  $\omega \in F$ . In other words, for every  $\omega \in \Omega$ , knowing  $Y_\gamma(\omega) \in \mathbb{R}$  for every  $\gamma \in \mathcal{C}$  is equivalent to knowing  $\mathbb{I}_F(\omega)$  for every  $F \in \sigma(\{Y_\gamma \mid \gamma \in \mathcal{C}\})$ .

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and the random variables  $X : \Omega \rightarrow \mathbb{R}$  and  $Y : \Omega \rightarrow \mathbb{R}$ . Suppose  $\sigma(Y) \subseteq \sigma(X)$ . For every  $\omega \in \Omega$ , knowing  $X(\omega)$  allows knowing  $\mathbb{I}_F(\omega)$  for every  $F \in \sigma(Y)$ . Therefore, knowing  $X(\omega)$  allows knowing  $Y(\omega)$ .

In fact, it is possible to show that for every function  $Z : \Omega \rightarrow \mathbb{R}$ , a function  $Y : \Omega \rightarrow \mathbb{R}$  is  $\sigma(Z)$ -measurable if and only if there is a Borel function  $f : \mathbb{R} \rightarrow \mathbb{R}$  such that  $Y = f \circ Z$ . Furthermore, if  $Z_1, Z_2, \dots, Z_n$  are functions from  $\Omega$  to  $\mathbb{R}$ , then a function  $Y : \Omega \rightarrow \mathbb{R}$  is  $\sigma(\{Z_1, Z_2, \dots, Z_n\})$ -measurable if and only if there is a Borel function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $Y(\omega) = f(Z_1(\omega), Z_2(\omega), \dots, Z_n(\omega))$  for every  $\omega \in \Omega$ .

Consider the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ , a random variable  $X : \Omega \rightarrow \mathbb{R}$  such that  $\mathbb{E}(|X|) < \infty$ , and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . A random variable  $Y : \Omega \rightarrow \mathbb{R}$  is called a version of the conditional expectation  $\mathbb{E}(X \mid \mathcal{G})$  of  $X$  given  $\mathcal{G}$  if and only if  $Y$  is  $\mathcal{G}$ -measurable,  $\mathbb{E}(|Y|) < \infty$ , and, for every set  $G \in \mathcal{G}$ ,

$$\int_G Y d\mathbb{P} = \int_G X d\mathbb{P}.$$

In that case, we say that  $Y = \mathbb{E}(X \mid \mathcal{G})$  almost surely. We will now show that a version  $Y$  of the conditional expectation  $\mathbb{E}(X \mid \mathcal{G})$  of  $X$  given  $\mathcal{G}$  always exists. Furthermore, if  $Y$  and  $\tilde{Y}$  are such versions, then  $\mathbb{P}(Y = \tilde{Y}) = 1$ .

First, suppose  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and recall that  $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$  is a complete vector space. Because  $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P}) \subseteq \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ , there is a version  $Y \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$  of the orthogonal projection of  $X$  onto  $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$  such that  $\|X - Y\|_2 = \inf\{\|X - W\|_2 \mid W \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})\}$  and  $\mathbb{E}((X - Y)Z) = 0$ , for every  $Z \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ . Clearly,  $Y$  is  $\mathcal{G}$ -measurable. By the monotonicity of norm,  $\mathbb{E}(|Y|) < \infty$ . For every  $G \in \mathcal{G}$ , we have  $\mathbb{I}_G \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ , so that  $\mathbb{E}((X - Y)\mathbb{I}_G) = 0$ . Therefore, by the linearity of expectation,  $\mathbb{E}(X\mathbb{I}_G) = \mathbb{E}(Y\mathbb{I}_G)$ , which completes this step.

Suppose that  $X$  is a bounded non-negative random variable, so that  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . As an auxiliary step, we will now show that if  $Y = \mathbb{E}(X \mid \mathcal{G})$  almost surely, then  $\mathbb{P}(Y \geq 0) = 1$ . In order to find a contradiction, suppose that  $\mathbb{P}(Y \geq 0) < 1$ , so that  $\mathbb{P}(Y < 0) > 0$ . Let  $A_n = \{Y < -n^{-1}\} = Y^{-1}((-\infty, -n^{-1}))$ , so that  $A_n \subseteq A_{n+1}$  and  $\cup_n A_n = \{Y < 0\}$ . Since  $A_n \uparrow \{Y < 0\}$ , the monotone-convergence property of measure guarantees that  $\mathbb{P}(A_n) \uparrow \mathbb{P}(Y < 0)$ . Because we supposed that  $\mathbb{P}(Y < 0) > 0$ , there is an  $n \in \mathbb{N}$  such that  $\mathbb{P}(A_n) = \mathbb{P}(Y < -n^{-1}) > 0$ . Consider the random variable  $Y\mathbb{I}_{A_n}$  given by

$$(Y\mathbb{I}_{A_n})(\omega) = Y(\omega)\mathbb{I}_{A_n}(\omega) = \begin{cases} Y(\omega), & \text{if } Y(\omega) < -n^{-1}, \\ 0, & \text{if } Y(\omega) \geq -n^{-1}. \end{cases}$$

Because  $Y\mathbb{I}_{A_n} < -n^{-1}\mathbb{I}_{A_n}$ , we know that  $\mathbb{E}(Y\mathbb{I}_{A_n}) \leq -n^{-1}\mathbb{P}(A_n) < 0$ . Because  $X \geq 0$ , we know that  $\mathbb{E}(X\mathbb{I}_{A_n}) \geq 0$ . However,  $A_n \in \mathcal{G}$ , so that  $\mathbb{E}(X\mathbb{I}_{A_n}) = \mathbb{E}(Y\mathbb{I}_{A_n})$ . Because this is a contradiction, we know that  $\mathbb{P}(Y \geq 0) = 1$ .

Next, suppose  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  is non-negative. For every  $n \in \mathbb{N}$ , let  $X_n = \alpha_n \circ X$ , where  $\alpha_n$  is the  $n$ -th staircase function, so that  $X_n \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ . Furthermore, let  $Y_n = \mathbb{E}(X_n \mid \mathcal{G})$  almost surely. Because  $X_n$  is a bounded non-negative random variable, we know that  $\mathbb{P}(Y_n \geq 0) = 1$ . For every  $n \in \mathbb{N}$  and  $G \in \mathcal{G}$ , note that

$$\mathbb{E}((Y_{n+1} - Y_n)\mathbb{I}_G) = \mathbb{E}(Y_{n+1}\mathbb{I}_G) - \mathbb{E}(Y_n\mathbb{I}_G) = \mathbb{E}(X_{n+1}\mathbb{I}_G) - \mathbb{E}(X_n\mathbb{I}_G) = \mathbb{E}((X_{n+1} - X_n)\mathbb{I}_G).$$

Because  $Y_n \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$  and  $Y_{n+1} \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$ , we know that  $Y_{n+1} - Y_n = \mathbb{E}(X_{n+1} - X_n \mid \mathcal{G})$  almost surely. Because  $X_{n+1} - X_n$  is non-negative and bounded for every  $n \in \mathbb{N}$ , we know that  $\mathbb{P}(Y_{n+1} - Y_n \geq 0) = 1$ .

Consider the set  $A^c = \bigcup_n \{Y_n < 0\} \cup \{Y_{n+1} - Y_n < 0\}$ . Note that  $A \in \mathcal{G}$  and  $\mathbb{P}(A) = 1$ , since

$$\mathbb{P}(A^c) = \mathbb{P}\left(\bigcup_n \{Y_n < 0\} \cup \{Y_{n+1} - Y_n < 0\}\right) \leq \sum_n \mathbb{P}(Y_n < 0) + \mathbb{P}(Y_{n+1} - Y_n < 0) = 0.$$

For every  $n \in \mathbb{N}$ , note that  $Y_n\mathbb{I}_A \geq 0$  and  $Y_{n+1}\mathbb{I}_A \geq Y_n\mathbb{I}_A$ . Let  $Y = \limsup_{n \rightarrow \infty} Y_n\mathbb{I}_A$ . For every  $G \in \mathcal{G}$ , because every non-decreasing sequence of real numbers converges (possibly to infinity), we know that  $Y_n\mathbb{I}_A\mathbb{I}_G \uparrow Y\mathbb{I}_G$ . By the monotone-convergence theorem, we know that  $\mathbb{E}(Y_n\mathbb{I}_A\mathbb{I}_G) \uparrow \mathbb{E}(Y\mathbb{I}_G)$ .

For every  $n \in \mathbb{N}$  and  $G \in \mathcal{G}$ , we have  $(A \cap G) \in \mathcal{G}$  and  $\mathbb{P}(X_n\mathbb{I}_G\mathbb{I}_{A^c} \neq 0) = 0$ , so that

$$\mathbb{E}(Y_n\mathbb{I}_A\mathbb{I}_G) = \mathbb{E}(Y_n\mathbb{I}_{A \cap G}) = \mathbb{E}(X_n\mathbb{I}_{A \cap G}) = \mathbb{E}(X_n\mathbb{I}_A\mathbb{I}_G) + \mathbb{E}(X_n\mathbb{I}_{A^c}\mathbb{I}_G) = \mathbb{E}(X_n\mathbb{I}_G),$$

which implies  $\mathbb{E}(X_n\mathbb{I}_G) \uparrow \mathbb{E}(Y\mathbb{I}_G)$ . Since  $X_n\mathbb{I}_G \uparrow X\mathbb{I}_G$ , we also know that  $\mathbb{E}(X_n\mathbb{I}_G) \uparrow \mathbb{E}(X\mathbb{I}_G)$ , so that  $\mathbb{E}(Y\mathbb{I}_G) = \mathbb{E}(X\mathbb{I}_G)$ . Because  $Y$  is  $\mathcal{G}$ -measurable and  $\Omega \in \mathcal{G}$ , we know that  $Y = \mathbb{E}(X \mid \mathcal{G})$  almost surely.

Finally, suppose  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $X = X^+ - X^-$ , where  $X^+ : \Omega \rightarrow [0, \infty]$  and  $X^- : \Omega \rightarrow [0, \infty]$ . Let  $Y^+ = \mathbb{E}(X^+ \mid \mathcal{G})$  almost surely and  $Y^- = \mathbb{E}(X^- \mid \mathcal{G})$  almost surely. For every  $G \in \mathcal{G}$ ,

$$\mathbb{E}(X\mathbb{I}_G) = \mathbb{E}((X^+ - X^-)\mathbb{I}_G) = \mathbb{E}(X^+\mathbb{I}_G) - \mathbb{E}(X^-\mathbb{I}_G) = \mathbb{E}(Y^+\mathbb{I}_G) - \mathbb{E}(Y^-\mathbb{I}_G) = \mathbb{E}((Y^+ - Y^-)\mathbb{I}_G),$$

so that  $Y^+ - Y^- = \mathbb{E}(X \mid \mathcal{G})$  almost surely.

It remains to show that if  $Y = \mathbb{E}(X \mid \mathcal{G})$  almost surely and  $\tilde{Y} = \mathbb{E}(X \mid \mathcal{G})$  almost surely then  $\mathbb{P}(Y = \tilde{Y}) = 1$ . For the purpose of finding a contradiction, suppose that  $\mathbb{P}(Y = \tilde{Y}) < 1$ , so that  $\mathbb{P}(Y \neq \tilde{Y}) > 0$ . In that case,  $\mathbb{P}(Y > \tilde{Y}) + \mathbb{P}(\tilde{Y} > Y) > 0$ , so that  $\mathbb{P}(Y > \tilde{Y}) > 0$  or  $\mathbb{P}(\tilde{Y} > Y) > 0$ . Suppose  $\mathbb{P}(Y > \tilde{Y}) > 0$ . Let  $A_n = \{Y > \tilde{Y} + n^{-1}\} = (Y - \tilde{Y})^{-1}((n^{-1}, \infty))$ , so that  $A_n \subseteq A_{n+1}$  and  $\cup_n A_n = \{Y > \tilde{Y}\}$ . By the monotone-convergence property of measure, we know that  $\mathbb{P}(A_n) \uparrow \mathbb{P}(Y > \tilde{Y})$ . Because  $\mathbb{P}(Y > \tilde{Y}) > 0$ , there is an  $n \in \mathbb{N}$  such that  $\mathbb{P}(A_n) = \mathbb{P}(Y > \tilde{Y} + n^{-1}) > 0$ . Note that  $(Y - \tilde{Y})\mathbb{I}_{A_n} > n^{-1}\mathbb{I}_{A_n}$ , since

$$(Y - \tilde{Y})(\omega)\mathbb{I}_{A_n}(\omega) = \begin{cases} (Y - \tilde{Y})(\omega), & \text{if } (Y - \tilde{Y})(\omega) > n^{-1}, \\ 0, & \text{if } (Y - \tilde{Y})(\omega) \leq n^{-1}. \end{cases}$$

Therefore,  $\mathbb{E}((Y - \tilde{Y})\mathbb{I}_{A_n}) \geq \mathbb{E}(n^{-1}\mathbb{I}_{A_n}) = n^{-1}\mathbb{P}(A_n) > 0$ . However, for every  $G \in \mathcal{G}$ , note that  $\mathbb{E}(Y\mathbb{I}_G) = \mathbb{E}(\tilde{Y}\mathbb{I}_G)$ , so that  $\mathbb{E}((Y - \tilde{Y})\mathbb{I}_G) = 0$ . Because  $A_n \in \mathcal{G}$ , we arrived at a contradiction. An analogous contradiction is found by supposing that  $\mathbb{P}(\tilde{Y} > Y) > 0$ . Therefore,  $\mathbb{P}(Y = \tilde{Y}) = 1$ .

Consider the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ , a random variable  $X : \Omega \rightarrow \mathbb{R}$  such that  $\mathbb{E}(|X|) < \infty$ , and a random variable  $Z : \Omega \rightarrow \mathbb{R}$ . A random variable  $Y : \Omega \rightarrow \mathbb{R}$  is called a version of the conditional expectation  $\mathbb{E}(X | Z)$  of  $X$  given  $Z$  if and only if it is a version of the conditional expectation  $\mathbb{E}(X | \sigma(Z))$  of  $X$  given  $\sigma(Z)$ . An analogous definition applies when  $Z$  is a set of random variables.

Suppose  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $Z : \Omega \rightarrow \mathbb{R}$  are random variables and let  $Y = \mathbb{E}(X | Z)$  almost surely. Recall that for every  $W \in \mathcal{L}^2(\Omega, \sigma(Z), \mathbb{P})$  there is a Borel function  $f : \mathbb{R} \rightarrow \mathbb{R}$  such that  $W = f \circ Z$  and that  $\mathbb{E}((X - Y)^2) \leq \mathbb{E}((X - W)^2)$ . In this sense, if  $g : \mathbb{R} \rightarrow \mathbb{R}$  is a Borel function such that  $Y = g \circ Z$ , then  $Y(\omega) = g(Z(\omega))$  is almost surely the best prediction about  $X(\omega)$  that can be made given  $Z(\omega)$ .

The next three examples illustrate the definition of conditional expectation.

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and the random variables  $X : \Omega \rightarrow \mathcal{X}$  and  $Z : \Omega \rightarrow \mathcal{Z}$ , where  $\mathcal{X} = \{x_1, \dots, x_m\}$  and  $\mathcal{Z} = \{z_1, \dots, z_n\}$ . Furthermore, suppose  $\mathbb{P}(Z = z) > 0$  for every  $z \in \mathcal{Z}$ .

Let  $\mathcal{P}(\mathcal{Z})$  denote the set of all subsets of  $\mathcal{Z}$  and consider the  $\mathcal{P}(\mathcal{Z})$ -measurable function  $E : \mathcal{Z} \rightarrow \mathbb{R}$  given by

$$E(z) = \sum_i x_i \frac{\mathbb{P}(X = x_i, Z = z)}{\mathbb{P}(Z = z)}.$$

We will now show that  $Y = E \circ Z$  is a  $\sigma(Z)$ -measurable function such that

$$\int_G Y d\mathbb{P} = \int_G X d\mathbb{P},$$

for every  $G \in \sigma(Z)$ , so that  $Y = \mathbb{E}(X | Z)$  almost surely.

For every  $B \in \mathcal{B}(\mathbb{R})$ , recall that  $Y^{-1}(B) = Z^{-1}(E^{-1}(B))$ . Because  $E^{-1}(B) \in \mathcal{P}(\mathcal{Z})$  and  $\mathcal{P}(\mathcal{Z}) \subseteq \mathcal{B}(\mathbb{R})$ , we know that  $Y^{-1}(B) \in \sigma(Z)$ . Therefore,  $Y$  is  $\sigma(Z)$ -measurable.

Because  $Y$  is a bounded  $\mathcal{F}$ -measurable function and  $\{Z = z\} \in \mathcal{F}$  for every  $z \in \mathcal{Z}$ ,

$$\int_{\{Z=z\}} Y d\mathbb{P} = \int_{\Omega} \mathbb{I}_{\{Z=z\}}(\omega) E(Z(\omega)) \mathbb{P}(d\omega) = \int_{\Omega} \mathbb{I}_{\{Z=z\}}(\omega) E(z) \mathbb{P}(d\omega) = E(z) \mathbb{P}(Z = z) = \sum_i x_i \mathbb{P}(X = x_i, Z = z).$$

By the definition of the integral of a simple function with respect to  $\mathbb{P}$ ,

$$\int_{\{Z=z\}} Y d\mathbb{P} = \int_{\Omega} \left( \sum_i x_i \mathbb{I}_{\{X=x_i, Z=z\}} \right) d\mathbb{P} = \int_{\Omega} \left( \mathbb{I}_{\{Z=z\}} \sum_i x_i \mathbb{I}_{\{X=x_i\}} \right) d\mathbb{P} = \int_{\Omega} \mathbb{I}_{\{Z=z\}} X d\mathbb{P} = \int_{\{Z=z\}} X d\mathbb{P}.$$

Because  $Z(\omega) \in \mathcal{Z}$  for every  $\omega \in \Omega$  and  $\mathcal{P}(\mathcal{Z}) \subseteq \mathcal{B}(\mathbb{R})$ ,

$$\sigma(Z) = \left\{ \bigcup_{z \in B} \{Z = z\} \mid B \in \mathcal{B}(\mathbb{R}) \right\} = \left\{ \bigcup_{z \in B} \{Z = z\} \mid B \in \mathcal{P}(\mathcal{Z}) \right\}.$$

Let  $G = \bigcup_{z \in B} \{Z = z\}$  for some  $B \in \mathcal{P}(\mathcal{Z})$ . For every  $\omega \in \Omega$ , note that  $\mathbb{I}_G(\omega) = \sum_{z \in B} \mathbb{I}_{\{Z=z\}}(\omega)$ , since  $G$  is a union of disjoint sets. Therefore, because  $Y$  is a bounded  $\mathcal{F}$ -measurable function and  $G \in \mathcal{F}$ ,

$$\int_G Y d\mathbb{P} = \int_{\Omega} \sum_{z \in B} \mathbb{I}_{\{Z=z\}}(\omega) Y(\omega) \mathbb{P}(d\omega) = \sum_{z \in B} \int_{\Omega} \mathbb{I}_{\{Z=z\}}(\omega) Y(\omega) \mathbb{P}(d\omega) = \sum_{z \in B} \int_{\Omega} \mathbb{I}_{\{Z=z\}}(\omega) X(\omega) \mathbb{P}(d\omega).$$

By the linearity of the integral with respect to  $\mathbb{P}$  and the fact that  $\mathbb{I}_G(\omega) = \sum_{z \in B} \mathbb{I}_{\{Z=z\}}(\omega)$ ,

$$\int_G Y d\mathbb{P} = \int_{\Omega} \mathbb{I}_G(\omega) X(\omega) \mathbb{P}(d\omega) = \int_G X d\mathbb{P},$$

which completes the proof.

Consider the probability triple  $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \text{Leb}) \times ([0, 1], \mathcal{B}([0, 1]), \text{Leb})$  and the bounded random variables  $X : \Omega \rightarrow \mathbb{R}$  and  $Z : \Omega \rightarrow [0, 1]$ , where  $Z(a, b) = a$ . Furthermore, consider the bounded  $\mathcal{B}([0, 1])$ -measurable function  $I_1^X : [0, 1] \rightarrow \mathbb{R}$  given by

$$I_1^X(a) = \int_{[0, 1]} X(a, b) \text{Leb}(db).$$

We will now show that  $Y = I_1^X \circ Z$  is a  $\sigma(Z)$ -measurable function such that

$$\int_G Y d\mathbb{P} = \int_G X d\mathbb{P},$$

for every  $G \in \sigma(Z)$ , so that  $Y = \mathbb{E}(X \mid Z)$  almost surely.

Recall that  $\sigma(Z) = \{A \times [0, 1] \mid A \in \mathcal{B}([0, 1])\}$ . For every  $B \in \mathcal{B}(\mathbb{R})$ , note that  $Y^{-1}(B) = Z^{-1}((I_1^X)^{-1}(B))$ . Because  $(I_1^X)^{-1}(B) \in \mathcal{B}([0, 1])$ , we know that  $Y$  is  $\sigma(Z)$ -measurable.

Let  $G = A \times [0, 1]$  for some  $A \in \mathcal{B}([0, 1])$ . Because  $Y$  is a bounded  $\mathcal{F}$ -measurable function and  $G \in \mathcal{F}$ ,

$$\int_G Y d\mathbb{P} = \int_{[0,1]} \left[ \int_{[0,1]} \mathbb{I}_{A \times [0,1]}(a, b) Y(a, b) \text{Leb}(db) \right] \text{Leb}(da) = \int_{[0,1]} \left[ \int_{[0,1]} \mathbb{I}_A(a) I_1^X(a) \text{Leb}(db) \right] \text{Leb}(da).$$

By the linearity of the integral with respect to  $\text{Leb}$  and using the fact that  $\text{Leb}([0, 1]) = 1$ ,

$$\int_G Y d\mathbb{P} = \left[ \int_{[0,1]} \text{Leb}(db) \right] \left[ \int_{[0,1]} \mathbb{I}_A(a) I_1^X(a) \text{Leb}(da) \right] = \int_{[0,1]} \mathbb{I}_A(a) \left[ \int_{[0,1]} X(a, b) \text{Leb}(db) \right] \text{Leb}(da).$$

Therefore, using the fact that  $\mathbb{I}_A(a) = \mathbb{I}_{A \times [0,1]}(a, b) = \mathbb{I}_G(a, b)$ ,

$$\int_G Y d\mathbb{P} = \int_{[0,1]} \left[ \int_{[0,1]} \mathbb{I}_G(a, b) X(a, b) \text{Leb}(db) \right] \text{Leb}(da) = \int_G X d\mathbb{P}.$$

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and the random variables  $X : \Omega \rightarrow \mathbb{R}$  and  $Z : \Omega \rightarrow \mathbb{R}$ . Suppose that  $f_{X,Z} : \mathbb{R}^2 \rightarrow [0, \infty]$  is a joint probability density function for  $X$  and  $Z$ . Let  $f_X : \mathbb{R} \rightarrow [0, \infty]$  be a probability density function for  $X$  and  $f_Z : \mathbb{R} \rightarrow [0, \infty]$  be a probability density function for  $Z$  such that

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{X,Z}(x, z) \text{Leb}(dz), \\ f_Z(z) &= \int_{\mathbb{R}} f_{X,Z}(x, z) \text{Leb}(dx). \end{aligned}$$

Furthermore, consider the elementary conditional probability density function  $f_{X|Z} : \mathbb{R}^2 \rightarrow [0, \infty]$  given by

$$f_{X|Z}(x, z) = \begin{cases} 0, & \text{if } f_Z(z) = 0, \\ f_{X,Z}(x, z)/f_Z(z), & \text{if } 0 < f_Z(z) < \infty, \\ 0, & \text{if } f_Z(z) = \infty. \end{cases}$$

Let  $h : \mathbb{R} \rightarrow \mathbb{R}$  be a Borel function such that  $\mathbb{E}(|h \circ X|) < \infty$ , so that

$$\mathbb{E}(h \circ X) = \int_{\Omega} (h \circ X) d\mathbb{P} = \int_{\mathbb{R}} h d\mathcal{L}_X = \int_{\mathbb{R}} h(x) f_X(x) \text{Leb}(dx),$$

where  $\mathcal{L}_X$  is the law of  $X$ . Finally, consider the function  $g : \mathbb{R} \rightarrow \mathbb{R}$  given by

$$g(z) = \begin{cases} 0, & \text{if } z \notin F_2^g, \\ \int_{\mathbb{R}} h(x) f_{X|Z}(x, z) \text{Leb}(dx), & \text{if } z \in F_2^g, \end{cases}$$

where  $F_2^g = \{z \in \mathbb{R} \mid \int_{\mathbb{R}} |h(x) f_{X|Z}(x, z)| \text{Leb}(dx) < \infty\}$ .

We will now show that  $Y = g \circ Z$  is a  $\sigma(Z)$ -measurable function such that  $\mathbb{E}(|Y|) < \infty$  and

$$\int_G Y d\mathbb{P} = \int_G (h \circ X) d\mathbb{P}$$

for every  $G \in \sigma(Z)$ , so that  $Y = \mathbb{E}((h \circ X) \mid Z)$  almost surely.

First, we will show that  $(h \circ \rho_1)f_{X|Z}$  is  $\mathcal{B}(\mathbb{R})^2$ -measurable. Let  $A_1 = \{z \in \mathbb{R} \mid f_Z(z) > 0\} \cap \{z \in \mathbb{R} \mid f_Z(z) < \infty\}$ . Because  $f_Z$  is Borel, we know that  $\mathbb{R} \times A_1 \in \mathcal{B}(\mathbb{R})^2$ . Furthermore, note that

$$f_{X|Z}(x, z) = \mathbb{I}_{\mathbb{R} \times A_1}(x, z) \frac{f_{X,Z}(x, z)}{f_Z(\rho_2(x, z)) + \mathbb{I}_{\mathbb{R} \times A_1^c}(x, z)}.$$

Because the function  $u : (0, \infty] \rightarrow [0, \infty)$  given by  $u(r) = 1/r$  is Borel, we know that  $f_{X|Z}$  is  $\mathcal{B}(\mathbb{R})^2$ -measurable. Because  $h$  is Borel, we also know that  $(h \circ \rho_1)f_{X|Z}$  is  $\mathcal{B}(\mathbb{R})^2$ -measurable.

We will now show that  $g$  is Borel. Because  $|(h \circ \rho_1)f_{X|Z}|$  is non-negative and  $\mathcal{B}(\mathbb{R})^2$ -measurable, we know that the function  $I_2 : \mathbb{R} \rightarrow [0, \infty]$  given by  $I_2(z) = \int_{\mathbb{R}} |h(x)f_{X|Z}(x, z)| \text{Leb}(dx)$  is Borel, so that  $F_2^g \in \mathcal{B}(\mathbb{R})$ . Furthermore,

$$g(z) = \mathbb{I}_{F_2^g}(z) \int_{\mathbb{R}} ((h \circ \rho_1)f_{X|Z})^+(x, z) \text{Leb}(dx) - \mathbb{I}_{F_2^g}(z) \int_{\mathbb{R}} ((h \circ \rho_1)f_{X|Z})^-(x, z) \text{Leb}(dx).$$

Since  $((h \circ \rho_1)f_{X|Z})^+$  and  $((h \circ \rho_1)f_{X|Z})^-$  are non-negative and  $\mathcal{B}(\mathbb{R})^2$ -measurable, we know that  $g$  is Borel, which also implies that  $Y = g \circ Z$  is a  $\sigma(Z)$ -measurable function.

We will now show that  $\mathbb{E}(|Y|) < \infty$ . Because  $|g(z)| \leq I_2(z)$  for every  $z \in \mathbb{R}$ ,

$$|g(z)|f_Z(z) \leq I_2(z)f_Z(z) = \int_{\mathbb{R}} |h(x)f_{X|Z}(x, z)|f_Z(z) \text{Leb}(dx) = \int_{\mathbb{R}} |h(x)|\mathbb{I}_{A_1}(z)f_{X,Z}(x, z) \text{Leb}(dx).$$

Because  $|g|f_Z$  and  $I_2f_Z$  are non-negative and Borel,

$$\int_{\mathbb{R}} |g(z)|f_Z(z) \text{Leb}(dz) \leq \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} |h(x)|\mathbb{I}_{A_1}(z)f_{X,Z}(x, z) \text{Leb}(dx) \right] \text{Leb}(dz).$$

Because a previous result for probability density functions extends to joint probability density functions,

$$\int_{\mathbb{R}} |g(z)|f_Z(z) \text{Leb}(dz) \leq \int_{\mathbb{R}^2} |h \circ \rho_1|(\mathbb{I}_{A_1} \circ \rho_2)f_{X,Z} d\text{Leb}^2 = \mathbb{E}(|h \circ X| \mathbb{I}_{Z^{-1}(A_1)}) < \infty,$$

since  $(\mathbb{I}_{A_1} \circ Z) = \mathbb{I}_{Z^{-1}(A_1)}$ . Because  $\text{Leb}(|g|f_Z) = \mathbb{E}(|g \circ Z|)$ , we know that  $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ .

Let  $\mathcal{L}_{X,Z} : \mathcal{B}(\mathbb{R})^2 \rightarrow [0, 1]$  denote the joint law of  $X$  and  $Z$ .

We will now show that  $\mathcal{L}_{X,Z}(\mathbb{I}_{\mathbb{R} \times A_1^c}) = 0$ . Because a previous result for laws extends to joint laws,

$$\int_{\mathbb{R}^2} \mathbb{I}_{\mathbb{R} \times A_1^c} d\mathcal{L}_{X,Z} = \int_{\mathbb{R}^2} \mathbb{I}_{\mathbb{R} \times A_1^c} f_{X,Z} d\text{Leb}^2 = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \mathbb{I}_{A_1^c}(z)f_{X,Z}(x, z) \text{Leb}(dx) \right] \text{Leb}(dz).$$

By rearranging terms,

$$\int_{\mathbb{R}^2} \mathbb{I}_{\mathbb{R} \times A_1^c} d\mathcal{L}_{X,Z} = \int_{\mathbb{R}} \mathbb{I}_{A_1^c}(z) \left[ \int_{\mathbb{R}} f_{X,Z}(x, z) \text{Leb}(dx) \right] \text{Leb}(dz) = \int_{\mathbb{R}} \mathbb{I}_{A_1^c}(z)f_Z(z) \text{Leb}(dz).$$

Because  $A_1^c = \{f_Z = 0\} \cup \{f_Z = \infty\}$  is a union of disjoint sets, we know that  $\mathbb{I}_{A_1^c} = \mathbb{I}_{\{f_Z=0\}} + \mathbb{I}_{\{f_Z=\infty\}}$ . Therefore,

$$\int_{\mathbb{R}^2} \mathbb{I}_{\mathbb{R} \times A_1^c} d\mathcal{L}_{X,Z} = \int_{\mathbb{R}} \mathbb{I}_{\{f_Z=0\}}(z)f_Z(z) \text{Leb}(dz) + \int_{\mathbb{R}} \mathbb{I}_{\{f_Z=\infty\}}(z)f_Z(z) \text{Leb}(dz) = 0,$$

since  $\mathbb{I}_{\{f_Z=0\}}f_Z = 0$  and  $\text{Leb}(f_Z) < \infty$ .

Let  $A_2 = \{z \in \mathbb{R} \mid \int_{\mathbb{R}} |h(x)|f_{X,Z}(x, z) \text{Leb}(dx) < \infty\}$ , so that  $A_2 \in \mathcal{B}(\mathbb{R})$ . We will now show that  $\mathcal{L}_{X,Z}(\mathbb{I}_{\mathbb{R} \times A_2^c}) = 0$ . From a previous result about probability density functions,

$$\mathbb{E}(|h \circ X|) = \int_{\mathbb{R}} |h(x)|f_X(x) \text{Leb}(dx) = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} |h(x)|f_{X,Z}(x, z) \text{Leb}(dz) \right] \text{Leb}(dx) = \int_{\mathbb{R}^2} |h \circ \rho_1|f_{X,Z} d\text{Leb}^2.$$

Because  $\mathbb{E}(|h \circ X|) < \infty$ , we know that  $\text{Leb}(A_2^c) = 0$ . Because a previous result about laws extends to joint laws,

$$\int_{\mathbb{R}^2} \mathbb{I}_{\mathbb{R} \times A_2^c} d\mathcal{L}_{X,Z} = \int_{\mathbb{R}^2} \mathbb{I}_{\mathbb{R} \times A_2^c} f_{X,Z} d\text{Leb}^2 = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \mathbb{I}_{A_2^c}(z)f_{X,Z}(x, z) \text{Leb}(dx) \right] \text{Leb}(dz).$$

By rearranging terms and the using fact that  $\text{Leb}(\mathbb{I}_{A_2^c}) = 0$  implies  $\text{Leb}(\{\mathbb{I}_{A_2^c} f_Z > 0\}) \leq \text{Leb}(\{\mathbb{I}_{A_2^c} > 0\}) = 0$ ,

$$\int_{\mathbb{R}^2} \mathbb{I}_{\mathbb{R} \times A_2^c} d\mathcal{L}_{X,Z} = \int_{\mathbb{R}} \mathbb{I}_{A_2^c}(z) f_Z(z) \text{Leb}(dz) = 0.$$

Finally, we will show that  $\mathbb{E}(Y\mathbb{I}_G) = \mathbb{E}((h \circ X)\mathbb{I}_G)$  for every  $G \in \sigma(Z)$ . Note that, for every  $G \in \sigma(Z)$ ,

$$\mathbb{I}_G(\omega) = \mathbb{I}_{Z^{-1}(B)}(\omega) = (\mathbb{I}_B \circ Z)(\omega) = \begin{cases} 1, & \text{if } Z(\omega) \in B, \\ 0, & \text{if } Z(\omega) \notin B, \end{cases}$$

for some  $B \in \mathcal{B}(\mathbb{R})$ . Let  $S = (\mathbb{R} \times A_1) \cap (\mathbb{R} \times A_2)$ , so that  $S^c = (\mathbb{R} \times A_1^c) \cup (\mathbb{R} \times A_2^c)$  and  $\mathcal{L}_{X,Z}(\mathbb{I}_{S^c}) = 0$ . Note that

$$\int_{\Omega} (h \circ X)\mathbb{I}_G d\mathbb{P} = \int_{\Omega} (h \circ X)(\mathbb{I}_B \circ Z) d\mathbb{P} = \int_{\mathbb{R}^2} (h \circ \rho_1)(\mathbb{I}_B \circ \rho_2) d\mathcal{L}_{X,Z} = \int_{\mathbb{R}^2} (h \circ \rho_1)(\mathbb{I}_B \circ \rho_2)\mathbb{I}_S d\mathcal{L}_{X,Z},$$

since  $(h \circ \rho_1)(\mathbb{I}_B \circ \rho_2)$  and  $(h \circ \rho_1)(\mathbb{I}_B \circ \rho_2)\mathbb{I}_S$  are  $\mathcal{L}_{X,Z}$ -integrable and equal almost everywhere.

Because a previous result for probability density functions extends to joint probability density functions,

$$\int_{\Omega} (h \circ X)\mathbb{I}_G d\mathbb{P} = \int_{\mathbb{R}^2} (h \circ \rho_1)(\mathbb{I}_B \circ \rho_2)\mathbb{I}_S f_{X,Z} d\text{Leb}^2.$$

Because  $\mathbb{I}_S(x, z) = \mathbb{I}_{A_1}(z)\mathbb{I}_{A_2}(z)$  for every  $(x, z) \in \mathbb{R}^2$ ,

$$\int_{\Omega} (h \circ X)\mathbb{I}_G d\mathbb{P} = \int_F \left[ \int_{\mathbb{R}} h(x)\mathbb{I}_B(z)\mathbb{I}_{A_1}(z)\mathbb{I}_{A_2}(z)f_{X,Z}(x, z) \text{Leb}(dx) \right] \text{Leb}(dz),$$

where  $F = \{z \in \mathbb{R} \mid \int_{\mathbb{R}} |h(x)|\mathbb{I}_B(z)\mathbb{I}_{A_1}(z)\mathbb{I}_{A_2}(z)f_{X,Z}(x, z) \text{Leb}(dx) < \infty\}$ .

Because  $A_2 \subseteq F$ , we know that  $\mathbb{I}_F\mathbb{I}_{A_2} = \mathbb{I}_{A_2}$ . Therefore,

$$\int_{\Omega} (h \circ X)\mathbb{I}_G d\mathbb{P} = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} h(x)\mathbb{I}_B(z)\mathbb{I}_{A_1}(z)\mathbb{I}_{A_2}(z)f_{X,Z}(x, z) \text{Leb}(dx) \right] \text{Leb}(dz).$$

Because  $f_{X,Z}(x, z)\mathbb{I}_{A_1}(z) = f_{X|Z}(x, z)f_Z(z)\mathbb{I}_{A_1}(z)$  for every  $(x, z) \in \mathbb{R}^2$ ,

$$\int_{\Omega} (h \circ X)\mathbb{I}_G d\mathbb{P} = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} h(x)\mathbb{I}_B(z)\mathbb{I}_{A_1}(z)\mathbb{I}_{A_2}(z)f_{X|Z}(x, z)f_Z(z) \text{Leb}(dx) \right] \text{Leb}(dz).$$

By rearranging terms,

$$\int_{\Omega} (h \circ X)\mathbb{I}_G d\mathbb{P} = \int_{\mathbb{R}} \mathbb{I}_B(z)f_Z(z)\mathbb{I}_{A_1 \cap A_2}(z) \left[ \int_{\mathbb{R}} h(x)f_{X|Z}(x, z) \text{Leb}(dx) \right] \text{Leb}(dz).$$

For any  $z \in (A_1 \cap A_2)$ , by the linearity of the integral with respect to  $\text{Leb}$ ,

$$\mathbb{I}_{A_1}(z) \int_{\mathbb{R}} |h(x)|f_{X,Z}(x, z) \text{Leb}(dx) = f_Z(z) \int_{\mathbb{R}} |h(x)|f_{X|Z}(x, z) \text{Leb}(dx) < \infty.$$

Because  $f_Z(z) > 0$ , we know that  $\int_{\mathbb{R}} |h(x)|f_{X|Z}(x, z) \text{Leb}(dx) < \infty$ , so that  $z \in F_2^g$ .

Because  $(A_1 \cap A_2) \subseteq F_2^g$  implies  $\mathbb{I}_{A_1 \cap A_2} = \mathbb{I}_{A_1 \cap A_2}\mathbb{I}_{F_2^g}$ ,

$$\int_{\Omega} (h \circ X)\mathbb{I}_G d\mathbb{P} = \int_{\mathbb{R}} \mathbb{I}_B(z)f_Z(z)\mathbb{I}_{A_1 \cap A_2}(z)\mathbb{I}_{F_2^g}(z) \left[ \int_{\mathbb{R}} h(x)f_{X|Z}(x, z) \text{Leb}(dx) \right] \text{Leb}(dz).$$

By the definition of  $g$ ,

$$\int_{\Omega} (h \circ X)\mathbb{I}_G d\mathbb{P} = \int_{\mathbb{R}} \mathbb{I}_B(z)f_Z(z)\mathbb{I}_{A_1 \cap A_2}(z)g(z) \text{Leb}(dz).$$

By once again applying results about probability density functions and joint laws,

$$\int_{\Omega} (h \circ X)\mathbb{I}_G d\mathbb{P} = \int_{\Omega} (\mathbb{I}_B \circ Z)(\mathbb{I}_{A_1 \cap A_2} \circ Z)(g \circ Z) d\mathbb{P} = \int_{\mathbb{R}^2} (\mathbb{I}_B \circ \rho_2)(\mathbb{I}_{A_1 \cap A_2} \circ \rho_2)(g \circ \rho_2) d\mathcal{L}_{X,Z}.$$



Because  $\mathbb{I}_S(x, z) = \mathbb{I}_{A_1}(z)\mathbb{I}_{A_2}(z)$  for every  $(x, z) \in \mathbb{R}^2$ ,

$$\int_{\Omega} (h \circ X) \mathbb{I}_G d\mathbb{P} = \int_{\mathbb{R}^2} (g \circ \rho_2)(\mathbb{I}_B \circ \rho_2) \mathbb{I}_S d\mathcal{L}_{X,Z}.$$

Because  $(g \circ \rho_2)(\mathbb{I}_B \circ \rho_2)$  and  $(g \circ \rho_2)(\mathbb{I}_B \circ \rho_2) \mathbb{I}_S$  are  $\mathcal{L}_{X,Z}$ -integrable functions that are equal almost everywhere,

$$\int_{\Omega} (h \circ X) \mathbb{I}_G d\mathbb{P} = \int_{\mathbb{R}^2} (g \circ \rho_2)(\mathbb{I}_B \circ \rho_2) d\mathcal{L}_{X,Z} = \int_{\Omega} (g \circ Z)(\mathbb{I}_B \circ Z) d\mathbb{P} = \int_{\Omega} Y \mathbb{I}_G d\mathbb{P},$$

which completes the proof.

Consider a random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . For the remainder of this text, we let  $\mathbb{E}(X | \mathcal{G})$  denote an arbitrary version of the conditional expectation of  $X$  given  $\mathcal{G}$ .

Consider a random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . Note that  $\mathbb{E}(\mathbb{E}(X | \mathcal{G})) = \mathbb{E}(\mathbb{E}(X | \mathcal{G}) \mathbb{I}_{\Omega}) = \mathbb{E}(X \mathbb{I}_{\Omega}) = \mathbb{E}(X)$ .

Consider a random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . Note that if  $X$  is  $\mathcal{G}$ -measurable, then  $X = \mathbb{E}(X | \mathcal{G})$  almost surely.

Consider a random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and let  $Y = \mathbb{E}(X) \mathbb{I}_{\Omega}$ . We will now show that  $Y = \mathbb{E}(X | \{\emptyset, \Omega\})$  almost surely. For every  $B \in \mathcal{B}(\mathbb{R})$ , we have  $Y^{-1}(B) = \emptyset$  if  $\mathbb{E}(X) \notin B$  and  $Y^{-1}(B) = \Omega$  if  $\mathbb{E}(X) \in B$ . Furthermore,  $\mathbb{E}(|Y|) = \mathbb{E}(|\mathbb{E}(X) \mathbb{I}_{\Omega}|) = \mathbb{E}(|X|) < \infty$ . Therefore,  $Y \in \mathcal{L}^1(\Omega, \{\emptyset, \Omega\}, \mathbb{P})$ . Finally,  $\mathbb{E}(Y \mathbb{I}_{\Omega}) = \mathbb{E}(\mathbb{E}(X) \mathbb{I}_{\Omega} \mathbb{I}_{\Omega}) = \mathbb{E}(X \mathbb{I}_{\Omega})$  and  $\mathbb{E}(Y \mathbb{I}_{\emptyset}) = 0 = \mathbb{E}(X \mathbb{I}_{\emptyset})$ .

Consider the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ , a random variable  $X : \Omega \rightarrow \mathbb{R}$ , and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . We will now show that if  $X = 0$  almost surely, then  $0 = \mathbb{E}(X | \mathcal{G})$  almost surely, where  $0$  denotes the zero function. Clearly,  $0 \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$ . For every  $G \in \mathcal{G}$ , because  $\mathbb{P}(X \mathbb{I}_G = 0) = 1$ , we know that  $\mathbb{E}(X \mathbb{I}_G) = 0 = \mathbb{E}(0 \mathbb{I}_G)$ .

Consider the random variables  $X_1 \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and  $X_2 \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . We will now show that  $a_1 \mathbb{E}(X_1 | \mathcal{G}) + a_2 \mathbb{E}(X_2 | \mathcal{G}) = \mathbb{E}(a_1 X_1 + a_2 X_2 | \mathcal{G})$  almost surely for every  $a_1, a_2 \in \mathbb{R}$ .

Because  $\mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$  is a vector space, we know that  $a_1 \mathbb{E}(X_1 | \mathcal{G}) + a_2 \mathbb{E}(X_2 | \mathcal{G}) \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$ . For every  $G \in \mathcal{G}$ ,

$$\mathbb{E}((a_1 \mathbb{E}(X_1 | \mathcal{G}) + a_2 \mathbb{E}(X_2 | \mathcal{G})) \mathbb{I}_G) = a_1 \mathbb{E}(\mathbb{E}(X_1 | \mathcal{G}) \mathbb{I}_G) + a_2 \mathbb{E}(\mathbb{E}(X_2 | \mathcal{G}) \mathbb{I}_G).$$

From the definition of a version of the conditional expectation,

$$\mathbb{E}((a_1 \mathbb{E}(X_1 | \mathcal{G}) + a_2 \mathbb{E}(X_2 | \mathcal{G})) \mathbb{I}_G) = a_1 \mathbb{E}(X_1 \mathbb{I}_G) + a_2 \mathbb{E}(X_2 \mathbb{I}_G) = \mathbb{E}((a_1 X_1 + a_2 X_2) \mathbb{I}_G).$$

Consider the random variables  $X_1 \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and  $X_2 \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . We will now show that if  $X_1 = X_2$  almost surely, then  $\mathbb{E}(X_1 | \mathcal{G}) = \mathbb{E}(X_2 | \mathcal{G})$  almost surely. Because  $\mathbb{P}(X_1 - X_2 = 0) = 1$ , we know that  $\mathbb{P}(\mathbb{E}(X_1 - X_2 | \mathcal{G}) = 0) = 1$ . Therefore, by linearity,  $\mathbb{P}(\mathbb{E}(X_1 | \mathcal{G}) = \mathbb{E}(X_2 | \mathcal{G})) = 1$ .

Consider a random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . We will again show that if  $X \geq 0$ , then  $\mathbb{P}(\mathbb{E}(X | \mathcal{G}) \geq 0) = 1$ .

In order to find a contradiction, suppose that  $\mathbb{P}(\mathbb{E}(X | \mathcal{G}) \geq 0) < 1$ , so that  $\mathbb{P}(\mathbb{E}(X | \mathcal{G}) < 0) > 0$ . Let  $A_n = \{\mathbb{E}(X | \mathcal{G}) < -n^{-1}\} = \mathbb{E}(X | \mathcal{G})^{-1}((-\infty, -n^{-1}))$ , so that  $A_n \subseteq A_{n+1}$  and  $\cup_n A_n = \{\mathbb{E}(X | \mathcal{G}) < 0\}$ . Since  $A_n \uparrow \{\mathbb{E}(X | \mathcal{G}) < 0\}$ , the monotone-convergence property of measure guarantees that  $\mathbb{P}(A_n) \uparrow \mathbb{P}(\mathbb{E}(X | \mathcal{G}) < 0)$ . Because we supposed that  $\mathbb{P}(\mathbb{E}(X | \mathcal{G}) < 0) > 0$ , there is an  $n \in \mathbb{N}$  such that  $\mathbb{P}(A_n) = \mathbb{P}(\mathbb{E}(X | \mathcal{G}) < -n^{-1}) > 0$ . Consider the random variable  $\mathbb{E}(X | \mathcal{G}) \mathbb{I}_{A_n}$  given by

$$(\mathbb{E}(X | \mathcal{G}) \mathbb{I}_{A_n})(\omega) = \mathbb{E}(X | \mathcal{G})(\omega) \mathbb{I}_{A_n}(\omega) = \begin{cases} \mathbb{E}(X | \mathcal{G})(\omega), & \text{if } \mathbb{E}(X | \mathcal{G})(\omega) < -n^{-1}, \\ 0, & \text{if } \mathbb{E}(X | \mathcal{G})(\omega) \geq -n^{-1}. \end{cases}$$

Because  $\mathbb{E}(X | \mathcal{G}) \mathbb{I}_{A_n} < -n^{-1} \mathbb{I}_{A_n}$ , we know that  $\mathbb{E}(\mathbb{E}(X | \mathcal{G}) \mathbb{I}_{A_n}) \leq -n^{-1} \mathbb{P}(A_n) < 0$ . Because  $X \geq 0$ , we know that  $\mathbb{E}(X \mathbb{I}_{A_n}) \geq 0$ . However,  $A_n \in \mathcal{G}$ , so that  $\mathbb{E}(X \mathbb{I}_{A_n}) = \mathbb{E}(\mathbb{E}(X | \mathcal{G}) \mathbb{I}_{A_n})$ . Because this is a contradiction, we know that  $\mathbb{P}(\mathbb{E}(X | \mathcal{G}) \geq 0) = 1$ .

Consider a random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . We will now show that  $|\mathbb{E}(X | \mathcal{G})| \leq \mathbb{E}(|X| | \mathcal{G})$  almost surely. By the linearity of conditional expectation,

$$\begin{aligned} \mathbb{P}(|\mathbb{E}(X | \mathcal{G})| = |\mathbb{E}(X^+ - X^- | \mathcal{G})| = |\mathbb{E}(X^+ | \mathcal{G}) - \mathbb{E}(X^- | \mathcal{G})|) &= 1, \\ \mathbb{P}(\mathbb{E}(|X| | \mathcal{G}) = \mathbb{E}(X^+ + X^- | \mathcal{G}) = \mathbb{E}(X^+ | \mathcal{G}) + \mathbb{E}(X^- | \mathcal{G})) &= 1. \end{aligned}$$

By the triangle inequality,  $|\mathbb{E}(X^+ | \mathcal{G}) - \mathbb{E}(X^- | \mathcal{G})| \leq |\mathbb{E}(X^+ | \mathcal{G})| + |\mathbb{E}(X^- | \mathcal{G})|$ .

Because  $\mathbb{P}(|\mathbb{E}(X^+ | \mathcal{G})| = \mathbb{E}(X^+ | \mathcal{G})) = 1$  and  $\mathbb{P}(|\mathbb{E}(X^- | \mathcal{G})| = \mathbb{E}(X^- | \mathcal{G})) = 1$ ,

$$\mathbb{P}(|\mathbb{E}(X | \mathcal{G})| \leq |\mathbb{E}(X^+ | \mathcal{G})| + |\mathbb{E}(X^- | \mathcal{G})| = \mathbb{E}(X^+ | \mathcal{G}) + \mathbb{E}(X^- | \mathcal{G}) = \mathbb{E}(|X| | \mathcal{G})) = 1.$$

Consider a sequence of non-negative random variables  $(X_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \mid n \in \mathbb{N})$ , a non-negative random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ , and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . The conditional monotone-convergence theorem states that if  $X_n \uparrow X$ , then  $\mathbb{P}(\mathbb{E}(X_n | \mathcal{G})\mathbb{I}_A \uparrow \mathbb{E}(X | \mathcal{G})) = 1$ , where  $A \in \mathcal{G}$  is a set such that  $\mathbb{P}(A) = 1$ .

We will now show this theorem. Because  $X_n$  is a non-negative random variable,  $\mathbb{P}(\mathbb{E}(X_n | \mathcal{G}) \geq 0) = 1$ . For every  $n \in \mathbb{N}$ , because  $X_{n+1} - X_n$  is non-negative and  $\mathbb{E}(X_{n+1} | \mathcal{G}) - \mathbb{E}(X_n | \mathcal{G}) = \mathbb{E}(X_{n+1} - X_n | \mathcal{G})$  almost surely,  $\mathbb{P}(\mathbb{E}(X_{n+1} | \mathcal{G}) - \mathbb{E}(X_n | \mathcal{G}) \geq 0) = 1$ .

Let  $A^c = \bigcup_n \{\mathbb{E}(X_n | \mathcal{G}) < 0\} \cup \{\mathbb{E}(X_{n+1} | \mathcal{G}) - \mathbb{E}(X_n | \mathcal{G}) < 0\}$ . Note that  $A \in \mathcal{G}$  and  $\mathbb{P}(A) = 1$ , since

$$\mathbb{P}(A^c) \leq \sum_n \mathbb{P}(\mathbb{E}(X_n | \mathcal{G}) < 0) + \mathbb{P}(\mathbb{E}(X_{n+1} | \mathcal{G}) - \mathbb{E}(X_n | \mathcal{G}) < 0) = 0.$$

For every  $n \in \mathbb{N}$ , note that  $\mathbb{E}(X_n | \mathcal{G})\mathbb{I}_A \geq 0$  and  $\mathbb{E}(X_{n+1} | \mathcal{G})\mathbb{I}_A \geq \mathbb{E}(X_n | \mathcal{G})\mathbb{I}_A$ .

Let  $Y = \limsup_{n \rightarrow \infty} \mathbb{E}(X_n | \mathcal{G})\mathbb{I}_A$ . For every  $G \in \mathcal{G}$ , because every non-decreasing sequence of real numbers converges (possibly to infinity), we know that  $\mathbb{E}(X_n | \mathcal{G})\mathbb{I}_A\mathbb{I}_G \uparrow Y\mathbb{I}_G$ , which also implies  $\mathbb{E}(X_n | \mathcal{G})\mathbb{I}_A \uparrow Y$ . By the monotone-convergence theorem, we know that  $\mathbb{E}(\mathbb{E}(X_n | \mathcal{G})\mathbb{I}_A\mathbb{I}_G) \uparrow \mathbb{E}(Y\mathbb{I}_G)$ .

For every  $n \in \mathbb{N}$  and  $G \in \mathcal{G}$ , we have  $(A \cap G) \in \mathcal{G}$  and  $\mathbb{P}(X_n\mathbb{I}_G\mathbb{I}_{A^c} \neq 0) = 0$ , so that

$$\mathbb{E}(\mathbb{E}(X_n | \mathcal{G})\mathbb{I}_A\mathbb{I}_G) = \mathbb{E}(\mathbb{E}(X_n | \mathcal{G})\mathbb{I}_{A \cap G}) = \mathbb{E}(X_n\mathbb{I}_{A \cap G}) = \mathbb{E}(X_n\mathbb{I}_A\mathbb{I}_G) + \mathbb{E}(X_n\mathbb{I}_{A^c}\mathbb{I}_G) = \mathbb{E}(X_n\mathbb{I}_G),$$

which implies  $\mathbb{E}(X_n\mathbb{I}_G) \uparrow \mathbb{E}(Y\mathbb{I}_G)$ . Since  $X_n\mathbb{I}_G \uparrow X\mathbb{I}_G$ , we also know that  $\mathbb{E}(X_n\mathbb{I}_G) \uparrow \mathbb{E}(X\mathbb{I}_G)$ , so that  $\mathbb{E}(Y\mathbb{I}_G) = \mathbb{E}(X\mathbb{I}_G)$ . Because  $Y$  is  $\mathcal{G}$ -measurable and  $\Omega \in \mathcal{G}$ , we know that  $Y = \mathbb{E}(X | \mathcal{G})$  almost surely.

Consider a sequence of non-negative random variables  $(X_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \mid n \in \mathbb{N})$  and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . The conditional Fatou lemma states that if  $\mathbb{E}(\liminf_{n \rightarrow \infty} X_n) < \infty$ , then

$$\mathbb{P}\left(\mathbb{E}\left(\liminf_{n \rightarrow \infty} X_n \mid \mathcal{G}\right) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n | \mathcal{G})\right) = 1.$$

We will now show this lemma. For any  $m \in \mathbb{N}$ , consider the function  $Z_m = \inf_{n \geq m} X_n$ , such that

$$\liminf_{n \rightarrow \infty} X_n = \lim_{m \rightarrow \infty} \inf_{n \geq m} X_n = \lim_{m \rightarrow \infty} Z_m.$$

Because  $Z_m \leq Z_{m+1}$  for every  $m \in \mathbb{N}$ , we have  $Z_m \uparrow \liminf_{n \rightarrow \infty} X_n$ . Furthermore,  $Z_m \geq 0$  and  $Z_m \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  for every  $m \in \mathbb{N}$ . Therefore, by the conditional monotone-convergence theorem,

$$\mathbb{P}\left(\mathbb{E}(Z_m | \mathcal{G})\mathbb{I}_A \uparrow \mathbb{E}\left(\liminf_{n \rightarrow \infty} X_n \mid \mathcal{G}\right)\right) = 1,$$

where  $A \in \mathcal{G}$  and  $\mathbb{P}(A) = 1$ .

For any  $n \geq m$ , note that  $X_n \geq Z_m$ . Therefore,  $\mathbb{P}(\mathbb{E}(X_n - Z_m | \mathcal{G}) \geq 0) = 1$  and  $\mathbb{P}(\mathbb{E}(X_n | \mathcal{G}) \geq \mathbb{E}(Z_m | \mathcal{G})) = 1$ . Furthermore, for every  $m \in \mathbb{N}$ , because  $\mathbb{P}(A^c) = 0$ ,

$$\mathbb{P}\left(\inf_{n \geq m} \mathbb{E}(X_n | \mathcal{G}) \geq \mathbb{E}(Z_m | \mathcal{G})\mathbb{I}_A\right) = 1.$$

By taking the limit of both sides of the previous inequation when  $m \rightarrow \infty$ ,

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} \mathbb{E}(X_n | \mathcal{G}) \geq \mathbb{E}\left(\liminf_{n \rightarrow \infty} X_n \mid \mathcal{G}\right)\right) = 1.$$

Consider a sequence of non-negative random variables  $(X_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \mid n \in \mathbb{N})$ , a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ , and a non-negative random variable  $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  such that  $X_n \leq Y$  for every  $n \in \mathbb{N}$ . The reverse conditional Fatou lemma states that

$$\mathbb{P}\left(\mathbb{E}\left(\limsup_{n \rightarrow \infty} X_n \mid \mathcal{G}\right) \geq \limsup_{n \rightarrow \infty} \mathbb{E}(X_n | \mathcal{G})\right) = 1.$$

We will now show this lemma. Because  $X_n \leq Y$  for every  $n \in \mathbb{N}$ , we know that  $\mathbb{E}(\limsup_{n \rightarrow \infty} X_n) \leq \mathbb{E}(Y) < \infty$ .

For every  $n \in \mathbb{N}$ , consider the non-negative function  $Z_n = Y - X_n$ , so that  $Z_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . From the conditional Fatou lemma, since  $\mathbb{E}(\liminf_{n \rightarrow \infty} Z_n) < \infty$ ,

$$\mathbb{P}\left(\mathbb{E}\left(\liminf_{n \rightarrow \infty} Y - X_n \mid \mathcal{G}\right) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(Y - X_n \mid \mathcal{G})\right) = 1.$$

For every  $n \in \mathbb{N}$ , by moving constants outside the corresponding limits and linearity,

$$\mathbb{P}\left(\mathbb{E}(Y \mid \mathcal{G}) + \mathbb{E}\left(\liminf_{n \rightarrow \infty} -X_n \mid \mathcal{G}\right) \leq \mathbb{E}(Y \mid \mathcal{G}) + \liminf_{n \rightarrow \infty} -\mathbb{E}(X_n \mid \mathcal{G})\right) = 1.$$

By the relationship between limit inferior and limit superior and linearity,

$$\mathbb{P}\left(\mathbb{E}(Y \mid \mathcal{G}) - \mathbb{E}\left(\limsup_{n \rightarrow \infty} X_n \mid \mathcal{G}\right) \leq \mathbb{E}(Y \mid \mathcal{G}) - \limsup_{n \rightarrow \infty} \mathbb{E}(X_n \mid \mathcal{G})\right) = 1.$$

The proof is completed by reorganizing terms in the inequation above.

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ , a sequence of random variables  $(X_n \mid n \in \mathbb{N})$ , a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ , a random variable  $X$ , and a non-negative random variable  $V \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  such that  $|X_n| \leq V$  for every  $n \in \mathbb{N}$ . The conditional dominated convergence theorem states that if  $\mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = 1$ , then  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \mathbb{E}(X_n \mid \mathcal{G}) \mathbb{I}_C = \mathbb{E}(X \mid \mathcal{G})\right) = 1.$$

where  $C \in \mathcal{G}$  is a set such that  $\mathbb{P}(C) = 1$ .

We will now show this theorem. Because  $|X_n| \leq V$  for every  $n \in \mathbb{N}$ , we know that  $\mathbb{E}(|X_n|) \leq \mathbb{E}(V) < \infty$ , which implies that  $X_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Because the function  $|\cdot|$  is continuous, we know that  $\mathbb{P}(\lim_{n \rightarrow \infty} |X_n| = |X|) = 1$ . Because  $\mathbb{P}(\lim_{n \rightarrow \infty} |X_n| \leq V) = 1$ , we know that  $\mathbb{P}(|X| \leq V) = 1$ . Because  $\mathbb{P}(|X| \neq |X| \mathbb{I}_{\{|X| \leq V\}}) = 0$ , we know that  $\mathbb{E}(|X|) = \mathbb{E}(|X| \mathbb{I}_{\{|X| \leq V\}}) \leq \mathbb{E}(V) < \infty$ , so that  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ .

Since  $\mathbb{P}(|X_n| \leq V) = 1$  and  $\mathbb{P}(|X| \leq V) = 1$ , we have  $\mathbb{P}(|X_n| + |X| \leq 2V) = 1$ . By the triangle inequality,

$$|X_n - X| = |X_n + (-X)| \leq |X_n| + |X|,$$

which implies that  $\mathbb{P}(|X_n - X| \leq 2V) = 1$ .

Let  $A = \{|X_n - X| \leq 2V\}$ , so that  $\mathbb{P}(|X_n - X| \mathbb{I}_A = |X_n - X|) = 1$  and  $\mathbb{E}(|X_n - X|) = \mathbb{E}(|X_n - X| \mathbb{I}_A)$ . Because  $|X_n - X| \mathbb{I}_A$  is an  $\mathcal{F}$ -measurable function and  $|X_n - X| \mathbb{I}_A \leq 2V$  for every  $n \in \mathbb{N}$ , where  $2V : \Omega \rightarrow [0, \infty]$  is an  $\mathcal{F}$ -measurable function such that  $\mathbb{E}(2V) = 2\mathbb{E}(V) < \infty$ , the reverse conditional Fatou lemma states that

$$\mathbb{P}\left(\mathbb{E}\left(\limsup_{n \rightarrow \infty} |X_n - X| \mathbb{I}_A \mid \mathcal{G}\right) \geq \limsup_{n \rightarrow \infty} \mathbb{E}(|X_n - X| \mathbb{I}_A \mid \mathcal{G})\right) = 1.$$

Since  $|\cdot|$  is continuous, we have  $\mathbb{P}(\lim_{n \rightarrow \infty} |X_n - X| \mathbb{I}_A = 0) = 1$ , where 0 is the zero function. Therefore,

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} |X_n - X| \mathbb{I}_A = \liminf_{n \rightarrow \infty} |X_n - X| \mathbb{I}_A = \lim_{n \rightarrow \infty} |X_n - X| \mathbb{I}_A = 0\right) = 1.$$

Because each of the random variables above is almost surely equal to zero,

$$\mathbb{P}\left(\mathbb{E}\left(\limsup_{n \rightarrow \infty} |X_n - X| \mathbb{I}_A \mid \mathcal{G}\right) = \mathbb{E}\left(\liminf_{n \rightarrow \infty} |X_n - X| \mathbb{I}_A \mid \mathcal{G}\right) = \mathbb{E}\left(\lim_{n \rightarrow \infty} |X_n - X| \mathbb{I}_A \mid \mathcal{G}\right) = 0\right) = 1.$$

Since  $(X_n - X) \mathbb{I}_A \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  for every  $n \in \mathbb{N}$ , we have  $\mathbb{P}(|\mathbb{E}((X_n - X) \mathbb{I}_A \mid \mathcal{G})| \leq \mathbb{E}(|X_n - X| \mathbb{I}_A \mid \mathcal{G})) = 1$ . By taking the limit superior of both sides of the previous inequation and employing the previous results,

$$\mathbb{P}\left(0 \leq \limsup_{n \rightarrow \infty} |\mathbb{E}((X_n - X) \mathbb{I}_A \mid \mathcal{G})| \leq \limsup_{n \rightarrow \infty} \mathbb{E}(|X_n - X| \mathbb{I}_A \mid \mathcal{G}) \leq \mathbb{E}\left(\limsup_{n \rightarrow \infty} |X_n - X| \mathbb{I}_A \mid \mathcal{G}\right) = 0\right) = 1.$$

Therefore, by the relationship between limits,

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} \mathbb{E}((X_n - X) \mathbb{I}_A \mid \mathcal{G}) = \limsup_{n \rightarrow \infty} \mathbb{E}((X_n - X) \mathbb{I}_A \mid \mathcal{G}) = 0\right) = 1.$$

Because  $\mathbb{P}((X_n - X)\mathbb{I}_A = (X_n - X)) = 1$  implies  $\mathbb{P}(\mathbb{E}((X_n - X)\mathbb{I}_A | \mathcal{G}) = \mathbb{E}(X_n - X | \mathcal{G})) = 1$ .

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} \mathbb{E}(X_n - X | \mathcal{G}) = \limsup_{n \rightarrow \infty} \mathbb{E}(X_n - X | \mathcal{G}) = 0\right) = 1.$$

By the linearity of conditional expectation,

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} \mathbb{E}(X_n | \mathcal{G}) = \limsup_{n \rightarrow \infty} \mathbb{E}(X_n | \mathcal{G}) = \mathbb{E}(X | \mathcal{G})\right) = 1.$$

Let  $C = \{\omega \in \Omega \mid \lim_{n \rightarrow \infty} \mathbb{E}(X_n | \mathcal{G})(\omega) \text{ exists in } \mathbb{R}\}$ . Because  $\mathbb{E}(X_n | \mathcal{G})$  is  $\mathcal{G}$ -measurable for every  $n \in \mathbb{N}$ , recall that  $C \in \mathcal{G}$ . Because  $\mathbb{E}(|\mathbb{E}(X | \mathcal{G})|) < \infty$ , recall that  $\mathbb{P}(|\mathbb{E}(X | \mathcal{G})| < \infty) = 1$ , so that  $\mathbb{P}(C) = 1$ . Furthermore,

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \mathbb{E}(X_n | \mathcal{G})\mathbb{I}_C = \mathbb{E}(X | \mathcal{G})\right) = 1.$$

Consider a random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ , a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ , and a convex function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ . The conditional Jensen's inequality states that if  $(\phi \circ X) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ , then  $\mathbb{P}((\phi \circ \mathbb{E}(X | \mathcal{G})) \leq \mathbb{E}((\phi \circ X) | \mathcal{G})) = 1$ .

We will now show this inequality. Because  $\phi$  is a convex function, it is possible to show that there is a sequence  $((a_n, b_n) \in \mathbb{R}^2 \mid n \in \mathbb{N})$  such that  $\phi(x) = \sup_n a_n x + b_n$  for every  $x \in \mathbb{R}$ . Therefore,  $\phi(x) \geq a_n x + b_n$  for every  $x \in \mathbb{R}$  and  $n \in \mathbb{N}$ . Furthermore, if  $(\phi \circ X) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ , then  $(\phi \circ X) - a_n X - b_n \geq 0$  for every  $n \in \mathbb{N}$  and

$$\mathbb{P}(\mathbb{E}((\phi \circ X) - a_n X - b_n | \mathcal{G}) \geq 0) = 1.$$

For every  $n \in \mathbb{N}$ , by the linearity of conditional expectation,

$$\mathbb{P}(\mathbb{E}((\phi \circ X) | \mathcal{G}) \geq a_n \mathbb{E}(X | \mathcal{G}) + b_n) = 1.$$

By taking the supremum of both sides of the previous inequation,

$$\mathbb{P}\left(\mathbb{E}((\phi \circ X) | \mathcal{G}) \geq \sup_n a_n \mathbb{E}(X | \mathcal{G}) + b_n = (\phi \circ \mathbb{E}(X | \mathcal{G}))\right) = 1.$$

Consider a random variable  $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ , where  $p \in [1, \infty)$ , and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . We will now show that  $\|\mathbb{E}(X | \mathcal{G})\|_p \leq \|X\|_p$ .

From the monotonicity of norm, we know that  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Consider the convex function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  given by  $\phi(x) = |x|^p$ , so that  $(\phi \circ X) = |X|^p$ . Because  $\mathbb{E}(|X|^p) < \infty$ , we know that  $|X|^p \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . From the conditional Jensen's inequality,  $\mathbb{P}(|\mathbb{E}(X | \mathcal{G})|^p \leq \mathbb{E}(|X|^p | \mathcal{G})) = 1$ . Let  $A = \{|\mathbb{E}(X | \mathcal{G})|^p \leq \mathbb{E}(|X|^p | \mathcal{G})\}$ .

Because  $|\mathbb{E}(X | \mathcal{G})|^p$  is non-negative and  $\mathcal{G}$ -measurable and  $\mathbb{E}(|X|^p | \mathcal{G}) \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$ ,

$$\mathbb{E}(|\mathbb{E}(X | \mathcal{G})|^p) = \mathbb{E}(|\mathbb{E}(X | \mathcal{G})|^p \mathbb{I}_A) \leq \mathbb{E}(\mathbb{E}(|X|^p | \mathcal{G}) \mathbb{I}_A) = \mathbb{E}(\mathbb{E}(|X|^p | \mathcal{G})) = \mathbb{E}(|X|^p).$$

Consider a random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ , a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ , and a  $\sigma$ -algebra  $\mathcal{H} \subseteq \mathcal{G}$ . The tower property states that  $\mathbb{E}(\mathbb{E}(X | \mathcal{G}) | \mathcal{H}) = \mathbb{E}(X | \mathcal{H})$  almost surely. We will now show this property.

Because  $\mathbb{E}(X | \mathcal{G}) \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$ , we know that  $\mathbb{E}(\mathbb{E}(X | \mathcal{G}) | \mathcal{H}) \in \mathcal{L}^1(\Omega, \mathcal{H}, \mathbb{P})$ . For every  $H \in \mathcal{H}$ , since  $H \in \mathcal{G}$ ,

$$\int_{\Omega} \mathbb{E}(\mathbb{E}(X | \mathcal{G}) | \mathcal{H}) \mathbb{I}_H d\mathbb{P} = \int_{\Omega} \mathbb{E}(X | \mathcal{G}) \mathbb{I}_H d\mathbb{P} = \int_{\Omega} X \mathbb{I}_H d\mathbb{P},$$

as we wanted to show. For the remainder of this text, we let  $\mathbb{E}(X | \mathcal{G} | \mathcal{H})$  denote  $\mathbb{E}(\mathbb{E}(X | \mathcal{G}) | \mathcal{H})$ .

Consider a random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ , a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ , and a  $\mathcal{G}$ -measurable random variable  $Z : \Omega \rightarrow \mathbb{R}$ . We will now show that if  $\mathbb{E}(|ZX|) < \infty$ , then  $\mathbb{E}(ZX | \mathcal{G}) = Z\mathbb{E}(X | \mathcal{G})$  almost surely.

We will start by assuming that  $X \geq 0$ .

First, suppose that  $Z = \mathbb{I}_A$ , where  $A \in \mathcal{G}$ . For every  $G \in \mathcal{G}$ , since  $ZX \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and  $A \cap G \in \mathcal{G}$ ,

$$\mathbb{E}(ZX \mathbb{I}_G) = \mathbb{E}(X \mathbb{I}_{A \cap G}) = \mathbb{E}(\mathbb{E}(X | \mathcal{G}) \mathbb{I}_{A \cap G}) = \mathbb{E}(Z \mathbb{E}(X | \mathcal{G}) \mathbb{I}_G).$$

Because  $Z\mathbb{E}(X | \mathcal{G})$  is  $\mathcal{G}$ -measurable and  $\mathbb{E}(Z\mathbb{E}(X | \mathcal{G})) = \mathbb{E}(ZX) < \infty$ , we know that  $Z\mathbb{E}(X | \mathcal{G}) = \mathbb{E}(ZX | \mathcal{G})$  almost surely.

Next, suppose that  $Z$  is a simple function that can be written as  $Z = \sum_{k=1}^m a_k \mathbb{I}_{A_k}$  for some fixed  $a_1, a_2, \dots, a_m \in [0, \infty]$  and  $A_1, A_2, \dots, A_m \in \mathcal{G}$ . By the linearity of the conditional expectation and the previous step,

$$\mathbb{P} \left( \mathbb{E}(ZX \mid \mathcal{G}) = \mathbb{E} \left( \sum_{k=1}^m a_k \mathbb{I}_{A_k} X \mid \mathcal{G} \right) = \sum_{k=1}^m a_k \mathbb{E}(\mathbb{I}_{A_k} X \mid \mathcal{G}) = \sum_{k=1}^m a_k \mathbb{I}_{A_k} \mathbb{E}(X \mid \mathcal{G}) = Z \mathbb{E}(X \mid \mathcal{G}) \right) = 1,$$

where we also used the fact that  $\mathbb{E}(\mathbb{I}_{A_k} X) \leq \mathbb{E}(X) < \infty$ .

Next, suppose that  $Z$  is a non-negative  $\mathcal{G}$ -measurable function. For any  $n \in \mathbb{N}$ , consider the simple function  $Z_n = \alpha_n \circ Z$ , where  $\alpha_n$  is the  $n$ -th staircase function.

For every  $G \in \mathcal{G}$ , since  $Z_n \uparrow Z$  and  $X \mathbb{I}_G \geq 0$ , note that  $Z_n X \mathbb{I}_G \uparrow Z X \mathbb{I}_G$ . For every  $G \in \mathcal{G}$ , since  $Z_n \uparrow Z$  and  $|\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_G \geq 0$ , note that  $Z_n |\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_G \uparrow Z |\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_G$ . Therefore, by the monotone-convergence theorem, we know that  $\mathbb{E}(Z_n X \mathbb{I}_G) \uparrow \mathbb{E}(Z X \mathbb{I}_G)$  and  $\mathbb{E}(Z_n |\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_G) \uparrow \mathbb{E}(Z |\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_G)$ .

Because  $Z_n$  is a simple  $\mathcal{G}$ -measurable function and  $\mathbb{E}(Z_n X) \leq \mathbb{E}(Z X) < \infty$ , note that  $\mathbb{E}(Z_n X \mid \mathcal{G}) = Z_n \mathbb{E}(X \mid \mathcal{G})$  almost surely. Because  $Z_n |\mathbb{E}(X \mid \mathcal{G})| = Z_n |\mathbb{E}(X \mid \mathcal{G})|$  almost surely,  $\mathbb{E}(Z_n X \mathbb{I}_G) = \mathbb{E}(Z_n |\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_G)$  for every  $G \in \mathcal{G}$ . Therefore, the previous result implies that  $\mathbb{E}(Z X \mathbb{I}_G) = \mathbb{E}(Z |\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_G)$  for every  $G \in \mathcal{G}$ , so that  $Z |\mathbb{E}(X \mid \mathcal{G})| = \mathbb{E}(Z X \mid \mathcal{G})$  almost surely. Because  $Z |\mathbb{E}(X \mid \mathcal{G})| = Z \mathbb{E}(X \mid \mathcal{G})$  almost surely, this step is complete.

Next, suppose that  $Z$  is a  $\mathcal{G}$ -measurable function. Recall that  $Z = Z^+ - Z^-$ , where  $Z^+$  and  $Z^-$  are non-negative  $\mathcal{G}$ -measurable functions. By the linearity of the conditional expectation and the previous step,

$$\mathbb{P}(\mathbb{E}(ZX \mid \mathcal{G}) = \mathbb{E}(Z^+ X \mid \mathcal{G}) - \mathbb{E}(Z^- X \mid \mathcal{G}) = Z^+ \mathbb{E}(X \mid \mathcal{G}) - Z^- \mathbb{E}(X \mid \mathcal{G}) = Z \mathbb{E}(X \mid \mathcal{G})) = 1,$$

where we have also used the fact that  $\mathbb{E}(Z^+ X) + \mathbb{E}(Z^- X) = \mathbb{E}((Z^+ + Z^-)X) = \mathbb{E}(|ZX|) < \infty$ .

Finally, suppose that  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Recall that  $X = X^+ - X^-$ , where  $X^+$  and  $X^-$  are non-negative  $\mathcal{F}$ -measurable functions. By the linearity of the conditional expectation,

$$\mathbb{P}(\mathbb{E}(ZX \mid \mathcal{G}) = \mathbb{E}(ZX^+ \mid \mathcal{G}) - \mathbb{E}(ZX^- \mid \mathcal{G}) = Z \mathbb{E}(X^+ \mid \mathcal{G}) - Z \mathbb{E}(X^- \mid \mathcal{G}) = Z \mathbb{E}(X \mid \mathcal{G})) = 1,$$

where we have also used the fact that  $\mathbb{E}(|Z|X^+) + \mathbb{E}(|Z|X^-) = \mathbb{E}(|Z|(X^+ + X^-)) = \mathbb{E}(|ZX|) < \infty$ .

Consider a random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ , a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ , and a  $\sigma$ -algebra  $\mathcal{H} \subseteq \mathcal{F}$ . We will now show that if  $\mathcal{H}$  and  $\sigma(\sigma(X) \cup \mathcal{G})$  are independent, then  $\mathbb{E}(X \mid \sigma(\mathcal{G} \cup \mathcal{H})) = \mathbb{E}(X \mid \mathcal{G})$  almost surely.

We will start by assuming that  $X \geq 0$ .

For every  $G \in \mathcal{G}$ , note that  $|\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_G$  is  $\mathcal{G}$ -measurable. Consider the Borel function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  given by  $f(a, b) = ab$ . Since  $(X \mathbb{I}_G)(\omega) = f(X(\omega), \mathbb{I}_G(\omega))$  for every  $\omega \in \Omega$ , we also know that  $X \mathbb{I}_G$  is  $\sigma(\sigma(X) \cup \mathcal{G})$ -measurable.

For every  $G \in \mathcal{G}$  and  $H \in \mathcal{H}$ , we know that  $X \mathbb{I}_G$  and  $\mathbb{I}_H$  are independent, since  $\mathbb{I}_H$  is  $\mathcal{H}$ -measurable. We also know that  $|\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_G$  and  $\mathbb{I}_H$  are independent, since  $\mathcal{G} \subseteq \sigma(\sigma(X) \cup \mathcal{G})$ .

For every  $G \in \mathcal{G}$  and  $H \in \mathcal{H}$ , because  $X \mathbb{I}_G \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ ,  $|\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_G \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ , and  $\mathbb{I}_H \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ ,

$$\mathbb{E}(X; G \cap H) = \mathbb{E}(X \mathbb{I}_G \mathbb{I}_H) = \mathbb{E}(X \mathbb{I}_G) \mathbb{E}(\mathbb{I}_H) = \mathbb{E}(|\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_G) \mathbb{E}(\mathbb{I}_H) = \mathbb{E}(|\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_G \mathbb{I}_H) = \mathbb{E}(|\mathbb{E}(X \mid \mathcal{G})|; G \cap H).$$

Consider the set  $\mathcal{I} = \{G \cap H \mid G \in \mathcal{G} \text{ and } H \in \mathcal{H}\}$ . Suppose that  $(G_1 \cap H_1) \in \mathcal{I}$  and  $(G_2 \cap H_2) \in \mathcal{I}$ , and note that  $(G_1 \cap H_1) \cap (G_2 \cap H_2) = (G_1 \cap G_2) \cap (H_1 \cap H_2)$ . Because  $(G_1 \cap G_2) \in \mathcal{G}$  and  $(H_1 \cap H_2) \in \mathcal{H}$ , we know that  $((G_1 \cap H_1) \cap (G_2 \cap H_2)) \in \mathcal{I}$ , so that  $\mathcal{I}$  is a  $\pi$ -system.

Since  $\Omega \in \mathcal{G}$ , we know that  $\mathcal{H} \subseteq \mathcal{I}$ . Since  $\Omega \in \mathcal{H}$ , we know that  $\mathcal{G} \subseteq \mathcal{I}$ . Therefore,  $\mathcal{G} \cup \mathcal{H} \subseteq \mathcal{I}$ , so that  $\sigma(\mathcal{G} \cup \mathcal{H}) \subseteq \sigma(\mathcal{I})$ . For every  $G \in \mathcal{G}$  and  $H \in \mathcal{H}$ , we know that  $(G \cap H) \in \sigma(\mathcal{G} \cup \mathcal{H})$ . Therefore  $\mathcal{I} \subseteq \sigma(\mathcal{G} \cup \mathcal{H})$ , so that  $\sigma(\mathcal{I}) \subseteq \sigma(\mathcal{G} \cup \mathcal{H})$ . In conclusion,  $\sigma(\mathcal{I}) = \sigma(\mathcal{G} \cup \mathcal{H})$ .

Consider the measure  $(X\mathbb{P}) : \mathcal{F} \rightarrow [0, \infty]$  given by  $(X\mathbb{P})(A) = \mathbb{E}(X; A)$  and the measure  $(|\mathbb{E}(X \mid \mathcal{G})| \mathbb{P}) : \mathcal{F} \rightarrow [0, \infty]$  given by  $(|\mathbb{E}(X \mid \mathcal{G})| \mathbb{P})(A) = \mathbb{E}(|\mathbb{E}(X \mid \mathcal{G})|; A)$ . For every  $I \in \mathcal{I}$ , we know that  $(X\mathbb{P})(I) = (|\mathbb{E}(X \mid \mathcal{G})| \mathbb{P})(I)$ . In particular, we know that  $(X\mathbb{P})(\Omega) = \mathbb{E}(X) = (|\mathbb{E}(X \mid \mathcal{G})| \mathbb{P})(\Omega) < \infty$ . Therefore, from a previous result, we know that  $\mathbb{E}(X \mathbb{I}_A) = \mathbb{E}(|\mathbb{E}(X \mid \mathcal{G})| \mathbb{I}_A)$  for every  $A \in \sigma(\mathcal{G} \cup \mathcal{H})$ . Because  $|\mathbb{E}(X \mid \mathcal{G})|$  is  $\sigma(\mathcal{G} \cup \mathcal{H})$ -measurable and  $\mathbb{E}(|\mathbb{E}(X \mid \mathcal{G})|) = \mathbb{E}(X) < \infty$ , we know that  $|\mathbb{E}(X \mid \mathcal{G})| = \mathbb{E}(X \mid \sigma(\mathcal{G} \cup \mathcal{H}))$  almost surely. Since  $|\mathbb{E}(X \mid \mathcal{G})| = \mathbb{E}(X \mid \mathcal{G})$  almost surely, this step is complete.

Finally, suppose  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Recall that  $X = X^+ - X^-$ , where  $X^+ \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and  $X^- \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  are non-negative. By the linearity of the conditional expectation,

$$\mathbb{P}(\mathbb{E}(X \mid \sigma(\mathcal{G} \cup \mathcal{H})) = \mathbb{E}(X^+ \mid \sigma(\mathcal{G} \cup \mathcal{H})) - \mathbb{E}(X^- \mid \sigma(\mathcal{G} \cup \mathcal{H})) = \mathbb{E}(X^+ \mid \mathcal{G}) - \mathbb{E}(X^- \mid \mathcal{G}) = \mathbb{E}(X \mid \mathcal{G})) = 1,$$

where we used the fact that  $\sigma(\sigma(X^+) \cup \mathcal{G}) \subseteq \sigma(\sigma(X) \cup \mathcal{G})$  and  $\sigma(\sigma(X^-) \cup \mathcal{G}) \subseteq \sigma(\sigma(X) \cup \mathcal{G})$ .

Consider a random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and a  $\sigma$ -algebra  $\mathcal{H} \subseteq \mathcal{F}$ . We will now show that if  $\mathcal{H}$  and  $\sigma(X)$  are independent, then  $\mathbb{E}(X \mid \mathcal{H}) = \mathbb{E}(X)$  almost surely.

Let  $\mathcal{G} = \{\emptyset, \Omega\}$ . Using the previous result, we know that  $\mathbb{E}(X \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{G})$  almost surely. Based on a previous result, we know that  $\mathbb{E}(X) = \mathbb{E}(X \mid \mathcal{G})$  almost surely.

Consider the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . For every  $F \in \mathcal{F}$ , we let  $\mathbb{P}(F \mid \mathcal{G})$  denote a version of the conditional expectation  $\mathbb{E}(\mathbb{I}_F \mid \mathcal{G})$  of  $\mathbb{I}_F$  given  $\mathcal{G}$ , so that  $\mathbb{P}(F \mid \mathcal{G}) = \mathbb{E}(\mathbb{I}_F \mid \mathcal{G})$  almost surely. Note that  $\mathbb{P}(F \mid \{\emptyset, \Omega\}) = \mathbb{E}(\mathbb{I}_F \mid \{\emptyset, \Omega\}) = \mathbb{E}(\mathbb{I}_F) = \mathbb{P}(F)$  almost surely.

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and the random variables  $\mathbb{I}_F : \Omega \rightarrow \{0, 1\}$  and  $Z : \Omega \rightarrow \mathcal{Z}$ , where  $F \in \mathcal{F}$  and  $\mathcal{Z} = \{z_1, \dots, z_n\}$ . Furthermore, suppose  $\mathbb{P}(Z = z) > 0$  for every  $z \in \mathcal{Z}$ . Recall that if  $E : \mathcal{Z} \rightarrow [0, 1]$  is given by

$$E(z) = \frac{\mathbb{P}(\mathbb{I}_F = 1, Z = z)}{\mathbb{P}(Z = z)} = \frac{\mathbb{P}(F \cap \{Z = z\})}{\mathbb{P}(Z = z)},$$

then  $E \circ Z = \mathbb{E}(\mathbb{I}_F \mid Z) = \mathbb{P}(F \mid Z)$  almost surely.

Consider a sequence of events  $(F_n \in \mathcal{F} \mid n \in \mathbb{N})$  such that  $F_n \cap F_m = \emptyset$  for every  $n \neq m$ . We will now show that  $\mathbb{P}(\bigcup_n F_n \mid \mathcal{G}) = \sum_n \mathbb{I}_A \mathbb{P}(F_n \mid \mathcal{G})$  almost surely, where  $A \in \mathcal{G}$  is a set such that  $\mathbb{P}(A) = 1$ .

For every  $k \in \mathbb{N}$ , by the linearity of conditional expectation,

$$\mathbb{P}\left(\mathbb{P}\left(\bigcup_{i=0}^k F_i \mid \mathcal{G}\right) = \mathbb{E}\left(\mathbb{I}_{\bigcup_{i=0}^k F_i} \mid \mathcal{G}\right) = \mathbb{E}\left(\sum_{i=0}^k \mathbb{I}_{F_i} \mid \mathcal{G}\right) = \sum_{i=0}^k \mathbb{E}(\mathbb{I}_{F_i} \mid \mathcal{G}) = \sum_{i=0}^k \mathbb{P}(F_i \mid \mathcal{G})\right) = 1.$$

Because  $\mathbb{I}_{\bigcup_{i=0}^k F_i} \uparrow \mathbb{I}_{\bigcup_n F_n}$  with respect to  $k$ , by the conditional monotone-convergence theorem,

$$\mathbb{P}\left(\sum_n \mathbb{I}_A \mathbb{P}(F_n \mid \mathcal{G}) = \lim_{k \rightarrow \infty} \sum_{i=0}^k \mathbb{I}_A \mathbb{P}(F_i \mid \mathcal{G}) = \lim_{k \rightarrow \infty} \mathbb{E}\left(\mathbb{I}_{\bigcup_{i=0}^k F_i} \mid \mathcal{G}\right) \mathbb{I}_A = \mathbb{E}(\mathbb{I}_{\bigcup_n F_n} \mid \mathcal{G}) = \mathbb{P}\left(\bigcup_n F_n \mid \mathcal{G}\right)\right) = 1,$$

where  $A \in \mathcal{G}$  is a set such that  $\mathbb{P}(A) = 1$ .

Consider the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ . A function  $\mathbb{P}_{\mathcal{G}} : \Omega \times \mathcal{F} \rightarrow [0, 1]$  is called a regular conditional probability given  $\mathcal{G}$  if

- There is a set  $A \in \mathcal{F}$  such that  $\mathbb{P}(A) = 1$  and, for every  $\omega \in A$ , the function  $\mathbb{P}_{\mathcal{G}}(\omega, \cdot) : \mathcal{F} \rightarrow [0, 1]$  is a probability measure on  $(\Omega, \mathcal{F})$ .
- For every  $F \in \mathcal{F}$ , the function  $\mathbb{P}_{\mathcal{G}}(\cdot, F) : \Omega \rightarrow [0, 1]$  is a version of the conditional expectation  $\mathbb{E}(\mathbb{I}_F \mid \mathcal{G})$  of  $\mathbb{I}_F$  given  $\mathcal{G}$ , so that  $\mathbb{P}_{\mathcal{G}}(\cdot, F) = \mathbb{P}(F \mid \mathcal{G}) = \mathbb{E}(\mathbb{I}_F \mid \mathcal{G})$  almost surely.

It can be shown that a regular conditional probability given  $\mathcal{G}$  exists under very permissive assumptions.

Consider the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ , a bounded Borel function  $h : \mathbb{R}^n \rightarrow \mathbb{R}$ , and the independent random variables  $X_1, X_2, \dots, X_n$ . Let  $h(X_1, X_2, \dots, X_n) : \Omega \rightarrow \mathbb{R}$  be given by

$$h(X_1, X_2, \dots, X_n)(\omega) = h(X_1(\omega), X_2(\omega), \dots, X_n(\omega)).$$

Furthermore, for every  $x_1 \in \mathbb{R}$ , let  $h(x_1, X_2, \dots, X_n) : \Omega \rightarrow \mathbb{R}$  be given by

$$h(x_1, X_2, \dots, X_n)(\omega) = h(x_1, X_2(\omega), \dots, X_n(\omega)).$$

Finally, let  $\gamma : \mathbb{R} \rightarrow \mathbb{R}$  be given by

$$\gamma(x_1) = \mathbb{E}(h(x_1, X_2, \dots, X_n)).$$

We will now show that  $\gamma(X_1) = \mathbb{E}(h(X_1, X_2, \dots, X_n) \mid X_1)$  almost surely, where  $\gamma(X_1) = \gamma \circ X_1$ .

For every  $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ , let  $h_{x_1} : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$  be given by  $h_{x_1}(x_2, \dots, x_n) = h(x_1, x_2, \dots, x_n)$ , and recall that  $h_{x_1}$  is a bounded Borel function. Furthermore, recall that the function  $Z : \Omega \rightarrow \mathbb{R}^n$  given by  $Z(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$  is  $\mathcal{F}/\mathcal{B}(\mathbb{R})^n$ -measurable and that the function  $Y : \Omega \rightarrow \mathbb{R}^{n-1}$  given by  $Y(\omega) = (X_2(\omega), \dots, X_n(\omega))$  is  $\mathcal{F}/\mathcal{B}(\mathbb{R})^{n-1}$ -measurable.

For every  $x_1 \in \mathbb{R}$ , note that  $h(X_1, X_2, \dots, X_n) = h \circ Z$  and  $h(x_1, X_2, \dots, X_n) = h_{x_1} \circ Y$ . Because  $h$  and  $h_{x_1}$  are Borel, for every  $B \in \mathcal{B}(\mathbb{R})$ , we know that  $Z^{-1}(h^{-1}(B)) \in \mathcal{F}$  and  $Y^{-1}(h_{x_1}^{-1}(B)) \in \mathcal{F}$ . Because  $h$  and  $h_{x_1}$  are bounded,  $h(X_1, X_2, \dots, X_n) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  and  $h(x_1, X_2, \dots, X_n) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ .

For every  $k \in \{1, \dots, n\}$ , let  $\mathcal{L}_k : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  denote the law of  $X_k$ . Because the random variables  $X_1, X_2, \dots, X_n$  are independent, recall that the joint law of  $X_i, X_{i+1}, \dots, X_n$  is given by  $\mathcal{L}_i \times \mathcal{L}_{i+1} \times \dots \times \mathcal{L}_n$ .

For every  $x_1 \in \mathbb{R}$ , because a previous result for laws extends to joint laws,

$$\gamma(x_1) = \int_{\Omega} h(x_1, X_2, \dots, X_n) d\mathbb{P} = \int_{\Omega} (h_{x_1} \circ Y) d\mathbb{P} = \int_{\mathbb{R}^{n-1}} h_{x_1} d(\mathcal{L}_2 \times \dots \times \mathcal{L}_n).$$

Because  $h_{x_1}$  is a bounded Borel function,

$$\gamma(x_1) = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} h(x_1, x_2, \dots, x_n) \mathcal{L}_n(dx_n) \dots \mathcal{L}_2(dx_2),$$

which also implies that  $\gamma$  is  $\mathcal{B}(\mathbb{R})$ -measurable, so that  $\gamma(X_1)$  is  $\sigma(X_1)$ -measurable.

For every  $B \in \mathcal{B}(\mathbb{R})$ , recall that  $\mathbb{I}_{X_1^{-1}(B)} = \mathbb{I}_B(X_1)$ . Therefore, for every  $X_1^{-1}(B) \in \sigma(X_1)$ ,

$$\int_{\Omega} h(X_1, X_2, \dots, X_n) \mathbb{I}_{X_1^{-1}(B)} d\mathbb{P} = \int_{\mathbb{R}^n} h \mathbb{I}_B(\rho_1) d(\mathcal{L}_1 \times \dots \times \mathcal{L}_n).$$

Because  $h \mathbb{I}_B(\rho_1)$  is bounded Borel function,

$$\int_{\Omega} h(X_1, X_2, \dots, X_n) \mathbb{I}_{X_1^{-1}(B)} d\mathbb{P} = \int_{\mathbb{R}} \mathbb{I}_B(x_1) \left[ \int_{\mathbb{R}} \dots \int_{\mathbb{R}} h(x_1, x_2, \dots, x_n) \mathcal{L}_n(dx_n) \dots \mathcal{L}_2(dx_2) \right] \mathcal{L}_1(dx_1).$$

Using the previous expression for  $\gamma(x_1)$  and a previous result for laws,

$$\int_{\Omega} h(X_1, X_2, \dots, X_n) \mathbb{I}_{X_1^{-1}(B)} d\mathbb{P} = \int_{\mathbb{R}} \mathbb{I}_B(x_1) \gamma(x_1) \mathcal{L}_1(dx_1) = \int_{\Omega} \gamma(X_1) \mathbb{I}_{X_1^{-1}(B)} d\mathbb{P}.$$

Because  $\mathbb{E}(\gamma(X_1)) = \mathbb{E}(h(X_1, X_2, \dots, X_n)) < \infty$ , the proof is complete.

Consider a measurable space  $(\Omega, \mathcal{F})$  and the sequence of  $\sigma$ -algebras  $(\mathcal{F}_n \subseteq \mathcal{F} \mid n \in \mathbb{N}^+)$ . For every  $n \in \mathbb{N}^+$ , let  $\mathcal{I}_n = \{\cap_{i=1}^n F_i \mid F_i \in \mathcal{F}_i \text{ for every } i \in \{1, \dots, n\}\}$ . We will now show that  $\mathcal{I} = \cup_n \mathcal{I}_n$  is a  $\pi$ -system on  $\Omega$  such that  $\sigma(\mathcal{I}) = \sigma(\mathcal{F}_1, \mathcal{F}_2, \dots)$ , where  $\sigma(\mathcal{F}_1, \mathcal{F}_2, \dots) = \sigma(\{F_1, F_2, \dots\}) = \sigma(\cup_n \mathcal{F}_n)$ .

For some  $n \in \mathbb{N}^+$ , consider the sets  $B \in \mathcal{I}_n$  and  $C \in \mathcal{I}_n$  such that  $B = \cap_{i=1}^n F_i$  and  $C = \cap_{i=1}^n F'_i$ , where  $F_i \in \mathcal{F}_i$  and  $F'_i \in \mathcal{F}_i$  for every  $i \in \{1, \dots, n\}$ . In that case,

$$B \cap C = \left( \bigcap_{i=1}^n F_i \right) \cap \left( \bigcap_{i=1}^n F'_i \right) = \bigcap_{i=1}^n (F_i \cap F'_i).$$

Because  $(F_i \cap F'_i) \in \mathcal{F}_i$  for every  $i \in \{1, \dots, n\}$ , we know that  $(B \cap C) \in \mathcal{I}_n$ . Therefore,  $\mathcal{I}_n$  is a  $\pi$ -system on  $\Omega$ . Because  $\Omega \in \mathcal{F}_n$  for every  $n \in \mathbb{N}^+$ , we know that  $\mathcal{I}_n \subseteq \mathcal{I}_{n+1}$ . Therefore,  $\mathcal{I} = \cup_n \mathcal{I}_n$  is also a  $\pi$ -system on  $\Omega$ .

Since  $\Omega \in \mathcal{F}_n$  for every  $n \in \mathbb{N}^+$ , we also know that  $\mathcal{F}_n \subseteq \mathcal{I}$  for every  $n \in \mathbb{N}^+$ . Therefore,  $\cup_n \mathcal{F}_n \subseteq \mathcal{I}$  and  $\sigma(\cup_n \mathcal{F}_n) \subseteq \sigma(\mathcal{I})$ . Consider a set  $(\cap_{i=1}^m F_i) \in \mathcal{I}$ , where  $m \in \mathbb{N}^+$  and  $F_i \in \mathcal{F}_i$  for every  $i \in \{1, \dots, m\}$ . Clearly,  $F_i \in \cup_n \mathcal{F}_n$  for every  $i \in \{1, \dots, m\}$ . Because  $\sigma(\cup_n \mathcal{F}_n)$  is a  $\sigma$ -algebra, we know that  $(\cap_{i=1}^m F_i) \in \sigma(\cup_n \mathcal{F}_n)$ , which implies  $\mathcal{I} \subseteq \sigma(\cup_n \mathcal{F}_n)$  and  $\sigma(\mathcal{I}) \subseteq \sigma(\cup_n \mathcal{F}_n)$ , completing the proof.

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and the sequence of independent  $\sigma$ -algebras  $(\mathcal{F}_n \subseteq \mathcal{F} \mid n \in \mathbb{N}^+)$ . We will now show that  $\sigma(\mathcal{F}_1, \dots, \mathcal{F}_k)$  and  $\sigma(\mathcal{F}_{k+1}, \mathcal{F}_{k+2}, \dots)$  are independent for every  $k \in \mathbb{N}^+$ .

From the previous proof, we know that  $\mathcal{I} = \{\cap_{i=1}^k F_i \mid F_i \in \mathcal{F}_i \text{ for every } i \in \{1, \dots, k\}\}$  is a  $\pi$ -system on  $\Omega$  such that  $\sigma(\mathcal{I}) = \sigma(\mathcal{F}_1, \dots, \mathcal{F}_k)$ . We also know that  $\mathcal{J} = \cup_n \{\cap_{i=k+1}^{k+n} F_i \mid F_i \in \mathcal{F}_i \text{ for every } i \in \{k+1, \dots, k+n\}\}$  is a  $\pi$ -system on  $\Omega$  such that  $\sigma(\mathcal{J}) = \sigma(\mathcal{F}_{k+1}, \mathcal{F}_{k+2}, \dots)$ .

Consider a set  $(\cap_{i=1}^k F_i) \in \mathcal{I}$ , where  $F_i \in \mathcal{F}_i$  for every  $i \in \{1, \dots, k\}$ , and a set  $(\cap_{i=k+1}^{k+n} F_i) \in \mathcal{J}$ , where  $n \in \mathbb{N}^+$  and  $F_i \in \mathcal{F}_i$  for every  $i \in \{k+1, \dots, k+n\}$ . Because  $\mathcal{F}_1, \dots, \mathcal{F}_{k+n}$  are independent,

$$\mathbb{P} \left( \left( \bigcap_{i=1}^k F_i \right) \cap \left( \bigcap_{i=k+1}^{k+n} F_i \right) \right) = \left( \prod_{i=1}^k \mathbb{P}(F_i) \right) \left( \prod_{i=k+1}^{k+n} \mathbb{P}(F_i) \right) = \mathbb{P} \left( \bigcap_{i=1}^k F_i \right) \mathbb{P} \left( \bigcap_{i=k+1}^{k+n} F_i \right),$$

which implies that  $\mathcal{I}$  and  $\mathcal{J}$  are independent. Because  $\sigma(\mathcal{I})$  and  $\sigma(\mathcal{J})$  are then independent, the proof is complete.

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a sequence of independent identically distributed random variables  $(X_n : \Omega \rightarrow \mathbb{R} \mid n \in \mathbb{N}^+)$ , each of which has the same law  $\mathcal{L}_X$  as the random variable  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $S_n : \Omega \rightarrow \mathbb{R}$  be a random variable given by  $S_n = X_1 + \dots + X_n$ . We will now show that

$$\mathbb{E}(X_k \mid S_n) = \mathbb{E}(X_k \mid S_n, S_{n+1}, \dots) = \frac{S_n}{n}$$

almost surely, where  $n \in \mathbb{N}^+$  and  $k \in \{1, \dots, n\}$ .

We will start by showing that  $\sigma(S_n, S_{n+1}, \dots) = \sigma(S_n, X_{n+1}, X_{n+2}, \dots)$  for every  $n \in \mathbb{N}^+$ . For every  $i \in \mathbb{N}^+$ , note that  $S_{n+i} = S_n + X_{n+1} + \dots + X_{n+i}$ , so that  $\sigma(S_{n+i}) \subseteq \sigma(S_n, X_{n+1}, X_{n+2}, \dots)$ . Therefore,  $\sigma(S_n, S_{n+1}, \dots) \subseteq \sigma(S_n, X_{n+1}, X_{n+2}, \dots)$ . For every  $i \in \mathbb{N}^+$ , note that  $X_{n+i} = S_{n+i} - S_{n+i-1}$ , so that  $\sigma(X_{n+i}) \subseteq \sigma(S_n, S_{n+1}, \dots)$ . Therefore,  $\sigma(S_n, X_{n+1}, X_{n+2}, \dots) \subseteq \sigma(S_n, S_{n+1}, \dots)$ .

Next, we will show that  $\sigma(S_n, X_k)$  and  $\sigma(X_{n+1}, X_{n+2}, \dots)$  are independent for every  $n \in \mathbb{N}^+$  and  $k \in \{1, \dots, n\}$ . Note that  $\sigma(S_n) \subseteq \sigma(X_1, \dots, X_n)$ . Therefore,  $\sigma(S_n, X_k) \subseteq \sigma(X_1, \dots, X_n)$ . From a previous result, we know that  $\sigma(X_1, \dots, X_n)$  and  $\sigma(X_{n+1}, X_{n+2}, \dots)$  are independent, so that  $\sigma(S_n, X_k)$  and  $\sigma(X_{n+1}, X_{n+2}, \dots)$  are independent.

By considering this independence, for every  $n \in \mathbb{N}^+$  and  $k \in \{1, \dots, n\}$ ,

$$\mathbb{E}(X_k \mid S_n, S_{n+1}, \dots) = \mathbb{E}(X_k \mid S_n, X_{n+1}, X_{n+2}, \dots) = \mathbb{E}(X_k \mid S_n)$$

almost surely.

For every  $n \in \mathbb{N}^+$ , recall that  $\mathbb{I}_{S_n^{-1}(B)} = \mathbb{I}_B(S_n)$  for all  $B \in \mathcal{B}(\mathbb{R})$ . Since  $X_k \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  for every  $k \in \{1, \dots, n\}$ ,

$$\int_{\Omega} X_k \mathbb{I}_{S_n^{-1}(B)} d\mathbb{P} = \int_{\Omega} X_k \mathbb{I}_B(S_n) d\mathbb{P} = \int_{\Omega} f_B(X_k, X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n) d\mathbb{P},$$

where  $f_B : \mathbb{R}^n \rightarrow \mathbb{R}$  is a Borel function given by  $f_B(x_1, \dots, x_n) = x_1 \mathbb{I}_B(x_1 + \dots + x_n)$ .

Because a previous result for laws extends to joint laws and  $X_1, \dots, X_n$  are independent,

$$\int_{\Omega} X_k \mathbb{I}_{S_n^{-1}(B)} d\mathbb{P} = \int_{\mathbb{R}^n} f_B d\mathcal{L}_{X_k, X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n} = \int_{\mathbb{R}^n} f_B d\mathcal{L}_X^n.$$

Therefore, for every  $n \in \mathbb{N}^+$ ,  $B \in \mathcal{B}(\mathbb{R})$ ,  $S_n^{-1}(B) \in \sigma(S_n)$ , and  $i, j \in \{1, \dots, n\}$ ,

$$\int_{\Omega} \mathbb{E}(X_i \mid S_n) \mathbb{I}_{S_n^{-1}(B)} d\mathbb{P} = \int_{\Omega} X_i \mathbb{I}_{S_n^{-1}(B)} d\mathbb{P} = \int_{\mathbb{R}^n} f_B d\mathcal{L}_X^n = \int_{\Omega} X_j \mathbb{I}_{S_n^{-1}(B)} d\mathbb{P} = \int_{\Omega} \mathbb{E}(X_j \mid S_n) \mathbb{I}_{S_n^{-1}(B)} d\mathbb{P},$$

so that  $\mathbb{E}(X_i \mid S_n) = \mathbb{E}(X_j \mid S_n)$  almost surely.

Finally, for every  $n \in \mathbb{N}^+$  and  $k \in \{1, \dots, n\}$ ,

$$n\mathbb{E}(X_k \mid S_n) = \sum_{i=1}^n \mathbb{E}(X_k \mid S_n) = \sum_{i=1}^n \mathbb{E}(X_i \mid S_n) = \mathbb{E}\left(\sum_{i=1}^n X_i \mid S_n\right) = \mathbb{E}(S_n \mid S_n) = S_n$$

almost surely, so that  $\mathbb{E}(X_k \mid S_n) = S_n/n$  almost surely.

## 10 Martingales

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ . A filtration  $(\mathcal{F}_n)_n$  is a sequence  $(\mathcal{F}_n \subseteq \mathcal{F} \mid n \in \mathbb{N})$  of  $\sigma$ -algebras such that  $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$  for every  $n \in \mathbb{N}$ . In that case, we let  $\mathcal{F}_{\infty} = \sigma(\mathcal{F}_0, \mathcal{F}_1, \dots) = \sigma(\cup_n \mathcal{F}_n)$ .

A filtered space  $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$  is composed of a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  and a filtration  $(\mathcal{F}_n)_n$ . Intuitively, at a given time  $n \in \mathbb{N}$ , for every  $\omega \in \Omega$ , recall that knowing  $\mathbb{I}_{F_n}(\omega)$  for every  $F_n \in \mathcal{F}_n$  allows knowing  $Z_n(\omega)$  for every  $\mathcal{F}_n$ -measurable random variable  $Z_n$ .

For any set  $\mathcal{C}$ , recall that a set (or sequence) of random variables  $Y = (Y_{\gamma} \mid \gamma \in \mathcal{C})$  on a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$  is called a stochastic process (parameterized by  $\mathcal{C}$ ).

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ . The natural filtration  $(\mathcal{F}_n)_n$  of the stochastic process  $(W_n \mid n \in \mathbb{N})$  is given by  $\mathcal{F}_n = \sigma(W_0, \dots, W_n)$  for every  $n \in \mathbb{N}$ . Intuitively, at a given time  $n \in \mathbb{N}$ , for every  $\omega \in \Omega$ , recall that knowing  $\mathbb{I}_{F_n}(\omega)$  for every  $F_n \in \sigma(W_0, \dots, W_n)$  is equivalent to knowing  $W_0(\omega), \dots, W_n(\omega)$ .

Consider a filtered space  $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ . A stochastic process  $(X_n \mid n \in \mathbb{N})$  is called adapted (to the filtration  $(\mathcal{F}_n)_n$ ) if  $X_n$  is  $\mathcal{F}_n$ -measurable for every  $n \in \mathbb{N}$ . Note that if  $(\mathcal{F}_n)_n$  is the natural filtration of the stochastic process  $(W_n \mid n \in \mathbb{N})$ , then there is a Borel function  $f_n : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  such that  $X_n = f_n(W_0, \dots, W_n)$ .

Consider a filtered space  $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ .

A stochastic process  $(X_n \mid n \in \mathbb{N})$  is called a martingale if  $(X_n \mid n \in \mathbb{N})$  is adapted;  $\mathbb{E}(|X_n|) < \infty$  for every  $n \in \mathbb{N}$ ; and  $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = X_{n-1}$  almost surely for every  $n \in \mathbb{N}^+$ .

A stochastic process  $(X_n \mid n \in \mathbb{N})$  is called a supermartingale if  $(X_n \mid n \in \mathbb{N})$  is adapted;  $\mathbb{E}(|X_n|) < \infty$  for every  $n \in \mathbb{N}$ ; and  $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \leq X_{n-1}$  almost surely for every  $n \in \mathbb{N}^+$ .



A stochastic process  $(X_n \mid n \in \mathbb{N})$  is called a submartingale if  $(X_n \mid n \in \mathbb{N})$  is adapted;  $\mathbb{E}(|X_n|) < \infty$  for every  $n \in \mathbb{N}$ ; and  $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \geq X_{n-1}$  almost surely for every  $n \in \mathbb{N}^+$ .

Consider an adapted stochastic process  $(X_n \mid n \in \mathbb{N})$  and suppose that  $\mathbb{E}(|X_n|) < \infty$  for every  $n \in \mathbb{N}$ . For every  $n \in \mathbb{N}^+$ , note that  $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = X_{n-1}$  almost surely if and only if  $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \leq X_{n-1} \leq \mathbb{E}(X_n \mid \mathcal{F}_{n-1})$  almost surely. Therefore,  $(X_n \mid n \in \mathbb{N})$  is a martingale if and only if  $(X_n \mid n \in \mathbb{N})$  is a supermartingale and a submartingale.

If  $(X_n \mid n \in \mathbb{N})$  is a supermartingale, then  $(-X_n \mid n \in \mathbb{N})$  is adapted;  $\mathbb{E}(|-X_n|) = \mathbb{E}(|X_n|) < \infty$  for every  $n \in \mathbb{N}$ ; and  $\mathbb{E}(-X_n \mid \mathcal{F}_{n-1}) \geq -X_{n-1}$  almost surely for every  $n \in \mathbb{N}^+$ . Therefore,  $(-X_n \mid n \in \mathbb{N})$  is a submartingale.

If  $(X_n \mid n \in \mathbb{N})$  is a submartingale, then  $(-X_n \mid n \in \mathbb{N})$  is adapted;  $\mathbb{E}(|-X_n|) = \mathbb{E}(|X_n|) < \infty$  for every  $n \in \mathbb{N}$ ; and  $\mathbb{E}(-X_n \mid \mathcal{F}_{n-1}) \leq -X_{n-1}$  almost surely for every  $n \in \mathbb{N}^+$ . Therefore,  $(-X_n \mid n \in \mathbb{N})$  is a supermartingale.

Consider an adapted stochastic process  $(X_n \mid n \in \mathbb{N})$  and suppose that  $\mathbb{E}(|X_n|) < \infty$  for every  $n \in \mathbb{N}$ . Furthermore, consider the stochastic process  $(X_n - X_0 \mid n \in \mathbb{N})$ . Because  $X_n - X_0$  is  $\mathcal{F}_n$ -measurable for every  $n \in \mathbb{N}$ , we know that  $(X_n - X_0 \mid n \in \mathbb{N})$  is adapted. Because  $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  is a vector space, we know that  $\mathbb{E}(|X_n - X_0|) < \infty$  for every  $n \in \mathbb{N}$ . By the linearity of conditional expectation,

$$\mathbb{E}(X_n - X_0 \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - \mathbb{E}(X_0 \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - X_0$$

almost surely for every  $n \in \mathbb{N}^+$ . Therefore:

- For every  $n \in \mathbb{N}^+$ ,  $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = X_{n-1}$  almost surely if and only if  $\mathbb{E}(X_n - X_0 \mid \mathcal{F}_{n-1}) = X_{n-1} - X_0$  almost surely. Therefore,  $(X_n \mid n \in \mathbb{N})$  is a martingale if and only if  $(X_n - X_0 \mid n \in \mathbb{N})$  is a martingale.
- For every  $n \in \mathbb{N}^+$ ,  $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \leq X_{n-1}$  almost surely if and only if  $\mathbb{E}(X_n - X_0 \mid \mathcal{F}_{n-1}) \leq X_{n-1} - X_0$  almost surely. Therefore,  $(X_n \mid n \in \mathbb{N})$  is a supermartingale if and only if  $(X_n - X_0 \mid n \in \mathbb{N})$  is a supermartingale.
- For every  $n \in \mathbb{N}^+$ ,  $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \geq X_{n-1}$  almost surely if and only if  $\mathbb{E}(X_n - X_0 \mid \mathcal{F}_{n-1}) \geq X_{n-1} - X_0$  almost surely. Therefore,  $(X_n \mid n \in \mathbb{N})$  is a submartingale if and only if  $(X_n - X_0 \mid n \in \mathbb{N})$  is a submartingale.

Consequently, it is common to assume that a stochastic process  $(X_n \mid n \in \mathbb{N})$  has  $X_0 = 0$  and  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ .

If  $(X_n \mid n \in \mathbb{N})$  is a martingale,  $n \in \mathbb{N}^+$ , and  $m < n$ , then

$$\mathbb{E}(X_n \mid \mathcal{F}_m) = \mathbb{E}(X_n \mid \mathcal{F}_{n-1} \mid \mathcal{F}_m) = \mathbb{E}(\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \mid \mathcal{F}_m) = \mathbb{E}(X_{n-1} \mid \mathcal{F}_m)$$

almost surely. Therefore, almost surely,

$$\mathbb{E}(X_n \mid \mathcal{F}_m) = \mathbb{E}(X_{n-1} \mid \mathcal{F}_m) = \dots = \mathbb{E}(X_{m+1} \mid \mathcal{F}_m) = \mathbb{E}(X_m \mid \mathcal{F}_m) = X_m.$$

If  $(X_n \mid n \in \mathbb{N})$  is a supermartingale,  $n \in \mathbb{N}^+$ , and  $m < n$ , then

$$\mathbb{E}(X_n \mid \mathcal{F}_m) = \mathbb{E}(X_n \mid \mathcal{F}_{n-1} \mid \mathcal{F}_m) = \mathbb{E}(\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \mid \mathcal{F}_m) \leq \mathbb{E}(X_{n-1} \mid \mathcal{F}_m)$$

almost surely. Therefore, almost surely,

$$\mathbb{E}(X_n \mid \mathcal{F}_m) \leq \mathbb{E}(X_{n-1} \mid \mathcal{F}_m) \leq \dots \leq \mathbb{E}(X_{m+1} \mid \mathcal{F}_m) \leq \mathbb{E}(X_m \mid \mathcal{F}_m) = X_m.$$

If  $(X_n \mid n \in \mathbb{N})$  is a submartingale,  $n \in \mathbb{N}^+$ , and  $m < n$ , then

$$\mathbb{E}(X_n \mid \mathcal{F}_m) = \mathbb{E}(X_n \mid \mathcal{F}_{n-1} \mid \mathcal{F}_m) = \mathbb{E}(\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \mid \mathcal{F}_m) \geq \mathbb{E}(X_{n-1} \mid \mathcal{F}_m)$$

almost surely. Therefore, almost surely,

$$\mathbb{E}(X_n \mid \mathcal{F}_m) \geq \mathbb{E}(X_{n-1} \mid \mathcal{F}_m) \geq \dots \geq \mathbb{E}(X_{m+1} \mid \mathcal{F}_m) \geq \mathbb{E}(X_m \mid \mathcal{F}_m) = X_m.$$

The next three examples illustrate the definition of martingales.

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ , a sequence of independent random variables  $(X_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})) \mid n \in \mathbb{N}^+$ , and suppose that  $\mathbb{E}(X_n) = 0$  for every  $n \in \mathbb{N}^+$ . Let  $S_n = X_1 + \dots + X_n$  for every  $n \in \mathbb{N}^+$  and  $S_0 = 0$ . We will now show that  $(S_n \mid n \in \mathbb{N})$  is a martingale.

Let  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$  for every  $n \in \mathbb{N}^+$  and  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ . Clearly,  $(S_n \mid n \in \mathbb{N})$  is adapted to the filtration  $(\mathcal{F}_n)_n$ . Because  $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  is a vector space,  $S_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  for every  $n \in \mathbb{N}$ . For every  $n \in \mathbb{N}^+$ ,

$$\mathbb{E}(S_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(S_{n-1} + X_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(S_{n-1} \mid \mathcal{F}_{n-1}) + \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = S_{n-1} + \mathbb{E}(X_n) = S_{n-1}$$

almost surely, where we used the fact that  $\sigma(X_n)$  is independent of  $\mathcal{F}_{n-1}$  for every  $n \in \mathbb{N}^+$ .

Consider a probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ , a sequence of independent random variables  $(X_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \mid n \in \mathbb{N}^+)$ , and suppose that  $\mathbb{E}(X_n) = 1$  for every  $n \in \mathbb{N}^+$ . Let  $M_n = X_1 \cdots X_n$  for every  $n \in \mathbb{N}^+$  and  $M_0 = 1$ . We will now show that  $(M_n \mid n \in \mathbb{N})$  is a martingale.

Let  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$  for every  $n \in \mathbb{N}^+$  and  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ . Clearly,  $(M_n \mid n \in \mathbb{N})$  is adapted to the filtration  $(\mathcal{F}_n)_n$ . Because  $X_1, \dots, X_n$  are independent,  $M_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  for every  $n \in \mathbb{N}$ . For every  $n \in \mathbb{N}^+$ ,

$$\mathbb{E}(M_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(M_{n-1}X_n \mid \mathcal{F}_{n-1}) = M_{n-1}\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = M_{n-1}\mathbb{E}(X_n) = M_{n-1}$$

almost surely, where we used the fact that  $\sigma(X_n)$  is independent of  $\mathcal{F}_{n-1}$  for every  $n \in \mathbb{N}^+$ .

Consider a filtered space  $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$  and a random variable  $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $M_n = \mathbb{E}(\xi \mid \mathcal{F}_n)$  almost surely for every  $n \in \mathbb{N}$ . We will now show that  $(M_n \mid n \in \mathbb{N})$  is a martingale.

Clearly,  $(M_n \in \mathcal{L}^1(\Omega, \mathcal{F}_n, \mathbb{P}) \mid n \in \mathbb{N})$  is adapted to the filtration  $(\mathcal{F}_n)_n$ . For every  $n \in \mathbb{N}^+$ ,

$$\mathbb{E}(M_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(\mathbb{E}(\xi \mid \mathcal{F}_n) \mid \mathcal{F}_{n-1}) = \mathbb{E}(\xi \mid \mathcal{F}_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(\xi \mid \mathcal{F}_{n-1}) = M_{n-1}$$

almost surely.

Consider a filtered space  $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ .

A stochastic process  $(C_n \mid n \in \mathbb{N})$  is called previsible if  $C_n$  is  $\mathcal{F}_{n-1}$  measurable for every  $n \in \mathbb{N}^+$ . Note that if  $(\mathcal{F}_n)_n$  is the natural filtration of the stochastic process  $(W_n \mid n \in \mathbb{N})$ , then there is a Borel function  $g_n : \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $C_n = g_n(W_0, \dots, W_{n-1})$  for every  $n \in \mathbb{N}^+$ .

The martingale transform  $(C \bullet X)$  of an adapted process  $X = (X_n \mid n \in \mathbb{N})$  by a previsible process  $C = (C_n \mid n \in \mathbb{N})$  is the adapted process  $((C \bullet X)_n \mid n \in \mathbb{N})$ , where  $(C \bullet X)_0 = 0$  and

$$(C \bullet X)_n = \sum_{k=1}^n C_k(X_k - X_{k-1})$$

for every  $n \in \mathbb{N}^+$ . Note that  $(C \bullet X)_n = (C \bullet X)_{n-1} + C_n(X_n - X_{n-1})$  for every  $n \in \mathbb{N}^+$ .

The following example illustrates the definition of martingale transform.

For every  $\omega \in \Omega$ , suppose that  $X_n(\omega) - X_{n-1}(\omega)$  represents the profit per unit stake in round  $n \in \mathbb{N}^+$  of a game. In that case,  $C_n(\omega)$  can be interpreted as the amount stake in round  $n \in \mathbb{N}^+$  by a particular gambling strategy  $C$ . For every  $n \in \mathbb{N}^+$  and  $\omega \in \Omega$ , the amount stake  $C_n(\omega)$  may rely on knowledge about  $\mathbb{I}_{\mathcal{F}_{n-1}}(\omega)$  for every  $\mathcal{F}_{n-1} \in \mathcal{F}_{n-1}$ , which includes at the very least knowledge about  $X_0(\omega), \dots, X_{n-1}(\omega)$  and  $C_0(\omega), \dots, C_{n-1}(\omega)$ . Finally, in this setting,  $(C \bullet X)_n(\omega)$  represents the profit after  $n \in \mathbb{N}^+$  rounds. Note that:

- If  $(X_n \mid n \in \mathbb{N})$  is a martingale, then  $\mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - X_{n-1} = 0$  almost surely for every  $n \in \mathbb{N}^+$ .
- If  $(X_n \mid n \in \mathbb{N})$  is a supermartingale, then  $\mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - X_{n-1} \leq 0$  almost surely for every  $n \in \mathbb{N}^+$ .
- If  $(X_n \mid n \in \mathbb{N})$  is a submartingale, then  $\mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - X_{n-1} \geq 0$  almost surely for every  $n \in \mathbb{N}^+$ .

Consider an adapted process  $X = (X_n \mid n \in \mathbb{N})$  and a previsible process  $C = (C_n \mid n \in \mathbb{N})$ . We will now show that if  $C_n \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  and  $X_n \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  for every  $n \in \mathbb{N}$ , then  $C_n(X_n - X_{n-1}) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  for every  $n \in \mathbb{N}^+$ .

Since  $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  is a vector space,  $(X_n - X_{n-1}) \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  for every  $n \in \mathbb{N}^+$ . By the Schwarz inequality,  $C_n(X_n - X_{n-1}) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ .

Consider an adapted process  $X = (X_n \mid n \in \mathbb{N})$  and a previsible process  $C = (C_n \mid n \in \mathbb{N})$ . We will now show that if  $|C_n| \leq K$  and  $\mathbb{E}(|X_n|) < \infty$  for every  $n \in \mathbb{N}$  and some  $K \in [0, \infty)$ , then  $C_n(X_n - X_{n-1}) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  for every  $n \in \mathbb{N}^+$ .

Since  $|C_n||X_n - X_{n-1}| \leq K|X_n - X_{n-1}|$  for every  $n \in \mathbb{N}^+$ , we know that  $\mathbb{E}(|C_n(X_n - X_{n-1})|) \leq K\mathbb{E}(|X_n - X_{n-1}|)$ . Because  $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  is a vector space, we know that  $C_n(X_n - X_{n-1}) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ .

Consider an adapted process  $X = (X_n \in \mathcal{L}^1(\Omega, \mathcal{F}_n, \mathbb{P}) \mid n \in \mathbb{N})$  and a previsible process  $C = (C_n \mid n \in \mathbb{N})$ . Furthermore, suppose that  $C_n(X_n - X_{n-1}) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  for every  $n \in \mathbb{N}^+$ .

First, recall that  $(C \bullet X)$  is adapted. Because  $(C \bullet X)_0 = 0$  and  $(C \bullet X)_n = (C \bullet X)_{n-1} + C_n(X_n - X_{n-1})$  for every  $n \in \mathbb{N}^+$ , we know that  $(C \bullet X)_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  for every  $n \in \mathbb{N}$ . Finally, for every  $n \in \mathbb{N}^+$ ,

$$\mathbb{E}((C \bullet X)_n \mid \mathcal{F}_{n-1}) = \mathbb{E}((C \bullet X)_{n-1} + C_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}) = (C \bullet X)_{n-1} + C_n\mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1})$$

almost surely. Therefore:

- If  $(X_n \mid n \in \mathbb{N})$  is a martingale, then,  $\mathbb{E}((C \bullet X)_n \mid \mathcal{F}_{n-1}) = (C \bullet X)_{n-1}$  almost surely for every  $n \in \mathbb{N}^+$ , so that  $(C \bullet X)$  is a martingale.
- If  $(X_n \mid n \in \mathbb{N})$  is a supermartingale and  $C$  is non-negative, then  $\mathbb{E}((C \bullet X)_n \mid \mathcal{F}_{n-1}) \leq (C \bullet X)_{n-1}$  almost surely for every  $n \in \mathbb{N}^+$ , so that  $(C \bullet X)$  is a supermartingale.
- If  $(X_n \mid n \in \mathbb{N})$  is a submartingale and  $C$  is non-negative, then  $\mathbb{E}((C \bullet X)_n \mid \mathcal{F}_{n-1}) \geq (C \bullet X)_{n-1}$  almost surely for every  $n \in \mathbb{N}^+$ , so that  $(C \bullet X)$  is a submartingale.

Consider a filtered space  $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ .

A function  $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$  is called a stopping time if  $\{T \leq n\} \in \mathcal{F}_n$  for every  $n \in \mathbb{N} \cup \{\infty\}$ . Intuitively, for every  $\omega \in \Omega$  and  $n \in \mathbb{N} \cup \{\infty\}$ , knowing  $\mathbb{I}_{F_n}(\omega)$  for every  $F_n \in \mathcal{F}_n$  allows knowing whether  $T(\omega) \leq n$ .

We will now show that  $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$  is a stopping time if and only if  $\{T = n\} \in \mathcal{F}_n$  for every  $n \in \mathbb{N} \cup \{\infty\}$ .

If  $T$  is a stopping time, then  $\{T \leq n\} \in \mathcal{F}_n$  and  $\{T \leq n-1\}^c \in \mathcal{F}_n$  for every  $n \in \mathbb{N}$ . Because  $\{T = n\} = \{T \leq n\} \cap \{T > n-1\}$ , we know that  $\{T = n\} \in \mathcal{F}_n$ . Furthermore,  $\{T = \infty\} = \bigcap_n \{T \leq n\}^c$ , so that  $\{T = \infty\} \in \mathcal{F}_\infty$ .

If  $\{T = n\} \in \mathcal{F}_n$  for every  $n \in \mathbb{N} \cup \{\infty\}$ , the fact that  $\{T \leq n\} = \bigcup_{k \leq n} \{T = k\}$  and  $\{T = k\} \in \mathcal{F}_n$  for every  $k \leq n$  implies that  $\{T \leq n\} \in \mathcal{F}_n$ .

The following example illustrates the definition of stopping time.

Consider an adapted process  $(A_n \mid n \in \mathbb{N})$  and a set  $B \in \mathcal{B}(\mathbb{R})$ . Let the function  $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$  be given by  $T(\omega) = \inf\{n \in \mathbb{N} \mid A_n(\omega) \in B\}$ , so that  $T(\omega) = \inf \emptyset = \infty$  if  $A_n(\omega) \notin B$  for every  $n \in \mathbb{N}$ . For every  $n \in \mathbb{N}$ ,

$$\{T \leq n\} = \{\omega \in \Omega \mid A_k(\omega) \in B \text{ for some } k \leq n\} = \bigcup_{k \leq n} \{\omega \in \Omega \mid A_k(\omega) \in B\} = \bigcup_{k \leq n} A_k^{-1}(B).$$

Because  $A_k$  is  $\mathcal{F}_n$ -measurable for every  $k \leq n$  and  $\{T \leq \infty\} \in \mathcal{F}_\infty$ , we know that  $T$  is a stopping time.

Consider an adapted process  $(X_n \mid n \in \mathbb{N})$  and a stopping time  $T$ . For some  $a \in \mathbb{R}$ , consider the set  $A$  given by

$$A = \{\omega \in \Omega \mid T(\omega) < \infty \text{ and } X_{T(\omega)}(\omega) \leq a\}.$$

We will now show that  $A \in \mathcal{F}_\infty$ . By definition,

$$A = \bigcup_{k \in \mathbb{N}} \{\omega \in \Omega \mid T(\omega) = k \text{ and } X_k(\omega) \leq a\} = \bigcup_{k \in \mathbb{N}} \{T = k\} \cap \{X_k \leq a\}.$$

Because  $\{T = k\} \cap \{X_k \leq a\} \in \mathcal{F}_k$  for every  $k \in \mathbb{N}$ , we know that  $A \in \mathcal{F}_\infty$ .

Consider an adapted process  $X = (X_n \mid n \in \mathbb{N})$  and a stopping time  $T$ . The stopped process  $X^T$  is the adapted process  $(X_n^T \mid n \in \mathbb{N})$ . For every  $n \in \mathbb{N}$ , the random variable  $X_n^T : \Omega \rightarrow \mathbb{R}$  is given by

$$X_n^T(\omega) = X_{\min(T(\omega), n)}(\omega) = \begin{cases} X_n(\omega), & \text{if } n \leq T(\omega), \\ X_{T(\omega)}(\omega), & \text{if } n > T(\omega). \end{cases}$$

We will now show that  $X^T$  is indeed adapted to the filtration  $(\mathcal{F}_n)_n$ . For every  $n \in \mathbb{N}$  and  $a \in \mathbb{R}$ ,

$$\{X_n^T \leq a\} = \{\omega \in \Omega \mid n \leq T(\omega) \text{ and } X_n(\omega) \leq a\} \cup \{\omega \in \Omega \mid n > T(\omega) \text{ and } X_{T(\omega)}(\omega) \leq a\}.$$

Let  $A_1$  denote the first set on the right side of the previous equation and  $A_2$  denote the second set.

Because  $\{n \leq T\} = \{n-1 \geq T\}^c \in \mathcal{F}_n$  and  $\{X_n \leq a\} \in \mathcal{F}_n$ , we know that  $A_1 \in \mathcal{F}_n$ . Regarding  $A_2$ , note that

$$A_2 = \{n > T\} \cap \{\omega \in \Omega \mid T(\omega) < \infty \text{ and } X_{T(\omega)}(\omega) \leq a\}.$$

Using a previous result,

$$A_2 = \{n > T\} \cap \bigcup_{k \in \mathbb{N}} \{T = k\} \cap \{X_k \leq a\} = \bigcup_{k \in \mathbb{N}} \{n > T\} \cap \{T = k\} \cap \{X_k \leq a\} = \bigcup_{k < n} \{T = k\} \cap \{X_k \leq a\}.$$

Because  $\{T = k\} \cap \{X_k \leq a\} \in \mathcal{F}_n$  for every  $k < n$ , we know that  $A_2 \in \mathcal{F}_n$ . Therefore,  $\{X_n^T \leq a\} \in \mathcal{F}_n$  for every  $n \in \mathbb{N}$  and  $a \in \mathbb{R}$ , so that  $X_n^T$  is  $\mathcal{F}_n$ -measurable.

Consider the adapted process  $X = (X_n \mid n \in \mathbb{N})$ , the stopping time  $T$ , and the process  $C = (C_n \mid n \in \mathbb{N})$ , where  $C_n = \mathbb{I}_{\{n \leq T\}}$  for every  $n \in \mathbb{N}$ . Note that  $C$  is previsible, since  $\{n \leq T\} = \{n-1 \geq T\}^c$  and  $\{n-1 \geq T\}^c \in \mathcal{F}_{n-1}$  for every  $n \in \mathbb{N}^+$ , which implies that  $\mathbb{I}_{\{n \leq T\}}$  is  $\mathcal{F}_{n-1}$ -measurable.

Now consider the martingale transform  $(C \bullet X) = ((C \bullet X)_n \mid n \in \mathbb{N})$ , so that  $(C \bullet X)_0 = 0$  and

$$(C \bullet X)_n(\omega) = \sum_{k=1}^n \mathbb{I}_{\{k \leq T\}}(\omega)(X_k(\omega) - X_{k-1}(\omega)) = \sum_{k=1}^{\min(T(\omega), n)} X_k(\omega) - X_{k-1}(\omega)$$

for every  $n \in \mathbb{N}^+$  and  $\omega \in \Omega$ . By reorganizing terms,

$$(C \bullet X)_n(\omega) = \sum_{k=1}^{\min(T(\omega), n)} X_k(\omega) - \sum_{k=0}^{\min(T(\omega), n)-1} X_k(\omega) = X_{\min(T(\omega), n)}(\omega) - X_0(\omega)$$

for every  $n \in \mathbb{N}^+$  and  $\omega \in \Omega$ . Therefore,  $(C \bullet X)_n = X_n^T - X_0 = X_n^T - X_0^T$  for every  $n \in \mathbb{N}$ .

When combined with previous results, this result implies the following:

- If  $(X_n \mid n \in \mathbb{N})$  is a martingale and  $T$  is a stopping time, then  $\mathbb{E}(X_n^T - X_0^T \mid \mathcal{F}_{n-1}) = X_{n-1}^T - X_0^T$  almost surely for every  $n \in \mathbb{N}^+$ , so that the stopped process  $X^T$  is a martingale.
- If  $(X_n \mid n \in \mathbb{N})$  is a supermartingale and  $T$  is a stopping time, then  $\mathbb{E}(X_n^T - X_0^T \mid \mathcal{F}_{n-1}) \leq X_{n-1}^T - X_0^T$  almost surely for every  $n \in \mathbb{N}^+$ , so that the stopped process  $X^T$  is a supermartingale.
- If  $(X_n \mid n \in \mathbb{N})$  is a submartingale and  $T$  is a stopping time, then  $\mathbb{E}(X_n^T - X_0^T \mid \mathcal{F}_{n-1}) \geq X_{n-1}^T - X_0^T$  almost surely for every  $n \in \mathbb{N}^+$ , so that the stopped process  $X^T$  is a submartingale.

Consider an adapted process  $X = (X_n \mid n \in \mathbb{N})$  and a stopping time  $T$ . Let  $X_T : \Omega \rightarrow \mathbb{R}$  be given by

$$X_T(\omega) = \begin{cases} X_{T(\omega)}(\omega), & \text{if } T(\omega) < \infty, \\ 0, & \text{if } T(\omega) = \infty. \end{cases}$$

We will now show that  $X_T$  is  $\mathcal{F}_\infty$ -measurable. For every  $a \in \mathbb{R}$ ,

$$\{X_T \leq a\} = \{\omega \in \Omega \mid T(\omega) < \infty \text{ and } X_{T(\omega)}(\omega) \leq a\} \cup \{\omega \in \Omega \mid T(\omega) = \infty \text{ and } 0 \leq a\}.$$

Let  $A_1$  denote the first set on the right side of the previous equation and  $A_2$  denote the second set. We have already shown that  $A_1 \in \mathcal{F}_\infty$ . If  $a \geq 0$ , then  $A_2 = \{T = \infty\}$ . Otherwise, if  $a < 0$ , then  $A_2 = \emptyset$ . In either case,  $A_2 \in \mathcal{F}_\infty$ . Therefore,  $\{X_T \leq a\} \in \mathcal{F}_\infty$  for every  $a \in \mathbb{R}$ , so that  $X_T$  is  $\mathcal{F}_\infty$ -measurable.

Consider a supermartingale  $X = (X_n \mid n \in \mathbb{N})$ , a stopping time  $T$ , and suppose at least one of the following:

1. The stopping time  $T$  is bounded, so that  $T \leq N$  for some  $N \in \mathbb{N}$ .
2. The stopping time  $T$  is almost surely finite, so that  $\mathbb{P}(T < \infty) = 1$ , and the stochastic process  $X$  is bounded, so that  $|X_n| \leq K$  for every  $n \in \mathbb{N}$  and some  $K \in [0, \infty)$ .
3. The stopping time  $T$  has finite expectation, so that  $\mathbb{E}(T) < \infty$ , and the stochastic process  $X$  has bounded increments, so that  $|X_n - X_{n-1}| \leq K$  for every  $n \in \mathbb{N}^+$  and some  $K \in [0, \infty)$ .

In that case, we will now show that  $\mathbb{E}(|X_T|) < \infty$  and  $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$ .

First, recall that the stopped process  $X^T$  is a supermartingale. Therefore,

$$\mathbb{E}(X_n^T) \mathbb{I}_\Omega = \mathbb{E}(X_n^T \mid \{\emptyset, \Omega\}) = \mathbb{E}(\mathbb{E}(X_n^T \mid \mathcal{F}_{n-1}) \mid \{\emptyset, \Omega\}) \leq \mathbb{E}(X_{n-1}^T \mid \{\emptyset, \Omega\}) = \mathbb{E}(X_{n-1}^T) \mathbb{I}_\Omega,$$

almost surely for every  $n \in \mathbb{N}^+$ , which implies that  $\mathbb{E}(X_n^T) \leq \mathbb{E}(X_{n-1}^T) \leq \dots \leq \mathbb{E}(X_1^T) \leq \mathbb{E}(X_0^T)$  for every  $n \in \mathbb{N}^+$ .

Suppose that the stopping time  $T$  is bounded, so that  $T \leq N$  for some  $N \in \mathbb{N}$ . In that case, for every  $\omega \in \Omega$ ,

$$X_T(\omega) = X_{T(\omega)}(\omega) = X_{\min(T(\omega), N)}(\omega) = X_N^T(\omega).$$

Because  $X_T = X_N^T$ , we know that  $\mathbb{E}(|X_T|) < \infty$ . From the previous result,  $\mathbb{E}(X_T) = \mathbb{E}(X_N^T) \leq \mathbb{E}(X_0^T) = \mathbb{E}(X_0)$ .

Suppose that the stopping time  $T$  is almost surely finite, so that  $\mathbb{P}(T < \infty) = 1$ , and the stochastic process  $X$  is bounded, so that  $|X_n| \leq K$  for every  $n \in \mathbb{N}$  and some  $K \in [0, \infty)$ . Because  $\mathbb{P}(T < \infty) = 1$ , we know that  $\mathbb{P}(\lim_{n \rightarrow \infty} X_n^T = X_T) = 1$ . Therefore, by the bounded convergence theorem, we know that  $\mathbb{E}(|X_T|) < \infty$  and  $\lim_{n \rightarrow \infty} \mathbb{E}(X_n^T) = \mathbb{E}(X_T)$ . Because  $\mathbb{E}(X_n^T) \leq \mathbb{E}(X_0)$  for every  $n \in \mathbb{N}^+$ , we know that  $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$ .

Finally, suppose that the stopping time  $T$  has finite expectation, so that  $\mathbb{E}(T) < \infty$ , and the stochastic process  $X$  has bounded increments, so that  $|X_n - X_{n-1}| \leq K$  for every  $n \in \mathbb{N}^+$  and some  $K \in [0, \infty)$ .

Because  $\mathbb{E}(T) < \infty$  implies  $\mathbb{P}(T < \infty) = 1$ , we know that  $\mathbb{P}(\lim_{n \rightarrow \infty} X_n^T = X_T) = 1$ . Therefore,

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_n^T - X_0 = X_T - X_0\right) = 1.$$

Note that  $|X_n^T - X_0| \leq KT$  for every  $n \in \mathbb{N}$ , since

$$|X_n^T(\omega) - X_0(\omega)| = \left| \sum_{k=1}^{\min(T(\omega), n)} X_k(\omega) - X_{k-1}(\omega) \right| \leq \sum_{k=1}^{\min(T(\omega), n)} |X_k(\omega) - X_{k-1}(\omega)| \leq \sum_{k=1}^{\min(T(\omega), n)} K \leq KT(\omega).$$

Because  $\mathbb{E}(KT) = K\mathbb{E}(T) < \infty$ , the dominated convergence theorem guarantees that  $\mathbb{E}(|X_T - X_0|) < \infty$  and

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n^T - X_0) = \mathbb{E}(X_T - X_0),$$

so that  $\lim_{n \rightarrow \infty} \mathbb{E}(X_n^T) = \mathbb{E}(X_T)$ . Because  $\mathbb{E}(X_n^T) \leq \mathbb{E}(X_0)$  for every  $n \in \mathbb{N}^+$ , we know that  $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$ .

Consider a martingale  $X = (X_n \mid n \in \mathbb{N})$ , a stopping time  $T$ , and suppose at least one of the following:

1. The stopping time  $T$  is bounded, so that  $T \leq N$  for some  $N \in \mathbb{N}$ .
2. The stopping time  $T$  is almost surely finite, so that  $\mathbb{P}(T < \infty) = 1$ , and the stochastic process  $X$  is bounded, so that  $|X_n| \leq K$  for every  $n \in \mathbb{N}$  and some  $K \in [0, \infty)$ .
3. The stopping time  $T$  has finite expectation, so that  $\mathbb{E}(T) < \infty$ , and the stochastic process  $X$  has bounded increments, so that  $|X_n - X_{n-1}| \leq K$  for every  $n \in \mathbb{N}^+$  and some  $K \in [0, \infty)$ .

In that case, we will now show that  $\mathbb{E}(|X_T|) < \infty$  and  $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ .

Because  $X$  is a supermartingale, we know that  $\mathbb{E}(|X_T|) < \infty$  and  $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$ . Because  $X$  is a submartingale, we know that  $-X = (-X_n \mid n \in \mathbb{N})$  is a supermartingale, so that  $\mathbb{E}(|(-X)_T|) < \infty$  and  $\mathbb{E}((-X)_T) \leq \mathbb{E}(-X_0)$ . Since  $(-X)_T = -X_T$ , we know that  $\mathbb{E}(X_T) \geq \mathbb{E}(X_0)$ , which implies  $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ .

Consider a martingale  $M = (M_n \mid n \in \mathbb{N})$  that has bounded increments, so that  $|M_n - M_{n-1}| \leq K_1$  for every  $n \in \mathbb{N}^+$  and some  $K_1 \in [0, \infty)$ . Consider also a previsible process  $C = (C_n \mid n \in \mathbb{N})$  that is bounded, so that  $|C_n| \leq K_2$  for every  $n \in \mathbb{N}$  and some  $K_2 \in [0, \infty)$ . Finally, consider a stopping time  $T$  with finite expectation, so that  $\mathbb{E}(T) < \infty$ . We will now show that  $\mathbb{E}((C \bullet M)_T) = 0$ .

Note that  $|C_n(M_n - M_{n-1})| = |C_n||M_n - M_{n-1}| \leq K_1 K_2$  for every  $n \in \mathbb{N}^+$ , so that  $\mathbb{E}(|C_n(M_n - M_{n-1})|) < \infty$ . Therefore, using a previous result, we know that  $(C \bullet M)$  is a martingale.

Because  $|(C \bullet M)_n - (C \bullet M)_{n-1}| = |C_n(M_n - M_{n-1})| \leq K_1 K_2$  for every  $n \in \mathbb{N}^+$ , we know that  $(C \bullet M)$  has bounded increments. Therefore, using a previous result, we know that  $\mathbb{E}(|(C \bullet M)_T|) < \infty$  and

$$\mathbb{E}((C \bullet M)_T) = \mathbb{E}((C \bullet M)_0) = \mathbb{E}(0) = 0.$$

Consider a supermartingale  $X = (X_n \mid n \in \mathbb{N})$  and a stopping time  $T$ . Furthermore, suppose that  $X_n \geq 0$  for every  $n \in \mathbb{N}$  and that  $\mathbb{P}(T < \infty) = 1$ . We will now show that  $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$ .

Because  $\mathbb{P}(T < \infty) = 1$ , we have  $\mathbb{P}(\lim_{n \rightarrow \infty} X_n^T = X_T) = 1$ . By the Fatou lemma,  $\mathbb{E}(X_T) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n^T)$ . Because  $\mathbb{E}(X_n^T) \leq \mathbb{E}(X_0)$  for every  $n \in \mathbb{N}$ , we know that  $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$ .

## Acknowledgements

I would like to thank Daniel Valesin for his guidance and the ideas behind many proofs found in these notes.

## License

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License .

## References

- [1] Williams, David. *Probability with Martingales*. Cambridge University Press, 1991.
- [2] Rosenthal, Jeffrey S. *A First Look at Rigorous Probability Theory*. World Scientific, 2006.
- [3] Pollard, D. *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, 2011.
- [4] Snoo, Henk de and Winkler, Henrik. *Measure and Integration, An Introduction*.