

INTRODUCTION TO PROBABILITY

CARSON JAMES

CONTENTS

1. Introduction	1
1.1. Purpose	1
2. Probability Framework	1

1. INTRODUCTION

1.1. Purpose.

2. PROBABILITY FRAMEWORK

3. PROBABILITY

3.1. Distributions.

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

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

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

Exercise 3.1.3. Let Ω be a set and \mathcal{L} a λ -system on Ω . Then

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

Proof. Straightforward. □

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

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

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

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

Exercise 3.1.5. Let Ω be a set and $\mathcal{C} \subset \mathcal{P}(\Omega)$. If \mathcal{C} is a λ -system and \mathcal{C} is a π -system, then \mathcal{C} is a σ -algebra.

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

Theorem 3.1.1. (*Dynkin's Theorem*)

Let Ω be a set.

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

Exercise 3.1.6. Let (Ω, \mathcal{F}) be a measurable space and μ, ν probability measures on (Ω, \mathcal{F}) . Put $\mathcal{L}_{\mu, \nu} = \{A \in \mathcal{F} : \mu(A) = \nu(A)\}$. Then $\mathcal{L}_{\mu, \nu}$ is a λ -system on Ω .

Proof.

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

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

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

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

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

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

Exercise 3.1.7. Let (Ω, \mathcal{F}) be a measurable space, μ, ν probability measures on (Ω, \mathcal{F}) and $\mathcal{P} \subset \mathcal{A}$ a π -system on Ω . Suppose that for each $A \in \mathcal{P}$, $\mu(A) = \nu(A)$. Then for each $A \in \sigma(\mathcal{P})$, $\mu(A) = \nu(A)$.

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

Definition 3.1.8. Let $F : \mathbb{R} \rightarrow \mathbb{R}$. Then F is said to be a **probability distribution function** if

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

Definition 3.1.9. Let P be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We define $F_P : \mathbb{R} \rightarrow \mathbb{R}$, by

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

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

Exercise 3.1.10. Let (Ω, \mathcal{F}, P) be a probability measure. Then F_P is a probability distribution function.

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

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

This implies that

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

. So F is right continuous.

(2) Clearly F_P is increasing.

(3) Continuity from below tells us that

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

and continuity from above tell us that

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

□

Exercise 3.1.11. Let μ, ν be probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then $F_\mu = F_\nu$ iff $\mu = \nu$.

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

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

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

Definition 3.1.12. Let (Ω, \mathcal{F}) be a measurable space and $X : \Omega \rightarrow \mathbb{R}$. Then X is said to be a **random variable** on (Ω, \mathcal{F}) if X is \mathcal{F} - $\mathcal{B}(\mathbb{R})$ measurable.

Definition 3.1.13. Let (Ω, \mathcal{F}, P) be a probability space and X a random variable on (Ω, \mathcal{F}) . We define the **probability distribution** of X , $P_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$, to be the measure

$$P_X = X_*P$$

so that for each $A \in \mathcal{B}(\mathbb{R})$,

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

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

$$F_X = F_{P_X}$$

Definition 3.1.14. Let (Ω, \mathcal{F}, P) be a probability space and X a random variable on (Ω, \mathcal{F}) . If $P_X \ll m$, we define the **probability density** of X , $f_X : \mathbb{R} \rightarrow \mathbb{R}$, by

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

Exercise 3.1.15. Let (Ω, \mathcal{F}, P) be a probability space and $(X_n)_{n \in \mathbb{N}}$ be a sequence of random variables on (Ω, \mathcal{F}) . Then for each $x \in \mathbb{R}$,

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

Proof. Let $\omega \in \left\{\liminf_{n \rightarrow \infty} X_n > x\right\}$. Then $x < \liminf_{n \rightarrow \infty} X_n(\omega) = \sup_{n \in \mathbb{N}} \left(\inf_{k \geq n} X_k(\omega)\right)$. So there exists $n^* \in \mathbb{N}$ such that $x < \inf_{k \geq n^*} X_k(\omega)$. Then for each $k \in \mathbb{N}$, $k \geq n^*$ implies that $x < X_k(\omega)$. So there exists $n^* \in \mathbb{N}$ such that for each $k \in \mathbb{N}$, $k \geq n^*$ implies that $\mathbf{1}_{\{X_k > x\}}(\omega) = 1$. Hence $\inf_{k \geq n^*} \mathbf{1}_{\{X_k > x\}}(\omega) = 1$. Thus $\liminf_{n \rightarrow \infty} \mathbf{1}_{\{X_k > x\}}(\omega) = \sup_{n \in \mathbb{N}} \left(\inf_{k \geq n} \mathbf{1}_{\{X_k > x\}}(\omega)\right) = 1$. Therefore $\omega \in \liminf_{n \rightarrow \infty} \{X_k > x\}$ and we have shown that

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

Then

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

□

Definition 3.1.16. Let (Ω, \mathcal{F}, P) be a probability space and $X \in L^+(\Omega) \cup L^1$. Define the **expectation of X**, $E[X]$, to be

$$\mathbb{E}[X] = \int X dP$$

3.2. Independence.

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

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

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

Note 3.2.1. We will explicitly say that for each $i = 1, \dots, n$, \mathcal{C}_i is independent when talking about the independence of the elements of \mathcal{C}_i to avoid ambiguity.

Definition 3.2.3. Let (Ω, \mathcal{F}, P) be a probability space and X_1, \dots, X_n random variables on (Ω, \mathcal{F}) . Then X_1, \dots, X_n are said to be **independent** if for each $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$, $X_1^{-1}B_1, \dots, X_n^{-1}B_n$ are independent.

Exercise 3.2.4. Let (Ω, \mathcal{F}, P) be a probability space and X_1, \dots, X_n random variables on (Ω, \mathcal{F}) . Then X_1, \dots, X_n are independent iff $\sigma(X_1), \dots, \sigma(X_n)$ are independent.

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

Exercise 3.2.5. Let (Ω, \mathcal{F}, P) be a probability space, X_1, \dots, X_n random variables on (Ω, \mathcal{F}) and $\mathcal{F}_1, \dots, \mathcal{F}_n \subset \mathcal{F}$ a collection of σ -algebras on Ω . Suppose that for each $i = 1, \dots, n$, X_i is \mathcal{F}_i -measurable. If $\mathcal{F}_1, \dots, \mathcal{F}_n$ are independent, then X_1, \dots, X_n are independent.

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

Exercise 3.2.6. Let (Ω, \mathcal{F}, P) be a probability space and $\mathcal{C}_1, \dots, \mathcal{C}_n \subset \mathcal{F}$. Suppose that for each $i = 1, \dots, n$, \mathcal{C}_i is a π -system and $\mathcal{C}_1, \dots, \mathcal{C}_n$ are independent, then $\sigma(\mathcal{C}_1), \dots, \sigma(\mathcal{C}_n)$ are independent.

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

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

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

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

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

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

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

Thus \mathcal{L} is a λ -system. Since $\mathcal{C}_1 \subset \mathcal{L}$ is a π -system, Dynkin's theorem tells us that $\sigma(\mathcal{C}_1) \subset \mathcal{L}$. Since $A_2 \in \mathcal{C}_2$ is arbitrary $\sigma(\mathcal{C}_1)$ and \mathcal{C}_2 are independent. The same reasoning implies that

$\sigma(\mathcal{C}_1)$ and $\sigma(\mathcal{C}_2)$ are independent. Let $A_2 \in \mathcal{C}_1, \dots, A_n \in \mathcal{C}_n$. We may do the same process with

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

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

Exercise 3.2.7. Let (Ω, \mathcal{F}, P) be a probability space, X_1, \dots, X_n random variables on (Ω, \mathcal{F}) . Then X_1, \dots, X_n are independent iff for each $x_1, \dots, x_n \in \mathbb{R}$,

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

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

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i). \text{ Conversely, suppose that for each}$$

$$x_1, \dots, x_n \in \mathbb{R}, P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i). \text{ Define } \mathcal{C} = \{(-\infty, x] : x \in \mathbb{R}\}.$$

Then $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$. For each $i = 1, \dots, n$, define $\mathcal{C}_i = X_i^{-1}\mathcal{C}$. Then for each $i = 1, \dots, n$, \mathcal{C}_i is a π -system and

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

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

Exercise 3.2.8. Let (Ω, \mathcal{F}, P) be a probability space and X_1, \dots, X_n random variables on (Ω, \mathcal{F}) . Define $X = (X_1, \dots, X_n)$. If X_1, \dots, X_n are independent, then

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

.

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

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

Put

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

Then \mathcal{P} is a π -system and

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

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

Exercise 3.2.9. Let (Ω, \mathcal{F}, P) be a probability space, X_1, \dots, X_n random variables on (Ω, \mathcal{F}) and $f_1, \dots, f_n : \mathbb{R} \rightarrow \mathbb{R} \in L^0$. Suppose that $f_1 \circ X_1, \dots, f_n \circ X_n \in L^+(\Omega)$ or $f_1 \circ X_1, \dots, f_n \circ X_n \in L^1(\Omega)$. If X_1, \dots, X_n are independent, then

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

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

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

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

3.3. L^p Spaces for Probability.

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

Definition 3.3.1. Let (Ω, \mathcal{F}, P) be a probability space and $X \in L^2$. Define the **variance of X** , $Var(X)$, to be

$$Var(X) = \mathbb{E}[(X - E[X])^2]$$

.

Definition 3.3.2. Let (Ω, \mathcal{F}, P) be a probability space and $X, Y \in L^2$. Define the

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

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

Exercise 3.3.4. Let (Ω, \mathcal{F}, P) be a probability space and $X, Y \in L^2$. Then the covariance is well defined and $Cov(X, Y)^2 \leq Var(X)Var(Y)$

Proof. By Holder's inequality,

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

So $Cov(X, Y)^2 \leq Var(X)Var(Y)$. □

Exercise 3.3.5. Let (Ω, \mathcal{F}, P) be a measure space and $X, Y \in L^2$. Then

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

Proof.

- (1) We have that

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

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

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

- (3) Part (1) implies that

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

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

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

(5) We have that

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

□

Definition 3.3.6. Let (Ω, \mathcal{F}, P) be a probability space and $X, Y \in L^2$. The **correlation of X and Y**, $\text{Cor}(X, Y)$, is defined to be

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

Exercise 3.3.7.

Exercise 3.3.8. Jensen's Inequality Let (Ω, \mathcal{F}, P) be a probability space, $X \in L^1$ and $\phi : \mathbb{R} \rightarrow \mathbb{R}$. If ϕ is convex, then

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

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

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

□

Exercise 3.3.9. Markov's Inequality: Let (Ω, \mathcal{F}, P) be a probability space and $X \in L^+$. Then for each $a \in (0, \infty)$,

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

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

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

Therefore

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

□

Exercise 3.3.10. Chebychev's Inequality: Let (Ω, \mathcal{F}, P) be a probability space and $X \in L^2$. Then for each $a \in (0, \infty)$,

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

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

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

□

Exercise 3.3.11. Chernoff's Bound: Let (Ω, \mathcal{F}, P) be a probability space and $X \in L^2$. Then for each $a, t \in (0, \infty)$,

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

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

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

□

Exercise 3.3.12. Weak Law of Large Numbers: Let (Ω, \mathcal{F}, P) be a probability space $(X_i)_{i \in \mathbb{N}} \subset L^2$. Suppose that $(X_i)_{i \in \mathbb{N}}$ are iid. Then

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

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

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

and

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

Let $\epsilon > 0$. Then

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

So

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

□

3.4. Borel Cantelli Lemma.

Definition 3.4.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$. We will define

$$P(A_n \text{ i.o.}) := P(\limsup_{n \rightarrow \infty} A_n)$$

and

$$P(A_n \text{ ev.}) := P(\liminf_{n \rightarrow \infty} A_n)$$

to be the **probability that A_n happens infinitely often** and the **probability that A_n happens eventually** respectively.

Exercise 3.4.2. Borel Cantelli Lemma: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$.

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

Proof.

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

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

Then

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

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

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

□

Exercise 3.4.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X_n)_{n \in \mathbb{N}} \subset L^0$ and $X \in L^0$.

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

Proof. (1)

□