

EE 226A NOTES

Ahmed Shakil

These notes were compiled while I took EE 226A, Random Processes in Systems. The class was taught by Professor Courtade in the fall of 2024.

CONTENTS

1	Elements of Probability Theory	1
1.1	Probability Spaces and Events	1
1.1.1	Infinitely often and Borel-Cantelli lemmas	2
1.1.2	Existence of probability measures	3
1.2	Random Variables and Expectation	4
1.2.1	Random variables and algebraic properties	4
1.2.2	Distribution functions and distributions	5

1 ELEMENTS OF PROBABILITY THEORY

1.1 Probability Spaces and Events

Definition 1.1.1: Kolmogorov's axioms

For any **probability space** (Ω, \mathcal{F}, P) , the function P is called a **probability measure**. It is assumed to satisfy Kolmogorov's axioms:

- i.) $P(A) \geq 0$ for all $A \in \mathcal{F}$;
- ii.) $P(\Omega) = 1$;
- iii.) if $A_1, A_2, \dots \in \mathcal{F}$ are disjoint events, then $P(\bigcup_{i \geq 1} A_i) = \sum_{i \geq 1} P(A_i)$.

The probability space we are working in encodes the model of our experiment, with the **measurable space** (Ω, \mathcal{F}) being the most fine-grained representation of outcomes we can hope to observe.

Theorem 1.1.2

For a probability space (Ω, \mathcal{F}, P) , the probability measure P enjoys the following properties:

- i.) Monotonicity: If $A \subset B$ are events, then $P(A) \leq P(B)$.
- ii.) Subadditivity (Union bound): If $(A_i)_{i \geq 1}$ is a sequence of events in \mathcal{F} and $A = \bigcup_{i \geq 1} A_i$, then $P(A) \leq \sum_{i \geq 1} P(A_i)$.
- iii.) Continuity from below: If $A_1 \subset A_2 \subset \dots$ are events in \mathcal{F} and $A = \bigcup_{i \geq 1} A_i$, then $P(A_i) \rightarrow P(A)$.
- iv.) Continuity from above: If $A_1 \supset A_2 \supset \dots$ are events in \mathcal{F} and $A = \bigcap_{i \geq 1} A_i$, then $P(A_i) \rightarrow P(A)$.

Proof.

- i.) Monotonicity follows from the following: $P(A) \leq P(A) + P(A^C \cap B) = P(B)$.
- ii.) Define $E_1 = A_1$ and $A_i = A_i \cap (\bigcup_{j < i} A_j)^C$ for $i \geq 2$. Then the E_i 's are disjoint, $E_i \subseteq A_i$ and $A = \bigcup_{i \geq 1} A_i = \bigcup_{i \geq 1} E_i$. Now we have $P(A) = P(\bigcup_{i \geq 1} E_i) = \sum_{i \geq 1} P(E_i) \leq \sum_{i \geq 1} P(A_i)$.
- iii.) Define E_i 's as above, and note that

$$P(A) = \sum_{i \geq 1} P(E_i) = \lim_{n \rightarrow \infty} \sum_{i \geq 1}^n P(E_i) = \lim_{n \rightarrow \infty} P(A_n).$$

- iv.) We now apply the previous part to $A_1^C \subset A_2^C \subset \dots$.

$$\lim_{n \rightarrow \infty} P(A_n) = 1 - \lim_{n \rightarrow \infty} P(A_n^C) = 1 - P(A^C) = P(A).$$

□

Theorem 1.1.3: Law of total probability

If events A_1, A_2, \dots partition Ω , then

$$P(B) = \sum_{i \geq 1} P(A_i \cap B), \quad B \in \mathcal{F}.$$

1.1.1 Infinitely often and Borel-Cantelli lemmas**Definition 1.1.4: Infinitely often**

$$\{A_n \text{ infinitely often}\} = \bigcap_{n \geq 1} \bigcup_{i \geq n} A_i.$$

We should understand $\{A_n \text{ i.o.}\}$ to be the set of samples $\omega \in \Omega$ such that $\omega \in A_i$ for infinitely many $i \geq 1$.

Lemma 1.1.5: Borel-Cantelli

Let A_1, A_2, \dots be a sequence of events. If

$$\sum_{i \geq 1} P(A_i) < \infty$$

then $P(\{A_i \text{ infinitely often}\}) = 0$.

Proof. Observe that $(\bigcup_{i \geq n} A_i)_{n \geq 1}$ is a decreasing sequence of events. Therefore, continuity from above and subadditivity together imply

$$P\left(\bigcap_{n \geq 1} \bigcup_{i \geq n} A_i\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{i \geq n} A_i\right) \leq \lim_{n \rightarrow \infty} \sum_{i \geq n} P(A_i) \rightarrow 0.$$

□

Definition 1.1.6: Independent events

A collection of events A_1, A_2, \dots are **independent** if

$$P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i)$$

for every finite subset $S \subset \{1, 2, 3, \dots\}$. If A_1, A_2, \dots are independent, then A_1^C, A_2^C, \dots are independent. By induction, the complements A_1^C, A_2^C, \dots are also independent.

Lemma 1.1.7: Converse to Borel-Cantelli

Let A_1, A_2, \dots be independent events. If

$$\sum_{i \geq 1} P(A_i) = \infty,$$

then $P(\{A_i \text{ infinitely often}\}) = 1$.

Proof. By definitions and continuity from above,

$$P(\{A_i \text{ i.o.}\}) = \lim_{n \rightarrow \infty} P\left(\bigcup_{i \geq n} A_i\right) = 1 - \lim_{n \rightarrow \infty} P\left(\bigcap_{i \geq n} A_i^c\right).$$

By independence, we have for any $m \geq n$

$$P\left(\bigcap_{i=n}^m A_i^c\right) = \prod_{i=n}^m P(A_i^c) = \prod_{i=n}^m (1 - P(A_i)) \leq \exp\left(-\sum_{i=n}^m P(A_i)\right),$$

where we made use of the inequality $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$. Since $\sum_{i \geq 1} P(A_i)$ doesn't converge, we must have that

$$P\left(\bigcap_{i \geq n} A_i^c\right) = \lim_{m \rightarrow \infty} P\left(\bigcap_{i=n}^m A_i^c\right) \leq \exp\left(-\sum_{i \geq n} P(A_i)\right) = 0.$$

□

1.1.2 Existence of probability measures

Some sets are simply not measurable, we'll call such sets non-measurable. A set which is non-measurable cannot be assigned a probability. In general, we tend to stick with σ -algebras, since they have nice properties. A deep theorem from probability theory called **Carathéodory's extension theorem** basically tells us that as long as we assign probabilities to a small collection of events in a consistent way, then there exists a natural and unique assignment of probabilities to the σ -algebra generated by the original collection.

Theorem 1.1.8: Carathéodory's extension theorem

Suppose \mathcal{G} is a family of subsets of Ω that satisfies the following (relatively modest) properties:

i.) $\emptyset, \Omega \in \mathcal{G}$;

ii.) if $A, B \in \mathcal{G}$, then $A \cap B \in \mathcal{G}$;

iii.) if $A, B \in \mathcal{G}$, then there is a *finite* number of *disjoint* sets $C_1, \dots, C_n \in \mathcal{G}$ such that $A \setminus B = \bigcup_{i=1}^n C_i$.

(Note: (iii) is weaker than imposing the assumption $A \in \mathcal{G} \implies A^c \in \mathcal{G}$.)

The extension theorem says that if we assign numbers (i.e., probabilities) $p(A)$ to the sets $A \in \mathcal{G}$ so that

A. $p(A) \geq 0$ for $A \in \mathcal{G}$;

B. $p(\Omega) = 1$;

C. if $B \in \mathcal{G}$ and $A_1, A_2, \dots \in \mathcal{G}$ are disjoint with $B = \bigcup_{i \geq 1} A_i$, then $p(B) = \sum_{i \geq 1} p(A_i)$,

then there exists a unique probability measure P on $\sigma(\mathcal{G})$ that satisfies A-C and has the property that $P(A) = p(A)$ for all $A \in \mathcal{G}$.

1.2 Random Variables and Expectation

1.2.1 Random variables and algebraic properties

Definition 1.2.1: Random Variable

We define a random variable to be a function $X : \Omega \rightarrow \overline{\mathbb{R}}$ that satisfies

$$\{\omega \in \Omega : X(\omega) \leq \alpha\} \in \mathcal{F} \text{ for each } \alpha \in \overline{\mathbb{R}}.$$

Note that $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

A function $X : \Omega \rightarrow \overline{\mathbb{R}}$ satisfying the definition above is said to be \mathcal{F} -measurable. If X does not take values $\pm\infty$ (say, with probability one), then we say it is a **real-valued random variable**.

Proposition 1.2.2

If X is a random variable, then pX and $|X|^p$ are random variables for $p \in \mathbb{R}$. Moreover, if X, Y are real-valued random variables, then $X + Y$, and XY are also random variables.

Proof. We leave the first claim as an exercise. For the second, note that we can write

$$\{\omega : X(\omega) + Y(\omega) > \alpha\} = \bigcup_{q \in \mathbb{Q}} \{\omega : X(\omega) > q\} \cap \{\omega : Y(\omega) > \alpha - q\}.$$

Since \mathcal{F} is closed under countable unions, intersections, and complements, it follows from the assumption that X, Y are random variables that $\{\omega : X(\omega) + Y(\omega) \leq \alpha\} \in \mathcal{F}$. The third claim now follows easily by writing $XY = [(X + Y)^2 - (X - Y)^2] / 4$. \square

Proposition 1.2.3

If $(X_n)_{n \geq 1}$ is a sequence of random variables defined on a common probability space (Ω, \mathcal{F}, P) , then

- $\sup_{n \geq 1} X_n$ and $\inf_{n \geq 1} X_n$ are random variables; and
- $\limsup_{n \rightarrow \infty} X_n$ and $\liminf_{n \rightarrow \infty} X_n$ are random variables; and
- if $\lim_{n \rightarrow \infty} X_n$ exists point wise, it is also a random variable.

Proof. Note that $\sup_{n \geq 1} X_n(\omega) \leq a$ if and only if $X_n(\omega) \leq a$ for each $n \geq 1$. It follows that

$$\left\{ \omega : \sup_{n \geq 1} X_n(\omega) \leq a \right\} = \bigcap_{n \geq 1} \{\omega : X_n(\omega) \leq a\} \in \mathcal{F},$$

where we used the fact that \mathcal{F} is closed under countable intersections. Hence, $\sup_{n \geq 1} X_n$ is a random variable. The rest of the claims are consequences of this. Indeed, since $\inf_{n \geq 1} X_n = -\sup_{n \geq 1} (-X_n)$, it holds that $\inf_{n \geq 1} X_n$ is also a random variable. So are $\limsup_{n \rightarrow \infty} X_n = \inf_{m \geq 1} \sup_{n \geq m} X_n$, and $\liminf_{n \rightarrow \infty} X_n$ by similar logic. As a result, if $\lim_{n \rightarrow \infty} X_n$ exists pointwise, then it is equal to $\limsup_{n \rightarrow \infty} X_n$ and is therefore also a random variable. \square

Definition 1.2.4: Almost sure equivalence of random variables

If X, Y are random variables and $P(\{\omega : X(\omega) \neq Y(\omega)\}) = 0$, then we say $X = Y$ almost surely (abbreviated a.s.), or $X = Y$ with probability one.

1.2.2 Distribution functions and distributions**Definition 1.2.5: Distribution function**

A random variable X on a probability space (Ω, \mathcal{F}, P) is described in part by its distribution function $F_X : \mathbb{R} \rightarrow [0, 1]$, defined as

$$F_X(x) := P\{X \leq x\}, \quad x \in \mathbb{R}.$$

Theorem 1.2.6: Properties of the distribution function

A function $F : \mathbb{R} \rightarrow [0, 1]$ is the distribution function of a random variable if and only if

- i.) F is nondecreasing
- ii.) F is right-continuous, that is $\lim_{y \downarrow x} F(y) = F(x)$, for all $x \in \mathbb{R}$.

Moreover, F is the distribution function of a real-valued random variable if and only if it further holds that

$$\lim_{x \rightarrow -\infty} F(x) = 0 \text{ and } \lim_{x \rightarrow \infty} F(x) = 1.$$

Proof. If F is a distribution function for X , then (i) follows by monotonicity of P since

$$\{\omega : X(\omega) \leq x\} \subset \{\omega : X(\omega) \leq x'\}, \quad x \leq x'.$$

Next, for $x \in \mathbb{R}$, note that

$$\{\omega : X(\omega) \leq x\} = \bigcap_{n \geq 1} \left\{ \omega : X(\omega) \leq x + \frac{1}{n} \right\}.$$

Continuity from above again gives $\lim_{n \rightarrow \infty} F(x + \frac{1}{n}) = F(x)$. This, together with (i), gives (ii).

If we have X real-valued, then we can write $\Omega = \bigcup_{n \geq 1} \{\omega : X(\omega) \leq n\}$. Hence, we find that $\lim_{n \rightarrow \infty} F(n) = 1$ follows by continuity from below and $P(\Omega) = 1$; together with (i), this implies $\lim_{x \rightarrow \infty} F(x) = 1$. Similarly, since $\emptyset = \bigcap_{n \geq 1} \{\omega : X(\omega) \leq -n\}$, continuity from above together with $P(\emptyset) = 0$ implies $\lim_{n \rightarrow \infty} F(-n) = 0$. As before, with (i) this yields $\lim_{x \rightarrow -\infty} F(x) = 0$.

For the other direction of the proof we refer to the textbook. □

Remark 1.2.7

Let X be a random variable with distribution F_X . It follows by continuity from above and below, respectively that

$$P\{X = -\infty\} = \lim_{x \rightarrow -\infty} F_X(x) \text{ and } P\{X = \infty\} = 1 - \lim_{x \rightarrow \infty} F_X(x).$$

These limits are always well-defined by monotonicity of F_X .

Definition 1.2.8: Law of a random variable

We define the law of a random variable to be

$$L_X(B) := P\{X \in B\}, \quad B \in \mathcal{B}_{\mathbb{R}}.$$

This defines a probability measure on \mathbb{R} equipped with the Borel σ -algebra. Indeed, $L_X([-\infty, x]) = F_X(x)$ for all $x \in \mathbb{R}$, where F_X is the distribution function of X . So, by Carathéodory's extension theorem and definition of the Borel σ -algebra, L_X is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. The **law** of X is synonymous with the distribution of X .

Definition 1.2.9: Discrete random variable

A real-valued random variable X is said to be a **discrete random variable** if its distribution function is a step function; that is, there is a countable set $(x_i)_{i \geq 1} \subset \mathbb{R}$ and a function $p_X : \mathbb{R} \rightarrow [0, 1]$ such that

$$F_X(x) = \sum_{i \geq 1: x_i \leq x} p_X(x_i) = \sum_{i \geq 1} p_X(x_i) 1_{[x_i, \infty)}(x), \quad x \in \mathbb{R}.$$

In this case, the function p_X is called the **probability mass function** (pmf) of X .

Definition 1.2.10: Continuous random variable

A real-valued random variable X is said to be a **continuous random variable** if its distribution function F_X is absolutely continuous. In other words, there exists a function $f_X : \mathbb{R} \rightarrow [0, \infty)$ such that

$$F_X(x) = \int_{-\infty}^x f_X(u) \, du, \quad x \in \mathbb{R}.$$

The function f_X is called the **probability density function** (pdf) of X .

Definition 1.2.11: Random vector

A **random vector** is a finite collection of random variables, and a similar story is true here.