1 Probability

1.1 Probability spaces

We will briefly look at the definition of a probability space, probability measures, conditional probability and independence of probability events.

Definition 1.1. *The set of all possible outcomes of an experiment is called the* sample space *and is denoted by* Ω*.*

Definition 1.2. *A collection* $\mathcal{F}$ *of subsets of the sample space* Ω *is called a* σ-algebra *(or a* σ*-field) if it satisfies the following conditions:*

1. *The empty set, denoted by* $\emptyset$*, is an element of* $\mathcal{F}$*. We write* $\emptyset \in \mathcal{F}$*.*
2. *If* $A \in \mathcal{F}$ *then* $A^c \in \mathcal{F}$*.*
3. *If the countable collection of sets* $A_1, A_2, \ldots$ *is in* $\mathcal{F}$ *(we write by* $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F}$*) then*
$$\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}.$$

Remark 1.3.

1. The smallest possible σ-algebra is $\{\emptyset, \Omega\}$.
2. If $A \subseteq \Omega$ then $\{\emptyset, A, A^c, \Omega\}$ is a σ-algebra.
3. The collection of all subsets of Ω is a σ-algebra. Sadly if Ω is uncountable then it is often not possible to define a probability measure on this σ-algebra and in this case it is not of much practical use.

We already mentioned probability measure, but what is it exactly?

Definition 1.4. *A* probability measure $\mathbb{P}$ *on* $(\Omega, \mathcal{F})$ *is a function* $\mathbb{P} : \mathcal{F} \to [0,1]$ *satisfying the following conditions:*

1. $\mathbb{P}(\emptyset) = 0$ *and* $\mathbb{P}(\Omega) = 1$*.*
2. *If* $\{A_i\}_{i\in\mathbb{N}}$ *is a collection of disjoint elements of* $\mathcal{F}$ *in the sense that* $A_i \cap A_j = \emptyset$ *whenever* $i \neq j$ *then*
$$\mathbb{P}\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mathbb{P}(A_i).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ *comprising a set* Ω *a* σ*-algebra* $\mathcal{F}$ *and a probability measure* $\mathbb{P}$ *is called a* probability space*.*

To check your understanding you may want to answer the following questions.

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space. Is it possible that there is $A \in \mathcal{F}$ such that $P(A) = 0$?
2. Let $\{A_i\}_{i\in\mathbb{N}}$ be a collection of disjoint elements of $\mathcal{F}$. Is there $c \in \mathbb{R}$ such that
$$\sum_{i\in\mathbb{N}} \mathbb{P}(A_i) \leq c\,?$$

Definition 1.5. *If* $\mathbb{P}(B) > 0$ *then the* conditional probability that A occurs given that B occurs *is defined to be*

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Definition 1.6. *Events A and B are called* independent *if* $P(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

1.2 Random variables

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

Definition 1.7. *Let $\mathcal{G}$ be a collection of subsets of Ω. The intersection of all σ-algebras contained in $\mathcal{F}$ that themselves contain $\mathcal{G}$ is called the σ-algebra* generated *by $\mathcal{G}$ and we will denote it by $\sigma(\mathcal{G})$. We can write*

$$\sigma(\mathcal{G}) := \bigcap \{\mathcal{H} \subseteq \mathcal{F} : \mathcal{H} \text{ is a } \sigma\text{-algebra and } \mathcal{G} \subseteq \mathcal{H}\}.$$

We take $d \in \mathbb{N}$ indicating the dimension of the space that our random variables take values in. Note that it is possible to define random variables taking values in more general spaces than $\mathbb{R}^d$ but we shall not need that.

Definition 1.8. *The* Borel *σ-algebra on $\mathbb{R}^d$ is the σ-algebra generated by the collection of all open sets in $\mathbb{R}^d$. We will denote it by $\mathcal{B}(\mathbb{R}^d)$.*

Since we know that the complement of an open set is a closed set it is not hard to see that this is equivalent to saying that the Borel σ-algebra is generated by all closed sets in $\mathbb{R}^d$. Furthermore, this definition extends to any metric spaces.

We can finally define random variables:

Definition 1.9. *A* random variable *is a function $X : \Omega \to \mathbb{R}^d$ with the property that*

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

for any $B \in \mathcal{B}(\mathbb{R}^d)$.

We say that X is $\mathcal{F}$-measurable.

Here we used the *pre-image* of set a B under X and denoted it by $X^{-1}(B)$. It is important to note that this has nothing to do with the inverse of a function. The function inverse is only defined for for one-to-one and onto functions, while the pre-image of a set under X *always* exists.

For any real valued random variable we can define its distribution function.

Definition 1.10. *The* distribution function *of a random variable $X : \Omega \to \mathbb{R}$ is the function $F_X : \mathbb{R} \to [0, 1]$ given by*

$$F_X(x) = \mathbb{P}(X \leq x).$$

We now look at two special types of random variables:

Definition 1.11. *Let $S \subset \mathbb{R}^d$ be a set containing only countably many elements. A random variable $X : \Omega \to S$ is then called* discrete.

Definition 1.12. *The random variable* $X : \Omega \to \mathbb{R}$ *is called* continuous *if there is a function* $f_X : \mathbb{R} \to [0, \infty)$ *such that its distribution function can be expressed, for any* $x \in \mathbb{R}$, *as*

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du.$$

The function f_X *is called the* probability density function.

The normal distribution. The probability density function of normal distribution with mean μ and variance σ^2 is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{|x-\mu|^2}{2\sigma^2}\right).$$

Note that there are random variables that are neither continuous nor discrete. Further note that saying that a random variable is continuous is not related to the continuity of the function X, it is instead related to the continuity of the distribution function of the random variable X.

We will now turn to what it means for random variables to be independent. It may be useful to recall that we call two events A and B independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Definition 1.13. *The σ-algebra* generated by a random variable $X : \Omega \to \mathbb{R}^d$ *is the collection of all pre-images of elements of the Borel σ-algebra* $\mathcal{B}(\mathbb{R}^d)$ *and is denoted by* $\sigma(X)$. *We can write*

$$\sigma(X) := X^{-1}(\mathcal{B}(\mathbb{R}^d)) := \left\{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^d)\right\}.$$

While in the definition we called the collection of sets $\sigma(X)$ a σ-algebra it is perhaps not immediately obvious that this collection is indeed a σ-algebra. It may be a good exercise to show that this is indeed the case.

Definition 1.14. *Two σ-algebras* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *over the same sample space* Ω *are called* independent *if for all* $A \in F_1$ *and* $B \in \mathcal{F}_2$ *the events* A *and* B *are independent.*

Two random variables $X : \Omega \to \mathbb{R}^d$ *and* $Y : \Omega \to \mathbb{R}^d$ *are called* independent *if the σ-algebras* $\sigma(X)$ *and* $\sigma(Y)$ *are independent.*

We will return to independence briefly once we have defined expectation, variance and covariance.

1.3 Integration and the Expectation Operator

Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be given. We would like to define the integral with respect to a probability measure. To this end, we first define the integral for special type of random variables.

Definition 1.15. *A random variable* $X : \Omega \to \mathbb{R}^d$ *is called* simple *if there exist real numbers* $x_1, x_2, \ldots, x_N$ *and elements of* $\mathcal{F}$ *denoted* $A_1, A_2, \ldots, A_N$ *such that*

$$X(\omega) = \sum_{k=1}^{N} x_k \mathbb{1}_{A_k}(\omega),$$

where we have used the indicator function of an event defined, for any event B, *as*

$$\mathbb{1}_B(\omega) := \begin{cases} 1 & \text{if } \omega \in B, \\ 0 & \text{otherwise.} \end{cases}$$

So simple functions can only take certain fixed values on certain sets. In a sense they are like they are like discrete random variables that you are no doubt familiar with. We all know that for a discre random variable Y, its expectation is

$$\mathbb{E}(Y) = \sum_{k=1}^{N} y_k \mathbb{P}(Y = y_k).$$

We define the expectation for simple random variables analogously.

Definition 1.16. *If $X : \Omega \to \mathbb{R}^d$ is a simple random variable then we can define its integral over Ω and thus its expectation as*

$$\mathbb{E}(X) := \int_\Omega X(\omega) d\mathbb{P}(\omega) := \sum_{k=1}^{N} x_k \mathbb{P}(A_k).$$

If $B \in \mathcal{F}$ and X is a simple random variable then we can easily check that $X\mathbb{1}_B$ is also a simple random variable.

Definition 1.17. *If $X : \Omega \to \mathbb{R}^d$ is a simple random variable then we can define its integral over $B \in \mathcal{F}$ as*

$$\int_B X(\omega) dP(\omega) := \int_\Omega X(\omega) \mathbb{1}_B(\omega) d\mathbb{P}(\omega).$$

From the definition we immediately see that

1. If X is a simple random variable then
$$|\mathbb{E}X| \leq \mathbb{E}(|X|).$$
2. If $\alpha, \beta \in \mathbb{R}$ then for any two simple random variables X and Y
$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}X + \beta \mathbb{E}Y$$
(and similarly for the notation with integrals). This is referred to as *linearity*.

Now we wish to define the (Lebesgue) integral for any random variable $X : \Omega \to \mathbb{R}$. We start by noting that we can split X into its positive and negative parts X^+ and X^- with $X^+ := X\mathbb{1}_{\{X\geq 0\}} \geq 0$ and $X^- := -X\mathbb{1}_{\{X<0\}} \geq 0$. Then $X = X^+ - X^-$. So it is enough to first define the integral for $X \geq 0$. This is done as follows:

Definition 1.18. *If $X : \Omega \to \mathbb{R}$ is non-negative, i.e. $X \geq 0$ then we define the expectation of X as*

$$\mathbb{E}X := \int_\Omega X(\omega) d\mathbb{P}(\omega) := \sup_{\xi \leq X \text{ and } \xi \text{ simple}} \int_\Omega \xi(\omega) d\mathbb{P}(\omega).$$

Note that it may happen that $\mathbb{E}X$ is infinite.

For a general $X : \Omega \to \mathbb{R}$ we now define

$$\mathbb{E}X := \mathbb{E}X^+ - \mathbb{E}X^-,$$

provided that either $\mathbb{E}X^+$ is finite (in which case $\mathbb{E}X$ is minus infinity) or that $\mathbb{E}X^-$ is finite (in which case $\mathbb{E}X$ is plus infinity). If both $\mathbb{E}X^+$ and $\mathbb{E}X^-$ are infinite then the expectation is not defined.

Definition 1.19. *We say* $X : \Omega \to \mathbb{R}$ *is* integrable *if* $\mathbb{E}(|X|) < \infty$ *and* square integrable *if* $\mathbb{E}(|X|^2) < \infty$.

If X is integrable then we say its *mean* is $\mathbb{E}X$.

If X is square integrable we say its *variance* is $\operatorname{Var}(X) := \mathbb{E}(X^2) - (EX)^2$. Exercise: show that $\operatorname{Var}(X) = \mathbb{E}\left((X - \mathbb{E}X)^2\right)$.

If X and Y are two square integrable random variables then their covariance is

$$\operatorname{Cov}(X, Y) = \mathbb{E}\left((X - \mathbb{E}X)(Y - \mathbb{E}Y)\right).$$

If X and Y are independent then $\mathbb{E}(XY) = 0$ and hence $\operatorname{Cov}(X, Y) = 0$. The converse is not true. Consider e.g. X normally distributed with mean 0 and variance 1 and $Y := X^2 - 1$. Clearly X and Y are not independent but $\mathbb{E}(XY) = 0$ and $(\mathbb{E}X)(\mathbb{E}Y) = 0$ and so $\operatorname{Cov}(X, Y) = 0$.

Lemma 1.20. *Let X be an integrable random variable. Let F_X denote the distribution of X. Let $g : \mathbb{R}^d \to \mathbb{R}$ be Borel measurable. Then*

$$\mathbb{E}\left(g(X)\right) = \int_{\mathbb{R}} g(x) dF_X(x).$$

Assume further that X is a continuous random variable with density f_X. Then

$$\mathbb{E}\left(g(X)\right) = \int_{\mathbb{R}} g(x) f_X(x) dx.$$

1.4 Conditional Expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be given.

Definition 1.21. *Let X be an integrable random variable. If $\mathcal{G} \subseteq \mathcal{F}$ is a σ-algebra then there exists a unique $\mathcal{G}$ measurable random variable Z such that*

$$\forall G \in \mathcal{G} \quad \int_G X d\mathbb{P} = \int_G Z d\mathbb{P}.$$

We say that Z is the conditional expectation *of X given $\mathcal{G}$ and write $\mathbb{E}(X|\mathcal{G}) := Z$.*

Of course it has to be proved that this new random variable exists, is unique and is $\mathcal{G}$ measurable for the definition to make sense. Here are some further important properties which we present without proof.

Theorem 1.22 (Properties of conditional expectations)**.** *Let X and Y be random variables. Let $\mathcal{G} \subseteq \mathcal{F}$.*

1. *For any $\alpha, \beta \in \mathbb{R}$*

   $$\mathbb{E}(\alpha X + \beta Y|\mathcal{G}) = \alpha \mathbb{E}(X|\mathcal{G}) + \beta \mathbb{E}(Y|\mathcal{G}).$$

   This is called linearity*.*

2. *Let $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$ be σ-algebras. Then*

   $$\mathbb{E}(X|\mathcal{G}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{G}_2)|\mathcal{G}_1).$$

   This is called the tower property*. A special case is $\mathbb{E}X = \mathbb{E}(\mathbb{E}(X|\mathcal{G}))$.*

3. *If* X *is* $\mathcal{G}$ *measurable then* $\mathbb{E}(X|\mathcal{G}) = X$*.*

4. *If* $\sigma(X)$ *is independent of* $\mathcal{G}$ *then* $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}X$*.*

5. *If* Y *is* $\mathcal{G}$ *measurable then* $\mathbb{E}(XY|\mathcal{G}) = YE(X|\mathcal{G})$*.*

Definition 1.23. *Let* X *and* Y *be two random variables. The* conditional expectation of X given Y *is defined as* $\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y))$*, that is, it is the conditional expectation of* X *given the* σ*-algebra generated by* Y*.*