Probability

20th May 2022

Contents

Chapter 1

Axioms of probability

The mathematical theory of probability is based on axioms much like Euclidean geometry, which has the mathematical objects of points and lines. In probability, we use **events** and **probabilities**. We use **set theory** to express these concepts.

1.1 Sets

Definition 1.1 (Set)**.** A **set** is an unordered collection of unique **elements**. If an element x belongs to the set S we write $x \in S$.

In this course, most of our sets are **finite**, meaning that we can express them as

$$S = \{x_1, x_2, \ldots, x_n\};$$

or countably infinite.

Definition 1.2 (Countable)**.** A set S is **countable** if either

(i) S is finite; or

(ii) there is a bijection (one-to-one mapping) between S and the natural numbers.

Example. $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ are countably infinite; however, $\mathbb{R}$ and $[0, 1]$ are not countable infinite.

Definition 1.3 (Empty set)**.** An important set is the **empty set**, we write

$$\varnothing = \{\}.$$

Definition 1.4 (Subset)**.** We write $A \subset B$ to say that A is a **subset** of B, or

$$\forall\, x \in A,\ x \in B.$$

Example.

$$\{1, 2, 3\} \subset \{1, 2, 3, 5, 7, 9\} \subset \mathbb{N} \subset \mathbb{Z}.$$

Remark. For any set A,

$$\varnothing \subset A.$$

Definition 1.5 (Power set)**.** The **power set** of the set A is the set of all subsets denoted as

$$2^A = \{B : B \subset A\}.$$

Example. Given $A = \{0, 1\}$, then

$$2^A = \{\varnothing, \{1\}, \{2\}, \{1, 2\}\}$$

1.2 Sample space and events

Consider a scenario with various different outcomes. We write Ω as the set of all possible outcomes. Ω is called the **sample space** and each $\omega \in \Omega$ is an **outcome**.

Example. Consider rolling a standard 6 sided dice. An obvious sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$; however, we could also have $\Omega = \{\text{odd}, \text{even}\}$ or $\{6, \text{not } 6\}$

Often, Ω is finite or countable infinite. In this case, we call it **discrete**.

Definition 1.6 (Events)**.** Associated to our sample space Ω is a collection $\mathcal{F}$ of all events:

$$A \subset \Omega\ \forall\, A \in \mathcal{F}.$$

We say that an event has occurred when the outcome at the end of the scenario is the set A.

Remark. $\varnothing$ is an impossible event, Ω is a certain event.

If Ω is discrete, we can always take $\mathcal{F} = 2^A$ so that every subset of Ω is an event; however, if Ω is not discrete we need to be more careful.

1.3 Event calculus

We may combine events using set theory operators.

Notation	Set theory language	Probability language	Meaning as events
$A \cup B$	A union B	A or B	A happens or B happens (or both)
$A \cap B$	A intersects B	A and B	A and B happens
A^c	A complement	not A	A does not happen
$A \setminus B$	A minus B	A but not B	A happens but B does not
$A \subset B$	A is a subset of B	A implies B	If A happens B must also happen

Table 1.1: Meaning of event notation in regards to set theory language, probability language, and events.

Definition 1.7. For an event $A \in \mathcal{F}$, define its **complement** A^c (not A)

$$A^c = \{\omega \in \Omega : \omega \notin A\}.$$

Remark. Notice that $(A^c)^c = A$, $A \cap A^c = \varnothing$, and $A \cup A^c = \Omega$. $\bar{A}$ is also valid notation for complement.

Let A, B be events. Table 1.1 shows the notation and meaning for events.

Proposition 1.8. Given events A, B, then

$$A \setminus B = A \cap B^c.$$

Proof.

$$\begin{aligned} A \setminus B &= \{\omega \in \Omega : \omega \in A, \omega \notin B\} \\ &= \{\omega \in \Omega : \omega \in A\} \cap \{\omega \in \Omega : \omega \notin B\} \\ &= A \cap B^c. \end{aligned}$$

□

Definition 1.9 (Disjoint)**.** Two sets A, B are **disjoint** if $A \cap B = \varnothing$.

Example. Consider $\Omega = \{1, 2, 3, 4, 5, 6\}$ and events

(i) $A = \{2, 4, 6\}$;

(ii) $B = \{1, 3, 5\}$; and

(iii) $C = \{1, 2, 3\}$.

Then $A \cup B = \Omega$, $A \cap B = \varnothing$, $A^c = B$, $C \setminus A = \{1, 3\}$, $A \cup C = \{1, 2, 3, 4, 6\}$.

Remark. For multiple events we use the notation

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n$$

and

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \ldots \cap A_n.$$

For infinite unions / intersections we can use the notation

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \ldots$$

and

$$\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \ldots.$$

Definition 1.10 (De Morgan's law)**.** Given sets A, B, then

$$(A \cap B)^c = A^c \cup B^c, \quad (A \cup B)^c = A^c \cap B^c.$$

1.4 Axioms of probability

Definition 1.11. A probability $\mathbb{P}$ on a sample space Ω with a collection of events $\mathcal{F}$ is a function mapping every event $A \in \mathcal{F}$ to a real number $\mathbb{P}(A)$ satisfying

A1 $\mathbb{P}(A) \geq 0$ for all A;

A2 $\mathbb{P}(A) = 1$;

A3 if A and B are disjoint (that is, $A \cap B = \varnothing$), then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B); \text{and}$$

A4 for an infinite sequence $A_1, A_2, \ldots$ of pairwise disjoint (such that all pairs are disjoint) events

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Example. Consider a finite sample space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ with size $|\Omega| = m$. We can define a valid probability $\mathbb{P}$ by taking any numbers $p_1, p_2, \ldots, p_m$

with $p_i \geq 0$ and $\sum_{i=1}^{m} p_i = 1$ and declare for any event $A \subset \Omega$

$$\mathbb{P}(A) = \sum_{i:w_i \in A} p_i.$$

Definition 1.12. A probability $\mathbb{P}$ on a sample space Ω with a collection of events $\mathcal{F}$ is a function mapping every event $A \in \mathcal{F}$ to a real number $\mathbb{P}(A)$ satisfying

A1 $\mathbb{P}(A) \geq 0$ for all A;

A2 $\mathbb{P}(A) = 1$;

A3 if A and B are disjoint (that is, $A \cap B = \varnothing$), then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B); \text{and}$$

A4 for an infinite sequence $A_1, A_2, \ldots$ of pairwise disjoint (such that all pairs are disjoint) events

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Example. Consider a finite sample space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ with size $|\Omega| = m$. We can define a valid probability $\mathbb{P}$ by taking any numbers $p_1, p_2, \ldots, p_m$ with $p_i \geq 0$ and $\sum_{i=1}^{m} p_i = 1$ and declare for any event $A \subset \Omega$

$$\mathbb{P}(A) = \sum_{i:w_i \in A} p_i.$$

Definition 1.13 (Partition)**.** Events $E_1, E_2, \ldots, E_k$ form a **partition** of a sample space Ω if

(i) $\mathbb{P}(E_i) > 0 \ \forall \ i$;

(ii) $E_i \cap E_j = \varnothing \ \forall \ i \neq j$; and

(iii) $\bigcup_{i=1}^{k} E_i = \Omega$.

Proposition 1.14 (Consequences of the axioms)**.** All of these follow from A1 - A4.

C1 $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}((A \cap B))$;

C2 $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;

C3 $\mathbb{P}(\varnothing) = 0$;

C4 $\mathbb{P}(A) \leq 1$;

C5 if $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$;

C6 $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B)$;

C7 if $A_1, A_2, \ldots, A_k$ are pairwise disjoint, then

$$\mathbb{P}\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \mathbb{P}(A_i);$$

C8 for any events $A_1, A_2, A_3, \ldots$

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i);$$

C9 if $A_1 \subset A_2 \subset \ldots$ is an increasing sequence of events then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty}\right) = \lim_{n \to \infty} \mathbb{P}(A_n); \text{ and}$$

C10 if $E_1, E_2, \ldots, E_k$ is a partition, then

$$\sum_{i=1}^{k} \mathbb{P}(E_i) = 1.$$

Example. Prove that for any events $A_1, A_2, \ldots$

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

given that for an infinite sequence $A_1, A_2, \ldots$ of pairwise disjoint (such that all pairs are disjoint) events

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

and that if $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

1.5 Sigma-algebras

Recall that we have a sample space Ω and a collection $\mathcal{F}$ of subsets of Ω called events. If Ω is discrete (that is, finite or infinitely countable), then we can take $\mathcal{F} = 2^\Omega$ such that all subsets of Ω are events. If Ω is uncountable, then it is too much to demand that $\mathbb{P}(A)$ is defined for all $A \in \Omega$.

Definition 1.15. A collection $\mathcal{F}$ is called a **σ-algebra** if

S1 $\Omega \in \mathcal{F}$;

S2 $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$; and

S3 if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{i=1}^\infty A_i \in \mathcal{F}$.

Example. The power set 2^Ω is the biggest σ-algebra over Ω. The set $\{\varnothing, \Omega\}$ is the trivial σ-algebra which is the smallest σ-algebra over Ω.

Example. Consider $\Omega = \{1, 2, 3, 4, 5, 6\}$. The choice of σ-algebra may depend on exactly what I am interested in. The following are all σ-algebras over Ω.

(i) $\mathcal{F}_0 = \{\varnothing, \Omega\}$;

(ii) $\mathcal{F}_1 = \{\varnothing, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$; and

(iii) $\mathcal{F}_2 = 2^\Omega$.

Note that $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2$.

Definition 1.16. If Ω is a set and the collection $\mathcal{F}$ of subsets of Ω is a σ-algebra, and $\mathbb{P}$ satisfies A1-A4 for all events in $\mathcal{F}$, then $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**.

In practise, we rarely make our σ-algebras explicit. They just sit there in the background.

Chapter 2

Equally likely outcomes and counting principles

2.1 Classical probability

One of the simplest scenarios we are faced with in probability is where we have A **finite sample space**

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$$

with m **equally likely** outcome, such that

$$\mathbb{P}(\omega_i) = \frac{1}{|\Omega|} = \frac{1}{m}.$$

For any event $A \subset \Omega$

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

Example. Roll a fair dice. Then

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

All outcomes are equally likely, so

$$\mathbb{P}(\text{odd score}) = \frac{|\{1, 3, 5\}|}{|\Omega|} = \frac{1}{2}.$$

2.2 Counting principles

Given a finite sample space and assuming that outcomes are equally likely, to determine probabilities of certain events comes down to counting. We have a number of principles that we can use to help make counting easier.

Proposition 2.1 (Multiplication principle)**.** Suppose we make k choices in succession with m_k representing the number of possibilities for the kth choice. Then the total number of possible selections is

$$\prod_{i=1}^{k} m_i.$$

Example. In a standard deck of cards, there are 13 denominations and 4 suits. So there are a total of 52 cards.

Proposition 2.2 (Ordered choices of distinct objects with replacement)**.** Select r from a collection of m distinct objects with replacement, that is, every object is available at each choice. Then the number of possible selections is

$$m^r.$$

Proposition 2.3 (Ordered choices of distinct objects without replacement)**.** Select $\leq n$ from a collection of n distinct objects without replacement. So no object can be used more than once. Then the total number of possible selections is

$$\frac{m!}{(m-r)!} = (m)_r.$$

Example. From a deck of cards, what is the probability of being dealt the king of diamonds, the queen of diamonds, then the jack of diamonds (in any order) in a deal of 3 cards.

Solution. The number different ordered hands of 3 cards is

$$52 \cdots 51 \cdot 50 = (52)_3.$$

The number of these hands that consist of the 3 cards we're looking for

$$3 \cdot 2 \cdot 1 = 6;$$

hence the probability is

$$\frac{6}{(52)_3} = 4.5 \cdot 10^{-5}.$$

$\square$

Example. Suppose there are $n < 365$ people in a room. Let B be the event that at least two people have the same birthday. What is $\mathbb{P}(B)$? How large does n have to be so that $\mathbb{P}(B) = \frac{1}{2}$.

Proposition 2.4 (Unordered choice without replacement)**.** Suppose that there is a collection of m distinct objects, we select $r \leq m$ of them, no object may be chosen more than once; and the order does not matter. The number of distinct selections of size r is

$$\binom{m}{r} = \frac{(m)_r}{r!} = \frac{m!}{r!(m-r)!}.$$

Example. There are 5 candidates for staff-student committee reps. 3 are to be chosen. How many ways are there to choose the reps?

Solution. There are

$$\binom{5}{3} = \frac{5!}{3!(2!)} = 10$$

ways to do this. □

Example. What is the change of having no aces in a four card hand?

Solution. The number of hands is $\binom{52}{4}$ and the number of hands with no ace is $\binom{48}{4}$. Therefore, the probability of having no aces in a four card hand is

$$\frac{\binom{48}{4}}{\binom{52}{4}} = \frac{(48!)^2}{(44!)(52!)}.$$

□

Remark. As a general rule of thumb for these kind of questions, use an unordered approach for questions involving decks of cards and an ordered approach for question involving dice.

Example. Suppose you are dealt 5 cards. The event A is that 4 out of the 5 cards have the same suit. What is $\mathbb{P}(A)$?

Solution. There are $\binom{52}{5}$ hands with all the same likeliness. We need to count the number of times there is 4 cards of the same suit. We describe a way of building the hand we are looking for:

(i) choose the suit (4 choices);

(ii) choose the denominations for the cards ($\binom{13}{4}$ choices); then

(iii) choose the last card (39 choices).

So

$$\mathbb{P}(A) = \frac{\left(\frac{4\cdot\binom{13}{4}\cdot 39}{(4!)(9!)}\right)}{\left(\binom{52}{5}\right)}.$$

□

Proposition 2.5 (Ordered choice from 2 types of object)**.** Consider that we have m objects, r of type 1 and $m-r$ of type 2, where objects are indistinguishable from others of their type. The number of **distinct**, **ordered** choices of the m objects is
$$\binom{m}{r}.$$

Proposition 2.6 (Ordered grouping of indistinguishable objects)**.** The number of ways to divide m indistinguishable objects into k distinct groups is
$$\binom{m+k-1}{m} = \binom{m+k-1}{k-1}$$
by a previous principle.

Proof. List all m objects in a line and insert $k-1$ fences to distinguish the group boundaries. We see that the number of ways to divide these numbers is the same as the number of objects and fences in the line, so
$$\binom{m+k-1}{m} = \binom{m+k-1}{k-1}$$
by a previous principle. □

Chapter 3

Random variables

3.1 Definitions

Definition 3.1 (Random variable)**.** A **random variable** X on a sample space Ω is a mapping $X : \Omega \to X(\Omega)$ given by $\omega \mapsto X(\omega)$.

$$X(\Omega) = \{X(\omega) : \omega \in \Omega\}.$$

We will typically find ourselves with either

(i) $X(\Omega) \subset \mathbb{R}$, univariate random variable; or

(ii) $X(\Omega) \subset \mathbb{R}^n$, multivariate random variable.

Example. Consider throwing two dice with sample space

$$\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}.$$

Let the random variable X be the sum of the numbers on the dice. Then

$$X(i, j) = i + j \ \forall \ (i, j) \in \Omega.$$

Also

$$(X = 10) = \{(4, 6), (5, 5), (6, 4)\}$$

and

$$(X \in \{0, 1, 2\}) = \{(1, 1)\}.$$

Definition 3.2 (Indicator function)**.** An **indicator function** is a common type of random variable that we may use, denoted $\mathbb{1}_A$ where $A \in \mathcal{F}$ is such that $\mathbb{1}_A : \Omega \to \{0, 1\}$ defined by

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \text{otherwise} \end{cases}.$$

Note that

$$\mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(A), \qquad \mathbb{P}(\mathbb{1}_A = 0) = \mathbb{P}(A') = 1 - \mathbb{P}(A).$$

Theorem 3.3. The function $\mathbb{P}_X : B \subset X(\Omega) \to \mathbb{R}$ defined by

$$\mathbb{P}_x(B) = \mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$$

is a probability on $X(\Omega)$.

3.2 Discrete random variables

Definition 3.4 (Discrete random variable)**.** A random variable $X : \Omega \to X(\Omega)$ is said to be **discrete** when there exists a finite or infinitely countable set $\mathcal{X} \subset X(\Omega)$ such that $\mathbb{P}(X \in \mathcal{X}) = 1$. The function $p : \mathcal{X} \to [0, 1]$ is defined by

$$p(x) = P(X = x) \quad \forall\, x \in \mathcal{X}$$

is called the **probability mass function**.

Theorem 3.5. Suppose X is a discrete random variable and $p : \mathcal{X} \to [0, 1]$ is the probability mass function. Then

$$\mathbb{P}(X \in B) = \sum_{x \in B} p(x) \quad \forall\, B \subset \mathcal{X}$$

and

$$\sum_{x \in \mathcal{X}} p(x) = 1.$$

The probability mass function summarises all information about the distribution of X.

Proof. This can be simply proved using the 4th axiom of probability at the start of this course and that

$$(X \in B) = \bigcup_{x \in B} \{X = x\}.$$

□

Theorem 3.6. A random variable $X : \Omega \to X(\Omega)$ is discrete whenever

(i) $X(\Omega)$ is finite or countable; or

(ii) Ω is finite or countable.

Example. Toss three coins and let X be the total number of heads. We can tabulate this as follows.

ω	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X(\omega)$	3	2	2	2	1	1	1	0

This is a discrete random variable as $\Omega is finite$. Now, we can find the probability mass function from the tabulation.

x	0	1	2	3
$p(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

We will shortly recognise this as being binomially distributed.

3.3 The binomial and geometric distributions

We call a scenario **binomial** if it meets the following criteria:

(i) consists of a fixed number of trials;

(ii) trials are independent;

(iii) each trial has two outcomes, *success* or *failure*; and

(iv) each *success* has a fixed probability.

We can model binomial scenarios using the binomial distribution.

Definition 3.7 (Binomial distribution)**.** We say that a discrete random variable X is **binomially distributed** with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$, denoted by $X \sim \text{Bin}(n, p)$, when $\mathcal{X} = \{0, 1, 2, \ldots, n\}$ and

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \forall\, x \in \{0, 1, 2, \ldots, n\}.$$

If $n = 1$ we call this a **Bernoulli trial** and we call $X \sim \text{Bin}(1, p)$ is known as a **Bernoulli random variable**.

Example. If we roll 4 fair dice and let X be the number of 6s, then $X \sim$

Bin $\left(4, \frac{1}{6}\right)$.

$$\mathbb{P}(X = x) = \binom{4}{x}\left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{4-x}.$$

Suppose we extend the binomial distribution to an infinite number of trials, this leads us to the following definition.

Definition 3.8 (Geometric distribution)**.** We say that a discrete random variable X is **geometrically distributed** with parameter $p \in (0, 1]$ and we write $X \sum \text{Geo}(p)$ when $\mathcal{X} = \mathbb{N}$ and

$$p(x) = (1 - p)^{x-1} p \quad \forall\, x \in \{1, 2, 3, \ldots\}.$$

Definition 3.9 (Poisson distribution)**.** We sya that a discrete random variable X is **Poisson distributed** with parameter $\lambda > 0$, denoted $X \in \text{Po}(\lambda)$, when $\mathcal{X} = \mathbb{Z}_+$ and

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \forall\, x \in \mathbb{Z}_+.$$

We use the Poisson distribution to model events with a constant average rate per unit time. That is,

$$\mathbb{P}(\text{ event in } (x, x+h)) \approx rh$$

where r is the average rate per unit time and for small h. Note that we can use the Poisson distribution over any median, such as the surface of an object, than time.

Example. Suppose in a particular fleet of aircraft there has been 32 crashes in 25 years. Let W, M, Y denote the number of crashed in the next week, month, and year respectively. Assume that a year has 365 days and a month has 30 days. Also suppose that crashes happen at random such that we can model it using a Poisson distribution.

(i) How are W, M, Y distributed?

(ii) Find $\mathbb{P}(\text{ no crashes next week })$.

(iii) Find $\mathbb{P}(\text{ no crashes next month})$.

(iv) Find $\mathbb{P}(\text{ no crashes next year})$.

Solution. (i) $W \sim \text{Po}\left(\frac{32}{25} \cdot \frac{1}{365} \cdot 7\right)$, $M \sim \text{Po}\left(\frac{32}{25} \cdot \frac{1}{365} \cdot 30\right)$, $Y \sim \text{Po}\left(\frac{32}{25}\right)$.

(ii) $\mathbb{P}(W = 0) = 0.976$.

(iii) $\mathbb{P}(M = 0) = 0.900$.

(iv) $\mathbb{P}(Y = 0) = 0.278$.

$\square$

Theorem 3.10 (Binomial → Poisson)**.** Consider $\lambda > 0$. Let $X_n \sim \mathrm{Bin}(n, p_n)$ such that

$$\lim_{n\to\infty} np_n = \lambda.$$

Let $Y \in \mathrm{Po}(\lambda)$. Then for all $x \in \mathbb{Z}_+$

$$\lim_{n\to\infty} p_{X_n}(x) = p_Y(x).$$

We describe this by saying that X_n **converges in distribution** to Y.

Proof. This is a simple look at the probability mass function for the binomial distribution and taking the limit as $n \to \infty$ to obtain the probability mass function for the Poisson distribution. $\square$

Example. Let $X \in \mathrm{Bin}(1000, 0.001)$. As our n is large,

$$X \sim \mathrm{Po}(1000 \cdot 0.001) = \mathrm{Po}(1).$$

Hence

$$\mathbb{P}(X \leq 2) \approx \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = \frac{5}{2e} = 0.920.$$

3.4 Continuous random variable

Definition 3.11 (Continuous random variable and probability density function)**.** Consider a real-valued random variable $X : \Omega \to \mathbb{R}$. We say that X is a **continuous random variable**, or that X has a continuous probability distribution, or that X is continuously distributed, when there is a non-negative function $f : \mathbb{R} \to \mathbb{R}$ such that

$$\mathbb{P}(X \in [a, b]) = \int_a^b f(t)\, dt,$$

for all $x \in [a, b] \subset \mathbb{R}$. f is called the **probability density function** of X.

Remark. Note that not all distributions have a unique probability density function, for example the probability density functions

$$f_1(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}, \qquad f_2(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

give the same distribution.

Theorem 3.12. Let X be continuously distributed with probability density function f, then foir $B \subset R$ that is a finite union of intervals,

$$\mathbb{P}(X \in B) = \int_B f(t)\,dt.$$

Like the probability mass function, the probability density function contains almost all information we have about a continuous random variable.

Example. Let X be a continuous random variable with probability density function $f(x) = kx^2$ for $0 < x < 1$. Find k and calculate $\mathbb{P}(x \in \left[0, \frac{1}{3}\right])$.

Solution.

$$\int_0^1 kx^2\,dx = 1 \Leftrightarrow \left[\frac{1}{3}kx^3\right]_0^1 = 1 \Leftrightarrow k = 3.$$

$$\mathbb{P}\left(X \in \left[0, \frac{1}{3}\right]\right) = \int_0^{\frac{1}{3}} 3x^2\,dx = \left[x^3\right]_0^{\frac{1}{3}} = \frac{1}{27}.$$

□

3.5 The uniform distribution

Definition 3.13 (Uniform distribution)**.** Let $a, b \in \mathbb{R}$ with $a < b$. We say a continuous random variable X is **uniformly distributed** on $[a, b]$, denoted $X \sim \mathrm{U}(a, b)$, when

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}.$$

Remark. Note that for $x, x+h \in [a, b]$, $\mathbb{P}(X \in [x, x+h])$ is constant regardless of x for $X \sim \mathrm{U}(a, b)$.

Example. Let $X \sim \mathrm{U}(0, 3)$. What is $\mathbb{P}(X \leq 1)$.

Solution.

$$\mathbb{P}(X \leq 1) = \int_0^1 \frac{1}{3-0}\,dx = \left[\frac{1}{3}x\right]_0^1 = \frac{1}{3}.$$

□

3.6 The exponential distribution

Definition 3.14 (Exponential distribution)**.** Let $\beta > 0$. We say a continuous random variable is **exponentially distributed** with parameter β, denoted $X \sim \text{Exp}(\beta)$, when

$$f(x) = \begin{cases} \beta e^{-\beta x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

Example. Let $X \sim \text{Exp}(\ln k)$, $k > 1$. Prove that

$$(X < k) = 1 - k^{-k}.$$

Solution. Proof follows from integrating the probability density function for the exponential distribution. □

3.7 The normal distribution

Definition 3.15 (Normal distribution)**.** Let $\mu, \sigma \in \mathbb{R}$ with $\sigma > 0$. We say a continuous random variable X is **normally distributed** with parameters μ, σ^2, denoted $X \sim \mathcal{N}(\mu, \sigma^2)$, when

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \forall\, x \in \mathbb{R}.$$

3.8 Cumulative distribution functions

Definition 3.16 (Cumulative distribution function)**.** For any real-valued random variable X, the function $F : \mathbb{R} \to [0, 1]$ defined by

$$F(x) = \mathbb{P}(X \leq x) \quad \forall\, x \in \mathbb{R}$$

is called the **cumulative distribution function**.

Theorem 3.17 (Relation between cumulative distribution function and probability density function)**.** Suppose that X is a continuous distributed random variable on $\mathbb{R}$ with probability density function f. Then F is continuous and for all $x \in \mathbb{R}$,

$$F(x) = \int_{-\infty}^{x} f(t)\, dt, \qquad f(x) = \frac{dF}{dx}(x)$$

when f is continuous at x.

Theorem 3.18. If X is a discrete real valued random variable with probability mass function p, then F is piecewise constant and

$$F(x) = \sum_{t:t\leq x} p(t)$$

and $p(x) = F(x) - F(x^-)$ where x^- is the limit from the left.

Theorem 3.19 (Proprieties of cumulative distribution functions)**.** Let F be a cumulative distribution function of a real valued random variable, then

(i) $\lim_{t\to-\infty} F(t) = 0$ and $\lim_{t\to\infty} F(t) = 1$;

(ii) monotonicity, $s \leq t \Rightarrow F(s) \leq F(t)$; and

(iii) right continuity, for all $t \in \mathbb{R}$, $F(t) = F(t^+)$ where t^+ is the limit from the right.

Theorem 3.20. The cumulative distribution function of a random variable completely determines its distribution.

3.9 Standard normal tables

Definition 3.21 (Standard normal distribution)**.** A continuous random variable Z is **standard normally distributed** if $Z \sim \mathcal{N}(0,1)$. We normally use Z for this random variable, ϕ for the probability density function of Z, and Φ for the cumulative density function of Z.

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}z^2}, \qquad \Phi(z) = \int_{-\infty}^{z} \phi(t)\, dt$$

Remark. $\Phi(z)$ has no analytical solution, we typically tabulate it.

Example. The following are the standard normal tables needed for this question.

z	0	1.28	1.64	1.96	2.58
$\Phi(z)$	0.5	0.9	0.95	0.975	0.995

Suppose $Z \sim \mathcal{N}(0,1)$. Calculate $\mathbb{P}(-1.28 \leq z \leq 1.64)$.

Solution.

$$\begin{aligned}\mathbb{P}(-1.28 \leq z \leq 1.64) &= \mathbb{P}(z \leq 1.64) - \mathbb{P}(z \leq -1.28) \\ &= \mathbb{P}(z \leq 1.64) + \mathbb{P}(z \leq 1.28) - 1 \\ &= \Phi(1.64) + \Phi(1.28) - 1 = 0.85.\end{aligned}$$

□

3.10 Functions of random variables

Theorem 3.22. Suppose $X : \Omega \to X(\Omega)$ is a random variable and $g : X(\Omega) \to S$is some function. Then $g(X)$ is also a random variable, that is

$$P(g(x) \in B) = \mathbb{P}(\{\omega \in \Omega : g(X(\omega)) \in B\}).$$

Example. Let $X \sim \text{Bin}(n, p)$. Show that $n - X \sim \text{Bin}(n, 1-p)$.

Solution.

$$\begin{aligned}\mathbb{P}(Y = x) &= \mathbb{P}(n - X = x) \\ &= \mathbb{P}(X = n - x) \\ &= \binom{n}{n-x} p^{n-x}(1-p)^x \\ &= \binom{n}{x}(1-p)^x p^{n-x};\end{aligned}$$

hence $n - X \sim \text{Bin}(n, 1-p)$. □

Theorem 3.23 (Standardising the normal distribution)**.** Suppose $\mu, \sigma \in \mathbb{R}$ with $\sigma > 0$. If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Z \sim \mathcal{N}(0, 1)$, then

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

and

$$\sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2).$$

Corollary 3.24. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Example. If $X \sim \mathcal{N}(2,4)$, then

$$\begin{aligned}\mathbb{P}(X \geq 5.28) &= \mathcal{P}\left(z \geq \frac{5.28-2}{2}\right)\\ &= 1-\mathbb{P}\left(z \leq \frac{5.28-2}{2}\right)\\ &= 1-\mathbb{P}(z \leq 1.64)\\ &= 1-\Phi(1.64) = 0.05.\end{aligned}$$