Lecture 2

Overview of probability theory and statistics

V.0.2, updated on 21.11.2023

CONTENTS

I. BASIC CONCEPTS OF PROBABILITY AND STATISTICS

A. Probability space

A probability space $(\Omega, \mathcal{F}, P)$ is a mathematical object that represents an experiment with random outcomes (e.g. the tossing of a fair coin). A probability space is made of three parts:

- The sample space Ω is the set of all the possible outcomes. The sample space can be discrete or continuous.

- The σ-algebra $\mathcal{F}$ is a collection of subsets of Ω with the following properties:

  1. $\emptyset \in \mathcal{F}$
  2. If $A \subset \Omega$ is a subset of Ω and $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$, where $A^C = \Omega - A$ is the complement of A
  3. If $A_i \in \mathcal{F}\ \forall i \Rightarrow A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

- The probability measure (or probability distribution) P returns the event's probability. It is a function $P : \mathcal{F} \to [0, 1]$, with the following properties:

  1. $P(\emptyset) = 0,\ P(\Omega) = 1.$
  2. If $\{A_i\}_{i=1}^{\infty}$ is a collection of disjoint sets then

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

A filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ is a probability space equipped with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, P)$, i.e. a collection of σ-algebras $\mathcal{F}_t \subset \mathcal{F}$ such that

$$\text{if } 0 \leq s < t \Rightarrow \mathcal{F}_s \subset \mathcal{F}_t\,. \tag{2}$$

Example. Consider a one coin toss experiment.

The sample space is $\Omega = \{\text{Head}, \text{Tail}\}$.

The power set 2^Ω, i.e. the set of all possible subsets of Ω, is the largest σ-algebra:

$2^\Omega = \{\emptyset, \{\text{Head}\}, \{\text{Tail}\}, \{\text{Head}, \text{Tail}\}\}$.

The notation 2^Ω is used because it contains 2^N elements, where N is the number of elements of Ω. Indeed, for each subset of Ω, each element of Ω can be present or absent.

The outcome of the probability measure applied to the power set 2^Ω is $P(2^\Omega) = \{0, 0.5, 0.5, 1\}$.

B. Probability axioms

There exist several ways to define probability. The classical definition is derived from experience. For example, if we consider a coin toss, we can define the probability of heads (or tails) as the ratio of the outcomes of heads (or tails) divided by a large number of tosses. Then, the concept of probability can be defined as the ratio of the number of favorable cases and the number of possible cases. This definition, however, is not exempt from problems of rigorousness.

There are several alternative definitions. Here we recall the axiomatic definition, also known as Kolmogorov's axioms, which defines probability as a measure of how likely an event is to occur. Given a subset $A \subset \Omega$, the probability measure (or probability distribution) $P(A)$ satisfies the following axioms:

1. $P(A) \geq 0\,,\ \forall A$
2. $P(\Omega) = 1$ and $P(\emptyset) = 0$
3. Given a collection of non-overlapping subsets $A_i \subset \Omega$ with $i = 1, 2, \ldots$, i.e. such that $A_i \cap A_j = \emptyset\,, \forall i \neq j$, then

$$P(\cup_i A_i) = \sum_i P(A_i)\,,$$

   and

$$P(A^c) = 1 - P(A)\,.$$

From this definitions, it follows that the probability is a measure such that $P : \mathcal{F} \rightarrow [0, 1]$

The probability axioms are completed by the following definitions.

a. *Joint probability* Given A and B subsets of Ω, then the joint probability is defined as

$$P(A \cap B) = \{(\omega \in A) \text{ and } (\omega \in B)\}\ . \tag{3}$$

b. *Conditional probability* Given A and B subsets of Ω, then the conditional probability is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}\,. \tag{4}$$

c. *Probability of independent events* Let A and B be independent subsets of Ω, i.e. the elements of A do not influence the occurrence of the events of B and vice versa, then the joint probability is

$$P(A \cap B) = P(A)P(B)\,. \tag{5}$$

d. Bayes' theorem Given A and B subsets of Ω, and the conditional probabilities

$$P(A|B) = \frac{P(A \cap B)}{P(B)}\,, \tag{6}$$

and

$$P(B|A) = \frac{P(A \cap B)}{P(A)}\,, \tag{7}$$

follows that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}\,. \tag{8}$$

C. Random variable

Given a probability space $(\Omega, \mathcal{F}, P)$, a real-valued random variable $X\colon \Omega \to \mathbb{R}^n$ is a measurable function that maps the elements of the sample space Ω to some real number. If Ω is discrete, then X is a discrete random variable, if Ω is continuous, then X is a continuous random variable. Given a subset $A \in \Omega$, then the probability that X returns an element of the subset is:

$$P(A) = \mathrm{Prob}(X \in A) = \mathrm{Prob}\left\{X = x \text{ with } x \in A\right\}\,. \tag{9}$$

The probability measure P can induce a probability density function $f_P : \Omega \to \mathbb{R}^n$ such that

$$f_P(x)dx = \mathrm{Prob}\left\{X = x \text{ with } x \in [x, x + dx]\right\}\,, \tag{10}$$

where dx is an infinitesimal change in x, which satisfies

$$\int_\Omega f_P(x)\,dx = 1\,. \tag{11}$$

If f_P exists, it is the Radon-Nikodym derivative of the measure P with respect to the Lebesgue measure λ:

$$f_P = \frac{dP}{d\lambda}\,, \tag{12}$$

and eq. 9 is written as

$$P(A) = \int_A f_P(x)\,dx\,. \tag{13}$$

Remarks. The capital letter X denotes the random variable, and the small letter x denotes the value that is assumed by the variable.

D. Moments

Moments are measurable quantities that describe properties of probability distributions: location, shape and scale. The moments are defined as

$$\langle X^k \rangle = \int_\Omega x^k f_P(x)\, dx\,. \tag{14}$$

Note that the notation $\langle \cdot \rangle$ is equivalent to $\mathbb{E}\left[\cdot\right]$.

The first 3 moments are of particular interest:

- 0th moment

  $$\langle X^0 \rangle = \int_\Omega f_P(x)\, dx = 1$$

  is the total mass. It captures the fact that the distribution is normalized.

- 1st moment

  $$\langle X \rangle = \int_\Omega x f_P(x)\, dx = \mu$$

  is the mean, (or expected value, expectation, ensemble average, ...). It is the equilibrium point of the distribution, it describes how far away from the origin is the center of mass.

- 2nd moment

  $$\langle (X-\mu)^2 \rangle = \int_\Omega (x-\mu) f_P(x)\, dx = \sigma^2$$

  is the variance. It describes how spread out a distribution is.

Given an arbitrary function $g : \mathbb{R}^n \to \mathbb{R}$, typically called observable function or measurable function, the moments of the function are written as

$$\langle g(X)^k \rangle = \int_{\mathbb{R}^n} g(x)^k\, f_P(x) dx\,. \tag{15}$$

Example. Let's consider a one coin toss experiment.

The random variable X can assume the values:

$$X(\omega) = \begin{cases} 1, & \text{if } \ \omega = \text{Head} \\ \\ 0, & \text{if } \ \omega = \text{Tail} \end{cases} \tag{16}$$

The related probability density function is:

$$f_P(x) = \begin{cases} \frac{1}{2}, & \text{if } x = 1 \\ \\ \frac{1}{2}, & \text{if } x = 0 \end{cases} . \tag{17}$$

The expected value is:

$$\mathbb{E}[X] = 0 \cdot \tfrac{1}{2} + 1 \cdot \tfrac{1}{2} = \tfrac{1}{2} \,. \tag{18}$$

E. Law of large numbers

Given a sequence of N real-valued random variables X_i, then

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i \,, \tag{19}$$

and

$$\lim_{N \to +\infty} \bar{X}_N = \langle X \rangle \,. \tag{20}$$

Likewise, the law of large numbers can be used to estimate the k-moments:

$$\bar{X}_N^k = \frac{1}{N} \sum_{i=1}^{N} X_i^k \,. \tag{21}$$

II. RANDOM WALK

Consider a one-dimensional lattice of equally spaced points. Assume that a random walker is at position $x_0 = 0$ at time 0, and that he either jumps to the right with probability p or to the left with probability $1 - p$. After N timesteps he arrives at position m, where $m \in \mathbb{Z}$ denotes the node of the grid that he reached: $m > 0$ if the walker reached a site on the right of the starting position; $m < 0$ if the walker reached a site on the left of the starting position.

The probability to reach the position m after N timesteps is given by the binomial distribution:

$$P_d(m, N) = \frac{N!}{\frac{N+m}{2}! \cdot \frac{N-m}{2}!} \, (p)^{\frac{N+m}{2}} \, (1-p)^{\frac{N-m}{2}} \quad . \tag{22}$$

The notation P_d denotes that the distribution is discrete.

If $p = 1/2$, then the Binomial distribution is written as

$$P_d(m, N) = \begin{cases} 0 & \text{if } N \text{ even and } m \text{ odd} \\ 0 & \text{if } N \text{ odd and } m \text{ even} \\ \frac{N!}{\frac{N+m}{2}! \cdot \frac{N-m}{2}!} \left(\frac{1}{2}\right)^N & \text{else} \end{cases} \tag{23}$$

As $N \to \infty$, P_d converges to the continuous distribution

$$P_c(m, N) = \sqrt{\frac{2}{N\pi}} \exp\left(-\frac{m^2}{2N}\right) . \tag{24}$$