

# Supervised Learning

## Supervised Learning Tools Refreshing Mathematical Tools



### Learning goals

- Refresher on the basics of probability theory

### Learning goals

- Refresher on the basics of probability theory

## PROBABILITY SPACE / 2

Examples:

- *Coin Tossing* — Possible outcomes are  $\Omega = \{H, T\}$  with **H** resp. **T** representing "heads" resp. "tails". The allowable events are contained in  $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$ . If the coin is fair, then  $P(H) = P(T) = 1/2$ .
- *Dice Rolling* — Possible outcomes are  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . The allowable events are contained in  $2^\Omega$ , i.e., the power set of  $\Omega$ . If the dice is fair, then  $P(i) = 1/6$  for  $i = 1, \dots, 6$ .

**Further properties.** For any probability space  $(\Omega, \mathcal{F}, P)$  the following properties hold

- *Monotonicity* —  $A, B \in \mathcal{F}$ , and  $A \subset B$ , then  $P(A) \leq P(B)$ .
- *Union bound* — For any finite or countably infinite sequence events  $E_1, E_2, \dots$ ,

$$P(\cup_{i \geq 1} E_i) \leq \sum_{i \geq 1} P(E_i).$$



## INDEPENDENCE / 2

**Conditional probability.** The *conditional probability* that event  $A \in \mathcal{F}$  occurs given that event  $B \in \mathcal{F}$  occurs is  $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ . The conditional probability is well-defined only if  $\mathbb{P}(B) > 0$ .

**Bayes rule.** For two events  $A, B \in \mathcal{F}$  it holds that

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

**The law of total probability.** Let  $E_1, \dots, E_n \in \mathcal{F}$  be mutually disjoint events, such that  $\cup_{i=1}^n E_i = \Omega$ , then  $\forall A \in \mathcal{F}$ ,

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A \cap E_i) = \sum_{i=1}^n \mathbb{P}(A|E_i)\mathbb{P}(E_i).$$



## RANDOM VARIABLES / 2

- Functions of random variables are again random variables, i.e., if  $f : \mathbb{R} \rightarrow \mathbb{R}$  is some (measurable) function, then  $Z = f(X)$  is also a random variable.
- Identically distributed — Two random variables  $X$  and  $Y$  are identically distributed if their probability distributions coincide, i.e.,  $P_X = P_Y$ .

One distinguishes between two types of random variables:

- A *discrete random variable* is a random variable that can take only a finite or countably infinite number of values. Its probability distribution is determined by the *probability mass function* which assigns a probability to each value in the image of  $X$ .
- A *continuous random variable* is a random variable which can take uncountably infinite number of values. Usually its probability distribution is determined by a *density function*, which assigns probabilities to intervals of the image of  $X$ .



# DISCRETE RANDOM VARIABLES

If the image  $\Omega_X$  of  $X$  is discrete (e.g., finite or countably infinite), then  $X$  is called a *discrete RV*.



For a discrete RV  $X$ , the function

$$p: \Omega_X \rightarrow [0, 1], x \mapsto \mathbb{P}(X \in \{x\}) = \mathbb{P}(X = x)$$

is called a probability function or probability mass function of  $X$ .

Obviously,  $p(x) \geq 0$  and  $\sum_{x \in \Omega_X} p(x) = 1$ .

Examples:

- *Bernoulli distribution*: For a binary RV with  $\Omega_X = \{0, 1\}$ ,  $X \sim \text{Ber}(\theta)$  if  $p(1) = \theta$  and  $p(0) = 1 - \theta$ .
- *Binomial distribution*:  $X \sim \text{Bin}(n, \theta)$  if

$$p(k) = \begin{cases} \binom{n}{k} \theta^k (1 - \theta)^{n-k} & \text{if } k \in \{0, \dots, n\} \\ 0 & \text{otherwise} \end{cases} .$$

## EXPECTED VALUE/EXPECTATION

Expectation is the most basic characteristic of a random variable. Let  $X$  be a random variable, then the expectation of  $X$ , denoted by  $E(X)$ , is

$$E(X) = \int x dF(x) = \begin{cases} \sum_{x \in \Omega_X} x p(x) & \text{if } X \text{ is discrete} \\ \int_{\Omega_X} x p(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

provided the sum resp. the integral is well-defined and exists.

Some important properties of the expected value are:

- Linearity — For any constants  $c_1, c_2 \in \mathbb{R}$  and any pair of random variables  $X$  and  $Y$  it holds that
$$E(c_1 X + c_2 Y) = c_1 E(X) + c_2 E(Y).$$
- Transformations — If  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a (measurable) function, then the expectation of  $f(X)$  is

$$E(f(X)) = \int f(x) dF(x) = \begin{cases} \sum_{x \in \Omega_X} f(x) p(x) & \text{if } X \text{ is discrete} \\ \int_{\Omega_X} f(x) p(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

(provided the sum resp. integral exists.)





# MULTIVARIATE RANDOM VARIABLES

RVs  $X_1, \dots, X_n$  over the same probability space can be combined into a *random vector*  $\mathbf{X} = (X_1, \dots, X_n)$ .

Their joint distribution is specified by the joint mass/density function  $p_{\mathbf{X}}$ , such that for any measurable set  $A$  it holds that

$$\mathbb{P}(\mathbf{X} \in A) = \begin{cases} \sum_{(x_1, \dots, x_n) \in A} p_{\mathbf{X}}(x_1, \dots, x_n) & \text{if } X_1, \dots, X_n \text{ are discrete} \\ \int_A p_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n & \text{if } X_1, \dots, X_n \text{ are continuous} \end{cases}$$

The marginal distribution  $p_{X_1}$  of  $X_1$  is given by

$$p_{X_1}(x_1) = \begin{cases} \sum_{(x_2, \dots, x_n) \in \Omega_{X_2} \times \dots \times \Omega_{X_n}} p_{\mathbf{X}}(x_1, \dots, x_n) & \text{if discrete} \\ \int_{\Omega_{X_2} \times \dots \times \Omega_{X_n}} p_{\mathbf{X}}(x_1, \dots, x_n) dx_2 \dots dx_n & \text{if continuous} \end{cases}$$

In the same way, the marginal distributions of  $X_2, \dots, X_n$  are defined. The same type of projection (summation/integration over all remaining variables) is used to define marginal distributions on subsets of variables  $(X_{i_1}, \dots, X_{i_k})$  with  $\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$ .





## INDEPENDENCE OF RANDOM VARIABLES / 2

Some important properties and concepts with respect to independence are:

- The iid assumption — Random variables  $X_1, \dots, X_n$  are called *independent and identically distributed* (iid) iff they are mutually independent and each random variable has the same probability distribution as the others.
- Independence under transformations — Let  $X$  and  $Y$  be independent random variables and  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  are (measurable) functions. Then,  $f(X)$  and  $g(Y)$  are independent as well.



## CONDITIONAL DISTRIBUTIONS

If  $(X, Y)$  have a joint distribution with mass function  $p_{X,Y}$ , then the *conditional probability mass function* for  $X$  given  $Y$  is defined by

$$p_{X|Y}(x | y) = \frac{p(X = x, Y = y)}{p(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

provided  $p(Y = y) > 0$ .



Likewise, in the continuous case, the *conditional probability density function* is given by

$$p_{X|Y}(x | y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

provided  $p_Y(y) > 0$ . Then,

$$p(X \in A | Y = y) = \int_A p_{X|Y}(x | y) dx.$$

The soundness of this definition is less obvious than in the discrete case, due to conditioning on an event of probability 0 here.

# CONDITIONAL EXPECTED VALUE/EXPECTATION

1/2  
 Some important properties of the conditional expected value are the following. For any random variables  $X, Y, Z$  it holds that



- Linearity — For any constants  $c_1, c_2 \in \mathbb{R}$  it holds that
  - Linearity — For any constants  $c_1, c_2 \in \mathbb{R}$  it holds that
- Independence — If  $X$  and  $Y$  are independent random variables, then  $E(X|Y) = E(X)$ .
- Independence — If  $X$  and  $Y$  are independent random variables, then  $E(X|Y) = E(X)$ .
- Transformations — If  $f: \mathbb{R} \rightarrow \mathbb{R}$  is a (measurable) function, then the conditional expectation of  $f(X)$  given  $Y = y$  is

$$E(f(X)|Y = y) = \begin{cases} \sum_{x \in \Omega_X} f(x) p_{X|Y}(x|y) & \text{discrete case} \\ \int_{\Omega_X} f(x) p_{X|Y}(x|y) dx & \text{continuous case} \end{cases}$$

- Law of total expectation —  $E(E(X|Y)) = E(X)$ .
- Law of total expectation —  $E(E(X|Y)) = E(X)$ .
- Tower property —  $E(E(X|Y, Z)|Y) = E(X|Y)$ .
- Tower property —  $E(E(X|Y, Z)|Y) = E(X|Y)$ .