

# I. Background on Probability

A prob. space is a triple  $(\Omega, \mathcal{F}, P)$  where

①  $\Omega$  is a set called the sample space

②  $\mathcal{F}$  is a collection of subsets of  $\Omega$  satisfying

$$(i) \Omega \in \mathcal{F}$$

"Questions we want to ask  
in our experiment"

$$(ii) \text{ If } A \in \mathcal{F} \text{ then } A^c \in \mathcal{F}$$

$$(iii) \text{ If } A_i \in \mathcal{F} \text{ then } \bigcup_i A_i \in \mathcal{F}$$

③  $P: \mathcal{F} \rightarrow [0, 1]$

$$(i) 0 \leq P(A) \leq 1 \quad \forall A \in \mathcal{F}$$

$$(ii) P(\Omega) = 1$$

(iii) If  $A_1, A_2, \dots$  is a sequence of mutually exclusive events then

Mutually Exclusive

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad A_i \cap A_j = \emptyset, \text{ if } i \neq j$$

## Specific Cases

### ① Discrete probability spaces

► When  $\Omega$  is countable (in particular, finite)

what is the  $\sigma$ -algebra?

$\mathcal{F} = 2^\Omega$  is the power set of  $\Omega$ , i.e. the set of all subsets of  $\Omega$

How to build  $P$ ?

We are going to use a probability mass function (PMF)

$$p: \Omega \rightarrow [0, 1]$$

s.t.

$$\sum_{\omega \in \Omega} p(\omega) = 1$$

Then the probability measure induced by the pmf  $p$  is:

$$P_p(A) := \sum_{\omega \in A} p(\omega) \quad A \in \mathcal{F}$$

► For discrete  $\Omega$ , and given a PMF  $p$ , we will always think of the probability space  $(\Omega, 2^\Omega, P_p)$

## ② Continuous Probability Space

When  $\Omega = \mathbb{R}^d$ , how to build  $\mathcal{F}$ ?

- Optimally, we would like to take  $\mathcal{F} = 2^{\mathbb{R}^d}$

! However, it is impossible to define any "natural" probability measure with  $\mathcal{F} = 2^{\mathbb{R}^d}$

$\mu(A) \neq \mu(x+A)$        $A \subseteq \mathbb{R}$   
 $\uparrow$  sets                           $x \in \mathbb{R}$   
 violates translation invariant  
 measure

Thus, we have to pick a smaller  $\mathcal{F}$ , that "cleverly" constructed to include all interesting subsets of  $\mathbb{R}^d$ , but exclude pathological ones.

### Borel $\sigma$ -Algebra

- When  $d=1$ ,  $\Omega = \mathbb{R}$ .

- Given a subset  $A \subseteq \Omega$ , the  $\sigma$ -Algebra generated by  $A$ , denoted by  $\sigma(A)$  is

$$\sigma(A) = \{ \emptyset, \Omega, A, A^c \} \quad \begin{matrix} \text{smallest } \sigma\text{-Algebra} \\ \text{generated by } A \end{matrix}$$

- For  $A, B \subseteq \Omega$ ,

$$\sigma(A, B) = \{\emptyset, \Omega, A, B, A^c, B^c, A \cap B, A \cap B^c, A^c \cap B, A^c \cap B^c, A \cap B \cap A^c, A \cap B \cap B^c, \dots\}$$

- Generally, given a collection  $\{A_i\}_{i \in I}$  of subsets  $A_i \subseteq \Omega$ ,

$\sigma(\{A_i\}_{i \in I})$  is the smallest  $\sigma$ -algebra that contains  $\{A_i\}_{i \in I}$ .

To generate  $\sigma(\{A_i\}_{i \in I})$  we collect all the (countable) unions / intersections / complement of the generating set  $\{A_i\}_{i \in I}$

For the Borel  $\sigma$ -algebra on  $\mathbb{R}$ , denoted by  $\mathcal{B}(\mathbb{R})$  this is generated by all half-infinite intervals of  $\mathbb{R}$ .

$$\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, a] \mid a \in \mathbb{R}\})$$

In general, for  $d \geq 1$

$$\mathcal{B}(\mathbb{R}^d) = \{(-\infty, a_1] \times (-\infty, a_2] \times \dots \times (-\infty, a_d) \mid a_1, \dots, a_d \in \mathbb{R}\}$$

In this class

Whenever  $\Omega = \mathbb{R}^d$ , we take  $\mathcal{F} = \mathcal{B}(\mathbb{R}^d)$

How to build  $P$ ?

Provided an integrable function  $f: \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$   
s.t.

$$\int_{\mathbb{R}^d} f(x) dx = 1 \quad \left( \begin{array}{l} \text{this function is called a } \\ \text{probability density function} \end{array} \right)$$

We can build a probability measure  $P_f$ :

$$P_f(B) = \int_B f(x) dx \quad B \in \mathcal{B}(\mathbb{R}^d)$$

$\Rightarrow$  The probability space is

$$(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P_f)$$

Notation: We will use  $\mathcal{P}(\Omega)$  to denote the set of all prob. measures over  $\Omega$

- If  $\Omega$  is countable then  $\mathcal{F} = 2^{\Omega}$

- If  $\Omega = \mathbb{R}^d$ , then  $\mathcal{F} = \mathcal{B}(\mathbb{R}^d)$

# Properties of probability space

(i)  $P(\emptyset) = 1 - P(\Omega) = 0$

(ii)  $P(A^c) = 1 - P(A)$  Law of Complement Probability

(iii)  $P(A) \leq P(B)$  for  $A \subseteq B$  Monotonicity

(iv)

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

(v)  $\sigma$ -increasing

Let  $A_1 \subseteq A_2 \subseteq \dots$  be increasing events to an event  $A \in \mathcal{F}$  ( $\bigcup_{i=1}^{\infty} A_i = A$ )

then

$$\lim_{n \rightarrow \infty} P(A_n) = P(A)$$

(vi) decreasing

let  $A_1 \supseteq A_2 \supseteq \dots$  be decreasing events  $A \in \mathcal{F}$   
 $(\bigcap_{n=1}^{\infty} A_n = A)$

then  $\lim_{n \rightarrow \infty} P(A_n) = P(A)$

## Conditional Probability

Given  $(\Omega, \mathcal{F}, P)$  and  $A \in \mathcal{F}$  w/  $P(A) > 0$ , can define

$$P(\cdot | A) : \mathcal{F} \rightarrow [0, 1] \quad \text{by}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \forall B \in \mathcal{F}$$

One can show that  $(\Omega, \mathcal{F}, P(\cdot | A))$  is a prob. space

## Bayes Theorem

Let  $A, B \in \mathcal{F}$ ,  $P(A), P(B) > 0$ , then

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