2 Probability

(2.1) Definition. Let Ω be a set. The *power set* of Ω is defined by:

$$\mathcal{P}(\Omega) \;:=\; \{S \mid S \subseteq \Omega\}$$

(2.2) Example. The power set of $\Omega = \{0, 1\}$ is given by: $\mathcal{P}(\Omega) = \{\emptyset, \{0\}, \{1\}, \Omega\}$

(2.3) Definition. Let Ω be a set and $\mathcal{A}$ be a subset of the power set of Ω:

$$\mathcal{A} \;\subseteq\; \mathcal{P}(\Omega)$$

$\mathcal{A}$ is called a *σ-algebra* over Ω, if the following holds:

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

(ii) $A_i \in \mathcal{A}$ for $i \in \mathbb{N} \quad \Longrightarrow \quad \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

(iii) $A \in \mathcal{A} \quad \Longrightarrow \quad \overline{A} := A^c := (\Omega \setminus A) \in \mathcal{A}$

(2.4) Examples.

(i) For every set Ω, $\mathcal{A} = \mathcal{P}(\Omega)$ is a σ-algebra over Ω.

(ii) For every set Ω, $\mathcal{A} = \{\emptyset, \Omega\}$ is a σ-algebra over Ω.

(iii) For $\Omega = \{1, 2, 3\}$, $\mathcal{A} = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ is a σ-algebra over Ω.

(2.5) Lemma. Let $\mathcal{A}$ be a σ-algebra over Ω. Then:

(i) $A_i \in \mathcal{A}$ for $i \in \mathbb{N} \quad \Longrightarrow \quad \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$

(ii) $A, B \in \mathcal{A} \quad \Longrightarrow \quad A \cup B,\ A \cap B,\ A \setminus B \in \mathcal{A}$

(2.6) Definition. Let Ω be a set of possible outcomes of a random experiment and $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be a σ-algebra. Then Ω is called the *sample space* of the experiment and $\mathcal{A}$ is the *set of events* considered in the experiment. For some $x \in \Omega$ it is said that the event $E \in \mathcal{A}$ occurs if and only if $x \in E$.

(2.7) Remark. If Ω is finite or countable infinite, all subsets of Ω are usually considered to be events, i.e. $\mathcal{A} = \mathcal{P}(\Omega)$.

(2.8) Example. Throwing a die can be considered to be a random experiment with sample space:

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

In this example, the subset $A = \{2, 4, 6\}$ of Ω is the event, that an even number has been thrown.

(2.9) Example. Throwing two dice can be considered to be a random experiment with sample space

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

where i denotes the result from the first die and j the result from the second die. (Assuming that the dice are distinguishable.)

(2.10) Example. Throwing a die, until a specific face, say six, shows up, can be considered to be a random experiment with sample space:

$$\Omega \;=\; \mathbb{N}$$

In this example, the subset $A = \{1, 2, 3, 4, 5, 6, 7\}$ of Ω is the event, that six shows up after the die has been thrown for at most seven times.

(2.11) Example. The durations of cell-phone calls can be considered to be samples of a random experiment with sample space:

$$\Omega \;=\; [0, \infty)$$

(2.12) Definition. A *probability measure* (or simply a *probability*) is a mapping

$$\Pr : \mathcal{A} \to \mathbb{R}$$

defined on a set of events $\mathcal{A}$ of a sample space Ω, such that:

(i) $\Pr(A) \;\geq\; 0 \quad \text{for all} \quad A \in \mathcal{A}$

(ii) $\Pr(\Omega) \;=\; 1$

(iii) For every countable sequence of pairwise disjoint events $A_i \in \mathcal{A}$ $(i \in \mathbb{N})$:

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

(2.13) Theorem. Let Pr be a probability measure defined on a set of events $\mathcal{A}$ of a sample space Ω and let A, B, $A_i \in \mathcal{A}$ $(i = 1, \ldots, n)$, then the following statements are true:

(i) $\Pr(\emptyset) = 0$

(ii) $\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i)$ if $A_i \cap A_j = \emptyset$ for all $i \neq j$

(iii) $\Pr(A^c) = 1 - \Pr(A)$

(iv) $A \subseteq B \implies \Pr(A) \leq \Pr(B)$

(v) $0 \leq \Pr(A) \leq 1$

(vi) $\Pr(A \setminus B) = \Pr(A \cap B^c) = \Pr(A) - \Pr(A \cap B)$

(vii) $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

(viii)

$$\begin{aligned}
\Pr\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} \Pr\left(\bigcap_{i \in I} A_i\right) \\
&= \sum_{i=1}^{n} \Pr(A_i) - \sum_{i_1 < i_2} \Pr(A_{i_1} \cap A_{i_2}) + \sum_{i_1 < i_2 < i_3} \Pr(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\
&\quad - \sum_{i_1 < i_2 < i_3 < i_4} \Pr(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap A_{i_4}) \\
&\quad + \cdots + (-1)^{n+1} \Pr(A_1 \cap A_2 \cap \cdots \cap A_n)
\end{aligned}$$

Proof of (viii): Proof by induction.

$n = 1: \quad \Pr(A_1) = \Pr(A_1)$

$n \to (n+1): \quad$ Set $A_i^* := A_i \cap A_{n+1}$ for $i = 1, 2, \ldots, n$. Then:

$$\begin{aligned}
\Pr\left(\bigcup_{i=1}^{n+1} A_i\right) &= \Pr\left(\left(\bigcup_{i=1}^{n} A_i\right) \cup A_{n+1}\right) \\
&= \Pr\left(\bigcup_{i=1}^{n} A_i\right) + \Pr(A_{n+1}) - \Pr\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap A_{n+1}\right) \\
&= \Pr\left(\bigcup_{i=1}^{n} A_i\right) + \Pr(A_{n+1}) - \Pr\left(\bigcup_{i=1}^{n} A_i^*\right) \\
&= \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} \Pr\left(\bigcap_{i \in I} A_i\right) + \Pr(A_{n+1}) - \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} \Pr\left(\bigcap_{i \in I} A_i^*\right) \\
&= \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} \Pr\left(\bigcap_{i \in I} A_i\right) + \Pr(A_{n+1}) - \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} \Pr\left(\bigcap_{i \in I} A_i \cap A_{n+1}\right)
\end{aligned}$$

$$= \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} \Pr\left(\bigcap_{i \in I} A_i\right) + \Pr(A_{n+1}) - \sum_{\substack{I \subseteq \{1,2,\ldots,n,n+1\} \\ (n+1) \in I \neq \{n+1\}}} (-1)^{|I|} \Pr\left(\bigcap_{i \in I} A_i\right)$$

$$= \sum_{\substack{\emptyset \neq I \subseteq \{1,2,\ldots,n,n+1\} \\ (n+1) \notin I}} (-1)^{|I|+1} \Pr\left(\bigcap_{i \in I} A_i\right) + \sum_{\substack{I \subseteq \{1,2,\ldots,n,n+1\} \\ (n+1) \in I}} (-1)^{|I|+1} \Pr\left(\bigcap_{i \in I} A_i\right)$$

$$= \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n,n+1\}} (-1)^{|I|+1} \Pr\left(\bigcap_{i \in I} A_i\right)$$

(2.14) Notation. Let Ω be a finite sample space

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$$

with probability measure Pr defined on $\mathcal{A} = \mathcal{P}(\Omega)$. The probabilities of the elementary events (defined by the elements of Ω) are then given as follows:

$$p_i := \Pr(\{\omega_i\}) \qquad (i = 1, 2, \ldots, N)$$

(2.15) Lemma. With the notation from (2.14), we have

$$\Pr(A) = \sum_{\{i | \omega_i \in A\}} p_i \qquad \text{for all } A \in \mathcal{A}$$

In particular:

$$\sum_{i=1}^{N} p_i = 1$$

On the other hand, given any sequence of non-negative numbers

$$p_1,\, p_2, \ldots,\, p_N$$

such that $\sum_{i=1}^{N} p_i = 1$, a probability measure Pr is defined on $\mathcal{A} = \mathcal{P}(\Omega)$ from the formula above.

(2.16) Definition. A sample space with equally likely outcomes is called a *simple sample space*. With the notation from (2.14), Ω is a simple sample space, iff:

$$p_i = \frac{1}{|\Omega|} = \frac{1}{N} \qquad \text{for all } i \in \{1, 2, \ldots, N\}$$

For a simple sample space Ω:

$$\Pr(A) = \frac{|A|}{|\Omega|} \qquad \text{for all } A \in \mathcal{A}$$

(2.17) Example. A fair die, used in the random experiment described in (2.8), is expressed by a uniform probability measure:

$$p_i \;=\; \Pr(i) \;:=\; \frac{1}{6} \qquad \text{for } \; i = 1, 2, \ldots, 6$$

(2.18) Notation. The notation and results introduced in (2.14) and (2.15) can also be used for countable infinite sample spaces

$$\Omega \;=\; \{\omega_i \mid i \in \mathbb{N}\}$$

to characterise the probability measures on $\mathcal{A} = \mathcal{P}(\Omega)$. These probability measures correspond to all sequences $(p_i)_{i\in\mathbb{N}}$ of non-negative numbers, such that

$$\sum_{i=1}^{\infty} p_i \;=\; 1$$

again by simply setting:

$$p_i \;:=\; \Pr(\{\omega_i\}) \qquad \text{for all } \; i \in \mathbb{N}$$

(2.19) Example. If the random experiment described in (2.10) is based on a fair die, the probabilities for the elementary events are given by:

$$p_i \;:=\; \left(\frac{5}{6}\right)^{i-1} \cdot \frac{1}{6} \qquad \text{for all } \; i \in \mathbb{N}$$

(2.20) Example. Generalizing (2.10) by considering the repetition of a simple random experiment with probability p for success, gives rise to a probability measure for $\mathbb{N}$ with:

$$p_i \;=\; p \cdot (1-p)^{i-1} \qquad \text{for all } \; i \in \mathbb{N}$$

The probability for the event, that success occurs after at most n repetitions, is given by:

$$s_n \;=\; \sum_{i=1}^{n} p_i \;=\; p \cdot \sum_{i=1}^{n} (1-p)^{i-1} \;=\; p \cdot \sum_{i=0}^{n-1} (1-p)^{i} \;=\; p \cdot \frac{1-(1-p)^n}{1-(1-p)} \;=\; 1-(1-p)^n$$

For example:

$$s_n \geq \frac{1}{2} \quad \Longleftrightarrow \quad \frac{1}{2} \leq 1-(1-p)^n \quad \Longleftrightarrow \quad (1-p)^n \leq \frac{1}{2} \quad \Longleftrightarrow \quad n \geq -\frac{\ln(2)}{\ln(1-p)}$$

For small p, the following holds:

$$\ln(1-p) \;\lesssim\; -p \quad \Longleftrightarrow \quad -\ln(1-p) \;\gtrsim\; p \quad \Longleftrightarrow \quad \frac{\ln(2)}{p} \;\gtrsim\; -\frac{\ln(2)}{\ln(1-p)}$$

Hence:

$$n \;\geq\; \frac{\ln(2)}{p} \quad \Longrightarrow \quad s_n \;\geq\; \frac{1}{2}$$

(2.21) Example (The matching problem). Let p_n denote the probability that a random permutation of $n \in \mathbb{N}$ elements has at least one fixed point. Then:

$$p_n \;=\; 1 - \sum_{i=0}^{n} \frac{(-1)^i}{i!} \quad \xrightarrow[n\to\infty]{} \quad 1 - \frac{1}{e} \;\approx\; 0{,}6321$$

<u>*Proof:*</u> For the set of permutations of n elements, let A_i denote the event, that the i'th element is fixed. Then (2.13)(viii) yields:

$$\begin{aligned}
p_n \;=\; \Pr\left(\bigcup_{i=1}^{n} A_i\right) &= n \cdot \frac{1}{n} - \binom{n}{2} \cdot \frac{1}{n\cdot(n-1)} + \binom{n}{3} \cdot \frac{1}{n\cdot(n-1)\cdot(n-2)} \mp \ldots \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \ldots + (-1)^{n+1}\frac{1}{n!} \\
&= -\sum_{i=1}^{n} \frac{(-1)^i}{i!} \;=\; 1 - \sum_{i=0}^{n} \frac{(-1)^i}{i!} \\
&\xrightarrow[n\to\infty]{} 1 - \sum_{i=0}^{\infty} \frac{(-1)^i}{i!} \;=\; 1 - e^{-1}
\end{aligned}$$

(2.22) Exercise. Let p_n be as in (2.21) and let $q_n(i)$ denote the probability that a random permutation of $n \in \mathbb{N}$ elements has exactly i fixed points ($i \in \{0, 1, \ldots, n\}$). Show:

$$\begin{aligned}
q_n(0) &= 1 - p_n \;=\; \sum_{i=0}^{n} \frac{(-1)^i}{i!} \;=\; q_{n-1}(0) + \frac{(-1)^n}{n!} \\
q_n(i) &= \frac{1}{i!} \cdot q_{n-i}(0) \;=\; \frac{1}{i!} \cdot (1 - p_{n-i}) \\
\lim_{n\to\infty} q_n(i) &= \frac{1}{i!} \cdot \frac{1}{e}
\end{aligned}$$

Moreover, calculate $p_2, p_3, p_4, p_5,\; q_3(1), q_4(1), q_5(1)$ and $q_5(2)$.

(2.23) Example. A needle is thrown randomly and its direction $\alpha \in [0, 2\pi)$ (with respect to some fixed orientation in the plane) is measured. The result can be considered to be a sample of the sample space $\Omega = [0, 2\pi)$. Now, if we want to define something like a uniform probability measure, the challenge is to define a reasonable set of events, i.e. a suitable σ-algebra $\mathcal{A}$ over $[0, 2\pi)$ and a probability measure on $\mathcal{A}$. Remarkably, no such definition is possible for $\mathcal{A} = \mathcal{P}(\Omega)$. Nevertheless, it can be shown that a probability measure can be defined on the σ-algebra $\mathcal{A}$ generated by all the subintervals of $\Omega = [0, 2\pi)$, such that

$$\Pr\big((a, b)\big) \;=\; \frac{b-a}{2\pi}$$

for all subintervals (a, b) of Ω.

Conditional probabilities

(2.24) Definition. Let Pr be a probability measure defined on a set of events $\mathcal{A}$ of a sample space Ω and let $A, B \in \mathcal{A}$. If $\Pr(B) \neq 0$, then the *conditional probability* of A given the event B is defined to be:

$$\Pr(A|B) \quad := \quad \frac{\Pr(A \cap B)}{\Pr(B)}$$

(2.25) Lemma. Let Pr be a probability measure defined on a set of events $\mathcal{A}$ of a sample space Ω and let $A, B \in \mathcal{A}$.

(i) If $\Pr(B) \neq 0$, then:

$$\Pr(A \cap B) \quad = \quad \Pr(A|B)\ \Pr(B)$$

(ii) If $\Pr(A) \neq 0$ and $\Pr(B) \neq 0$, then:

$$\Pr(B|A) \quad = \quad \frac{\Pr(A|B)\ \Pr(B)}{\Pr(A)}$$

(iii) If $\Pr(B) \neq 0$ and $\Pr(B^c) \neq 0$, then:

$$\Pr(A) \quad = \quad \Pr(A|B)\ \Pr(B) \ + \ \Pr(A|B^c)\ \Pr(B^c)$$

Proof of (iii):

$$\Pr(A) \quad = \quad \Pr(A \cap B) \ + \ \Pr(A \cap B^c) \ \stackrel{\text{(i)}}{=} \ \Pr(A|B)\ \Pr(B) \ + \ \Pr(A|B^c)\ \Pr(B^c)$$

(2.26) Theorem. Let Pr be a probability measure defined on a set of events $\mathcal{A}$ of a sample space Ω. Furthermore, let $B_i \in \mathcal{A}$ $(i = 1, \ldots, n)$ with $\Pr(B_i) \neq 0$ for $i = 1, \ldots, n$ and:

- $\Pr(B_i \cap B_j) = 0$ for $i \neq j$
- $\displaystyle\sum_{i=1}^{n} \Pr(B_i) \ = \ 1$

Then for any $A \in \mathcal{A}$ the following holds:

(i) *Law of total probability*:

$$\Pr(A) \quad = \quad \sum_{i=1}^{n} \Pr(A|B_i)\ \Pr(B_i)$$

(ii) *Bayes' rule*:

$$\Pr(B_j|A) \quad = \quad \frac{\Pr(A|B_j)\ \Pr(B_j)}{\displaystyle\sum_{i=1}^{n} \Pr(A|B_i)\ \Pr(B_i)} \qquad \text{if} \ \ \Pr(A) \neq 0$$

<u>*Proof:*</u> For $i = 1, \ldots, n$ set

$$\tilde{B}_i \;:=\; B_i \setminus \bigcup_{j\neq i} B_j, \quad B_i^0 \;:=\; B_i \setminus \tilde{B}_i \;=\; \bigcup_{j\neq i}(B_i \cap B_j)$$

and define:

$$\tilde{B}_{n+1} \;:=\; \Omega \setminus \bigcup_{i=1}^{n} \tilde{B}_i$$

Then, for $i = 1, \ldots, n$ and any $A \in \mathcal{A}$, the following holds:

$$B_i \;=\; \tilde{B}_i \,\dot\cup\, B_i^0, \qquad \Pr(B_i^0) \;=\; 0, \qquad \Pr(\tilde{B}_i) \;=\; \Pr(B_i), \qquad \Pr(A \cap B_i) \;=\; \Pr(A \cap \tilde{B}_i)$$

Furthermore

$$\Omega \;=\; \dot{\bigcup_{i=1}^{n+1}} \tilde{B}_i$$

and

$$\sum_{i=1}^{n} \Pr(\tilde{B}_i) \;=\; 1 \;=\; \Pr(\Omega) \;=\; \sum_{i=1}^{n+1} \Pr(\tilde{B}_i)$$

implies:

$$\Pr(\tilde{B}_{n+1}) \;=\; 0$$

(i):

$$\begin{aligned} \Pr(A) \;&=\; \Pr(A \cap \Omega) \;=\; \Pr\left(A \cap \left(\dot{\bigcup_{i=1}^{n+1}} \tilde{B}_i\right)\right) \;=\; \Pr\left(\dot{\bigcup_{i=1}^{n+1}} (A \cap \tilde{B}_i)\right) \\ &=\; \sum_{i=1}^{n+1} \Pr\left(A \cap \tilde{B}_i\right) \;=\; \sum_{i=1}^{n} \Pr\left(A \cap B_i\right) \;=\; \sum_{i=1}^{n} \Pr(A|B_i)\,\Pr(B_i) \end{aligned}$$

(ii):

$$\Pr(B_j|A) \;=\; \frac{\Pr(B_j \cap A)}{\Pr(A)} \;\overset{\text{(i)}}{=}\; \frac{\Pr(A|B_j)\,\Pr(B_j)}{\sum_{i=1}^{n} \Pr(A|B_i)\,\Pr(B_i)}$$

(2.27) Definition. Let Pr be a probability measure defined on a set of events $\mathcal{A}$ of a sample space Ω. $A, B \in \mathcal{A}$ are called *(stochastically) independent* if:

$$\Pr(A \cap B) \;=\; \Pr(A)\,\Pr(B)$$

(2.28) Remark. If $A, B \in \mathcal{A}$ are events with $\Pr(A) \neq 0$ and $\Pr(B) \neq 0$. Then the following statements are equivalent:

- A and B are stochastically independent.
- $\Pr(A|B) \;=\; \Pr(A)$
- $\Pr(B|A) \;=\; \Pr(B)$

(2.29) Definition. Let Pr be a probability measure defined on a set of events $\mathcal{A}$ of a sample space Ω. Then, $A_i \in \mathcal{A}$ $(i \in I)$ are called *independent* if

$$\mathrm{Pr}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} Pr(A_j)$$

for every finite subset $J \subseteq I$.