Probability Theory

FACULTY OF SCIENCE AND TECHNOLOGY, BNU-HKBU UNITED INTERNAITONAL COLLEGE

CHAPTER 1

Events and Their Probabilities

1.1. Basic notions of set theory

Let Ω be a set.

DEFINITION 1. The collection of all subsets of Ω, denoted by

$$\mathcal{P}(\Omega) \;=\; \{A \mid A \subset \Omega\},$$

is called the potential set of Ω.

REMARK 2. *a*) $\emptyset \in \mathcal{P}(\Omega)$

b) *For* $A \in \mathcal{P}(\Omega)$ *define*

$$A^c \;=\; \{\omega \in \Omega \mid \omega \notin A\} \;\in\; \mathcal{P}(\Omega)$$

as the complement of A.

c) *For* $A, B \in \mathcal{P}(\Omega)$ *define*

$$A \setminus B \;=\; \{\omega \in \Omega \mid \omega \in A \text{ and } \omega \notin B\} \;\in\; \mathcal{P}(\Omega).$$

d) *Let* I *be an index set (e.g.* $I = \{1, \ldots, n\}, I = \mathbb{N}, I = \mathbb{R}$ *or any other non-empty set) and* $A_i \subset \Omega$ *for each* $i \in I$. *Define*

$$\bigcup_{i \in I} A_i = \{\omega \in \Omega | \exists i_0 \in I \text{ such that } \omega \in A_{i_0}\}$$

and

$$\bigcap_{i \in I} A_i \;=\; \{\omega \in \Omega \mid \omega \in A_i,\ \forall i \in I\}.$$

THEOREM 3 (Distributive and De Morgan's laws). *Let I be an index set and $A_i \subset \Omega$ for each $i \in I$. Then*

(1) for any $B \subset \Omega$,

$$\bigcup_{i\in I}(B \cap A_i) = B \cap \left(\bigcup_{i\in I} A_i\right) \quad \textit{and} \quad \bigcap_{i\in I}(B \cup A_i) = B \cup \left(\bigcap_{i\in I} A_i\right);$$

(2)

$$\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c \quad \text{and} \quad \left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c.$$

DEFINITION 4. (i) Two sets X and Y are called equinumerous or equipotent, written $X \sim Y$, if there is a bijection from X to Y.

(ii) A set X is called countably infinite if $X \sim \mathbb{N}$, and we say X is countable if $X \sim \mathbb{N}$ or X is finite. Finally, X is uncountable if X is not countable.

A set is countable if you can count its elements. Of course if the set is finite, you can easily count its elements. If the set is infinite, being countable means that you are able to put the elements of the set in order just like natural numbers are in order. Of course, the set of natural numbers is countably infinite. More interesting is the observation that proper subsets of countably infinite sets can themselves be countably infinite, as the example of the set of even natural numbers $2\mathbb{N} = \{2n; n \in \mathbb{N}\}$ shows.

THEOREM 5. *The set $\mathbb{Q}$ of rational numbers is countably infinite.*

PROOF. We first show that the set of all positive rational numbers,

$$\mathbb{Q}_{>0} = \left\{\frac{r}{s} \,|\, r, s \in \mathbb{N}\right\}$$

is a countable set, i.e., we will arrange the rational numbers into a particular order that allows us to count them.

We begin by creating a chart with the numerators ascending from left to right and denominators ascending from top to bottom:

$$\begin{array}{ccccc} 1/1 & 2/1 & 3/1 & 4/1 & \cdots \\ 1/2 & 2/2 & 3/2 & 4/2 & \cdots \\ 1/3 & 2/3 & 3/3 & 4/3 & \cdots \\ 1/4 & 2/4 & 3/4 & 4/4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{array}$$

Now, we can see that it is possible to assign the number 1 to the top left corner. Then the number 2 to the 1/2 beneath 1/1, and 3 to the 2/1 above and to the right. We will continue by ordering each number along ascending diagonals starting in the first column and moving up and to the right, ignoring any fraction that has already been included on the list already. This effectively counts the positive rational numbers. In other words, we can write

$$\mathbb{Q}_{>0} = \{a_1, a_2, a_3, \ldots\}.$$

By writing

$$\mathbb{Q} = \{0, a_1, -a_1, a_2, -a_2, a_3, -a_3, \ldots\},$$

we see that all rational numbers are also put in order, which means $\mathbb{Q}$ is countabe. □

1.2. Events as sets

The set of all possible outcomes of a random experiment is known as the sample space of the experiment and is denoted by Ω.

Example 6.

(a) If the experiment consists of tossing two coins, then the sample space consists of the following four outcomes:

$$\Omega = \{HH, HT, TH, TT\}.$$

(b) If the experiment consists of drawing two numbers from $\{1,2,3,4\}$ without replacement, then the sample space consists of 6 outcomes:

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

(c) If the experiment consists of tossing two dice, then the sample space consists of 36 outcomes:

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

where the outcome (i,j) is said to occur if i appears on the first dice and j on the second dice.

(d) If the experiment consists of measuring (in hours) the lifetime of a transistor, then the sample space consists of all nonnegative real numbers; that is,

$$\Omega = \{x : 0 \leq x < \infty\}.$$

An "event" is a property which can be observed either to hold or not to hold *after* the experiment is done. In Example 6.(a), if

$$E = \{HH, HT\},$$

then E is the event that a head appears on the first coin. In Example 6.(c), if

$$E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}\,,$$

then E is the event that the sum of the dice equals 7. In Example 6.(d), if

$$E = \{x : 0 < x \leq 5\},$$

then E is the event that the transistor does not last longer than 5 hours.

In mathematical terms, an event is a subset of Ω. If A and B are two events, then

- the *contrary* event of A is interpreted as the complement set A^c;
- the event "A or B" is interpreted as the union $A \cup B$;

• the event "A and B" is interpreted as the intersection $A \cap B$;

• the *sure* event is Ω;

• the *impossible* event is the empty set $\emptyset$.

Thus events are subsets of Ω, but need all the subsets of Ω be events? The answer is no, but some of the reasons for this are too difficult to be discussed here. We denote by $\mathcal{A}$ the family of all events. It suffices for us to think of $\mathcal{A}$ as a collection of subsets of Ω. This collection $\mathcal{A}$ should have at least the following properties:

1. $\emptyset \in \mathcal{A}$;

2. If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$, where A^c denotes the complement of A;

3. $\mathcal{A}$ is closed under countable unions: that is, if $A_1, A_2, A_3, \ldots$ is a countably infinite sequence of events in $\mathcal{A}$, then $\cup_{i=1}^{\infty} A_i$ is also in $\mathcal{A}$.

DEFINITION 7. Any collection $\mathcal{A}$ of subsets of Ω which satisfies the above three conditions is called a σ-algebra (*or a σ-field*).

PROPOSITION 8. *Suppose $\mathcal{A}$ is a σ-algebra, $A_1, A_2, \ldots$ are in $\mathcal{A}$, and $m \in \mathbb{N}$. Then each of the sets*

$$\Omega,\ A_1 \backslash A_2,\ \bigcup_{j=1}^{m} A_j,\ \bigcap_{j=1}^{m} A_j,\ \bigcap_{j \in \mathbb{N}} A_j$$

also belongs to $\mathcal{A}$.

PROOF. Set

$B_k := \begin{cases} A_k\ , & k \leq m, \\ A_m\ , & k > m, \end{cases}$ so that $B_k \in \mathcal{A}$, $k \in \mathbb{N}$ and therefore

$$\bigcup_{k \in \mathbb{N}} B_k = \bigcup_{j=1}^{m} A_j \ \in \mathcal{A}\ .$$

The remaining statements follow from de Morgan's laws. □

We now look at a few examples.

EXAMPLE 9. a) *$\mathcal{P}(\Omega)$ is a σ-algebra over Ω.*

b) *$\{\Omega,\ \emptyset\}$ is the trivial and smallest σ-algebra over Ω.*

c) *The collection* $\mathcal{F} = \{\emptyset,\ \{a\},\ \{b\},\ \{c,\ d\},\ \{a,\ c,\ d\},\ \{b,\ c,\ d\},\ \Omega\}$ *is* not *a* σ*-algebra over* $\Omega = \{a,\ b,\ c,\ d\}$, since $\{a,\ b\} \notin \mathcal{F}$.

To recapitulate, with any random experiment we may associate a pair $(\Omega, \mathcal{A})$, where Ω is the set of all possible outcomes and $\mathcal{A}$ is a σ-algebra of subsets of Ω which contains all the events in whose occurrences we may be interested; henceforth, to call a set A an event is equivalent to asserting that A belongs to the σ-algebra in question.

DEFINITION 10. A probability measure *defined on a* σ*-algebra* $\mathcal{A}$ *of* Ω *is a function* $\mathbb{P} : \mathcal{A} \to [0, 1]$ *that satisfies*:

1. $\mathbb{P}(\Omega) = 1$*;*

2. $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ *whenever the* A_i *are in* $\mathcal{A}$ *and are pairwise disjoint (i.e.* $A_n \cap A_m = \emptyset$ *if* $n \neq m$*).*

Axiom (2) above is called σ*-additivity*; the number $\mathbb{P}(A)$ for $A \in \mathcal{A}$ is called the *probability* of the event A. The triple $(\Omega, \mathcal{A}, \mathbb{P})$ is called a *probability space*.

Whenever we have an abstract definition such as this one, the first thing to do is to look for examples. Here are some.

EXAMPLE 11. $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$, and

$$\mathbb{P}(A) = \frac{|A|}{6}, \quad A \in \mathcal{A},$$

where $|A|$ denotes the number of elements in A. The random experiment here is rolling a fair dice. Clearly, this can be generalized to any finite set with equally likely outcomes.

The sample space can also be infinite, as shown in the next example.

EXAMPLE 12. $\Omega = \{1, 2,\ \ldots\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ and you have numbers $p_1, p_2, \ldots \geq 0$ with $p_1 + p_2 + \ldots = 1$. For any $A \subset \Omega$,

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

For example, toss a fair coin repeatedly until the first Head. Your outcome is the number of tosses. Here, $p_i = 2^{-i}$. Note that p_i cannot be chosen to be equal, as you cannot make the sum of infinitely many equal numbers to be 1.

Let's turn to an example of geometric nature. In practice, we can often use the geometric concepts as length, area or volume to constuct probabilty measures.

EXAMPLE 13. Pick a point from inside the unit circle centered at the origin. Here,

$$\Omega = \{(x, y) : x^2 + y^2 < 1\}$$

and

$$\mathbb{P}(A) = \frac{\text{(area of } A)}{\pi}.$$

The problem of finding an appropriate σ-algebra on Ω is highly non-trivial and is skipped here. However, we don't take $\mathcal{P}(\Omega)$, since this σ-algebra is too large. Intuitively, simple sets like points or segments within Ω should be events. It is important to observe that if A is a singleton (a set whose element is a single point), then $P(A) = 0$. Only "fatter" sets with positive area can be hitten with positive probability.

We now introduce some notation. Let A *and* $A_i, i = 1, 2, \ldots$, be subsets of Ω.

(i) $A_n \uparrow A$ means that $A_1 \subset A_2 \subset \ldots$ and $\cup_{n=1} A_n = A$.

(ii) $A_n \downarrow A$ means that $A_1 \supset A_2 \supset \ldots$ and $\cap_{n=1} A_n = A$.

We gather here the most important rules for working with probability measures.

PROPOSITION 14 (Properties of probability measures). *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space. For* $A, B \in \mathcal{A}$ *and* $A_i \in \mathcal{A}$, $i = 1, 2, \ldots$, *we have:*

(i) $\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$ *if* $A_1, \ldots, A_n$ *are pairwise disjoint. This means that* $\mathbb{P}$ *is additive.*

(ii) $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$. *This property is called the inclusion-exclusion principle.*

(iii) $\mathbb{P}(B \backslash A) = \mathbb{P}(B) - \mathbb{P}(A)$ *if* $A \subset B$.

(iv) $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$.

(v) $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$*, that is,* $\mathbb{P}$ *is monotone.*

(vi) (continuity from below) $\mathbb{P}(A_n)\uparrow \mathbb{P}(A)$ *if* $A_n \uparrow A$*.*

(vii) (continuity from above) $\mathbb{P}(A_n)\downarrow \mathbb{P}(A)$ *if* $A_n \downarrow A$*.*

Proof. (i) Set

$B_k := \begin{cases} A_k\ , & k \leq n, \\ \emptyset\ , & k > n, \end{cases}$ so that $B_k \in \mathcal{A}$, $k \in \mathbb{N}$ and

$$\bigcup_{k\in\mathbb{N}} B_k = \bigcup_{j=1}^{n} A_j \ \in \mathcal{A}.$$

Therefore,

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

(ii) From $A \cup B = A \cup (B\backslash A)$ and $A \cap (B\backslash A) = \emptyset$ it follows by additivity that

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B\backslash A). \tag{1.1}$$

Analogously, we get from $B = (A \cap B) \cup (B\backslash A)$ and $(A \cap B) \cap (B\backslash A) = \emptyset$ that

$$\mathbb{P}(A \cap B) + \mathbb{P}(B\backslash A) = \mathbb{P}(B). \tag{1.2}$$

By adding (1.1) and (1.2) we find

$$\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) + \mathbb{P}(B\backslash A) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(B\backslash A)$$

Subtracting $\mathbb{P}(B\backslash A)$ from both sides, the claim follows.

(iii) Since $A \subset B$, we have $B = A \cup (B\backslash A)$. But A and $B\backslash A$ are disjoint, so $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B\backslash A)$. Therefore, $\mathbb{P}(B) - \mathbb{P}(A) = \mathbb{P}(B\backslash A)$.

(iv) Choosing $B = \Omega$ in (iii), we get $\mathbb{P}(A^c) = \mathbb{P}(\Omega \setminus A) = \mathbb{P}(\Omega) - \mathbb{P}(A) = 1 - \mathbb{P}(A)$.

(v) We have $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B\backslash A)$ and thus $\mathbb{P}(B) \geq \mathbb{P}(A)$.

(vi) We set $A_0 := \emptyset$ and $B_k := A_k \backslash A_{k-1}$ for $k \in \mathbb{N}$. By assumption, (B_k) is a disjoint sequence in $\mathcal{A}$ with $\bigcup_{k=1}^{\infty} B_k = \bigcup_{j=1}^{\infty} A_j$ and $\bigcup_{k=1}^{m} B_k = A_m$. The countable additivity of $\mathbb{P}$ therefore implies

$$\mathbb{P}(\bigcup_j A_j) = \mathbb{P}(\bigcup_k B_k) = \lim_{m\to\infty} \sum_{k=1}^{m} \mathbb{P}(B_k) = \lim_{m\to\infty} \mathbb{P}(\cup_{k=1}^{m} B_k) = \lim_{m\to\infty} \mathbb{P}(A_m).$$

(vii) If (A_k) is a decreasing sequence in $\mathcal{A}$, then $A_k^c = (\Omega \backslash A_k)$ is increasing. Further,

$$(\bigcap_k A_k)^c = \bigcup_k (A_k^c).$$

Using (iii) and (v), we get

$$\begin{aligned} 1 - \mathbb{P}(\bigcap_k A_k) &= \mathbb{P}((\bigcap_k A_k)^c) = \mathbb{P}(\bigcup_k (A_k^c)) \\ &= \lim_{m\to\infty} \mathbb{P}(A_m^c) = 1 - \lim_{m\to\infty} \mathbb{P}(A_m), \end{aligned}$$

from which the claim follows. □

EXAMPLE 15. Among a large group of patients recovering from shoulder injuries, it is found that 22% visit both a physical therapist and a chiropractor, whereas 12% visit neither of these. The probability that a patient visits a chiropractor exceeds by 0.14 the probability that a patient visits a physical therapist. Calculate the probability that a randomly chosen member of this group visits a physical therapist.

SOLUTION. Let C = event that patient visits a chiropractor and T = event that patient visits a physical therapist. We are given that

$$\mathbb{P}(C) = \mathbb{P}(T) + 0.14, \quad \mathbb{P}(C \cap T) = 0.22, \quad \mathbb{P}((C \cup T)^c) = 0.12$$

Therefore,

$$
\begin{aligned}
0.88 &= 1 - \mathbb{P}((C \cup T)^c) \\
&= \mathbb{P}(C \cup T) \\
&= \mathbb{P}(C) + \mathbb{P}(T) - \mathbb{P}(C \cap T) \\
&= \mathbb{P}(T) + 0.14 + \mathbb{P}(T) - 0.22 \\
&= 2\mathbb{P}(T) - 0.08,
\end{aligned}
$$

which implies $\mathbb{P}(T) = 0.48$.

1.3. Equally likely outcomes

We have seen in Example 11 how we model the random experiment of rolling a fair die. In general, if the sample space is finite: $|\Omega| = N < \infty$, and each outcome of our experiment has equal probability, then necessarily this probability equals $1/N$:

$$\mathbb{P}(\{w\}) = \frac{1}{N}, \quad \omega \in \Omega.$$

For such an experiment, as in Example 11, we can simply choose $\mathcal{A} = \mathcal{P}(\Omega)$ and

$$\mathbb{P}(A) = \frac{\text{(number of elements in } A)}{N}, \quad A \in \mathcal{A}.$$

Example 16.

(a) Rolling two dice, what is the probability that the sum of the numbers shown is 7? Defining $E = \{\text{sum is } 7\} = \{(1,6), (2,5), ..., (6,1)\}$ in the sample space $\Omega = \{(i,j) : i,j = 1, ..., 6\}$, and noting that each pair of numbers is equally likely, we have

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}.$$

(b) An urn contains 6 red and 5 blue balls. We draw three balls at random, at once (that is, without replacement). What is the chance of drawing one red and two blue balls? Define the sample space as unordered choices of 3 out of the 11 balls. Then each choice is equally likely, and

$$|\Omega| = \binom{11}{3} = \frac{11 \cdot 10 \cdot 9}{6}.$$

Now, our event E consists of picking 1 out of the 6 red balls and 2 out of the 5 blue balls, with no respect to order. Thus,

$$|E| = \binom{6}{1} \cdot \binom{5}{2} = 6 \cdot 10,$$

and the answer is $\mathbb{P}(E) = |E|/|\Omega| = 6 \cdot 10/(11 \cdot 10 \cdot 9/6) = 4/11$.

(c) Out of n people, what is the probability that there are no coinciding birthdays? Assume 365 days in the year. The answer is of course zero, if $n > 365$ (this is called the pigeonhole principle). Otherwise, our sample space Ω is a possible birthday for all n people, $|\Omega| = 365^n$. The event E of no coinciding birthdays can occur in

$$|E| = 365 \cdot 364 \cdots (365 - n + 1) = \frac{365!}{(365 - n)!}$$

many ways. The answer is

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{365!}{(365 - n)! \cdot 365^n}.$$

This is

$$\begin{aligned} 88\% &\quad \text{for } n = 10, \\ 59\% &\quad \text{for } n = 20, \\ 29\% &\quad \text{for } n = 30, \\ 11\% &\quad \text{for } n = 40, \\ 3\% &\quad \text{for } n = 50, \\ 0.00003\% &\quad \text{for } n = 100. \end{aligned}$$

How to count? Listing all outcomes is very inefficient, especially if their number is large. We will, therefore, learn a few counting techniques, starting with a trivial, but conceptually important fact.

Basic principle of counting. If an experiment consists of two stages and the first stage has m outcomes, while the second stage has n outcomes regardless of the outcome at the first stage, then the experiment as a whole has mn outcomes.

EXAMPLE 17. Roll a dice 4 times. What is the probability that you get different numbers? At least at the beginning, you should divide every solution into the following three steps:

Step 1: Identify the set of equally likely outcomes. In this case, this is the set of all ordered 4-tuples of numbers $1, ..., 6$. That is, $\{(a, b, c, d) \, : \, a, b, c, d \in \{1, \ldots, 6\}\}$.

Step 2: Compute the number of outcomes. In this case, it is therefore 6^4.

Step 3: Compute the number of good outcomes. In this case it is $6 \cdot 5 \cdot 4 \cdot 3$. Why? We have 6 options for the first roll, 5 options for the second roll since its number must differ from the number on the first roll; 4 options for the third roll since its number must not appear on the first two rolls, etc. Note that the *set* of possible outcomes changes from stage to stage (roll to roll in this case), but their *number* does not!

The answer then is $\dfrac{6 \cdot 5 \cdot 4 \cdot 3}{6^4} = \dfrac{5}{18} \approx 0.2778.$

EXAMPLE 18. Let us now compute probabilities for de Méré's games. In Game 1, there are 4 rolls and he wins with at least one 6. The number of good outcomes is $6^4 - 5^4$, as the number of *bad* outcomes is 5^4. Therefore

$$\mathbb{P}(\text{win}) = 1 - (\frac{5}{6})^4 \approx 0.5177.$$

In Game 2, there are 24 rolls of two dice and he wins by at least one pair of 6's rolled. The number of outcomes is 36^{24}, the number of bad ones is 35^{24}, thus the number of good outcomes equals $36^{24} - 35^{24}$. Therefore

$$\mathbb{P}(\text{win}) = 1 - (\frac{35}{36})^{24} \approx 0.4914.$$

Chevalier de Méré overcounted the good outcomes in both cases. His count $4{\cdot}6^3$ in Game 1 selects a die with a 6 and arbitrary numbers for other dice. However, many outcomes have more than one six and are hence counted more than once.

One should also note that both probabilities are barely different from 1/2, so de Méré was gambling *a lot* to be able to notice the difference.

Permutations.

Assume you have n objects. The number of ways to fill n ordered slots with them is

$$n \cdot (n-1) \dots 2 \cdot 1 = n!,$$

while the number of ways to fill $k \le n$ ordered slots is

$$n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}.$$

EXAMPLE 19. Shuffle a deck of cards.

- $\mathbb{P}(\text{top card is an Ace}) = \dfrac{1}{13} = \dfrac{4 \cdot 51!}{52!}$.
- $\mathbb{P}(\text{all cards of the same suit end up next to each other}) = \dfrac{4! \cdot (13!)^4}{52!} \approx 4.5 \cdot 10^{-28}$. This event never happens in practice.
- $\mathbb{P}(\text{hearts are together}) = \dfrac{40!13!}{52!} = 6 \cdot 10^{-11}$.

To compute the last probability, for example, collect all hearts into a block; a good event is specified by ordering 40 items (the block of hearts plus 39 other cards) and ordering the hearts within their block.

Before we go on to further examples, let us agree that when the text says *without further elaboration*, that a random choice is made, this means that *all available choices are equally likely*. Also, in the next problem (and in statistics in general) *sampling with replacement* refers to choosing, at random, an object from a population, noting its properties, putting the object back into the population, and then repeating. *Sampling without replacement* omits the putting back part.

EXAMPLE 20. A bag has 6 pieces of paper, each with one of the letters, E, E, P, P, P, and R, on it. Pull pieces at random out of the bag (1) without, and (2) with replacement. What is the probability that these pieces, in order, spell *PEPPER*?

There are two problems to solve. For sampling without replacement:

1. An outcome is an ordering of the pieces of paper $E_1, E_2, P_1, P_2, P_3, R$.
2. The number of outcomes thus is 6!.
3. The number of good outcomes is $3!2!$.

The probability is $\frac{3!2!}{6!} = \frac{1}{60}$.

For sampling with replacement, the answer is $\frac{3^3 \cdot 2^2}{6^6} = \frac{1}{2 \cdot 6^3}$, quite a lot smaller.

EXAMPLE 21. Sit 3 men and 3 women at random (1) in a row of chairs and (2) around a table. Compute $\mathbb{P}$(all women sit together). In case (2), also compute $\mathbb{P}$(men and women alternate).

In case (1), the answer is $\frac{4!3!}{6!} = \frac{1}{5}$.

For case (2), pick a man, say John Smith, and sit him first. Then, we reduce to a row problem with 5! outcomes; the number of good outcomes is $3! \cdot 3!$. The answer is $\frac{3}{10}$. For the last question, the seats for the men and women are fixed after John Smith takes his seat and so the answer is $\frac{3!2!}{5!} = \frac{1}{10}$.

EXAMPLE 22. A group consists of 3 Norwegians, 4 Swedes, and 5 Finns, and they sit at random around a table. What is the probability that all groups end up sitting together?

The answer is $\frac{3! \cdot 4! \cdot 5! \cdot 2!}{11!}$. Pick, say, a Norwegian (Arne) and sit him down. Here is how you count the good events. There are 3! choices for ordering the group of Norwegians (and then sit them down to one or both sides of Arne, depending on the ordering). Then, there are 4! choices for arranging the Swedes and 5! choices for arranging the Finns. Finally, there are 2! choices to order the two blocks of Swedes and Finns.

Combinations.

Let $\binom{n}{k}$ be the number of different subsets with k elements of a set with n elements. Then,

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}.$$

To understand why the above is true, first choose a subset, then order its elements in a row to fill k ordered slots with elements from the set with n objects. Then, distinct choices of a subset and its ordering will end up as distinct orderings. Therefore,

$$\binom{n}{k} k! = n(n-1)\cdots(n-k+1).$$

We call $\binom{n}{k}$ *"n choose k"* or a *binomial coefficient* (as it appears in the binomial theorem: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$). Also, note that

$$\binom{n}{0} = \binom{n}{n} = 1 \quad \text{and} \quad \binom{n}{k} = \binom{n}{n-k}.$$

The *multinomial coefficients* are more general and are defined next.

The number of ways to divide a set of n elements into r (distinguishable) subsets of $n_1, n_2, \ldots, n_r$ elements, where $n_1 + \ldots + n_r = n$, is denoted by $\binom{n}{n_1,\ldots,n_r}$ and

$$\binom{n}{n_1,\ldots,n_r} = \binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-\cdots-n_{r-1}}{n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}.$$

To understand the slightly confusing word *distinguishable*, just think of painting n_1 elements red, then n_2 different elements blue, etc. These colors distinguish among the different subsets.

EXAMPLE 23. A fair coin is tossed 10 times. What is the probability that we get exactly 5 Heads?

$$\mathbb{P}(\text{exactly 5 Heads}) = \frac{\binom{10}{5}}{2^{10}} \approx 0.2461,$$

as one needs to choose the position of the five heads among 10 slots to fix a good outcome.

EXAMPLE 24. We have a bag that contains 100 balls, 50 of them red and 50 blue. Select 5 balls at random. What is the probability that 3 are blue and 2 are red?

The number of outcomes is $\binom{100}{5}$ and all of them are equally likely, which is a reasonable interpretation of "select 5 balls at random." The answer is

$$\mathbb{P}(\text{3 are blue and 2 are red}) = \frac{\binom{50}{3}\binom{50}{2}}{\binom{100}{5}} \approx 0.3189.$$

Why should this probability be less than a half? The probability that 3 are blue and 2 are red is equal to the probability of 3 are red and 2 are blue and they cannot both exceed $\frac{1}{2}$, as their sum cannot be more than 1. It cannot be exactly $\frac{1}{2}$ either, because other possibilities (such as all 5 chosen balls red) have probability greater than 0.

EXAMPLE 25. Shuffle a standard deck of 52 cards and deal 13 cards to each of the 4 players.

What is the probability that each player gets an Ace? We will solve this problem in two ways to emphasize that you often have a choice in your set of equally likely outcomes.

The first way uses an outcome to be an ordering of 52 cards:

1. There are $52!$ equally likely outcomes.

2. Let the first 13 cards go to the first player, the second 13 cards to the second player, etc. Pick a slot within each of the four segments of 13 slots for an Ace. There are 13^4 possibilities to choose these four slots for the Aces.

3. The number of choices to fill these four positions with (four different) Aces is 4!.

4. Order the rest of the cards in $48!$ ways.

The probability, then, is $\frac{13^4 4! 48!}{52!}$.

The second way, via a small leap of faith, assumes that each set of the four *positions of the four Aces* among the 52 shuffled cards is equally likely. You may choose to believe this intuitive fact or try to write down a formal proof: the number of permutations that result in a given set F of four positions is independent of F. Then:

1. The outcomes are the positions of the 4 Aces among the 52 slots for the shuffled cards of the deck.

2. The number of outcomes is $\binom{52}{4}$.

3. The number of good outcomes is 13^4, as we need to choose one slot among 13 cards that go to the first player, etc.

The probability, then, is $\frac{13^4}{\binom{52}{4}}$, which agrees with the number we obtained the first way.

EXAMPLE 26. Roll a die 12 times. What is $\mathbb{P}$(each number appears exactly twice)?

1. An outcome consists of filling each of the 12 slots (for the 12 rolls) with an integer between 1 and 6 (the outcome of the roll).

2. The number of outcomes, therefore, is 6^{12}.

3. To fix a good outcome, pick two slots for 1, then pick two slots for 2, etc., with $\binom{12}{2}\binom{10}{2}\cdots\binom{2}{2}$ choices.

The probability, then, is $\frac{\binom{12}{2}\binom{10}{2}\cdots\binom{2}{2}}{6^{12}}$.

What is $\mathbb{P}$(1 appears exactly 3 times, 2 appears exactly 2 times)?

To fix a good outcome now, pick three slots for 1, two slots for 2, and fill the remaining 7 slots by numbers 3, . . . , 6. The number of choices is $\binom{12}{3}\binom{9}{2}4^7$ and the answer is $\frac{\binom{12}{3}\binom{9}{2}4^7}{6^{12}}$.

EXAMPLE 27. We have 14 rooms and 4 colors, white, blue, green, and yellow. Each room is painted at random with one of the four colors. There are 4^{14} equally likely outcomes, so, for example,

$$\mathbb{P}(\text{5 white, 4 blue, 3green, 2 yellow rooms}) = \frac{\binom{14}{5}\binom{9}{4}\binom{5}{3}\binom{2}{2}}{4^{14}}.$$

EXAMPLE 28. A middle row on a plane seats 7 people. Three of them order chicken (C) and the remaining four pasta (P). The flight attendant returns with the meals, but has forgotten who ordered what and discovers that they are all asleep, so she puts the meals in front of them at random. What is the probability that they all receive correct meals?

A reformulation makes the problem clearer: we are interested in

$$\mathbb{P}(\text{3 people who ordered C get C}).$$

Let us label the people $1, ..., 7$ and assume that 1, 2, and 3 ordered C. The outcome is a selection of 3 people from the 7 who receive C, the number of them is $\binom{7}{3}$, and there is a *single* good outcome. So, the answer is $\frac{1}{\binom{7}{3}} = \frac{1}{35}$. Similarly,

$$\mathbb{P}(\text{no one who ordered C gets C}) = \frac{\binom{4}{3}}{\binom{7}{3}} = \frac{4}{35},$$

$$\mathbb{P}(\text{a single person who ordered C gets C}) = \frac{3 \cdot \binom{4}{2}}{\binom{7}{3}} = \frac{18}{35},$$

$$\mathbb{P}(\text{two persons who ordered C get C}) = \frac{\binom{3}{2} \cdot 4}{\binom{7}{3}} = \frac{12}{35}.$$

1.4. Independence

DEFINITION 29. (a) Two events A and B are independent if

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

(b) A (possibly infinite) collection of events

$$\{E_i\}_{i \in I}$$

is an independent collection if for every finite subset J of I, one has

$$\mathbb{P}\left(\bigcap_{i \in J} E_i\right) = \prod_{i \in J} \mathbb{P}(E_i).$$

In this case, the collection

$$\{E_i\}_{i \in I}$$

is often said to be mutually independent.

EXAMPLE 30. One chooses a card at random from a deck of 52 cards. A = the card is a heart, and B = the card is Queen. A natural model for this experiment consists in prescribing the probability

$$\frac{1}{52}$$

for picking any one of the cards. We have

$$\mathbb{P}(A) = \frac{13}{52}, \quad \mathbb{P}(B) = \frac{4}{52}$$

and

$$\mathbb{P}(A \cap B) = \frac{1}{52},$$

hence A and B are independent.

We now give an example of events that are not mutually independent.

EXAMPLE 31. Let

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

Let

$$\mathbb{P}(\{i\}) = \frac{1}{4}, \quad i = 1, 2, 3, 4.$$

Consider $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{1, 4\}$. Then A, B and C are pairwise independent, i.e.,

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{1\}) = \frac{1}{4} = \left(\frac{1}{2}\right)^2 = \mathbb{P}(\{1, 2\})\mathbb{P}(\{1, 3\}) = \mathbb{P}(A)\mathbb{P}(B).$$

Similarly,

$$\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C), \quad \mathbb{P}(C \cap A) = \mathbb{P}(C)\mathbb{P}(A).$$

However, A, B and C are not mutually independent because

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{1\}) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

REMARK 32. The above example shows that a collection of events may not be mutually independent even if they are pairwise independent.

1.5. Conditional Probability

If the event E occurs, then, in order for F to occur, it is necessary that the actual occurrence is a point both in E and in F; that is, it must be in

$$E \cap F.$$

Now, since we know that E has occurred, it follows that E becomes our new, or reduced, sample space; hence, the probability that the event

$$E \cap F$$

occurs given that E occurs will equal the probability of

$$E \cap F$$

relative to the probability of E. That is, we have the following definition.

DEFINITION 33. Let A, B be events, $\mathbb{P}(B) > 0$. The conditional probability of *A* given *B* is

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

EXAMPLE 34. (a) Two fair dice are thrown. Given that the first shows 3, what is the probability that the total exceeds 6? The answer is obviously $1/2$, since the second must show 4, 5, or 6. However, let us labour the point. Clearly $\Omega = \{(i, j) : i, j = 1, 2, 3, 4, 5, 6\}$, and we can take $\mathcal{A}$ to be the set of all subsets of Ω, with $\mathbb{P}(A) = |A|/36$ for any $A \subset \Omega$. Let B be the event that the first dice shows 3, and A be the event that the total exceeds 6. Then

$$B = \{(3, b) : 1 \leq b \leq 6\},\ A = \{(a, b) : a + b > 6\},\ A \cap B = \{(3, 4), (3, 5), (3, 6)\},$$

and

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{|A \cap B|}{|B|} = \frac{3}{6} = \frac{1}{2}.$$

(b) A family has two children. What is the probability that both are boys, given that at least one is a boy? The older and younger child may each be male or female, so there are four possible combinations of sexes, which we assume to be equally likely. Hence we can represent the sample space in the obvious way as

$$\Omega = \{\mathrm{GG}, \mathrm{GB}, \mathrm{BG}, \mathrm{BB}\},$$

where $\mathbb{P}(\mathrm{GG}) = \mathbb{P}(\mathrm{BB}) = \mathbb{P}(\mathrm{GB}) = \mathbb{P}(\mathrm{BG}) = \frac{1}{4}$. From the definition of conditional probability,

$$\begin{aligned}
\mathbb{P}(\mathrm{BB} \mid \text{one boy at least}) &= \mathbb{P}(\mathrm{BB}|\mathrm{GB} \cup \mathrm{BG} \cup \mathrm{BB}) \\
&= \frac{\mathbb{P}(\mathrm{BB} \cap (\mathrm{GB} \cup \mathrm{BG} \cup \mathrm{BB}))}{\mathbb{P}(\mathrm{GB} \cup \mathrm{BG} \cup \mathrm{BB})} \\
&= \frac{\mathbb{P}(\mathrm{BB})}{\mathbb{P}(\mathrm{GB} \cup \mathrm{BG} \cup \mathrm{BB})} = \frac{1}{3}.
\end{aligned}$$

THEOREM 35. *Suppose* $\mathbb{P}(B) > 0$.

(a) A *and* B *are independent if and only if* $\mathbb{P}(A|B) = \mathbb{P}(A)$.

(b) The operation $A \to \mathbb{P}(A|B)$ *from* $\mathcal{A} \to [0, 1]$ *defines a new probability measure on* $\mathcal{A}$*, called the "conditional probability measure given* B*".*

PROOF. (a) It holds that

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

So A *and* B *are independent if and only if* $\mathbb{P}(A|B) = \mathbb{P}(A)$.

For (b), define $\mathbb{Q}(A) = \mathbb{P}(A|B)$, with B fixed. We must show $\mathbb{Q}$ satisfies properties (1) and (2) in the Definition of a probability measure. But

$$\mathbb{Q}(\Omega) = \mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

Therefore, Q satisfies (1). As for (2), note that if $(A_n)_{n\geq 1}$ is a sequence of elements of $\mathcal{A}$ which are pairwise disjoint, then

$$\mathbb{Q}(\cup_{n=1}^{\infty} A_n) = \mathbb{P}(\cup_{n=1}^{\infty} A_n | B) = \frac{\mathbb{P}((\cup_{n=1}^{\infty} A_n) \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\cup_{n=1}^{\infty} (A_n \cap B))}{\mathbb{P}(B)}$$

and also the sequence $(A_n \cap B)_{n\geq 1}$ is pairwise disjoint as well; thus

$$\mathbb{Q}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n=1}^{\infty} \mathbb{P}(A_n | B) = \sum_{n=1}^{\infty} \mathbb{Q}(A_n).$$

□

THEOREM 36. *If $A_1, \ldots, A_n \in \mathcal{A}$ and $\mathbb{P}(A_1 \cap \ldots \cap A_{n-1}) > 0$, then*

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1}).$$

PROOF. We use induction. For $n = 2$, the theorem is simply Definition 33. Suppose the theorem holds for $n-1$ events. Let $B = A_1 \cap \ldots \cap A_{n-1}$. Then by Definition 33, $\mathbb{P}(B \cap A_n) = \mathbb{P}(A_n|B)P(B)$; next we replace $\mathbb{P}(B)$ by its value given in the inductive hypothesis:

$$\mathbb{P}(B) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1) \ldots \mathbb{P}(A_{n-1}|A_1 \cap \ldots \cap A_{n-2}).$$

So

$$\begin{aligned}
\mathbb{P}(A_1 \cap \ldots \cap A_n) &= \mathbb{P}(B \cap A_n) \\
&= \mathbb{P}(A_n|B)\mathbb{P}(B) \\
&= \mathbb{P}(A_n|B)\mathbb{P}(A_1)\mathbb{P}(A_2|A_1) \ldots \mathbb{P}(A_{n-1}|A_1 \cap \ldots \cap A_{n-2}) \\
&= \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1}).
\end{aligned}$$

□

A (finite or countable) collection of disjoint evens $\{E_n\}_n$ such that $\cup_n E_n = \Omega$ is said to be a partition of Ω.

THEOREM 37. *(Law of total probability). Let* $E_1, E_2, \dots, E_n$ *be a partition of* Ω*, and* $\mathbb{P}(E_i) > 0$ *for all* $i \in \{1, \dots, n\}$*. If* $A \in \mathcal{A}$*, then*

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

PROOF. Note that

$$A = A \cap \Omega = A \cap (\cup_{i=1}^{n} E_i) = \cup_{i=1}^{n}(A \cap E_i).$$

Since the E_i are pairwise disjoint so also are $(A \cap E_i)_{i \geq 1}$, hence

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

□

REMARK 38. With a very similar proof, we can easily extend the previous theorem to the case where the partition is countable. That is, suppose $E_1, E_2, \dots, E_n, \dots$ is a partition of Ω, and $\mathbb{P}(E_i) > 0$ for all $i \in \mathbb{N}$. If $A \in \mathcal{A}$, then

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

EXAMPLE 39. We are given two urns, each containing a collection of coloured balls. Urn I contains two white and three blue balls, whilst urn II contains three white and four blue balls. A ball is drawn at random from urn I and put into urn II, and then a ball is picked at random from urn II and examined. What is the probability that it is blue? We assume unless otherwise specified that a ball picked randomly from any urn is equally likely to be any of those present. The reader will be relieved to know that we no longer need to describe $(\Omega, \mathcal{A}, \mathbb{P})$ in detail; we are confident that we could do so if necessary. Clearly, the colour of the final ball depends on the colour of the ball picked from urn I. So let us 'condition' on this. Let A be the event that the final ball is blue, and let B be the event that the first one picked was blue. Then, by using Theorem 37 for the partition

$\Omega = B \cup B^c$, we get

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c).$$

We can easily find all these probabilities:

$$\mathbb{P}(A|B) = \mathbb{P}(A|\text{ urn II contains three white and five blue balls}) = \frac{5}{8},$$

$$\mathbb{P}(A|B^c) = \mathbb{P}(A|\text{ urn II contains four white and four blue balls}) = \frac{1}{2},$$

$$\mathbb{P}(B) = \frac{3}{5},\ \mathbb{P}(B^c) = \frac{2}{5}.$$

Hence

$$\mathbb{P}(A) = \frac{5}{8} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{2}{5} = \frac{23}{40}.$$

THEOREM 40. *(Bayes' Theorem) Let $E_1, E_2, \dots, E_n$ be a partition of Ω with $\mathbb{P}(E_i) > 0$ for all i, and suppose $\mathbb{P}(A) > 0$. Then for each m,*

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

PROOF. It holds

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

□

EXAMPLE 41. In a certain assembly plant, three machines, I, II and III, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Suppose that a product is randomly selected and it is defective, then what is the conditional probability the defective product is made by machine III?

SOLUTION. Let A be the event of the selected product is defective, E_i the product made by the i-th machine.

$$\mathbb{P}(E_3|A) = \frac{\mathbb{P}(A|E_3)\mathbb{P}(E_3)}{\sum_{n=1}^{3}\mathbb{P}(A|E_n)\mathbb{P}(E_n)} = \frac{0.25(0.02)}{0.3(0.02)+0.45(0.03)+0.25(0.02)} = 0.204.$$