1. Probability Notes

In this course, we are going to study core concepts of Probability, which shall help us in analyzing 'randomness'/uncertainty in real world situations.

Example 1.1 (Toss a coin)**.** If we toss a coin, then either a head or a tail will appear. For simplicity, we do not consider the unlikely event in which the coin lands on its edge. Here the results/outcomes are non-numerical.

Example 1.2 (Throw a die)**.** If we throw/roll a standard six-sided die and observe the number that appears on the top, then we would have one of the numbers $1, 2, 3, 4, 5, 6$ as the result/outcome. Here, the outcome is numerical and the values are in $\{1, 2, 3, 4, 5, 6\}$.

Example 1.3 (Lifetime of an electric bulb)**.** Switch on a new electric bulb and wait till the time it fails. The duration in which the bulb was working gives us the lifetime of the bulb. The result/outcome is some real number in $[0, \infty)$.

Remark 1.4 (What is Probability?). Probability theory is a branch of Pure Mathematics, and deals with objects involving 'randomness'/uncertainty. As in Pure Mathematics, certain Axioms/hypotheses shall be assumed and results shall be derived from these assumptions. However, it turns out that in many real world situations we may take appropriate models in probability and it will replicate the intrinsic features from the real world – in this sense, Probability may also be considered as 'applicable'. In Example 1.3, probability may represent the law according to which the lifetimes vary across multiple electric bulbs.

Remark 1.5 (What is Statistics?). In Statistics we are faced with data/sample from an underlying population (for example, consider the lifetimes of 5 electric bulbs from a batch of 100 bulbs in Example 1.3). These data/sample typically consists of measurements in an experiment, responses in a survey etc.. We would like to make various kinds of inferences (involving characteristics) about the underlying population from the data/sample provided. We are also interested in the procedures through which such an analysis may be done and the effectiveness of such procedures. These topics are not part of this course.

Note 1.6. A reasonable approach in studying any new random phenomena is to perform experiments under controlled situations, by repeating the phenomena under identical conditions. After a sufficient number of repetitions, we may have some idea about outcomes/'events' which are more likely to occur than other such outcomes/events. We must note, however, that each experiment terminates in an outcome, which cannot be specified in advance, i.e. before performing the experiment.

Definition 1.7 (Random experiment)**.** A random experiment is an experiment in which

(a) all possible outcomes of the experiment are known in advance,
(b) outcome of a particular trial/performance of the experiment cannot be specified in advance,
(c) the experiment can be repeated under identical conditions.

Notation 1.8. A random experiment shall be denoted by $\mathcal{E}$.

Definition 1.9 (Sample Space)**.** The collection of all possible outcomes of a random experiment $\mathcal{E}$ is called its sample space.

Notation 1.10. A sample space shall be denoted by Ω. It is a set containing all possible outcomes.

Example 1.11 (Examples of Random experiments and corresponding Sample spaces)**.** The experiments mentioned in Examples 1.1, 1.2 and 1.3 are all examples of random experiments. The corresponding sample spaces are $\{H, T\}, \{1, 2, 3, 4, 5, 6\}$ and $[0, \infty)$ respectively. Here, H and T denotes a head and a tail respectively.

Example 1.12 (Tossing two coins simultaneously)**.** If we write the result/outcome from the first coin as x and the second coin as y, then the result of the experiment may be written as an ordered pair (x, y). Here, x is either a head or a tail. Similarly, y is either a head or a tail. The sample space is therefore,

$$\Omega = \{(x, y) : x, y \in \{H, T\}\} = \{(H, H), (H, T), (T, H), (T, T)\}.$$

Notation 1.13. Given two sets A and B, we write $A \times B := \{(x, y) : x \in A, y \in B\}$. In Example 1.12, $\Omega = \{H, T\} \times \{H, T\}$. The set $\mathbb{R}^2$ is nothing but $\mathbb{R} \times \mathbb{R}$.

Example 1.14 (Throwing a die three times)**.** In this case, we record the outcome of all three throws taken together. If x, y and z represent the result/outcome of the first, second and the third throws respectively, then the outcome may be represented as the ordered triple (x, y, z). The sample space is therefore

$$\Omega = \{(x, y, z) : x, y, z \in \{1, 2, 3, 4, 5, 6\}\}.$$

Note 1.15. We are interested in specific outcomes or more generally, specific subsets of the sample space Ω, which are more likely to appear than other such subsets. In the case where we deal with specific outcomes, we shall consider them as singleton subsets of Ω.

Definition 1.16 (Events)**.** If the outcome of a random experiment $\mathcal{E}$ is an element of a subset E of Ω, then we say that the event E has occured.

Notation 1.17. As mentioned in the previous definition, we are interested in specific subsets of Ω, to be referred to as events. The collection of all events shall be denoted as $\mathcal{F}$.

Note 1.18. The empty set $\emptyset$ and the sample space Ω will always be an element in $\mathcal{F}$.

Remark 1.19. In many situations, we shall take the event space $\mathcal{F}$ as the power set 2^Ω of Ω. Recall that the power set of Ω is the collection of all subsets of Ω. Later on, we shall discuss specific situations in which we may restrict our attention to a smaller collection than 2^Ω.

Notation 1.20. We may refer to a collection of sets as a class of sets. The event space $\mathcal{F}$ is a class of subsets of the sample space Ω.

Example 1.21 (Examples of Events)**.**

(a) $\{H\}$ and $\{T\}$ are events in Example 1.1. The event space $\mathcal{F}$ may be taken as $\mathcal{F} = 2^\Omega = \{\emptyset, \{H\}, \{T\}, \Omega\}$ with $\Omega = \{H, T\}$.

(b) $\{4\}, [5, \infty), [2, 3], [1, 100)$ are events in Example 1.3.

(c) $\{(1, 4, 5), (2, 2, 2), (3, 6, 2)\}$ is an event in Example 1.14.

Remark 1.22. Observe that complementation of an event E gives us the subset E^c. The set E^c may be interpretated as the non-occurrence of the event E. Thus, we treat E^c as another event.

Similarly, finite or countably infinite unions and intersections of events give us further events. Therefore, we can consider standard set theoretic operations, viz. complementation, finite and countably infinite unions and intersections on the event space $\mathcal{F}$.

Note 1.23. For technical reasons, we do not consider uncountable unions or intersections of events.

Remark 1.24. As mentioned earlier in Note 1.15, we would like to identify special subsets or events which are more likely to occur than the others. This is where Probability enters the discussion. Probability is a measure of uncertainty and we are interested in associating numerical quantities to events/outcomes thereby quantifying the uncertainty related to these events/outcomes. This is achieved by assigning probabilities to the events.

Definition 1.25 (A priori or Classical definition of probability)**.** Suppose that a random experiment results in n (a finite number) outcomes. Given an event $A \in \mathcal{F}$, if it appears in m $(0 \leq m \leq n)$ outcomes, then the probability of A is $\frac{m}{n}$.

Note 1.26. Given a random experiment, we are already aware of all possible outcomes. Therefore, without performing the experiment, we can discuss about the a priori definition of probability.

Note 1.27. The classical definition works only when there are finitely many outcomes. Due to the limitations of this definition, we look for other ways to understand the notion of probability.

Remark 1.28 (A posteriori or Relative Frequency definition of probability)*.* If a random experiment $\mathcal{E}$ is repeated a large number, say n, of times and an event A occurs m many times, then the relative frequency $\frac{m}{n}$ may be taken as an approximate value of the probability of A. This concept of probability assumes that in the long run there is a regularity of occurence of the event.

Note 1.29. The a posteriori definition of probability works only after performing the random experiment.

Note 1.30. We now discuss an axiomatic definition of probability. We shall recover the classical definition as part of the axiomatic definition and also justify the relative frequency definition of probability.

Note 1.31. At this moment, we do not focus on how the probabilities of events are assigned, i.e. how a probability model is developed. Our interest is in the properties of probability as a measure of uncertainty/'randomness'.

Definition 1.32 (Set function)**.** A set function is a function whose domain is a collection/class of sets.

Definition 1.33 (Probability function/measure)**.** Suppose that Ω and $\mathcal{F}$ are the sample space and the event space of a random experiment $\mathcal{E}$ respectively. A real valued set function $\mathbb{P}$, defined on the event space $\mathcal{F}$, is said to be a probability function/measure if it satisfies the following axioms/properties, viz.

(a) $\mathbb{P}(\Omega) = 1$.
(b) (non-negativity) $\mathbb{P}(E) \geq 0$ for any event E in $\mathcal{F}$.
(c) (Countable additivity) If $\{E_n\}_n$ is a sequence of events in $\mathcal{F}$ such that $E_i \cap E_j = \emptyset, \forall i \neq j$, then $\mathbb{P}(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mathbb{P}(E_n)$.

Definition 1.34 (Probability space)**.** If $\mathbb{P}$ is a probability function defined on the event space $\mathcal{F}$ of a random experiment $\mathcal{E}$, then the triple $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a probability space. Here, Ω denotes the sample space of $\mathcal{E}$.

Definition 1.35 (Mutually Exclusive or Pairwise disjoint events)**.** Let $\mathcal{I}$ be an indexing set. A collection of events $\{E_i : i \in \mathcal{I}\}$ is said to be mutually exclusive or pairwise disjoint if $E_i \cap E_j = \emptyset, \forall i \neq j$.

Note 1.36. We first study some basic properties of probability functions and then look at some explicit examples.

Proposition 1.37. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space associated with a random experiment $\mathcal{E}$.*

(a) $\mathbb{P}(\emptyset) = 0$.

Proof. Take the sequence of events $\{E_n\}_n$ in $\mathcal{F}$ given by $E_1 := \Omega$ and $E_n = \emptyset, \forall n \geq 2$. Then $\bigcup_n E_n = \Omega$ and the E_n's are pairwise disjoint. By Definition 1.33,

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\bigcup_n E_n) = \sum_{n=1}^{\infty} \mathbb{P}(E_n) = \mathbb{P}(E_1) + \sum_{n=2}^{\infty} \mathbb{P}(E_n) = 1 + \sum_{n=2}^{\infty} \mathbb{P}(E_n).$$

Therefore,

$$0 = \sum_{n=2}^{\infty} \mathbb{P}(E_n) = \lim_{m\to\infty} \sum_{n=2}^{m} \mathbb{P}(E_n) = \lim_{m\to\infty} [(m-1)\mathbb{P}(\emptyset)].$$

The result follows. $\square$

(b) *(Finite additivity) Let $E_1, E_2, \cdots, E_n \in \mathcal{F}$ for some integer $n \geq 2$ be mutually exclusive events. Then $\mathbb{P}(\bigcup_{i=1}^n E_i) = \sum_{i=1}^n \mathbb{P}(E_i)$.*

Proof. Consider the events $E_i = \emptyset, \forall i > n$. Then $\mathbb{P}(E_i) = 0, \forall i > n$ and the sequence of events $\{E_m\}_m$ is mutually exclusive. Now,

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

$\square$

(c) *$\mathbb{P}(E) + \mathbb{P}(E^c) = 1$ for all events $E \in \mathcal{F}$.*

Proof. Note that $E \cap E^c = \emptyset$, i.e. the events E and E^c are mutually exclusive. Then by finite additivity, $\mathbb{P}(E) + \mathbb{P}(E^c) = \mathbb{P}(E \bigcup E^c) = \mathbb{P}(\Omega) = 1$. $\square$

(d) *$0 \leq \mathbb{P}(E) \leq 1$ for all events E in $\mathcal{F}$.*

Proof. The inequality $\mathbb{P}(E) \geq 0$ follows from the definition. Again $\mathbb{P}(E^c) \geq 0$. Using $\mathbb{P}(E) + \mathbb{P}(E^c) = 1$, we have $\mathbb{P}(E) \leq \mathbb{P}(E) + \mathbb{P}(E^c) = 1$. $\square$

(e) *(Monotonicity) Suppose $A, B \in \mathcal{F}$ with $A \subseteq B$. Then $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B)$. In particular, $\mathbb{P}(A) \leq \mathbb{P}(B)$. If, in addition $\mathbb{P}(B) = 0$, then $\mathbb{P}(A) = 0$.*

Proof. Observe that the sets A and $A^c \cap B$ are mutually exclusive and that $B = A \bigcup (A^c \cap B)$. By finite additivity, we have $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B)$. As $\mathbb{P}(A^c \cap B) \geq 0$, hence $\mathbb{P}(A) \leq \mathbb{P}(B)$.

If $\mathbb{P}(B) = 0$, then $0 \leq \mathbb{P}(A) \leq \mathbb{P}(B) = 0$. Hence, $\mathbb{P}(A) = 0$. $\square$

(f) *(Inclusion-Exclusion principle for two events) For $A, B \in \mathcal{F}$, we have*

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

Proof. Observe that $A = (A \cap B) \bigcup (A \cap B^c)$ and the events $A \cap B, A \cap B^c$ are mutually exclusive. Then

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

Similarly, $B = (A \cap B) \bigcup (A^c \cap B)$ and hence

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

Then,

$$\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \cap B) + [\mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)].$$

Observe that the events $A \cap B^c, A \cap B, A^c \cap B$ are mutually exclusive and

$$(A \cap B^c) \bigcup (A \cap B) \bigcup (A^c \cap B) = A \bigcup B.$$

Then $\mathbb{P}(A \bigcup B) = \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$ and hence

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

The result follows. $\square$

(g) *(Boole's inequality for two events) For $A, B \in \mathcal{F}$, we have $\mathbb{P}(A \bigcup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.*

Proof. Since $\mathbb{P}(A \cap B) \geq 0$, using the Inclusion-Exclusion principle, we have $\mathbb{P}(A \bigcup B) \leq \mathbb{P}(A \cap B) + \mathbb{P}(A \bigcup B) = \mathbb{P}(A) + \mathbb{P}(B)$. $\square$

(h) *(Bonferroni's inequality for two events) For $A, B \in \mathcal{F}$, we have $\mathbb{P}(A \cap B) \geq \max\{0, \mathbb{P}(A) + \mathbb{P}(B) - 1\}$.*

Proof. By definition, we have $\mathbb{P}(A \cap B) \geq 0$. Again, using $\mathbb{P}(A \bigcup B) \leq 1$ and the Inclusion-Exclusion principle, we have $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \bigcup B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$. The result follows. $\square$

Example 1.38 (Probability space associated with a coin toss)**.** Recall from Example 1.21 that in the random experiment of tossing a coin, we have the sample space $\Omega = \{H, T\}$ and the event space $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}$. If $\mathbb{P}$ is a probability function defined on $\mathcal{F}$, then we have $\mathbb{P}(\Omega) = 1, \mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\{H\}) + \mathbb{P}(\{T\}) = 1$. The last relation follows from the observation that $\{T\} = \{H\}^c$. If $\mathbb{P}(\{H\}) = p \in [0, 1]$, then $\mathbb{P}(\{T\}) = 1 - p$. These are necessary conditions derived from the axioms/properties. Now we can ask the following: given a function $\mathbb{P}$ on $\mathcal{F}$ defined by

$$\mathbb{P}(\emptyset) := 0,\ \mathbb{P}(\{H\}) := p,\ \mathbb{P}(\{T\}) := 1 - p,\ \mathbb{P}(\Omega) := 1$$

is $\mathbb{P}$ a probability function for any $p \in [0, 1]$? If you have a fair coin, you would expect that the probability of occurrence of a head and a tail should be the same – in which case we have $p = 1 - p$, i.e. $p = \frac{1}{2}$.

Note 1.39. In the next week, we shall see further examples.

Note 1.40. In the examples discussed in the previous week, we have the corresponding sample spaces are either finite or uncountably infinite.

Example 1.41 (Throw/Roll a die until 6 appears)**.** Suppose we take a standard six-sided die and count the number of rolls required to obtain the first 6. In this case, our sample space is $\Omega = \{1, 2, \cdots\}$, which is countably infinite.

Remark 1.42. If $(\Omega, \mathcal{F} = 2^\Omega, \mathbb{P})$ is a probability space, with Ω being a finite or a countably infinite set, then all its subsets $A \in \mathcal{F}$ are also finite or countably infinite. By finite/countable additivity of $\mathbb{P}$, we have

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}), \forall A \in \mathcal{F}.$$

Note 1.43. In the discussion below, we consider a set Ω, assumed to be a finite or a countably infinite set and discuss structural properties of probability spaces on Ω. The observations here are then applicable to situations where we have a random experiment with only finitely many or countably infinite many outcomes, i.e. the sample space is finite or countably infinite. We are going to see that specifying the probability of singleton events can describe the probability function/measure on the event space $\mathcal{F}$.

Let Ω be any finite or countably infinite set. Consider $\mathcal{F} = 2^\Omega$ the power set. Let $p : \Omega \to [0,1]$ be a function such that

$$\sum_{\omega \in \Omega} p_\omega = 1.$$

Now consider the real valued set function $\mathbb{P}$ on $\mathcal{F}$ defined by

$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega, \forall A \in \mathcal{F}.$$

Note 1.44. Observe that $\mathbb{P}(\{\omega\}) = p_\omega, \forall \omega \in \Omega$.

Proposition 1.45. *The set function $\mathbb{P}$, as defined above, is a probability function/measure on $\mathcal{F}$.*

Proof. We verify the axioms in Definition 1.33. By definition, $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} p_\omega = 1$.

Since, $p_\omega \geq 0, \forall \omega \in \Omega$, we have $\mathbb{P}(A) = \sum_{\omega \in A} p_\omega \geq 0, \forall A \in \mathcal{F}$.

Let $\{A_n\}_n$ be a sequence of mutually exclusive events. Then each element of $\bigcup_n A_n$ belongs to exactly one A_n. Then,

$$\sum_{n=1}^\infty \mathbb{P}(A_n) = \sum_{n=1}^\infty \sum_{\omega \in A_n} p_\omega = \sum_{\omega \in \bigcup_{n=1}^\infty A_n} p_\omega = \mathbb{P}(\bigcup_{n=1}^\infty A_n).$$

This completes the proof. $\square$

Definition 1.46 (Discrete Probability spaces)**.** Let Ω be a finite or countable set. We refer to a probability space of the form $(\Omega, 2^\Omega, \mathbb{P})$ as a discrete probability space.

Remark 1.47. A Probability function/measure on a discrete probability space is determined by the probability of singleton events.

Notation 1.48. We may refer to the singleton events in a discrete probability space as elementary events.

Example 1.49 (Examples of discrete probability spaces)**.** The following are some examples of discrete probability spaces. Here we only specify the probability of singleton sets/events.

(a) $\Omega = \{H, T\}$ with $\mathbb{P}(\{H\}) = p, \mathbb{P}(\{T\}) = 1 - p$ for some fixed $p \in [0,1]$.

(b) $\Omega = \{1,2,3,4,5,6\}$ with $\mathbb{P}(\{i\}) = \frac{1}{6}, \forall i \in \Omega$.

(c) Let Ω be the set of all natural numbers, i.e. $\Omega = \{1, 2, 3, \cdots\}$. Take $\mathbb{P}(\{n\}) = \frac{1}{2^n}, \forall n \in \Omega$.

Remark 1.50 (Equally likely probability models on finite sample spaces). Let $\mathcal{E}$ be a random experiment with the sample space $\Omega = \{\omega_1, \cdots, \omega_k\}$, a finite set with k elements. Here, any probability function/measure $\mathbb{P}$ on $\mathcal{F} = 2^\Omega$ is determined by the values $p_{w_i} = \mathbb{P}(\{\omega_i\}), i = 1, \cdots, k$. Assume that the elementary events $\{\omega_i\}$ are equally likely, i.e. $p_{w_i} = \mathbb{P}(\{\omega_i\}) = p_{w_j} = \mathbb{P}(\{\omega_j\}), \forall i \neq j$. Since, $\sum_{\omega \in \Omega} p_\omega = 1$, we have $p_{w_i} = \mathbb{P}(\{\omega_i\}) = \frac{1}{k}, \forall i = 1, \cdots, k$. For any set/event $A \in \mathcal{F}$, we have

$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega = \frac{\#A}{k},$$

where $\#A$ denotes the cardinality of A, i.e. the number of elements in A. We can rewrite the above observation in terms of the following interpretation.

$$\mathbb{P}(A) = \frac{\text{number of ways favourable to the event } A}{\text{number of ways in which the random experiment can terminate}}.$$

Definition 1.51 (At random)**.** Let $\mathcal{E}$ be a random experiment with finite sample space. We say that the experiment has been performed at random to imply that all the elementary/singleton events are equally likely. Identifying singleton events with the corresponding outcomes, we may also say that the outcomes are equally likely. In this case, the number of ways in which the random experiment can terminate is exactly the cardinality of the sample space.

Note 1.52. While tossing a coin or rolling a die, if the outcomes are equally likely, then we say that the coin/die is 'fair'.

Example 1.53. Example 1.49(b) has been performed at random.

Example 1.54. A box contains 3 red balls and 2 green balls. Balls of the same colour are assumed to be identical. Draw a ball at random. If A denotes the event that the ball drawn is red, then $\mathbb{P}(A) = \frac{3}{5}$.

Note 1.55. Consider a random experiment $\mathcal{E}$ with the sample space Ω being the set of all natural numbers, i.e. $\Omega = \{1, 2, 3, \cdots\}$. Then for any probability function/measure $\mathbb{P}$ on $\mathcal{F} = 2^\Omega$, we have $1 = \mathbb{P}(\Omega) = \sum_{n=1}^{\infty} p_n, \forall A \in \mathcal{F}$. Consequently, $\lim_n p_n = 0$ and all p_n's cannot be equal. Hence we cannot have natural numbers drawn at random. By a similar argument, we cannot have any random experiment performed at random if the sample space is countably infinite.

Remark 1.56. When multiple draws from a box are involved in a single trial of a random experiment, then there are two broad categories of problems, viz. sampling with replacement and sampling without replacement. In the first case, the outcome of each draw is returned to the box before the next draw. In the second case, the outcome is removed from the possibilities in the next draw. Following examples illustrate these concepts.

Example 1.57. Example 1.14 where we throw/roll a die thrice is an example of sampling with replacement. The cardinality of the sample space is $6 \times 6 \times 6 = 6^3$. If A denote the event that all the rolls result in an even number, then number of ways favourable to A is $3 \times 3 \times 3 = 3^3$. Thus, $\mathbb{P}(A) = \frac{3^3}{6^3} = \frac{1}{8}$.

Example 1.58. Draw 2 cards at random from a standard deck of 52 cards. Here, the cardinality of the sample space is $\binom{52}{2}$. Since, we are looking at the 2 cards in hand together, the order in which they have been obtained does not matter. Consider the event A that both cards are from the Club ($\clubsuit$) suit. Since a standard deck of cards contain 13 cards from the Club suit, we have $\mathbb{P}(A) = \binom{13}{2}/\binom{52}{2}$. This is an example of sampling without replacement.

Example 1.59 (Placing r balls in m bins)**.** Fix two positive integers r and m. Suppose that there are r labelled balls and m labelled bins/boxes/urns. Assume that each bin can hold all the balls, if required. One by one, we put the balls into the bins 'at random'. Then, by letting ω_i be the bin-number into which the i-th ball is placed, we can capture the full configuration by the vector $\underline{\omega} = (\omega_1, \ldots, \omega_r)$. Let Ω be the list of all configurations. Therefore, Ω is the sample space of this random experiment. We have

$$\Omega = \{\underline{\omega} : \underline{\omega} = (\omega_1, \ldots, \omega_r) \text{ with } 1 \leq \omega_i \leq m \text{ for each } 1 \leq i \leq r\}.$$

The cardinality of Ω is m^r (since each ball may be placed in one of the m bins). Since the experiment has been performed at random, we have $\mathbb{P}(\{\underline{\omega}\}) = p_{\underline{\omega}} = m^{-r}, \forall \underline{\omega} \in \Omega$. We now consider the probabilities of the following events.

(a) Let A be the event that the r-th ball is placed in the first bin. Then $A = \{\underline{\omega} \in \Omega : \omega_r = 1\}$. Here, balls numbered 1 to $r-1$ can be placed in any of the m bins. Therefore, the number of outcomes $\underline{\omega}$ favourable to A is m^{r-1}. Hence, $\mathbb{P}(A) = \frac{m^{r-1}}{m^r} = \frac{1}{m}$.

(b) Let B be the event that the first bin is empty. Then $B = \{\underline{\omega} \in \Omega : \omega_i \neq 1, \forall i = 1, 2, \cdots, r\}$. Here, each ball can be placed in any of the remaining bins numbered 2 to m. Since there are $m-1$ choices for each ball, the number of outcomes $\underline{\omega}$ favourable to B is $(m-1)^r$. Hence $\mathbb{P}(B) = \frac{(m-1)^r}{m^r}$.

(c) Consider $r \leq m$ and let C be the event that all the balls are placed in distinct bins, i.e. no bins contain more than one ball (a bin may remain empty). Then, $C = \{\underline{\omega} \in \Omega : \omega_i \neq \omega_j, \forall i \neq j\}$. Here, we are choosing/sampling bins for each ball and the sampling is being done without replacement. Hence, the number of outcomes $\underline{\omega}$ favourable to C is ${}^mP_r$. Hence $\mathbb{P}(C) = {}^mP_r \, m^{-r} = \frac{m(m-1)\cdots(m-r+1)}{m^r} = \frac{(m-1)\cdots(m-r+1)}{m^{r-1}}$.

Example 1.60 (Birthday Paradox)**.** There are n people at a party. What is the chance that two of them have the same birthday? Assume that none of them was born on a leap year and that days are equally likely to be a birthday of a person. The problem structure remains the same as in the previous balls in bin problem, where the bins are labelled as $1, 2, \ldots, 365$ (days of the year), and the balls are labelled as $1, 2, \ldots, n$ (people). In the notations of the previous example, $r = n$ and $m = 365$. Here, we wish to find the probability of the following event

$$D = \{\underline{\omega} \in \Omega : \omega_i = \omega_j, \text{for some } i \neq j\}.$$

Note that $D = C^c$, where C is as in the previous example. Therefore, $\mathbb{P}(D) = 1 - \mathbb{P}(C) = 1 - \frac{(365-1)\cdots(365-n+1)}{365^{n-1}}$. The reason this is called a 'paradox' is that even for n much smaller than 365, the probability becomes significantly large. For example, $n = 25$ gives $\mathbb{P}(D) > 0.5$.

We now discuss a generalization of the Inclusion-Exclusion principle for two events discussed in Proposition 1.37.

Proposition 1.61 (Inclusion Exclusion Principle). *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* $A_1, \ldots, A_n$ *be events. Then,*

$$\mathbb{P}(\bigcup_{i=1}^{n} A_i) = S_{1,n} - S_{2,n} + S_{3,n} - \cdots + (-1)^{n-1} S_{n,n},$$

where

$$S_{k,n} := \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}).$$

Proof. For the case $n = 2$, the result has already been discussed in Proposition 1.37. We prove the result for general n by an application of the principle of Mathematical Induction.

Suppose the result is true for $n = 2, 3, \cdots, k$. We want to establish the result for $n = k + 1$.

Using the result for $n = 2$, we have

$$\mathbb{P}(\bigcup_{i=1}^{k+1} A_i) = \mathbb{P}((\bigcup_{i=1}^{k} A_i) \cup A_{k+1}) = \mathbb{P}(\bigcup_{i=1}^{k} A_i) + \mathbb{P}(A_{k+1}) - \mathbb{P}((\bigcup_{i=1}^{k} A_i) \cap A_{k+1}).$$

Consider

$$T_{j,k} := \sum_{1 \le i_1 < i_2 < \ldots < i_j \le k} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_j} \cap A_{k+1}), \ j = 1, 2, \cdots, k.$$

Applying the result for $n = k$ on the set $\bigcup_{i=1}^{k}(A_i \cap A_{k+1})$, we have

$$\mathbb{P}((\bigcup_{i=1}^{k} A_i) \cap A_{k+1}) = \sum_{j=1}^{k} (-1)^{j-1} T_{j,k}.$$

Then,

$$\begin{aligned}
&\mathbb{P}(\bigcup_{i=1}^{k+1} A_i) \\
&= (S_{1,k} + \mathbb{P}(A_{k+1})) - (S_{2,k} + T_{1,k}) + (S_{3,k} + T_{2,k}) \cdots + (-1)^{k-1}(S_{k,k} + T_{k-1,k}) + (-1)^{(k+1)-1} T_{k,k} \\
&= \sum_{j=1}^{k+1} (-1)^{j-1} S_{j,k+1}.
\end{aligned}$$

Hence the result is true for the case $n = k+1$. Applying the principle of Mathematical Induction, the result is true for any positive integer n. $\square$

Example 1.62. Consider placing r labelled balls in m labelled bins at random (Example 1.59). Let E denote the event that at least one bin is empty. Now, for $j = 1, 2, \cdots, m$, consider the event E_j that none of the balls are placed in the j-th bin, i.e. j-th bin is empty. Then

$$E_j = \{\underline{\omega} \in \Omega : \omega_i \neq j, \forall i = 1, 2, \cdots, r\}, \forall j = 1, 2, \cdots, m$$

and $E = \bigcup_{j=1}^m E_j$. Not all the bins can be empty and hence $\bigcap_{j=1}^m E_j = \emptyset$. For $1 \leq k \leq m - 1$, $E_{j_1} \cap E_{j_2} \cap \cdots \cap E_{j_k}$ denotes the event that the bins numbered $j_1 < j_2 < \cdots < j_k$ are empty. Here, each ball can be placed in the remaining $m - k$ bins. Therefore,

$$\mathbb{P}(E_{j_1} \cap E_{j_2} \cap \cdots \cap E_{j_k}) = \frac{(m-k)^r}{m^r}.$$

By the Inclusion-Exclusion Principle,

$$\mathbb{P}(E) = \mathbb{P}\left(\bigcup_{j=1}^m E_j\right) = \sum_{k=1}^{m-1} (-1)^{k-1} \binom{m}{k} \frac{(m-k)^r}{m^r}.$$

Remark 1.63 (Bonferroni's inequality). In the notations of Proposition 1.61, it can be shown that $S_{1,n} - S_{2,n} \leq \mathbb{P}(\bigcup_{i=1}^n A_i) \leq S_{1,n}$. More generally,

$$\mathbb{P}(\bigcup_{i=1}^n A_i) \leq S_{1,n} - S_{2,n} + \cdots + S_{m,n} \quad \text{if } m \text{ is odd,}$$
$$\mathbb{P}(\bigcup_{i=1}^n A_i) \geq S_{1,n} - S_{2,n} + \cdots - S_{m,n} \quad \text{if } m \text{ is even.}$$

We do not discuss the proof. These inequalities are sometimes referred to as Bonferroni's inequalities in the literature.

Note 1.64. During the performance of a random experiment, if an event A is observed, then it may also provide some information regarding other events. The next concept attempts to formalize this information.