2. Week 2

Example 2.1. 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 2.2. 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 2.3. 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 2.4. 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 2.5. 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 2.6 (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

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

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 $r \le m$ 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 2.7 (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 2.8 (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

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

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

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

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,

$$\mathbb{P}(\bigcup_{i=1}^{k+1} A_i)$$

$$= (S_{1,k} + \mathbb{P}(A_{k+1})) - (S_{2,k} + T_{1,k}) + (S_{3,k} + T_{2,k}) \cdots + (-1)^{k-1}(S_{k,k} + T_{k-1,k}) + (-1)^{(k+1)-1}T_{k,k}$$
$$= \sum_{j=1}^{k+1} (-1)^{j-1} S_{j,k+1}.$$

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. □

Example 2.9. Consider placing r labelled balls in m labelled bins at random (Example 2.6). 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 2.10 (Bonferroni's inequality). In the notations of Proposition 2.8, 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 2.11. 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.

Definition 2.12 (Conditional Probability)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let A be an event with $\mathbb{P}(A) > 0$. For any event B, we define

$$\mathbb{P}(B \mid A) := \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$$

to be the conditional probability of B given A.

Example 2.13. Consider rolling a fair die twice. The sample space is $\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}$ with $\mathbb{P}((i, j)) = \frac{1}{36}, \forall (i, j) \in \Omega$. Consider the events $A = \{(i, j) : i \text{ is odd}\}$ and $B = \{(i, j) : i + j = 3\} = \{(1, 2), (2, 1)\}$. Here, number of outcomes favourable to A is $3 \times 6 = 18$ and hence $\mathbb{P}(A) = \frac{18}{36} = \frac{1}{2} > 0$. The event $A \cap B = \{(1, 2)\}$ and hence $\mathbb{P}(A \cap B) = \frac{1}{36}$. Then $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{1}{18}$.

Note 2.14. We note some basic properties of conditional probability. If $\mathbb{P}(A) > 0$, then

(a) $\mathbb{P}(\Omega \mid A) = \frac{\mathbb{P}(\Omega \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1$.
(b) $\mathbb{P}(A \mid A) = 1$
(c) $B \mapsto \mathbb{P}(B \mid A)$ defines a non-negative set function on $\mathcal{F}$.

Proposition 2.15. $(\Omega, \mathcal{F}, \mathbb{P}(\cdot \mid A))$ *is a probability space.*

Proof. We verify the axioms in Definition 1.33. We have already observed $\mathbb{P}(\Omega \mid A) = 1$ and non-negativity in the above note.

To establish the countable additivity. If $\{E_n\}_n$ is a sequence of mutually exclusive events, then so are $\{E_n \cap A\}_n$. By the countable additivity of $\mathbb{P}$, we have

$$\mathbb{P}(\bigcup_{n=1}^{\infty} E_n \mid A) = \frac{\mathbb{P}(\bigcup_{n=1}^{\infty} E_n \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(\bigcup_{n=1}^{\infty} (E_n \cap A))}{\mathbb{P}(A)} = \sum_{n=1}^{\infty} \frac{\mathbb{P}(E_n \cap A)}{\mathbb{P}(A)} = \sum_{n=1}^{\infty} \mathbb{P}(E_n \mid A).$$

This completes the proof. $\square$

Proposition 2.16 (Multiplication Rule)**.** *Let $E_1, E_2, \cdots, E_n$ be events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) > 0$. Then*

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) = \mathbb{P}(E_1)\ \mathbb{P}(E_2 \mid E_1)\ \mathbb{P}(E_3 \mid E_1 \cap E_2) \cdots\ \mathbb{P}(E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}).$$

Proof. Note that $E_1 \cap E_2 \cap \cdots \cap E_n \subseteq E_1 \cap E_2 \cap \cdots \cap E_{n-1} \subseteq E_1 \cap E_2 \subseteq E_1$ and hence by the hypothesis, all the conditional probabilities in the statement are well-defined.

For the case $n = 2$, the result follows from the definition of $\mathbb{P}(E_2 \mid E_1)$. For general n, apply result for two events repeatedly in the following manner:

$$\begin{aligned}
&\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) \\
&= \mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_{n-1})\ \mathbb{P}(E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}) \\
&= \mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_{n-2})\ \mathbb{P}(E_{n-1} \mid E_1 \cap E_2 \cap \cdots \cap E_{n-2})\ \mathbb{P}(E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}) \\
&= \cdots \\
&= \mathbb{P}(E_1)\ \mathbb{P}(E_2 \mid E_1)\ \mathbb{P}(E_3 \mid E_1 \cap E_2) \cdots\ \mathbb{P}(E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}).
\end{aligned}$$

This completes the proof. □

Example 2.17. Suppose that an urn contains 3 red balls and 5 green balls. All balls of the same colour are identical. Suppose 2 balls are drawn successively at random from the urn without replacement. Let A and B denote the events that the first ball is red and second ball is green, respectively. Then $\mathbb{P}(A) = \frac{3}{8}$. If the event A has already happened, then at the time of the second draw the urn contains 2 red balls and 5 green balls. As such $\mathbb{P}(B \mid A) = \frac{5}{7}$. By the Multiplication rule, the probability that the first ball drawn is red and the second ball drawn is green is $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B \mid A) = \frac{15}{56}$.

Definition 2.18 (Exhaustive events)**.** Let $\mathcal{I}$ be an indexing set. A collection of events $\{E_i : i \in \mathcal{I}\}$ is said to be exhaustive if $\cup_{i \in \mathcal{I}} E_i = \Omega$.

Theorem 2.19 (Theorem of Total Probability)**.** *Let $\mathcal{I}$ be a finite or countably infinite indexing set. Let $\{E_i : i \in \mathcal{I}\}$ be a collection of mutually exclusive and exhaustive events in a probability* *space* $(\Omega, \mathcal{F}, \mathbb{P})$ *such that* $\mathbb{P}(E_i) > 0, \forall i$. *Then*

$$\mathbb{P}(E) = \sum_{i \in \mathcal{I}} \mathbb{P}(E \cap E_i) = \sum_{i \in \mathcal{I}} \mathbb{P}(E_i)\ \mathbb{P}(E \mid E_i).$$

Proof. Since E_i's are exhaustive, we have $\mathbb{P}(\cup_{i \in \mathcal{I}} E_i) = \mathbb{P}(\Omega) = 1$. Then $\mathbb{P}(E) = \mathbb{P}(E \cap (\cup_{i \in \mathcal{I}} E_i))$ (see practice problem set 1). Since E_i's are mutually exclusive, so are $E \cap E_i$'s. Hence by the finite or countable additivity of $\mathbb{P}$ (depending on whether $\mathcal{I}$ is finite or countably infinite), we have

$$\mathbb{P}(E) = \mathbb{P}(E \cap (\cup_{i \in \mathcal{I}} E_i)) = \sum_{i \in \mathcal{I}} \mathbb{P}(E \cap E_i) = \sum_{i \in \mathcal{I}} \mathbb{P}(E_i)\ \mathbb{P}(E \mid E_i).$$

This completes the proof. $\square$

Remark 2.20. The practice problem referred in the above proof does not require the fact that $\cup_{i \in \mathcal{I}} E_i = \Omega$, but rather we need $\mathbb{P}(\cup_{i \in \mathcal{I}} E_i) = 1$. In the hypothesis of the previous theorem, we may replace the exhaustiveness of E_i's by the condition $\mathbb{P}(\cup_{i \in \mathcal{I}} E_i) = 1$.

Example 2.21. Suppose we perform a random experiment with the following steps.

(a) Suppose that there are two urns. The first urn contains 3 red balls and 5 green balls. The second urn contains 6 red balls and 3 green balls. All balls of the same colour are identical.
(b) Suppose a fair die is rolled and if the outcome is 1 or 6, then the first urn is chosen. Otherwise, the second urn is chosen.
(c) Finally, 2 balls are drawn at random from the chosen urn.

We want to find the probability that both the balls drawn are red. Let E denote this event. Suppose U_1 and U_2 denote the events that the first urn and the second urn is chosen respectively. Then the events $U_i, i = 1, 2$ are mutually exclusive and exhaustive. Moreover, $\mathbb{P}(U_1) = \frac{2}{6} = \frac{1}{3}$ and $\mathbb{P}(U_2) = \frac{4}{6} = \frac{2}{3}$. Further,

$$\mathbb{P}(E \mid U_1) = \frac{\binom{3}{2}}{\binom{8}{2}} = \frac{3}{28}, \mathbb{P}(E \mid U_2) = \frac{\binom{6}{2}}{\binom{9}{2}} = \frac{15}{36} = \frac{5}{12}.$$

Then the required probability can be computed as an application of Theorem of Total Probability as

$$\mathbb{P}(E) = \mathbb{P}(U_1)\ \mathbb{P}(E \mid U_1) + \mathbb{P}(U_2)\ \mathbb{P}(E \mid U_2) = \frac{1}{3}\frac{3}{28} + \frac{2}{3}\frac{5}{12} = \frac{1}{28} + \frac{5}{18} = \frac{79}{252}.$$

Theorem 2.22 (Bayes' Theorem)**.** *Let $\mathcal{I}$ be a finite or countably infinite indexing set. Let $\{E_i : i \in \mathcal{I}\}$ be a collection of mutually exclusive and exhaustive events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(E_i) > 0, \forall i$. Then for any event $E \in \mathcal{F}$ with $\mathbb{P}(E) > 0$, we have*

$$\mathbb{P}(E_j \mid E) = \frac{\mathbb{P}(E_j)\ \mathbb{P}(E \mid E_j)}{\sum_{i \in \mathcal{I}} \mathbb{P}(E_i)\ \mathbb{P}(E \mid E_i)}, \forall j \in \mathcal{I}.$$

Proof. For any $j \in \mathcal{I}$, we have

$$\mathbb{P}(E_j \mid E) = \frac{\mathbb{P}(E_j \cap E)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E_j)\ \mathbb{P}(E \mid E_j)}{\sum_{i \in \mathcal{I}} \mathbb{P}(E_i)\ \mathbb{P}(E \mid E_i)}.$$

In the last step of the calculation above, we have used the Multiplication rule and the Theorem of Total Probability. ☐

Remark 2.23. As discussed in Remark 2.20, in the statement of Theorem 2.22, we may replace the exhaustiveness of E_i's by the condition $\mathbb{P}(\cup_{i \in \mathcal{I}} E_i) = 1$.

Remark 2.24. In the setup of Theorem 2.19 and Theorem 2.22, information about the 'standard' events E_i's may be known beforehand and we want to understand the probability of occurrence of an arbitrary event E, treated as an 'effect' caused by the E_i's. These two results, therefore, allows us to understand/quantify the relationship between 'cause' and 'effect', in terms of conditional probability.

Notation 2.25. In the setup of the Bayes' Theorem, we shall refer to $\mathbb{P}(E_i), i \in \mathcal{I}$ as prior probabilities and $\mathbb{P}(E_i \mid E), i \in \mathcal{I}$ as posterior probabilities.

Example 2.26. Consider a rare disease X that affects one in a million people. A medical test is used to test for the presence of the disease. The test is 99% accurate in the sense that if a person does not have this disease, the chance that the test shows positive is 1% and if the person has this disease, the chance that the test shows negative is also 1%. Suppose a person is tested for the disease and the test result is positive. Let A be the event that the person has the disease X. Let B be the event that the test shows positive. As per the given information, the given data may be summarized as follows.

$$\mathbb{P}(A) = 10^{-6},\ \mathbb{P}(B^c \mid A) = 0.01,\ \mathbb{P}(B \mid A^c) = 0.01.$$

Then

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A) = 1 - 10^{-6},\ \mathbb{P}(B \mid A) = 1 - \mathbb{P}(B^c \mid A) = 0.99.$$

We are interested in the conditional probability that the person has the disease, given that the test result is positive. Here, A and A^c are mutually exclusive and exhaustive. By the Bayes' Theorem, we have

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid A^c)\mathbb{P}(A^c)} = \frac{0.99 \times 10^{-6}}{0.99 \times 10^{-6} + 0.01 \times (1 - 10^{-6})} = 0.000099.$$

Definition 2.27 (Independence of Two events)**.** Let A, B be events in a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Note 2.28. If $A = \emptyset$, then A is independent of any other event B. To see this, observe that

$$\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(A)\mathbb{P}(B) = 0 \times \mathbb{P}(B) = 0.$$

Again, if $A = \Omega$, then A is independent of any other event. Observe that

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

Remark 2.29 (Independence and Mutually Exclusiveness (Pairwise disjointness) of two events). Independence should not be confused with pairwise disjointness! If A and B are disjoint, $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$ and hence A and B can be independent if and only if at least one of $\mathbb{P}(A)$ or $\mathbb{P}(B)$ equals 0. If A and B are disjoint, then $A \subseteq B^c$ and $B \subseteq A^c$. If we know that A has occurred, then we immediately conclude that B did not occur. Independence is not to be expected in such situations. On the other hand, if $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$ and A, B are independent, then $\mathbb{P}(A \cap B) > 0$. In this situation, A and B cannot be mutually exclusive.

Remark 2.30 (Conditional probability and Independence of two events). If A, B are independent and $\mathbb{P}(A) > 0$, then $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(B)$.

Example 2.31. Recall Example 2.13, where we considered rolling a standard six-sided die twice at random. The sample space is $\Omega = \{(i,j) : i,j \in \{1,2,3,4,5,6\}\}$ with $\mathbb{P}((i,j)) = \frac{1}{36}, \forall (i,j) \in \Omega$. Consider the events $A = \{(i,j) : i \text{ is odd}\}$, $B = \{(i,j) : i + j = 4\}$ and $C = \{(i,j) : j = 2\}$. Observe that

$$\mathbb{P}(A) = \frac{1}{2}, \mathbb{P}(B) = \frac{1}{12}, \mathbb{P}(A \cap B) = \frac{1}{18}.$$

Here, A and B are not independent. Again,

$$\mathbb{P}(A) = \frac{1}{2}, \mathbb{P}(C) = \frac{1}{6}, \mathbb{P}(A \cap C) = \frac{1}{12}.$$

Here, A and C are independent.

Definition 2.32 (Mutual Independence of a collection of events)**.**

(a) Let $\{E_1, E_2, \cdots, E_n\}$ be a finite collection of events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that this collection of events is mutually independent or equivalently, the events are mutually independent, if for all $k \in \{2, 3, \cdots, n\}$ and indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, we have

$$\mathbb{P}\left(\bigcap_{j=1}^{k} E_{i_j}\right) = \prod_{j=1}^{k} \mathbb{P}\left(E_{i_j}\right).$$

(b) Let $\mathcal{I}$ be any indexing set and let $\{E_i : i \in \mathcal{I}\}$ be a collection of events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that this collection of events is mutually independent or equivalently, the events are mutually independent, if all finite subcollections $\{E_{i_i}, E_{i_2}, \cdots, E_{i_k}\}$ are mutually independent.

Definition 2.33 (Pairwise Independence of a collection of events)**.** Let $\mathcal{I}$ be any indexing set and let $\{E_i : i \in \mathcal{I}\}$ be a collection of events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that this collection of events is pairwise independent or equivalently, the events are pairwise independent, if for all distinct indices i and j, the events E_i and E_j are independent.

Note 2.34. To check that events $E_1, E_2, \cdots, E_n$ are mutually independent, we need to check $\binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n} = 2^n - n - 1$ conditions. However, to check that these events are pairwise independent we need to check $\binom{n}{2}$ conditions.

Remark 2.35 (Mutual independence and Pairwise independence of a collection of events)*.* If a collection of events is mutually independent, then by definition the events are also pairwise independent. We consider an example to see that the converse need not be true. Consider a random experiment $\mathcal{E}$ with sample space $\Omega = \{1, 2, 3, 4\}$ and event space $\mathcal{F} = 2^\Omega$. If the outcomes are equally likely, then we have the probability function/measure $\mathbb{P}$ determined by the information $\mathbb{P}(\{\omega\}) = \frac{1}{4}, \forall \omega \in \Omega$. Consider the events $A = \{1, 4\}, B = \{2, 4\}, C = \{3, 4\}$. Then $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$. Moreover, $A \cap B = B \cap C = C \cap A = \{4\}$ and hence $\mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \mathbb{P}(C \cap A) = \frac{1}{4}$. Therefore, the collection of events $\{A, B, C\}$ is pairwise independent. However, $A \cap B \cap C = \{4\}$ and hence $\mathbb{P}(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. Here, the collection $\{A, B, C\}$ is not mutually independent.

Notation 2.36. We say that a collection of events is independent or equivalently, some events are independent to mean that the collection is (equivalently, the events are) mutually independent.

Remark 2.37*.* Consider random experiments where we roll a standard six-sided die thrice or draw two balls from a bin/box/urn. In such experiments multiple draws/trials of the same operations are being performed. In such situations, if the events depending purely on different trials are independent, then we say that the trials are being performed independently. In Example 2.31, the two throws/rolls of the die are independent. Here, the events A and C which depend on the first and the second throws respectively, are independent. If we toss a fair coin twice independently, then the sample space is $\Omega = \{HH, HT, TH, TT\}$ with $\mathbb{P}(\{HH\}) = \mathbb{P}(\{HT\}) = \mathbb{P}(\{TH\}) = \mathbb{P}(\{TT\}) = \frac{1}{4}$.

Note 2.38 (Functions defined on sample spaces)**.** While studying a random phenomena with the framework of a random experiment, in most situations we shall be interested in numerical quantities associated with the outcomes. To elaborate, consider the following two examples.

(a) Consider the random experiment of tossing a coin once. As discussed earlier, the sample space is $\Omega = \{H, T\}$. Suppose we think of the occurrence of head as winning a rupee and the occurrence of a tail as losing a rupee. This information may be captured by a function $X : \Omega \to \mathbb{R}$ given by

$$X(H) := 1, \quad X(T) := -1.$$

(b) Consider the random experiment of tossing a coin until head appears. The sample space may be written as $\Omega = \{H, TH, TTH, TTTH, \cdots\}$. If we are interested in the number of tosses required to obtain the first head, then the information can be captured by the function $X : \Omega \to \mathbb{R}$ given by

$$X(H) := 1, \quad X(TH) := 2, \quad X(TTH) := 3, \quad X(TTTH) := 4, \cdots$$

Now, we focus on analysis of such functions X defined on the sample space Ω of some random experiment $\mathcal{E}$.

Notation 2.39 (Pre-image of a set under a function)**.** Let Ω be a non-empty set and let $X : \Omega \to \mathbb{R}$ be a function. Given any subset A of $\mathbb{R}$, we consider the subset $X^{-1}(A)$ of Ω defined by

$$X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\}.$$

The set $X^{-1}(A)$ shall be referred to as the pre-image of A under the function X.

Remark 2.40*.* In Notation 2.39, we do not know whether the function X is bijective. As such, we cannot identify X^{-1} as the 'inverse' function of X. To avoid any confusion, treat $X^{-1}(A)$ as one symbol referring to the set as defined above and do not identify it as a combination of symbols X^{-1} and A.