Introduction to Probability

Definition : A *random experiment* is an experiment or a process for which the outcome cannot be predicted with certainty.

Definition : The *sample space* (denoted S) of a random experiment is the set of all possible outcomes. An element $x \in S$ is called an outcome.

Example 1:

Here are examples of random experiments, along with their corresponding sample spaces:

(a) Selection of a plastic component from a collection of 15 compliant pieces and 12 non-compliant pieces. Here is one possible sample space:

$$S = \{\text{piece } 1, \text{piece } 2, \ldots, \text{piece } 27\}$$

(pieces 1-15 are compliant, and pieces 16-27 are non-compliant.)

(b) Lifetime of an electronic component. Here is one possible sample space:

$$S = \{t \in \mathbb{R} \,:\, t \geq 0\} = [0, \infty).$$

(c) Number of calls to a communication system from 3:00pm-4:00pm. Here is one possible sample space:

$$S = \{0, 1, 2, 3, \ldots\}.$$

(e) The **ordered** selection of two tools **without replacement** from a box of three tools $\{A, B, C\}$. Here is one possible sample space:

$$S = \{(A, B), (A, C), (B, A), (B, C), (C, A), (C, B)\}.$$

(f) The **ordered** selection of two tools **with replacement** from a box of two tools. Here is one possible sample space:

$$S = \{(A, A), (A, B), (B, A), (B, B)\}.$$

Definition: An event E is a subset of the sample space S. We say that E has occurred if the observed outcome x is an element of E.
Notation: The notation $x \in E$ means "x is an element of the set E".

Remarks [A few special events]:

- We say that the sample space S is a *certain* event, since we know for sure that the observed outcome will be an element of S.

- $\emptyset$ is the *empty set*. We say that the empty set is an *impossible* event since the observed outcome can never be an element of the empty set. (we introduce the empty set for technical reasons)

Example 2: Consider the random experiments of Example 1.
(a) Let S be the event that the plastic component is compliant, i.e.,

$$A = \{\text{piece } 1, \text{piece } 2, \cdots, \text{piece } 15\}.$$

(recall that we said pieces 1-15 were compliant and pieces 16-27 were non-compliant)
(c) Recall the event with the number of calls from 3:00pm-4:00pm with sample space $S = \{0, 1, 2, 3, \cdots\}$. Here are some possible events:

- Let A be the event that there are less than two calls, i.e.,
$$A = \{0, 1\}.$$
- Let B be the event that there are **more** than 8 calls, i.e.,
$$B = \{9, 10, 11, \ldots\}.$$
- Let C be the event that there are **at most** 3 calls, i.e.,
$$C = \{0, 1, 2, 3\}.$$
- Let D be the event that there are **at least** 6 calls, i.e.,
$$D = \{6, 7, 8, 9, 10, \ldots\}.$$

Interpretation of a Probability:

Goal: To define a measure of (i.e. to quantify) the likelihood or the chances that an event E will occur.

Note We introoduce three interpretations of the probability $P(E)$, that is, the probability that the event E occurs:

1. Subjective Probability;
2. Equally Likely Model;
3. Relative Frequency Model.

Subjective Probability: We associate a real number $P(E)$ between 0 and 1 in a subjective manner to the event E. Numbers closer to 0 are interpreted as less likely and numbers closer to 1 are interpreted as more likely. $P(E)$ can be viewed as our personal belief regarding the likelihood of the outcome of the event E.

Example:

(a) Suppose we ask a fan of the Montreal Canadians or the Ottawa Senators what is the probability that his favorite team will win the president's cup in 2017 (the team that ends the season with the most points wins the cup).

Using his judgement and his knowledge of hockey, the fan might reply that there is a 45% chance of his favorite team winning the president's cup in 2017. We interpret the 45% like a belief that the event will happen.

(b) We ask an engineer what is the probability of a major failure of a nuclear reactor within the next 10 years. Using her judgment and her technical knowledge, the engineer could give a number between 0 and 1 to describe her personal opinion regarding the likelihood of this event occuring.

Classical Approach: The concepts of probability have existed for millennia, but the theory of probability arose as a discipline of mathematics only in the 17th century. The first definition of a probability is found below. The model is called the equally likely model.

[**Equally Likely Model**]**:** Consider a random experiment with a finite sample space S such that each result has the same chance to occur. The probability that E will occur is

$$P(E) = \frac{N(E)}{N(S)} = \frac{\text{\# favourable outcomes}}{\text{\# total possible outcomes}}$$

where $N(E) = \#$ of outcomes in E.

Remark: We consider a **random selection** of an object among N objects as an experiment with equally likely outcomes.

Example: Consider the selection of a plastic component from a collection of 15 compliant pieces and et 12 non-compliant pieces. What is the probability that the selected piece will be compliant?

Assuming that each piece has the same chance of being selected, using the classical approach, we have

$$P(C) = \frac{N(C)}{N(S)} = \frac{15}{27} = 0.5556,$$

where C is the event that the selected piece is compliant.

Frequentist approach :

Consider the use of a gasoline supplement. What is the probability that the gasoline supplement will improve the fuel efficiency of the car? Let's say that an improvement would mean using less than 35 liters per 100 km.

One solution: We could conduct an experiment : We fill the car with supplemented gasoline and we observe whether the car uses less than 35 liters per 100km. We could then repeat the experiment many times. Suppose that in 75 trials of our experiment, there were 65 trials where the car used less than 35 liters per 100km.

The frequency of our even E, where E is the event that the car uses less than 35 liters per 100 km, is 65, and its relative frequency is 65/75=0.86667=86.667%. We can use the relative frequency as an approximation of the probability of the occurrence of the event E. Therefore,

$$P(E) \approx \frac{65}{75} = 0.86667.$$

[Relative Frequency Model]: Consider a random experiment with a sample space S. We repeat the experiment n times. The probability that the event E will occur is

$$P(E) = \lim_{n\to\infty} \frac{f_n(E)}{n},$$

where $f_n(E)$ is the number of times (the frequency) that event E occurs among the n trials of the experiment.

Example :

Among the last 100,000 packets passing through a particular communication channel, there were 500 corrupted packets. What is the probability that any given packet passing through the communication channel will be corrupted?

Solution: Using the frequentist approach, the probability that any given packet will be corrupted is (approximately) 500/100,000=0.005.

Remarks:

- We have multiple approaches to probability. If we want to develop mathematical tools for doing probability we would need to verify them using each of the approaches. We need a more general definition of probability.
- The modern theory of probability is based on three fundamental principles that are called the axioms of probability theory. The equally likely model and the relative frequency model satisfy these axioms. Thus, the axiomatic approach is more general and any consequences of the axioms are true for both models.
- First, we will present methods of enumerating the possible outcomes.

Enumerating Techniques

To answer more complicated probability problems, such as those involving a sequence of steps, we will need to develop techniques for counting.

Probability trees: If an experiment can be described by a sequence of k steps, then we can illustrate all the possible outcomes in the sample space by use of a tree. Each path in the tree represents an outcome.

Example 3: We first select an operator, say Arthur, Beatrice, and Celine, and then we select one of the two machines for them to operate. Construct a probability tree to enumerate the possible outcomes.

Remark: While it is very easy to construct a tree, the tree can become very large rather quickly. We will therefore need other enumerating techniques.

Multiplication principle: If an experiment can be described by a sequence of k steps, and there are n_i ways to accomplish each step i, then

$$\#\text{ possible outcomes } = n_1 \times n_2 \times \cdots \times n_k.$$

Example 4: We first select one of three operator, one of two machines, and then we take one of five measurements using this machine. How many ways can we perform this experiment?

n Factorial ($n!$)

Definition:

Let n be a non-negative integer, that is $n = 0, 1, 2, \ldots$. We define n *factorial* by

$$n! = \begin{cases} n\,(n-1)\,(n-2) \times \cdots \times 1, & \text{if } n \geq 1 \\ 1, & \text{if } n = 0 \end{cases}$$

Example 5: $0! = 1$; $1! = 1$; $2! = 2 \cdot 1 = 2$; $3! = 3 \cdot 2 \cdot 1 = 6$

$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$; $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$

Many experiments can be described by an arrangement or a selection of objects. Since we will frequently encounter these types of experiments, we will give them special names:

Definition: Consider n distinct objects. An arrangement of r of these objects is called a **permutation**.

Remark: For a permutation the order matters.

Notation:

$P_r^n = \#$ of different permutations of r objects chosen from a collection of n distinct objects.

Definition: Consider n distinct objects. A selection of r objects is called a **combination**.

Remark: For a combination the order does **not** matter.

Notation:

$$C_r^n = \binom{n}{r} = \#\text{ of different combinations of size } r \text{ that can be chosen from a collection of distinct } n \text{ objects.}$$

Example 6:

Consider the set of objects $\{a, b, c, d\}$. Here is a list of all permutations of size 2 of these objets:

$$ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc.$$

So, $P_2^4 = 12$

And here is a list of their combinations of size 2:

$$ab, ac, ad, bc, bd, cd.$$

So $\binom{4}{2} = 6$

Note: For a combination, what is matters is which objects we select, not the order in which we select them. So ab et ba represent the same combination.

Formulas:

$$P_r^n = n\,(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

et

$$\binom{n}{r} = C_r^n = \frac{n!}{r!\,(n-r)!}$$

Proof: Consider the ordered selection of r objects among n objects. We have n choices for the first object, then $n-1$ choices for the second object, $n-2$ choices for the third object, and so on. For the last choice, the is the r'th object, we are left with $n-(r-1)$ object to choose from, so the number of choices will be $n-(r-1)=n-r+1$. Therefore, by the multiplication principle, we have

$$\begin{aligned} P_r^n &= n\,(n-1)\cdots(n-r+1) \\ &= n\,(n-1)\cdots(n-r+1)\frac{(n-r)(n-r-1)\cdots 1}{(n-r)(n-r-1)\cdots 1} \\ &= \frac{n!}{(n-r)!}. \end{aligned}$$

We can also construct the above permutation in a different way. Choose r from the collection of n objects, irrespective of order. There are C_r^n possible unordered selections. Then, arrange our selection in order. We have r choices for the first object, then $r-1$ choices for the second object, $r-2$ choices for the third object, and so on. Therefore, there are $r\,(r-1)\,(r-2)\cdots 1 = r!$ different ways of arranging these r objects. Therefore, by the multiplication principle, we have

$$P_r^n = C_r^n\, r!.$$

But, $P_r^n = n!/(n-r)!$, so

$$C_r^n = \frac{n!/(n-r)!}{r!} = \frac{n!}{(n-r)!r!}.$$

Example 7: We have a group of 10 Engineers. We will select 3 engineers from this group to lead the committee on little projects, the committee on medium projects, and the committee on big projects, respectively (each committee must have exactly one leader, and no two committees can have the same leader). In how many different ways can we choose these three committee leaders?

Example 8: Consider a collection of 50 articles, of which 3 are defective. We select 5 articles at random.
(a) In how many different ways can we choose the articles (ignoring the order in which they are selected)?
(b) What is the probability that there will be **exactly** 1 defective article from among the 5 chosen articles?
(c) What is the probability that there will be **at most** 1 defective article from among the 5 chosen articles?

Example 9: A machine makes three types of holes: small, medium and large. Suppose we can program the machine to perform a sequence of operations. In how many different ways can we program the machine if we want 2 small, 2 medium, and 3 large holes. That is, we want to count the number of arrangements of the holes where holes of the same type are indistinguishable.

Example 10: Suppose we have 12 students and that we want to divide the 12 students into 4 groups of size 3. Determine the number of different ways that we could distribute these 12 students into 4 groups.

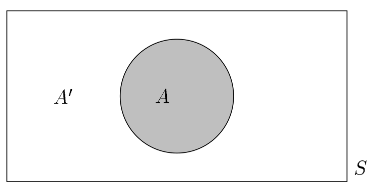
Remark: We are now laying the groundwork for the general theory of probability. To do so, we need to define the following operations on events: *union*, *intersection*, and *complement*.

Operations on events

Note: The following opperations permit us to represent events in terms of other events.

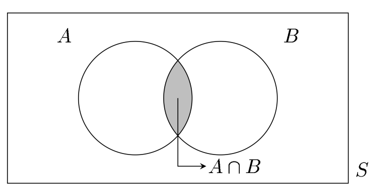
Complement:

We say that the event A' occurs if the event A does **not** occur.



Intersection:

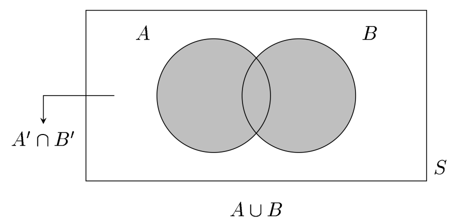
a) We say that $A \cap B$ occurs if both A **and** B occur.



b) More generally, we say that $E_1 \cap E_2 \cap \cdots \cap E_n$ occurs is **all** the events $E_1, E_2, \ldots, E_n$ occur.

Union:

a) We say that $A \cup B$ occurs if A occurs **or** B occurs, (including if both A and B occur).



b) More generally, we say that $E_1 \cup E_2 \cup \cdots \cup E_n$ occurs if **at least one** of the events $E_1, E_2, \cdots, E_n$ occurs.

De Morgan's law:
a)

$$
\begin{aligned}
& (E_1 \cup E_2 \cup \cdots \cup E_n)' \text{ occurs} \\
= \; & \text{none of the events } E_1, E_2, \ldots, E_n \\
& \text{ occurs} \\
= \; & E_1' \cap E_2' \cap \ldots \cap E_n' \text{ occurs}
\end{aligned}
$$

b)

$$
\begin{aligned}
& (E_1 \cap E_2 \cap \cdots \cap E_n)' \text{occurs} \\
= \; & \text{at least on of the events } E_1, E_2, \ldots, E_n \\
& \text{ occurs} \\
= \; & E_1' \cup E_2' \cup \ldots \cup E_n' \text{occurs}
\end{aligned}
$$

Remark: We will see examples of De Morgan's law later, together with other probability rules.

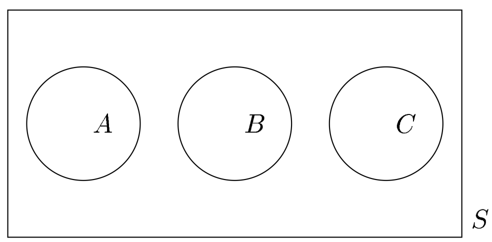
Mutually Exclusive Events:

Definition: The events $E_1, E_2, \ldots, E_n$ are said to be **mutually exclusive** if

$$E_i \cap E_j = \emptyset \quad \text{for } i \neq j.$$

Remark: If A and B are mutually exclusive, this means that A and B cannot occur simultaneously.

An illustration of three mutually exclusive events:

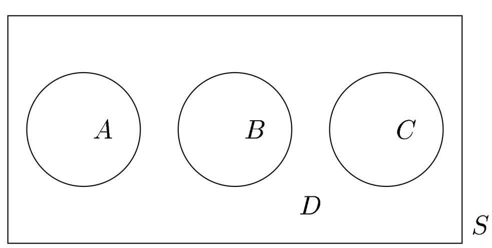


Exhaustive Events:

Definition: The events $E_1, E_2, \ldots, E_n$ are said to be **exhaustive** if

$$E_1 \cup E_2 \cup \cdots \cup E_n = S.$$

An illustration of 4 *mutually exclusive* and *exhaustive* events:



Example 11: Suppose we have three factories : in Ontario, inn Quebec and in Michigan. Suppose that the article is produced in only one of these three factories. Let O, Q and M, be the events that the article in produces in Ontario, in Quebec and in Michigan, resepctively. Then the events O, Q et M are both mutually exclusive and exhaustive.

Remark: We are now ready to define a probability function.

Axioms of Probability: Consider an experiment with the sample space S. For each event E, we can associate a real number $P(E)$ such that:

Positivity :

(a) $P(E) \geq 0$,

Certainty :

(b) $P(S) = 1$,

Additivity :

(c) For each sequence of events $E_1, E_2, \ldots$ that are mutually exclusive (that is $E_i \cap E_j = \emptyset$ (the empty set), if $i \neq j$), we have

$$P\left(E_1 \bigcup E_2 \bigcup \ldots\right) = P\left(E_1\right) + P\left(E_2\right) + \ldots.$$

Note: $P(E)$ is called the probability that E occurs.

Here are some consequences of the axioms of probability:

Theorem:

1. $P(\emptyset) = 0$

2. $E_1 \subseteq E_2 \Rightarrow P(E_1) \leq P(E_2)$.

3. $0 \leq P(E) \leq 1$.

4. $P(E') = 1 - P(E)$.

Here are some rules of addition :

a)

$$P(A \cap B') = P(A) - P(A \cap B)$$

b)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

c)

$$\begin{aligned} & P(E_1 \cup E_2 \cup \ldots \cup E_n) \\ = & \ 1 - P\left[(E_1 \cup E_2 \cup \ldots \cup E_n)'\right] \\ = & \ 1 - P(E_1' \cap E_2' \cap \ldots \cap E_n') \end{aligned}$$

Example 12: The probability that a piece of integrated circuit will have a defective etching is 0.12, the probability that it will have a defective slot is 0.29 and the probability that it will have both defects is 0.07.

(a) What is the probability that a piece of integrated circuit will have a defective etching but not a defective slot.

(b) What is the probability that a piece of integrated circuit will have a defective etching or a defective slot?

(c) What is the probability that it will have neither a defective etching nor a defective slot?

Conditional Probability

Definition: If $P(B) > 0$, the *conditional probability* that A will occur given that B occurs is

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

Remark:

1. We can show that the function $P(\cdot|B)$ satisfies the three probability axioms. Thus, the consequences of the axioms also apply to conditional probabilities. For example,

   (a) $P(\emptyset|B) = 0$.

   (b) If $A \subseteq C$, then $P(A|B) \leq P(C|B)$.

   (c) $0 \leq P(A|B) \leq 1$, for all events A.

   (d) $P(A'|B) = 1 - P(A|B)$, for all events A.

2. For the <u>equally likely model</u>, the definition of a conditional probability is equivalent to

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

   For the latter approach, we interpret $P(A|B)$ as the proportion of outcomes from B that satisfy criteria for A.

Remark: We can also interpret the context for which we use a conditional probability as a change in the experimental conditions.

Multiplication Rule :

$$P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B)$$

This rule follows from the definition of conditional probability.

Example 13: Your company offers two contracts. You believe that the probability that you will obtain contract A is 0.8. If you do obtain contract A, the probability that you will then obtain contract B is 0.2. However, if you don't obtain contract A, then the probability that you will obtain contract B is 0.3.

(a) What is the probability that you will get both contracts?

(b) What is the probability of getting exactly one of the two contracts?

Law of Total Probability (general case): If $E_1, ...E_k$ are mutually exclusive and exhaustive (i.e. $E_i \cap E_j = \emptyset$ if $i \neq j$ and $E_1 \cup .... \cup E_k = S$), then for every event B,

$$\begin{aligned} P(B) &= P(B \cap E_1) + ... + P(B \cap E_k) \\ &= P(B \mid E_1)P(E_1) + ... + P(B \mid E_k)P(E_k) \end{aligned}$$

Bayes' Theorem:

$$P(E|A) = \frac{P(E \cap A)}{P(A)} = \frac{P(A|E)P(E)}{P(A)}$$

Example 14: Samples taken from a river near an industrial plant are tested for toxic levels of lead and mercury: 32% contain toxic levels of mercury, 16% contain toxic levels of lead, and 38% contain toxic levels of at least one of the two substances.

1. A sample chosen at random contains toxic levels of mercury. What is the probability that it also contains toxic levels of lead?
2. It is known that 45% of the samples containing toxic levels of mercury also contain toxic levels of arsenic. What is that probability that a sample contains toxic levels of both mercury and arsenic?
3. From these samples that do not contain toxic levels of mercury, only 2% contain toxic levels of arsenic. What is the probability that the sample contains toxic levels of arsenic?

Example 15: Nissan sold three models of cars in 1999 : Sentras, Maximas and Pathfinders. Of the vehicles sold, 50% were Sentras, 30% were Maximas and 20% were Pathfinders. For that same year, 12% of the Sentras, 15% of the Maximas and 25% of the Pathfinders had a problem in the ignition system.

1. What is the probability that a Nissan sold in 1999 had a problem with the ignition system?
2. A 1999 Nissan has problems with the ignition system, what is the probability that it is a Pathfinder?

Independence of Events

Definition: Two events A and B are *independant*, if

$$P(A \cap B) = P(A)P(B). \tag{1}$$

Question: Why do we use Equation (1) as a definition for independence? The following theorem will provide a motivation for this definition:

Theorem: Let A and B be events of nonzero probability, that is $P(A) > 0$ and $P(B) > 0$. Then the following statements are equivalent :

1. $P(A \cap B) = P(A)P(B)$
2. $P(B \mid A) = P(B)$
3. $P(A \mid B) = P(A)$

Remark: So, if A and B are independant, then the probability of A does not depend on whether B occurs and vice-versa.

Example 16: Here is a summary of the analysis of whether each of 370 gears satisfy to certain specifications:

Surface is good	Roundness is good yes	no	total
yes	345	5	350
no	12	8	20
total	357	13	370

Let A be the event the surface of a gear is good and let B be the event that its roundness is good. Are A and B independent events? If there is a dependence, please describe it.

Remark: Do not confuse the concept of independent events with the concept of mutually exclusive events. Both concepts imply different things.

- If A and B are **independent events**, then

$$P(A \cap B) = P(A)\, P(B)$$

and

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A)\, P(B) \end{aligned}$$

- If A and B are **mutually exclusive events**, then

$$P(A \cap B) = P(\emptyset) = 0$$

et

$$P(A \cup B) = P(A) + P(B).$$

We generalize the concept of independent events to collections of events:

Definition: The events $E_1, ..., E_k$ are **mutually independent** if for all collections of events $E_{i_1}, E_{i_2}, ..., E_{i_k}$, we have

$$P(E_{i_1} \cap E_{i_2} \cap ... \cap E_{i_k}) = P(E_{i_1}) \times P(E_{i_2}) \times ... \times P(E_{i_k})$$

Remark: If we have a collection of independent events and we replace one of those events with its complement, then the new collection of events is also mutually independent.

For example: If A et B are independent, then

- A and B' are independent;
- A' et B are independent;
- A' et B' are independent.

Remark: Oftentimes we will assume (when reasonable) that events are independent and we will use independence to calculate the probability of more complicated events.

Example 17: Suppose that an electronic component will work is 0.90. Consider 5 independent components.

(a) What is the probability that all 5 will work?

(b) Compute the probability that at least 1 will work?

(c) What is the probability that only the third component will work?

Example 18: Consider the following circuit. We say that it is functional if there is a path of functional components from left to right. The probability that the component is functional is illustrated. Suppose components operate independently of each other. What is the probability that the circuit is functional?