Set theory for probability

by Marco Taboga, PhD

Understanding the basics of set theory is a prerequisite for studying probability.
This lecture presents a concise introduction to set membership and inclusion, unions, intersections and complements. These are all concepts that are frequently used in the calculus of probabilities.

Table of contents

Sets

A set is a collection of objects. Sets are usually denoted by a letter and the objects (or elements) belonging to a set are usually listed within curly brackets.

Example Denote by the letter s the set of the natural numbers less than or equal to 5 . Then, we can write

$$S = \{1, 2, 3, 4, 5\}$$

Example Denote by the letter A the set of the first five letters of the alphabet. Then, we can write

$$A = \{a, b, c, d, e\}$$

Note that a set is an unordered collection of objects, i.e. the order in which the elements of a set are listed does not matter.
and

$$\{b, d, a, c, e\}$$

are considered identical.

Sometimes a set is defined in terms of one or more properties satisfied by its elements. For example, the set

$$S = \{1, 2, 3, 4, 5\}$$

could be equivalently defined as

$$S = \{n \in \mathbb{N} : n \leq 5\}$$

which reads as follows: " S is the set of all natural numbers n such that n is less than or equal to 5 ", where the colon symbol (:) means "such that" and precedes a list of conditions that the elements of the set need to satisfy.

Example The set

$$S = \left\{n \in \mathbb{N} : \frac{n}{4} \in \mathbb{N}\right\}$$

is the set of all natural numbers n such that n divided by 4 is also a natural number, that is,

$$S = \{4, 8, 12, \ldots\}$$

Set membership

When an element a belongs to a set A, we write

$$a \in A$$

which reads " a belongs to A " or " a is a member of A ".
On the contrary, when an element a does not belong to a set A, we write

$$a \notin A$$

which reads " a does not belong to A " or " a is not a member of A ".
Example Let the set s be defined as follows:

$$A = \{2, 4, 6, 8, 10\}$$

Then, for example,

$$4 \in A$$

and

$$7 \notin A$$

Set inclusion

If A and B are two sets and if every element of A also belongs to B, then we write
and we read " B includes A ". We also say that A is a subset of B.
Example The set

$$A = \{2, 3\}$$

is included in the set

$$B=\{1,2,3,4\}$$

because all the elements of A also belong to B. Thus, we can write $A \subseteq B$

When $A \subseteq B$ but A is not the same as B (i.e., there are elements of B that do not belong to A), then we write

$$A \subset B$$

which reads " A is strictly included in B " or

$$B \supset A$$

We also say that A is a proper subset of B.
Example Given the sets

$$\begin{aligned} & A=\{2,3\} \\ & B=\{1,2,3,4\} \\ & C=\{2,3\} \end{aligned}$$

we have that

$$\begin{aligned} & A \subset B \\ & A \subseteq C \end{aligned}$$

but we cannot write

$$A \subset C$$

Union

Let A and B be two sets. Their union is the set of all elements that belong to at least one of them and it is denoted by

$$A \cup B$$

Example Define two sets A and B as follows:

$$\begin{aligned} & A=\{a, b, c, d\} \\ & B=\{c, d, e, f\} \end{aligned}$$

Their union is

$$A \cup B=\{a, b, c, d, e, f\}$$

If $A_1, A_2, \ldots, A_n$ are n sets, their union is the set of all elements that belong to at least one of them and it is denoted by

$$\bigcup_{i=1}^n A_i=A_1 \cup A_2 \cup \ldots \cup A_n$$

Example Define three sets A_1, A_2 and A_3 as follows:

$$A_1 = \{a, b, c, d\}$$
$$A_2 = \{c, d, e, f\}$$
$$A_3 = \{c, f, g\}$$

Their union is

$$\bigcup_{i=1}^{3} A_i = A_1 \cup A_2 \cup A_3 = \{a, b, c, d, e_2, f, g\}$$

Intersection

Let A and B be two sets. Their intersection is the set of all elements that belong to both of them and it is denoted by

$$A \cap B$$

Example Define two sets A and B as follows:

$$A = \{a, b, c, d\}$$
$$B = \{c, d, e, f\}$$

Their intersection is

$$A \cap B = \{c, d\}$$

If $A_1, A_2, \ldots, A_n$ are n sets, their intersection is the set of all elements that belong to all of them and it is denoted by

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \ldots \cap A_n$$

Example Define three sets A_1, A_2 and A_3 as follows:

$$A_1 = \{a, b, c, d\}$$
$$A_2 = \{c, d, e, f\}$$
$$A_3 = \{c, f, g\}$$

Their intersection is

$$\bigcap_{i=1}^{3} A_i = A_1 \cap A_2 \cap A_3 = \{c\}$$

Complement

Complementation is another concept that is fundamental in probability theory.
Suppose that our attention is confined to sets that are all included in a larger set Ω, called universal set. Let A be one of these sets. The complement of A is the set of all elements of Ω that do not belong to A and it is indicated by
A^c

Example Define the universal set Ω as follows:
and the two sets

$$A = \{b, c, d\}$$
$$B = \{c, d, e\}$$

The complements of A and B are

$$A^c = \{a, e, f, g, h\}$$
$$B^c = \{a, b, f, g, h\}$$

Also note that, for any set A, we have

$$(A^c)^c = A$$

De Morgan's Laws

De Morgan' Laws are

$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

and can be extended to collections of more than two sets:

$$(\mathrm{U}_4 4)^{\circ} - \mathrm{n}^4$$
$$(\mathrm{n}_1 \mathrm{Al}^2 - \mathrm{U}^n \mathrm{A}^2$$

Solved exercises

Below you can find some exercises with explained solutions.

Exercise 1

Define the following sets:

$$A_1 = \{a, b, c\}$$
$$A_2 = \{b, c, d, e, f\}$$
$$A_3 = \{b, f\}$$
$$A_4 = \{a, b, d\}$$

Find the following union:

$$A = \bigcup_{i=2}^{4} A_i$$

Solution

The union can be written as

$$A = A_2 \cup A_3 \cup A_4$$

The union of the three sets A_2, A_3 and A_4 is the set of all elements that belong to at least one of them:

$$\begin{aligned} A &= A_2 \cup A_3 \cup A_4 \\ &= \{a, b, c, d, e, f\} \end{aligned}$$

Exercise 2

Given the sets defined in the previous exercise, find the following intersection:

$$A = \bigcap_{i=1}^{4} A_i$$

Solution

The intersection can be written as

$$A = A_1 \cap A_2 \cap A_3 \cap A_4$$

The intersection of the four sets A_1, A_2, A_3 and A_4 is the set of elements that are members of all the four sets:

$$\begin{aligned} A &= A_1 \cap A_2 \cap A_3 \cap A_4 \\ &= \{b\} \end{aligned}$$

Exercise 3

Suppose that A and B are two subsets of a universal set Ω and that

$$\begin{aligned} A^c &= \{a, b, c\} \\ B^c &= \{b, c, d\} \end{aligned}$$

Find the following union:

$$(A \cup B)^c$$

Solution

By using De Morgan's laws, we obtain

$$\begin{aligned} (A \cup B)^c &= A^c \cap B^c \\ &= \{a, b, c\} \cap \{b, c, d\} \\ &= \{b, c\} \end{aligned}$$

Applications to probability

Now that you are familiar with the basics of set theory, you can see how it is used in probability theory.

Read the lectures on:

1. the mathematics of probability;
2. conditional probability;
3. independent events.

How to cite

Please cite as:

Taboga, Marco (2021). "Set theory", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/mathematical-tools/set-theory.

Permutations

by Marco Taboga, PhD

This lecture introduces permutations, one of the most important concepts in combinatorial analysis.
We first deal with permutations without repetition, also called simple permutations, and then with permutations with repetition.

Table of contents

1. Permutation of n elements without repetition

A permutation without repetition of n objects is one of the possible ways of ordering the n objects.
A permutation without repetition is also simply called a permutation.
The following subsections give a slightly more formal definition of permutation and deal with the problem of counting the number of possible permutations of n objects.

2. Definition of permutation without repetition

Let $a_1, a_2, \ldots, a_n$ be n objects. Let $s_1, s_2, \ldots, s_n$ be n slots to which the n objects can be assigned. A permutation (or permutation without repetition or simple permutation) of $a_1, a_2, \ldots, a_n$ is one of the possible ways to fill each of the n slots with one and only one of the n objects (with the proviso that each object can be assigned to only one slot).

Example Consider three objects a_1, a_2 and a_3. There are three slots (s_1, s_2 and s_3) to which we can assign the three objects (a_1, a_2 and a_3). There are six possible permutations of the three objects (six possible ways to fill the three slots with the three objects):

Slots	s_1	s_2	s_3
Permutation 1	a_1	a_2	a_3
Permutation 2	a_1	a_3	a_2
Permutation 3	a_2	a_1	a_3
Permutation 4	a_2	a_3	a_1
Permutation 5	a_3	a_2	a_1
Permutation 6	a_3	a_1	a_2

3. Number of permutations without repetition

Denote by P_n the number of possible permutations of n objects. How much is P_n in general? In other words, how do we count the number of possible permutations of n objects?

The number P_n can be derived by noting that filling the n slots is a sequential problem:

1. First, we assign an object to the first slot. There are n objects that can be assigned to the first slot, so there are

4. n possible ways to fill the first slot

2. Then, we assign an object to the second slot. There were n objects, but one has already been assigned to a slot. So, we are left with $n-1$ objects that can be assigned to the second slot. Thus, there are
   $n-1$ possible ways to fill the second slot
   and

$$n \cdot (n-1) \text{ possible ways to fill the first two slots}$$

3. Then, we assign an object to the third slot. There were n objects, but two have already been assigned to a slot. So, we are left with $n-2$ objects that can be assigned to the third slot. Thus, there are

$$n-2 \text{ possible ways to fill the third slot}$$

and

$$n \cdot (n-1) \cdot (n-2) \text{ possible ways to fill the first three slots}$$

4. An so on, until only one object and one free slot remain.
5. Finally, when only one free slot remains, we assign the remaining object to it. There is only one way to do this. Thus, there are

$$1 \text{ possible way to fill the last slot}$$

and

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1 \text{ possible ways to fill all the } n \text{ slots}$$

Therefore, by the above sequential argument, the total number of possible permutations of n objects is

$$P_n = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1$$

The number P_n is usually indicated as follows:

$$P_n = n!$$

where n ! is read " n factorial", with the convention that

$$0! = 1$$

Example The number of possible permutations of 5 objects is

$$P_5 = 5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

5. Permutation of n elements with repetition

A permutation with repetition of n objects is one of the possible ways of selecting another set of n objects from the original one. The selection rules are:

1. each object can be selected more than once;
2. the order of selection matters (the same n objects selected in different orders are regarded as different permutations).

Thus, the difference between simple permutations and permutations with repetition is that objects can be selected only once in the former, while they can be selected more than once in the latter.

The following subsections give a slightly more formal definition of permutation with repetition and deal with the problem of counting the number of possible permutations with repetition.

6. Definition of permutation with repetition

Let $a_1, a_2, \ldots, a_n$ be n objects. Let $s_1, s_2, \ldots, s_n$ be n slots to which the n objects can be assigned. A permutation with repetition of $a_1, a_2, \ldots, a_n$ is one of the possible ways to fill each of the n slots with one and only one of the n objects (with the proviso that an object can be assigned to more than one slot).

Example Consider two objects a_1 and a_2. There are two slots to fill (s_1 and s_2). There are four possible permutations with repetition of the two objects (four possible ways to assign an object to each slot, being allowed to assign the same object to more than one slot):

Slots	s_1	s_2
Permutation 1	a_1	a_1
Permutation 2	a_1	a_2
Permutation 3	a_2	a_1
Permutation 4	a_2	a_2

7. Number of permutations with repetition

Denote by P'_n the number of possible permutations with repetition of n objects. How much is P'_n in general? In other words, how do we count the number of possible permutations with repetition of n objects?

We can derive a general formula for P'_n by using a sequential argument:

1. First, we assign an object to the first slot. There are n objects that can be assigned to the first slot, so there are
   more than once. So, there are n objects that can be assigned to the second slot and n possible ways to fill the second slot
   and

$$n \cdot n \text{ possible ways to fill the first two slots}$$

3. Then, we assign an object to the third slot. Even if two objects have been assigned to a slot in the previous two steps, we can still choose among n objects, because we are allowed to choose an object more than once. So, there are n objects that can be assigned to the third slot and
   n possible ways to fill the third slot
   and
   $n \cdot n \cdot n$ possible ways to fill the first three slots
4. An so on, until we are left with only one free slot (the n-th).
5. When only one free slot remains, we assign one of the n objects to it. Thus, there are:
   n possible ways to fill the last slot
   and

$$\underbrace{n \cdot n \cdot \ldots \cdot n}_{n \text{ times}} \text{ possible ways to fill the } n \text{ available slots}$$

Therefore, by the above sequential argument, the total number of possible permutations with repetition of n objects is

$$P'_n = n^n$$

Example The number of possible permutations with repetition of 3 objects is

$$3^3 = 27$$

8. Solved exercises

Below you can find some exercises with explained solutions.

9. Exercise 1

There are 5 seats around a table and 5 people to be seated at the table. In how many ways can they seat themselves?

Solution
Sitting 5 people at the table is a sequential problem. We need to assign a person to the first chair. There are 5 possible ways to do this. Then we need to assign a person to the second chair. There are 4 possible ways to do this, because one person has already been assigned. An so on, until there remain one free chair and one person to be seated. Therefore, the number of ways to seat the 5 people at the table is equal to the number of permutations of 5 objects (without repetition). If we denote it by P_5, then

$$\begin{aligned} P_5 &= 5! \\ &= 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \\ &= 120 \end{aligned}$$

10. Exercise 2

Bob, John, Luke and Tim play a tennis tournament. The rules of the tournament are such that at the end of the tournament a ranking will be made and there will be no ties. How many different rankings can there be?

Solution
Ranking 4 people is a sequential problem. We need to assign a person to the first place. There are 4 possible ways to do this. Then we need to assign a person to the second place. There are 3 possible ways to do this, because one person has already been assigned. An so on, until there remains one person to be assigned. Therefore, the number of ways to rank the 4 people participating in the tournament is equal to the number of permutations of 4 objects (without repetition). If we denote it by P_4, then

$$\begin{aligned} P_4 &= 4! \\ &= 4 \cdot 3 \cdot 2 \cdot 1 \\ &= 24 \end{aligned}$$

11. Exercise 3

A byte is a number consisting of 8 digits that can be equal either to 0 or to 1 . How many different bytes are there?

12. Solution

To answer this question we need to follow a line of reasoning similar to the one we followed when we derived the number of permutations with repetition. There are 2 possible ways to choose the first digit and 2 possible ways to choose the second digit. So, there are 4 possible ways to choose the first two digits. There are 2 possible ways two choose the third digit and 4 possible ways to choose the first two. Thus, there are 8 possible ways to choose the first three digits. An so on, until we have chosen all digits. Therefore, the number of ways to choose the 8 digits is equal to

$$\underbrace{2 \cdot \ldots \cdot 2}_{8 \text{ times}} = 2^8 = 256$$

13. How to cite

Please cite as:
Taboga, Marco (2021). "Permutations", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix.
https://www.statlect.com/mathematical-

k-permutations

by Marco Taboga, PhD

This lecture introduces k-permutations, a basic concept in combinatorial analysis. Before reading this lecture, you should read the lecture on permutations.

We first deal with k-permutations without repetition and then with k-permutations with repetition.

Table of contents

1. Solved exercises

2. k-permutation without repetition

A k-permutation without repetition of n objects is a way of selecting k objects from a list of n. The selection rules are:

1. the order of selection matters (the same k objects selected in different orders are regarded as different k-permutations);
2. each object can be selected only once.

A k-permutation without repetition is also simply called k-permutation.
The following subsections give a slightly more formal definition of k-permutation and deal with the problem of counting the number of possible k-permutations.

3. Definition of k-permutation without repetition

Let $a_1, a_2, \ldots, a_n$ be n objects. Let $s_1, s_2, \ldots, s_k$ be $k (k \leq n)$ slots to which k of the n objects can be assigned. A k-permutation (or k-permutation without repetition or simple k-permutation) of n objects from $a_1, a_2, \ldots, a_n$ is one of the possible ways to choose k of the n objects and fill each of the k slots with one and only one object. Each object can be chosen only once.

two of the three objects. There are six possible 2-permutations of the three objects (six possible ways to choose two objects and fill the two slots with the two objects):

Slots	s_1	s_2
2-permutation 1	a_1	a_2
2-permutation 2	a_1	a_3
2-permutation 3	a_2	a_1
2-permutation 4	a_2	a_3
2-permutation 5	a_3	a_1
2-permutation 6	a_3	a_2

4. Number of k-permutations without repetition

Denote by $P_{n,k}$ the number of possible k-permutations of n objects. How much is $P_{n,k}$ in general? In other words, how do we count the number of possible k-permutations of n objects?

We can derive a general formula for $P_{n,k}$ by filling the k slots in a sequential manner:

1. First, we assign an object to the first slot. There are n objects that can be assigned to the first slot, so there are

$$n \text{ possible ways to fill the first slot}$$

2. Then, we assign an object to the second slot. There were n objects, but one has already been assigned to a slot. So, we are left with $n-1$ objects that can be assigned to the second slot. Thus, there are

$$n-1 \text{ possible ways to fill the second slot}$$

and

$$n \cdot (n-1) \text{ possible ways to fill the first two slots}$$

3. Then, we assign an object to the third slot. There were n objects, but two have already been assigned to a slot. So, we are left with $n-2$ objects that can be assigned to the third slot. Thus, there are
   $n-2$ possible ways to fill the third slot
   and
   $n \cdot (n-1) \cdot (n-2)$ possible ways to fill the first three slots
4. An so on, until we are left with $n-k+1$ objects and only one free slot (the k-th).
5. Finally, when only one free slot remains, we assign one of the remaining $n-k+1$ objects to it. Thus, there are
   $n-k+1$ possible ways to fill the last slot

Therefore, by the above sequential argument, the total number of possible k-permutations of n objects is

$$P_{n,k} = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)$$

$P_{n,k}$ can be written as

$$P_{n,k} = \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) \cdot (n-k) \cdot (n-k-1) \cdot \ldots \cdot 2 \cdot 1}{(n-k) \cdot (n-k-1) \cdot \ldots \cdot 2 \cdot 1}$$

Remembering the definition of factorial, we can see that the numerator of the above ratio is n ! while the denominator is $(n-k)$!, so the number of possible k-permutations of n objects is

$$P_{n,k} = \frac{n!}{(n-k)!}$$

The number $P_{n,k}$ is usually indicated as follows:

$$P_{n,k} = n^k$$

Example The number of possible 3-permutations of 5 objects is

$$P_{5,3} = \frac{5!}{2!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = 5 \cdot 4 \cdot 3 = 60$$

5. k-permutation with repetition

A k-permutation with repetition of n objects is a way of selecting k objects from a list of n. The selection rules are:

1. the order of selection matters (the same k objects selected in different orders are regarded as different k-permutations);
2. each object can be selected more than once.

Thus, the difference between k-permutations without repetition and k-permutations with repetition is that objects can be selected more than once in the latter, while they can be selected only once in the former.

The following subsections give a slightly more formal definition of k-permutation with repetition and deal with the problem of counting the number of possible k-permutations with repetition.

6. Definition of k-permutation with repetition

Let $a_1, a_2, \ldots, a_n$ be n objects. Let $s_1, s_2, \ldots, s_k$ be $k(k \leq n)$ slots to which k of the n objects can be assigned. A k-permutation with repetition of n objects from $a_1, a_2, \ldots, a_n$ is one of the possible ways to choose k of the n objects and fill each of the k slots with one and only one object. Each object can be chosen more than once.

Example Consider three objects a_1, a_2 and a_3 and two slots (s_1 and s_2). There are nine possible 2permutations with repetition of the three objects (nine possible ways to choose two objects and fill the two slots with the two objects, being allowed to pick the same object more than once):

Slots	s_1	s_2
2-permutation 1	a_1	a_1
2-permutation 2	a_1	a_2
2-permutation 3	a_1	a_3
2-permutation 4	a_2	a_1
2-permutation 5	a_2	a_2
2-permutation 6	a_2	a_3
2-permutation 7	a_3	a_1
2-permutation 8	a_3	a_2
2-permutation 9	a_3	a_3

7. Number of k-permutations with repetition

Denote by $P'_{n,k}$ the number of possible k-permutations with repetition of n objects. How much is $P'_{n,k}$ in general? In other words, how do we count the number of possible k-permutations with repetition of n objects?

We can derive a general formula for $P'_{n,k}$ by filling the k slots in a sequential manner:

1. First, we assign an object to the first slot. There are n objects that can be assigned to the first slot, so there are
   n possible ways to fill the first slot
2. Then, we assign an object to the second slot. Even if one object has been assigned to a slot in the previous step, we can still choose among n objects, because we are allowed to choose an object more than once. So, there are n objects that can be assigned to the second slot and
   n possible ways to fill the second slot
   and
   $$n \cdot n \text{ possible ways to fill the first two slots}$$
3. Then, we assign an object to the third slot. Even if two objects have been assigned to a slot in the previous two steps, we can still choose among n objects, because we are allowed to choose an object more than once. So, there are n objects that can be assigned to the second slot and
   n possible ways to fill the third slot
   and
   $n \cdot n \cdot n$ possible ways to fill the first three slots
4. An so on, until we are left with only one free slot (the k-th).
5. When only one free slot remains, we assign one of the n objects to it. Thus, there are:
   n possible ways to fill the last slot

and
$\underbrace{n \cdot n \cdot \ldots \cdot n}_{k \text{ times}}$ possible ways to fill the k available slots

Therefore, by the above sequential argument, the total number of possible k-permutations with repetition of n objects is

$$P'_{n,k} = n^k$$

Example The number of possible 2-permutations of 4 objects is

$$P'_{4,2} = 4^2 = 16$$

8. Solved exercises

Below you can find some exercises with explained solutions.

9. Exercise 1

There is a basket of fruit containing an apple, a banana and an orange and there are five girls who want to eat one fruit. How many ways are there to give three of the five girls one fruit each and leave two of them without a fruit to eat?

10. Solution

Giving the 3 fruits to 3 of the 5 girls is a sequential problem. We first give the apple to one of the girls. There are 5 possible ways to do this. Then we give the banana to one of the remaining girls. There are 4 possible ways to do this, because one girl has already been given a fruit. Finally, we give the orange to one of the remaining girls. There are 3 possible ways to do this, because 2 girls have already been given a fruit. Summing up, the number of ways to assign the three fruits is equal to the number of 3permutations of 5 objects (without repetition). If we denote it by $P_{5,3}$, then

$$\begin{aligned} P_{5,3} &= \frac{5!}{(5-3)!} \\ &= \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} \\ &= 5 \cdot 4 \cdot 3 \\ &= 60 \end{aligned}$$

11. Exercise 2

An hexadecimal number is a number whose digits can take sixteen different values: either one of the
ten numbers from 0 to 9 , or one of the six letters from A to F. How many different 8-digit

hexadecimal numbers are there, if an hexadecimal number is allowed to begin with any number of zeros?

Solution
Choosing the 8 digits of the hexadecimal number is a sequential problem. There are 16 possible ways to choose the first digit and 16 possible ways to choose the second digit. So, there are 16×16 possible ways to choose the first two digits. There are 16 possible ways two choose the third digit and 16×16 possible ways to choose the first two. Thus, there are $16 \times 16 \times 16$ possible ways to choose the first three digits. An so on, until we have chosen all digits. Therefore, the number of ways to choose the 8 digits is equal to the number of 8-permutations with repetition of 16 objects:

$$P'_{16,8} = 16^8$$

12. Exercise 3

An urn contains ten balls, each representing one of the ten numbers from 0 to 9 . Three balls are drawn at random from the urn and the corresponding numbers are written down to form a 3-digit number, writing down the digits from left to right in the order in which they have been extracted. When a ball is drawn from the urn it is set aside, so that it cannot be extracted again. If one were to write down all the 3-digit numbers that could possibly be formed, how many would they be?

13. Solution

The 3 balls are drawn sequentially. At the first draw there are 10 balls, hence 10 possible values for the first digit of our 3-digit number. At the second draw there are 9 balls left, hence 9 possible values for the second digit of our 3-digit number. At the third and last draw there are 8 balls left, hence 8 possible values for the third digit of our 3-digit number. In summary, the number of possible 3-digit numbers is equal to the number of 3-permutations of 10 objects (without repetition). If we denote it by $P_{10,3}$, then

$$\begin{aligned} P_{10,3} &= \frac{10!}{(10-3)!} \\ &= \frac{10 \cdot 9 \cdot \ldots \cdot 2 \cdot 1}{7 \cdot 6 \cdot \ldots \cdot 2 \cdot 1} \\ &= 10 \cdot 9 \cdot 8 \\ &= 720 \end{aligned}$$

14. How to cite

Please cite as:
Taboga, Marco (2021). "k-permutations", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/mathematical-tools/kpermutations.

Combinations

by Marco Taboga, PhD

This lecture introduces combinations, one of the most important concepts in combinatorial analysis. Before reading this lecture, you should be familiar with the concept of permutation.

We first deal with combinations without repetition and then with combinations with repetition.

1. :=

Table of contents

2. Combinations without repetition

A combination without repetition of k objects from n is a way of selecting k objects from a list of n. The selection rules are:

1. the order of selection does not matter (the same objects selected in different orders are regarded as the same combination);
2. each object can be selected only once.

A combination without repetition is also called a simple combination or, simply, a combination.
The following subsections give a slightly more formal definition of combination and deal with the problem of counting the number of possible combinations.

3. Definition of combination without repetition

Let $a_1, a_2, \ldots, a_n$ be n objects. A simple combination (or combination without repetition) of k objects from the n objects $a_1, a_2, \ldots, a_n$ is one of the possible ways to form a set containing k of the n objects. To form a valid set, any object can be chosen only once. Furthermore, the order in which the objects are chosen does not matter.

Example Consider three objects, a_1, a_2 and a_3. There are three possible combinations of two objects from a_1, a_2 and a_3, that is, three possible ways to choose two objects from this set of three:

Combination 1	a_1 and a_2
Combination 2	a_1 and a_3
Combination 3	a_2 and a_3

Other combinations are not possible, because, for example, $\{a_2, a_1\}$ is the same as $\{a_1, a_2\}$
.

4. Number of combinations without repetition

Denote by $C_{n,k}$ the number of possible combinations of k objects from n. How much is $C_{n,k}$ in general? In other words, how do we count the number of possible combinations of k objects from n ?

To answer this question, we need to recall the concepts of permutation and k-permutation introduced in previous lectures.

Like a combination, a k-permutation of n objects is one of the possible ways of choosing k of the n objects. However, in a k-permutation the order of selection matters: two k-permutations are regarded as different if the same k objects are chosen, but they are chosen in a different order. On the contrary, in the case of combinations, the order in which the k objects are chosen does not matter: two combinations that contain the same objects are regarded as equal.
number of possible combinations ($C_{n,k}$) from the number of possible k-permutations ($P_{n,k}$). Consider a combination of k objects from n. This combination will be repeated many times in the set of all possible k-permutations. It will be repeated one time for each possible way of ordering the k objects. So, it will be repeated $P_k = k!$ times (P_k is the number of all possible ways to order the k objects - the number of permutations of k objects). Therefore, if each combination is repeated P_k times in the set of all possible k-permutations, dividing the total number of k-permutations $(P_{n,k})$ by P_k, we obtain the number of possible combinations:

$$C_{n,k} = \frac{P_{n,k}}{P_k} = \frac{n!}{(n-k)!k!}$$

The number of possible combinations is often denoted by

$$C_{n,k} = \binom{n}{k}$$

and $\binom{n}{k}$ is called a binomial coefficient.
Example The number of possible combinations of 3 objects from 5 is

$$C_{5,3} = \frac{5!}{2!3!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(2 \cdot 1)(3 \cdot 2 \cdot 1)} = \frac{5 \cdot 4}{2 \cdot 1} = 10$$

5. Combinations with repetition

A combination with repetition of k objects from n is a way of selecting k objects from a list of n. The selection rules are:

1. the order of selection does not matter (the same objects selected in different orders are regarded as the same combination);
2. each object can be selected more than once.

Thus, the difference between simple combinations and combinations with repetition is that objects can be selected only once in the former, while they can be selected more than once in the latter.

The following subsections give a slightly more formal definition of combination with repetition and deal with the problem of counting the number of possible combinations with repetition.

6. Definition of combination with repetition

A more rigorous definition of combination with repetition involves the concept of multiset, which is a generalization of the notion of set (see the lecture entitled Set theory). Roughly speaking, the difference between a multiset and a set is the following: the same object is allowed to appear more than once in the list of members of a multiset, while the same object is allowed to appear only once in the list of members of an ordinary set. Thus, for example, the collection of objects

$$\{a, b, c, a\}$$

is a valid multiset, but not a valid set, because the letter a appears more than once. Like sets, multisets are unordered collections of objects, i.e. the order in which the elements of a multiset are

7. listed does not matter.

Let $a_1, a_2, \ldots, a_n$ be n objects. A combination with repetition of k objects from the n objects $a_1, a_2, \ldots, a_n$ is one of the possible ways to form a multiset containing k objects taken from the set $\{a_1, \ldots, a_n\}$.

Example Consider three objects, a_1, a_2 and a_3. There are six possible combinations with repetition of two objects from a_1, a_2 and a_3, that is, six possible ways to choose two objects from this set of three, allowing for repetitions:

Combination 1	a_1 and a_2
Combination 2	a_1 and a_3
Combination 3	a_2 and a_3
Combination 4	a_1 and a_1
Combination 5	a_2 and a_2
Combination 6	a_3 and a_3

Other combinations are not possible, because, for example, $\{a_2, a_1\}$ is the same as $\{a_1, a_2\}$.

8. Number of combinations with repetition

Denote by $C'_{n,k}$ the number of possible combinations with repetition of k objects from n. How much is $C'_{n,k}$ in general? In other words, how do we count the number of possible combinations with repetition of k objects from n ?

To answer this question, we need to use a slightly unusual procedure, which is introduced by the next example.

Example We need to order two scoops of ice cream, choosing among four flavours: chocolate, pistachio, strawberry and vanilla. It is possible to order two scoops of the same flavour. How many different combinations can we order? The number of different combinations we can order is equal to the number of possible combinations with repetition of 2 objects from 4 . Let us represent an order as a string of crosses (x) and vertical bars ($|$), where a vertical bar delimits two adjacent flavours and a cross denotes a scoop of a given flavour. For example,

```
* | | \ chocolate, 1 vanilla
\| \| \ \| \times 1 \text { strawberry, 1 vanilla}
x }\times| |2\mathrm{ chocolate
\| \| \times \times \| 2 \text { strawberry}
```

where the first vertical bar (the leftmost one) delimits chocolate and pistachio, the second one delimits pistachio and strawberry and the third one delimits strawberry and vanilla. Each string contains three vertical bars, one less than the number of flavours, and two crosses, one for each scoop. Therefore, each string contains a total of five symbols. Making an order is equivalent to choosing which two of the five symbols will be a cross (the remaining will be vertical bars). So, to make an order, we need to choose 2 objects from 5 . The number of possible ways to choose 2 objects from 5 is equal to the number of possible combinations without repetition of 2 objects from
5. Therefore, there are

$$\binom{5}{2} = \frac{5!}{(5-2)!2!} = 10$$

In general, choosing k objects from n with repetition is equivalent to writing a string with $n+k-1$ symbols, of which $n-1$ are vertical bars (I) and k are crosses ($\times$). In turn, this is equivalent to choose the k positions in the string (among the available $n+k-1$) that will contain a cross (the remaining ones will contain vertical bars). But choosing k positions from $n+k-1$ is like choosing a combination without repetition of k objects from $n+k-1$. Therefore, the number of possible combinations with repetition is

$$\begin{aligned}
C'_{n,k} &= C_{n+k-1,k} \\
&= \binom{n+k-1}{k} \\
&= \frac{(n+k-1)!}{(n+k-1-k)!k!} \\
&= \frac{(n+k-1)!}{(n-1)!k!}
\end{aligned}$$

The number of possible combinations with repetition is often denoted by

$$C'_{n,k} = \left(\binom{n}{k}\right)$$

and $\left(\binom{n}{k}\right)$ is called a multiset coefficient.
Example The number of possible combinations with repetition of 3 objects from 5 is

$$\begin{aligned}
C'_{5,3} &= \frac{(5+3-1)!}{(5-1)!3!} \\
&= \frac{7!}{4!3!} \\
&= \frac{7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{(4\cdot 3\cdot 2\cdot 1)(3\cdot 2\cdot 1)} \\
&= \frac{7\cdot 6\cdot 5}{3\cdot 2\cdot 1} \\
&= 7\cdot 5 = 35
\end{aligned}$$

9. More details

The following sections contain more details about combinations.

10. Binomial coefficients and binomial expansions

The binomial coefficient is so called because it appears in the binomial expansion:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

where $n \in \mathbb{N}$.

11. Recursive formula for binomial coefficients

The following is a useful recursive formula for computing binomial coefficients:

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$$

Proof

It is proved as follows:

$$\begin{aligned}
\binom{n}{k} + \binom{n}{k-1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n+1-k)!} \\
&= \frac{n!}{(k-1)!(n-k)!k} + \frac{n!}{(k-1)!(n-k)!(n+1-k)} \\
&= \frac{n!(n+1-k+k)}{(k-1)!(n-k)!k(n+1-k)} \\
&= \frac{n!(n+1)}{k!(n+1-k)!} \\
&= \frac{(n+1)!}{k!(n+1-k)!} \\
&= \binom{n+1}{k}
\end{aligned}$$

12. Solved exercises

Below you can find some exercises with explained solutions.

13. Exercise 1

3 cards are drawn from a standard deck of 52 cards. How many different 3-card hands can possibly be drawn?

14. Solution

First of all, the order in which the 3 cards are drawn does not matter (the same cards drawn in different orders are regarded as the same 3-card hand). Furthermore, each card can be drawn only once. Therefore the number of different 3-card hands that can possibly be drawn is equal to the number of possible combinations without repetition of 3 objects from 52 . If we denote it by $C_{52,3}$, then

$$
\begin{aligned}
C_{52,3} &= \binom{52}{3} \\
&= \frac{52!}{(52-3)!3!} \\
&= \frac{52!}{49!3!} \\
&= \frac{52 \cdot 51 \cdot 50}{3!} \\
&= \frac{52 \cdot 51 \cdot 50}{3 \cdot 2 \cdot 1} \\
&= 22100
\end{aligned}
$$

15. Exercise 2

John has got 1 dollar, with which he can buy green, red and yellow candies. Each candy costs 50 cents. John will spend all the money he has on candies. How many different combinations of green,
red and yellow candies can he buy?

16. Solution

First of all, the order in which the 3 different colours are chosen does not matter. Furthermore, each colour can be chosen more than once. Therefore, the number of different combinations of coloured candies John can choose is equal to the number of possible combinations with repetition of 2 objects from 3. If we denote it by $C'_{3,2}$, then

$$
\begin{aligned}
C'_{3,2} &= \left(\binom{3}{2}\right) \\
&= \binom{3+2-1}{2} \\
&= \binom{4}{2} \\
&= \frac{4!}{(4-2)!2!} \\
&= \frac{4!}{2!2!} \\
&= \frac{4 \cdot 3}{2!} \\
&= \frac{4 \cdot 3}{2 \cdot 1} \\
&= 6
\end{aligned}
$$

17. Exercise 3

The board of directors of a corporation comprises 10 members. An executive board, formed by 4 directors, needs to be elected. How many possible ways are there to form the executive board?

18. Solution

First of all, the order in which the 4 directors are selected does not matter. Furthermore, each director can be elected to the executive board only once. Therefore, the number of different ways to form the executive board is equal to the number of possible combinations without repetition of 4 objects from 10 . If we denote it by $C_{10,4}$, then

$$\begin{aligned}
C_{10,4} &= \binom{10}{4} \\
&= \frac{10!}{(10-4)!4!} \\
&= \frac{10!}{6!4!} \\
&= \frac{10 \cdot 9 \cdot 8 \cdot 7}{4!} \\
&= \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1} \\
&= 210
\end{aligned}$$

19. How to cite

20. Please cite as:

Taboga, Marco (2021). "Combinations", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/mathematicaltools/combinations.

Partitions into groups

by Marco Taboga, PhD

A partition of n objects into k groups is one of the possible ways of subdividing the n objects into k groups $(k \leq n)$. The rules are:

1. the order in which objects are assigned to a group does not matter;
2. each object can be assigned to only one group.

The following subsections give a slightly more formal definition of partition into groups and deal with the problem of counting the number of possible partitions into groups.

Table of contents

1. Definition of partition into groups

Let $a_1, a_2, \ldots, a_n$ be n objects. Let $g_1, g_2, \ldots, g_k$ be k (with $k \leq n$) groups to which we can assign the n objects. n_1 objects can be assigned to group g_1, n_2 objects can be assigned to group g_2 and so on. $n_1, n_2, \ldots, n_k$ are such that:

$$n_1 + n_2 + \ldots + n_k = n$$

A partition of $a_1, a_2, \ldots, a_n$ into the k groups $g_1, g_2, \ldots, g_k$ is one of the possible ways to assign the n objects to the k groups.

Example Consider three objects a_1, a_2 and a_3 and two groups g_1 and g_2, with

$$n_1 = 2$$
$$n_2 = 1$$

There are three possible partitions of the three objects into the two groups:

Groups	g_1	g_2
Partition 1	$\{a_1, a_2\}$	a_3
Partition 2	$\{a_1, a_3\}$	a_2
Partition 3	$\{a_2, a_3\}$	a_1

Note that the order of objects belonging to a group does not matter, so, for example, $\{a_1, a_2\}$ in Partition 1 is the same as $\{a_2, a_1\}$.

2. Counting the number of partitions into groups

Denote by $P_{n_1,n_2,\ldots,n_k}$ the number of possible partitions into the k groups (where group i contains n_i objects). How much is $P_{n_1,n_2,\ldots,n_k}$ in general?

The number $P_{n_1,n_2,\ldots,n_k}$ can be derived with the following sequential procedure:

1. First, we assign n_1 objects to the first group. There is a total of n objects to choose from. The number of possible ways to choose n_1 of the n objects is equal to the number of combinations of n_1 elements from n. So there are

$$\binom{n}{n_1} = \frac{n!}{n_1!(n-n_1)!} \text{ possible ways to form the first group}$$

2. Then, we assign n_2 objects to the second group. There were n objects, but n_1 have already been assigned to the first group. So, there are $n - n_1$ objects left, that can be assigned to the second group. The number of possible ways to choose n_2 of the remaining $n - n_1$ objects is equal to the number of combinations of n_2 elements from $n - n_1$. So there are

$$\binom{n-n_1}{n_2} = \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \text{ possible ways to form the second group}$$

$$\begin{aligned}\binom{n}{n_1}\binom{n-n_1}{n_2} &= \frac{n!}{n_1!(n-n_1)!}\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \\ &= \frac{n!}{n_1!n_2!(n-n_1-n_2)!} \text{ possible ways to form the first two groups}\end{aligned}$$

3. Then, we assign n_3 objects to the third group. There were n objects, but $n_1 + n_2$ have already been assigned to the first two groups. So, there are $n - n_1 - n_2$ objects left, that can be assigned to the third group. The number of possible ways to choose n_3 of the remaining $n - n_1 - n_2$ objects is equal to the number of combinations of n_3 elements from $n - n_1 - n_2$. So there are

$$\binom{n-n_1-n_2}{n_3} = \frac{(n-n_1-n_2)!}{n_3!(n-n_1-n_2-n_3)!} \text{ possible ways to form the third group}$$

and

$$\begin{aligned}&\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3} \\ =&\frac{n!}{n_1!n_2!(n-n_1-n_2)!}\frac{(n-n_1-n_2)!}{n_3!(n-n_1-n_2-n_3)!} \\ =&\frac{n!}{n_1!n_2!n_3!(n-n_1-n_2-n_3)!} \text{ possible ways to form the first three groups}\end{aligned}$$

4. An so on, until we are left with n_k objects and the last group. There is only one way to form the last group, which can also be written as

$$\binom{n-n_1-n_2-\ldots-n_{k-1}}{n_k} = \frac{(n-n_1-n_2-\ldots-n_{k-1})!}{n_k!(n-n_1-n_2-\ldots-n_k)!} \text{ possible ways to form the last group}$$

Therefore, there are

$$
\begin{aligned}
&\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdot\ldots\cdot\binom{n-n_1-n_2-\ldots-n_{k-1}}{n_k} \\
=&\frac{n!}{n_1!n_2!\cdot\ldots\cdot n_{k-1}!(n-n_1-n_2-\ldots-n_{k-1})!}\frac{(n-n_1-n_2-\ldots-n_{k-1})!}{n_k!(n-n_1-n_2-\ldots-n_k)!} \\
=&\frac{n!}{n_1!n_2!\cdot\ldots\cdot n_k!(n-n_1-n_2-\ldots-n_k)!} \\
=&\frac{n!}{n_1!n_2!\cdot\ldots\cdot n_k!0!} \\
=&\frac{n!}{n_1!n_2!\cdot\ldots\cdot n_k!} \text{ possible ways to form all the groups}
\end{aligned}
$$

Therefore, by the above sequential argument, the total number of possible partitions into the k groups is

$$P_{n_1,n_2,\ldots,n_k}=\frac{n!}{n_1!n_2!\cdot\ldots\cdot n_k!}$$

The number $P_{n_1,n_2,\ldots,n_k}$ is often indicated as follows:

$$P_{n_1,n_2,\ldots,n_k}=\binom{n}{n_1,n_2,\ldots,n_k}$$

and $\left(n^{n}_{n_1,n_2,\ldots,n_k}\right)$ is called a multinomial coefficient.
Sometimes the following notation is also used:

$$P_{n_1,n_2,\ldots,n_k}=(n_1,n_2,\ldots,n_k)!$$

Example The number of possible partitions of 4 objects into 2 groups of 2 objects is

$$P_{2,2}=\binom{4}{2,2}=\frac{4!}{2!2!}=\frac{4\cdot 3\cdot 2\cdot 1}{(2\cdot 1)(2\cdot 1)}=6$$

3. More details

4. Multinomial coefficients and multinomial expansions

The multinomial coefficient is so called because it appears in the multinomial expansion:
Index > Mathematical tools
where $n\in\mathbb{N}$ and the summation is over all the k-tuples $n_1,n_2,\ldots,n_k$ such that:

$$n_1+n_2+\ldots+n_k=n$$

5. Solved exercises

Below you can find some exercises with explained solutions.

6. Exercise 1

John has a basket of fruit containing one apple, one banana, one orange and one kiwi. He wants to give one fruit to each of his two little sisters and two fruits to his big brother. In how many different ways can he do this?

7. Solution

John needs to decide how to partition 4 objects into 3 groups, where the first two groups will contain one object and the third one will contain two objects. The total number of partitions is

$$\begin{aligned}
P_{1,1,2} &= \binom{4}{1,1,2} \\
&= \frac{4!}{1!!2!} \\
&= \frac{4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 1 \cdot 2 \cdot 1} \\
&= \frac{24}{2} = 12
\end{aligned}$$

8. Exercise 2

Ten friends want to play basketball. They need to divide into two teams of five players. In how many different ways can they do this?

9. Solution

They need to decide how to partition 10 objects into 2 groups, where each group will contain 5 objects. The total number of partitions is

$$\begin{aligned}
P_{5,5} &= \binom{10}{5,5} \\
&= \frac{10!}{5!5!} \\
&= \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)} \\
&= \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5 \cdot 4 \cdot 3 \cdot 7 \cdot 1} \\
&= \frac{9 \cdot 8 \cdot 7 \cdot 6}{4 \cdot 3} \\
&= 9 \cdot 2 \cdot 7 \cdot 2 = 252
\end{aligned}$$

Sequences

by Marco Taboga, PhD

Let A be a set of objects (e.g., real numbers, events, random variables). A sequence of elements of A is a function from the set of natural numbers $\mathbb{N}$ to the set A, i.e., a correspondence that associates one and only one element of A to each natural number $n \in \mathbb{N}$. In other words, a sequence of elements of A is an ordered list of elements of A, where the ordering is provided by the natural numbers.

A sequence is usually indicated by enclosing a generic element of the sequence in curly brackets:

$$\{a_n\}$$

where a_n is the n-th element of the sequence. Alternative notations are

$$\begin{aligned}
&\{a_n\}_{n-1}^{\infty} \\
&\{a_1, a_2, \ldots, a_n, \ldots\} \\
&a_n, n \in \mathbb{N} \\
&a_1, a_2, \ldots, a_n, \ldots
\end{aligned}$$

Thus, if $\{a_n\}$ is a sequence, a_1 is its first element, a_2 is its second element, a_n is its n-th element, and so on.

Example Define a sequence $\{a_n\}$ by characterizing its n-th element a_n as follows:

$$a_n = \frac{1}{n}$$

$\{a_n\}$ is a sequence of rational numbers. The elements of the sequence are $a_1 = 1, a_2 = \frac{1}{2}, a_3 = \frac{1}{3}$ $a_4 = \frac{1}{4}$ and so on.

Example Define a sequence $\{a_n\}$ by characterizing its n-th element a_n as follows:

$$a_n = \begin{cases} 1 \text{ if } n \text{ is even} \\ 0 \text{ if } n \text{ is odd} \end{cases}$$

$\{a_n\}$ is a sequence of 0 and 1 . The elements of the sequence are $a_1 = 0, a_2 = 1, a_3 = 0, a_4 = 1$ and so on.

Example Define a sequence $\{a_n\}$ by characterizing its n-th element a_n as follows:

$$a_n = \left[\frac{1}{n+1}, \frac{1}{n}\right]$$

$\{a_n\}$ is a sequence of closed subintervals of the interval $[0, 1]$. The elements of the sequence are $a_1 = \left[\frac{1}{2}, 1\right], a_2 = \left[\frac{1}{3}, \frac{1}{2}\right], a_3 = \left[\frac{1}{4}, \frac{1}{3}\right], a_4 = \left[\frac{1}{5}, \frac{1}{4}\right]$ and so on.

:

Table of contents

1. Countable and uncountable sets

Let A be a set of objects. A is a countable set if all its elements can be arranged into a sequence, i.e., if there exists a sequence $\{a_n\}$ such that
$\forall a \in A, \exists n \in \mathbb{N} : a_n = a$
In other words, A is a countable set if there exists at least one sequence $\{a_n\}$ such that every element of A belongs to the sequence. A is ancountable set if such a sequence does not exist. The most important example of an uncountable set is the set of real numbers $\mathbb{R}$.

2. Limit of a sequence

The concept of limit of a sequence is discussed in the lecture entitled Limit of a sequence.

3. How to cite

Please cite as:
Taboga, Marco (2021). "Sequences", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/mathematical-tools/sequences.

Limit of a sequence

by Marco Taboga, PhD

In this lecture we introduce the notion of limit of a sequence $\{a_n\}$. We start from the simple case in which $\{a_n\}$ is a sequence of real numbers, then we deal with the general case in which $\{a_n\}$ can be a sequence of objects that are not necessarily real numbers.

Table of contents

The limit of a sequence of real numbers

We first give an informal definition and then a more formal definition of the limit of a sequence of real numbers.

Informal definition of the limit of a sequence of real numbers

Let $\{a_n\}$ be a sequence of real numbers. Let $n_0 \in \mathbb{N}$. Denote by $\{a_n\}_{n>n_0}$ a subsequence of $\{a_n\}$ obtained by dropping the first n_0 terms of $\{a_n\}$, i.e.,

$$\{a_n\}_{n\infty n_0} = \{a_{n_0+1}, a_{n_0+2}, a_{n_0+3}, \ldots\}$$

The following is an intuitive definition of limit of a sequence.

Definition (informal) Let a be a real number. We say that a is a limit of a sequence $\{a_n\}$ of real numbers if, by appropriately choosing n_0, the distance between a and any term of the subsequence $\{a_n\}_{nDn_0}$ can be made as close to zero as we like. If a is a limit of the sequence $\{a_n\}$, we say that the sequence $\{a_n\}$ is a convergent sequence and that it converges to a. We indicate the fact that a is a limit of $\{a_n\}$ by

$$a = \lim_{n\to\infty} a_n$$

Thus, a is a limit of $\{a_n\}$ if, by dropping a sufficiently high number of initial terms of $\{a_n\}$, we can make the remaining terms of $\{a_n\}$ as close to a as we like. Intuitively, a is a limit of

$\{a_n\}$ if a_n becomes closer and closer to a by letting n go to infinity.

Formal definition of the limit of a sequence of real numbers

The distance between two real numbers is the absolute value of their difference. For example, if $a \in \mathbb{R}$ and a_n is a term of a sequence $\{a_n\}$, the distance between a_n and a, denoted by $d(a_n, a)$, is

$$d(a_n, a) = |a_n - a|$$

By using the concept of distance, the above informal definition can be made rigorous. Definition (formal) Let $a \in \mathbb{R}$. We say that a is a limit of a sequence $\{a_n\}$ of real numbers if

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} : d(a_{n_2} a) < \varepsilon, \forall n > n_0$$

If a is a limit of the sequence $\{a_n\}$, we say that the sequence $\{a_n\}$ is a convergent sequence and that it converges to a. We indicate the fact that a is a limit of $\{a_n\}$ by

$$a = \lim_{n \to \infty} a_n$$

For those unfamiliar with the universal quantifiers $\forall$ (any) and $\exists$ (exists), the notation

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} : d(a_n, a) < \varepsilon, \forall n > n_0$$

reads as follows: "For any arbitrarily small number ε, there exists a natural number n_0 such that the distance between a_n and a is less than ε for all the terms a_n with $n > n_0$ ", which can also be restated as "For any arbitrarily small number ε, you can find a subsequence $\{a_n\}_{nnn_0}$ such that the distance between any term of the subsequence and a is less than ε " or as "By dropping a sufficiently high number of initial terms of $\{a_n\}$, you can make the remaining terms as close to a as you wish".

It is also possible to prove that a convergent sequence has a unique limit, i.e., if $\{a_n\}$ has a limit a, then a is the unique limit of $\{a_n\}$.

Example Define a sequence $\{a_n\}$ by characterizing its n-th element a_n as follows:

$$a_n = \frac{1}{n}$$

The elements of the sequence are $a_1 = 1, a_2 = \frac{1}{2}, a_3 = \frac{1}{3}, a_4 = \frac{1}{4}$ and so on. The higher n is, the smaller a_n is and the closer it gets to 0 . Therefore, intuitively, the limit of the sequence should be 0 :

It is straightforward to prove that 0 is indeed a limit of $\{a_n\}$ by using the above definition. Choose any $\varepsilon > 0$. We need to find an $n_0 \in \mathbb{N}$ such that all terms of the subsequence $\{a_n\}_{n \infty n_0}$ have distance from zero less than ε :

$$d(a_n, 0) < \varepsilon, \forall n > n_0$$

Note first that the distance between a generic term of the sequence a_n and 0 is

$$d(a_n, 0) = |a_n - 0| = |a_n| = a_n$$

where the last equality obtains from the fact that all the terms of the sequence are positive (hence they are equal to their absolute values). Therefore, we need to find an $n_0 \in \mathbb{N}$ such that all terms of the subsequence $\{a_n\}_{n\Delta n_0}$ satisfy

$$a_n < \varepsilon, \forall n > n_0$$

Since

$$a_n < a_{n_0}, \forall n > n_0$$

the condition

$$a_n < \varepsilon, \forall n > n_0$$

is satisfied if $a_{n_0} < \varepsilon$, which is equivalent to $\frac{1}{n_0} < \varepsilon$. Therefore, it suffices to pick any n_0 such that $n_0 > \frac{1}{\varepsilon}$ to satisfy the condition

$$d(a_n, 0) < \varepsilon, \forall n > n_0$$

In summary, we have just shown that, for any ε, we are able to find $n_0 \in \mathbb{N}$ such that all terms of the subsequence $\{a_n\}_{nDn_0}$ have distance from zero less than ε. As a consequence 0 is the limit of the sequence $\{a_n\}$.

The limit of a sequence in general

We now deal with the more general case in which the terms of the sequence $\{a_n\}$ are not necessarily real numbers. As before, we first give an informal definition, then a more formal one.

Informal definition of limit - The general case

Let A be a set of objects (e.g., real numbers, events, random variables) and let $\{a_n\}$ be a sequence of elements of A. The limit of $\{a_n\}$ is defined as follows.

Definition (informal) Let $a \in A$. We say that a is a limit of a sequence $\{a_n\}$ of elements of A, if, by appropriately choosing n_0, the distance between a and any term of the subsequence $\{a_n\}_{n\Delta n_0}$ can be made as close to zero as we like. If a is a limit of the sequence $\{a_n\}$, we say that the sequence $\{a_n\}$ is a convergent sequence and that it converges to a. We indicate the fact that a is a limit of $\{a_n\}$ by

The definition is the same we gave above, except for the fact that now both a and the terms of the sequence $\{a_n\}$ belong to a generic set of objects A.

Metrics and the definition of distance

In the definition above, we have implicitly assumed that the concept of distance between elements of A is well-defined. Thus, for the above definition to make any sense, we need to properly define distance.

We need a function $d : A \times A \rightarrow \mathbb{R}$ that associates to any couple of elements of A a real number measuring how far these two elements are. For example, if a and a' are two elements of $A, d(a, a')$ needs to be a real number measuring the distance between a and a'.

A function $d : A \times A \rightarrow \mathbb{R}$ is considered a valid distance function (and it is called a metric on A) if it satisfies some properties, listed in the next proposition.

Definition Let A be a set of objects. Let $d : A \times A \rightarrow \mathbb{R} \cdot d$ is considered a valid distance function (in which case it is called a metric on A) if, for any a, a' and a'' belonging to A :

1. non-negativity: $d(a, a') \geq 0$;
2. identity of indiscernibles: $d(a, a') = 0$ if and only if $a = a'$;
3. symmetry: $d(a, a') = d(a', a)$;
4. triangle inequality: $d(a, a') + d(a', a'') \geq d(a, a'')$.

All four properties are very intuitive: property 1) says that the distance between two points cannot be a negative number; property 2) says that the distance between two points is zero if and only if the two points coincide; property 3) says that the distance from a to a' is the same as the distance from a' to a; property 4) says (roughly speaking) that the distance you cover when you go from a to a'' directly is less than (or equal to) the distance you cover when you go from a to a'' passing from a third point a' (if a' is not on the way from a to a'' you are increasing the distance covered).

Example (Euclidean distance) Consider the set of K-dimensional real vectors $\mathbb{R}^K$. The metric usually employed to measure the distance between elements of $\mathbb{R}^K$ is the so-called Euclidean distance. If a and b are two vectors belonging to $\mathbb{R}^K$, then their Euclidean distance is

$$d(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \ldots + (a_K - b_K)^2}$$

where $a_{1,}, \ldots, a_K$ are the K components of a and $b_1, \ldots, b_K$ are the K components of b. It is possible to prove that the Euclidean distance satisfies all the four properties that a metric needs to satisfy.
Furthermore, when $K = 1$, it becomes

$$d(a, b) = \sqrt{(a - b)^2} = |a - b|$$

which coincides with the definition of distance between real numbers already given above.

Whenever we are faced with a sequence of objects and we want to assess whether it is convergent, we need to first define a distance function on the set of objects to which the terms of the sequence belong and verify that the proposed distance function satisfies all the properties of a proper distance function (a metric). For example, in probability theory and statistics, we often deal with sequences of random variables. To assess whether these sequences are convergent, we need to define a metric to measure the distance between two random variables. As we will see in the lecture entitled Sequences of random variables and their convergence, there are several ways of defining the concept of distance between two random variables. All these ways are legitimate and are useful in different situations.

Formal definition of limit - The general case

Having defined the concept of a metric, we are now ready to state the formal definition of a limit of a sequence.

Definition (formal) Let A be a set of objects. Let $d : A \times A \to \mathbb{R}$ be a metric on A. We say that $a \in A$ is a limit of a sequence $\{a_n\}$ of objects belonging to A if

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} : d(a_n, a) < \varepsilon, \forall n > n_0$$

If a is a limit of the sequence $\{a_n\}$, we say that the sequence $\{a_n\}$ is a convergent sequence and that it converges to a. We indicate the fact that a is a limit of $\{a_n\}$ by

$$a = \lim_{n\to\infty} a_n$$

Also in this case, it is possible to prove (see below) that a convergent sequence has a unique limit, i.e., if $\{a_n\}$ has a limit a, then a is the unique limit of $\{a_n\}$.

Proof
The proof is by contradiction. Suppose that a and a' are two limits of a sequence $\{a_n\}$ and $a \neq a'$. By combining property 1) and 2) of a metric (see above) it must be that

$$d(a, a') > 0$$

i.e., $d(a, a') = \bar{d}$ where $\bar{d}$ is a strictly positive constant. Pick any term a_n of the sequence. By property 4) of a metric (the triangle inequality), we have

$$d(a, a_n) + d(a_n, a') \geq d(a, a')$$

Considering that $d(a, a') = \bar{d}$, the previous inequality becomes

$$d(a_2, a_n) + d(a', a_n) \geq \bar{d} > 0$$

Now, take any $\varepsilon < \bar{d}$. Since a is a limit of the sequence, we can find n_0 such that $d(a, a_n) < \varepsilon, \forall n > n_0$, which means that

$$\varepsilon + d(a', a_n) \geq d(a_, , a_n) + d(a', a_n) \geq \bar{d} > 0, \forall n > n_0$$

and

$$d(a', a_n) \geq \bar{d} - \varepsilon > 0, \forall n > n_0$$

Therefore, $d(a', a_n)$ can not be made smaller than $\bar{d} - \varepsilon$ and as a consequence a' cannot be a limit of the sequence.

Convergence criterion

In practice, it is usually difficult to assess the convergence of a sequence using the above definition. Instead, convergence can be assessed using the following criterion.

Lemma (criterion for convergence) Let A be a set of objects. Let $d : A \times A \to \mathbb{R}$ be a metric on A. Let $\{a_n\}$ be a sequence of objects belonging to A and $a \in A$. $\{a_n\}$ converges to a if and only if

$$\lim_{n\to\infty} d(a_{n,a}a) = 0$$

Proof
This is easily proved by defining a sequence of real numbers $\{d_n\}$ whose generic term is

$$d_n = d(a_n, a)$$

and noting that the definition of convergence of $\{a_n\}$ to a, which is

$$\forall\varepsilon > 0, \exists n_0 \in \mathbb{N} : d(a_n, a) < \varepsilon, \forall n > n_0$$

can be written as

$$\forall\varepsilon > 0, \exists n_0 \in \mathbb{N} : |d_n - 0| < \varepsilon, \forall n > n_0$$

which is the definition of convergence of $\{d_n\}$ to 0 .

So, in practice, the problem of assessing the convergence of a generic sequence of objects is simplified as follows:

1. find a metric $d(a_n, a)$ to measure the distance between the terms of the sequence a_n and the candidate limit a ;
2. define a new sequence $\{d_n\}$, where $d_n = d(a_n, a)$;
3. study the convergence of the sequence $\{d_n\}$, which is a simple problem, because $\{d_n\}$ is a sequence of real numbers.

How to cite

Please cite as:
Taboga, Marco (2021). "Limit of a sequence", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/mathematicaltools/limit-of-a-sequence.

Derivatives - Review

by Marco Taboga, PhD

This review page contains a summary of differentiation rules, that is, of rules for computing the derivative of a function. If $f(x)$ is a function, its first derivative is denoted by $\frac{d}{dx}f(x)$.

$$:=$$

Derivative of a constant function

If $f(x)$ is a constant function

$$f(x) = c$$

where $c \in \mathbb{R}$, then its first derivative is

$$\frac{d}{dx}f(x) = 0$$

Derivative of a power function

If $f(x)$ is a power function

$$f(x) = x^n$$

then its first derivative is

$$\frac{d}{dx}f(x) = nx^{n-1}$$

where $n \in \mathbb{R}$ is a constant.

Derivative of a logarithmic function

If $f(x)$ is the natural logarithm of x, that is,

$$f(x) = \ln(x)$$

then its first derivative is

$$\frac{d}{dx}f(x) = \frac{1}{x}$$

If $f(x)$ is the logarithm to base b of x, that is,

$$f(x) = \log_b(x)$$

then its first derivative is

$$\frac{d}{dx}f(x) = \frac{1}{x\ln(b)}$$

(remember that $\log_b(x) = \frac{\ln(x)}{\ln(b)}$).

Derivative of an exponential function

If $f(x)$ is the exponential function

$$f(x) = \exp(x)$$

then its first derivative is

$$\frac{d}{dx}f(x) = \exp(x)$$

If the exponential function $f(x)$ does not have the natural base e, but another positive base b, that is, if

$$f(x) = b^x$$

then its first derivative is

$$\frac{d}{dx}f(x) = \ln(b)b^x$$

(remember that $b^x = \exp(x\ln(b))$).

Derivative of a linear combination of functions

If $f_1(x)$ and $f_2(x)$ are two functions and $c_1, c_2 \in \mathbb{R}$ are two constants, then

$$\frac{d}{dx}(c_1 f_1(x) + c_2 f_2(x)) = c_1 \frac{d}{dx} f_1(x) + c_2 \frac{d}{dx} f_2(x)$$

In other words, the derivative of a linear combination is equal to the linear combinations of the derivatives. This property is called "linearity of the derivative".

Two special cases of this rule are

$$\frac{d}{dx}(c_1 f_1(x)) = c_1 \frac{d}{dx} f_1(x)$$
$$\frac{d}{dx}(f_1(x) + f_2(x)) = \frac{d}{dx} f_1(x) + \frac{d}{dx} f_2(x)$$

Derivative of a product of functions

If $f_1(x)$ and $f_2(x)$ are two functions, then the derivative of their product is

$$\frac{d}{dx}(f_1(x)f_2(x)) = \left(\frac{d}{dx}f_1(x)\right)f_2(x) + f_1(x)\left(\frac{d}{dx}f_2(x)\right)$$

Derivative of a composition of functions (chain rule)

If $f(x)$ and $g(y)$ are two functions, then the derivative of their composition is

$$\frac{d}{dx}(g(f(x))) = \left(\frac{d}{dy}g(y)\Big|_{y=f(x)}\right)\frac{d}{dx}f(x)$$

What does this chain rule mean in practice? It means that first you need to compute the derivative of $g(y)$:

$$\frac{d}{dy}g(y)$$

Then, you substitute y with $f(x)$:

$$\frac{d}{dy}g(y)\Big|_{y-f(x)}$$

Finally, you multiply it by the derivative of $f(x)$:

$$\frac{d}{dx}f(x)$$

Derivatives of trigonometric functions

The trigonometric functions have the following derivatives:

$$\begin{aligned}
\frac{d}{dx}\sin(x) &= \cos(x) \\
\frac{d}{dx}\cos(x) &= -\sin(x) \\
\frac{d}{dx}\tan(x) &= \frac{1}{\cos^2(x)}
\end{aligned}$$

while the inverse trigonometric functions have the following derivatives:

$$\frac{d}{dx}\arcsin(x) = \frac{1}{\sqrt{1-x^2}}$$
$$\frac{d}{dx}\arccos(x) = -\frac{1}{\sqrt{1-x^2}}$$
$$\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2}$$

Derivative of an inverse function

If $y = f(x)$ is a function with derivative

$$\frac{d}{dx}f(x)$$

then its inverse $x = f^{-1}(y)$ has derivative

$$\frac{d}{dy}f^{-1}(y) = \left(\left.\frac{d}{dx}f(x)\right|_{x-f^{-1}(y)}\right)^{-1}$$

How to cite

Please cite as:
Taboga, Marco (2021). "Derivatives - Review", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/mathematicaltools/derivatives-review.

Integrals - Review

by Marco Taboga, PhD

This review page contains a summary of integration rules, that is, of rules for computing definite and indefinite integrals of a function.

Table of contents

Indefinite integrals

If $f(x)$ is a function of one variable, an indefinite integral of $f(x)$ is a function $F(x)$ whose first derivative is equal to $f(x)$:

$$\frac{d}{dx}F(x) = f(x)$$

An indefinite integral $F(x)$ is denoted by

$$F(x) = \int f(x)dx$$

Indefinite integrals are also called antiderivatives or primitives.

Example Let

$$f(x) = x^3$$

The function

$$F(x) = \frac{1}{4}x^4$$

is an indefinite integral of $f(x)$ because

$$\begin{aligned}\frac{d}{dx}F(x) &= \frac{d}{dx}\left(\frac{1}{4}x^4\right) \\ &= \frac{1}{4}\frac{d}{dx}\left(x^4\right) \\ &= \frac{1}{4}4x^3 \\ &= x^3\end{aligned}$$

Also the function

$$G(x) = \frac{1}{2} + \frac{1}{4}x^4$$

is an indefinite integral of $f(x)$ because

$$\begin{aligned}\frac{d}{dx}G(x) &= \frac{d}{dx}\left(\frac{1}{2} + \frac{1}{4}x^4\right) \\ &= \frac{d}{dx}\left(\frac{1}{2}\right) + \frac{1}{4}\frac{d}{dx}\left(x^4\right) \\ &= 0 + \frac{1}{4}4x^3 \\ &= x^3\end{aligned}$$

Note that if a function $F(x)$ is an indefinite integral of $f(x)$ then also the function

$$G(x) = F(x) + c$$

is an indefinite integral of $f(x)$ for any constant $c \in \mathbb{R}$ because

$$\begin{aligned}\frac{d}{dx}G(x) &= \frac{d}{dx}(F(x) + c) \\ &= \frac{d}{dx}(F(x)) + \frac{d}{dx}(c) \\ &= f(x) + 0\end{aligned}$$

This is also the reason why the adjective indefinite is used: because indefinite integrals are defined only up to a constant.

The following subsections contain some rules for computing the indefinite integrals of functions that are frequently encountered in probability theory and statistics. In all these subsections, c will denote a constant and the integration rules will be reported without a proof. Proofs are trivial and can be easily performed by the reader: it suffices to compute the first derivative of $F(x)$ and verify that it equals $f(x)$.

Indefinite integral of a constant function

If $f(x)$ is a constant function

$$f(x) = a$$

where $a \in \mathbb{R}$, then an indefinite integral of $f(x)$ is

$$F(x) = ax + c$$

Indefinite integral of a power function

If $f(x)$ is a power function

$$f(x) = x^n$$

then an indefinite integral of $f(x)$ is

$$F(x) = \frac{1}{n+1} x^{n+1} + c$$

when $n \neq -1$. When $n = -1$, that is, when

$$f(x) = \frac{1}{x}$$

the integral is

$$F(x) = \ln(x) + c$$

Indefinite integral of a logarithmic function

If $f(x)$ is the natural logarithm of x, that is,

$$f(x) = \ln(x)$$

then its indefinite integral is

$$F(x) = x\ln(x) - x + c$$

If $f(x)$ is the logarithm to base b of x, that is,

$$f(x) = \log_b(x)$$

then its indefinite integral is
(remember that $\log_b(x) = \frac{\ln(x)}{\ln(b)}$.

Indefinite integral of an exponential function

If $f(x)$ is the exponential function

$$f(x) = \exp(x)$$

then its indefinite integral is

$$F(x) = \exp(x) + c$$

If the exponential function $f(x)$ does not have the natural base e, but another positive base b, that is,

$$f(x) = b^x$$

then its indefinite integral is

$$F(x) = \frac{1}{\ln(b)} b^x + c$$

(remember that $b^x = \exp(x \ln(b))$).

Indefinite integral of a linear combination of functions

If $f_1(x)$ and $f_2(x)$ are two functions and $c_1, c_2 \in \mathbb{R}$ are two constants, then

$$\int (c_1 f_1(x) + c_2 f_2(x)) dx = c_1 \int f_1(x) dx + c_2 \int f_2(x) dx$$

In other words, the integral of a linear combination is equal to the linear combinations of the integrals. This property is called "linearity of the integral".

Two special cases of this rule are

$$\int c_1 f_1(x) dx = c_1 \int f_1(x) dx$$
$$\int (f_1(x) + f_2(x)) dx = \int f_1(x) dx + \int f_2(x) dx$$

Indefinite integrals of trigonometric functions

The trigonometric functions have the following indefinite integrals:

$$\int \sin(x) dx = -\cos(x) + c$$
$$\int \cos(x) dx = \sin(x) + c$$
$$\int \tan(x) dx = \ln\left(\left|\frac{1}{\cos(x)}\right|\right) + c$$

Let $f(x)$ be a function of one variable and $[a, b]$ an interval of real numbers. The definite integral (or, simply, the integral) from a to b of $f(x)$ is the area of the region in the xy-plane bounded by the graph of $f(x)$, the x-axis and the vertical lines $x = a$ and $x = b$, where regions below the x-axis have negative sign and regions above the x-axis have positive sign.

The integral from a to b of $f(x)$ is denoted by

$$\int_a^b f(x) dx$$

$f(x)$ is called the integrand function and a and b are called upper and lower bound of integration. The following subsections contain some properties of definite integrals, which are also often utilized to actually compute definite integrals.

Fundamental theorem of calculus

The fundamental theorem of calculus provides the link between definite and indefinite integrals. It has two parts.

On the one hand, if you define

$$F(x) = \int_a^x f(t)dt$$

then, the first derivative of $F(x)$ is equal to $f(x)$, that is,

$$\frac{d}{dx}F(x) = f(x)$$

In other words, if you differentiate a definite integral with respect to its upper bound of integration, then you obtain the integrand function.

Example Define

$$F(x) = \int_a^x \exp(2t)dt$$

Then,

$$\frac{d}{dx}F(x) = \exp(2x)$$

On the other hand, if $F(x)$ is an indefinite integral (an antiderivative) of $f(x)$, then

$$\int_a^b f(x)dx = F(b) - F(a)$$

In other words, you can use the indefinite integral to compute the definite integral.
The following notation is often used:

$$\int_a^b f(x)dx = [F(x)]_a^b$$

where

$$[F(x)]_a^b = F(b) - F(a)$$

Sometimes the variable of integration x is explicitly specified and we write

$$[F(x)]_{x-a}^{x_x-b}$$

Example Consider the definite integral

$$\int_0^1 x^2 dx$$

The integrand function is

$$f(x) = x^2$$

An indefinite integral of $f(x)$ is

$$F(x) = \frac{1}{3}x^3$$

Therefore, the definite integral from 0 to 1 can be computed as follows.

$$\begin{aligned}\int_0^1 x^2 dx &= \left[\frac{1}{3}x^3\right]_0^1 \\ &= \frac{1}{3} \cdot 1^3 - \frac{1}{3} \cdot 0^3 \\ &= \frac{1}{3}\end{aligned}$$

Definite integral of a linear combination of functions

Like indefinite integrals, also definite integrals are linear. If $f_1(x)$ and $f_2(x)$ are two functions and $c_1, c_2 \in \mathbb{R}$ are two constants, then

$$\int_a^b (c_1 f_1(x) + c_2 f_2(x)) dx = c_1 \int_a^b f_1(x) dx + c_2 \int_a^b f_2(x) dx$$

with the two special cases

$$\begin{aligned}\textstyle\int_a^b c_1 f_1(x) dx &= c_1 \textstyle\int_a^b f_1(x) dx && \text{(multiplica)} \\ \textstyle\int_a^b (f_1(x) + f_2(x)) dx &= \textstyle\int_a^b f_1(x) dx + \textstyle\int_a^b f_2(x) dx && \text{(addition)}\end{aligned}$$

Example For example,

$$\int_0^1 \left(3x + 2x^2\right) dx = 3 \int_0^1 x dx + 2 \int_0^1 x^2 dx$$

Change of variable

If $f(x)$ and $g(x)$ are two functions, then the integral

$$\int_a^b f(g(x)) \left(\frac{d}{dx} g(x)\right) dx$$

can be computed by a change of variable, with the variable

The change of variable is performed in the following steps:

1. Differentiate the change of variable formula

$$t = g(x)$$

and obtain

$$dt = \frac{d}{dx} g(x) dx$$

2. Recompute the bounds of integration:

$$\begin{aligned}x = a &\Rightarrow t = g(x) = g(a) \\ x = b &\Rightarrow t = g(x) = g(b)\end{aligned}$$

3. Substitute $g(x)$ and $\frac{d}{dx}g(x)dx$ in the integral:

$$\int_a^b f(g(x))\left(\frac{d}{dx}g(x)\right)dx = \int_{g(a)}^{g(b)} f(t)dt$$

Example The integral

$$\int_1^2 \frac{\ln(x)}{x}dx$$

can be computed performing the change of variable

$$t = \ln(x)$$

By differentiating the change of variable formula, we obtain

$$dt = \frac{d}{dx}\ln(x)dx = \frac{1}{x}dx$$

The new bounds of integration are

$$\begin{aligned} x = 1 &\Rightarrow t = \ln(1) = 0 \\ x = 2 &\Rightarrow t = \ln(2) \end{aligned}$$

Therefore the integral can be written as follows:

$$\int_1^2 \frac{\ln(x)}{x}dx = \int_0^{\ln(2)} tdt$$

Integration by parts

Let $f(x)$ and $g(x)$ be two functions and $F(x)$ and $G(x)$ their indefinite integrals. The following integration by parts formula holds:

$$\int_a^b f(x)G(x)dx = [F(x)G(x)]_a^b - \int_a^b F(x)g(x)dx$$

Example The integral

$$\int_0^1 \exp(x)xdx$$

can be integrated by parts, by setting

An indefinite integral of $f(x)$ is

$$F(x) = \exp(x)$$

and $G(x)$ is an indefinite integral of

$$g(x) = 1$$

or, said differently, $g(x) = 1$ is the derivative of $G(x) = x$. Therefore,

$$\begin{aligned}\int_0^1 \exp(x)x dx &= [\exp(x)x]_0^1 - \int_0^1 \exp(x)dx \\ &= \exp(1) - 0 - \int_0^1 \exp(x)dx \\ &= \exp(1) - [\exp(x)]_0^1 \\ &= \exp(1) - [\exp(1) - 1] \\ &= 1\end{aligned}$$

Exchanging the bounds of integration

Given the integral

$$\int_a^b f(x)dx$$

exchanging its bounds of integration is equivalent to changing its sign:

$$\int_b^a f(x)dx = -\int_a^b f(x)dx$$

Subdividing the integral

Given the two bounds of integration a and b, with $a \leq b$, and a third point m such that $a \leq m \leq b$, then

$$\int_a^b f(x)dx = \int_a^m f(x)dx + \int_m^b f(x)dx$$

Leibniz integral rule

Given a function of two variables $f(x,y)$ and the integral

$$I(y) = \int_{a(y)}^{b(y)} f(x,y)dx$$

where both the lower bound of integration a and the upper bound of integration b may depend on y, under appropriate technical conditions (not discussed here) the first derivative of the function $I(y)$ with respect to y can be computed as follows:

$$\frac{d}{dy}I(y) = \left(\frac{d}{dy}b(y)\right)f(b(y),y) - \left(\frac{d}{dy}a(y)\right)f(a(y),y) + \int_{a(y)}^{b(y)} \frac{\partial}{\partial y}f(x,y)dx$$

where $\frac{\partial}{\partial y}f(x,y)$ is the first partial derivative of $f(x,y)$ with respect to y.

Example The derivative of the integral

$$I(y) = \int_{y^2}^{y^2+1} \exp(xy)dx$$

is

$$
\begin{aligned}
\frac{d}{dy}I(y) &= \left(\frac{d}{dy}\left(y^2+1\right)\right)\exp\left(\left(y^2+1\right)y\right) - \left(\frac{d}{dy}y^2\right)\exp\left(y^2 y\right) + \int_{y^2}^{y^2+1}\frac{\partial}{\partial y}(\exp(xy))dx \\
&= 2y\exp\left(y^3+y\right) - 2y\exp\left(y^3\right) + \int_{y^2}^{y^2+1} x\exp(xy)dx
\end{aligned}
$$

Solved exercises

Below you can find some exercises with explained solutions.

Exercise 1

Compute the following integral:

$$
\int_0^\infty \cos(x)\exp(-x)dx
$$

Hint: perform two integrations by parts.

Solution

By performing two integrations by parts, we obtain

$$
\begin{aligned}
&\int_0^\infty \cos(x)\exp(-x)dx \\
&= [\sin(x)\exp(-x)]_0^\infty - \int_0^\infty \sin(x)(-\exp(-x))dx \\
&= 0 - 0 + \int_0^\infty \sin(x)\exp(-x)dx \\
&= [-\cos(x)\exp(-x)]_0^\infty - \int_0^\infty (-\cos(x))(-\exp(-x))dx \\
&= 0 - (-1) - \int_0^\infty \cos(x)\exp(-x)dx \\
&= 1 - \int_0^\infty \cos(x)\exp(-x)dx \\
&\text{(integrating by parts)}
\end{aligned}
$$

Therefore,

$$
\int_0^\infty \cos(x)\exp(-x)dx = 1 - \int_0^\infty \cos(x)\exp(-x)dx
$$

which can be rearranged to yield

$$
2\int_0^\infty \cos(x)\exp(-x)dx = 1
$$

or

$$
\int_0^\infty \cos(x)\exp(-x)dx = \frac{1}{2}
$$

Use Leibniz integral rule to compute the derivative with respect to y of the following integral:

$$I(y) = \int_0^{y^2} \exp(-xy)dx$$

Solution

Leibniz integral rule is

$$\frac{d}{dy}\int_{a(v)}^{b(v)} f(x,y)dx = \left(\frac{d}{dy}b(y)\right)f(b(y),y) - \left(\frac{d}{dy}a(y)\right)f(a(y),y) + \int_{a(y)}^{b(y)} \frac{\partial}{\partial y}f(x,y)dx$$

We can apply it as follows:

$$\begin{aligned}
\frac{d}{dy}I(y) &= \frac{d}{dy}\int_0^{y^2} \exp(-xy)dx \\
&= \left(\frac{d}{dy}y^2\right)\exp\left(-y^2 y\right) + \int_0^{y^2} \frac{\partial}{\partial y}\exp(-xy)dx \\
&= 2y\exp\left(-y^3\right) - \int_0^{y^2} x\exp(-xy)dx \\
&= 2y\exp\left(-y^3\right) - \left\{\left[x\left(-\frac{1}{y}\exp(-xy)\right)\right]_0^{y^2} + \frac{1}{y}\int_0^{y^2} \exp(-xy)dx\right\} \qquad \text{(integrating by parts)} \\
&= 2y\exp\left(-y^3\right) + y\exp\left(-y^3\right) - \frac{1}{y}\left[-\frac{1}{y}\exp(-xy)\right]_0^{y^2} \\
&= 3y\exp\left(-y^3\right) + \frac{1}{y^2}\exp\left(-y^3\right) - \frac{1}{y^2}
\end{aligned}$$

Exercise 3

Compute the following integral:

$$\int_0^1 x\left(1+x^2\right)^{-2}dx$$

Solution

This integral can be solved by using the change of variable technique:

$$\begin{aligned}
&\int_0^1 x\left(1+x^2\right)^{-2}dx \\
=&\int_0^1 \frac{1}{2}(1+t)^{-2}dt \qquad \left(\text{by changing variable: } t = x^2\right) \\
=&\left[-\frac{1}{2}(1+t)^{-1}\right]_0^1 \\
=&-\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot 1 = \frac{1}{4}
\end{aligned}$$

How to cite

Please cite as:

Taboga, Marco (2021). "Integrals - Review", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/mathematicaltools/integrals-review.

Gamma function

by Marco Taboga, PhD

The Gamma function is a generalization of the factorial function to non-integer numbers. It is often used in probability and statistics, as it shows up in the normalizing constants of important

In this lecture we define the Gamma function, we present and prove some of its properties, and we discuss how to calculate its values.

$:=$

Table of contents

Introduction and motivation

Recall that, if $n \in \mathbb{N}$, its factorial n ! is

$$n! = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n$$

so that n ! satisfies the following recursion:

$$n! = (n-1)! \cdot n$$

The Gamma function $\Gamma(z)$ satisfies a similar recursion:

$$\Gamma(z) = \Gamma(z-1) \cdot (z-1)$$

but it is defined also when z is not an integer.

The following is a possible definition of the Gamma function.
Definition The Gamma function Γ is a function $\Gamma : \mathbb{R}_{++} \to \mathbb{R}_{++}$ satisfying the following

equation:

$$\Gamma(z) = \int_0^\infty x^{x^{z-1}} \exp(-x) dx$$

The domain of definition of the Gamma function can be extended beyond the set $\mathbb{R}_{++}$ of strictly positive real numbers (for example to complex numbers).

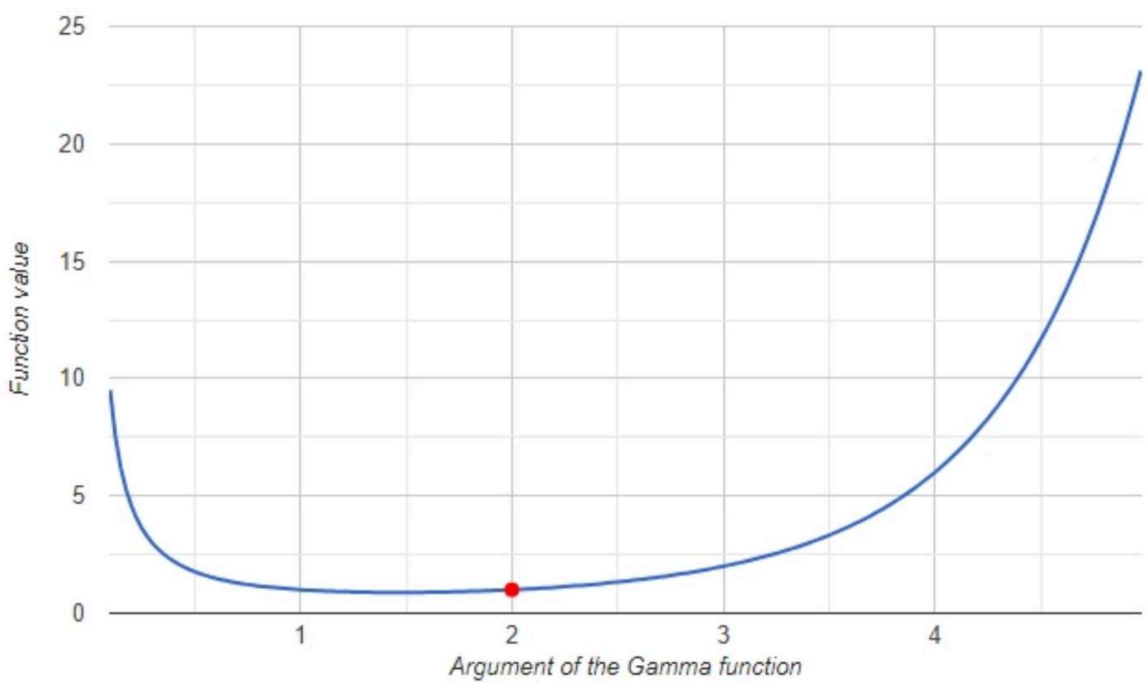
However, the somewhat restrictive definition given above is sufficient to address the great majority of statistics problems that involve the Gamma function.

How to compute the values of the function

We will show below some special cases in which the value of the Gamma function can be derived analytically.

However, in general, it is not possible to express $\Gamma(z)$ in terms of elementary functions for every z.
As a consequence, one often needs to resort to numerical algorithms to compute $\Gamma(z)$.
We include here a calculator that implements one of these algorithms and we refer the reader to Abramowitz and Stegun (1965) for a thorough discussion of the main methods to compute numerical approximations of $\Gamma(z)$.

Plot of the gamma function with interactive calculator



0
Argument z

If you play with the calculator, you will notice several properties of the Gamma function:

- it tends to infinity as z approaches 0 ;
- it quickly tends to infinity as z increases;
- for large values of $z, \Gamma(z)$ is so large that an overflow occurs: the true value of $\Gamma(z)$ is replaced by infinity; however, we are still able to correctly store the natural logarithm of $\Gamma(z)$ in the computer memory.

The last point has great practical relevance. When we manipulate quantities that depend on a value taken by the Gamma function, we should always work with logarithms.

Recursive formula

Given the above definition, it is straightforward to prove that the Gamma function satisfies the following recursion:

$$\Gamma(z) = \Gamma(z-1) \cdot (z-1)$$

Proof

The recursion can be derived by using integration by parts:

$$\begin{aligned}\Gamma(z) &= \int_0^\infty x^{z-1} \exp(-x) dx \\ &= \left[-x^{z-1} \exp(-x)\right]_0^\infty + \int_0^\infty (z-1) x^{z-2} \exp(-x) dx \\ &= (0-0) + (z-1) \int_0^\infty x^{(z-1)-1} \exp(-x) dx \\ &= (z-1) \Gamma(z-1)\end{aligned}$$

Relation to the factorial function

When the argument of the Gamma function is a natural number $n \in \mathbb{N}$ then its value is equal to the factorial of $n-1$:

$$\Gamma(n) = (n-1)!$$

Proof

First of all, we have that

$$\begin{aligned}\Gamma(1) &= \int_0^\infty x^{1-1} \exp(-x) dx \\ &= \int_0^\infty \exp(-x) dx \\ &= [-\exp(-x)]_0^\infty \\ &= 1\end{aligned}$$

Using the recursion $\Gamma(z) = \Gamma(z-1) \cdot (z-1)$, we obtain

$$
\begin{aligned}
\Gamma(1) =& 1 = 0! \\
\Gamma(2) =& \Gamma(2-1)\cdot(2-1) = \Gamma(1)\cdot 1 = 1 = 1! \\
\Gamma(3) =& \Gamma(3-1)\cdot(3-1) = \Gamma(2)\cdot 2 = 1\cdot 2 = 2! \\
\Gamma(4) =& \Gamma(4-1)\cdot(4-1) = \Gamma(3)\cdot 3 = 1\cdot 2\cdot 3 = 3! \\
& \vdots \\
\Gamma(n) =& \Gamma(n-1)\cdot(n-1) = 1\cdot 2\cdot 3\cdot\ldots\cdot(n-1) = (n-1)!
\end{aligned}
$$

Some known values

A well-known formula, which is often used in probability theory and statistics, is the following:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Proof

By using the definition and performing a change of variable, we obtain

$$
\begin{aligned}
\Gamma\left(\frac{1}{2}\right) &= \int_0^\infty x^{1/2-1}\exp(-x)dx \\
&= \int_0^\infty x^{-1/2}\exp(-x)dx \\
&= 2\int_0^\infty \exp\left(-t^2\right)dt \quad \left(\text{change of variable: } t = x^{1/2}\right) \\
&= 2\left(\int_0^\infty \exp\left(-t^2\right)dt \int_0^\infty \exp\left(-t^2\right)dt\right)^{1/2} \\
&= 2\left(\int_0^\infty \exp\left(-t^2\right)dt \int_0^\infty \exp\left(-s^2\right)ds\right)^{1/2} \\
&= 2\left(\int_0^\infty \int_0^\infty \exp\left(-t^2-s^2\right)dtds\right)^{1/2} \\
&= 2\left(\int_0^\infty \int_0^\infty \exp\left(-s^2u^2-s^2\right)sduds\right)^{1/2} \quad (\text{ change of variable: } t = su) \\
&= 2\left(\int_0^\infty \int_0^\infty \exp\left(-\left(1+u^2\right)s^2\right)sdsdu\right)^{1/2} \\
&= 2\left(\int_0^\infty \left[-\frac{1}{2\left(1+u^2\right)}\exp\left(-\left(1+u^2\right)s^2\right)\right]_0^\infty du\right)^{1/2} \\
&= 2\left(\int_0^\infty \left[0+\frac{1}{2\left(1+u^2\right)}\right]du\right)^{1/2} \\
&= 2^{1/2}\left(\int_0^\infty \frac{1}{1+u^2}du\right)^{1/2} \\
&= 2^{1/2}([\arctan(u)]_0^\infty)^{1/2} \\
&= 2^{1/2}(\arctan(\infty)-\arctan(0))^{1/2} \\
&= 2^{1/2}\left(\frac{\pi}{2}-0\right)^{1/2} \\
&= \pi^{1/2}
\end{aligned}
$$

By using this fact and the recursion formula previously shown, it is immediate to prove that

$$
\Gamma\left(n+\frac{1}{2}\right) = \sqrt{\pi}\prod_{j-0}^{n-1}\left(j+\frac{1}{2}\right)
$$

Proof

The result is obtained by iterating the recursion formula:

$$
\begin{aligned}
&\Gamma\left(n+\frac{1}{2}\right)\\
=&\left(n-1+\frac{1}{2}\right)\Gamma\left(n-1+\frac{1}{2}\right)\\
=&\left(n-1+\frac{1}{2}\right)\left(n-2+\frac{1}{2}\right)\Gamma\left(n-2+\frac{1}{2}\right)\\
&\vdots\\
=&\left(n-1+\frac{1}{2}\right)\left(n-2+\frac{1}{2}\right)\cdots\left(n-n+\frac{1}{2}\right)\Gamma\left(n-n+\frac{1}{2}\right)\\
=&\left(n-1+\frac{1}{2}\right)\left(n-2+\frac{1}{2}\right)\cdots\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)\\
=&\sqrt{\pi}\prod_{j-0}^{n-1}\left(j+\frac{1}{2}\right)
\end{aligned}
$$

Lower and upper incomplete Gamma functions

The definition of the Gamma function

$$\Gamma(z)=\int_0^\infty x^{x^{z-1}}\exp(-x)dx$$

can be generalized in two ways:

1. by substituting the upper bound of integration $(x=\infty)$ with a variable $(x=y)$:

$$\gamma_L(z,y)=\int_0^y x^{z-1}\exp(-x)dx$$

2. by substituting the lower bound of integration with a variable:

$$\gamma_U(z,y)=\int_y^\infty x^{2-1}\exp(-x)dx$$

The functions $\gamma_L(z,y)$ and $\gamma_U(z,y)$ thus obtained are called lower and upper incomplete Gamma functions.

Clearly, they have the property that

$$\Gamma(z)=\gamma_L(z,y)+\gamma_U(z,y)$$

for any y, which is equivalent to

$$\frac{\gamma_L(z,y)}{\Gamma(z)}+\frac{\gamma_U(z,y)}{\Gamma(z)}=1$$

The two ratios

$$\frac{\gamma_L(z,y)}{\Gamma(z)}$$

and

$$\frac{\gamma_U(z,y)}{\Gamma(z)}$$

They are numerically more stable and easier to deal with because they take values between 0 and 1 , while the values taken by the two functions $\gamma_L(z,y)$ and $\gamma_U(z,y)$ can easily overflow.

The lower incomplete function is particularly important in statistics, as it appears in the distribution function of the Chi-square and Gamma distributions.

Interactive calculator for lower and upper incomplete Gamma functions

3	Argument z	2	Bound of integration x
0.3233	Standardized lower incomplete	0.6767	Standardized upper incomplete
2.0000	Gamma(z)	0.6931	$\log(\text{Gamma}(\text{z}))$
0.6466	Value of lower incomplete	1.3534	Value of upper incomplete
-0.4360	Logarithm of lower incomplete	0.3026	Logarithm of upper incomplete

Solved exercises

Below you can find some exercises with explained solutions.

Exercise 1

Compute the following ratio:

$$\frac{\Gamma\left(\frac{16}{3}\right)}{\Gamma\left(\frac{10}{3}\right)}$$

Solution
We need to repeatedly apply the recursive formula

$$\Gamma(z) = (z-1)\Gamma(z-1)$$

to the numerator of the ratio:

$$
\begin{aligned}
\frac{\Gamma\left(\frac{16}{3}\right)}{\Gamma\left(\frac{10}{3}\right)} &= \frac{\left(\frac{16}{3}-1\right)\Gamma\left(\frac{16}{3}-1\right)}{\Gamma\left(\frac{10}{3}\right)} \\
&= \frac{13}{3}\frac{\Gamma\left(\frac{13}{3}\right)}{\Gamma\left(\frac{10}{3}\right)} \\
&= \frac{13}{3}\frac{\left(\frac{13}{3}-1\right)\Gamma\left(\frac{13}{3}-1\right)}{\Gamma\left(\frac{10}{3}\right)} \\
&= \frac{13}{3}\frac{10}{3}\frac{\Gamma\left(\frac{10}{3}\right)}{\Gamma\left(\frac{10}{3}\right)} \\
&= \frac{130}{9}
\end{aligned}
$$

Exercise 2

Compute

$$\Gamma(5)$$

Solution
We need to use the relation of the Gamma function to the factorial function:

$$\Gamma(n) = (n-1)!$$

which, for $n = 5$, becomes

$$
\begin{aligned}
\Gamma(5) &= (5-1)! \\
&= 4! \\
&= 4 \cdot 3 \cdot 2 \cdot 1 \\
&= 24
\end{aligned}
$$

Exercise 3

Express the following integral in terms of the Gamma function:

$$\int_0^\infty x^{9/2} \exp\left(-\frac{1}{2}x\right) dx$$

Solution

This is accomplished as follows:

$$
\begin{aligned}
&\int_0^\infty x^{9/2}\exp\left(-\frac{1}{2}x\right)dx \\
=&\int_0^\infty (2t)^{9/2}\exp(-t)2dt \qquad \left(\text{by changing variables: } t=\frac{1}{2}x\right) \\
=&2^{11/2}\int_0^\infty t^{11/2-1}\exp(-t)dt \\
=&2^{11/2}\Gamma(11/2)
\end{aligned}
$$

where in the last step we have just used the definition of Gamma function.

References

Abramowitz, M. and I. A. Stegun (1965) Handbook of mathematical functions: with formulas, graphs, and mathematical tables, Courier Dover Publications.

How to cite

Please cite as:
Taboga, Marco (2021). "Gamma function", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/mathematical-tools/gammafunction.

Beta function

by Marco Taboga, PhD

The Beta function is a function of two variables that is often found in probability theory and mathematical statistics (for example, as a normalizing constant in the probability density functions of the F distribution and of the Student's t distribution). We report here some basic facts about the Beta function.

$:=$

Table of contents

Definition

The following is a possible definition of the Beta function:
Definition The Beta function is a function $B : \mathbb{R}^2_{++} \rightarrow \mathbb{R}$ defined as follows:

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

where $\Gamma(\quad)$ is the Gamma function.

While the domain of definition of the Beta function can be extended beyond the set $\mathbb{R}^2_{++}$ of couples of strictly positive real numbers (for example to couples of complex numbers), the somewhat restrictive definition given above is more than sufficient to address all the problems involving the Beta function that are found in these lectures.

Integral representations

The Beta function has several integral representations, which are sometimes also used as a definition of the Beta function, in place of the definition we have given above. We report here two often used representations.

Integral between zero and infinity

The first representation involves an integral from zero to infinity:

$$B(x,y) = \int_0^\infty t^{x-1}(1+t)^{-x-y}dt$$

Proof

Given the definition of the Beta function as a ratio of Gamma functions (see above), the equality holds if and only if

$$\int_0^\infty t^{x-1}(1+t)^{-x-y}dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

or

$$\Gamma(x+y)\int_0^\infty t^{t^{x-1}}(1+t)^{-x-y}dt = \Gamma(x)\Gamma(y)$$

That the latter equality indeed holds is proved as follows:

$$\begin{aligned}
\Gamma(x)\Gamma(y) &= \textstyle\int_0^\infty u^{x-1}\exp(-u)du \int_0^\infty v^{\nu-1}\exp(-v)dv \quad \text{(by the definition of Gamma funct.)} \\
&= \textstyle\int_0^\infty v^{\nu-1}\exp(-v)\int_0^\infty u^{x-1}\exp(-u)dudv \\
&= \textstyle\int_0^\infty v^{\nu-1}\exp(-v)\int_0^\infty (vt)^{x-1}\exp(-vt)vdtdv \qquad \text{(changing variable: } u = vt\text{)} \\
&= \textstyle\int_0^\infty v^{\nu-1}\exp(-v)\int_0^\infty v^x t^{x-1}\exp(-vt)dtdv \\
&= \textstyle\int_0^\infty v^{x+v-1}\exp(-v)\int_0^\infty t^{x-1}\exp(-vt)dtdv \\
&= \textstyle\int_0^\infty \int_0^\infty v^{x+y-1}t^{x-1}\exp(-(1+t)v)dtdv \\
&= \textstyle\int_0^\infty t^{x-1}\int_0^\infty v^{x+y-1}\exp(-(1+t)v)dvdt \\
&= \textstyle\int_0^\infty t^{x-1}\int_0^\infty \left(\frac{s}{1+t}\right)^{x+y-1}\exp(-s)\frac{1}{1+t}dsdt \quad \text{(changing variable: } s = (1+t)v\text{)} \\
&= \textstyle\int_0^\infty t^{x-1}(1+t)^{-x-v}\int_0^\infty s^{x^{x+v-1}}\exp(-s)dsdt \\
&= \textstyle\int_0^\infty t^{x-1}(1+t)^{-x-y}\Gamma(x+y)dt \quad \text{(by the definition of Gamma funct.)} \\
&= \textstyle\Gamma(x+y)\int_0^\infty t^{x-1}(1+t)^{-x-v}dt
\end{aligned}$$

(by the definition of Gamma funct.)

(changing variable: $u = vt$)

$$\begin{aligned} B(x,y) &= \int_0^\infty t^{x-1}(1+t)^{-x-y}dt \\ &= \int_0^1 \left(\frac{s}{1-s}\right)^{x-1}\left(1+\frac{s}{1-s}\right)^{-x-y}\left(\frac{1}{1-s}\right)^2 ds \\ &= \int_0^1 \left(\frac{s}{1-s}\right)^{x-1}\left(\frac{1}{1-s}\right)^{-x-y}\left(\frac{1}{1-s}\right)^2 ds \\ &= \int_0^1 s^{x-1}\left(\frac{1}{1-s}\right)^{x-1-x-y+2} ds \\ &= \int_0^1 s^{x-1}\left(\frac{1}{1-s}\right)^{1-y} ds \\ &= \int_0^1 s^{x-1}(1-s)^{y-1} ds \end{aligned}$$

Note that the two representations above involve improper integrals that converge if $x > 0$ and $y > 0$: this might help you to see why the arguments of the Beta function are required to be strictly positive.

More details

The following sections contain more details about the Beta function.

Incomplete Beta function

The integral representation of the Beta function

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$$

can be generalized by substituting the upper bound of integration $(t = 1)$ with a variable $(t = z \leq 1)$:

$$B(z,x,y) = \int_0^z t^{x^{-1}}(1-t)^{y-1}dt$$

The function $B(z,x,y)$ thus obtained is called incomplete Beta function.

Solved exercises

Below you can find some exercises with explained solutions.

Exercise 1

Compute the following product:

$$\Gamma\left(\frac{5}{2}\right)B\left(\frac{3}{2},1\right)$$

where $\Gamma()$ is the Gamma function and $B(\quad)$ is the Beta function.

Solution

We need to write the Beta function in terms of Gamma functions:

$$\begin{aligned}
& \Gamma\left(\tfrac{5}{2}\right)B\left(\tfrac{3}{2},1\right) \\
= & \Gamma\left(\tfrac{5}{2}\right)\frac{\Gamma\left(\frac{3}{2}\right)\Gamma(1)}{\Gamma\left(\frac{3}{2}+1\right)} \\
= & \Gamma\left(\tfrac{5}{2}\right)\frac{\Gamma\left(\frac{3}{2}\right)\Gamma(1)}{\Gamma\left(\frac{5}{2}\right)} \\
= & \Gamma\left(\tfrac{3}{2}\right)\Gamma(1) \\
= & \Gamma\left(\tfrac{3}{2}\right) && \text{(because } \Gamma(1)=1\text{)} \\
= & \left(\tfrac{3}{2}-1\right)\Gamma\left(\tfrac{3}{2}-1\right) \\
= & \tfrac{1}{2}\Gamma\left(\tfrac{1}{2}\right) \\
= & \tfrac{1}{2}\sqrt{\pi} && \text{(becursive formula for the Gamma function) } \Gamma\left(\tfrac{1}{2}\right)=\sqrt{\pi})
\end{aligned}$$

where we have used several elementary facts about the Gamma function, that are explained in the lecture entitled Gamma function.

Exercise 2

Compute the following ratio

$$\frac{B\left(\frac{7}{2},\frac{9}{2}\right)}{B\left(\frac{5}{2},\frac{11}{2}\right)}$$

where $B(\quad)$ is the Beta function.

Solution

This is achieved by rewriting the numerator of the ratio in terms of Gamma functions and using the recursive formula for the Gamma function:

$$
\begin{aligned}
&\frac{B\left(\frac{7}{2},\frac{9}{2}\right)}{B\left(\frac{5}{2},\frac{11}{2}\right)}\\
=&\frac{1}{B\left(\frac{5}{2},\frac{11}{2}\right)}\frac{\Gamma\left(\frac{7}{2}\right)\Gamma\left(\frac{9}{2}\right)}{\Gamma\left(\frac{7}{2}+\frac{9}{2}\right)}\\
=&\frac{1}{B\left(\frac{5}{2},\frac{11}{2}\right)}\frac{\left(\frac{7}{2}-1\right)\Gamma\left(\frac{7}{2}-1\right)\Gamma\left(\frac{9}{2}\right)}{\Gamma\left(\frac{7}{2}+\frac{9}{2}\right)}\\
=&\frac{1}{B\left(\frac{5}{2},\frac{11}{2}\right)}\frac{\frac{5}{2}\Gamma\left(\frac{7}{2}-1\right)\left[\Gamma\left(\frac{11}{2}\right)/\left(\frac{11}{2}-1\right)\right]}{\Gamma\left(\frac{7}{2}-1+\frac{9}{2}+1\right)} \qquad \text{(recursive formula for the Gamma function)}\\
=&\frac{5}{2}\frac{1}{9}\frac{1}{B\left(\frac{5}{2},\frac{11}{2}\right)}\frac{\Gamma\left(\frac{5}{2}\right)\Gamma\left(\frac{11}{2}\right)}{\Gamma\left(\frac{5}{2}+\frac{11}{2}\right)}\\
=&\frac{5}{9}\frac{1}{B\left(\frac{5}{2},\frac{11}{2}\right)}B\left(\frac{5}{2},\frac{11}{2}\right) \qquad \text{(by the definition of Beta function)}\\
=&\frac{5}{9}
\end{aligned}
$$

Exercise 3

Solution

We need to use the integral representation of the Beta function:

$$
\begin{aligned}
&\int_0^\infty x^{3/2}(1+2x)^{-5}dx\\
=&\int_0^\infty \left(\frac{1}{2}t\right)^{3/2}(1+t)^{-5}\frac{1}{2}dt \qquad \left(\text{by a change of variables: } t=2x\right)\\
=&\left(\frac{1}{2}\right)^{5/2}\int_0^\infty t^{3/2}(1+t)^{-5}dt\\
=&\left(\frac{1}{2}\right)^{5/2}\int_0^\infty t^{5/2-1}(1+t)^{-5/2-5/2}dt\\
=&\left(\frac{1}{2}\right)^{5/2}B\left(\frac{5}{2},\frac{5}{2}\right) \qquad \text{(using the integral representation of the Beta function)}
\end{aligned}
$$

Now, write the Beta function in terms of Gamma functions:

$$
\begin{aligned}
B\left(\frac{5}{2},\frac{5}{2}\right) &= \frac{\Gamma\left(\frac{5}{2}\right)\Gamma\left(\frac{5}{2}\right)}{\Gamma\left(\frac{5}{2}+\frac{5}{2}\right)} \\
&= \frac{\left[\Gamma\left(\frac{5}{2}\right)\right]^2}{\Gamma(5)} \\
&= \frac{\left[\frac{3}{2}\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\right]^2}{4\cdot 3\cdot 2\cdot 1} \qquad \text{(recursive formula for the Gamma function)} \\
&= \frac{1}{24}\left(\frac{3}{4}\right)^2\left[\Gamma\left(\frac{1}{2}\right)\right]^2 \\
&= \frac{9}{384}\pi \qquad \left(\text{because } \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\ \right)
\end{aligned}
$$

Substituting this number into the previous expression for the integral, we obtain

$$
\begin{aligned}
&\int_0^\infty x^{3/2}(1+2x)^{-5}dx \\
&= \left(\frac{1}{2}\right)^{5/2} B\left(\frac{5}{2},\frac{5}{2}\right) \\
&= \left(\frac{1}{2}\right)^{5/2} \frac{9}{384}\pi \\
&= \frac{1}{8}\sqrt{2}\frac{9}{384}\pi \\
&= \frac{9}{3072}\sqrt{2}\pi \\
&= \frac{3}{1024}\sqrt{2}\pi
\end{aligned}
$$

If you wish, you can check the above result by using the following MATLAB commands:

$$
\begin{gathered}
\text{syms } x \\
f = \left(x^\wedge(3/2)\right)^\star\left((1+2^\star x)^\wedge - 5\right) \\
\operatorname{int}(f, 0,\ \text{Inf}\ )
\end{gathered}
$$

How to cite

Please cite as:
Taboga, Marco (2021). "Beta function", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/mathematical-tools/betafunction.