Unit 2 A: Vectors, Matrices and $\mathbb{R}^n$

Prof.Dr.P.M.Bajracharya

School of Mathematical Sciences
T.U., Kirtipur

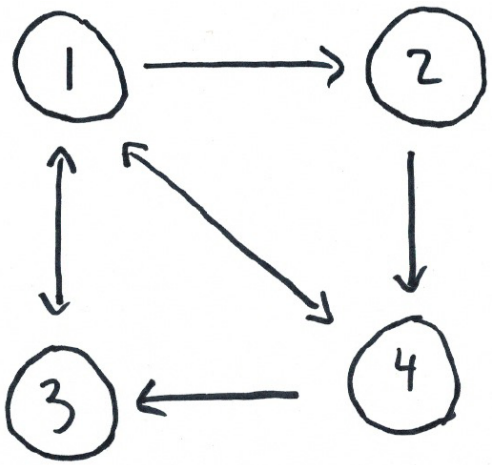
February 16, 2025

Summary

1. Random walk matrices
2. Ways of viewing a matrix
3. What is $\mathbb{R}^n$?
4. Some very basic subsets of $\mathbb{R}^n$

Random walk matrices

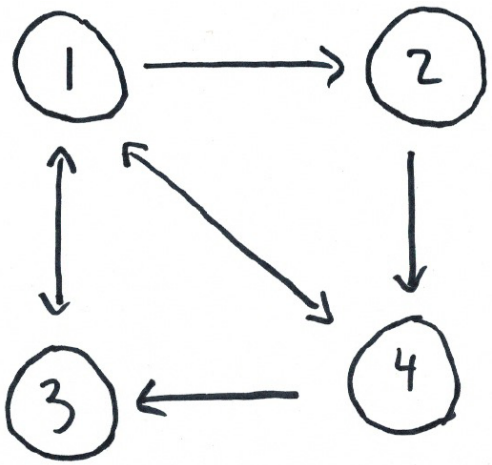
Consider a directed graph:



Its adjacency matrix:

$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

Consider a directed graph:



Its adjacency matrix:

$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

One can define the adjacency matrix of a directed graph either such that a non-zero element A_{ij} indicates

1. an edge from i to j or
2. an edge from j to i.

- The former definition is commonly used in graph theory and social network analysis (e.g., sociology, political science, economics, psychology).

- The former definition is commonly used in graph theory and social network analysis (e.g., sociology, political science, economics, psychology).
- The latter is more common in other applied sciences (e.g., dynamical systems, physics, network science) where the resulting matrix is sometimes used to describe linear dynamics on graphs.

Dividing each column by the sum of the entries in that column, we obtain a matrix W given by

$$W = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 0 & 0 \end{pmatrix}$$

Dividing each column by the sum of the entries in that column, we obtain a matrix W given by

$$W = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 0 & 0 \end{pmatrix}$$

It is called a **random walk matrix**.
This matrix is quite important, and it has many uses.

Let $x \in \mathbb{R}^4$ be a column vector. Put

$$y = Ax.$$

Note that

1. The original vector x is a vector in $\mathbb{R}^4$ can be interpreted as a score or importance score for each node of the graph.

Let $x \in \mathbb{R}^4$ be a column vector. Put

$$y = Ax.$$

Note that

1. The original vector x is a vector in $\mathbb{R}^4$ can be interpreted as a score or importance score for each node of the graph.
2. The interpretation of $y = Wx$ is that we take one vector in $\mathbb{R}^4$ and we apply W to get another vector in $\mathbb{R}^4$, so we have applied some sort of function that transforms one 4-vector into another 4-vector.

Moreover, if the following is true:

$$0 \leq x_i \leq 1 \text{ and } \sum_{i=1}^{4} x_i = 1,$$

then x is what we will call a probability distribution (in the sense that it generalizes the probabilities of 50% heads and 50% tails from flipping a fair coins to more complicated dependency situations).

- The reason is that, in this case, if x is a probability distribution, then because of the way we have normalized the columns of W, it will be the case that y is also a probability distribution.

- The reason is that, in this case, if x is a probability distribution, then because of the way we have normalized the columns of W, it will be the case that y is also a probability distribution.
- This will help us to tap into ideas from probability theory to get better algorithms and better statistical properties.

So, in addition to having an informal interpretation in terms of importance scores, this process will have a more rigorous interpretation in terms of random walks and diffusion processes that are central to probability theory.

For example, if we start with our probability mass x at one node and compute

$$y = Wx$$

and then

$$z = W(Wx) = W^2x,$$

and so on, then we might hope that probability mass will spread out.

For example, if we start with our probability mass x at one node and compute

$$y = Wx$$

and then

$$z = W(Wx) = W^2x,$$

and so on, then we might hope that probability mass will spread out. This is known as a **Markov chain**, and it is central to probability theory.

In fact, if you iterate that process many steps, then we will get a vector, call it $z*$ (sometimes this is called the **spectral ranking vector** or the **Pagerank vector**) that has been used to rank all sorts of things from web pages to sports teams to academic departments, and so on.

In fact, if you iterate that process many steps, then we will get a vector, call it $z*$ (sometimes this is called the **spectral ranking vector** or the **Pagerank vector**) that has been used to rank all sorts of things from web pages to sports teams to academic departments, and so on.

By the way, what we just discussed is true if we work with a directed graph. If, on the other hand, we ignore the direction, then we get a vector that can be used for classification, clustering, and other common data science tasks.

Ways of viewing a matrix

There are **four** major ways to view a matrix. They are

1. As a set of points
2. As a linear transformation
3. As a quadratic form
4. In terms of linear equations

1. As a set of points

m points in $\mathbb{R}^n$ can be written as an $m \times n$ matrix.

1. As a set of points

m points in $\mathbb{R}^n$ can be written as an $m \times n$ matrix.

Examples.

- One point on the real line, i.e., on $\mathbb{R}$ as 1×1 matrix
- One point on the plane as 1×2 matrix
- Two points on the line as 2×1 matrix
- Two points on the plane as 2×2 matrix

These examples are so simple that this is not usually done.

2. As a linear transformation.

Matrix-vector products, and matrix-matrix products, e.g., the functions

$$y \leftarrow Ax, \;\; z \leftarrow Ay, \;\; \text{etc.}$$

can be viewed as defining linear functions taking vectors x as inputs and returning vectors y as outputs.

- If the matrix is $m \times n$, we will see that this involves linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$.

- If the matrix is $m \times n$, we will see that this involves linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$.
- This is a generalization of the familiar $y = ax$, the equation of a line with slope a going through the origin, where y, a, x are all real numbers.

- If the matrix is $m \times n$, we will see that this involves linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$.
- This is a generalization of the familiar $y = ax$, the equation of a line with slope a going through the origin, where y, a, x are all real numbers.
- We are thinking of y as a simple function of x. In this case, given the value of a, one may want to compute y as a function of the input x.

3. As a quadratic form.

- This is s a generalization of $y = ax^2$, the familiar equation for a simple quadratic form involving one variable $x \in \mathbb{R}$.

3. As a quadratic form.

- This is s a generalization of $y = ax^2$, the familiar equation for a simple quadratic form involving one variable $x \in \mathbb{R}$.
- For example, if we have a symmetric matrix A and a vector x, then we can define the transpose of x^T.

3. As a quadratic form.

- This is s a generalization of $y = ax^2$, the familiar equation for a simple quadratic form involving one variable $x \in \mathbb{R}$.
- For example, if we have a symmetric matrix A and a vector x, then we can define the transpose of x^T.
- Then we can extend the matrix-vector multiplication to write $y = x^T Ax$, which is a quadratic form in

$$x = (x_1, ..., x_n)^T \in \mathbb{R}^n.$$

- We know that the one-dimensional quadratic form $y = ax^2$ has very different properties depending on whether $a > 0$ or $a = 0$ or $a < 0$.

- We know that the one-dimensional quadratic form $y = ax^2$ has very different properties depending on whether $a > 0$ or $a = 0$ or $a < 0$.
- We will see that for higher-dimensional quadratic forms, i.e., involving vectors $x \in \mathbb{R}^n$ rather than $x \in \mathbb{R}^2$, there is a rich array of properties that are possible.

3. In terms of linear equations.

- Given a vector y and a matrix A, we might want to find a vector x such that $y = Ax$.

3. In terms of linear equations.

- Given a vector y and a matrix A, we might want to find a vector x such that $y = Ax$.
- This could be a vector where the random walk process leaves the vectors unchanged, if such a vector exists, but it could also be all sorts of other things.

3. In terms of linear equations.

- Given a vector y and a matrix A, we might want to find a vector x such that $y = Ax$.
- This could be a vector where the random walk process leaves the vectors unchanged, if such a vector exists, but it could also be all sorts of other things.
- This has similarities with the linear transformation perspective, except when one thinks about matrices in terms of linear equation solving one typically knows y and A and one wants to find x.

When restricted to the familiar case, one wants to solve for x in the equation $y = ax$, and we know that the solution is $x = a^{-1}y = y/a$, assuming that $a \neq 0$.

When restricted to the familiar case, one wants to solve for x in the equation $y = ax$, and we know that the solution is $x = a^{-1}y = y/a$, assuming that $a \neq 0$. Again, when we generalize to higher-dimensional linear equations, a rich array of properties are possible.

While the linear transformation and linear equation perspectives both involve expressions of the form $y = Ax$, there is an important difference in terms of what we assume we know.

- For the linear transformation perspective, we know A and x and we are interested in computing y. This is sometimes known as a **forward process**.

- For the linear transformation perspective, we know A and x and we are interested in computing y. This is sometimes known as a **forward process**.
- For the linear equation perspective, we know A and y and we are interested in computing x. This is sometimes known as a **backward** or **inverse process**, since at least formally this involves the inverse of A.

- Inverse processes tend to be more difficult.

- Inverse processes tend to be more difficult.
- Most linear algebra classes start with this inverse process of solving linear equations as the main motivation.

- Inverse processes tend to be more difficult.
- Most linear algebra classes start with this inverse process of solving linear equations as the main motivation.
- We will get to that eventually as an important application, but we won't start with it.

- Viewing matrices from each of these perspectives has its advantages and disadvantages.

- Viewing matrices from each of these perspectives has its advantages and disadvantages.
- In particular, the last perspective (i.e., in terms of linear equations) is most common and is the most important for many traditional applications of linear algebra, e.g., in engineering, physics, and scientific computation, and so it gets the greatest emphasis in many traditional linear algebra classes.

- Viewing matrices from each of these perspectives has its advantages and disadvantages.
- In particular, the last perspective (i.e., in terms of linear equations) is most common and is the most important for many traditional applications of linear algebra, e.g., in engineering, physics, and scientific computation, and so it gets the greatest emphasis in many traditional linear algebra classes.
- This approach leads to an emphasis on things like *Reduced Row Echelon Forms*, *QR decompositions*, and so on.

- The other perspectives are more useful in *data science*, *data analysis*, and *machine learning*, and so in this class we will be much more interested in those perspectives.

- The other perspectives are more useful in *data science*, *data analysis*, and *machine learning*, and so in this class we will be much more interested in those perspectives.
- In particular, we will start by viewing matrices as consisting of a bunch of points in a high-dimensional vector space, in which case the linear transformation and quadratic form perspectives will arise naturally.

What is $\mathbb{R}^n$?

One-dimensional Euclidean space $\mathbb{R}$:

- This is just the set of real numbers.

One-dimensional Euclidean space $\mathbb{R}$:

- This is just the set of real numbers.
- This is typically visualized as the **real number line**.

One-dimensional Euclidean space $\mathbb{R}$:

- This is just the set of real numbers.
- This is typically visualized as the **real number line**.
- The elements of $\mathbb{R}$ can be added and multiplied, and both of these operations have an inverse, except for multiplication by 0 which doesn't have an inverse, etc.

One-dimensional Euclidean space $\mathbb{R}$:

- This is just the set of real numbers.
- This is typically visualized as the **real number line**.
- The elements of $\mathbb{R}$ can be added and multiplied, and both of these operations have an inverse, except for multiplication by 0 which doesn't have an inverse, etc.

By the last point, we mean that $y = ax$ has the solution $x = a^{-1}y$ for all $a \neq 0$; and if $a = 0$, then there is in general no solution.

Two-dimensional Euclidean space $\mathbb{R}^2$:

- This is just the set of ordered pairs of real numbers. That is,

$$x \in \mathbb{R}^2 \Leftrightarrow x = (x_1, x_2),$$

where $x_1, x_2 \in \mathbb{R}$.

Two-dimensional Euclidean space $\mathbb{R}^2$:

- This is just the set of ordered pairs of real numbers. That is,

$$x \in \mathbb{R}^2 \Leftrightarrow x = (x_1, x_2),$$

  where $x_1, x_2 \in \mathbb{R}$.
- This is visualized as the **plane**.

Two-dimensional Euclidean space $\mathbb{R}^2$:

- This is just the set of ordered pairs of real numbers. That is,

$$x \in \mathbb{R}^2 \Leftrightarrow x = (x_1, x_2),$$

  where $x_1, x_2 \in \mathbb{R}$.
- This is visualized as the **plane**.
- We will explore operations like *addition* and *multiplication* for points of the plane.

Three-dimensional Euclidean space $\mathbb{R}^3$:

- This is the set of ordered triples of real numbers. That is,

$$x \in \mathbb{R}^3 \Leftrightarrow x = (x_1, x_2, x_3),$$

where $x_1, x_2, x_3 \in \mathbb{R}$.

Three-dimensional Euclidean space $\mathbb{R}^3$:

- This is the set of ordered triples of real numbers. That is,

$$x \in \mathbb{R}^3 \Leftrightarrow x = (x_1, x_2, x_3),$$

  where $x_1, x_2, x_3 \in \mathbb{R}$.
- This is visualized as the **space**. We will explore operations like *addition* and *multiplication* for points of the space.

Four-dimensional Euclidean space $\mathbb{R}^4$:

- This is the set of ordered n-tuples of real numbers. That is,

$$x \in \mathbb{R}^n \Leftrightarrow x = (x_1, x_2, ..., x_n),$$

where $x_1, x_2, ..., x_n \in \mathbb{R}$.

Four-dimensional Euclidean space $\mathbb{R}^4$:

- This is the set of ordered n-tuples of real numbers. That is,

$$x \in \mathbb{R}^n \Leftrightarrow x = (x_1, x_2, ..., x_n),$$

  where $x_1, x_2, ..., x_n \in \mathbb{R}$.
- This is harder to visualize.
- However, here too we will be able to define operations on it in a way that cleanly generalizes familiar operations on $\mathbb{R}^2$ and $\mathbb{R}^3$.

- Importantly, as n gets larger, e.g., for 10 or 20, many properties of $\mathbb{R}^n$ are very different than the corresponding properties of $\mathbb{R}^2$ or $\mathbb{R}^3$, but a few properties do generalize cleanly.

- Importantly, as n gets larger, e.g., for 10 or 20, many properties of $\mathbb{R}^n$ are very different than the corresponding properties of $\mathbb{R}^2$ or $\mathbb{R}^3$, but a few properties do generalize cleanly.
- We can use these latter properties to analyze data that are modeled as points in $\mathbb{R}^n$ and thus understand better data modeled as matrices and graphs.

Some very basic properties of $\mathbb{R}^n$

As a first step toward the goal of defining operations that cleanly generalize familiar operations from $\mathbb{R}^2$ and $\mathbb{R}^3$ to $\mathbb{R}^n$, let's ask:

As a first step toward the goal of defining operations that cleanly generalize familiar operations from $\mathbb{R}^2$ and $\mathbb{R}^3$ to $\mathbb{R}^n$, let's ask:

What are some of the most basic notions that we have about points on a line or on a plane or in three dimensional space?

As a first step toward the goal of defining operations that cleanly generalize familiar operations from $\mathbb{R}^2$ and $\mathbb{R}^3$ to $\mathbb{R}^n$, let's ask:

What are some of the most basic notions that we have about points on a line or on a plane or in three dimensional space?

Ans.: Size, Distance, Angle(for $\mathbb{R}^2$ and $\mathbb{R}^3$).

In $\mathbb{R}$:

Size or absolute value. Given a real number x, some are larger, and some are smaller, and this is quantified by the absolute value

$$|x|.$$

In $\mathbb{R}$:

Size or absolute value. Given a real number x, some are larger, and some are smaller, and this is quantified by the absolute value

$$|x|.$$

Using it, we can define a distance between two points.

In $\mathbb{R}$:

Size or absolute value. Given a real number x, some are larger, and some are smaller, and this is quantified by the absolute value

$$|x|.$$

Using it, we can define a distance between two points.

Distance between two points

Given $x, y \in \mathbb{R}$, we can define

$$\text{dist}(x, y) = |x - y|.$$

In $\mathbb{R}$:

Size or absolute value. Given a real number x, some are larger, and some are smaller, and this is quantified by the absolute value

$$|x|.$$

Using it, we can define a distance between two points.

Distance between two points

Given $x, y \in \mathbb{R}$, we can define

$$\text{dist}(x, y) = |x - y|.$$

In many cases, nearby points are more alike.

In $\mathbb{R}^2$:

Size or norm. Given a point on the plane, let's join it to an origin, in which case we will call it a **vector**. Then, we might want to measure the size of the vector.

In $\mathbb{R}^2$:

Size or norm. Given a point on the plane, let's join it to an origin, in which case we will call it a **vector**. Then, we might want to measure the size of the vector.

We can associate a real number to any point on the plane which measures it's size. There are actually many ways to do this, and we will discuss several in detail below.

In $\mathbb{R}^2$:

Size or norm. Given a point on the plane, let's join it to an origin, in which case we will call it a **vector**. Then, we might want to measure the size of the vector.

We can associate a real number to any point on the plane which measures it's size. There are actually many ways to do this, and we will discuss several in detail below.

The usual version of this is the **Euclidean norm**, which we will denote as $\|\cdot\|_2$.

Euclidean norm

Given $x = (x_1, x_2) \in \mathbb{R}^2$, we define that

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2}.$$

This generalizes the notion of size or absolute value to $\mathbb{R}$.

Using a norm, we can define the distance between two vectors.

Using a norm, we can define the distance between two vectors.

Euclidean distance

Given $x, y \in \mathbb{R}^2$, we define that

$$\begin{aligned} \text{dist}(x, y) &= \|x - y\|_2 \\ &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}. \end{aligned}$$

Using a norm, we can define the distance between two vectors.

Euclidean distance

Given $x, y \in \mathbb{R}^2$, we define that

$$\begin{aligned} \text{dist}(x, y) &= \|x - y\|_2 \\ &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}. \end{aligned}$$

Note that

Since there are other notions of size, and since we can associate a distance with a norm as the norm of the vector difference, there are other notions of distance, even for $\mathbb{R}^2$.

Using a norm, we can define the distance between two vectors.

Euclidean distance

Given $x, y \in \mathbb{R}^2$, we define that

$$\begin{aligned} \text{dist}(x, y) &= \|x - y\|_2 \\ &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}. \end{aligned}$$

Note that

Since there are other notions of size, and since we can associate a distance with a norm as the norm of the vector difference, there are other notions of distance, even for $\mathbb{R}^2$.

One reason this is of interest is that nearby points are more alike.

In addition, for $x, y \in \mathbb{R}^2$, by using this particular norm (i.e., the Euclidean norm), we can define an angle θ between x and y as

In addition, for $x, y \in \mathbb{R}^2$, by using this particular norm (i.e., the Euclidean norm), we can define an angle θ between x and y as

Angle

Let $x, y \in \mathbb{R}^2$. The angle θ between x and y is defined by

$$\cos(\theta) = \frac{x^T y}{\|x\|_2 \|y\|_2} = \frac{x_1 y_1 + x_2 y_2}{\|x\|_2 \|y\|_2}.$$

In $\mathbb{R}^3$:

Norms, distances, and angles. Here too, the notion of norms, distances, and angles arise.

In $\mathbb{R}^3$:

Norms, distances, and angles. Here too, the notion of norms, distances, and angles arise.

We will see that norms and angles do generalize to $\mathbb{R}^n$, for $n > 3$. These are very useful in data science, and so we will spend a great deal of time on them.

Problem

Let $n \in \mathbb{N}$. Write the norm of a vector $x \in \mathbb{R}^n$, the distance between two points in $\mathbb{R}^n$ and the angle betweet two vectors in $\mathbb{R}^n$.

Some very basic subsets of $\mathbb{R}^n$

It is often best to have an intuition of what $\mathbb{R}^n$ "looks like," by understanding the properties of very simple subsets of $\mathbb{R}^n$.

It is often best to have an intuition of what $\mathbb{R}^n$ "looks like," by understanding the properties of very simple subsets of $\mathbb{R}^n$.

Here are three very simple subsets of the plane:

- Unit ball,
- Positive orthant,
- Probability Simplex.

Unit ball.

This is the set of points (x_1, x_2) defined by the inequality: $x_1^2 + x_2^2 \leq 1$. In other words, this is the set

$$\left\{(x_1, x_2) \in \mathbb{R}^2 : \ x_1^2 + x_2^2 \leq 1\right\}.$$

Unit ball.

This is the set of points (x_1, x_2) defined by the inequality: $x_1^2 + x_2^2 \leq 1$. In other words, this is the set

$$\left\{(x_1, x_2) \in \mathbb{R}^2 : \ x_1^2 + x_2^2 \leq 1\right\}.$$

We can rewrite it as

$$\left\{x \in \mathbb{R}^2 : \ \|x\| \leq 1\right\}.$$

Generalization to $\mathbb{R}^n$

$$\left\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : \ \sum_1^n x_i^2 \leq 1\right\}$$

$$= \{x \in \mathbb{R}^n : \ \|x\| \leq 1\}$$

An orthant is an n-dimensional generalization of quadrant or octant.

An orthant is an n-dimensional generalization of quadrant or octant.

A positive orthant in $\mathbb{R}^2$ is defined as follows:

$$\left\{(x_1, x_2) \in \mathbb{R}^2 : \ x_1 \geq 0, x_2 \geq 0\right\}.$$

An orthant is an n-dimensional generalization of quadrant or octant.

A positive orthant in $\mathbb{R}^2$ is defined as follows:

$$\left\{(x_1, x_2) \in \mathbb{R}^2 : \ x_1 \geq 0, x_2 \geq 0\right\}.$$

Generalization to $\mathbb{R}^n$

$$\left\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : \ x_i \geq 0, i = 1, 2, ..., n\right\}.$$

Suppose an experiment has a finite set $X = \{x_1, x_2, ..., x_n\}$ of n possible outcomes. Each time the experiment is performed exactly one on the n outcomes happens. Assign each outcome a real number between 0 and 1, called the probability of that outcome.

Suppose an experiment has a finite set $X = \{x_1, x_2, ..., x_n\}$ of n possible outcomes. Each time the experiment is performed exactly one on the n outcomes happens. Assign each outcome a real number between 0 and 1, called the probability of that outcome.

Write $p(x_i)$ for the probability of the outcome x_i. The sum of the probabilities of all outcomes in X must be 1 because the outcomes in X are the only possible outcomes, so one of them must happen.

So far we have $0 \leq p(x_i) \leq 1$ for each i and

$$\sum_{i=1}^{n} p(x_i) = 1.$$

Probability Simplex

A **probability simplex** is a mathematical space where each point represents a probability distribution between a finite number of mutually exclusive events. Each event is often called a **category** (**trial**).

We use the variable K to denote the number of categories.

A point on a probability simplex can be represented by K non-negative numbers that add up to 1.

A point on a probability simplex can be represented by K non-negative numbers that add up to 1. Here are some examples:

- A point in a simplex where $K = 2:\ (0.6, 0.4)$.
- A point in a simplex where $K = 3:\ (0.1, 0.1, 0.8)$.
- A point in a simplex where $K = 6$: $(0.05, 0.2, 0.15, 0.1, 0.3, 0.2)$.

- A simplex is a generalization of the concept of a triangle or tetrahedron to arbitrary dimensions.
- A simplex is so named because it represents the simplest possible polyhedron in any dimension.

- When $K = 1$, this space consisting of only one number 1 is a point.
- When $K = 2$, this space is a line.
- When $K = 3$, it is a triangle.
- When $K = 4$, it is a tetrahedron.

In each case, the simplex is a $K - 1$ dimensional object.

- When $K = 1$, this space consisting of only one number 1 is a point.
- When $K = 2$, this space is a line.
- When $K = 3$, it is a triangle.
- When $K = 4$, it is a tetrahedron.

In each case, the simplex is a $K - 1$ dimensional object. The requirement that the numbers sum to 1 reduces the dimensionality by 1.

$K - 1$ simplex

A probability simplex represented by K non-negative numbers is called a $K - 1$ simplex.

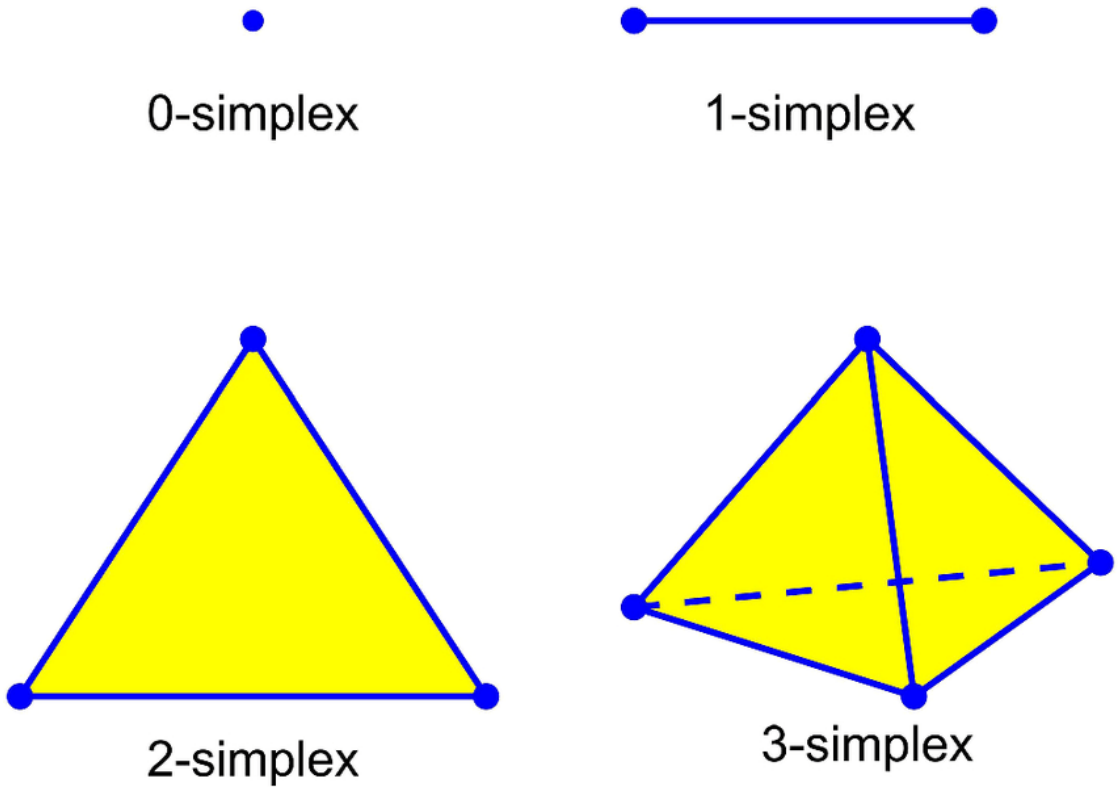


Figure 1: Probability simplexes when $K = 1, 2, 3, 4$

Thus, we have the following definition.

Probability simplex space

A probability simplex is the set of points (x_1, x_2) defined by

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 : \; x_1 + x_2 = 1, x_1 \geq 0, x_2 \geq 0 \right\}.$$

Thus, we have the following definition.

Probability simplex space

A probability simplex is the set of points (x_1, x_2) defined by

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 : \; x_1 + x_2 = 1, x_1 \geq 0, x_2 \geq 0 \right\}.$$

Generalization to $\mathbb{R}^n$

$$\left\{ x \in \mathbb{R}^n : \sum_{1}^{n} x_i = 1, x_i \geq 0, \right\}.$$