An introduction to majorization and some personal thoughts about it

Maxence Jauberty

October 15, 2024

Abstract

The following article is *somehow* an introduction to the theory of majorization. Its objective is to present the main ideas of the theory, as well as some personal thoughts about it. The reader can a complete introduction to the theory in [?].

I would assume some basic knowledge in linear algebra, probability theory and familiarity with information theory.

1 Introduction

Majorization is a concept that appears in various fields of mathematics whenever we need to talk about non-uniformity, or *randomness*. Many approaches are possible depending on the field, you are working on. A satisfying definition, giving a general intuition of it, is given using the concept of Lorenz curve.

Definition 1 (Lorenz Curve)**.** *Let $x \in \Delta_n$, the Lorenz curve $\mathcal{L}_x$ of x is the plot of the points $\left(\frac{i}{n}, \sum_{j=1}^{i} x_{(j)}\right)$, where $x_{(1)} \geq x_{(2)} \geq \ldots \geq x_{(n)}$ are the components of x in decreasing order.*

A Lorenz curve is somehow quantifying the weight of the largest components. In figure 1, the Lorenz curve of the vector $(0.5, 0.25, 0.25, 0)$ is plotted.

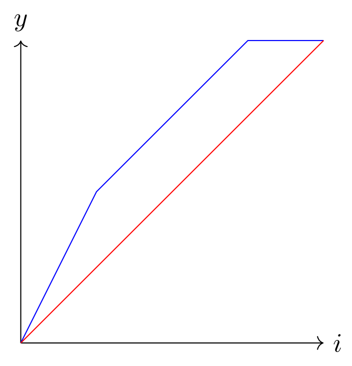
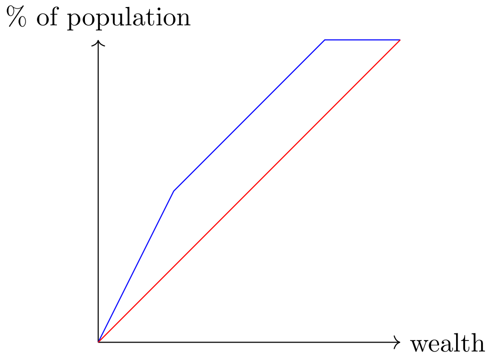


Figure 1: The Lorenz curve of the vector $(0.5, 0.25, 0.25, 0)$ in blue and the Lorenz curve of the vector $(0.25, 0.25, 0.25, 0.25)$ in red.



As you can see, the Lorenz curve of the vector $x = (0.5, 0.25, 0.25, 0)$ is *more* concave than the Lorenz curve of the vector $u = (0.25, 0.25, 0.25, 0.25)$. How can we interpret informally this fact? According to the figure, the greatest component $x_(1)$ weights more than the greatest component of u. A natural way of reprensenting this is to say that these vectors represent the wealth of a population. x would be a situation where the 25% of the richest people would own the 50% of the wealth, while u would be a situation of perfect equality. In other words, the Lorenz curve of a vector x is one way of representing its non-uniformity. It is quite easy to state that a vector x is *less* uniform than the uniform distribution but how can we quantify this comparison of non-uniformity? This is where the concept of majorization comes in.

Definition 2 (Majorization)**.** *Let $x, y \in \mathbb{R}^n$ be two vectors. We say that x is majorized by y, denoted $x \prec y$, if $\mathcal{L}_x$ is below $\mathcal{L}_y$, that is,*

$$\sum_{i=1}^{k} x_{(i)} \leq \sum_{i=1}^{k} y_{(i)} \quad \text{for all } k \in \{1, \dots, n\} \tag{1}$$

$$\sum_{i=1}^{n} x_{(i)} = \sum_{i=1}^{n} y_{(i)}. \tag{2}$$

Remark 1. *Here, we have defined majorization for vectors in $\mathbb{R}^n$ which is not the context of the introduction. However, the definition can be extended and some equivalent definitions can give more intuition of what happens in the general case.*

We will, most of the time, work with non-negative vectors. In this case, the introduction is still relevant.

Example 1. *Let $x = (0.5, 0.25, 0.25, 0)$ and $y = (0.25, 0.25, 0.25, 0.25)$. We have $\mathcal{L}_x$ below $\mathcal{L}_y$, so $x \prec y$.*

Example 2. *Consider the set of angles Θ_Δ to construct a triangle. Let $\theta = (\theta_1, \theta_2, \theta_3)$ be a vector of angles. Basic geometry says that $\theta \in \Theta_\Delta$ if and only if $\theta_1 + \theta_2 + \theta_3 = \pi$. In other words, if and only if $\theta \prec (\pi, 0, 0)$.*

This example is quite interesting because it illustrates the comparison of non-uniformity through the comparison of triangles and we will see that this is somehow related to a characterization of majorization.

Here are some basics properties of majorization.

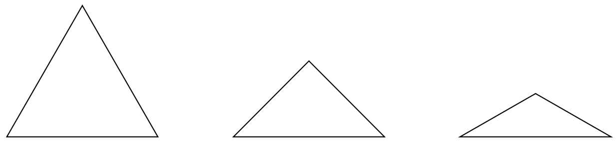


Figure 2: We have respectively $(\pi/3, \pi/3, \pi/3) \prec (\pi/2, \pi/4, \pi/4) \prec (2\pi/3, \pi/6, \pi/6)$

Proposition 1. *Let $x, y \in \mathbb{R}^n$, $\prec$ defines a partial order on $\mathbb{R}^n$. Moreover,*

- *If $x \prec y$ and $y \prec x$, then there exists a permutation π such that $\pi \cdot x = y$.*
- *If $x \prec y$, then for any permutations π, π' we have $\pi \cdot x \prec \pi' \cdot y$.*
- *We have $|x|u \prec x \prec |x|\delta$ where u is the uniform distribution and δ is the Dirac distribution.*

Note that $\prec$ is *indeed* a partial order. There exists vectors x, y such that neither $x \prec y$ nor $y \prec x$.
Before going further, I will present a more information theoretic approach of majorization.

Definition 3. *A matrix $A \in \mathbb{R}_+^{n\times n}$ is said to be stochastic if $u^T A = u^T$. Moreover, we say that A is doubly stochastic if $u^T A = u^T$ and $Au = u$.*

A stochastic matrix A can be considered in the finite case as a so-called *transition matrix* of a Markov chain. These are considered as the most basic representation of channels in information theory. source $\xrightarrow{A}$ destination

2 Generalized (?) majorization

We consider $(\mathfrak{X}, \mathcal{X})$ be a measurable space.

We want to generalize the concept of doubly stochastic matrices. As a reminder, a doubly stochastic matrix is a matrix $A \in \mathbb{R}^{n\times n}$ such that

$$u^T A = u^T \tag{3}$$

$$Au = u. \tag{4}$$

The first condition is only defining A as a proper Markov kernel in the finite case. The second condition must be interpreted as the fact that $A \cdot \mathcal{U}$ where $\mathcal{U}$ is the uniform distribution on $\{1, ..., n\}$. Therefore, we can generalize the concept of doubly stochastic matrices as follows.

For a given probability measure $\mathcal{U}$ on $(\mathfrak{X}, \mathcal{X})$, K is said to be doubly stochastic kernel if it is $\mathcal{U}-$invariant.

Example 3. *Let $\mathfrak{X} = [a, b]$ and $\mathcal{U} = \mathbf{1}_{[a,b]}$, then K is a double stochastic kernel if and only if K verifies*

$$\mathcal{U}(A) = \int_A \mathcal{U}(dx)K(x, A) = \frac{1}{b-a}\int_a^b K(x, A)dx \tag{5}$$

Hence

$$\int_a^b K(x, A)dx = \mathrm{Vol}(A). \tag{6}$$

Definition 4. *We say that μ is majorized by ν, denoted $\mu \prec \nu$, if there exists a doubly stochastic kernel K such that*

$$\mu = K\nu. \tag{7}$$

Proposition 2. *$\prec$ is partial order on the set of probability measures on $(\mathfrak{X}, \mathcal{X})$.*