[MATH-14] Convexity

Miguel-Angel Canela
Associate Professor, IESE Business School

Convex sets

We say that a subset A of $\mathbb{R}^n$ is a **convex set** when for every pair of points of A, the **segment** joining these two points is entirely contained in A. For $n = 1$, the only convex sets are the intervals and the half lines, so that convexity does not bring anything new, but, for $n > 1$, it provides a key mathematical concept.

Example 1. A simple example of a convex set in the plane would be the triangle

$$A = \big\{X \in \mathbb{R}^2 : x_1 \geq 0,\ x_2 \geq 0,\ x_1 + x_2 \leq 1\big\}. \ \square$$

Convex and concave functions

Let D be an interval of $\mathbb{R}$, and $f : D \to \mathbb{R}$ a function. We say that f is a convex function when the set

$$\big\{(x, y) \in \mathbb{R}^2 : x \in D,\ y \geq f(x)\big\}$$

that corresponds to the graph of f and all the points above it, is convex. Reversing the inequalities gives the definition of concave function. Note that f is concave when $-f$ is convex. Also, note that the notions of convex and concave functions are derived from that of convex set, but the expression "concave set" is meaningless.

Example 2. A linear function is both convex and concave. For a linear function, the sets used in the above definitions are half-planes. $\square$

Example 3. A quadratic function $f(x) = ax^2 + bx + c$ is convex when $a > 0$, and concave when $a < 0$. This is easily seen by plotting the parabola $y = ax^2 + bx + c$. $\square$

Alternative definitions

It follows from the definition given above that f is convex when

$$x_1, x_2 \in D, \quad 0 \leq \alpha \leq 1 \Longrightarrow f\big(\alpha x_1 + (1 - \alpha)x_2\big) \leq \alpha f(x_1) + (1 - \alpha)f(x_2).$$

Concavity is obtained by changing signs. This has a geometric interpretation: f is convex (resp. concave) when the curve of equation $y = f(x)$ is below (resp. above) any segment joining two points of the curve (a picture helps to see this). The following theorem provides a method of checking the convexity of a function in an interval.

Theorem. Let D be an interval. $f : D \to \mathbb{R}$ is convex (resp. concave) if and only if f' is non-decreasing. Therefore, when $f''(x) \geq 0$ in D, f is convex.

Example 4. The second derivative of the exponential, which is the exponential itself, is positive. Therefore, the exponential is convex. What about the logarithm? □

Inflection points

A point x_0 such that, for some $\delta > 0$, f is convex on $x_0 - \delta < x < x_0$ and concave on $x_0 < x < x_0 + \delta$, or conversely, is called an **inflection point**. Inflection points are easily detected, because, since f'' changes its sign at x_0, it must be $f''(x_0) = 0$.

Example 5. A rich example is given by the the **standard normal density** function

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2},$$

The first derivative is $f'(x) = -\dfrac{x}{\sqrt{2\pi}}\, e^{-x^2/2}$. Therefore:

- $f'(x) > 0$ for $x < 0$ (increasing).
- $f'(x) = 0$ for $x = 0$ (local maximum).
- $f'(x) < 0$ for $x > 0$ (decreasing).

The second derivative is $f''(x) = \dfrac{x^2 - 1}{\sqrt{2\pi}}\, e^{-x^2/2}$. We have, then:

- $f''(x) > 0$ for $|x| > 1$ (convex).
- $f''(x) = 0$ for $x = \pm 1$ (inflection points).
- $f''(x) < 0$ for $|x| < 1$ (concave).

The picture is completed by remarking that $f(x) \to 0$ as $x \to \infty$, so that the graph of f is asymptotic on both sides, forming the tails of the Gaussian distribution. □

Replacing $f(x)$ by the quadratic approximation, we have

$$f(x) \geq f(x_0) + f'(x_0)(x - x_0)$$

when f is convex near x_0 and the opposite when it is concave. This leads to an alternative (local) definition of convex and concave functions: f is convex (resp. concave) near x_0 when the curve $y = f(x)$ is above (resp. below) any tangent at x_0. What about inflection points? That the tangent crosses the curve (check this for $f(x) = x^3$ at $x_0 = 0$).

Jensen inequality

The convexity (or concavity) of a function is not only a nice geometric property. It gives raise to interesting inequalities. The **Jensen inequality** is a famous one. First, note that the convexity inequality can be easily extended,

$$\alpha_i \geq 0, \quad \sum_{i=1}^{n} \alpha_i = 1 \Longrightarrow f\left[\sum_{i=1}^{n} \alpha_i x_i\right] \leq \sum_{i=1}^{n} \alpha_i f(x_i).$$

In particular, taking $\alpha_i = 1/n$, we get

$$f\left[\frac{x_1 + \cdots + x_n}{n}\right] \leq \frac{f(x_1) + \cdots + f(x_n)}{n}.$$

If f is concave, the inequality is reversed (if it is linear, we have an equality). This means that, if we have a collection of observations $x_1, \ldots, x_n$ of some variable X, and $Y = f(X)$ is a rescaling of X given by a convex function f (e.g. the exponential), the mean $\bar{y}$ of the transformed values $y_1, \ldots, y_n$ is less or equal (in general, less) than the transformed mean $f(\bar{x})$.

Homework

A. The Weibull probability distribution, widely used in duration studies, can be presented by the function

$$S(t) = e^{-(t/\alpha)^\beta}, \quad t > 0.$$

Here, S stands for "survival" and t for "time". α and β are positive parameters. α is the **scale** and β the **shape** parameter. This function is used as a model for the proportion of survivors (the meaning of this depends on the application) at time t.

(a) Keeping α constant (the value does not matter), draw Weibull curves for $\beta = 0.5, 1, 2$. Do you guess why β is called shape?

(b) Which is the proportion of survivors at $t = \alpha$? Why is α called scale?

(c) Find the local extremum and inflection points of the Weibull curve, for the various cases of α and β.

B. For positive variables with a right-skewed distribution, the **geometric mean**

$$\big(x_1 \cdots x_n\big)^{1/n},$$

is frequently preferred to the common (arithmetic) mean as a central value. Show that the geometric mean is lower or equal than the common mean.