

# Jensen's Inequality

Claudiu Reditu

31-07-2019

## Jensen's inequality

In mathematics, Jensen's inequality, named after the Danish mathematician Johan Jensen, relates the value of a convex function of an integral to the integral of the convex function. It was proven by Jensen in 1906. Given its generality, the inequality appears in many forms depending on the context, some of which are presented below. In its simplest form the inequality states that the convex transformation of a mean is less than or equal to the mean applied after convex transformation; it is a simple corollary that the opposite is true of concave transformations.

Jensen's inequality generalizes the statement that the secant line of a convex function lies above the graph of the function, which is Jensen's inequality for two points: the secant line consists of weighted means of the convex function (for  $t \in [0,1]$ ).

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2) \quad (1)$$

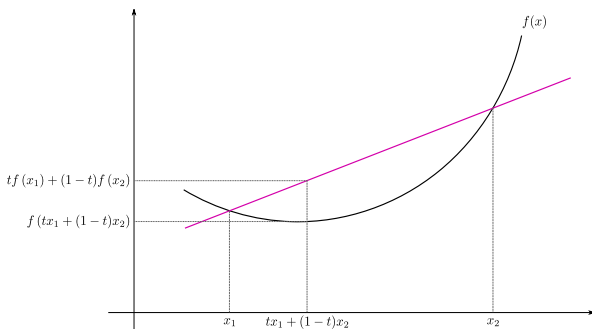


Figure 1: A secant line of a convex function

In the context of Probability Theory it is used usually in this way:

$$f\left(\sum_{i=1}^n p_i x_i\right) \leq \sum_{i=1}^n p_i f(x_i) \quad (2)$$

*Proof.* Let's show that  $f$  is convex on an interval if and only if for all  $x_1$  and  $x_2$  in the interval we have

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2), \quad 0 < t < 1$$

$$\frac{f(tx_1 + (1-t)x_2)}{tx_1 + (1-t)x_2} \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

After working the inequality a bit, we end up with Jensen's Inequality

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

We're going to need a different approach to obtain Jensen's Inequality that is used mostly in Probability Theory. We will start with  $\sum_{i=1}^n p_i = 1$  for  $n > 1$ . We're going to show that for any numbers  $x_1, \dots, x_n$  the sum  $\sum_{i=1}^n p_i x_i$  is between the smallest and largest  $x_i$

We pick  $x_\alpha$  to be the smallest one and  $x_\beta$  to be the largest value. Then we end up with the inequality:

$$\sum_{i=1}^n p_i x_\alpha < \sum_{i=1}^n p_i x_i < \sum_{i=1}^n p_i x_\beta$$

$\sum_{i=1}^n p_i x_\alpha$  is equal to  $x_\alpha$  and  $\sum_{i=1}^n p_i x_\beta$  to  $x_\beta$ . Then this is done.

The next step is to show the same for  $\frac{1}{t} \sum_{i=1}^n p_i x_i$  where  $t = \sum_{i=1}^{n-1} p_i$ .

What we've just shown applied to  $\frac{p_1}{t}, \dots, \frac{p_n}{t}$  shows that  $\frac{1}{t} \sum_{i=1}^n p_i x_i$  lies between the smallest and the largest  $x_1, \dots, x_{n-1}$ , so it certainly lies between  $x_1, \dots, x_n$ .

We now can prove Jensen's Inequality in the context of Probability Theory

$$\begin{aligned} f\left(\sum_{i=1}^n p_i x_i\right) &= f\left(t * \frac{1}{t} \sum_{i=1}^n p_i x_i + (1-t)x_n\right) \\ &\leq t f\left(\sum_{i=1}^{n-1} \frac{p_i}{t} x_i\right) + (1-t)f(x_n) \\ &\leq t \sum_{i=1}^{n-1} \frac{p_i}{t} f(x_i) + p_n f(x_n) \\ &= \sum_{i=1}^n p_i f(x_i) \end{aligned}$$

We end our proof with  $f\left(\sum_{i=1}^n p_i x_i\right) \leq \sum_{i=1}^n p_i f(x_i)$  ■