
6.2.5 Jensen's Inequality

Remember that variance of every random variable X is a positive value, i.e.,

$$\text{Var}(X) = EX^2 - (EX)^2 \geq 0.$$

Thus,

$$EX^2 \geq (EX)^2.$$

If we define $g(x) = x^2$, we can write the above inequality as

$$E[g(X)] \geq g(E[X]).$$

The function $g(x) = x^2$ is an example of **convex** function. **Jensen's inequality** states that, for any convex function g , we have $E[g(X)] \geq g(E[X])$. So what is a convex function? Figure 6.2 depicts a convex function. A function is convex if, when you pick any two points on the graph of the function and draw a line segment between the two points, the entire segment lies above the graph. On the other hand, if the line segment always lies below the graph, the function is said to be **concave**. In other words, $g(x)$ is convex if and only if $-g(x)$ is concave.

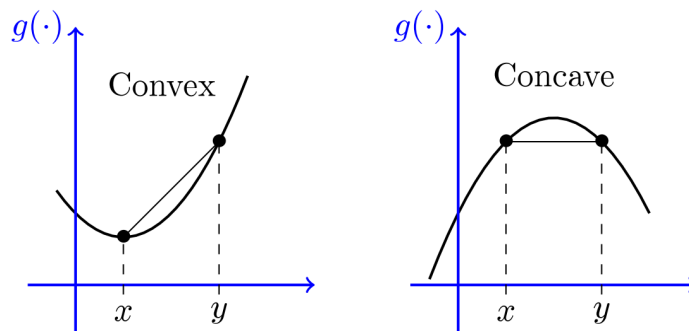


Fig.6.2 - Pictorial representation of a convex function and a concave function.

We can state the definition for convex and concave functions in the following way:

Definition 6.3

Consider a function $g : I \rightarrow \mathbb{R}$, where I is an interval in \mathbb{R} . We say that g is a **convex** function if, for any two points x and y in I and any $\alpha \in [0, 1]$, we have

$$g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y).$$

We say that g is **concave** if

$$g(\alpha x + (1 - \alpha)y) \geq \alpha g(x) + (1 - \alpha)g(y).$$

Note that in the above definition the term $\alpha x + (1 - \alpha)y$ is the weighted average of x and y . Also, $\alpha g(x) + (1 - \alpha)g(y)$ is the weighted average of $g(x)$ and $g(y)$. More generally, for a convex function $g : I \rightarrow \mathbb{R}$, and x_1, x_2, \dots, x_n in I and nonnegative real numbers α_i such that $\alpha_1 + \alpha_2 + \dots + \alpha_n = 1$, we have

$$g(\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n) \leq \alpha_1 g(x_1) + \alpha_2 g(x_2) + \dots + \alpha_n g(x_n) \quad (6.4)$$

If $n = 2$, the above statement is the definition of convex functions. You can extend it to higher values of n by induction.

Now, consider a discrete random variable X with n possible values x_1, x_2, \dots, x_n . In Equation 6.4, we can choose $\alpha_i = P(X = x_i) = P_X(x_i)$. Then, the left-hand side of 6.4 becomes $g(EX)$ and the right-hand side becomes $E[g(X)]$ (by LOTUS). So we can prove the Jensen's inequality in this case. Using limiting arguments, this result can be extended to other types of random variables.

Jensen's Inequality:

If $g(x)$ is a convex function on R_X , and $E[g(X)]$ and $g(E[X])$ are finite, then

$$E[g(X)] \geq g(E[X]).$$

To use Jensen's inequality, we need to determine if a function g is convex. A useful method is the second derivative.

A twice-differentiable function $g : I \rightarrow \mathbb{R}$ is convex if and only if $g''(x) \geq 0$ for all $x \in I$.

For example, if $g(x) = x^2$, then $g''(x) = 2 \geq 0$, thus $g(x) = x^2$ is convex over \mathbb{R} .

Example 6.24

Let X be a positive random variable. Compare $E[X^a]$ with $(E[X])^a$ for all values of $a \in \mathbb{R}$.

Solution

First note

$$\begin{aligned} E[X^a] &= 1 = (E[X])^a, & \text{if } a = 0, \\ E[X^a] &= EX = (E[X])^a, & \text{if } a = 1. \end{aligned}$$

So let's assume $a \neq 0, 1$. Letting $g(x) = x^a$, we have

$$g''(x) = a(a-1)x^{a-2}.$$

On $(0, \infty)$, we can say $g''(x)$ is positive, if $a < 0$ or $a > 1$. It is negative, if $0 < a < 1$.

Therefore we conclude that $g(x)$ is convex, if $a < 0$ or $a > 1$. It is concave, if $0 < a < 1$.

Using Jensen's inequality we conclude

$$\begin{aligned} E[X^a] &\geq (E[X])^a, & \text{if } a < 0 \text{ or } a > 1, \\ E[X^a] &\leq (E[X])^a, & \text{if } 0 < a < 1. \end{aligned}$$
