ECE 302: Lecture 6.3 Jensen's Inequality

Prof Stanley Chan

School of Electrical and Computer Engineering
Purdue University

Jensen's inequality

Theorem (Jensen's Inequality)

Let X be a random variable, and let $g : \mathbb{R} \to \mathbb{R}$ be a **convex** *function. Then*

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X]). \tag{1}$$

Where does it come from?

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Since $\mathrm{Var}[X] \geq 0$ for any X, it follows that

$$\underbrace{\mathbb{E}[X^2]}_{=\mathbb{E}[g(X)]} \geq \underbrace{\mathbb{E}[X]^2}_{=g(\mathbb{E}[X])}. \tag{2}$$

Convex functions

Definition

A function f is convex if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \quad (3)$$

for any $0 \leq \lambda \leq 1$.

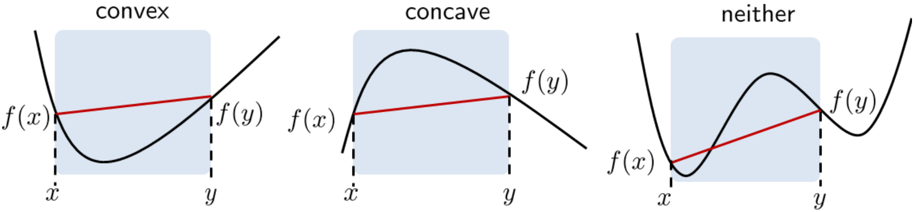

Figure: Illustration of a convex function, a concave function, and a function that is neither convex or concave.

Convex functions

For 1D functions, they are convex if

$$f''(x) \geq 0. \tag{4}$$

Example. The following functions are convex/concave:

- $f(x) = \log x$ is concave, because $f'(x) = \frac{1}{x}$ and $f''(x) = -\frac{1}{x^2} \leq 0$ for all x.
- $f(x) = x^2$ is convex, because $f'(x) = 2x$ and $f''(x) = 2$ which is positive.
- $f(x) = e^{-x}$ is convex, because $f'(x) = -e^{-x}$ and $f''(x) = e^{-x} \geq 0$.

Convexity for Jensen's inequality

$$\underbrace{f(\lambda x + (1-\lambda)y)}_{=f(\mathbb{E}[X])} \leq \underbrace{\lambda f(x) + (1-\lambda) f(y)}_{=\mathbb{E}[f(X)]}, \tag{5}$$

Proof of Jensen's inequality

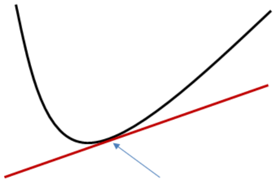

Figure: Proof of Jensen's inequality

Proof. Consider $L(X)$ as defined above. Since g is convex, $g(X) \geq L(X)$ for all X. Therefore,

$$\mathbb{E}[g(X)] \geq \mathbb{E}[L(X)] = \mathbb{E}[aX + b] = a\mathbb{E}[X] + b = L(\mathbb{E}[X]) = g(\mathbb{E}[X]),$$

where the last equality holds because L is a tangent line evaluated at $\mathbb{E}[X]$ which should coincide with $g(\mathbb{E}[X])$. $\square$

Remark for Proof

What are (a, b) in the proof? By Taylor expansion, we can show that

$$g(X) \approx g(\mathbb{E}[X]) + g'(\mathbb{E}[X])(X - \mathbb{E}[X]) \overset{\text{def}}{=} L(X).$$

Therefore, if we want to be precise, then $a = g'(\mathbb{E}[X])$ and $b = g(\mathbb{E}[X]) - g'(\mathbb{E}[X])\mathbb{E}[X]$.

Summary

Theorem (Jensen's Inequality)

Let X be a random variable, and let $g : \mathbb{R} \to \mathbb{R}$ be a **convex** *function. Then*

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X]). \tag{6}$$

Example. By Jensen's inequality, we have that

(a) $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$

(b) $\mathbb{E}\left[\frac{1}{X}\right] \geq \frac{1}{\mathbb{E}[X]}$

(c) $\mathbb{E}[\log X] \leq \log \mathbb{E}[X]$

Questions?