Jensen's Inequality

Jensen's inequality applies to convex functions. Intuitively a function is convex if it is "upward bending". $f(x) = x^2$ is a convex function. To make this definition precise consider two real numbers x_1 and x_2. f is convex if the line between $f(x_1)$ and $f(x_2)$ stays above the function f. To make this even more precise consider $p \in [0,1]$ and consider the weighted average $px_1 + (1-p)x_2$. This is a number between x_1 and x_2. For a given function f we can consider the same weighted average between $f(x_1)$ and $f(x_2)$. As p goes from 1 to 0 we have that the x-y pairs $\langle px_1 + (1-p)x_2,\ \ pf(x_1) + (1-p)f(x_2)\rangle$ trace a line from the point $\langle x_1, f(x_1)\rangle$ to the point $\langle x_2, f(x_2)\rangle$. Jensen's inequality states that this line is everywhere at least as large as $f(x)$.

> **Definition:** A function f from the reals to the reals is convex if for every x_1 and x_2 and every $p \in [0,1]$ we have
>
> $$pf(x_1) + (1-p)f(x_2) \geq f(px_1 + (1-p)x_2).$$

If f is (doubly) differentiable then f is convex if and only if $d^2f/dx^2 \geq 0$.

Now consider a probability distribution P on a set M and a function X assigning real values $X(m)$ for $m \in M$.

Theorem 1 (Jensen's Inequality) *If f is convex then for any distribution P on M we have the following.*

$$\mathrm{E}_{\mathrm{m}\sim\mathrm{P}}\left[\mathrm{f}(\mathrm{X}(\mathrm{m}))\right] \geq \mathrm{f}\left(\mathrm{E}_{\mathrm{m}\sim\mathrm{P}}\left[\mathrm{X}(\mathrm{m})\right]\right)$$

Usually the right hand side above — f of an expectation — is simpler than the left hand side — the expectation of f. Jensen's inequality is used to bound the "complicated" expression $\mathrm{E}\left[\mathrm{f}(\mathrm{X})\right]$ by the simpler expression $f(\mathrm{E}\left[\mathrm{X}\right])$. Often these expression are actually very close to each other. (Assuming that these expressions are equal is called the mean field approximation).

We prove Jensen's inequality only for the case where M is a finite set $\{m_1, \ldots, m_k\}$. Let x_i abbreviates $X(m_i)$ and p_i abbreviates $P(m_i)$. First consider the case where M contains only two elements. In this case we have the following.

$$\begin{aligned} \mathrm{E}_{\mathrm{m} \sim \mathrm{M}}\left[\mathrm{f}(\mathrm{X}(\mathrm{m})]\right. & = p_1 f(x_1) + p_2 f(x_2) \\ & \geq f(p_1 x_1 + p_2 x_2) \\ & = f(\mathrm{E}_{\mathrm{m} \sim \mathrm{M}}[\mathrm{X}(\mathrm{m})]) \end{aligned}$$

Note that the definition of convexity is simply the statement that Jensen's inequality holds for two point distributions. We prove Jensen's inequality for finite M by induction on the number of elements of M. Suppose M contains k elements and assume that Jensen's inequality holds for distributions on $k-1$ points. We now have the following where the fourth line follows from the induction hypothesis.

$$\begin{aligned} & \mathrm{E}_{\mathrm{m} \sim \mathrm{P}}[\mathrm{f}(\mathrm{X}(\mathrm{m}))] \\ = & \; p_1 f(x_1) + p_2 f(x_2) + p_3 f(x_3) + \cdots + p_k f(x_k) \\ = & \; (p_1 + p_2)\left(\left(\frac{p_1}{p_1 + p_2}\right) f(x_1) + \left(\frac{p_2}{p_1 + p2}\right) f(x_2)\right) + p_3 f(x_3) + \cdots + p_k f(x_k) \\ \leq & \; (p_1 + p_2) f\left(\left(\frac{p_1}{p_1 + p_2}\right) x_1 + \left(\frac{p_2}{p_1 + p_2}\right) x_2\right) + p_3 f(x_3) + \cdots + p_k f(x_k) \\ \leq & \; f\left((p_1 + p_2)\left(\frac{p_1 x_1}{p_1 + p_2} + \frac{p_2 x_2}{p_1 + p_2}\right) + p_3 x_3 + \cdots p_k x_k\right) \\ = & \; f(p_1 x_1 + p_2 x_2 + p_3 x_3 + \cdots p_k x_k) \\ = & \; f(\mathrm{E}_{\mathrm{m} \sim \mathrm{M}}[\mathrm{X}](\mathrm{m})) \end{aligned}$$

The definition of convexity generalizes to the case where f is a function from vectors to reals and x_1 and x_2 are taken to be vectors. Jensen's inequality also generalizes to the case where $X(m)$ is a vector. In this case $\mathrm{E}_{\mathrm{m} \sim \mathrm{P}}[\mathrm{X}(\mathrm{m})]$ is an average vector. In the vector case the above definitions and derivations go through unchanged.

1 Problems

1. Which of the following functions are convex. (Hint: compute the second derivative.)

- x^3 (on all reals)
- $2x^2 - 3x + 1$ (on all reals)
- $x^2 - \ln x$ (on positive reals only)
- $\ln\left(\frac{1}{1+e^x}\right)$ (on all reals)

2. Show that if f and g are convex then so is $f + g$ and $\max(f, g)$.

3. Use Jensen's inequality to show that if f is convex and $a_i > 0$ then we have the following.

$$\sum_i a_i f(x_i) \geq \left(\sum_i a_i\right) f\left(\frac{\sum_i a_i x_i}{\sum_i a_i}\right)$$