INEQUALITIES IN ANALYSIS

Darren Yao

ABSTRACT

Inequalities find important uses in algebra and analysis, either to find extreme values of functions, or to bound values that cannot be computed exactly. In this paper, we prove some basic inequalities like the AM-GM and Cauchy-Schwarz, generalize Cauchy-Schwarz to integrals, and discuss the various forms and applications of Jensen's inequality.

1 Introduction

When we first learn about inequalities, the first thing we learn is that all perfect squares are nonnegative. This is known as the trivial inequality, which many other inequalities can be reduced to.

Theorem 1.1 (Trivial Inequality)**.** *For any real number x, we have $x^2 \geq 0$.*

Next, we look at two other well-known inequalities: the arithmetic mean-geometric mean inequality, and the discrete form of the Cauchy-Schwarz inequality.

Theorem 1.2 (AM-GM inequality)**.** *If $a_1, a_2, \dots, a_n$ are positive real numbers, then*

$$\sqrt[n]{a_1 a_2 \cdots a_n} \leq \frac{a_1 + a_2 + \cdots + a_n}{n}$$

with equality if and only if all a_i's are equal.

We present Cauchy's proof, using a technique known as forward-backward induction. Where normal induction shows that k implies $k+1$, forward-backward induction breaks the inductive step into two parts: showing that k implies $2k$ and showing that k also implies $k-1$.

Proof. Let $P(n)$ represent the inequality in n variables, where we take both sides to the n^{th} power:

$$a_1 a_2 \cdots a_n \leq \left(\frac{a_1 + a_2 + \cdots + a_n}{n}\right)^n.$$

Our base case is $n = 2$, for which

$$a_1 a_2 \leq \left(\frac{a_1 + a_2}{2}\right)^2 \iff (a_1 - a_2)^2 \geq 0.$$

- Backward step: going from $P(n)$ to $P(n-1)$: Let

  $$A = \sum_{k=1}^{n-1} \frac{a_k}{n-1}.$$

  Then

  $$\left(\prod_{k=1}^{n-1}\right) A \overset{P(n)}{\leq} \left(\frac{\sum_{k=1}^{n-1} a_k + A}{n}\right)^n = \left(\frac{(n-1)A + A}{n}\right)^n = A^n.$$

  It follows that

  $$\prod_{k=1}^{n-1} a_k \leq A^{n-1} = \left(\frac{\sum_{k=1}^{n-1} a_k}{n-1}\right) \leq A^{n-1} = \left(\frac{\sum_{k=1}^{n-1} a_k}{n-1}\right)^{n-1}.$$

- Forward step: going from $P(n)$ to $P(2n)$:

$$\prod_{k=1}^{2n} a_k = \left(\prod_{k=1}^{n} a_k\right)\left(\prod_{k=n+1}^{2n} a_k\right) \overset{P(n)}{\leq} \left(\sum_{k=1}^{n} \frac{a_k}{n}\right)^n \left(\sum_{k=n+1}^{2n} \frac{a_k}{n}\right)^n$$
$$\overset{P(2)}{\leq} \left(\frac{\sum_{k=1}^{2n} \frac{a_k}{n}}{2}\right)^{2n} = \left(\frac{\sum_{k=1}^{2n} a_k}{2n}\right)^{2n}.$$

This completes the proof. Showing the equality case follows similarly. ■

Remark 1.3. We can extend this to show that the harmonic mean is always less than or equal to the geometric mean, by considering $\frac{1}{a_1}, \frac{1}{a_2}, \ldots, \frac{1}{a_n}$.

Theorem 1.4 (Discrete Cauchy-Schwarz Inequality)**.** *If $a_1, a_2, \ldots a_n$ and $b_1, b_2, \ldots b_n$ are real numbers, then*

$$\left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right) \geq \left(\sum_{i=1}^{n} a_i b_i\right)^2.$$

Remark 1.5. The discrete Cauchy-Schwarz inequality can also be stated in this alternate form: if $\langle \mathfrak{a}, \mathfrak{b}\rangle$ is the inner product (dot product) of two vectors in $\mathbb{R}^n$, then $\langle \mathfrak{a}, \mathfrak{b}\rangle^2 \leq |\mathfrak{a}|^2|\mathfrak{b}|^2$, with equality only if $\mathfrak{a}$ and $\mathfrak{b}$ are linearly dependent.

We present the following proof from the original Proofs from THE BOOK:

Proof. Consider the following function, which is quadratic in x:

$$|x\mathfrak{a} + \mathfrak{b}|^2 = x^2|\mathfrak{a}|^2 + 2x\langle \mathfrak{a}, \mathfrak{b}\rangle + |\mathfrak{b}|^2.$$

We trivially assume $\mathfrak{a} \neq \vec{0}$ (that case is just the zero equality case). If $\mathfrak{a}$ and $\mathfrak{b}$ are linearly dependent then $\mathfrak{b} = \lambda\mathfrak{a}$ and $\langle \mathfrak{a}, \mathfrak{b}\rangle^2 = |\mathfrak{a}|^2|\mathfrak{b}|^2$. Otherwise, $|x\mathfrak{a} + \mathfrak{b}|^2 > 0$ for all x. Since this is always positive, it has no real roots, so the discriminant $4(\langle \mathfrak{a}, \mathfrak{b}\rangle^2 - |\mathfrak{a}|^2|\mathfrak{b}|^2)$ is negative. Dividing by 4 and rearranging finishes the proof. ■

2 Integral Form of Cauchy-Schwarz

When we take a sum of infinitely many small rectangles under a curve, we get an integral. Similarly, extending the discrete form of the Cauchy-Schwarz inequality to the case of $n = \infty$, gives us the integral form.

Theorem 2.1 (Continuous Cauchy-Schwarz Inequality)**.** *Let f and g be functions for which integrals are defined, and S be a space we can integrate over, such as an interval, union of intervals, etc. Then, we have*

$$\int_S |fg| \leq \left(\int_S |f|^2\right)^{\frac{1}{2}} \left(\int_S |g|^2\right)^{\frac{1}{2}}.$$

Remark 2.2. The more formal measure-theoretic statement of the inequality is as follows:

Let $f, g : S \to \mathbb{C}$ be measurable functions on a measure space S.

The proof is very similar to the discrete version:

Proof. For any real x, we have $0 \leq (x|f| + |g|)^2$.

Then, integrating over the space S, we have

$$0 \leq \int_S (x|f| + |g|)^2.$$

Expanding gives

$$x^2 \int_S |f|^2 + 2x \int_S |f||g| + \int_S |g|^2$$
$$= Ax^2 + 2Bx + C,$$

where

$$A = \int_S |f|^2; B = \int_S |f||g|; C = \int_S |g|^2.$$

By the same reasoning as in the discrete inequality, $(2B)^2 - 4AC$ must be nonpositive, thus $B^2 \leq AC$ and hence the result. The equality case follows similarly. ■

3 Jensen's Inequality

The previous inequalities that we studied were all derived from the trivial inequality. Now, we look at Jensen's inequality, which is derived from convexity of functions.

Definition 3.1. A function f on an n-dimensional interval is *convex* if the line segment between any two points on the graph of the function lies on or above the graph. More rigorously: Let f be a function on an interval S. Then, f is convex if for any two points $x, y \in S$, and any $0 < t < 1$, we have

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y).$$

We say f is *strictly convex* if the inequality is strict.

The definitions of *concave* and *strictly concave* functions follow from reversing the direction of the inequality.

Remark 3.2. When $t = \frac{1}{2}$, the inequality reduces to the well-known form

$$f\left(\frac{x+y}{2}\right) \le \frac{f(x)+f(y)}{2}.$$

Convex functions also satisfy several useful properties. We will not prove any of them, but they will soon come into use:

- Convex functions are always continuous, but not always differentiable. The absolute value function is a counterexample to differentiability.
- If a function $f : S \to \mathbb{R}$ is twice differentiable over all of S, then f is convex if and only if $f''(x) \ge 0$ throughout all of S.
- $f(\frac{x+y}{2}) \le \frac{f(x)+f(y)}{2}$ does not guarantee convexity, unless we know that f is also continuous over the entire region.

With this, we can define Jensen's inequality, which is essentially an n-dimensional generalization of the definition of convexity.

Theorem 3.3 (Jensen's inequality, general version)**.** *Let $f : S \to \mathbb{R}$ be a convex function. Given $x_1, x_2, \ldots, x_n \in S$ and positive real numbers $t_1 \ldots t_n$ such that $\sum t_i = 1$, we have:*

$$f(t_1x_1 + t_2x_2 + \cdots + t_nx_n \le t_1f(x_1) + t_2f(x_2) + \cdots + t_nf(x_n).$$

The inequality is reversed if f is concave.

Remark 3.4. If we let $t_1 = t_2 = \cdots = t_n = \frac{1}{n}$, then the inequality becomes the simple form

$$f\left(\frac{x_1+x_2+\cdots+x_n}{n}\right) \le \frac{f(x_1)+f(x_2)+\cdots+f(x_n)}{n}.$$

Proof. We apply induction on n.

The base case is $n = 2$, which is trivial by the definition of convexity.

Assume the result holds for some integer $n \ge 2$. Given $x_1, x_2, \ldots, x_n, x_{n+1} \in S$ and $t_1, t_2, \ldots, t_n, t_{n+1}$ such that $\sum t_i = 1$, we can write

$$\begin{aligned} f\left(\sum_{i=1}^{n+1} t_ix_i\right) &= f\left(t_{n+1}x_{n+1} + (1-t_{n+1})\sum_{i=1}^{n}\frac{t_i}{1-t_{n+1}}x_i\right) \\ &\le t_{n+1}f(x_{n+1}) + (1-t_{n+1})f\left(\sum_{i=1}^{n}\frac{t_i}{1-t_{n+1}}x_i\right). \end{aligned}$$

Since $\sum_{i=1}^{n}\frac{t_i}{(1-t_{n+1})} = 1$, by the inductive hypothesis we have

$$f\left(\sum_{i=1}^{n+1} t_ix_i\right) \le t_{n+1}f(x_{n+1}) + (1-t_{n+1})\sum_{i=1}^{n}\frac{t_i}{1-t_{n+1}}f(x_i) = \sum_{i=1}^{n+1} t_if(x_i).$$

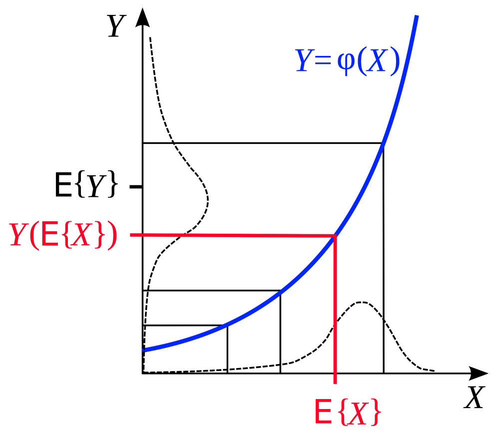
■

Similar to how the Cauchy-Schwarz inequality can be generalized to an integral form in infinite dimensions, Jensen's inequality can be generalized to an integral form with infinite number of terms in the sum. This lends itself to a convenient application for probability functions. Unfortunately, we will not prove the inequality here as it is beyond the scope of this paper.

Theorem 3.5 (Jensen's inequality for integration on probability spaces)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function, and S be a measure space with total measure (length, area, etc.) equal to 1. Given an integrable function $g : X \to \mathbb{R}$, we have:*

$$f\left(\int_S g\right) \leq \int_S f \circ g.$$

Here is a graphical explanation of the integral form of Jensen's [5]:



Example (Standard combinatorics problem). Suppose we assign one of 10 colors to each of the integers $\{1, 2, 3, \ldots, 100\}$. At least how many unordered pairs $\{a, b\} \subset \{1, 2, 3, \ldots, 100\}$ must be the same color?

Let x_i be the number of integers that get color i. We know that $\sum x_i = 100$. If color i has x_i integers, then there are $\frac{x_i(x_i-1)}{2}$ pairs of integers that both have color i. Thus, we want to minimize $\sum \binom{x_i}{2}$. Since the function $\binom{x}{2}$ is convex, by the equality case of Jensen's, this is minimized when $x_1 = x_2 = \cdots = x_10 = 10$. Then there are $\binom{10}{2} = 45$ pairs of each color, and 450 pairs total.

4 Proving other inequalities using Jensen

In this section, we give alternate proofs of AM-GM and Cauchy-Schwarz using Jensen's inequality.

Alternate proof of Theorem 1.2:

Proof. Note that the function $f(x) = \ln x$ is concave. By Jensen's inequality,

$$\ln\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) \geq \sum_{i=1}^{n} \frac{1}{n} \ln x_i = \sum_{i=1}^{n} \left(lnx_i^{1/n}\right) = \ln\left(\prod_{i=1}^{n} x_i^{1/n}\right).$$

Taking the exponential of the far left and far right terms makes the $\ln$s go away, and we are left with the AM-GM inequality. ∎

Alternate proof of Theorem 1.4

Proof. We let $p_1, p_2, \ldots, p_n$ be such that $p_i > 0$ for all i, and $\sum_{i=1}^{n} p_i = 1$. Since $f(x) = x^2$ is convex, by Jensen's we have

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

Suppose that none of the b_i's are zero. Letting $x_i = \frac{a_i}{b_i}$ and $p_i = \frac{b_i^2}{\sum_{i=1}^n b_i^2}$, we have

$$\left(\frac{\sum_{i=1}^n a_i b_i}{\sum_{i=1}^n b_i^2}\right)^2 \leq \frac{\sum_{i=1}^n a_i^2}{\sum_{i=1}^n b_i^2}$$

Clearing the denominator, we have

$$\left(\sum_{i=1}^n a_i b_i\right)^2 \leq \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right),$$

as desired.

Now, if at least one of the b_i's are zero, suppose $b_{i_1} = b_{i_2} = \cdots = b_{i_k} = 0$. Then,

$$\left(\sum_{i=1}^n a_i b_i\right)^2 = \left(\sum_{1\leq i\leq n, b_i\neq 0} a_i b_i\right)^2 \leq \left(\sum_{1\leq i\leq n, b_i\neq 0} a_i^2\right)\left(\sum_{1\leq i\leq n, b_i\neq 0} b_i^2\right) \leq \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right).$$

■

5 Probabilistic Applications

Jensen's inequality can be applied to solve many problems in probability and statistics. In fact, there is yet another form of Jensen's for random variables:

Definition 5.1. A *random variable* is a variable whose value depends on the outcome of a random phenomenon. Formally, a random variable is a function from the sample space (the set of all possible outcomes) Ω to the real numbers $\mathbb{R}$. In the measure-theoretic definition, X is also required to be a measurable function.

Theorem 5.2 (Jensen's inequality for a random variable)**.** *Let X be an integrable random variable. If $g : \mathbb{R} \to \mathbb{R}$ is a convex function such that $g(X)$ is also integrable, we have*

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

with the inequality reversed if g is concave.

Proof. A function g is convex if for any point x_0, the graph of g lies above its tangent at the point x_0:

$$g(x) \geq g(x_0) + b(x - x_0)\text{for all x.}$$

If we let $x = X$ and $x_0 = E[X]$, the inequality becomes

$$g(X) \geq g(E[X]) + b(X - E[X]).$$

Taking expected value of both sides, we have

$$E[g(X)] \geq E[g(E[X])] + b(X - E[X]).$$

By linearity of expectation:

$$E[g(X)] \geq g(E[X]) + b(E[X] - E[X]);$$
$$E[g(x)] \geq g(E[X]).$$

■

Example. Suppose a strictly positive nonconstant random variable X has expected value $E[X] = 1$. What bound can we find on the expected value of $\ln X$?

We know that the function $g(x) = \ln x$ is concave, because its second derivative $g''(x) = -\frac{1}{x^2}$ is always negative.

By Jensen's inequality, we have

$$E[\ln X] \leq \ln E[X] = \ln 1 = 0.$$

Therefore, $E[\ln X] \leq 0$.

References

[1] Martin, Charles: A Primer on Inequalities. UC Santa Barbara http://web.math.ucsb.edu/~cmart07/inequalities.pdf

[2] Aigner, Martin and Ziegler, Gunter M.: Proofs from THE BOOK, Sixth Edition

[3] Pavic, Z., Pecaric, J., and Peric, I.: Integral, Discrete, and Functional Variants of Jensen's Inequality. Journal of Mathematical Inequalities http://files.ele-math.com/articles/jmi-05-23.pdf

[4] Tao, Terence. An Introduction to Measure Theory. Department of Mathematics, UCLA https://terrytao.files.wordpress.com/2011/01/measure-book1.pdf

[5] Unknown public-domain image. Retrieved from https://en.wikipedia.org/wiki/Jensen%27s_inequality#/media/File:Jensen_graph.svg