The Pearson and Cauchy-Schwarz Inequalities

David Rose (drose@flsouthern.edu), Florida Southern College, Lakeland, FL 33801

Recently, there has been renewed interest in the Cauchy-Schwarz inequality and particularly in its applications outside vector algebra [1]. The Cauchy-Schwarz inequality for vectors in $\mathbb{R}^n$ asserts that for any two vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$, $(\mathbf{x} \cdot \mathbf{y})^2 \leq \|\mathbf{x}\|^2 \|\mathbf{y}\|^2$, where $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} (x_i y_i)$, $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} (x_i)^2}$, and $\|\mathbf{y}\| = \sqrt{\sum_{i=1}^{n} (y_i)^2}$.

For a sample of $n$ data pairs $(x_1, y_1), (x_1, y_2), \ldots, (x_n, y_n)$, Pearson’s coefficient of linear correlation is defined as (see [2] for example)

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}},$$

where as usual $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$. Pearson’s inequality, $|r| \leq 1$, is immediate from the Cauchy-Schwarz inequality applied to the vectors $\mathbf{d}_x = (x_1 - \bar{x}, \ldots, x_n - \bar{x})$ and $\mathbf{d}_y = (y_1 - \bar{y}, \ldots, y_n - \bar{y})$.

The purpose of this note is to prove the converse.

**Theorem 1.** For any two vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$, if $|r| \leq 1$, then $\left|\sum_{i=1}^{n} (x_i y_i)\right| \leq \sqrt{\sum_{i=1}^{n} (x_i)^2} \sqrt{\sum_{i=1}^{n} (y_i)^2}$.

**Proof.** We apply Pearson’s coefficient formula to two new vectors

$$\mathbf{x}^* = \left(\frac{-x_1}{\sqrt{2}}, \frac{-x_2}{\sqrt{2}}, \ldots, \frac{-x_n}{\sqrt{2}}, \frac{x_1}{\sqrt{2}}, \frac{x_2}{\sqrt{2}}, \ldots, \frac{x_n}{\sqrt{2}}\right)$$

and

$$\mathbf{y}^* = \left(\frac{-y_1}{\sqrt{2}}, \frac{-y_2}{\sqrt{2}}, \ldots, \frac{-y_n}{\sqrt{2}}, \frac{y_1}{\sqrt{2}}, \frac{y_2}{\sqrt{2}}, \ldots, \frac{y_n}{\sqrt{2}}\right)$$

in $\mathbb{R}^{2n}$. Since $\bar{x} = \bar{y} = 0$,

$$\left|\frac{\sum_{i=1}^{n} (x_i^* y_i^*)}{\sqrt{\sum_{i=1}^{n} (x_i^*)^2} \sqrt{\sum_{i=1}^{n} (y_i^*)^2}}\right| \leq 1$$

implies that

$$\left|2 \sum_{i=1}^{n} \frac{(x_i y_i)}{2}\right| \leq \sqrt{2 \sum_{i=1}^{n} \frac{(x_i)^2}{2}} \sqrt{2 \sum_{i=1}^{n} \frac{(y_i)^2}{2}},$$

from which the desired result follows. $\blacksquare$

Thus Pearson’s inequality is equivalent to the Cauchy-Schwarz inequality.

**References**