# 1. Cauchy–Schwarz Inequality

For any two real (or complex) sequences (or vectors)

$$a=(a_1,a_2,\ldots,a_n),\ b=(b_1,b_2,\ldots,b_n)$$

the Cauchy–Schwarz inequality states:

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

This inequality is fundamental in linear algebra, functional analysis, and probability theory. In statistics, it guarantees that the **correlation coefficient**

$$r = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

always satisfies $|r| \leq 1$.

## Proof (Quadratic Function Method)

We construct a quadratic function that must always be nonnegative:

$$f(t) = \sum_{i=1}^{n} (a_i - t b_i)^2.$$

1. **Nonnegativity**:
   Since squares are always nonnegative, we have $f(t) \geq 0$ for all real t.
2. **Expand the expression**:

$$f(t) = \sum_{i=1}^{n} a_i^2 - 2t \sum_{i=1}^{n} a_i b_i + t^2 \sum_{i=1}^{n} b_i^2.$$

This is a quadratic in t:

$$f(t) = At^2 - 2Bt + C,$$

With

$$A = \sum_{i=1}^{n} b_i^2, \quad B = \sum_{i=1}^{n} a_i b_i, \quad C = \sum_{i=1}^{n} a_i^2.$$

3. **Discriminant condition**:
   For f(t) to be nonnegative for all t, the discriminant must satisfy:

$$\Delta = (2B)^2 - 4AC = 4B^2 - 4AC \leq 0,$$

which simplifies to

$$B^2 \leq AC.$$

4. **Conclusion**:

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

5. **Equality condition**:
   Equality holds if and only if there exists a constant λ such that

$$a_i = \lambda b_i \text{ for all } i,$$

i.e., when the vectors a and b are **linearly dependent**.

## Interpretation in Statistics

If we let $a_i = X_i - \mathbb{E}[X]$ and, $b_i = Y_i - \mathbb{E}[Y]$ then the inequality directly shows that the

**Pearson correlation coefficient** satisfies:

$$|r| \leq 1.$$

Thus, Cauchy–Schwarz is the mathematical foundation behind why correlation must always lie between −1 and +1.

# 2. Concepts of Independence and Uncorrelation

## Independence

Two random variables X and Y are **independent** if knowledge of one gives no information about the other. Formally, for all measurable sets A,B:

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B).$$

- This is a **strong condition**: it means that the joint distribution factors into the product of the marginals.
- Independence eliminates **all forms of dependence**—both linear and nonlinear.

## Uncorrelation

Two random variables are **uncorrelated** if their covariance is zero:

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = 0.$$

- This condition only rules out **linear relationships**.
- However, X and Y may still be related **nonlinearly**.

## Key Differences

- **Strength**:
  - Independence $\implies$ uncorrelation (if variances are finite).
  - Uncorrelation $\implies$ independence.
- **Implications**:
  - Independence guarantees that **all possible dependencies vanish**.
  - Uncorrelation only ensures the absence of **linear dependence**.

**Example**:
Let X∼Uniform[−1,1], and define $Y=X^2$.

- Then $\mathbb{E}[X] = 0$ and Cov (X,Y)=0.
- So X and Y are **uncorrelated**, but they are clearly **dependent** (since knowing X determines Y).

## Other Measures of Dependence

Since uncorrelation is limited, other measures are often used:

- **Mutual Information (MI):**
  Captures the overall dependency (linear and nonlinear). $MI(X;Y)= 0$ if and only if X and Y are independent.
- **Distance Correlation (dCor):**
  Detects both linear and nonlinear dependencies. It is zero if and only if X and Y are independent.

## Summary

1. **Cauchy–Schwarz Inequality**:
   - Proved by constructing a nonnegative quadratic function.
   - Ensures that correlation coefficients always lie in $[-1,1]$.
   - Equality holds only when the two vectors are proportional.
2. **Independence vs. Uncorrelation**:
   - Independence is much stronger, eliminating **all** dependencies.
   - Uncorrelation only eliminates **linear dependence**.
   - Example: X uniform on $[-1,1]$, $Y= X^2$: uncorrelated but not independent.
   - More general dependence can be measured using **mutual information** and **distance correlation.**