



Università  
di Genova



# Evaluating Two-Sample Tests for validating generators in precision sciences

Samuele Grossi

European AI for Fundamental Physics Conference

Cagliari, 17<sup>th</sup> June 2025

Based on [2409.16336](#) and on a work in progress in collaboration with  
M. Letizia and R. Torre

# Two-sample test in high energy physics

**Two-Sample Test:** understand if two independent data samples are drawn from the same probability density function (PDF)

- **PARAMETRIC:** Some assumptions on the underlying distributions of the samples are needed to perform the test
- **NON-PARAMETRIC:** Only the data are used to perform the test, without any assumption on the underlying distributions

**In high energy physics:** two-sample test to compare data provided by two different generators.  
Example: Powheg/MadGraph vs Neural Networks

# Objectives and procedure

**Purpose of the work:** provide a systematic analysis of non-parametric two-sample test using different evaluation metrics, including both traditional statistical and machine learning-based approaches.

Particle physics  high dimensional datasets

- Ensure a fair comparison between different test statistics across a wide range of scenarios
- Propose simple, 1D-distance-based metrics for fast validation.  
Include a kernel based classifier to assess performance gains
- Enable the comparison with other metrics with the same procedure:  
the whole framework is available online at <https://github.com/TwoSampleTests>