EUROPEAN AI FOR
FUNDAMENTAL PHYSICS
CONFERENCE
EuCAIFCon 2024

Contribution ID: **106**   Type: **Flashtalk with Poster**

Evaluating Generative Models with non-parametric two-sample tests

Tuesday, 30 April 2024 17:16 (3 minutes)

The problem of comparing two high-dimensional samples to test the null hypothesis that they are drawn from the same distribution is a fundamental question in statistical hypothesis testing. This study presents a comprehensive comparison of various non-parametric two-sample tests, specifically focusing on their statistical power in high-dimensional settings. The tests are built from univariate tests and are selected for their computational efficiency, as they all possess closed-form expressions as functions of the marginal empirical distributions. We use toy mixture of Gaussian models with dimensions ranging from 5 to 100 to evaluate the performance of different test-statistics: mean of 1D Kolmogorov-Smirnov (KS) tests-statistics, sliced KS test-statistic, and sliced-Wasserstein distance. We also add to the comparison two recently proposed multivariate two-sample tests, namely the Fr\'echet and kernel physics distances and compare all test-statistics against a likelihood ratio test, which serves as the gold standard due to the Neyman-Pearson lemma. All tests are implemented in Python using \textsc{TensorFlow2} and made available on \textsc{GitHub} \href{https://github.com/NF4HEP/GenerativeModelsMetrics}{\faGithub}. This allows us to leverage hardware acceleration for efficient computation of the test-statistic distribution under the null hypothesis on Graphic Processing Units. Our findings reveal that while the likelihood ratio test-statistic remains the most powerful, certain non-parametric tests exhibit competitive performance in specific high-dimensional scenarios. This study provides valuable insights for practitioners in selecting the most appropriate two-sample test for evaluating generative models, thereby contributing to the broader field of model evaluation and statistical hypothesis testing.

Primary authors: TORRE, Riccardo (INFN, Sezione di Genova);   Dr GROSSI, Samuele (University of Genova)

Presenter: Dr GROSSI, Samuele (University of Genova)

Session Classification: 3.3 Hardware acceleration, FPGAs & Uncertainty quantification

Track Classification: Session B