

# Comparative benchmarks for statistical analysis frameworks in Python

*Wednesday 11 October 2023 15:00 (30 minutes)*

The statistical models that are used in modern HEP research are independent of their specific implementations. As a consequence, many different tools have been developed to perform statistical analyses in HEP. These implementations differ both in their performance, but also in their usability. In this scenario, comparative benchmarks are essential to aid users in choosing a library and to support developers in resolving bottlenecks or usability shortcomings.

This contribution showcases a comparison of some of the most used tools in statistical analyses. The python package `pyhf` focuses on template histogram fits, which are the backbone of many analyses at the ATLAS and CMS experiments. In contrast, `zfit` mainly focuses on unbinned fits to analytical functions. The third tool, RooFit, is part of the ROOT analysis framework, implemented in C++ but with Python bindings. RooFit can perform both template histogram and unbinned fits and is thus compared to both other packages.

To benchmark `pyhf` and RooFit, the "Simple Hypothesis Testing" example from the `pyhf` tutorial is used. The complexity of this example is scaled up by increasing the number of bins and measurement channels. In the benchmark of `zfit` and RooFit, the same unbinned fits are used as in the original `zfit` paper.

In all frameworks, the minimizer Minuit2 is used with identical settings, leading to a comparison of the different likelihood evaluation backends. While working on these benchmarks, we were in contact with developers of all three benchmarked frameworks. This talk will focus on the results of the comparative benchmarks between the different frameworks.

**Authors:** WERNER, Daniel; REMBSER, Jonas (CERN); MONETA, Lorenzo (CERN)

**Presenter:** WERNER, Daniel

**Session Classification:** Plenary Session Wednesday