The Journal of Open Source Software

pyhf: pure-Python implementation of HistFactory statistical models

Lukas Heinrich[1], **Matthew Feickert**[*2], **Giordon Stark**[3], **and Kyle Cranmer**[4]

1 CERN **2** University of Illinois at Urbana-Champaign **3** SCIPP, University of California, Santa Cruz **4** New York University

DOI: 10.21105/joss.02823

Software

- Review
- Repository
- Archive

Editor: Eloisa Bentivegna

Reviewers:

- @suchitakulkarni
- @bradkav

Submitted: 07 October 2020
Published: 04 February 2021

Summary

Statistical analysis of High Energy Physics (HEP) data relies on quantifying the compatibility of observed collision events with theoretical predictions. The relationship between them is often formalised in a statistical model $f(\mathbf{x}|\phi)$ describing the probability of data $\mathbf{x}$ given model parameters ϕ. Given observed data, the likelihood $\mathcal{L}(\phi)$ then serves as the basis for inference on the parameters ϕ. For measurements based on binned data (histograms), the `HistFactory` family of statistical models (Cranmer et al., 2012) has been widely used in both Standard Model measurements (ATLAS Collaboration, 2013) as well as searches for new physics (ATLAS Collaboration, 2018). `pyhf` is a pure-Python implementation of the `HistFactory` model specification and implements a declarative, plain-text format for describing `HistFactory`-based likelihoods that is targeted for reinterpretation and long-term preservation in analysis data repositories such as HEPData (Maguire et al., 2017). The source code for `pyhf` has been archived on Zenodo with the linked DOI: (Heinrich, Lukas and Feickert, Matthew and Stark, Giordon, 2020). At the time of writing this paper, the most recent release of `pyhf` is `v0.5.4`.

Statement of Need

Through adoption of open source "tensor" computational Python libraries, `pyhf` decreases the abstractions between a physicist performing an analysis and the statistical modeling without sacrificing computational speed. By taking advantage of tensor calculations, `pyhf` outperforms the traditional `C++` implementation of `HistFactory` on data from real LHC analyses. `pyhf`'s default computational backend is built from NumPy and SciPy, and supports TensorFlow, PyTorch, and JAX as alternative backend choices. These alternative backends support hardware acceleration on GPUs, and in the case of JAX JIT compilation, as well as auto-differentiation allowing for calculating the full gradient of the likelihood function — all contributing to speeding up fits.

Impact on Physics

In addition to enabling the first publication of full likelihoods by an LHC experiment (ATLAS Collaboration, 2019), `pyhf` has been used by the `SModelS` library to improve the reinterpretation of results of searches for new physics at LHC experiments (Abdallah & others, 2020; Alguero et al., 2020; Khosa et al., 2020).

*Corresponding author.

Future work

Future development aims to provide support for limit setting through pseudoexperiment generation in the regimes in which asymptotic approximations (Cowan et al., 2011) are no longer valid. Further improvements to the performance of the library as well as API refinement are also planned.

Acknowledgements

We would like to thank everyone who has made contributions to `pyhf` and thank our fellow developers in the Scikit-HEP community for their continued support and feedback. Matthew Feickert and Kyle Cranmer have received support for work related to `pyhf` provided by NSF cooperative agreement OAC-1836650 (IRIS-HEP) and grant OAC-1450377 (DIANA/HEP).

References

Abdallah, W., & others. (2020). Reinterpretation of LHC Results for New Physics: Status and Recommendations after Run 2. *SciPost Phys.*, *9*(2), 022. https://doi.org/10.21468/SciPostPhys.9.2.022

Alguero, G., Kraml, S., & Waltenberger, W. (2020). *A SModelS interface for pyhf likelihoods*. http://arxiv.org/abs/2009.01809

ATLAS Collaboration. (2019, August). *Reproducing searches for new physics with the ATLAS experiment through publication of full statistical likelihoods* (ATL-PHYS-PUB-2019-029). CERN. https://cds.cern.ch/record/2684863

ATLAS Collaboration. (2013). Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC. *Phys. Lett. B*, *726*, 88. https://doi.org/10.1016/j.physletb.2014.05.011

ATLAS Collaboration. (2018). *Search for supersymmetry in final states with missing transverse momentum and multiple b-jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. ATLAS-CONF-2018-041. https://cds.cern.ch/record/2632347

Cowan, G., Cranmer, K., Gross, E., & Vitells, O. (2011). Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, *71*, 1554. https://doi.org/10.1140/epjc/s10052-011-1554-0

Cranmer, K., Lewis, G., Moneta, L., Shibata, A., & Verkerke, W. (2012). *HistFactory: A tool for creating statistical models for use with RooFit and RooStats* (CERN-OPEN-2012-016). New York U. https://cds.cern.ch/record/1456844

Heinrich, Lukas and Feickert, Matthew and Stark, Giordon. (2020). *pyhf: v0.5.4* (Version 0.5.4) [Computer software]. https://doi.org/10.5281/zenodo.1169739

Khosa, C. K., Kraml, S., Lessa, A., Neuhuber, P., & Waltenberger, W. (2020). SModelS database update v1.2.3. *LHEP*, *158*, 2020. https://doi.org/10.31526/lhep.2020.158

Maguire, E., Heinrich, L., & Watt, G. (2017). HEPData: a repository for high energy physics data. *J. Phys. Conf. Ser.*, *898*(10), 102006. https://doi.org/10.1088/1742-6596/898/10/102006