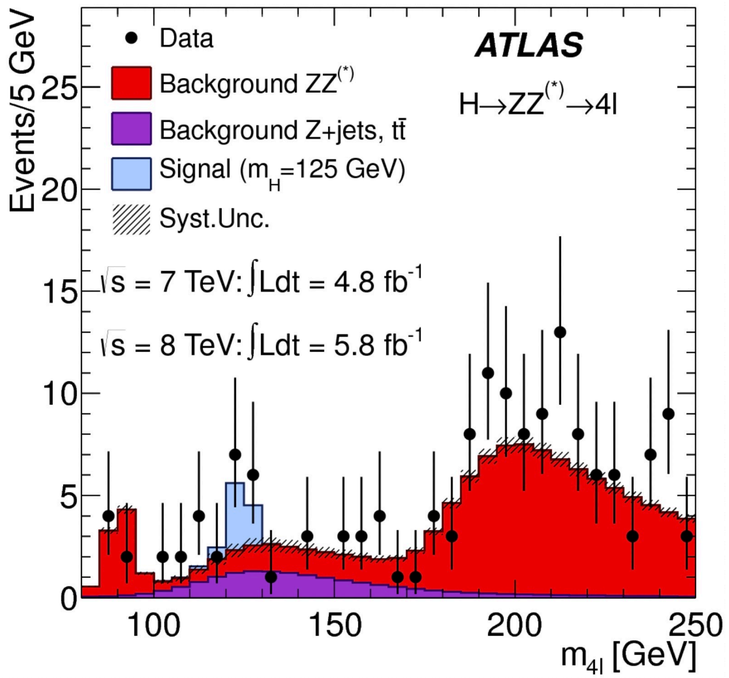
Bayesian and Frequentist Methodologies in Collider Physics with `pyhf`

Matthew Feickert, Lukas Heinrich, Malin Horstmann (malin.horstmann@tum.de)

11.09.2023, PyHEP23

~~What's pyhf?~~ What's `HistFactory`?



→ Tool box for creating data-generating models for binned analyses

https://cds.cern.ch/record/1456844?ln=en

https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2012-27/

~~What's pyhf?~~ What's `HistFactory`?

$\vec{\eta}$: Unconstrained Parameters
$\vec{\chi}$: Nuisance parameters

Constraint terms

$$p\left(\vec{n},\vec{a}\,|\,\vec{\eta},\vec{\chi}\right) = \prod_{c\,\in\,\text{Channels}} \prod_{b\,\in\,\text{Bins}} \text{Poiss}\left(n_{cb}\,|\,\nu_{cb}(\vec{\eta},\vec{\chi})\right) \prod_{\chi\,\in\,\vec{\chi}} c_\chi\left(a_\chi\,|\,\chi\right)$$

Poisson-distributed bin counts

$\vec{n}$: Bin counts
$\vec{a}$: Auxiliary data

https://cds.cern.ch/record/1456844?ln=en

What's `pyhf`?

- Pure Python implementation of the `HistFactory` framework
- Support auto-differentiation via multiple back-ends
- Already supports a wide range of frequentist analysis tools
- Also see Matthew's contribution

https://github.com/scikit-hep/pyhf

Statistics Exploration: Frequentist Analysis Methodologies

→ Likelihood: $L\left(\theta \,|\, x\right) = L_x(\theta)$

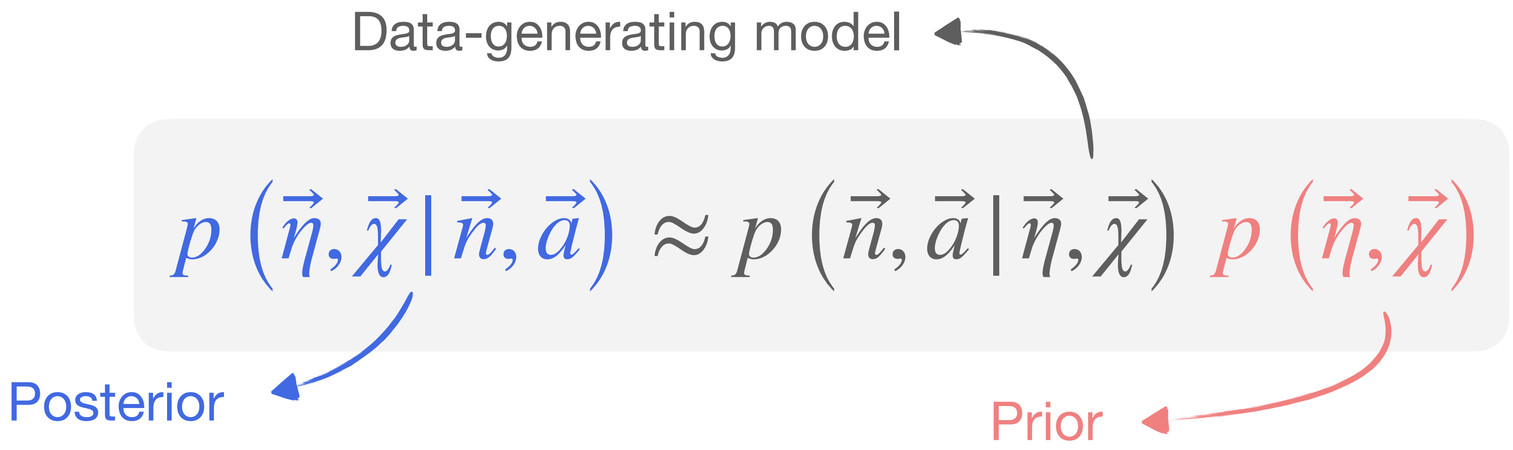
→ Maximum Likelihood Estimate:

$$\hat{\theta} = \text{argmax}_\theta \, \text{L}\left(\theta \,|\, x\right)$$

https://pyhf.github.io/pyhf-tutorial/HelloWorld.html

Statistics Exploration: Bayesian Analysis Methodologies

→ Bayes' Theorem:



→ Auxiliary measurements $\vec{a}$: Use your prior knowledge!

https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html

Statistics Exploration: Bayesian Analysis Methodologies

Step 1:
Use $\vec{a}$ to build priors for the constraint parameters $\vec{\chi}$

$$p\left(\vec{\eta},\vec{\chi}\,|\,\vec{n},\vec{a}\right) \approx p\left(\vec{n}\,|\,\vec{\eta},\vec{\chi}\right)\; p\left(\vec{\eta}\right)\;\; p\left(\vec{\chi}\,|\,\vec{a}\right)$$

https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html

Statistics Exploration: Bayesian Analysis Methodologies

Step 2:
Make an educated guess on the prior for the unconstrained parameters $\vec{\eta}$

Step 1:
Use $\vec{a}$ to build priors for the constraint parameters $\vec{\chi}$

$$p\left(\vec{\eta},\vec{\chi}\,|\,\vec{n},\vec{a}\right) \approx p\left(\vec{n}\,|\,\vec{\eta},\vec{\chi}\right)\; p\left(\vec{\eta}\right)\;\; p\left(\vec{\chi}\,|\,\vec{a}\right)$$

https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html

Statistics Exploration: Bayesian Analysis Methodologies

Step 3:
Update only with the actual observations

Step 2:
Make an educated guess on the prior for the unconstrained parameters $\vec{\eta}$

Step 1:
Use $\vec{a}$ to build priors for the constraint parameters $\vec{\chi}$

$$p\left(\vec{\eta},\vec{\chi}\,|\,\vec{n},\vec{a}\right) \approx p\left(\vec{n}\,|\,\vec{\eta},\vec{\chi}\right)\; p\left(\vec{\eta}\right)\; p\left(\vec{\chi}\,|\,\vec{a}\right)$$

https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html

Bayesian_pyhf

- Tools to enable Bayesian methodologies for `pyhf HistFactory` models
- Bayesian analysis is based on MCMC methods implemented with `PyMC`
- Plan to fully integrate with `pyhf` in the future

launch binder

DOI 10.5281/zenodo.7886632

https://github.com/malin-horstmann/Bayesian_pyhf

https://doi.org/10.21105/joss.00024

https://arxiv.org/abs/2309.17005

Backup

https://jupyterhub.ssl-hep.org/user/malin.horstmann@tum.de/malin-horstmann-pyhep23-zfm7r5yb/doc/tree/talk.ipynb

https://jupyterhub.ssl-hep.org/user/malin.horstmann@tum.de/malin-horstmann-pyhep23-kj9q42u8/doc/tree/talk.ipynb