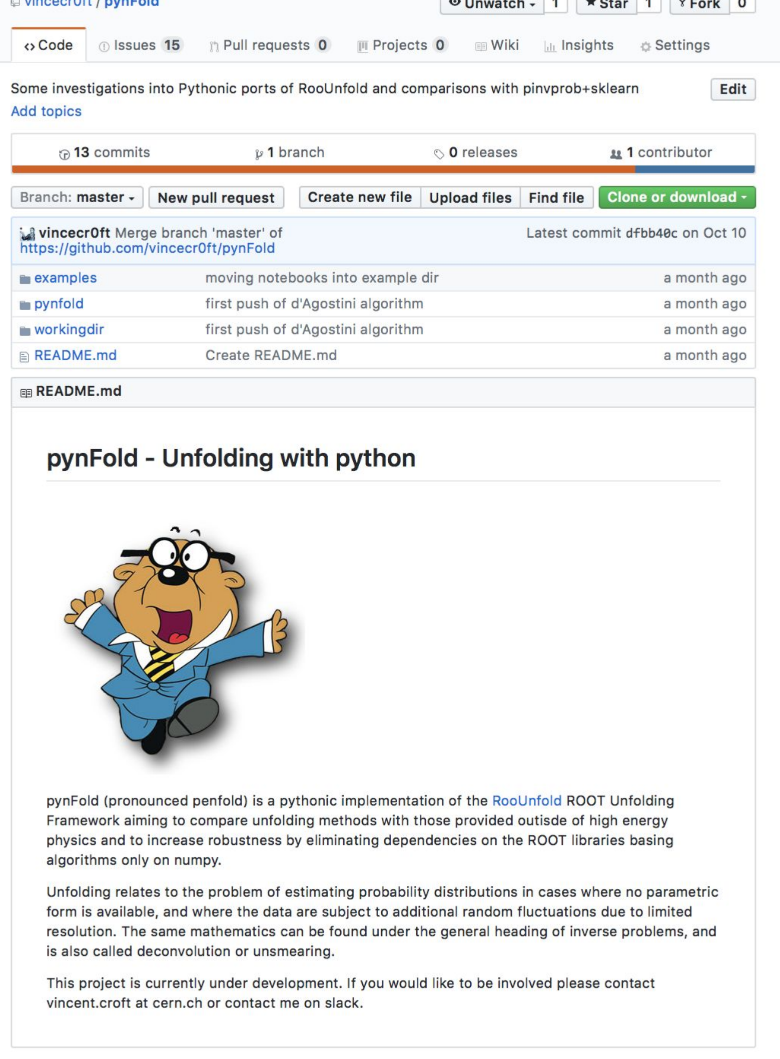
Histogram Unfolding in Python

Vince Croft

An Investigation into implementations of unfolding outside of the ROOT environment.

PynFold

- Used for first investigations:
  - ROOT -> pyROOT -> numpy
- Use cases:
  - Any ROOT independent analysis setup
- Independent of RooUnfold
  - Possible to use without any knowledge of ROOT
  - Decoupled from upstream algorithm implementations and fixes
- Useful for comparisons with existing non-hep implementations:
  - Other implementations do not use histograms.
  - 1:1 comparison not always possible
  - Useful for cross checks.



Some investigations into Pythonic ports of RooUnfold and comparisons with pinvprob+sklearn

vincecr0ft Merge branch 'master' of https://github.com/vincecr0ft/pynFold Latest commit dfbb40c on Oct 10

examples	moving notebooks into example dir	a month ago
pynfold	first push of d'Agostini algorithm	a month ago
workingdir	first push of d'Agostini algorithm	a month ago
README.md	Create README.md	a month ago

pynFold - Unfolding with python

pynFold (pronounced penfold) is a pythonic implementation of the RooUnfold ROOT Unfolding Framework aiming to compare unfolding methods with those provided outisde of high energy physics and to increase robustness by eliminating dependencies on the ROOT libraries basing algorithms only on numpy.

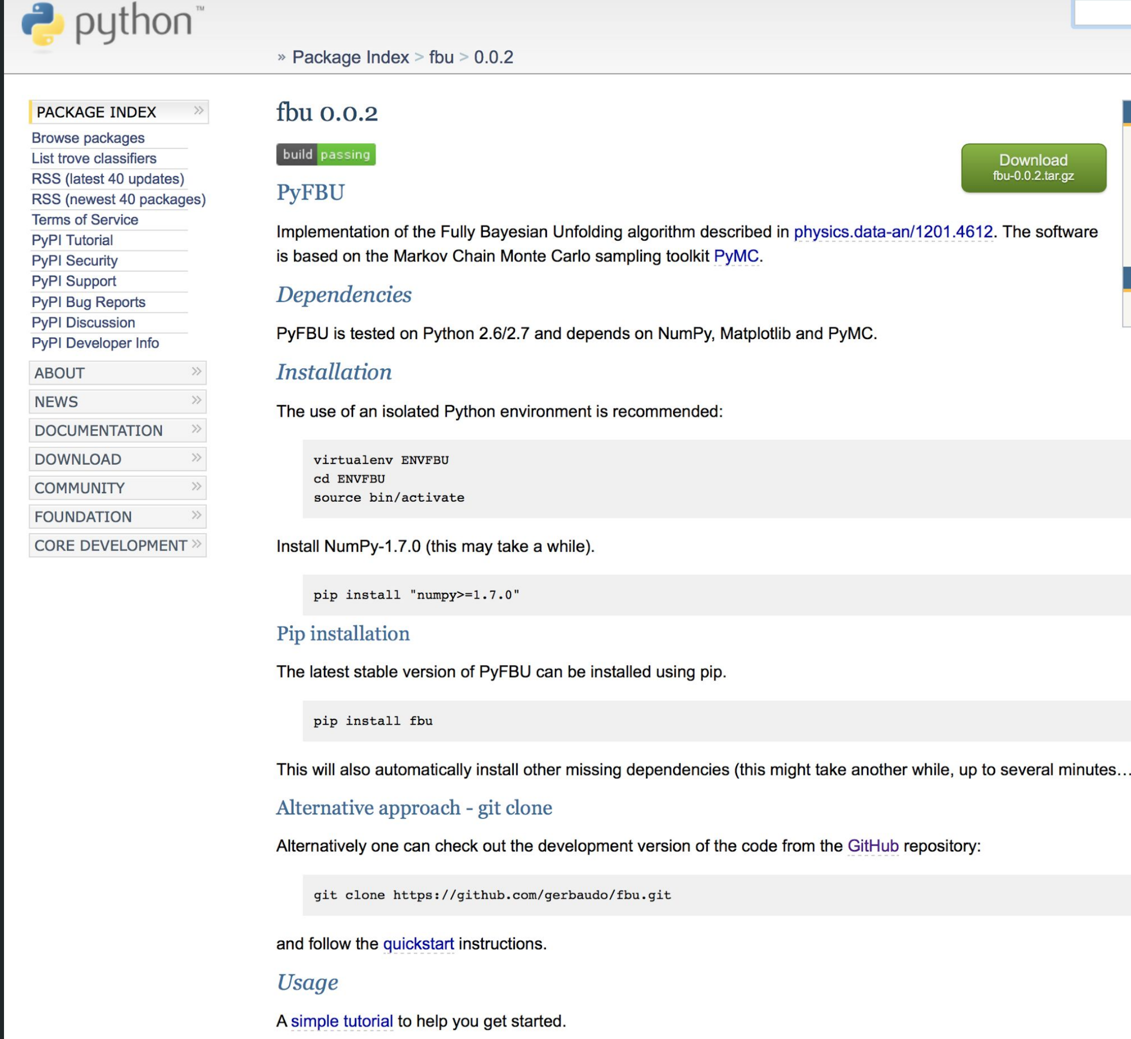
Unfolding relates to the problem of estimating probability distributions in cases where no parametric form is available, and where the data are subject to additional random fluctuations due to limited resolution. The same mathematics can be found under the general heading of inverse problems, and is also called deconvolution or unsmearing.

This project is currently under development. If you would like to be involved please contact vincent.croft at cern.ch or contact me on slack.

Fully Bayesian Unfolding

- Available on pypy
- Depends on:
  - Numpy
  - Matplotlib
  - PyMC
- Implements new algorithm:
  - https://arxiv.org/abs/1201.4612
- Already used for several ATLAS top quark measurements.

https://pypi.python.org/pypi/fbu/0.0.2



fbu 0.0.2

PyFBU

Implementation of the Fully Bayesian Unfolding algorithm described in physics.data-an/1201.4612. The software is based on the Markov Chain Monte Carlo sampling toolkit PyMC.

Dependencies

PyFBU is tested on Python 2.6/2.7 and depends on NumPy, Matplotlib and PyMC.

Installation

The use of an isolated Python environment is recommended:

```
virtualenv ENVFBU
cd ENVFBU
source bin/activate
```

Install NumPy-1.7.0 (this may take a while).

```
pip install "numpy>=1.7.0"
```

Pip installation

The latest stable version of PyFBU can be installed using pip.

```
pip install fbu
```

This will also automatically install other missing dependencies (this might take another while, up to several minutes…

Alternative approach - git clone

Alternatively one can check out the development version of the code from the GitHub repository:

```
git clone https://github.com/gerbaudo/fbu.git
```

and follow the quickstart instructions.

Usage

A simple tutorial to help you get started.

Inverse Problem

$$x = A^{-1}b$$

- Infer information about x from measurement A using response b

Or solve $$\beta = (X^TX)^{-1}(X^TY)$$

- Most problems are large ill-conditioned linear systems
- Lots of solutions exist

- Edwin

  https://pypi.python.org/pypi/edwin/0.1.0

  - Package for bayesian inversion for numerical computation of the inverse problem
- pinvprob

  https://github.com/HajimeKawahara/pinvprob

  - Python codes for the linear inverse problem including the generalized inverse matrix trucated SVD Tikonov regularization, L-curve criterion
- Inverseproblem

  https://pypi.python.org/pypi/InverseProblem/1.0

  - Iterative approach to using Tikhonov regularisation for inverting a matrix

Summary

- Several codes exist for unfolding outside of ROOT for both HEP and non-HEP uses.
    - Up to date?
    - Variety
    - Verified correct
- Specific mathematical considerations must be taken into account when using each case.
    - Can we build in checks to control these cases
- Histograms -
    - Commonplace in HEP
    - Allow for use of 2D histogram for response matrix.
    - Introduces new issues.