

Development of a Pyhf - Combine interface

Santiago Duque-Escobar

Instituto de Física
Universidad de Antioquia

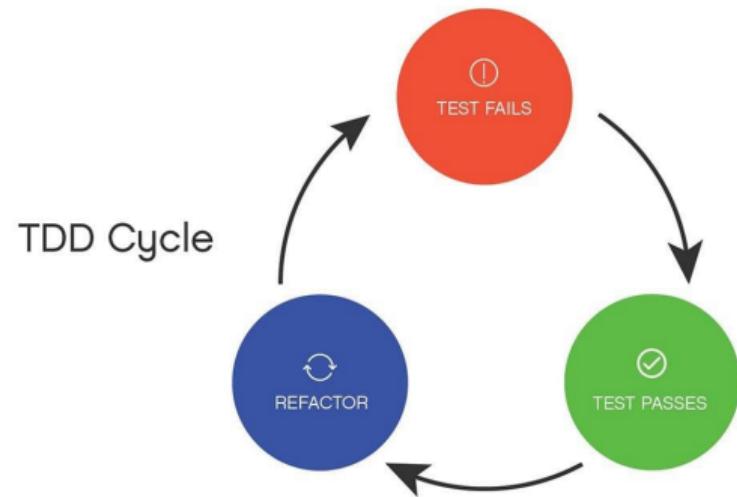
December 01, 2021

Motivation

- There are many tools used to perform the statistical analysis.
- Combine tool is heavily used to produce binned statistical models.
- The proper Combine's configuration is too difficult.
- Pyhf is a pure-python tool.
- The main objective of this work is develop an intermediary software that enable the use of Combine datacards to perform statistical analysis using Pyhf.

Work flow

- Build confidence code from beginning
- Allows an incremental building
- Allows increase the complexity of code without risk of degrade the performance of previous block of code builded



```
imax 1  
jmax 1  
kmax 1
```

```
bin bin1  
observation 7
```

```
bin          vbf      vbf  
process     sig       bck  
process     0         1  
rate        4         3
```

```
lumi_8TeV    lnN     1.04   —
```

```
{  
  "channels": [],  
  "observations": [],  
  "measurements": [],  
  "version": "1.0.0"  
}
```

```
imax 1  
jmax 1  
kmax 1
```

```
bin bin1  
observation 7
```

```
bin          vbf      vbf  
process     sig       bck  
process     0         1  
rate        4         3
```

```
lumi_8TeV    lnN    1.04   -
```



```
"observations": [  
  {  
    "name": "singlechannel",  
    "data": [7.0]  
  }  
,
```

imax	1		
jmax	1		
kmax	1		
<hr/>			
bin	bin1		
observation	7		
<hr/>			
bin	vbf	vbf	
process	sig	bck	
process	0	1	
rate	4	3	
<hr/>			
lumi_8TeV	lnN	1.04	-

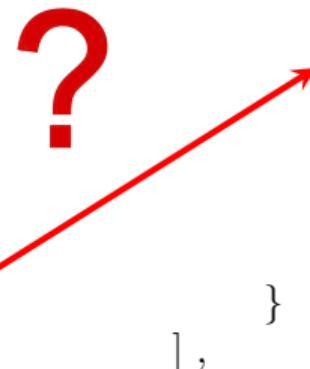
```
"channels": [
    {
        "name": "singlechannel",
        "samples": [
            {
                "name": "signal",
                "data": [4.0],
                "modifiers": []
            },
            {
                "name": "background",
                "data": [3.0],
                "modifiers": []
            }
        ]
    },
]
```

imax 1
jmax 1
kmax 1

bin bin1
observation 7

bin	vbf	vbf
process	sig	bck
process	0	1
rate	4	3

lumi_8TeV lnN 1.04 —



```
    "channels": [  
        { "name": "singlechannel",  
          "samples": [  
              { "name": "signal",  
                "data": [4.0],  
                "modifiers": [] }  
            ],  
            { "name": "background",  
              "data": [3.0],  
              "modifiers": [] }  
        ] }]
```

Description of systematical uncertainties for Combine:

- **lnN**: Log-normal
- **gmN**: Gamma
- **lnU**: Log-uniform

Channel modifiers in Pyhf:

- **histosys**: Correlated Shape
- **lumi**: Luminosity
- **normfactor**: Unconstrained Normalisation
- **normsys**: Normalisation Uncertainty
- **shapefactor**: Data-driven Shape
- **shapesys**: Uncorrelated Shape
- **statterror**: MC Statistical Uncertainty

- HistFactory: $L(\mu) = \text{Pois}(n|\mu S + B) \left[\prod_{e=0}^n \frac{\mu S f_S(x_e) + B f_B(x_e)}{\mu S + B} \right]$
- Combine: $L(\mu, \delta_S, \delta_B) = \text{Pois}(n_{obs}|\mu S + B) e^{-[\mu n_S(\delta_s) + n_B(\delta_B)]} e^{-\frac{1}{2}(\delta_S - \delta_S^{\ln})^2} e^{-\frac{1}{2}(\delta_B - \delta_B^{\ln})^2}$
- Pyhf: $L(\mu, \theta) = \prod_{j=1}^N \text{Pois}(n_{obs}|\mu s_j + b_j) \prod_{k=1}^M \text{Pois}(m_k|u_k)$

Summary

- This project is in an early stage.
- It's necessary doing more research in the deeps of the code of both tools.
- The main source of problems translating Combine datacards to Pyhf JSON is differences in the way these tools handle systematic uncertainties.

Thanks!