Data analysis in high-energy physics

Search for top—anti-top quark resonances in proton-proton collisions with ATLAS data

CSU NUPAC Tutorials 2019

Gabriel Palacino

Thursday, April 25, 2019
What’s new in AnalysisCode?

- New component to the package: **Plotting**
- Contains text file with information about the MC samples
- There is macro for plotting histograms correctly reweighted to luminosity and cross section
- There are also some macros for giving nice professional style to your plots

<table>
<thead>
<tr>
<th>Name</th>
<th>Last commit</th>
</tr>
</thead>
<tbody>
<tr>
<td>AnalysisCode</td>
<td>Update to AnalysisCode</td>
</tr>
<tr>
<td>Plotting</td>
<td>Update to AnalysisCode</td>
</tr>
<tr>
<td>Root</td>
<td>Update to AnalysisCode</td>
</tr>
<tr>
<td>cmt</td>
<td>Initial commit</td>
</tr>
<tr>
<td>util</td>
<td>Update to Analysis code</td>
</tr>
<tr>
<td>README.md</td>
<td>update to steering macro</td>
</tr>
</tbody>
</table>

- include
  - Updates to Plotting
- HistPlotter.C
  - Updates to Plotting
- samples.txt
  - Update to AnalysisCode
### MC sample information

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Cross section [pb]</th>
<th>Reduction efficiency</th>
<th>Sum of weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYeeM08to15</td>
<td>92.15</td>
<td>0.893071694</td>
<td>4772549.624</td>
</tr>
<tr>
<td>DYeeM10to40</td>
<td>279.19</td>
<td>0.945799529</td>
<td>1389547.42</td>
</tr>
<tr>
<td>DMumuM08to15</td>
<td>92.12</td>
<td>0.968298189</td>
<td>4977032.008</td>
</tr>
<tr>
<td>DMumuM10to40</td>
<td>279.2</td>
<td>0.999654588</td>
<td>11510844.32</td>
</tr>
<tr>
<td>DYTautauM08to15</td>
<td>92.12</td>
<td>1</td>
<td>2489620.309</td>
</tr>
<tr>
<td>DYTautauM10to40</td>
<td>279.11</td>
<td>1</td>
<td>13704752.81</td>
</tr>
<tr>
<td>WW</td>
<td>20.90286</td>
<td>0.388</td>
<td>2489550.66</td>
</tr>
<tr>
<td>WZ</td>
<td>6.9673</td>
<td>0.96668</td>
<td>996210.5</td>
</tr>
<tr>
<td>ZZ</td>
<td>1.5376</td>
<td>0.95088</td>
<td>243674.74</td>
</tr>
<tr>
<td>WenuJetsBVeto</td>
<td>591.624</td>
<td>0.981554126</td>
<td>1719244.25</td>
</tr>
<tr>
<td>WenuNoJetsBVeto</td>
<td>11324.5</td>
<td>0.946929361</td>
<td>13179098.28</td>
</tr>
<tr>
<td>WenuWithB</td>
<td>154.374</td>
<td>0.129496439</td>
<td>4260058.061</td>
</tr>
<tr>
<td>WmunuJetsBVeto</td>
<td>513.117</td>
<td>0.985793003</td>
<td>1726755.472</td>
</tr>
<tr>
<td>WmunuNoJetsBVeto</td>
<td>11404.8</td>
<td>0.943743915</td>
<td>11600960.24</td>
</tr>
<tr>
<td>WmunuWithB</td>
<td>154.429</td>
<td>0.133223309</td>
<td>4010756.4</td>
</tr>
<tr>
<td>WtauuJetsBVeto</td>
<td>557.095</td>
<td>0.913648001</td>
<td>1776192.932</td>
</tr>
<tr>
<td>WtauuNoJetsBVeto</td>
<td>11359.7</td>
<td>0.926932574</td>
<td>13316520.62</td>
</tr>
<tr>
<td>WtauuWithB</td>
<td>154.374</td>
<td>0.188978183</td>
<td>4343529.568</td>
</tr>
</tbody>
</table>

- **Reduction efficiency**: efficiency of all the kinematic requirements applied at generation
- **Sum of weights**: effective number of events in the simulated sample before generator cuts
HistPlotter

- Allows for plotting any histogram at any selection level
- Draws all the MC backgrounds grouped in six categories
- Draws the MC statistical uncertainty
- Overlays collision data on top of MC background
- Draws a $Z'$ signal mass point of your choice
- Prints a legend
- Everything is fully customizable
HistPlotter: basic configuration

```
std::string histname = "h_wmt";
float lumi = 1000.;
int rebin = 4; //20 //4
bool logy = true;
std::string signalmass = "2000";
```
HistPlotter: basic configuration

You need to specify the luminosity for automatic rescaling, but need to manually adjust it in the legend.
HistPlotter: basic configuration

```
std::string histname = "h_wmt";
float lumi = 1000.;
int rebin = 4; 1/20 //4
bool logy = true;
std::string signalmass = "2000";
```
HistPlotter: basic configuration

Histogram name
Luminosity
Rebinning factor
y-axis log scale

MC statistical uncertainties only visible in log scale!
HistPlotter: basic configuration

```cpp
class HistPlotter {
    std::string histname = "h_wmt";
    float lumi = 1000.;
    int rebin = 4; // 20 // 4
    bool logy = true;
    std::string signalmass = "2000";
};
```

**Histogram name**

Luminosity

Rebinning factor

y-axis log scale

signal mass point

---

HistPlotter::basic configuration

- **Histogram name**
- **Luminosity**
- **Rebinning factor**
- **y-axis log scale**
- **signal mass point**

**Graph Details**

- **x-axis**:
  - $M_{\text{inv}}(l, E_T^{\text{miss}}, 4j)$ [GeV]

- **y-axis**:
  - Events

**Legend**

- Data
- $t\bar{t}$
- Single top
- $Z$
- $W$
- Drell-Yan
- Diboson
- MC stat. uncertainty

**Legend Details**

- **$\sqrt{s} = 8 \text{ TeV, } 1 \text{ fb}^{-1}$**
- **Luminosity**
- **Rebinning factor**
- **y-axis log scale**
- **signal mass point**
HistPlotter: adding a $Z'$ signal mass point

- The signal is often too small to be seen
- We need no multiply it by a factor $> 1$
- Change the legend too!

Note the different factor for linear and log scales!
HistPlotter: plotting a different histogram

- On top of changing the name of the histogram, a few other settings have to be changes
- The rebinning factor has to be changed too
HistPlotter: plotting a different histogram

- Adjust the range of the x-axis for each histogram
What else is new in AnalysisCode?

Improved event selection

- It is explicitly required to pass either the Egamma OR the Muon trigger

- The expression below finishes processing the current event and moves on to the next.

```
return EL::StatusCode::SUCCESS;
```

Whatever is below it isn't executed!!
What else is new in AnalysisCode?

Improved event selection

- passGRL = the event is part of a the Good Run List
- Good Run List: is a lust of data taking runs deemed as good for physics
  - All detector components working properly
  - No data quality defects
  - The data has been collected under "nominal" conditions, i.e., ATLAS isn’t performing any specific test
What else is new in AnalysisCode?

- We can see the contributions to the total background from each SM process
- The effect of each cut on each background can be seen clearly
- Note the discrepancy between data and MC
- It becomes smaller as we apply cuts
  - why?
  - Are we considering all possible backgrounds?
What else is new in AnalysisCode?

New observable: invariant mass of the lepton, MET and 4 jet system

- If we wanted to reject a large amount of background, this would probably be one of the most efficient ways to do it
To run the plotting macro HistPlotter.C you need to copy all the output data over to the Plotting directory

```
cd YOURTOPDIR/run
RunAnalysis OUTPUTDIR
cd ../AnalysisCode/Plotting
cp .././run/OUTPUTDIR_*/hist-*.*root.
```

Modify the histogram name string and all the other configurables, then run

```
root -l -q -b HistPlotter.C
```

A pdf file named HISTOGRAMNAME.pdf or HISTOGRAMNAME_log.pdf should appear now in YOURTOPDIR/AnalysisCode/Plotting.
Let’s talk about uncertainties: for just a bit

• We discussed how we can avoid statistical fluctuations in the MC by simulating more events that what is expected

• The events are later rescaled to the integrated luminosity and cross section of the process

• Statistical uncertainties are not the only uncertainties

• We also have systematic uncertainties
  ‣ Experimental uncertainties
  ‣ Theory uncertainties

• We won’t be including systematic uncertainties in the analysis.

• But we will experiment with the effect of inflating the uncertainties

Smother behavior in MC than in collision data in particular in the tails
Systematic uncertainties: experimental uncertainties

- Come from a variety of sources
- Example: the energy of a jet measured by the detector does not correspond to the actual energy of the jet
- A Jet Energy Scale (JES) calibration is needed: this comes with an associated uncertainty
- Such uncertainty is applied individually to each jet: it is $p_T$ and $\eta$ dependent
Systematic uncertainties: experimental uncertainties

• There are uncertainties related to the identification of leptons
• Also applied to individual electrons and muons
• Simulating the complex processes that occur in a pp collision at the LHC requires a lot of approximations
• Some of the theory parameters are also not precisely known
• Each event generator generates slightly different events: some generators are preferred for some processes
• We can extract a theory uncertainty by comparing event generators
The final project is due Tuesday, May 9 at midnight!

Part 1

- Go to http://opendata.atlas.cern/books/current/openatlasdatatools/_book/analyses.html and select one analysis. (It can’t be Z’ as this is the example we are working on in class!)
  - All the necessary collision datasets and MC samples exist in http://opendata.atlas.cern/extendedanalysis/datasets.php

- Identify the final state of your signal.
  - How many leptons do you expect?
  - How many jets?
  - Do you expect neutrinos or other source of missing transverse energy?

- What other Standard Model processes share the same final state?
  - List the background processes of you analysis

- Which samples from the list of datasets would you use. List them all and explain why.

Create a new private project in your git repo called AnalysisProject.

Answer the above questions in a .txt file.
The final project is due Tuesday, May 9 at midnight!

Part 2

• Implement your own analysis class. Use ZPrimeAnalysis as an example.
• Make histograms of the main kinematic variables for your signal and backgrounds.
• Decide on a list of event selection criteria based on the histograms of the kinematic variables.
• Make stack histograms of your kinematic variables and event selection

Create a new private project in your git repo called AnalysisProject. Implement the points above and push your analysis framework to your repository.
Thanks!