Data analysis in high-energy physics

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

CSU Tutorials 2020

Gabriel Palacino

Thursday, April 28, 2020
What does cross section mean?

- It gives a measure of the probability for a process to occur.
- Given two protons at 8 TeV:
  - What is the pp collision cross section?
  - What is the W boson production cross section?
  - What is the Higgs boson production cross section?
What does cross section mean?

- Each process has a specific cross section which determines how likely they are to occur.
- We need to accumulate more data to be able to observe processes with smaller cross section.
- Most processes that include new particles beyond the standard model have very small cross sections.

![Cross Section Table]

<table>
<thead>
<tr>
<th>Process</th>
<th>Cross Section [pb]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZPrime400 nts</td>
<td>4.2589</td>
</tr>
<tr>
<td>ZPrime500 nts</td>
<td>3.925</td>
</tr>
<tr>
<td>ZPrime750 nts</td>
<td>1.243</td>
</tr>
<tr>
<td>ZPrime1000 nts</td>
<td>0.3937</td>
</tr>
<tr>
<td>ZPrime1250 nts</td>
<td>0.1387</td>
</tr>
<tr>
<td>ZPrime1500 nts</td>
<td>0.0524</td>
</tr>
<tr>
<td>ZPrime1750 nts</td>
<td>0.0211</td>
</tr>
<tr>
<td>ZPrime2000 nts</td>
<td>0.00894</td>
</tr>
<tr>
<td>ZPrime2250 nts</td>
<td>0.00394</td>
</tr>
<tr>
<td>ZPrime2500 nts</td>
<td>0.0018</td>
</tr>
<tr>
<td>ZPrime3000 nts</td>
<td>0.000434</td>
</tr>
</tbody>
</table>

![Cross Section Chart]
How many MC events we need to simulate?

- We have over 1,400,000 events in the ttbar sample in which there is at least one W decaying leptonically
  - Cross section of the process: $\sigma = 137.297$ pb
  - Integrated luminosity: $L = 1$ fb$^{-1}$
  - Number of expected events: $N = \sigma L = 137,297$
- We simulate more events that actually expected in order to avoid statistical fluctuation in the MC sample
We need to rescale our histograms

- It is necessary to rescale the MC histograms to match what we would expect given the integrated luminosity
- Each process is rescaled using its cross section
- We DO NOT re-scale collision data

\[ k = \frac{\sigma L}{N_{\text{MC}}} \]

This rescaling is done to the histograms after the full event selection/analysis.
Event by event scale factors

- There are also scale factors that are applied on an event-by-event basis
- There are related, for instance, to trigger efficiency, lepton reconstruction and identification, pielup, etc.
- These are also only applied to MC since they account for discrepancies between MC and collision data

```cpp
void ZprimeAnalysis :: getEventInfo()
{
    eventInfo.nvertices = pvxp_n;
    eventInfo.passGamma = trigE;
    eventInfo.passMuon = trigM;
    eventInfo.passGRL = passGRL;
    eventInfo.hasGoodVertex = hasGoodVertex;
    eventInfo.evtWeight = mcWeight * scaleFactor_PILEUP * scaleFactor_ZVERTEX;
    eventInfo.evtSF = scaleFactor_ELE * scaleFactorMuon * scaleFactor_TRIGGER;

    return;
}
```
How do we apply the scale factors?

- Scale factors are applied when filling the histograms in ZprimeAnalysis::execute()

```c
float weight = isData ? 1.0 : eventInfo.evtWeight*eventInfo.evtSF;

// WmT
double WmT = (goodLeptons.at(0).tlv + MET.tlv).Mt();

if(!eventInfo.hasGoodVertex && goodLeptons.size() == 1 && goodJets.size() >= 4 && bJets.size() >= 1)
   h_event_selection->Fill(passWmt30, weight);

h_wmt->Fill(WmT*GeV, weight);

h_wmt_met->Fill(WmT*GeV + MET.tlv.Et()*GeV, weight);
```
Which samples are we running over?

<table>
<thead>
<tr>
<th>Sample</th>
<th>Cross section [pb]</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYeeM08to15</td>
<td>92.15</td>
</tr>
<tr>
<td>DYeeM15to40</td>
<td>279.19</td>
</tr>
<tr>
<td>DYMumuM08to15</td>
<td>92.12</td>
</tr>
<tr>
<td>DYMumuM15to40</td>
<td>279.2</td>
</tr>
<tr>
<td>DYtautauM08to15</td>
<td>92.12</td>
</tr>
<tr>
<td>DYtautauM15to40</td>
<td>279.11</td>
</tr>
<tr>
<td>WW</td>
<td>20.98286</td>
</tr>
<tr>
<td>WZ</td>
<td>6.9673</td>
</tr>
<tr>
<td>ZZ</td>
<td>1.5376</td>
</tr>
<tr>
<td>WenuJetsBVeto</td>
<td>591.624</td>
</tr>
<tr>
<td>WenuNoJetsBVeto</td>
<td>11324.5</td>
</tr>
<tr>
<td>WenuWithB</td>
<td>154.374</td>
</tr>
<tr>
<td>WmunuJetsBVeto</td>
<td>513.117</td>
</tr>
<tr>
<td>WmunuNoJetsBVeto</td>
<td>11404.8</td>
</tr>
<tr>
<td>WmunuWithB</td>
<td>154.429</td>
</tr>
<tr>
<td>WtaunuJetsBVeto</td>
<td>557.095</td>
</tr>
<tr>
<td>WtaunuNoJetsBVeto</td>
<td>11359.7</td>
</tr>
<tr>
<td>WtaunuWithB</td>
<td>154.374</td>
</tr>
<tr>
<td>Zee</td>
<td>1241.2072</td>
</tr>
<tr>
<td>Zmumu</td>
<td>1241.2072</td>
</tr>
<tr>
<td>Ztautau</td>
<td>1240.8988</td>
</tr>
<tr>
<td>stop_schan</td>
<td>1.817694</td>
</tr>
<tr>
<td>stop_tchan_antitop</td>
<td>9.95976</td>
</tr>
<tr>
<td>stop_tchan_top</td>
<td>18.39495</td>
</tr>
<tr>
<td>stop_wtchan</td>
<td>22.3014</td>
</tr>
<tr>
<td>ttbar_had</td>
<td>115.518854</td>
</tr>
<tr>
<td>ttbar_lep</td>
<td>137.29749</td>
</tr>
</tbody>
</table>

- The cross sections are in pb
- These samples can be classified in six groups
  - Drell-Yan
  - Diboson
  - W+jets
  - Z
  - Single top
  - ttbar
- Results should be shown grouped in the six categories
Background processes

Drell-Yan production

\[
\begin{align*}
\text{DYeeM08to15} & : 92.15 \\
\text{DYeeM15to40} & : 279.19 \\
\text{DYmumuM08to15} & : 92.12 \\
\text{DYmumuM15to40} & : 279.2 \\
\text{DYtautauM08to15} & : 92.12 \\
\text{DYtautauM15to40} & : 279.11 \\
\text{WW} & : 20.90286 \\
\text{WZ} & : 6.9673 \\
\text{ZZ} & : 1.5376 \\
\text{WenuJetsBVeto} & : 591.624 \\
\text{WenuNoJetsBVeto} & : 11324.5 \\
\text{WenuWithB} & : 154.374 \\
\text{WmumuJetsBVeto} & : 513.117 \\
\text{WmumuNoJetsBVeto} & : 11404.8 \\
\text{WmumuWithB} & : 154.429 \\
\text{WtaunuJetsBVeto} & : 557.095 \\
\text{WtaunuNoJetsBVeto} & : 11359.7 \\
\text{WtaunuWithB} & : 154.374 \\
\text{Zee} & : 1241.2072 \\
\text{Zmumu} & : 1241.2072 \\
\text{Ztautau} & : 1240.8988 \\
\text{stop_schan} & : 1.817694 \\
\text{stop_tchan_antitop} & : 9.95976 \\
\text{stop_tchan_top} & : 18.39495 \\
\text{stop_wtchan} & : 22.3014 \\
\text{ttbar_had} & : 115.518854 \\
\text{ttbar_lep} & : 137.29749
\end{align*}
\]
## Background processes

<table>
<thead>
<tr>
<th>Process</th>
<th>Cross Section (fb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYeeM08to15</td>
<td>92.15</td>
</tr>
<tr>
<td>DYeeM15to40</td>
<td>279.19</td>
</tr>
<tr>
<td>DymumuM08to15</td>
<td>92.12</td>
</tr>
<tr>
<td>DymumuM15to40</td>
<td>279.2</td>
</tr>
<tr>
<td>DYtautauM08to15</td>
<td>92.12</td>
</tr>
<tr>
<td>DYtautauM15to40</td>
<td>279.11</td>
</tr>
</tbody>
</table>

### Drell-Yan production

\[
\nu + \nu + l^+ + W^+ + b + q + q' + Z
\]

### Diboson

\[
\begin{align*}
q & \rightarrow W^+ l^- \\
\bar{q} & \rightarrow W^- d \\
\bar{u} & \rightarrow Z q \\
\bar{d} & \rightarrow Z q
\end{align*}
\]
Background processes

Drell-Yan production

Diboson

W+jets
## Background processes

### Drell-Yan production

<table>
<thead>
<tr>
<th>Process</th>
<th>Cross Section (fb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYeeM08to15</td>
<td>92.15</td>
</tr>
<tr>
<td>DYeeM15to40</td>
<td>279.19</td>
</tr>
<tr>
<td>DYMumuM08to15</td>
<td>92.12</td>
</tr>
<tr>
<td>DYMumuM15to40</td>
<td>279.2</td>
</tr>
<tr>
<td>DYTautauM08to15</td>
<td>92.12</td>
</tr>
<tr>
<td>DYTautauM15to40</td>
<td>279.11</td>
</tr>
</tbody>
</table>

### Diboson

- WW: 20.90286 fb
- WZ: 6.9673 fb
- ZZ: 1.5376 fb

### W+jets

- WenuJetsBVeto: 591.624 fb
- WenuNoJetsBVeto: 11324.5 fb
- WenuWithB: 154.374 fb
- WmunuJetsBVeto: 513.117 fb
- WmunuNoJetsBVeto: 11404.8 fb
- WmunuWithB: 154.429 fb
- WtaunuJetsBVeto: 557.995 fb
- WtaunuNoJetsBVeto: 11359.7 fb
- WtaunuWithB: 154.374 fb

### Z+jets

- Zee: 1241.2072 fb
- Zmumu: 1241.2072 fb
- Ztautau: 1240.8986 fb

### Diagrams

- Drell-Yan production
- Diboson
- W+jets
- Z+jets
Background processes

<table>
<thead>
<tr>
<th>Process</th>
<th>Cross Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYeeM08to15</td>
<td>92.15</td>
</tr>
<tr>
<td>DYeeM15to40</td>
<td>279.19</td>
</tr>
<tr>
<td>DYmumuM08to15</td>
<td>92.12</td>
</tr>
<tr>
<td>DYmumuM15to40</td>
<td>279.2</td>
</tr>
<tr>
<td>DYtautauM08to15</td>
<td>92.12</td>
</tr>
<tr>
<td>DYtautauM15to40</td>
<td>279.11</td>
</tr>
</tbody>
</table>

Drell-Yan production

Diboson

W+jets

Z

Single top
Drell-Yan production

Diboson

W+jets

Z

Single top

ttbar
Background processes

Drell-Yan production

W+jets

- W+jets is by far the most common process out of the six described above
- But it is ttbar the most important background as it has exactly the same final state as the signal
  - Main differences come from the kinematics of the top pair from the Z' decay
  - Since the Z' is very massive, the top quarks have a high transverse momentum
Normalizing to integrated luminosity

### Drell-Yan production

<table>
<thead>
<tr>
<th>Process</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WW</td>
<td>20.90286</td>
</tr>
<tr>
<td>WZ</td>
<td>6.9673</td>
</tr>
<tr>
<td>ZZ</td>
<td>1.5376</td>
</tr>
</tbody>
</table>

### Diboson

- ATLAS data (DataEgamma + DataMuon) corresponds to 1 fb\(^{-1}\) (1000 pb\(^{-1}\))
- You should rescale each the histograms corresponding to each sample as

\[
h \rightarrow \text{Scale} (\text{crossSection} \times 1000 / h \rightarrow \text{Integral}(0, h \rightarrow \text{GetNbinsX}()+1))
\]

### W+jets

### Z

### Single top

### ttbar

- Always take the integral including the underflow and overflow bins (range [0,\(N_{\text{bins}}+1\)])
Diboson example

- Read output ROOT files from AnalysisCode
- Read histograms
- Cross section and luminosity information
- Rescale histograms to cross section and luminosity
- Clone one of the diboson histograms
- Add the other diboson histograms
- Do the same for all the other five categories
• Create a THStack

• Change color of each background group (see previous slide)

• Add each background group to the THStack

• Clone one of the collision data histograms and add it to the other (here muon is added to a clone of electron)

• Create a TCanvas

• Draw the backgrounds and overlay the collision data
The final project is due Thursday, May 7 at midnight!

Send me an email with your chosen analysis by Monday!!

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! And try not to choose the tt either)
  ‣ 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.
Thanks!