From Raw Data to Physics Results

Jamie Boyd
CERN
CERN Summer Student Lectures
2014
Outline

- Summary
  - Brief overview of the full lecture course
  - A simple example
    - Measuring the $Z^0$ cross-section
  - Reconstruction & Simulation
    - Track reconstruction
    - Calorimeter reconstruction
    - Physics object reconstruction
    - Simulation
  - Physics Analysis
    - Data Quality
    - $Z' \rightarrow ll$
    - $H \rightarrow yy$
    - $H \rightarrow ZZ \rightarrow 4l$
  - Computing infrastructure
  - The End!

Disclaimer: Much of the content based on previous years lectures
Thanks to G. Dissertori
Section 1: Summary
Data Analysis Chain

- Have to collect data from many channels on many sub-detectors (millions)
- Decide to read out everything or throw event away (Trigger)
- Build the event (put info together)
- Store the data

Analyze them
  - reconstruction, user analysis algorithms, data volume reduction

- do the same with a simulation
  - correct data for detector effects

Compare data and theory

This lecture course!!
DAQ chain
(see lectures by W. Vandelli & B. Dahmes)

Detector Front-End

Front-End Electronics 1

Front-End Electronics 2

Rate: MHz

Trigger

Main event Builder, High Level Trigger online reconstr.

Upper Level Readout 1

Upper Level Readout 2

Yes/no

Rate: ~100 Hz

Data storage
Offline Analysis Chain

- **Data storage**
  - Data size: PByte
- **Offline reconstruction, calibration, alignment**
- **Further data storage**
  - Data size: TByte
- **User Analysis Program**
  - Data size: kByte
**Data reduction/abstraction**

- **Digitization/Reconstruction**
  - Analog signals
  - Store the info for every event and every track

- **Track finding + Track fit**
  - Track momentum: $p_x, p_y, p_z$

- **Magnetic field B:**
  - Helix $(R, d_0, z_0)$

- **Events:**
  - Event 1
    - Track 1
    - Track 2
  - Event 2

- **Hits:**
  - $(x_1, y_1, z_1, t_1)$
  - $(x_2, y_2, z_2, t_2)$
  - ...
High Level Data Storage

- Data are stored sequentially in files...

Event 1

\(N_{ch}\) (charged tracks) : 2

\(P_{cha}\)
(Momentum of each track):

\[
\{\begin{array}{ccc}
-7.65698 & 42.9725 & 14.3404 \\
7.54101 & -42.1729 & -14.0108
\end{array}\}
\]

\(Q_{cha}\)
(Charge of each track):
\{-1,1\}

Event 2

\(N_{ch}\) (charged tracks) : 3

\(P_{cha}\)
(Momentum of each track):

\[
\{\begin{array}{ccc}
-12.9305 & 12.2713 & 40.5615 \\
12.2469 & -11.606 & -38.7182 \\
0.143435 & -0.143435 & -0.497444
\end{array}\}
\]

\(Q_{cha}\)
(Charge of each track):
\{-1,1,-1\}

File A
Simulation

Exactly the same steps as for the data

Simulation of many (billions) of events
- simulate physics process
e.g. \( p p \rightarrow Z \)
or \( p p \rightarrow H \)
- plus the detector response to the produced particles
- understand detector response and analysis parameters (lost particles, resolution, efficiencies, backgrounds)
- and compare to real data
- Note: simulations present from beginning to end of experiment, needed to make design choices
Our Task

We use experiments to inquire about what “reality” (nature) does

We intend to fill this gap

The goal is to understand in the most general; that’s usually also the simplest.

- A. Eddington
Theory...

eg.

the Standard Model

has parameters

coupling constants

masses

predicts:
cross sections, branching ratios, lifetimes, ...
Experiment...

eg. 1/30\(^{th}\) of an event in the BaBar detector

- get about 100 evts/sec

"Address" :
  - which detector element took the reading

"Value(s)" :
  - what the electronics wrote out
Making the connection

Reality

Raw Data

The imperfect measurement of a (set of) interactions in the detector

Reconstructed Events

A unique happening:

eg. Run 23458, event 1345 which contains a $Z \rightarrow \mu^+ \mu^-$ decay

Analysis: We “confront theory with experiment” by comparing the measured quantity (observable) with the prediction.

Observables

cross sections (probabilities for interactions), branching ratios (BR), ratios of BRs, specific lifetimes, ...

Theory

A small number of general equations, with some parameters (poorly or not known at all)

$S = i \int d^4x \mathcal{L}(x)$
Measuring $Z^0$ cross-section at LHC

- $Z^0$ boson decays to lepton or quark pairs
  - We can reconstruct it in the $e^+e^-$ or $\mu^+\mu^-$ decay modes
- Discovery and study of the $Z^0$ boson was a critical part of understanding the electroweak force
- Measuring the $Z^0$ cross-section at the LHC important test of theory
  - Does the measurement agree with the theoretical prediction at LHC collision energy?
- Now we use the $Z^0$ as a tool for studying electron and muon reconstruction and deriving calibrations (have now recorded millions of Z decays)

$Z^0$ cross-section is related to the probability that we will produce a $Z^0$ at the LHC
CMS Experiment at LHC, CERN
Run 136087 Event 39967482
Lumi section: 314
Mon May 24 2010, 15:31:58 CEST

Z->μμ event in CMS

Muon $p_T = 27.3, 20.5$ GeV/c
Inv. mass $= 85.5$ GeV/c$^2$
Reconstructing $Z^0$’s

How do we know if it’s a $Z^0$:

Identify $Z$ decays using the invariant mass of the 2 leptons

$$M^2 = (L_1 + L_2)^2$$  where $L_i = (E_i, \mathbf{p}_i) = 4$-vector for lepton $i$

Under assumption that lepton is massless compared to mass of $Z^0$

$$=> M^2 = 2E_1E_2(1-\cos\theta_{12})$$  where $\theta_{12} = \text{angle between the leptons}$

So need to reconstruct the electron and muon energy and direction. Then can calculate the mass.

Select $Z^0$ events with ‘analysis cuts’:
- Events with 2 high momentum electrons or muons
- Require the electrons or muons are of opposite charge
- With di-lepton mass close to the $Z^0$ mass (e.g. $70 < m_{l^+l^-} < 110$ GeV)

Very little background in the $Z^0$ mass region
UA1: observation of $Z \rightarrow e^+ e^-$
(May 1983)

Two energy clusters ($p_T > 25$ GeV)
in electromagnetic calorimeters;
energy leakage in hadronic calorimeters
consistent with electrons

Isolated track with $p_T > 7$ GeV
pointing to at least one cluster

Isolated track with $p_T > 7$ GeV
pointing to both clusters

Slide taken from: “The achievements of the CERN proton – antiproton collider”, Luigi DiLella, 2004
Measuring the $Z^0$ cross-section

**Theoretically:**
Cross-section calculated for:
- Specific production mechanism ($pp$, $p\bar{p}$, $e^+e^-$)
- Centre-of-Mass of the collisions (7TeV at LHC)

**Experimentally:**
$\sigma(pp->Z) = (N_{OBS} - N_{BKG})/ L \varepsilon$

Where:
- $N_{OBS} =$ Number of observed events passing the selection
- $N_{BKG} =$ Estimate of the number of background events
- $L =$ Luminosity of the data samples (amount of data)
- $\varepsilon =$ Efficiency of the selection on $Z^0$ events
  (how often would we select a true $Z^0$ event with our selection?)

Can use simulated data to evaluate $\varepsilon$ and $N_{BKG}$
Measuring the $Z^0$ cross-section

$\sigma(pp\rightarrow Z) = (N_{\text{OBS}} - N_{\text{BKG}})/L \varepsilon$

Looks like simple counting experiment. But need to also calculate uncertainty on the cross-section – measurement without an uncertainty is useless.

Two components to the uncertainty:
Statistical: $\sim \sqrt{N_{\text{OBS}}}$
Systematic:
- How well do we know the background?
- How well do we know the efficiency?
- How well do we know the luminosity?

Most of the work in the physics analysis is trying to understand the systematic uncertainties related to the above questions.

Measurements of the Z cross-section were one of the first physics measurements from ATLAS and CMS.
Analysis flow in Z cross-section measurement

1. Select events with 2 oppositely charged El/Mu
2. Calculate mass
3. Select events with mass close to Z mass

Reconstruct El and Mu candidates

Detector & Trigger → Reconstruction → Physics Analysis → Compare theory and experiment

Simulated data → Reconstruction
Analysis flow in Z cross-section measurement

1. Detector & Trigger
2. Reconstruction
3. Physics Analysis
4. Simulated data
5. Reconstruction
6. Physics Analysis

Centrally produced by the collaboration
Carried out by individuals or small analysis teams

Compare theory and experiment
• Path from Raw data to physics results contains many steps
  • Online path (Trigger and DAQ)
  • Offline path
    • Reconstruction
    • Physics Analysis
  • Use simulation in order to compare data with theoretical predictions
• Above points illustrated with the example of the $Z^0$ cross-section measurement at the LHC
• More details on Monday
Section 3: Reconstruction
Reconstruction

• Detector reconstruction
  – Tracking
    • finding path of charged particles through the detector
  – Calorimeter reconstruction
    • finding energy deposits in calorimeters from charged and neutral particles

• Combined reconstruction
  – Electron/Photon identification
  – Muon identification
  – Jet finding

• Calibrations and alignments applied at nearly every step
Important figures of merit for reconstructed objects

• **Efficiency**
  
  – how often do we reconstruct the object – e.g. tracking efficiency

\[
\text{Efficiency} = \frac{\text{Number of Reconstructed Tracks}}{\text{Number of True Tracks}}
\]
Important figures of merit for reconstructed objects

• **Efficiency**
  – how often do we reconstruct the object – e.g. tracking efficiency

• **Resolution**
  – how accurately do we reconstruct it – e.g. energy resolution

Energy resolution = \((\text{Measured\_Energy} - \text{True\_Energy})/\text{True\_Energy}\)
Important figures of merit for reconstructed objects

• **Efficiency**
  – how often do we reconstruct the object – e.g. tracking efficiency

• **Resolution**
  – how accurately do we reconstruct a quantity – e.g. energy resolution

• **Fake rate**
  – how often we reconstruct a different object as the object we are interested in – e.g. a jet faking a electron

Fake rate = \( \frac{\text{Number of jets reconstructed as an electron}}{\text{Number of jets}} \)
Important figures of merit for reconstructed objects

- **Efficiency**
  - how often do we reconstruct the object – e.g. tracking efficiency

- **Resolution**
  - how accurately do we reconstruct a quantity – e.g. energy resolution

- **Fake rate**
  - how often we reconstruct a different object as the object we are interested in – e.g. a jet faking a electron

These quantities depend on the detector, but also on the reconstruction and calibrations and alignment!
Important figures of merit for reconstructed objects

• **Efficiency**
  – how often do we reconstruct the object – e.g. tracking efficiency

• **Resolution**
  – how accurately do we reconstruct a quantity – e.g. energy resolution

• **Fake rate**
  – how often we reconstruct a different object as the object we are interested in – e.g. a jet faking a electron

For physics analysis it is important
i) to have high efficiency, good resolution, and low fake rates
ii) to be able to measure the efficiencies, resolutions and fake rates and their uncertainties (not easy)
Reconstruction Goals

• High efficiency
• Good resolution
• Low fake rate
• Robust against detector problems
  – Noise
  – Dead regions of the detector
• Be able to run within the computing resources limitations
  – CPU time per event
  – Memory use