Hands on HEP data analysis

Measurement of the W charge asymmetry at LHC using early CMS data at 7 TeV

IDPASC school, LPNHE Paris

February 2015

Julie Malclès, José Ocariz

This tutorial is based on public ntuples from CMS data kindly made available by **Christian Sander and Alexander Schmidt**. Their tutorial on top physics is available here: http://opendata.cern.ch/record/50

Introduction

In this tutorial, we propose to perform a measurement of the W charge asymmetry in the muon channel W$\rightarrow$ μ ν_μ, using 50 pb^{-1} of CMS data recorded in 2010 at a center of mass energy of 7 TeV. The W charge asymmetry was measured both by ATLAS and CMS collaborations [1, 2, 3]. We propose to perform this measurement as a function of the muon pseudorapidity η, and for two different cuts on the muon p_{T}, like it was done in the CMS publication [2]. The dataset used in this tutorial includes the dataset used in this publication. The CMS result is depicted in figure 1.

W charge asymmetry: definition and physics case

In pp collisions, W bosons are produced primarily via the processes u$\bar{\mathrm{d}}$ $\rightarrow$ W$^+$ and d$\bar{\mathrm{u}}$ $\rightarrow$ W$^-$. The first quark is a valence quark from one of the protons, and the second one is a sea antiquark from the other proton. Due to the presence of two valence u quarks in the proton, there is an overall excess of W$^+$ over W$^-$ bosons.

Measurement of this production asymmetry between W$^+$ and W$^-$ bosons as a function of boson rapidity can provide new insights on the u/d ratio and the sea antiquark densities [4, 5] in the ranges of the Björken parameter x [6] probed in pp collisions at 7 TeV. However, due to the presence of neutrinos in leptonic W decays the boson rapidity is not directly accessible. The experimentally accessible quantity is the lepton charge asymmetry, defined to be

$$\mathcal{A}(\eta) = \frac{\mathrm{d}\sigma/\mathrm{d}\eta(\mathrm{W}^+ \rightarrow \ell^+\nu) - \mathrm{d}\sigma/\mathrm{d}\eta(\mathrm{W}^- \rightarrow \ell^-\bar{\nu})}{\mathrm{d}\sigma/\mathrm{d}\eta(\mathrm{W}^+ \rightarrow \ell^+\nu) + \mathrm{d}\sigma/\mathrm{d}\eta(\mathrm{W}^- \rightarrow \ell^-\bar{\nu})},$$

where ℓ is the daughter charged lepton, η is the charged lepton pseudorapidity in the CMS lab frame ($\eta = -\ln[\tan(\frac{\theta}{2})]$ where θ is the polar angle), and $\mathrm{d}\sigma/\mathrm{d}\eta$ is the differential cross section for charged leptons from W boson decays. The lepton charge asymmetry can be used to test SM

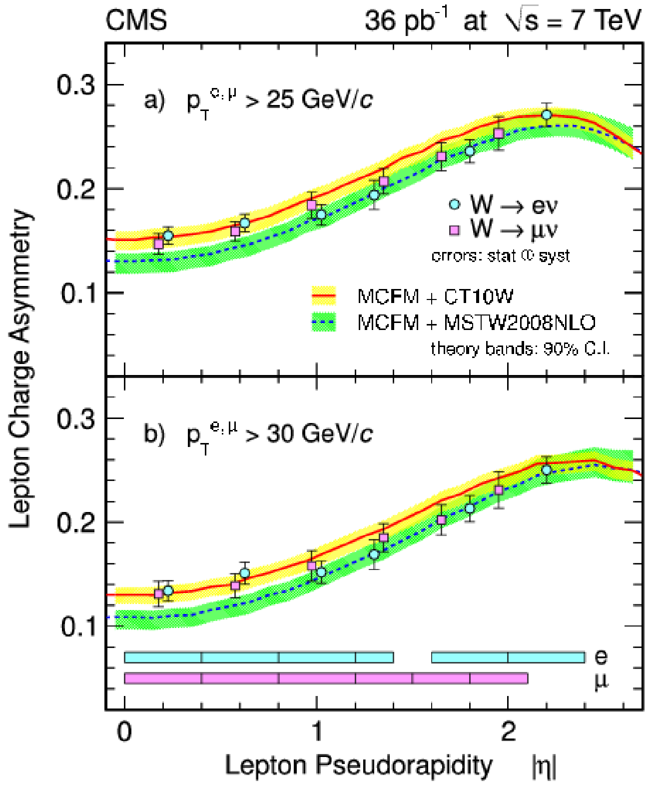

Figure 1: W charge asymmetry measured by CMS in 2011 [2].

predictions with high precision. Due to the structure of the W boson couplings to fermions, theoretical predictions of the charge asymmetry depend on the transverse momentum (p_T) threshold applied on the daughter leptons. For this reason, we measure $\mathcal{A}(\eta)$ for two different charged lepton p_{T} thresholds, 25 GeV/c and 30 GeV/c.

Data and code provided for this tutorial

For the documentation on the data files and on the basic structure of the code, first read: DocumentationDataAndCode.pdf. The program MyAnalysis.{C,h} will be used in section 1.1. On top of the code described in this document, we provide for this tutorial few more pieces of code to ease your work: CutAndCount.{C,h} and METSmearing.{C,h}, which you will use respectively in sections 1.2 and 2.1. The executable using these different pieces of code is example.C, where you have several functions, and the main function. Finally, a macro example for the final fit is given: simFitMET.cc, used in section 2.2. Disclaimer: there are holes in this code that you are supposed to fill. If you don't, some parts of the code will crash. You can find the places to complete searching the keyword "TODO" in the code.

1 Warmup

The signature of a W in the muon channel is a muon plus large missing transverse energy from the undetected neutrino. The signature of a Z boson decaying to muons is a pair of muons of opposite charges and small missing transverse energy as there is no neutrino.

1.1 Selection definition

1. Perform basic plots of several variables of interest (namely the number of muons, the dimuon invariant mass, the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$, the lepton transverse momentum p_{T}) with data and MC, including the backgrounds and the signal (See example code: MyAnalysis.{C,h} for the definition and filling of histograms and function BasicStackPlots() in the main program example.C for plotting these histograms.)

2. Using these distributions, define selection criteria for the signal (cuts to enrich the data in signal) and also selection criteria to have an almost pure QCD sample or Z sample from the data.

1.2 Very basic cut and count analysis

1. Compute the raw W charge asymmetry by counting plus and minus charges in a sample where the background contribution is small (with a cut on $E_{\mathrm{T}}^{\mathrm{miss}}$). (The skeleton code to do that is given in CutAndCount.{C,h} and the functions to use it in example.C are getAsym() and DrawCutAndCountAsym(). The binning in eta is already defined there. Search the word "TODO" in the code for further instructions.)

2. Compare to CMS result (you have, in example.C, functions to plot the 2011 CMS result). Do the same computation on signal MC for different cuts on the $E_{\mathrm{T}}^{\mathrm{miss}}$ and compare it to the results without cut on the $E_{\mathrm{T}}^{\mathrm{miss}}$. You can for example write new functions getAsymMC() and DrawCutAndCountMC() using the class CutAndCount in a different way. What are the two main sources of bias in this very simple analysis?

Those two limitations would imply large systematic uncertainties if this very simple method was used to compute the asymmetry. We propose a much better method in the second part.

2 W asymmetry determination

In this section, we propose to determine the W charge asymmetry for each η bin with a maximum likelihood fit on data using the $E_{\mathrm{T}}^{\mathrm{miss}}$ variable with three species: signal, drell-yan, and QCD background, as it was done in [2] for the electron channel. As in this paper, we propose to fix in the fit the ratio between the total number of Z events and the total number of W events. The fit will be done using the RooFit framework. The probability density function will be determined from corrected MC for signal and drell-yan, and from data for QCD.

2.1 MET calibration

If we want to perform a ML fit along $E_{\mathrm{T}}^{\mathrm{miss}}$ to determine the yields for the different species, we need to know the probability density functions of the $E_{\mathrm{T}}^{\mathrm{miss}}$ variable for those 3 species. As you probably notice during the warmup, the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions from MC do not match the distributions in data, so they cannot be used as they are. We propose here to first determine a smearing to be applied to the MC for it to match the data using a Z sample (double muon sample). Then the smeared MC will be used for the Z and W PDFs definition.

1. Define the smearing to be applied to MC to match the data on a Z sample, by minimizing the χ^2 between data histogram and smeared MC histogram. The class METSmearing.{C,h} provides $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions before and after smearing, for a given sample, a given smearing value, and a given number of isolated muons. Do you understand from this code how the smearing is done? For the best smearing determination, look at the example function fitMetSmearing() in example.C and complete it.

2. Apply this smearing to single muon MC events from Wjets and drell-yan (Z) samples to get probability density functions (PDFs) for signal and drell-yan (with 1 reconstructed muon) species. Save your PDFs in a root file as root histograms (TH1F). Look at the example function applyMetSmearing() in example.C and complete it.

3. Plot the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in data for events with 0 muon. Note that the QCD MC sample does not have enough statistics to be trusted. What are those events mainly? Define the probability density function for QCD using data with 0 muon and removing the background from MC (at first order, as the backgroud is very small, you can use MC directly without smearing). You can also do this in the same function applyMetSmearing() taking advantage of the histograms provided by METSmearing class. Save your PDFs in a root file as root histograms (TH1F). You can draw them with the function DrawPDF() provided in example.C.

2.2 Maximum likelihood (ML) fit

1. Preliminary work: compute the ratio between the total expected Z and W yields as a function of η using MC. MC can be trusted for this ratio and it will be fixed in the fit. You can write in example.C a function getZfrac() computing this fraction. (Note that if you modify CutAndCount to make it work with Z, you can compute this very easily. A boolean doW is already included in the class for this purpose.)

2. If we assume Z and QCD backgrounds have no charge asymmetry (which can be checked eventually in the MC), how many parameters do you have in your fit?

3. Perform a simultaneous fit on $\mathrm{W}^+$and W^-samples to get the W asymmetry using these PDFs. You can fix the ratio between the total Z and W yields using MC and assume no asymmetry for Z and QCD backgrounds. You can use directly histograms for PDFs using RooHistPdf in RooFit. For this last step, we will use a standalone macro and run it directly in root. You can have a look at the roofit tutorial rf501_simultaneouspdf.C included here to know how to perform a simultaneous fit. To ease your work, look at the code skeleton simFitMET.C, where the dataset to fit and the PDFs are already defined. To run it, open root and do: ".L simFitMET.cc+;" and then run the function you want. In this file, you have to write the function simFitMET(ptCut,etabin), performing the fit for a given p_{T} cut and a given eta bin.

4. Compare your results to CMS results and to the theory prediction. The function doFits(ptCut) will plot your results automatically and compare them with CMS results.

3 Extension for fast students: systematic uncertainties

3.1 Detection efficiency differences for positive and negative charges

Compute the cut and count asymmetry on Z data to estimate this detection systematic uncertainty. What is your conclusion? What systemetic uncertainty would you quote?

3.2 Other source of systematic uncertainties

What are the other possible sources of systematic uncertainties? How could you estimate them?

References

[1] ATLAS Collaboration, "Measurement of the Muon Charge Asymmetry from W Bosons Produced in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector," Phys. Lett. B **701** (2011) 31 [arXiv:1103.2929 [hep-ex]].

[2] CMS Collaboration, "Measurement of the lepton charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV," JHEP **1104** (2011) 050 [arXiv:1103.3470 [hep-ex]].

[3] CMS Collaboration, "Measurement of the muon charge asymmetry in inclusive $pp \to W + X$ production at $\sqrt{s} = 7$ TeV and an improved determination of light parton distribution functions," Phys. Rev. D **90** (2014) 3, 032004 [arXiv:1312.6283 [hep-ex]].

[4] A.D. Martin *et al.*, Eur. Phys. J. **C64**, 653 (2009); A.D. Martin *et al.*, Eur. Phys. J. **C63**, 189 (2009).

[5] H.L. Lai *et al.*, Phys. Rev. **D82**, 074024 (2010).

[6] J.D. Björken, Phys. Rev. **179**, 1547 (1969); J.D. Björken and E.A. Paschos, Phys. Rev. **D1**, 3151 (1970).