Institut für Experimentelle Kernphysik (EKP)
Prof. Dr. Th. Müller
Dr. M. Mozer, Dr. Th. Chwalek
N. Faltermann, R. Friese, G. Sieber, M. Gelb

3. Dezember 2015

Computerpraktikum zur Vorlesung Teilchenphysik I Top-Quark-Praktikum

Introduction

In the following exercises, we will study top quarks produced at the LHC, using data that has been collected with the CMS detector. Top quarks are the heaviest quarks known as of now. At the LHC, top quarks are mostly produced in pairs of quark and antiquark. These quarks decay under weak interaction almost exclusively into a W boson and a b quark. W bosons decay in about $\frac{2}{3}$ of all cases hadronically into two quarks and in about $\frac{1}{3}$ of all cases leptonically into a charged lepton and the corresponding neutrino. Combining these decays, there are 3 possibilities for a $t\bar{t}$ pair to decay:

- all-hadronic: both W bosons decay into quarks
- dileptonic: both W bosons decay into lepton and neutrino
- semileptonic: one W boson decays into quarks and the other decays into lepton and neutrino

In this exercise, we will study the semileptonic channel, more precisely the μ+jets channel. This channel has the advantage of a high branching ratio compared to the dileptonic channel and a small background contribution compared to the all-hadronic channel.
The used files (simulated Monte Carlo and real data events) are already preselected by cutting on the transverse momentum (p_T) and the pseudorapidity (η) of both the jets and the muon. The used cut values are common for $t\bar{t}$-analyses at CMS. In addition, the events are required to have at least two jets and exactly one muon fulfilling the p_T and η requirements.

object	p_T[GeV/c]	$\mid \eta \mid$
μ	> 45	< 2.1
jet	> 30	< 2.4

Figure 1: Summary of the preselection cuts that are applied to the samples used in this tutorial

Manual

There are two programs to use. The first one is a C++ program (`RunTTBar.cc`) to analyse the samples as well as create and fill histograms. Compile the source code of `RunTTBar.cc` with `make` after every change. This program writes the created histograms into a new `root`-file which will appear in the directory `results/`. The second program `(plotHistos.C)` is a `root` macro for plotting the histograms that were created by the C++ program. Both programs already contain the bulk of the code, but you will have to add specific passages yourself.

Poolraum

Copy the file `ttbar.tar` from the web page into your working directory and unpack it with `tar -xf ttbar.tar`

Virtual machine

Download the file `ttbar.tar` from the web page into your working directory and unpack it with `tar -xf ttbar.tar`

Files

You also will need the following files, which contain events from either simulation or measured data. In the computer pool you can find the samples under
`/home/staff/chwalek/WS1516/ttbar_CMS/samples/`
but for the VM you can also get them from the web page. In either case you will need to change the path in the file `RunTTBar.cc` accordingly:
`pathName="/path/to/your/samples/"`

- `data_NJets4.root`: dataset with an integrated luminosity of about $5\,\mathrm{fb}^{-1}$ containing events with one muon and 4 or more jets
- `MC_TTBar_NJets2.root`: Monte Carlo sample of simulated $t\bar{t}$ events with one muon and 2 or more jets
- `MC_WJets_NJets2.root`: Monte Carlo sample of simulated W+jets events with one muon and 2 or more jets
- `MC_TTBar_NJets4.root`: Monte Carlo sample of simulated $t\bar{t}$ events with one muon and 4 or more jets

- `MC_WJets_NJets4.root`: Monte Carlo sample of simulated W+jets events with one muon and 4 or more jets

List of commands for RunTTBar.cc

Command	Meaning
`ch->GetEntry(iE)`	go to event iE
Monte Carlo Events	KITAGenTopEvent* KitaGen
`KitaGen->tops[0].vec.Pt()`	get transverse momentum of first top quark
`KitaGen->tops[1].vec.Eta()`	get pseudorapidity of second top quark
Measured jets	vector<KITAJet>* KitaJets
`KitaJets->size()`	number of jets
`KitaJets->at(i).btag_combSV`	output value of the combined secondary vertex b-tagger for jet i
`KitaJets->at(i).vec`	get momentum 4-vector vector of jet i
`KitaJets->at(i).vec.x()`	get momentum in x-direction of jet i
`KitaJets->at(i).vec.Eta()`	get pseudorapidity of jet i
`KitaJets->at(i).vec.Pt()`	get the transverse momentum of jet i
Measured missing transverse energy	KITAPFMet* KitaMet
`KitaMet->vec`	get 4-vector of missing transverse energy
`KitaMet->vec.Pt()`	get sum of missing transverse energy
Vector for M_3 **calculation**	KITA4Vector tvec
`tvec.M()`	get invariant mass of 4-vector *tvec*
Measured muon	vector<KITAMuon>* KitaMuon
`KitaMuon->at(0).vec`	get muon momentum 4-vector
`KitaMuon->at(0).vec.x()`	get momentum in x-direction of the muon
`KitaMuon->at(0).vec.Pt()`	get the transverse momentum of the muon
Neutrino 4-vector	KITA4Vector p4Neutrino
`p4Neutrino.SetPz(pz)`	set z-momentum of the neutrino to pz
`p4Neutrino.SetE(energy)`	set energy of the neutrino

Exercises

1 Kinematics of $t\bar{t}$ events

Have a look at the **simulated** $t\bar{t}$ events. Plot the distributions of p_T and η of top quarks and top antiquarks. Do you see why $t\bar{t}$ events are called high-p_T events? Also have a look at the correlation of p_T and η.

Instructions:
Here we use the Monte Carlo sample `MC_TTBar_NJets2.root`. All events in this sample are simulated $t\bar{t}$ events. The code to declare and fill the histograms for top quarks already exists in `RunTTBar.cc`. Compile this program with `make` and run it with `./RunTTBar MC_TTBar_NJets2.root`. This program creates a new `root`-file with the histograms (`Histos_MC_TTBar_NJets2.root`) in the directory `results/`. You can open this `root`-file with the `TBrowser` in `root`. To plot the first three histograms start `root` and run the plot-macro with `.x plotHistos.C(1)`. The three histograms for the top antiquarks are also declared. Edit the program code in `RunTTBar.cc` to fill these histograms for the top antiquark. Compile and run it again. Edit the macro `plotHistos.C` to also plot the histograms of the top antiquarks.

2 Number of jets

How many jets would you expect in a semileptonically decaying $t\bar{t}$ event? Plot the number of jets for the simulated $t\bar{t}$ events and for the simulated W+jets background. When studying $t\bar{t}$ events there are many more backgrounds to consider. For this exercise we use only the contribution of W+jets, because this is the most important background. Which cut on the number of jets would you use in order to reject W+jets background and keep $t\bar{t}$ events?

Instructions:
Write the code to fill the histogram with the number of jets in `RunTTBar.cc`. Compile and run this program with `./RunTTBar MC_TTBar_NJets2.root` for $t\bar{t}$ events and `./RunTTBar MC_WJets_NJets2.root` for W+jets background. Add code to `plotHistos.C` to normalize the histograms to 1.0 and draw the normalized signal and background histograms into one plot. "Normalized" means both here and later that the sum over all bins is 1. To plot the histograms open `root` and execute the plot script with `.x plotHistos.C(2)`.

3 Shape variables

For the following exercises we use the samples with 4 or more jets.
Study which event shape variable could be used to discriminate background (W+jets) from signal ($t\bar{t}$). Study the missing transverse energy $\not{E}_T$ in an event, the transverse momentum (p_T) of the muon and the pseudorapidity (η) of the muon. Also look at the sum of all transverse momenta (H_T) and the mass of the 3 jets with the highest transverse momentum (M_3). Which of these variables performs best in distinguishing W+jets background from signal?

Instructions:
Add code to fill the corresponding histograms in `RunTTBar.cc`. Sum up the absolute values of the transverse momenta of all jets, of the muon and of the neutrino to calculate H_T.
Construct three loops to run over all combinations of 3 jets. For each combination calculate the p_T of the resulting 4-vector. Choose that combination that yields the largest p_T, and calculate the invariant mass of this combination. This – the invariant mass of the summed jets with the highest p_T – is M_3.
The maximum of this distribution should lie around the top quark mass. The last bin is the overflow bin. It includes all events with $M_3 > 700\mathrm{GeV/c}^2$. The code to move all events with $M_3 > 700\mathrm{GeV/c}^2$ into the last bin can be found at the end of `RunTTBar.cc`.
Run this program for the $t\bar{t}$ Monte Carlo sample with `./RunTTBar MC_TTBar_NJets4.root` and for the W+jets sample with `./RunTTBar MC_WJets_NJets4.root`.
At the end of `plotHistos.C` is a function called `shapeVar` that is used in the (already prepared) plotting of the signal and background distributions. In this function you again have to add code to normalize the histograms to 1.0. Load the script and plot these histograms with `.x plotHistos.C(3)`.

4 Signal fraction and S/B

For calculating the $t\bar{t}$ cross-section, you need to know the number of $t\bar{t}$ events – so you have to extract the signal fraction (f_{sig}) of the data. It is defined as the number of $t\bar{t}$ events divided by the total number of observed data events. Use the variable M_3 to determine the signal fraction in a template fit. In the template fit the normalizations of the signal and background shapes are varied until the combination of both fits optimally to the data.
Calculate the signal to background ratio (S/B) in your data sample from the result of the fit. Check whether a cut on the number of b-tagged jets ($N_{b-tag} \geq 1$) in an event increases the signal fraction and S/B.

Instructions:
Run `./RunTTBar data_NJets4.root` to get the distribution of M_3 in data, in addition

to the distributions for $t\bar{t}$ and W+jets events.
First, normalize the M_3 histograms from $t\bar{t}$ and W+jets to the total number of observed data events in `plotHistos.C`. To fit the signal fraction the function `PerformFit` is called, which is defined in `m3Fit.h`. We will use a χ^2 function to check how well a given combination of the two templates describes the data:

$$\chi^2 = \sum_i^{N_{bins}} \frac{(N_i^{pred} - N_i^{data})^2}{\sigma^2(N_i^{pred})}$$

In this, N_i^{data} is the number of data events in bin i and N_i^{pred} is the corresponding number of predicted events. Since the $t\bar{t}$ and W+jets distributions are normalized to the data, N_i^{pred} can be calculated as the sum of the number of $t\bar{t}$ events in bin i ($N_i^{MC-t\bar{t}}$) multiplied by the signal fraction (f_{sig}) and the number W+jets events in bin i ($N_i^{MC-W+jets}$) multiplied by $(1 - f_{sig})$. For the statistical error you should use Poisson errors $\left(\sigma(N_i^{pred}) = \sqrt{N_i^{pred}}\right)$. Add the code to calculate N_i^{pred} for each bin to `m3Fit.h` and implement the χ^2 function. The rest of the code has already been written. It uses `TMinuit` to find the signal fraction that minimizes the χ^2 function.

Use the signal fraction from the fit to calculate the signal to background ratio (S/B). Furthermore, modify the code creating a stack plot in `plotHistos.C`. In the stack plot the signal and background distributions are normalized to the normalisation determined in the fit and drawn stacked. Draw the data distribution into this plot, too. Run `.x plotHistos.C(4)` to perform the fits and plot the histograms.

Finally, repeat the same exercise using only events containing at least one b-tagged jet. To calculate the number of b-tagged jets write a loop over all jets in `RunTTBar.cc` and count the tagged jets. Use the so-called medium working point of the combined secondary vertex (CSV) b-tagger for this, i.e., count a jet as tagged if the b-tagger output fulfills `btag_combSV > 0.679`. There are usually at least three working points for b-tagging: loose, medium and tight. These working points are defined by their mistag rate, which is the fraction of non-b jets that are erroneously tagged as b-jets. At the medium working point the mistag rate is 1% while the b-tagging efficiency for b-jets is about 70%.
Fill the histograms of M_3 again – but this time for events with at least one b-tagged jet and add code to `plotHistos.C` to repeat the fit with these histograms.

5 $t\bar{t}$ cross section

Calculate the cross section for $t\bar{t}$ production at $\sqrt{s} = 7\,\text{TeV}$ using the following formula

$$\sigma_{t\bar{t}} = \frac{N^{data} \cdot f_{sig}}{L \cdot \epsilon_{eff}}.$$

The selection efficiency ϵ_{eff} is the fraction of $t\bar{t}$ events that is left after all selection cuts (one muon, 4 jets, b-tag, etc.) have been applied. For $t\bar{t}$ the selection efficieny

$\epsilon_{eff} = 0.0259$ has been calculated from MC. The dataset corresponds to an integrated luminosity of $L = 5\,\mathrm{fb}^{-1}$.
Compare your result to the predicted theoretical cross-section[1]:
$\sigma_{t\bar{t}}^{NNLOapprox}(m = 173\,\mathrm{GeV/c}^2, \sqrt{s} = 7\,\mathrm{TeV}) = 163^{+7+9}_{-5-9}\mathrm{pb}$

6 Reconstruction of the top quark 4-vectors

Reconstruct the semileptonically decaying top quark pair from four measured jets, the muon and the missing transverse momentum. One big problem is that you do not know which jet comes from which quark, and in many events there are also more than four jets. To get the best combination for an event reconstruct all possible jet combinations and select the combination with the smallest χ^2 value (as defined below). Ensure that jets interpreted as b-jets are b-tagged at the loose working point (`btag_combSV > 0.244`) if possible. Reconstruct the mass of the leptonically ($m_{t,lep}$) and hadronically ($m_{t,had}$) decaying top quarks. For each $t\bar{t}$ pair calculate the average top quark mass ($m_{t,avg} = \frac{m_{t,lep}+m_{t,had}}{2}$), too.

Instructions:
Add code to `RunTTBar.cc` to reconstruct the hadronically decaying ($t_{had} \to b + W$, $W \to q, q'$ with $q, q' = u, d, s, c$) and the leptonically decaying ($t_{lep} \to b + W$, $W \to l + \bar{\nu}$) top quarks for each event.
First, apply an event selection with $N^{med}_{b-tag} \geq 1$. Use the medium working point (`btag_combSV > 0.679`) for this. Write a loop to count the number of b-tagged jets (N^{loose}_{b-tag}) with the loose working point (`btag_combSV > 0.244`). Keep in mind that any jet that passes the medium working point will also be counted among these loosely tagged jets.
Add loops to get all possible assignments of four jets to the four quarks of the $t\bar{t}$ decay. Ensure that if only one jet is b-tagged ($N^{loose}_{b-tag} = 1$), this jet is assigned to one of the two b-quarks. In events with additional b-tagged jets ($N^{loose}_{b-tag} > 1$), make sure that only b-tagged jets are assigned to the two b quarks. Otherwise skip the combination.

Reconstruct the 4-vector of the hadronically decaying W boson as the sum of the 4-vectors of two jets. Then, reconstruct t_{had} from this W boson candidate and another jet, which is interpreted as coming from the b-quark of the top decay.
Calculate the z-component of the neutrino momentum. Use the function `calcNuPz(KitaMuon->at(0).vec,KitaMet->vec)` for this, which needs the 4-vector of the muon and the x- and y-momentum of the missing transverse momentum. With the calculated z-momentum and the missing transverse momentum vector you can calculate the 4-vector of the neutrino. For this, save the x- and y momentum of the neutrino with `p4Neutrino = KitaMet->vec`. Then add the calculated z-momentum with `p4Neutrino.SetPz(pznu)`. Finally, calculate and save the energy of the neutrino

[1]Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution, from Nikolaos Kidonakis, arXiv:1009.4935v2

`p4Neutrino.SetE(p4Neutrino.P())`.
Add the 4-vector of the muon and the 4-vector of the neutrino to get the 4-vector of the leptonically decaying W boson. Assign a fourth selected jet to the b quark that comes from the leptonically decaying top quark.
Calculate the χ^2 for every possible assignment of jets.

$$\chi^2 = \frac{(m_{W_{had}} - M_{W_{had},exp})^2}{\sigma^2_{M_{W_{had}},exp}} + \frac{(m_{t_{lep}} - m_{t_{had}} - \Delta M)^2}{\sigma^2_{\Delta M_{t,exp}}}$$

The reconstructed mass of the hadronically decaying W boson is $m_{W_{had}}$. $m_{t_{lep}}$ is the reconstructed mass of the leptonically decaying and $m_{t_{had}}$ the reconstructed mass of the hadronically decaying top quark. The remaining variables in the χ^2 function were calculated from simulated $t\bar{t}$ events, using the combination that comes closest to the truth for each simulated event . The mean value of the expected hadronically decaying W boson mass distribution is $M_{W_{had},exp}$ with a distribution width of $\sigma_{M_{W_{had}},exp}$. ΔM is the mean of the distribution of the mass difference between the hadronically and the leptonically decaying top quark in an event – this difference arises from imperfections in measurement and reconstruction. $\sigma_{\Delta M_{t,exp}}$ is the width of that distribution.

$M_{W_{had},exp}$	$78.1\,\mathrm{GeV/c^2}$
$\sigma_{M_{W_{had}},exp}$	$11.2\,\mathrm{GeV/c^2}$
$\sigma_{\Delta M_{t,exp}}$	$45.2\,\mathrm{GeV/c^2}$
ΔM	$-7.5\,\mathrm{GeV/c^2}$

For each event, choose the combination that yields the smallest χ^2 value. Fill the reconstructed top quark masses (hadronically/leptonically decaying and average top quark mass for each event) of the chosen combination into the corresponding histogram. The average top quark mass is the average between the mass of the leptonically and hadronically decaying top quark ($m_{avg} = \frac{m_{t_{lep}}+m_{t_{had}}}{2}$).
Compile `RunTTBar.cc` and run `./RunTTBar data_NJets4.root`. Plot the three reconstructed top quark masses by running `.x plotHistos.C(5)`. Add code to fit the reconstructed top quark masses with a Landau function in `plotHistos.C`. Plot the three Landau functions together with the reconstructed top quark masses by running `.x plotHistos(5)` again. As a last step, run `./RunTTBar MC_WJets_NJets4.root` and `./RunTTBar MC_TTBar_NJets4.roo` for the Monte Carlo sample – the output will be used for the next exercise.

7 Determination of the top quark mass

First, run `RunTTBar.cc` over all samples with at least 4 jets. Have a look at the simulated $t\bar{t}$ events. The Monte Carlo sample (`MC_TTBar_NJets4.root`) was generated with a top quark mass of $m_{t,true} = 172.5\,\mathrm{GeV/c^2}$. When you fit the reconstructed average top quark mass, it will be different to the top quark mass which was used by the generator. This difference between true and reconstructed top quark mass exists also in the observed data.

Subtract the background from the data to get the distribution of the top quark mass for a pure $t\bar{t}$ signal. Fit a Landau function to this distribution and add the calculated difference from Monte Carlo to the maximum of the fit function. This is the true top quark mass in data. Compare your result to the PDG mass[2]: $m_t = 173.5 \pm 0.6 \pm 0.8\,\mathrm{GeV/c^2}$

Instructions:
Calculate first the difference between true and average reconstructed top quark mass from the simulated $t\bar{t}$ sample. Fit a Landau function to the distribution of the reconstructed top quark mass in a range around the signal peak (50 - 400 GeV/c^2). The maximum of this function is the reconstructed mass m_{rec}.
Use `Landau_function->GetMaximumX()` for this. Calculate the difference $m_{t,true} - m_{t,rec}$.
Plot the distribution of the average reconstructed top quark mass with the corresponding fit function. Add the code to plot into this same canvas a line with the true top quark mass and a line with the average reconstructed top quark mass. Perform the plot by running `.x plotHistos.C(6)`
Add code to `plotHistos.C` to subtract the scaled background distribution from the distribution of the reconstructed average top quark mass in data. Use the signal fraction measured in exercise 4 for this. Fit the background-subtracted data distribution in the same range around the signal peak (50 - 400 GeV/c^2) using a Landau function. Add the difference between reconstructed and true top quark masses you calculated for the simulated $t\bar{t}$ events to the position of the maximum of this function. This is the true reconstructed top quark mass. Plot the distribution of the reconstructed top quark mass and data without background with their corresponding Landau function by running `.x plotHistos.C(6)`.

[2]Particle Physics Booklet: July 2012