Systematic Uncertainty Study Part1

Four top

Anshul Kapoor Huiling Hua[1] Hongbo Liao[1]

[1]IHEP

May 15, 2023

Outline

Section 1

Systematic uncertainties

Systematic uncertainties

- Theoretical uncertainties
  - QCD scale variations and PDF
  - ISR, FSR
  - how to get them???
- Experimental uncertainties
  - Luminosity(✓)
  - Pileup reweighting(✓)
  - Trigger efficiency
  - Lepton reconstruction & identification(✓); lepton energy scale
  - Hardronic tau identification (✓); tau energy scale
  - B-tag shape correction
  - JES and JER
  - Fake rate uncertainty(only 1tau0l)
- https://twiki.cern.ch/twiki/bin/view/CMS/TopSystematics

Intergrated luminosity(log uncertainty)

- 1.2%(36.31), 2.3%(41.48), 2.5% (59.83)
- Correlation?
- https://twiki.cern.ch/twiki/bin/view/CMS/LumiRecommendationsRun2

Pileup reweighting(shape)

- Centrally provided weight files both for the nominal weights of the simulated samples, as well as to generate the varied distributions according to the cross section uncertainty(4.5% on 69.2 mb)

Prefiring(shape)

- Implementation
- Available in NanoAOD

Leptons: Efficiency of lepton, energe scale(shape)

- Efficiency SF and systematic variation provided by TOP MVA
- Energy scale

Jet: JES and JER

Uncertainties in combine

- Log normal uncertainties
- Shape uncertainties
  - shapes will be interpolated with a 6th order polynomial for shifts below 1 σ and linearly beyond
  - normalizations are interpolated linearly in log scale just like we do for log-normal uncertainties

Section 2

Datacard

Datacard

- Following convention: https://twiki.cern.ch/twiki/bin/viewauth/CMS/HiggsWG/HiggsCombinationConventions

```
imax * number of bins
jmax * number of processes minus 1
kmax 6 number of nuisance parameters
------------------------------------------------------------------------------------------------------------------------
shapes *   SR1tau1l_2017  /publicfs/cms/user/huahuil/tauOfTTTT_NanoAOD/forMVA/2017/v4baselineBtagRUpdated_v57ovelapWithTausF/mc/variableHists_v1sysVariation1tau1l/combine/
templatesForCombine1tau1l.root $PROCESS_SR_BDT $PROCESS_$SYSTEMATIC_BDT
------------------------------------------------------------------------------------------------------------------------
bin          SR1tau1l_2017
observation  -1
------------------------------------------------------------------------------------------------------------------------
bin         SR1tau1l_2017  SR1tau1l_2017  SR1tau1l_2017  SR1tau1l_2017
process      tttt           singleTop      tt             ttX
process      0              1              2              3
rate         -1             -1             -1             -1
------------------------------------------------------------------------------------------------------------------------
lumi_13TeV_2017 lnN        1.023      1.023        1.023         1.023
CMS_pileup_2017 shape      1          1            1             1
CMS_prefiring_2017 shape   1          1            1             1
CMS_eff_t_vsJet2017 shape  1          1            1             1
CMS_tttt_eff_e_2017 shape  1          1            1             1
CMS_tttt_eff_m_2017 shape  1          1            1             1
------------------------------------------------------------------------------------------------------------------------
SR1tau1l_2017 autoMCStats 10 0 1
```

Section 3

Impact

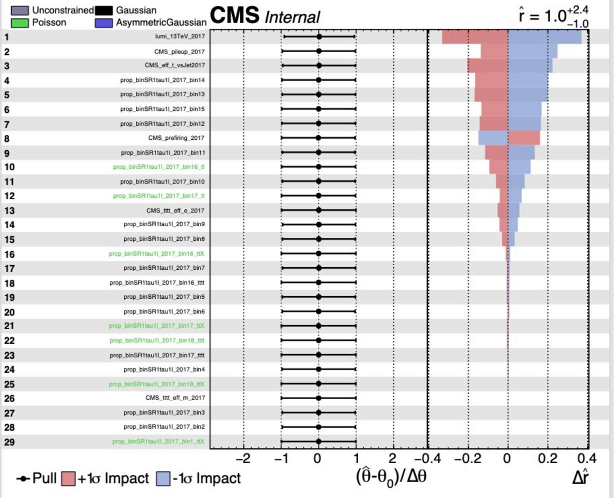
Impact plot

- "impact" of each uncertainty
- the shift in the signal strength, with respect to the best-fit, that is induced if a given nuisance parameter is shifted by its $\pm 1\sigma$ post-fit uncertainty values. with all other parameters profiled as normal
- Effectively a measure of the correlation between the NP and the POI, useful for determining which NPs have the largest effect on the POI uncertainty
- $(\widehat{\theta} - \theta_I)/\sigma_I$
  - where θ and θ_0 are the post and pre-fit values of the nuisance parameter and $\delta\theta$is the pre-fit uncertainty
  - post-fit uncertainty divided by the pre-fit uncertainty meaning that parameters with error bars smaller than $\pm\ 1\sigma$ are constrained in the fit.

calculate impact with combine

- combineTool.py -M Impacts -d workspace_part3.root -m 200 –rMin -1 –rMax 2 –robustFit 1 –doInitialFit
  - perform an initial fit for the signal strength and its uncertainty
- combineTool.py -M Impacts -d workspace_part3.root -m 200 –rMin -1 –rMax 2 –robustFit 1 –doFits
  - run the impacts for all the nuisance parameters
- combineTool.py -M Impacts -d workspace_part3.root -m 200 –rMin -1 –rMax 2 –robustFit 1 –output impacts.json
  - collect all the output and convert it to a json file
- plotImpacts.py -i impacts.json -o impacts
- How to be blind? the data_obs is actually MC(s+b)
- http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/part3/nonstandard/#nuisance-parameter-impacts

Impact plots



Section 4

1tau1l Results

1tau1l

- BDT score as discriminant
- Shape uncertainties,
  - Event weight uncertainties
    - Evaluate BDT score normally, fill histogram with weight up and down
  - Energy scale uncertainties
    - Reproduce input variables with energy scale up or down, evaluate BDT score

Results before and after (2017)

- Before
  - significance: 0.49; limit: 4.72;
- After 6 systematic uncertainties
  - significance: 0.48; ;limit: 4.92

Section 5

Back up

Questions

- What should be our binning strategy and why?
  - Maybe group low statistic bins together because the best result is a tradeoff between statistic uncertainty and systematic uncertainties
- Should we group the subprocesses together or not?
  - When add subprocesses, the uncertainty is recalculated for each bin using error propagation formula, this could be problematic
- How to treat correlated and uncorrelated uncertainties?
- How to intepre tate the pull?
- How to connect the impact to the uncertainty of significance?

How to do systematic study?

- Expected significance and limit after considering systematic uncertainties
  - How does them change and why?
- Impact plot
- https://twiki.cern.ch/twiki/bin/view/CMS/TopSystematics

Look elsewhere effect