COMPUTING METHODS IN HEP Home exam Spring 2019

(To be returned by 24.00 on Monday 20.5.2019)

The CMS experiment was designed to be able to measure muons with a good precision. Your task is to study whether the Higgs boson can be detected in the LHC Run3 with 14 TeV center-of-mass energy with the discovered mass $m_H = 125$ GeV/c^2 in the $H \to \mu\mu$ decay channel.

The main backgrounds for this channel are the γ^*/Z and $t\bar{t}$ production. Assume that the collected amount of data will be ~300 fb^{-1}.

1. Create generator-level datasets for

   a) signal

   b) background

   The events are triggered with a HLT_DoubleIsoMu20_eta2p1 trigger with symmetric thresholds: $|\eta| < 2.1$ and $p_T > 20$ GeV/c for both muons. Ignore the isolation (Iso) at this point. Fill the datasets with events passing the trigger. What is the trigger efficiency?
   (6 points)

2. Analyze your data.

   a) Identify the signal muons. Take as the muon candidates the muons in the event. Apply the following operations and standard selection criteria to each of the muon candidates:

      – Simulate the muon measurement error by applying 1% Gaussian smearing to their momenta and 2 mrad Gaussian smearing to their angles θ and ϕ.

      – Selection of the two signal muons: Require that the muon candidates satisfy $p_T > 30$ GeV/c.

      – Track isolation: Require that the sum of the momenta of charged pions within $\Delta R < 0.3$ of a muon candidate is smaller than 1.5 GeV/c in p_T; after this criterion at least two muon candidates have to exist in the event

      Report the number of events passing the selection. (6 points)

   b) Reconstruct the invariant mass $M_{\mu\mu}$. Produce a histogram of $M_{\mu\mu}$ with correct normalization (i.e. that each entry in the histogram corresponds to a correct amount of cross-section in fb). Plot two invariant mass distributions in the same figure: the background only, and the sum of the signal and background. (6 points)

   c) Fit the signal+background and background only $M_{\mu\mu}$ histograms with appropriate functions over an appropriate mass range.

      Choose a reasonable mass window and calculate the number of signal and background events. Estimate the statistical significance of the signal peak with the naive expression $N_S/\sqrt{N_B}$, where N_S and N_B are the number of signal and background events, respectively. (6 points)

3. If the data were real data, you'd have to present your work to the world-wide scientific community. Reason why the scientific community should believe in your results. Can you convince yourself? Why/why not? How would you improve your study to make it more realistic? (6 points)

Please make a gzipped tar-ball all your files (except the datasets) and return it by email to: Sami.Lehti@cern.ch. Include in your answer a written description how you proceed and where you argument any non-trivial choice you make.