Exercises 4: **Advanced hypothesis testing**

Toni Šćulac (Faculty of Science, University of Split, Croatia)

You are now given experimental data from the Large Hadron Collider in the "LHC_data_2023.txt" file. This dataset was collected with the CMS experiment in the period from 2021 to 2023. **Once again, it is very important not to look at the data before you perform all the necessary steps in advance!**

Problem 1

Scientific community got excited from the results of 2021 data analysis and they all agreed that if there is a new particle in the $m_{e^+e^-}$ spectrum it has to be the long proposed **CSC particle** that would explain the origin of the matter-antimatter asymmetry. From first principles they can calculate that the cross section for the production and decay of this particle predicts a total of 1% events to be results of decay of the CSC particle. They also know that after taking into account detector resolution the PDF of the signal is a simple Gaussian distribution. You and your colleagues calculated based on the callibrations of the CMS detector that the resolution of the Gaussian function has to be 0.2 GeV. In addition, theory community is sure that the mass of the particle has to be in the range $20.0 < m_{\mathrm{CSC}} < 40$ GeV. Using MC simulation draw an example of what it would look like if CSC exists in the given sample. In order to do so, choose any value of m_{CSC} that you want!

Problem 2

Using the test statistic definition from previous exercise calculate what is your expected sensitivity. That means to calculate what would be the significance if the signal exists with properties predicted by the theory community. You still can't use real data, but instead do it on the MC you produced in problem 1! Think carefully if you need a new $g(t|H_0)$ or you can reuse the one from exercise 3! Do you encounter any problems with expected significance? How would you solve them?

Problem 3

The long-awaited time has arrived. You are ready to unblind the data and calculate the observed significance. Did you discover a new particle?

Problem 4*

Perform a p-value scan. That means that you need to calculate the PDF of the test statistic for each bin and then calculate observed p-value as a function of invariant mass. Where is the minimum of your distribution?

Problem 5*

Is the newly discovered particle the long awaited CSC particle or not? Try figuring it out using statistics!