Exercise 3: **Hypothesis testing**

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

You are given experimental data from the Large Hadron Collider in the "LHC_data.txt" file. Data includes the reconstructed invariant mass of the two photons in search for a new particle. **Do not look at the data before you perform all the necessary steps in advance!**

Problem 1

Your colleagues from the theory department have studied the production of two photons in the Standard Model in detail and they have concluded that they are described by the exponential PDF $f(m_{\gamma\gamma};\alpha) = \alpha \exp(-\alpha m_{\gamma\gamma})$ where the parameter $\alpha = 2.5 \cdot 10^{-2}$ is defined by the theory. Draw the theoretical PDF and generate a sample of random data using it. Define the null hypothesis H_0 for the process of searching for a new particle in the LHC data.

Problem 2

Define the test statistic t as the χ^2/N value of the fit when fitting the data distribution of the invariant mass $m_{\gamma\gamma}$ to the theoretical model $f(m_{\gamma\gamma};\alpha = 2.5 \cdot 10^{-2})$ defined in Problem 1. N is the total number of events and $\chi^2 = \sum_{i=1}^{N} \left[f(m_{\gamma\gamma};\alpha) - y_i\right]^2$; where $f(m_{\gamma\gamma};\alpha$ is the predicted number of events and y_i is the observed number of events in bin i. Is this a good choice for the test statisic? How would you find a better test statistic? Using the null hypothesis derive and draw the probability density distribution of the test statistic $g(t|H_0)$. In order to do so you will have to generate many experiments under the assumption that the null hypothesis H_0 is valid. What do you expect, how will this distribution look?

Problem 3

Starting from $g(t|H_0)$ define the critical region for which you will reject the null hypothesis. What significance you want to use and why? Derive the t_{critical} value. How many pseudo-experiments you have to throw in order to get a reliable statistics in the $g(t|H_0)$ distribution? Is there any other way than just to throw more pseudo-experiments?

Problem 4

Using data from the LHC derive the observed value of test statistic t_{obs}. Did you find a new particle? Elaborate your answer and calculate the p-value.

Problem 5*

Redefine the test statistic t as the χ^2 value of the fit in a given mass range. Using this approach perform an observed p-value scan as a function of the hypothetical mass of the particle. How can you use this kind of scan to measure the mass of the newly found particles?