1 Search for a New Particle

You will be provided with a distribution with a bin size of 1 GeV (each student will receive his or her own histogram) emulating a possible experimental measurement. The hypothetical measurement is an invariant mass distribution of two photon candidates measured by an XYZ experiment as it has been searching for a new particle that has an unknown mass and an unknown production rate. However, if this particle exist, in events where this particle has been produced the invariant mass of two photons should appear as a gaussian bump of the form:

\[
s(m) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(m - m_0)^2}{2\sigma^2}\right)
\]  

Theoretically, the width of the particle itself is expected to be very small, therefore \( \sigma \) in the formula above is determined by the experimental accuracy in the measurement of the invariant mass of two photon candidates, which is known to scale as:

\[
\sigma(m) = 0.05 \times \sqrt{m}
\]
From Monte Carlo simulations of the signal process and detector response, you know that the acceptance of your analysis (which is the full efficiency of all selections) for your signal events follows the following parametric dependence:

$$\alpha = 1 - \frac{10 \text{ GeV}}{m}$$  \hspace{1cm} (3)

and is known to 3% accuracy (i.e. if for particular mass $m_0$ the calculation yields $\alpha_0 = 0.2$, you expect the true acceptance to be within $\alpha = 0.200 \pm 0.006$ at 68% C.L.)

In addition to signal events, you know that there is a large background contamination due to $pp \rightarrow \gamma\gamma$ process, which has very large cross-section but is not expected to have a bump. Furthermore, from simulation studies you know the expected shape of the background to be:

$$dN/dm = 40000/m + 90,000/m^2$$  \hspace{1cm} (4)

and you have reasons to believe that any systematic uncertainties are negligible.

The integrated luminosity of the dataset you are analyzing is $10 \text{ fb}^{-1}$.

## 2 Assignment

You task is to analyze the distribution and report your findings as to whether you believe there is sufficient evidence that the particle exists.

The necessary components of your report:

- The 95% C.L. upper limit on the production cross-section of the new particle as a function of mass
- Numeric evaluation of the probability that the observed distribution is consistent with the background only hypothesis
- Determine the most likely values of the mass for the new particle if it exists
- For the most suspicious value of $m$, measure the signal cross-section.
- Assume for a minute that there is signal at the suspicious mass $m$ and it has the cross section you just calculated. Provide quantitative evaluation of the probability that the observed data is consistent with the background plus the described signal hypothesis.
- Compare the two hypotheses. Expand your calculations on a more general comparison of the background only and signal plus background hypotheses for arbitrary mass. Note that you will need to use a test statistic that “knows” about both signal and background shapes, e.g. like the statistic used in the $CL_S$ method. Note that you will need to provide a distribution of whatever the metrics you decide to use for this probability as a function of signal particle mass.
Based on available information, make a conclusion of whether you can claim a discovery or quantify the significance of a potential excess in your data.

3 Hints

The idea with this assignment is that you are pretty much on your own, you can use any literature you want and can ask any questions you want as you are working on this assignment (except you can’t have anyone writing code for you or doing your work for you). However, various hints will be provided over the next weeks as you work on this assignment and these hints will be added to this section and distributed to everyone.

Hint 1: With the experience you already have, you should be able to do pretty much everything required of you in the assignment, the only point so far not discussed at all is the test statistic for the case where you compare two hypotheses against each other (in your previous assignments, you calculated a p-value for a known expectation, so there was never any second hypothesis in sight). The trick is to define a new test statistic that would carry information on how much probable one option is over the other, the $CL_S$ method provides such statistic.

Hint 2: For your first round of evaluations, ignore the systematic uncertainty in your signal acceptance, you can add it later.