DEPARTMENT OF PHYSICS & ASTRONOMY
3459 Final Examination
14:00 - 17:00 : 10th December 2012

Please read the exam guidelines, rules, instructions and marking criteria at
`http://moodle.ucl.ac.uk/mod/wiki/view.php?id=13963&page=Final+exam`
(linked from the *Exams and Coursework* page).

This exam is worth 50% of your final mark for the course. The duration of the exam is 3 hours. Students should upload the Java source code files for their solution using Moodle under the section headed “Exam 2”.

Note : no knowledge of nuclear or particle physics is required to answer this question.

In this examination you will be processing (simulated) data from two channels in the Higgs search of the ATLAS experiment at the Large Hadron Collider at CERN. You will read data from a web directory, store the data in (a) suitable container(s) and perform statistical analyses.

Three files are provided in the following web directory:
`http://www.hep.ucl.ac.uk/undergrad/3459/exam-data/2011-12/`.

- The file backgroundGG.txt contains the predicted number of background (non-Higgs) events for the analysis channel $h \to gg$ in 1 GeV bins of energy between 100 and 200 GeV. Each line of the file contains the following fields:
  - The low edge of the energy bin in GeV
  - The high edge of the energy bin in GeV
  - The predicted number of events with energy greater than or equal to the low edge but less than the high edge.
- The file backgroundZZ.txt contains the predicted number of background (non-Higgs) events for the analysis channel $h \to ZZ$ in 1 GeV bins of energy between 100 and 200 GeV. The format is the same as the backgroundGG.txt file.
- The file higgsData.txt contains one line for every Higgs candidate event in the ATLAS data sample. Each line contains the following fields:
  - The channel identification code, GG for $h \to gg$ or ZZ for $h \to ZZ$
  - The event energy in GeV

Part 1: 20/50 marks

Write a program to do the following:

- Read all the data from the `backgroundGG.txt` and `backgroundZZ.txt` files and store it in an(some) appropriately designed data structure(s).
- Determine the number of expected background events for the two channels, $h \to gg$ and $h \to ZZ$, in the energy range 120-140 GeV.
- Read in all the candidate Higgs data events from the `higgsData.txt` file. For each of the channels, determine and store the number of candidate events in each of the energy bins listed in the `backgroundGG.txt` and `backgroundZZ.txt` files.
- For each of the channels, determine the number of candidate events in the energy range 120-240 GeV.
- Calculate the log-likelihood for each of the channels using the following equation:

$$LL = \sum_{i=1}^{i=100} (y_i - n_i) + n_i \times \ln \frac{n_i}{y_i} \tag{1}$$

  where i is the energy bin number, y_i is the predicted number of background events in the bin and n_i is the measured number of candidate events in the bin.

Part 2: 15/50 marks Make the following enhancements to your code:

- Define an interface to return the predicted number of Higgs signal events in 1 GeV energy bins between 100 and 200 GeV, for any arbitrary distribtuion of Higgs signal events.
- Create an implementation of this interface that models the Higgs signal as a Gaussian distribution with a given normalisation (N), Higgs Mass (m_H) and width (σ_H) according to:

$$f(E) = \frac{N}{\sigma_H \sqrt{(2\pi)}} e^{-\frac{(E-m_H)^2}{2\sigma_H^2}} \tag{2}$$

  You may need to use the trapesium rule to calculate the number of events in a given bin:

$$\text{Number of Events} = 0.5 * (f(E_{high}) + f(E_{low})) * (E_{high} - E_{low}) \tag{3}$$

- Use the Gaussian implementation to produce 80 signal predictions for the channel $h \to gg$ for m_H between 110.5 GeV and 179.5 GeV in 1 GeV steps with $N = 100$ and $\sigma_H = 2$.
- Use the Gaussian implementation to produce 80 signal predictions for the channel $h \to ZZ$ for m_H between 110.5 GeV and 179.5 GeV in 1 GeV steps with $N = 6$ and $\sigma_H = 1$.

Part 3: 15/50 marks Make the following enhancements to your code to perform a likelihood analysis and determine the most likely value of the Higgs mass.

- For each channel and for each of the 80 values of the Higgs mass, m_H, determine the sum of the expected number background (from Part 1) and signal (from Part 2) events in each bin.
- Use this sum as y_i in Equation 1 to determine the log-Likelihood for each channel for each value of the 80 values of the Higgs mass
- Print out the Higgs mass with the lowest value of the sum of the log-likelihood in each channel.
- To determine whether you have discovered the Higgs you need to compare the log-likelihoods between the signal and background cases to see if you are above the five sigma discovery threshold. The number of sigma can be determined from the following equation:

$$\sigma = \sqrt{-2(\sum LL_{signal} - \sum LL_{background})} \tag{4}$$

- Determine the number of sigma and print to screen whether you have discovered the Higgs boson

Uploading your work

If you use your own classes from earlier modules, make sure you upload them as well as any new classes you create during the exam.

END OF PAPER