Year 3 Particle Physics Computing
Multivariate Analysis Project Description

Resource 1 *Function Minimization* by Fred James

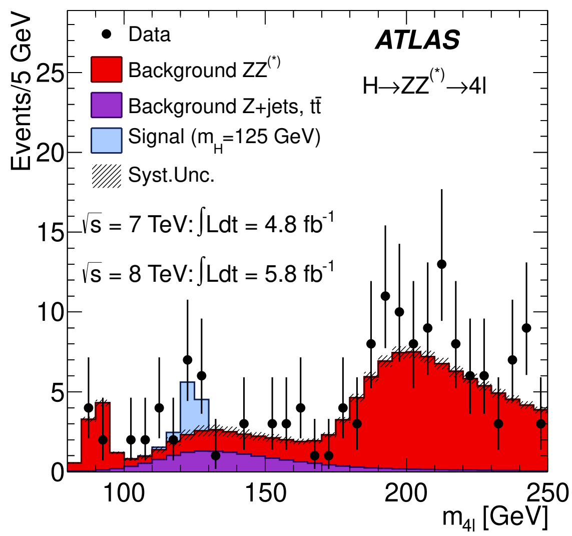
Resource 2 *A simplex method for function minimization* by Nelder and Mead

Resource 3 *A new algorithm for computing a single root of a real continuous function* by Ridder

Resource 4 *TMVA Users Guide* by Albertsson, *et al.*

Resource 5 *Numerical Analysis* by Burden and Faires; chapter 2 and 10.4

When searching for the Higgs, one of the final states used was $H \to Z(\to \ell^+\ell^-)Z(\to \ell^+\ell^-$, a final state with two lepton/anti-lepton pairs. Unfortunately, the decay of the Higgs boson is not the only process at the Large Hadron Collider (LHC) which produces this final state. There are a number of other processes which produce this final state, including continuum ZZ or $Z\gamma$ production where both Z bosons, or the Z boson and the photon, decay into lepton pairs. This is a common problem in particle physics; the final state of the signal for which we are looking will almost always have backgrounds which produce the same final state. A solution to this problem is to implement requirements in the data, oftentimes called called cuts, which help reduce the number of background events, while maintaining the number of signal events. The goal of this project is to develop algorithms to optimise a set of analysis cuts.



A very good example of how cuts can be used to eliminate background in the case of the Higgs is through an invariant mass cut of the four leptons. In the plot on the right, the number of muon pairs observed by the ATLAS experiment in the paper *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC* is plotted as a function of the four lepton invariant mass. If we were performing a simple counting experiment, and counted all the events within this plot there might be roughly 200 background events and only 5 signal events. Assuming the number of expected background events is Poisson distributed, then the fluctuation on the number of observed background events will be roughly $\sqrt{200} \approx 14$ and so an additional 5 signal events would just appear as an upward statistical fluctuation. If we saw 205 events rather than 200 events, this would only be 0.3 standard deviations above the background only hypothesis, and we would not be able to claim the discovery of the Higgs.

However, if we instead cut on the invariant mass of the two lepton pairs, we could significantly reduce the number of background events while keeping all the signal events. In this case, the four lepton invariant mass could be required to fall within the range $120 < m_{4\ell} < 130$ GeV. Now, the expected number of signal events is still the same, ≈ 5, but the expected number of background events has also been reduced to ≈ 5. If we observe a total of 10 events in the data, this is nearly 2.5 standard deviations away from the expected background only hypothesis, much better than the previous 0.3. Typically in particle physics if we see a 3 standard deviation excess we claim to see evidence for a particle, and if we see a 5 standard deviation excess we claim the discovery of a new particle.

Unfortunately we do not always have such clean variables as is the case for this Higgs analysis. Actually, this invariant mass plot for the Higgs boson is somewhat misleading; cuts on a number of other variables have already been implemented, prior to the invariant mass cut, to provide such a clean mass peak. We are only seeing the end result of a multi-variate analysis, where a set of cuts has been optimised to minimise the background while maximising the signal. More precisely, the metric,

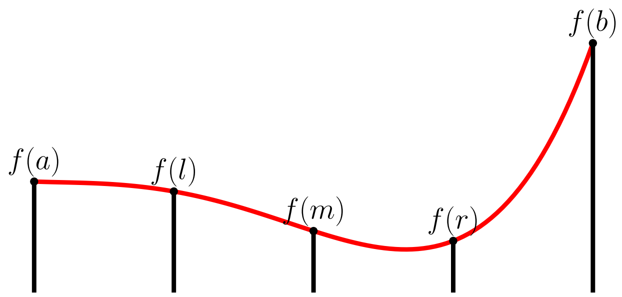
$$N_s/\sqrt{N_b} \tag{1}$$

is being maximised, where N_s is the number of expected signal events and N_b is the expected number of background events. This metric maximises the significance of observing a signal with respect to the background only hypothesis. So the idea is to determine a number of variables which separate signal events from background events, and then determine the set of cuts on these variables which maximises the significance of the signal.

The tricky part is the maximisation of the significance, which is an optimisation problem. Depending on the number of parameters, this optimisation can be accomplished a number of ways. The Toolkit for Multivariate Data Analysis (TMVA) is a software package available in the Root framework and is oftentimes used in particle physics to optimise cut selection. See the Canvas page for this course on how to run TMVA through Root on the computing clusters. A number of optimisation techniques are implemented including machine learning algorithms like boosted decision trees (BDT) and artificial neural nets (ANN). More traditional methods are also implemented, *e.g.* a modified version of gradient descent. The TMVA user guide provides a nice introduction to many of these techniques.

One of the simplest yet most robust optimisation algorithms for a single dimension is the bisection method. Given an interval of a variable x between a and b, and a function $f(x)$ that is unimodal, *e.g.* it has only a single maximum over the interval, the bisection method is as follows.

1. Define the midpoint $m = (a + b)/2$, and calculate $f(a)$, $f(b)$, and $f(m)$.
2. Define the midpoint for the left interval $[a, m]$ as $l = (a + m)/2$ and for the right interval $[m, b]$ as $r = (b + m)/2$. Calculate $f(l)$ and $f(r)$.
3. Find the maximum of $f(a)$, $f(b)$, $f(m)$, $f(l)$, and $f(r)$.
4. If the maximum is $f(a)$ or $f(l)$, redefine $b = m$.
5. If the maximum is $f(b)$ or $f(r)$, redefine $a = m$.
6. If the maximum is $f(m)$, redefine $a = l$ and $b = r$.
7. If $b - a$ is small enough or a maximum number of iterations has been reached, terminate. Otherwise, return to Step 1.



In the example to the right, the first maximum interval chosen would be $[m, b]$, and iterations would continue until approximately the endpoint b was reached as the maximum. Alternatively, if minimisation was performed, the minimum interval selected would also be $[m, b]$. The bisection method does not require the calculation of a first derivative, and consequently can be quite robust, although it may converge slowly. Modifications of the bisection method like the regula falsi or Ridder's method may converge more quickly. A similar scheme to the bisection method can be generalised to more dimensions in the simplex Nelder-Meade algorithm. Again this method does not require the calculation of derivatives. Method that require derivatives such as Newton's method in a single dimension, or gradient descent in multiple dimension, may converge more quickly but are typically not as robust.

Goal 1 Create a toy distribution of background and signal events in a single variable. For example, the background distribution might be a Gaussian with a mean of 10 and a width of 2, while the signal distribution is a Gaussian with a mean of 15 and a width of 5. Make sure this is implemented in a class so this step can be repeated.

Goal 2 Implement the bisection method and maximise the significance for the toy variable of Goal 1. Plot the function being maximised and demonstrate the validity of the method.

Goal 3 Given an expected N_s and N_b, and assuming a Poisson distribution, generate 1000 toy experiments and calculate the significance for each experiment, given the optimised cut value. Plot the histogram of this distribution.

Goal 4 Implement the Nelder-Meade method.

Goal 5 Repeat Goal 2 and Goal 3, but now allow a lower cut and an upper cut, *e.g.* maximise the significance given two variables.

Goal 6 Make a more complicated toy distribution, now in multiple variables and optimise the selection for this distribution.

Goal 7 Investigate other minimisation methods, check convergence times, *etc.*