Multivariate Analysis

Matthew Rollings

January 24, 2020

Abstract

This investigation looks into the use of significance maximisation and cutting alongside the use of the interval bisection method in order extract a signal from a toy distribution. We find that interval bisection is a robust and efficient optimisation method when used on one dimensional distributions.

1 Introduction

Multivariate analysis is a hugely popular technique, not just in the scope of physical sciences, used for analysing complicated data sets. Multivariate analytical techniques have such an advantage over other techniques purely as a result of their functionality. Allowing a researcher to look at more than one or two variables at a time; dependent (DV) or independent variables (IV), and see the relationships between those variables, is fundamental to discovery in all fields of science. Due to the difficulty in addressing complex research question with other techniques, such as univariate analysis, and because of an extensive increase in the availability of canned software able to perform multivariate analyses, these techniques have become widely used[1].

In particle physics, particularly in experiments dealing with colliding particles such as CERN, multivariate analytical techniques allow researchers to break down the data obtained in the experiment and pick out the information useful to the hypothesis they are looking to answer. One method, commonly used for slimming down the data received in large experiments like the LHC[2], is known as analytical cutting and often implemented by calculating the significance of the events. The significance of an event; the mathematical definition of which will be covered in the theory section, is defined to behave in a similar way to when a researcher might make a 'significant' discovery. This means that cuts are made to the data set in order to reduce the background intensity and boost the signal intensity relative to it, thus distinguishing whether the bump in the data is indeed a point of interest or simply a statistical fluctuation in the number of background events that occurred[3].

The point at which a researcher knows from where to calculate the significance is when the function is maximised. The method of maximising this function is when multivariate analytical techniques start to become immensely favourable. The first maximisation method to be implemented in this investigation will be the interval bisection method. This is one of the simplest yet most robust optimisation algorithms for a single dimension in which data points are selected at five equidistant intervals; the highest point determined and five more smaller, equidistant intervals are determined based around this high point. Although this process may be robust, for large data ranges, this process converges much more slowly than other optimisation algorithms.

For this reason, other methods are often preferred by large experiments with the resources and personnel to implement them. This leads to the second optimisation algorithm to be investigated in this report; the Nelder-Mead algorithm. This method is an extension of the idea introduced by Spendley *et al.*[4] for tracking the optimal operating conditions by evaluating the output from a system at a set of points forming a simplex in the factor-space. Then continually forming new simplices by reflecting one point in the hyperplane of the remaining points[5]. The method results in the simplex adapting itself to the local landscape and contracting onto the final minimum, which raises a small issue in wanting a maximisation algorithm. This is easily adapted to become a maximisation algorithm by minimising the negative of the desired function to be optimised.

2 Background Theory

2.1 Significance and Cutting

As in this investigation, we are only performing a simple counting experiment, the theory behind calculating the significance and cutting is relatively simple. Assuming the number of expected background events follows a Poisson distribution, then the fluctuation on this number of background events, N_b, will be equal to $\sqrt{N_b}$. So say the number of observed background events is 225, the fluctuation in that number is ± 15, and so an additional 10 signal events would only make this appear as an upward fluctuation in the number of background events. However, if instead the invariant mass (x value in this investigation) is cut to be required to fall into a certain range suited to the expected position of signal events, the number of background events would be reduced dramatically. Now we might have the same number of signal events, 10, but also only 10 background events, which is about 3 standard deviations away from the expected value of background events. Typically in particle physics if a standard deviation of 3 is seen, we claim to see evidence for a particle, and if a standard deviation of 5 or more is seen, we claim discovery of a new particle.

Unfortunately cutting is not always this simple and lots of cuts and adaptations to the data are required to get the data into a position to extract useful information from it. This is where significance calculation comes into play, as it maximises the significance of observing a signal with respect to the background only hypothesis. This significance metric to be maximised in this investigation is:

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

where N_s is the number of expected signal events and N_b is the number of expected background events.

2.2 Interval Bisection

This method is applicable for numerically solving $f'(x)$ for a real variable x, where f is a continuous function defined on an interval [a, b]. The bisection method works best when a function is unimodal, i.e. the function only has one maximum over the interval, otherwise there is always a possibility of the method converging to a local maximum instead of the global maximum.

The bisection method iterates through a set of instructions to determine the maximum point 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 and right intervals [a, m] and [m, b] as $l = (a + m)/2$ and $r = (m + b)/2$ respectively. Now calculate $f(l)$ and $f(r)$.
3. Find the maximum value out of $f(a)$, $f(l)$, $f(m)$, $f(r)$ and $f(b)$.

4. If the maximum is $f(a)$ or $f(l)$, redefine $b = m$.

5. If the maximum is $f(r)$ or $f(b)$, redefine $a = m$.

6. If the maximum $f(m)$, redefine $a = l$ and $b = r$.

7. If the interval [a, b] is small enough or a maximum number of iterations has been reached, terminate the iteration process and return the x-value corresponding to the maximum point. Otherwise return to step 1.

2.3 Nelder-Mead

The Nelder-Mead method approximates a local optimum of a problem with n variables when the function operating on varies smoothly and is unimodal[**wikiNM**]. Typically it is used to minimise a function $f(x)$ but this optimisation can be changed to maximising a function by minimising $-f(x)$.

In the case of this experiment, the function $-f(x)$ is being minimised, where $x \in \mathbb{R}^n$, and the current set of test points is $x_1, ..., x_{n+1}$. The method operates by iterating through a set of instructions as follows:

2.3.1 Order

1. Order the set of x points according to their values at the vertices: $-f(x_1) \leqslant -f(x_2) \leqslant ... \leqslant -f(x_{n+1})$. Check the termination criteria to see if the method should stop.

2. Calculate the centroid, $x_0 = (\sum_{i=1}^{n} x_i)/n$, of all the points except x_{n+1}.

2.3.2 Reflection

3. Compute the reflected point, $x_r = x_0 + \alpha(x_0 - x_{n+1})$, with $\alpha > 0$.

4. If the reflected point is better than second best but not better than the best, $-f(x_1) \leqslant -f(x_r) < -f(x_{n+1})$, then replace x_{n+1} with x_r and go to step 1.

2.3.3 Expansion

5. If the reflected point is the best so far, $-f(x_r) < -f(x_1)$, then compute the expanded point, $x_e = x_0 + \gamma(x_r - x_0)$, with $\gamma > 1$.

6. If the expanded point is better than the reflected point, $-f(x_e) < -f(x_r)$, then replace x_{n+1} with x_e and go to step 1.

7. Otherwise replace x_{n+1} with x_r and go to step 1.

2.3.4 Contraction

8. Now it is certain that $-f(x_r) \geqslant -f(x_n)$. Compute the contracted point, $x_c = x_0 + \rho(x_{n+1} - x_0)$, with $0 < \rho \leqslant 0.5$.

9. If the contracted point is better than the worst point, $-f(x_c) < -f(x_{n+1}$, then replace x_{n+1} with x_c and go to step 1.

2.3.5 Shrink

10. Replace all the points except x_1 by computing $x_i = x_1 + \sigma(x_i - x_1)$ and go to step 1.

2.3.6 Termination

Some termination criteria are required to break the iterative cycle. Nelder and Mead chose to use the standard deviation of the sample function values of the current simplex, so the process was terminated when this value dropped below a predetermined value ϵ. The calculation for the standard deviation is:

$$\sigma^2 = \sum_{i=1}^{n+1} (f_i - \bar{f})^2/(n+1) \tag{2}$$

where $\bar{f} = \sum f_i/(n+1)$. In the case where $\sigma < \epsilon$ all function values are very close together, and so hopefully are the points, near the maximum x_1.

2.3.7 Coefficients

α, γ, ρ and σ are the reflection, expansion, contraction and shrink coefficients, respectively. Standard values of these constants were used in the investigation, which are: $\alpha = 1$, $\gamma = 2$, $\rho = \frac{1}{2}$ and $\sigma = \frac{1}{2}$.

3 Algorithm

3.1 Distribution

At the beginning of the 'Project.py' code a class is implemented for the creation of the toy distribution of background and signal events in a single variable. The background distribution is created to be a Gaussian distribution with a mean of 10 and a standard deviation of 6, while the signal distribution is created to be a Gaussian distribution of mean 15 and standard deviation 5. To create the Gaussian distributions 'random.gauss' is used. When the 'sample' function in the class is called, this toy distribution is created and saved in the form of a dictionary where the keys are the x values (0 through 25) and the values corresponding to the keys are the number of events, of that x value, that occur. This can be visualised by the dictionary being plotted on a scatter graph when the 'plot' function is called.

3.2 Cutting

The 'Optimisation' function represents the calculation of the significance values to be used and referenced by the interval bisection and Nelder-Mead methods. It operates by iterating through all possible low bound cuts in the x values between the minimum x value in the signal distribution and the maximum x value in the background distribution as this is the interval in which the maximum significance will occur. Within this iteration the total number of background counts, N_b, and the total number of signal counts, N_s, are recorded and plugged into equation 1 to find the value of significance at the low bound value. Each of these significance values is then appended to a list to be used in the optimisation method of interval bisection.

The 'Optimisation' function also allows for the possibility of a single x value, 'low_bound', being input for which the significance is found for that x value being the low bound. This was implemented only in trying to implement the Nelder-Mead method.

3.3 Interval Bisection

The 'Bisection_Iteration' function gives the option of selecting the number of iterations for it to run for, done by providing the function with an argument, or for the iteration process to stop when the $[a, b]$ interval is sufficiently small. When run the function iterates through the steps outlined in the theory section above by calling on the function 'Single_Bisection' which calculates m, l and r from the a and b values passed to it as arguments and then adjusts the plotting of the function accordingly. The returned m, l and r values are then used alongside the a and b values to find their corresponding significance value and decide which method to take of either step 4, 5, 6 or 7 explained in the theory.

3.4 Optimised Significance

The 'Opimised_significance_plot' function operates based on the number of toy experiments the user wishes to iterate over and changes the histogram output dependent on the number of bins input as an argument. This function finds the optimised low bound returned from calling the 'Bisection_Iteration' function and uses it to calculate the significance value from each of the experiments it iterates through. To start, it uses the optimum low cut value to create new dictionaries for each of those created in the 'Distribution' class with x values only above the optimum low bound and with each point having been Poisson distributed (using 'numpy.random.poisson' to calculate the Poisson fluctuations). These new distributions are then used to generate significance values, for each of the experiments, with low bounds being at the optimum value to maximise significance. These significance values are appended to a list and then plotted on a histogram.

3.5 Nelder-Mead

The 'Nelder_Mead' function starts by passing x values of integers 5 through 25 and their corresponding significance values through the Nelder-Mead method described in the theory section. However, some trouble occurred in getting the algorithm to behave with the desired effect. I feel with another week or so I may have been able to implement the method correctly but in the time frame allowed and other lab report deadlines coinciding with this one, it was not able to be done.

4 Results and Analysis

As expected the toy distribution for the total events was correspondent with that of the superposition of the two individual Gaussian signal and background distributions, as can be seen in Figure 1. The interval bisection method optimised the significance values perfectly, correctly obtaining

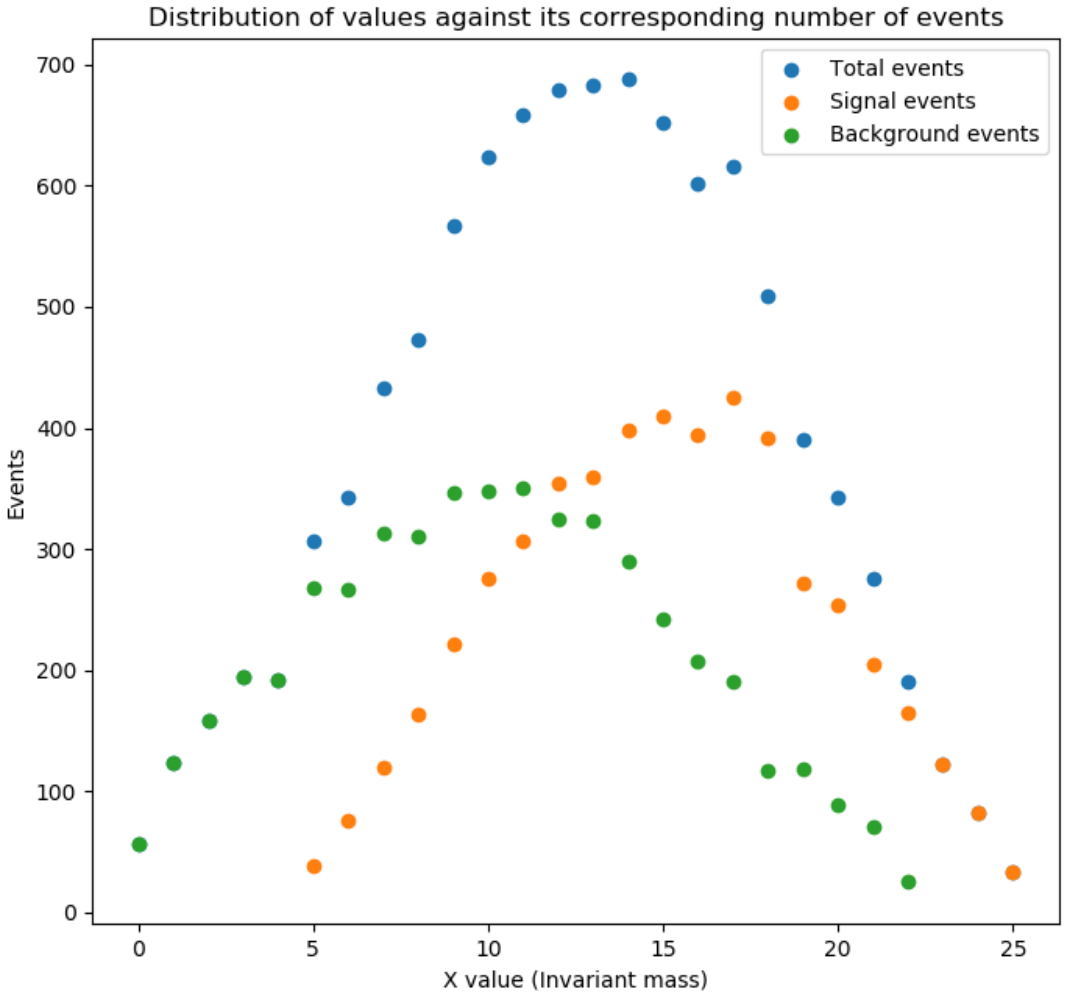

Figure 1: Toy distribution for signal, background and total events.

the maximum value of significance and the corresponding x value every time. Figure 2 shows the

method in progress after 4 iterations, about to obtain the x value, for the maximum significance value, of 12. This value is also the point at which the background distribution and signal distribution meet, which makes sense as long as the number of events in the background distribution is decreasing and signal distribution is increasing, with each increase in x value, at the point of crossover. Figure 2 does highlight the efficiency of the interval bisection method, having almost already reached the optimum value after only 4 iterations, so this method is ideal for optimisation of simple 1 dimensional distributions and very robust as an analytical technique.

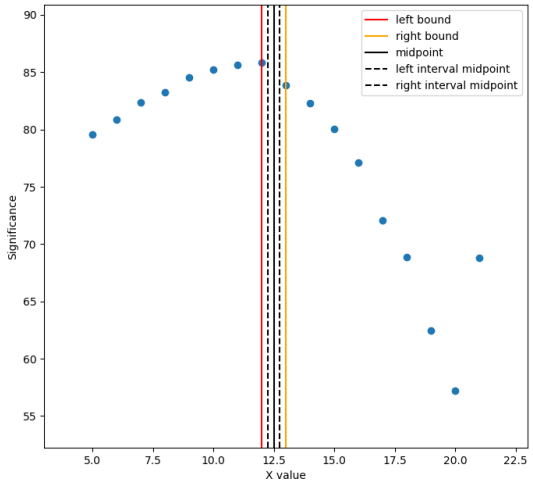

Figure 2: Interval Bisection method in progress (after 4 iterations).

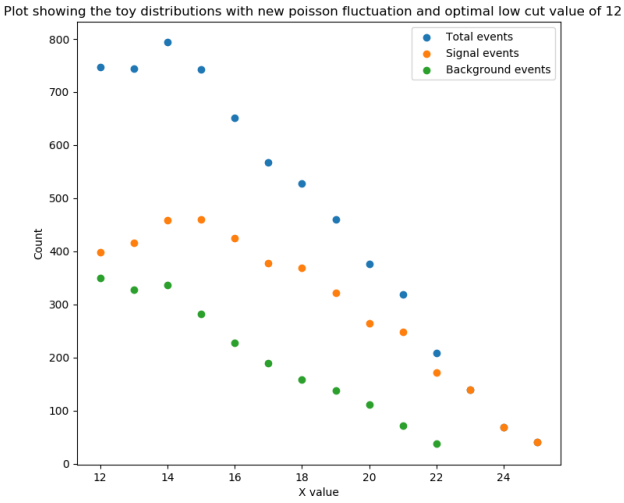

Figure 3: Poisson fluctuated toy distribution with low bound optimum cut off value of 12.

The Poisson fluctuation of the toy distributions alongside implementing the low bound optimum cut off value found through the interval bisection, shown in Figure 2, produced the distributions shown in Figure 3.

The significance plot in Figure 4 shows the significance follows a Gaussian distribution around a mean value of around 89.4. This is clearly a very high value for significance compared to usual particle experiments, with it commonly being more of order 10^{-6}[3]. However, considering it is known that there is a clear signal in the distribution to begin with, the high significance value does not seem out of place.

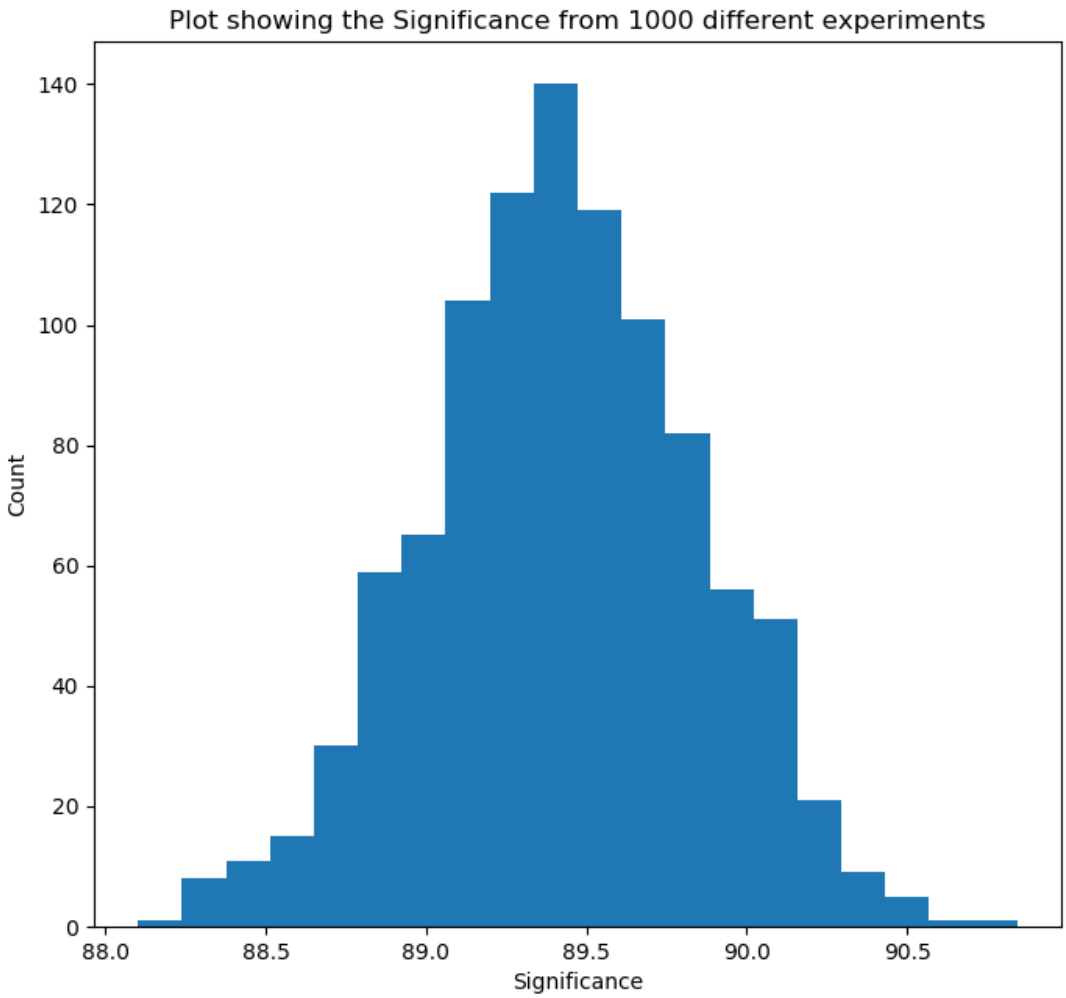

Figure 4: Plot of the significance found from 1000 different experiments, using the optimum low cut value: 12.

5 Conclusion

This experiment was successful in demonstrating the benefits of using multivariate analytical techniques to obtain data more tailored to the hypotheses being investigated. The use of one of the simplest analytical techniques alone was enough to extract meaningful data from a toy distribution. The Nelder-Mead method would have demonstrated that this ability of analytical techniques was not confined to only 1 dimensional distributions of data but multiple dimensional distributions; making them optimal for use in the most complex problems in all ranges of applications.

It is clear to see from this introduction to multivariate analysis that these techniques are crucial in a particle physics setting. The ability to filter information received from a detector allows researchers to spend time focused on the data important to the hypothesis being tested, rather than filtering through all of the data obtained themselves.

6 Acknowledgements

I would like to thank Dr. P. Ilton; Dr. F. Gonella; Dr. N. Lurkin; Dr. E. Reynolds; Mr. J. Plews; Mr. R. Valence and Mr. H. Cooke for their guidance throughout this experiment and their patience when answering questions. I would also like to thank Miss A. Rollings for her patience in helping understand the particle physics concepts used in this investigation.

References

[1] Laura L. Myers Karen A. Randolph. *Basic statistics in multivariate analysis*. eng. Pocket guide to social work research methods. 2013.

[2] *The Large Hadron Collider*. URL: https://home.cern/science/accelerators/large-hadron-collider.

[3] Pekka K. Sinervo. "Signal Significance in Particle Physics". In: *High Energy Physics - Experiment* (August 2002).

[4] G. R. Spendley W.; Hext and F. R. Himsworth. "Sequential Application of Simplex Designs in Optimisation and Evolutionary Operation". In: *Technometrics* (1962).

[5] John Nelder and Roger Mead. "A simplex method for function minimization". In: *Computer Journal* (1965).