Likelihood Exercise

1. Generate a data set consisting of 100 random numbers between 0 and 10 (representing a uniform background over some arbitrary energy range) and 15 “signal events,” centred on the value 6 and characterised by a Gaussian distribution with a standard deviation of 1 (representing the energy resolution).

2. Construct a likelihood function for a mixed signal and background hypothesis and then write a routine to maximise the likelihood (or minimise -2ln(likelihood)) to find the best overall position of the signal and the number of signal events. Using an Asimov data set, construct a contour plot of -2ln(maximum likelihood) as a function of the two fit parameters. Also plot the profile likelihood for each parameter separately.

3. Make the same plots for several fluctuated data sets and compare.

4. Estimate the uncertainty in each parameter based on Wilks’ Theorem using both the Asimov and fluctuated data sets above. Verify this by generating 1000 fluctuated data sets and maximising the likelihoods to find how often the fit values fall within 1σ and 2σ of the true values for each parameter.

5. Separately make a scatter plot of fit signal position vs fit number of signal events along with the 1σ and 2σ contours based on Wilks’ Theorem, now assuming 2 degrees of freedom. Verify the fraction of events within each contour.

6. Draw 1σ and 2σ Bayesian contours on the scatter plot of step 4, assuming priors that are uniform in position and event rate.

7. **(Bonus question)** Repeat the generation and fitting in step 3 for true signal values of 5, 10, 20, 30 and plot the average significance of signal detection, in terms of standard deviations based on Wilks’ Theorem, as a function of the signal strength. Similarly, repeat with the number of signal fixed to 15 again, but with the number of backgrounds taken as 50, 200, 400, 1000. Again, plot the average significance as a function background number.