Final Project Pre-proposal

DSC 291 Statistical Analysis: Physics

Nate Sheibley (A69046267), Jet Yue (A17406581),
Tristan Cooper (A15824397), Kevin Lin (A69043483)

Summary

From 2 calibration detectors, an unknown detector, and the expected Gaussian model of Neutrinoless Double-Beta Decay (NLDBD) of particle energies, we will first compute their respective probability distribution functions (PDF). Using these 4 PDF's, we will calculate their combined fitting parameters with respect to a target detector's PDF, as well as compute the overall impact of the NLDBD distribution and its sensitivity to the target.

Methodology Plan

1. **Convert CSVs into energy spectra**
   a. Read the respective data files: detectorA.csv, detectorB.csv, detectorC.csv, and DetectorTarget.csv
   b. Use np.histogram to return count and bin_edges (Sturges's rule / Auto / Manual testing for optimal bin numbers)
      i. Use np.histogram to search for optimal bin count
   c. Plot all energy spectra of all 4 detectors
      i. Use seaborn.kdeplot for energy spectra
2. **Use the 1592 keV peak in Detector A to set a cutting threshold and evaluate ◎true positive rate.**
   a. Subtract baseline (simple linear baseline + gaussian estimation using minuit)
   b. Use classification scores provided + PDF/histogram to generate an ROC curve to establish a cut threshold
      i. `"Across all detectors, the 1592 keV peak contains mostly signal events"` - we will assume this dataset is a set of all positives, true and false
   c. Evaluate to cut threshold value that maximizes TPR
3. **Use the 2103 keV peak in Detector B to estimate ◎false positive rate after applying cut.**
   a. Subtract baseline (simple linear baseline + gaussian estimation using minuit)
   b. Use classification scores provided + PDF/histogram to generate an ROC curve
      i. `"the 2103 keV peak contains mostly background events"` - we will assume this dataset is a set of all negatives, true and false
   c. Apply cut threshold from DetectorA
   d. Backward propagate and iterate on the cut threshold to minimize the FPR of detectorB while maximizing the TPR of detectorA
4. **NLDBD is expected to be a Gaussian peak of Signal Events at 2039 keV with width ± 1 keV. Generate the PDF for NLDBD.**

    a. For Detector C, model gaussian and baseline similar to HW3 for an arbitrary signal region
    b. Do this for a 12 keV region (± 1 keV times 6 sigma) around expectation energy level to avoid selection bias. 6 sigma is an arbitrary number, we will tweak this as necessary.

5. **Apply cut, then generate PDF of detectors A, B, and C and DetectorTarget histogram.**
    a. Generate energy histograms + pdfs after cuts to use for final fit
    b. Create combined pdf model for fitting to detectorTarget
6. **Using either the Frequentist or Bayesian method, fit all generated PDFs from steps 4-5 to the DetectorTarget histogram to obtain ◎θA, ◎θB, ◎θC, and ⚛θNLDBD.**
    a. Prefer Bayesian fit, but if time allots, we will do both for numerical rigor
    b. Fit pdfs to calculate theta vector (minuit fine here)