Final Project Description

Aobo Li*

(Dated: November 6, 2025)

I. PROJECT STRUCTURE AND EVALUATION

Students must complete this project in groups of 2-4 members. Any deviation from this group size requirement requires explicit instructor approval. The project is structured in two stages with specific deliverables and evaluation criteria.

For the first stage, students must submit a pre-proposal document of 1-2 pages by **Nov. 16th** on Gradescope. For the pre-proposal, students should write a document of no more than 2 pages. For Option A, this document should include information up to Step 6 of the fitting procedure. For Option B, it should include information before the 90% Confidence Level (See Section II and III for details).

The second stage is due on **December 12th** and accounts for 50% of the final grade. It includes three components: a 10-page final report containing the theoretical framework, technical details, and all results outlined in this document (30%); a final presentation (15%); and a GitHub repository with the complete project codebase (5%). In total, this project constitutes 60% of the course grade.

II. OPTION A: MJD DATA RELEASE ANALYSIS

Option A requires students to work with the MJD data release as described in the paper available at https://arxiv.org/abs/2308.10856. The dataset consists of four CSV files:

- detectorA.csv
- detectorB.csv
- detectorC.csv
- detectorTarget.csv

Each file contains multiple event records with the structure shown in Table I. The `Event ID` is a unique integer identifier assigned to each event for tracking purposes. The `Classification Score` is a floating-point value between 0 and 1, produced by a neural network, where values closer to 1 indicate more signal-like events and values closer to 0 indicate more background-like events. The `Energy` column represents the measured energy of each event in keV[1].

* aol002@ucsd.edu

TABLE I. CSV Structure of Detector Data Files

Event ID	**Classification Score**	**Energy**
101	0.873	1423.5
102	0.342	987.2
103	0.921	1632.8
⋮	⋮	⋮

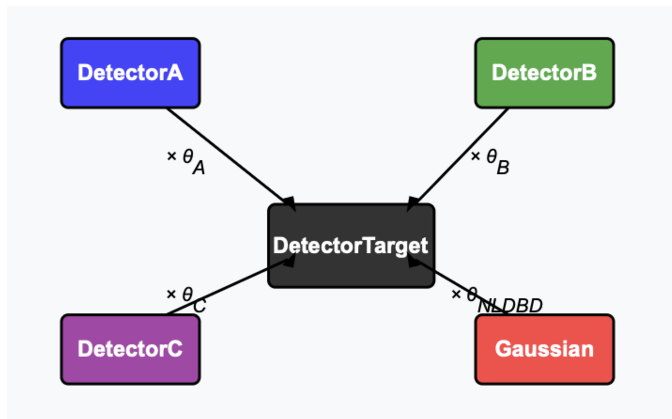

FIG. 1. A Flow Chart for the Fitting Project with MJD Dataset

Detectors A and B are calibration detectors with independently measured fitting parameters $\theta_A = 1350 \pm 100$ and $\theta_B = 770 \pm 270$, respectively. Detector C has an unknown parameter θ_C that must be determined through fitting. The DetectorTarget file contains a linear combination of spectrums from Detectors A, B, and C and serves as the target of fitting. Across all detectors, the 1592 keV peak contains mostly signal events and the 2103 keV peak contains mostly background events.

The analysis project should follow this procedure:

1. Convert the .csv files into energy spectrum.
2. Use the 1592 keV peak in Detector A to set a cutting threshold and evaluate true positive rate.
3. Use the 2103 keV peak in Detector B to estimate false positive rate after applying cut.
4. Neutrinoless Double-Beta Decay (NLDBD) is expected to be a Gaussian peak of **Signal Events** at 2039 keV with width ± 1 keV. Generate the PDF for NLDBD.
5. Apply cut, then generate PDF of detectors A, B, and C and DetectorTarget histogram.

6. Using either the Frequentist or Bayesian method, fit all generated PDFs from steps 4-5 to the DetectorTarget histogram to obtain ◎θ_A, ◎θ_B, ◎θ_C, and ⚙θ_{NLDBD}.

7. Calculate the upper limit of ⚙θ_{NLDBD} at 90% confidence level.

8. Calculate the experimental sensitivity of ⚙ θ_{NLDBD}.

Icon ◎represents the metric you should obtain to support the major fitting result, and ⚙represents the major fitting result. All ⚙s must appear in both the final report and the presentation, while all ◎must appear within the final report with clear explanation of how they are obtained and how they contribute to validating your primary conclusions. The pre-proposal should include a rough discussion about how to obtain each ◎s and ⚙s.

Not that an integer N amount of NLDBD events are injected into the DetectorTarget dataset as ground truth. Two awards will be given based on exceptional final project performance:

- **Best Result Award:** Awarded to the group whose analysis yields results closest to N.

- **Best Analysis Techniques Award:** Awarded to the group presenting the most rigorous and comprehensive physics analysis, incorporating techniques from within the course scope or advanced methods beyond the scope of this class.

III. OPTION B: CUSTOM DATASET ANALYSIS

Students have the option to conduct fitting analyses using their own datasets, provided they follow a structured approval process. To pursue this path, students must first submit a detailed preproposal that thoroughly explains their dataset and its relevance to the course objectives. The preproposal should include the data's origin, structure, and scientific significance. Only after receiving explicit instructor approval may students proceed with analyzing their chosen dataset.

The complexity and scope of this self-directed option must be comparable to the standard assignment (Option A). Simple curve-fitting exercises will not meet the course requirements. Students must explicitly define their fitting parameters and demonstrate how these parameters connect to specific scientific questions they aim to address through their analysis.

The analytical approach must align with the statistical methods covered in the lecture series, utilizing either Frequentist or Bayesian fitting methodologies. The final report must present a complete mathematical formulation of the fitting procedure, including all relevant equations, assumptions, and statistical considerations. As part of the analysis, students are required to calculate and interpret the 90% confidence intervals for their fitted parameters, demonstrating their understanding of statistical uncertainty in scientific analysis.

---

[1] KeV is the unit of energy.