Modeling Beta Decay Spectra to Analyze the Sensitivity of a Neutrino Mass Experiment

by

Talia E. Weiss

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Bachelor of Science in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

© Massachusetts Institute of Technology 2018. All rights reserved.

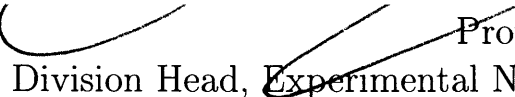Author . . . . **Signature redacted** . . . .

Department of Physics
May 11, 2018

Certified by . . **Signature redacted** . . . .

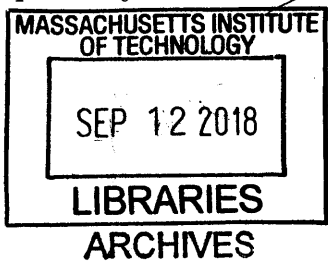Professor Joseph A. Formaggio
Division Head, Experimental Nuclear and Particle Physics
Thesis Supervisor

Accepted by . . . . **Signature redacted** . . . .

Professor Scott Hughes
Associate Head, Department of Physics

Modeling Beta Decay Spectra to Analyze the Sensitivity of a Neutrino Mass Experiment

by

Talia E. Weiss

Abstract

The Project 8 experiment aims to measure the electron neutrino mass by obtaining and analyzing β spectra from tritium decay. Using an inferential model of the experiment's anticipated data, I evaluate its projected sensitivity to certain parameters of interest. I focus on the precision and accuracy with which Project 8 can expect to resolve the β-decay spectrum's endpoint in an upcoming stage of the experiment. I also present an initial prediction of Project 8's eventual expected sensitivity to the electron neutrino mass. This analysis involved generating and analyzing β-decay spectral data using a model implemented in *Stan*, a platform for Bayesian statistical inference. The sensitivity analysis was designed to account for the anticipated distribution of results (mass and endpoint measurements) produced by the potential variation in a number of physical and experimental parameters. In addition, the method used here allows for a calibration of the consequences of inferences and decisions made in reaching those results. I find that, using one year of Project 8 Phase II data, the T_2 endpoint can be resolved within a 13.7 eV window (90% C.I.) with 62% coverage (or accuracy), corresponding to a 4.1 eV posterior standard deviation. Preliminarily, using one year of Phase IV data, the electron neutrino mass can be resolved within a 0.051 eV window (90% C.I.) with 56% coverage. I also outline a way that model-based sensitivity procedures and calibration of inference can be extended to the neutrino mass hierarchy problem.

Thesis Supervisor: Professor Joseph A. Formaggio
Title: Division Head, Experimental Nuclear and Particle Physics

Acknowledgments

I would first like to acknowledge my thesis advisor Joseph Formaggio, who has not only provided me with extensive and valuable mentorship as I worked on this thesis, but also played an integral role in making the past three and a half years of my time at MIT enriching and enjoyable. I also want to thank Michael Betancourt, who generously devoted much time to helping me understand the statistics concepts, modeling techniques and programming methods that underlie this project. Thanks very much to Mathieu Guigue and Joseph Johnston for working with me on both software development and β-spectrum modeling, and for assisting me any with challenges that arose on either of those fronts. Finally, thank you to Valerian Sibille and Evan Zayas for always being willing to answer my questions and engage in interesting discussion.

Contents

List of Figures

List of Tables

Chapter 1

Introduction: The Neutrino Mass Problem and Project 8

The Project 8 Neutrino Mass Experiment aims to determine both the magnitude of the electron neutrino mass and the ordering of the three neutrino mass eigenvalues [1, 2]. Project 8 is expected to yield a *direct* mass measurement, meaning that experimenters will derive information regarding the neutrino mass from the shape of the electron spectrum produced when a particle β-decays [3]. As Project 8 works to obtain a spectrum that can be used to resolve the neutrino mass, it is valuable to evaluate the experiment's expected sensitivity to certain parameters of interest—that is, the precision and accuracy with which those parameters can be determined.

Here, I assess Project 8's sensitivity to the endpoint, or energy at which the β-decay spectral rate vanishes. The Project 8 Collaboration must measure the endpoint at an intermediate stage to develop a fuller understanding of the experiment's setup and performance [2]. I also assess experimental sensitivity to the electron neutrino mass, itself. I perform this analysis by generating spectra resembling Project 8's anticipated data sets, then analyzing that data to extract parameter distributions via Bayesian inference. Such a sensitivity study enables the Project 8 Collaboration to make predictions regarding the effect of presently unknown physical and experimental parameters (including runtime) on the Collaboration's ability to resolve the endpoint and neutrino mass.

In this chapter, I provide background regarding the neutrino mass scale and ordering, as well as Project 8's approach to addressing the neutrino mass problem.

1.1   The Neutrino Mass Problem

A neutrino is produced in one of three flavor states: electron, muon, and tau, denoted ν_e, ν_μ and ν_τ, respectively. The discovery of neutrino oscillations demonstrated that each of these flavor states is a superposition of mass eigenstates ν_1, ν_2 and ν_3; this superposition allows neutrinos to change flavors by evolving in time. The neutrino oscillation rate is related to the differences between the mass eigenvalues m_1, m_2 and m_3, so at least two of these masses must be non-zero for neutrinos to oscillate.
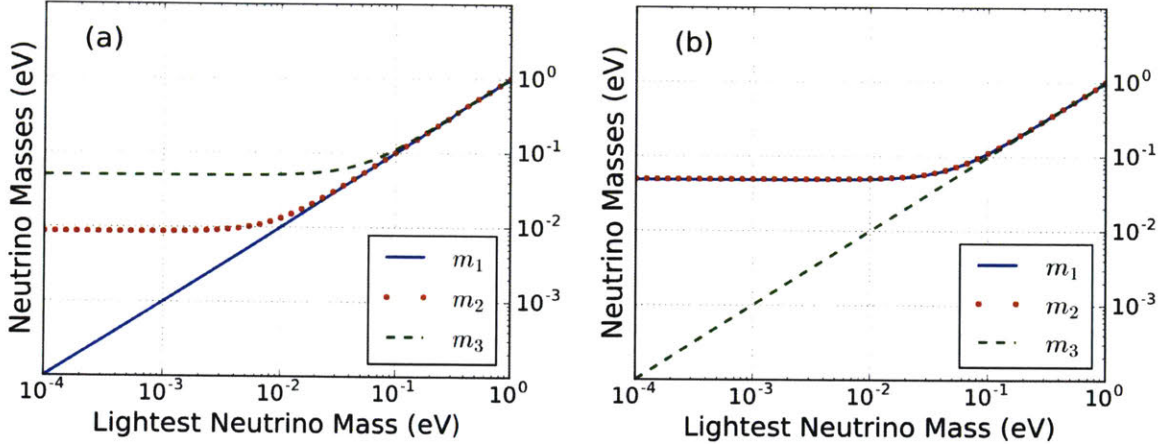
Figure 1-1: Neutrino masses corresponding to normal and inverted orderings
Three ν masses m_i according to (a) normal and (b) inverted hierarchies as functions of the lightest mass. Points computed using mass splitting values obtained from [8].

Oscillations have now been observed by solar [4, 5], atmospheric [6] and reactor [7] neutrino experiments, confirming that neutrinos are massive.

The relationship between the flavor eigenstates ν_α and mass eigenstates ν_i of the neutrino is given by

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \cdot \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix},$$

where U_{ij} are the elements of the unitary Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix. Using this mixing matrix, we can express the approximate electron neutrino mass m_β as a function of the mass eigenvalues:

$$m_\beta = \sqrt{\sum_i |U_{ei}|^2 m_i^2}.$$

This formula applies when the differences between the mass eigenvalues are much smaller than the overall energy scale being considered.

Neutrino physicists formulate the mass and mixing problem in terms of two sets of parameters that have now been precisely measured: mixing angles and mass splittings. In terms of mixing angles θ_{12} and θ_{13}, the PMNS elements are defined as

$$\begin{aligned} |U_{e1}|^2 &= \cos^2(\theta_{12})\cos^2(\theta_{13}), \\ |U_{e2}|^2 &= \sin^2(\theta_{12})\cos^2(\theta_{13}), \\ |U_{e3}|^2 &= \sin^2(\theta_{13}). \end{aligned} \qquad (1.1)$$

Neutrino flavor oscillation frequencies depend on mass splittings $\Delta m_{ij}^2 = m_i^2 - m_j^2$.

10
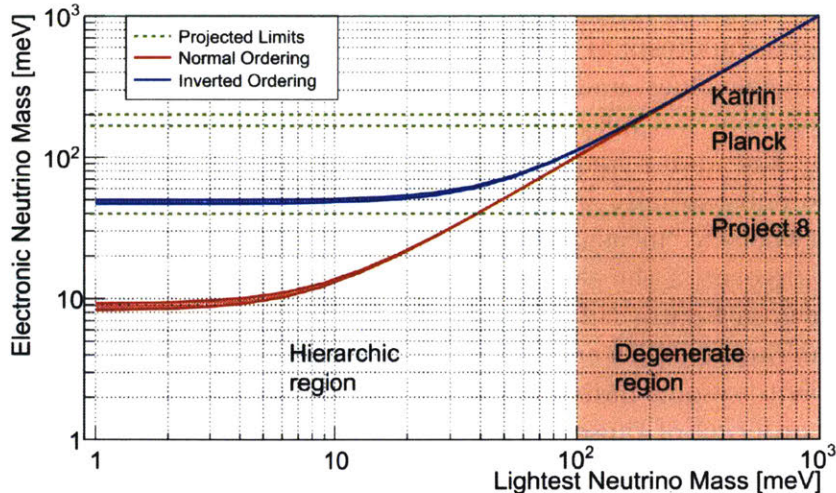
Figure 1-2: Allowed electron neutrino mass values, as a function of lightest mass
The allowed values of m_β are constrained by measured mixing parameters. If the lightest mass falls within the hierarchic region, then a precise m_β measurement can be used to resolve the mass hierarchy. In the degenerate region, it is more difficult to distinguish between the two hierarchies. The green dashed lines represent projected 90% confidence limits from the KATRIN and Project 8 direct neutrino mass experiments, and well as the Planck spacecraft. Figure from [2].

While the smaller mass splitting, Δm_{21}^2, is known from solar oscillation experiments, only the magnitude of the larger splitting, $|\Delta m_{31}^2| \approx |\Delta m_{32}^2|$, is known from atmospheric experiments [8]. Thus, two orderings of the mass spectrum are possible: if $\Delta m_{31}^2 > 0$, then $m_1 < m_2 < m_3$ and the masses obey a "normal hierarchy"; alternatively, if $\Delta m_{31}^2 < 0$, then $m_3 < m_1 < m_2$ and the masses follow an "inverted hierarchy." So the *neutrino mass problem* involves two unknowns: the overall scale of the mass eigenvalues and their ordering. Table 1.1 shows physicists' knowledge of the mixing angles and mass splittings, to date.

Parameter	Best fit	3σ
$\Delta m_{21}^2 \ [10^{-5} \ \mathrm{eV}^2]$	7.37	$6.93 - 7.97$
$|\Delta m^2| \ [10^{-3} \ \mathrm{eV}^2]$	2.50 (2.46)	$2.37 - 2.63 \ (2.33 - 2.60)$
$\sin^2 \theta_{12}$	0.297	$0.250 - 0.354$
$\sin^2 \theta_{23}$	0.437 (0.569)	$0.379 - 0.616 \ (0.383 - 0.637)$
$\sin^2 \theta_{13}$	0.0214 (0.0218)	$0.0185 - 0.0246 \ (0.0186 - 0.0248)$

Table 1.1: Current knowledge of mixing parameters from the Particle Data Group
Best fit values and 3σ confidence intervals from a global fit of oscillation data, as reported in [8]. Parameter values (values in parentheses) are reported assuming a normal (inverted) mass ordering. $\Delta m^2 \equiv m_3^2 - (m_2^2 + m_1^2)$ is the large mass splitting.

Neutrinos are currently the only elementary particles in the standard model with unknown masses, but tests of cosmological models place an upper limit on the sum on their mass eigenvalues, and nuclear physics experiments place an analogous limit on the electron neutrino mass [3]. Currently, the upper bound on m_β is 2 eV, as determined by the Mainz and Troitsk experiments [9, 10]. Project 8's major goal is not only to substantially improve the nuclear physics-based bound, but also to measure m_β (where a "measurement" can be expressed as a pair of bounds, upper and lower). Moreover, if the electron neutrino mass is sufficiently small (below about 0.1 eV, as shown in Figure 1-2), Project 8's direct mass measurement could rule out the inverted hierarchy. With enough precision, by analyzing β-decay spectra, it might also be possible to resolve individual neutrino masses, and thereby to determine both the scale and ordering of those masses.

The neutrino mass problem has implications for cosmology, because the neutrino mass contribution to the matter density of the universe substantially affects late-times large structure formation [3]. In addition, knowledge of the neutrino mass hierarchy would enable better predictions of the rate of neutrino-less double beta ($0\nu\beta\beta$) decay [11]. Experiments seeking to detect that decay process could potentially provide an explanation for the universe's matter-antimatter asymmetry.

1.2 The Project 8 Experiment: A Frequency-Based Approach

Project 8 aims to solve the neutrino mass problem by analyzing the electron energy spectrum produced by the the nuclear beta decay of tritium (T) into ^{3}He, an electron and electron antineutrino. Project 8's unique addition to past and ongoing direct mass experiments is its use of cyclotron radiation emission spectroscopy (CRES). CRES involves measuring the cyclotron frequencies of electrons in a magnetic field to determine their energies to high precision.

1.2.1 Motivation for a Direct Neutrino Mass Measurement

Direct mass measurements exploit radioactive decay kinematics to determine the neutrino mass scale. The differential decay rates of both β-decay and electron capture depend on neutrino mass values, so by examining either of these rates, an experimenter can determine the mass scale [3]. For example, in the case of β-decay (the process of interest to Project 8), we expect the size of m_β to correspond to the size of a shift in the decay spectrum's endpoint energy. Furthermore, as I will show in Chapter 3, a direct mass measurement combined with external data from reactor neutrino experiments can yield an independent determination of the mass hierarchy (no additional neutrino mass or mixing information is needed) [12]. This is possible because we model and analyze the full spectral shape, instead of only considering the endpoint location (see Figure 1-3).

Direct mass measurements are so called because the approach of studying decay features imposed by energy conservation is relatively model-independent. Unlike other
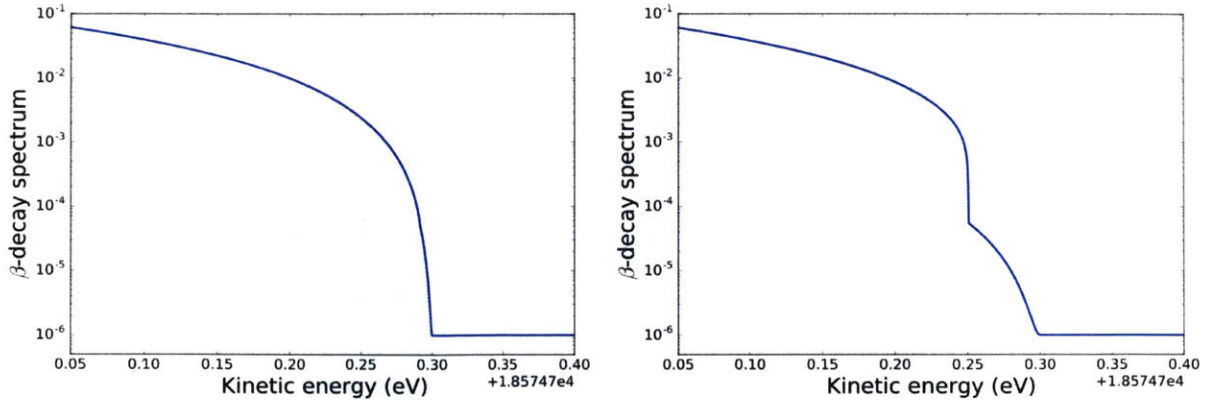
Figure 1-3: Models of beta decay spectra near the endpoint

Spectra are modeled assuming that the neutrinos are ordered according to the normal (*left*) and inverted (*right*) hierarchies. In both cases, the lightest neutrino mass is zero, $Q = 18575$ eV, mean mixing parameters come from [8], and a flat background is included. With sufficient energy resolution, it should be possible to determine whether data is better described by a "normal" or "inverted" spectral model.

techniques, a decay kinematics-based method does not depend on cosmological models or whether neutrinos are their own antiparticle. Direct measurement is therefore a particularly promising approach to resolving the neutrino mass scale. Currently, the Karlsruhe Tritium Neutrino (KATRIN) experiment is working to employ MAC-E (Magnetic Adiabatic Collimation with Electrostatic) Filtering to obtain β-decay spectra and probe neutrino mass limits as low as 0.2 eV [13]. Project 8 aims to reach sensitivities as low as $m_\beta \lesssim 40$ meV [2].

1.2.2 Cyclotron Radiation Emission Spectroscopy

The Project 8 Collaboration developed the technique of cyclotron radiation emission spectroscopy (CRES) for obtaining a β-decay spectrum, as originally proposed by Monreal and Formaggio [14]. CRES involves confining an isotope gas in a relatively uniform magnetic field generated by a solenoid-like magnetic trap. The field induces electrons produced by the isotope decay to travel in a spiral path with cyclotron frequency f_c, and to emit coherent radiation at the same frequency. For a relativistic electron, f_c depends on electron kinetic energy K and magnetic field strength B:

$$f_c = \frac{1}{2\pi} \cdot \frac{eB}{m + K/c^2},$$

where e is the electron charge, m is the electron mass, and c is the speed of light in vacuum. By measuring the total power radiated by each electron and computing f_c using the Larmor formula, it is possible to determine electron kinetic energies and construct a β-decay spectrum [2, 14].

CRES is a particularly valuable technique because it evades the need to transport electrons out of a source [14]. As a result of that necessity, the size of a MAC-E

13
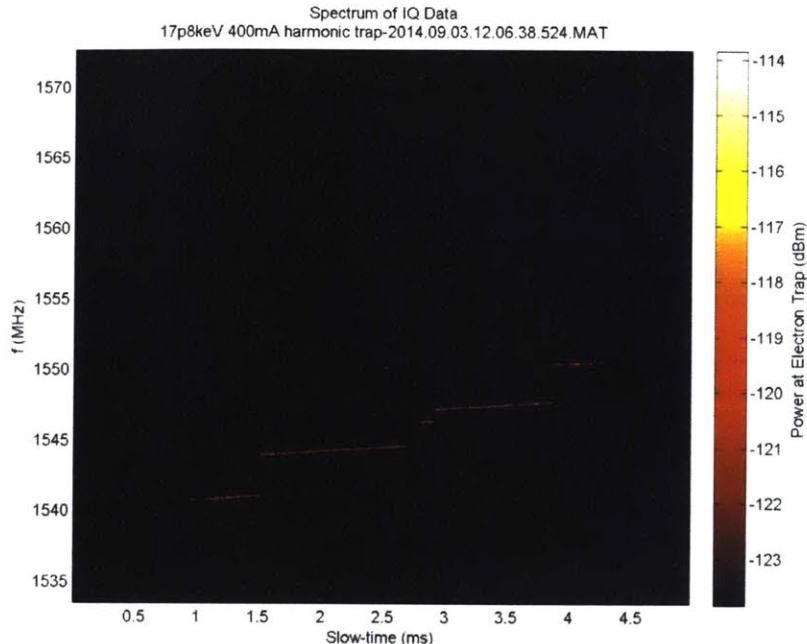
Figure 1-4: An electron spectogram from demonstrating the success of CRES From Project 8 Phase I. The frequency jumps upward repeatedly as the electron scatters with hydrogen, losing energy [2].

spectrometer (like KATRIN's) would have to scale up dramatically in order to further improve sensitivity to the neutrino mass. The Project 8 Collaboration successfully implemented CRES for the decay of gaseous ^{83m}Kr during Phase I of its experiment [2]. Figure 1-4 shows the cyclotron frequency evolution for a single krypton CRES event.

1.2.3 Extracting Neutrino Masses from Tritium Beta Decay Spectra

The β-decay spectrum at high energies can be written[1]

$$\mathcal{P}(K) \equiv \frac{\mathrm{d}N}{\mathrm{d}K} \simeq A \sum_i \left[|U_{ei}|^2 (Q - K) \cdot \sqrt{(Q-K)^2 - m_i^2} \right] \cdot \Theta(Q - K - m_i), \quad (1.2)$$

where N is the number of electrons as a function of energy K, A is a constant that depends on source activity, U_{ei} are the mixing matrix elements given by Eq. 1.2, and Q is the endpoint. This endpoint is defined as the largest possible electron energy for the case in which all three neutrino masses are zero. A particular decay's endpoint can be calculated given energy conservation constraints [15].

Direct neutrino mass experiments have not yet reached sensitivities that would

---

[1]The exact form of $\mathcal{P}(K)$ depends on $p_e E_e$, where p_e and E_e are the electron momentum and energy, respectively. However, near the endpoint, $p_e E_e$ is essentially constant.

allow them to detect the each neutrino mass' individual contribution to the spectrum. Past experiments have therefore worked to extract the effective electron neutrino mass m_β using a simplified spectral model:

$$\mathcal{P}(K) = A(Q - K) \cdot \sqrt{(Q - K)^2 - m_\beta^2} \cdot \Theta(Q - K - m_\beta), \qquad (1.3)$$

In this case, the size of the shift of the spectrum's highest energy from Q can be used to determine m_β.

However, Project 8 aims to probe energy scales at which the full spectral shape becomes relevant. This shape is a composite of three terms with different maximum energies, as determined by the Θ functions in Eq. 1.2. As a result, the region in the last ≈ 0.1 eV of the spectrum includes "kinks" at three energies, corresponding to the three neutrino mass eigenvalues. The location of the more prominent kink differs between the two neutrino mass orderings, as shown in Figure 1-3. Thus, by modeling and analyzing the shape of a spectrum near the endpoint, it should be possible to distinguish between the normal and inverted hierarchies.

Before the Project 8 Collaboration can extract neutrino masses from tritium β-decay spectra, it faces a few major challenges. Only 2×10^{-13} of the spectrum's events occur in the last 1 eV of the spectrum, so a large volume of tritium gas is required to obtain sufficient data [2]. Furthermore, while the experiment's current phase (Phase II) uses T_2, since tritium is easiest to manage in its natural molecular form, the final states of T_2 present an irreducible systematic. Specifically, the uncertainty introduced by vibrational and rotational states of ^{3}HeT+ yields a best-case energy resolution of 100 meV, preventing Project 8 from probing neutrino mass scales below that energy [16, 15]. Thus, in its final phase (Phase IV), Project 8 will use an atomic tritium source. This should enable it to potentially attain an energy resolution of $\approx$40 meV, assuming that magnetic field strength uncertainty is the limiting systematic [2, 16].

One key goal of Phase II is to measure the endpoint of a β-decay spectrum and compare it with theoretical predictions. Chapter 4 details a new model-based analysis of Project 8's Phase II sensitivity to the endpoint. Chapter 5 details a similar sensitivity analysis that yields information regarding Project 8's ability to resolve m_β during Phase IV.

1.3   Modeling Spectra to Assess Sensitivity

This study aims to answer two sets of questions:

1. With what precision can Project 8 (Phase II) resolve the β-decay endpoint? How accurate can we expect that endpoint determination to be?

2. Under what conditions will Project 8 (Phase IV) be able to resolve the electron neutrino mass, and to what precision (as a function of runtime)? How accurate can we expect that mass determination to be?

To address these questions, I modeled β-decay spectra using the Stan probabilistic programming language [17, 18], as described in Chapter 2. I employed a modified spectral function, approximated to facilitate its use in an inferential model (see Chapter 3). With this model, I generated many pseudo-data sets, each of which assumed a particular configuration (set of values) of experimental and physical parameters. This allowed me to maximize the degree to which I incorporated available knowledge regarding statistical and systematic uncertainties in the system. I then analyzed each data set, performing Bayesian inference to extract posterior distributions for the endpoint and neutrino mass.

In Chapter 6.2, I lay out a similar approach for assessing Project 8's sensitivity to the mass hierarchy.

Chapter 2

A Model-Based Sensitivity Analysis with Stan and Morpho

This chapter discusses the methods applied to perform a Bayesian sensitivity analysis for Project 8. The approach described here involves statistical modeling and analysis using the *Stan* platform and probabilistic programming language [17, 18]. The choice to perform a sensitivity analysis with Stan was motivated by the benefits of bayesian modeling. A Bayesian approach directly illuminates the consequences of making inferences from data and decisions (e.g. to claim a result) based on those inferences. It also permits one to consider pseudo-data corresponding to wide regions of parameter space (or "model configurations"), instead of selecting a small number of "best guess" values [19]. Stan is particularly well suited to a model-based sensitivity analysis because it deals well with high dimensional problems and highly correlated posteriors [20]. At a basic level, Stan allows users to mostly abstract from difficult computation and focus on modeling systems — in this case, tritium β-decay. I implement this sensitivity analysis using *morpho*, a software tool developed with collaborators at MIT and Pacific Northwest National Laboratory that organizes information inflow to and outflow from Stan [21].

2.1   Calibrating Inferences and Decisions

Sensitivity analyses predict whether, given some data, an experiment will be able to claim a particular result — for example, "The Higgs boson exists," "The neutrino mass hierarchy is normal," or "The 90% credible interval for m_β is 50-90 meV." In Bayesian analyses (unlike frequentist ones), the decision or *action* of claiming a result is decoupled from the process of *inference*.

Bayesian inference yields posterior distributions $\pi(\theta|y)$ for parameters θ given data y. Such inference is based on Bayes' rule:

$$\pi(\theta|y) \propto \pi(y|\theta) \cdot \pi(\theta), \tag{2.1}$$

where $\pi(y|\theta)$ is the likelihood function of y given θ and $\pi(\theta)$ are priors on θ. Claims can be made about a data generating process — that is, the system that produces

the data—by computing expectations from posteriors (e.g. posterior means and intervals). Experimenters can decide whether or how to make a claim (what action to take) by constructing an inferential *loss function* L or a *utility function* $U \equiv -L$. A loss function quantifies the loss incurred by selecting a particular action a via some inferential decision making process [19].

For example, L could equal the percentage of time that a reported posterior interval on the parameter θ_a fails to include the the true value $\tilde{\theta}_a$—one of the possibilities discussed below. In that case, the loss function yields a Bayesian analogue to a typically frequentist quantity, since it is based on the distribution of outcomes expected when one repeats an experiment many times. When performing a sensitivity analysis, it is important to *calibrate* the consequences of claiming some sensitivity by computing the expected loss associated with that claim [19].

2.1.1   Calibrating a Credible Interval or Limit Claim

Calibrating model-based inferences produces a measure of how likely it is that reported results will be consistent with true parameter values. For this Project 8 sensitivity analysis, a calibration addresses the question: "If we repeatedly generate data given "true" (inputted) parameter values, how often does the true endpoint or neutrino mass fall within some reported energy window?" The *coverage* C of a model is the frequency with which a parameter of interest falls within a reported window [19]. This coverage is related to the relevant loss function by $C = 1 - L$.

The reported energy window can consist either of an upper limit (with a minimum value of zero) or of a credibility interval (upper and lower bounds). Either way, the energy window is constructed given some *credibility* α—meaning that some fraction α of the posterior mass falls within the window. Reporting a limit or interval amounts to claiming that an experiment and associated model has some sensitivity to the parameter of interest (Q or m_β). The coverage reflects the expected accuracy of that claim.

I follow the same procedure to calibrate claims of sensitivity to both m_β and Q. For a parameter $\theta_a \in \theta$, the process of estimating the coverage of a credible interval or limit is as follows [19]:

1. Develop a model of the data generating process, or system being studied, that depends on parameters θ.

2. Select "true" (input) parameter values $\tilde{\theta}$ by sampling from priors $\pi(\theta)$. Those priors should incorporate as much external knowledge about θ as is reasonable.[1]

3. Generate data (in this case, a β spectrum) using the model devised in step 1, with true values $(\tilde{\theta})$ as inputs.

---

[1]Statistical uncertainty on a parameter is incorporated into the model used for both data generation and analysis (steps 3 and 4), as it reflects an expected variation or fluctuation in the data itself. Systematic uncertainty is incorporated into priors used for pre-generation sampling and analysis (steps 2 and 4), as it represents a lack of clarity in our knowledge of naturally fixed parameters.

4. Analyze the data using the model devised in step 1 to infer a posterior distribution on θ_a.

5. Determine the posterior value(s) of θ_a corresponding to an interval or limit with credibility α.

6. Check whether $\tilde{\theta}_a$ (the true value of θ_a) falls within the chosen interval or limit.

7. Repeat steps 2-6 many times.

8. From the results of step 6, estimate the probability that $\tilde{\theta}_a$ will be consistent with a reported interval or limit.

The results of executing this procedure for Project 8's Phases II and IV are presented in Chapters 4 and 5, respectively.

2.1.2 Evaluating Model Performance

Bayesian modeling allows for an assessment of the extent to which an inferential process yields reasonable outcomes. Two quantities can be computed for each "experiment"—that is, for each data set (spectrum) generated and analyzed according to the above procedure. First, the *posterior z-score Z* quantifies the compatibility between a posterior and true value of a parameter θ_i. The z-score is defined as [19]

$$Z \equiv \frac{|\mu_{\text{post}}(\theta_i) - \tilde{\theta}_i|}{\sigma_{\text{post}}(\theta_i)}, \tag{2.2}$$

where $\mu_{\text{post}}(\theta_i)$ is the posterior mean and $\sigma_{\text{post}}(\theta_i)$ is the posterior standard deviation. As indicated by Eq. 2.2, Z reflects the number of posterior standard deviations difference between an expectation and the outcome. An experiment with a low z-score is highly accurate [19].

Second, the *posterior shrinkage S* quantifies how much the posterior on θ_i shrinks as compared to its prior. It is given by [19]

$$S \equiv 1 - \frac{\sigma^2_{\text{post}}(\theta_i)}{\sigma^2_{\text{prior}}(\theta_i)}, \tag{2.3}$$

where $\sigma^2_{\text{prior}}(\theta_i)$ is the prior standard deviation. A large shrinkage (close to 1) signals that an experiment is highly informative.

For an ensemble of experiments, each with inputs $\tilde{\theta}$ independently sampled from priors, one can identify pathologies that arise during the process of inference by considering together the distributions of z-scores and shrinkages (see Figure 2-1) [19]. This enables one to evaluate the performance of a sensitivity analysis model.
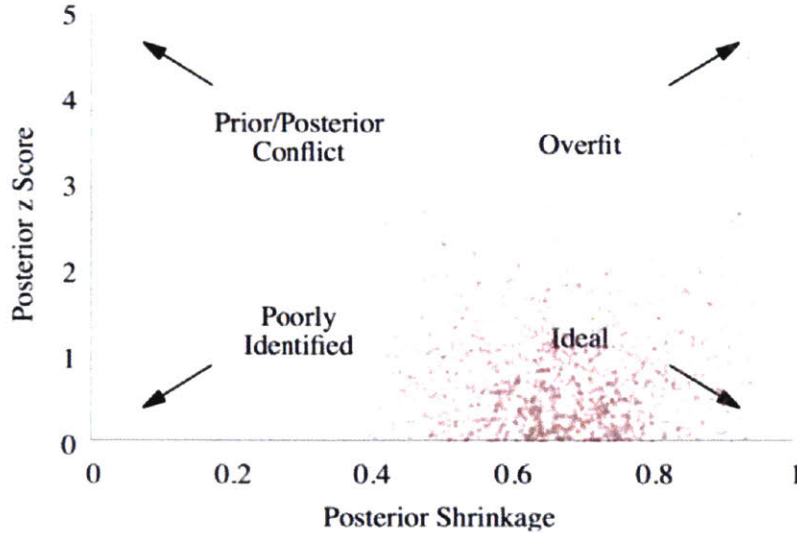
Figure 2-1: An example scatter plot of shrinkage vs. z-score

The model that produced this plot demonstrates good behavior (low z-score, high shrinkage) most of the time, but it occasionally overfits to the data (high z-score, high shrinkage). Plot from [19].

2.2 Tools for Modeling and Bayesian Inference

2.2.1 The Stan Statistical Software Platform

This sensitivity analysis uses Stan, a platform for Bayesian statistical modeling and computation, to model a β-decay spectrum, generate pseudo-data and infer posteriors. Stan performs Bayesian inference by way of Markov Chain Monte Carlo (MCMC), which uses Markov chains to explore a probability density parameter space. By doing so, the chains map out *typical sets*—the only regions of parameter space with densities that contribute significantly to parameter expectations. If a Markov chain moves stochastically for an infinite amount of time, it will completely explore the typical set, yielding exact posteriors. In reality, MCMC supplies estimates of posteriors [22].

However, Stan provides diagnostics that help users ascertain whether the Markov chains have explored the typical set in a representative way—or, in other words, whether the estimated posteriors closely approach exact posteriors. The most generic of these convergence diagnostics is the split $\hat{R}$ metric, which quantifies the consistency between the regions explored by a group of Markov chains. $\hat{R} = 1$ for perfectly consistent chains, and $\hat{R} > 1.1$ is a sign of problems in the fit. Another indication of a problem with the fit is that $n_{\text{eff}}/N < 0.001$, where N is the number of Markov chain iterations and n_{eff} is the effective sample size—that is, the effective size of each posterior array. A small n_{eff} indicates that the chains are exploring too slowly [23].[2]

---

[2]There are other important implications of n_{eff}. The effective sample size can limit the allowed credibility of reported limits and intervals. For my analysis, I required at least 150 effective samples

A particularly powerful algorithm in Stan, and the one used for this analysis, is Hamiltonian Monte Carlo (HMC). HMC harnesses information about the typical set—specifically, the gradient of the probability density function at each point in space—to direct Markov chains to move efficiently. (The mathematics that describes the differential geometry underlying HMC also describes classical physics, hence the name "Hamiltonian.") There are additional convergence diagnostics associated with HMC. Numerical trajectories of the HMC integrator can shoot of toward the boundaries of parameter space, producing "divergences." These divergences indicate that a chain has failed to properly explore a high-curvature region. In addition, the E-BFMI metric quantifies the success of a random walk between HMC iterations [22].

When diagnostics indicate pathological behavior, it is almost always possible to resolve the issue by more fully constraining or otherwise adapting one's model. Good diagnostic outcomes do not guarantee convergence, but diagnostic problems suggest that the algorithm has not fully converged [23]. However, given successful diagnostic tests, one can be reasonably confident that posteriors are accurate if they have the approximate shapes, central values, and correlations with other parameters that one would expect.

2.2.2   The Morpho Analysis Tool

Morpho is an analysis tool that organizes how information is inputted to and outputted by Stan via a python inferface. It is especially useful for generating pseudo-data and performing Bayesian statistical inference on real or fake data. Morpho was developed for this analysis and to serve other needs of Project 8; in fact, the Collaboration intends to analyze Phase II data using this software. However, it is designed to be employed by general Stan users, as well.

Morpho is a useful tool for several reasons. It streamlines Stan analyses by enabling users to control a range of processes through a single configuration file. These include loading data, running Stan, saving results, performing convergence diagnostic tests, and creating plots of posteriors and their correlations. Morpho also provides a framework for allowing Stan models to share input data, user-defined functions in Stan, and other information. Furthermore, it minimizes the need to recompile Stan models, sometimes a time-consuming task. Importantly, morpho automatically performs and displays the results of convergence checks after running Stan, helping users to quickly identify whether a model displays pathological behavior. It also offers options for convergence analysis and plotting, beyond these initial checks.

Using morpho and Stan, I implement the sensitivity analysis and calibration procedure described in Section 2.1. That procedure begins with devising a model of the β spectrum that can be implemented in Stan.

---

to fall outside of each bound. For example, for a 95% credible interval, the standard would be $0.025 \cdot n_{\text{eff}} \geq 150 \rightarrow n_{\text{eff}} \geq 6000$.

Chapter 3

A Simplified Beta Decay Spectral Model

Stan computes posteriors by inducing Markov chains to explore the likelihood space defined by a model. One can specify features of that likelihood space by adding log probability density functions (PDF) to a total log probability. To make inferences about which configurations of a model (i.e. parameter values) are consistent with a data set, it is necessary to formulate the model in terms of PDFs. In this Chapter, I show how a β-decay spectral model can be expressed as a PDF and thus implemented in Stan, following the derivation by Formaggio in [12].

3.1 A Two Neutrino Approximation

In working to devise a probability density version of a spectral function, we face a few challenges. The first is that the full β-decay spectrum, in terms of the three neutrino masses, does not explicitly depend on or relate simply to the probability that a particular mass ordering scheme (normal or inverted) is correct. That makes it difficult to infer posteriors for the hierarchy probabilities, preventing us from assessing a Project 8-like experiment's sensitivity to the mass hierarchies. However, it is possible to simplify and re-parameterize the spectrum so that it depends on a variable that relates directly to the ordering probabilities.

We make two simplifications to the spectral phase space function near the endpoint, introduced in Eq. 1.2:

$$\mathcal{P}(K) \simeq \sum_i \left[|U_{ei}|^2 (Q - K) \cdot \sqrt{(Q - K)^2 - m_i^2} \right] \cdot \Theta(Q - K - m_i). \qquad (3.1)$$

First, we adopt a two-mass-state approximation, where $m_a \simeq m_1 \simeq m_2$ and $m_b = m_3$. This is justified because distinguishing between m_1 and m_2 — that is, observing the less prominent "kink" in the spectrum — would require an energy resolution less than 8 meV (see Figure 1-3). Eight meV resolution is not necessary for a direct mass

measurement to reveal the mass hierarchy, and Project 8's objective is to reach ≈ 40 meV resolution, so the small mass splitting can be neglected. Second, we expand the neutrino momenta $p_{\nu i} = \sqrt{(Q-K)^2 - m_i^2}$ to first order in m_i^2 — the same order as the two-neutrino approximation. Thus, we would expect the effect of the expansion to similarly be small [12].

With these two simplifications, Eq. 3.1 reduces to

$$\mathcal{P} \simeq \eta \cdot \left[(Q-K)^2 - \frac{1}{2}m_a^2 \right] \cdot \Theta(Q - K - m_a)$$
$$+ (1 - \eta) \cdot \left[(Q-K)^2 - \frac{1}{2}m_b^2 \right] \cdot \Theta(Q - K - m_b),$$

where $\eta = |U_{ea}|^2 \approx |U_{e1}|^2 + |U_{e2}|^2$, and $|U_{ea}|^2 + |U_{eb}|^2 = 1$. To highlight this spectrum's dependence on the mass splitting, we can substitute $\Delta \equiv m_b^2 - m_a^2$ and rewrite the phase space equation in terms of the *lightest* and *heaviest* masses (m_L and m_H, respectively):

$$\mathcal{P}(\eta, m_L, m_H) \simeq \eta \cdot \left[(Q-K)^2 - \frac{1}{2}m_L^2 \right] \cdot \Theta(Q - K - m_L)$$
$$(1 - \eta) \cdot \left[(Q-K)^2 - \frac{1}{2}m_L^2 \right] \cdot \Theta(Q - K - m_H) \qquad (3.2)$$
$$\frac{|\Delta|}{2} \cdot (1 - \eta) \cdot \Theta(Q - K - m_H).$$

This expression describes spectra generated assuming both normal ($\mathcal{P}_{\mathrm{normal}}$) and inverted ($\mathcal{P}_{\mathrm{inverted}}$) neutrino mass orderings, provided that

$$\mathcal{P}_{\mathrm{normal}} = \mathcal{P}(\eta_N, m_L, m_H) \qquad\qquad \mathcal{P}_{\mathrm{inverted}} = \mathcal{P}(1 - \eta_N, m_L, m_H),$$

where η_N is the fractional contribution of the m_L-dependent term to the spectrum, if the masses are normally ordered. Note that $|\Delta|$ is equivalent to the neutrino mass splitting $\Delta m_{31}^2 \approx \Delta m_{32}^2$ as measured by vacuum oscillations. In addition, the (m_L, m_H) formulation is well suited to this analysis because model "fitters" (including Stan) prefer sets of parameters to be explicitly ordered, instead of having built-in ordering ambiguities.

A normalized version of Eq 3.2 can be used to model a β spectrum in Stan. Information regarding the mass hierarchy can be extracted from the posterior distribution on η — the posterior should center around η_N if the hierarchy is normal and around $[1 - \eta_N]$ if the hierarchy is inverted. But what is the expected value of η_N? If the solar splitting Δm_{21}^2 is considered negligible, then

$$\eta_N = |U_{e1}|^2 + |U_{e2}|^2 = 1 - |U_{e3}|^2 = \cos^2(\theta_{13}).$$

The ordering question can thus be formulated solely in terms of the large mass splitting and θ_{13}, both of which are measured by reactor anti-neutrino disappearance experiments [12]. Using the reported mean value of $\cos^2(\theta_{13})$ (μ_{c13}) and uncertainty

23

in that value (σ_{c13}), we can construct a prior on η (π_η) in probability density form:

$$\pi_\eta = P_N \cdot \mathcal{N}(\eta|\mu_{c13}, \sigma_{c13}) + (1 - P_N) \cdot \mathcal{N}(\eta|[1 - \mu_{c13}], \sigma_{c13}), \qquad (3.3)$$

where $\mathcal{N}$ is the normal function and P_N is the probability that the neutrino masses are normally ordered. The log of this PDF can be added to the total log likelihood in Stan to model P_N given some spectral data. The actual implementation of Eq. 3.3 employs a Stan function called `log_mix` that allows a user to combine two PDFs in a desired ratio — as determined by P_N, in this case.

Inferring a posterior on P_N provides information regarding the probability that each hierarchy exists. Hence, the above model enables a mass hierarchy determination using only a β spectrum and priors from reactor experiments. This is of particular advantage because reactor experiment priors are hierarchy-independent (unlike some neutrino mixing priors from other sources). It should be possible to perform a mass hierarchy sensitivity analysis with this model by inferring posteriors on P_N, then assessing the coverage of potential hierarchy claims made based on those posteriors. I describe this approach in more detail in section 6.2.

3.2   Incorporating a Finite Energy Resolution

In practice, the data obtained by a direct mass experiment cannot fully be described by Eq. 3.1 because of a "smearing" effect, quantified by energy resolution σ. For Project 8, this smearing derives primarily from the uncertainty in parameters used to convert an electron frequency measurement to an energy value [12]. Accounting for smearing, the spectral form becomes

$$\mathcal{F}(K) \equiv \int \mathcal{P}(K')\mathcal{N}(K'|K, \sigma)dK', \qquad (3.4)$$

where $\mathcal{N}$ is a normal distribution. This convolution is neither analytic nor normalizable, posing a problem: $\mathcal{F}$ cannot be used for an inferential model, as it cannot be expressed as a normalized probability density function.

This problem can be addressed by again expanding p_ν to first order in m_i^2, then considering spectral data within an energy window $[a, b]$. In that case, for a single neutrino mass m (or for one term of the full spectrum),

$$\mathcal{P}(K) = \left[(Q - K)^2 - \frac{1}{2}m^2\right] \cdot \Theta(Q - K - a)\Theta(b - Q + K).$$

The convolution of that expression with $\mathcal{N}$ yields a smooth analytic function:

$$\mathcal{F}(K|Q, a, m, \sigma) = N \cdot \left[\gamma(Q - K, m, m, \sigma) - \gamma(Q - K, Q - a, m, \sigma)\right], \qquad (3.5)$$
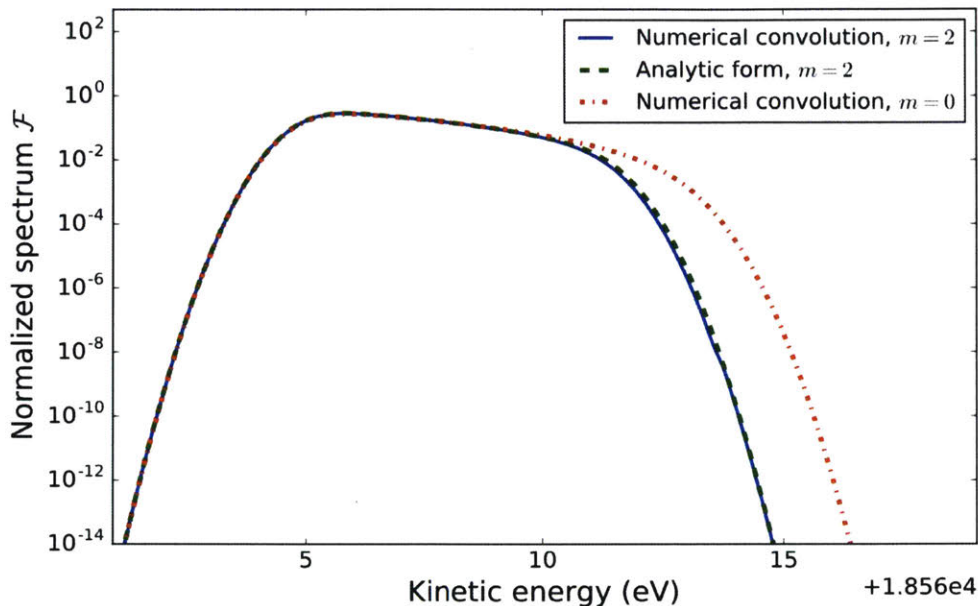
Figure 3-1: Comparison between approximate and exact smeared spectra
The approximate form given in Eq 3.5 is plotted alongside a numerical convolution of the exact spectral phase space function with a gaussian. In each case, $Q = 18573.25$ eV, $a = 18565$ eV and $\sigma = 0.5$ eV.

where

$$\gamma(u, v, m, \sigma) = (u + v)\sigma^2 \mathcal{N}(u|v, \sigma) + \frac{1}{2}\left(u^2 + \sigma^2 - \frac{m^2}{2}\right) \cdot \mathrm{erfc}\left(\frac{v - u}{\sqrt{2}\sigma}\right)$$

$$N = \frac{6}{m^3 - 3m^2(Q - a) + 2(Q - a)^3}.$$

(3.6)

Furthermore, the result is normalizable in the limit that $\sigma \to 0$ [12]. Figure 3-1 illustrates that this approximation shows good agreement with a numerical convolution of the single neutrino mass spectrum (not expanded in m^2) against a gaussian, normalized using the analytic expression for N in Eq. 3.6.

3.3 The Spectrum as a Probability Density Function

Combining the results above, we arrive at an approximate two neutrino model of a β-decay spectrum given finite energy resolution:

$$\mathcal{F}'(K) = \eta \cdot \mathcal{F}(K|Q, a, m_L, \sigma) + (1 - \eta) \cdot \mathcal{F}(K|Q, a, m_H, \sigma).$$

(3.7)

This function is analytic, well-behaved, and normalized, so it can be used to specify a likelihood space in Stan—again using the log_mix function. Posteriors on m_L, m_H,

Q, and σ can then be extracted. The parameter η is modeled according to Eq. 3.3, facilitating a mass hierarchy determination.

The Project 8 sensitivity analysis results presented in Chapters 4 and 5 do not consider the mass hierarchy problem. They therefore rely on the one neutrino spectral model in Eq. 3.5, with the substitution $m \to m_\beta$. In Chapter 6.2, I return to the two neutrino model in Eq. 3.7 and describe how it would be used in a mass hierarchy sensitivity analysis.

Chapter 4

Sensitivity to the Endpoint During Phase II

The main goal of this Phase II sensitivity analysis is to assess how precisely Project 8 can expect to resolve the T_2 β-decay endpoint Q_{T_2}. In addition, the analysis involves calibrating the consequences of claiming that Q_{T_2} falls within some credible interval, based on inferred posteriors. That calibration yields a *coverage* metric that indicates how accurately the endpoint can be resolved, if the model described here is used to analyze Phase II data (see Chapter 2). I also extract posteriors and report results for other parameters of interest: the electron neutrino mass m_β, the energy resolution σ, and the signal fraction (fraction of total events produced by tritium decays) f_s.

4.1 The Model

I generate and analyze data using the same model of Project 8 Phase II data. The main feature of this model is the approximated probability density form of the β-decay spectrum $\mathcal{F}(K)$ that was derived in Chapter 3 (see Eq. 3.5). I consider the function $\mathcal{F}(K|Q_{T_2}, a, m_\beta, \sigma)$. It is reasonable to substitute m_β for individual neutrino masses because this analysis considers a wide energy window and assumes σ is of $\mathcal{O}(1\,\text{eV})$; it should not be possible to distinguish between the mass eigenstatesat that resolution.

The data will also include a background $\mathcal{B}(K)$, within the same energy window $[a, b]$ as the signal $\mathcal{F}(K)$. A smeared flat background (convolved with the normal function $\mathcal{N}(K, \sigma)$) is described by the function

$$\mathcal{B}(K|a, b, \sigma) = \frac{1}{2(b-a)} \left[\text{erf}\left(\frac{b-K}{\sqrt{2}\sigma}\right) - \text{erf}\left(\frac{a-K}{\sqrt{2}\sigma}\right) \right]. \qquad (4.1)$$

The full model, then, takes the form

$$\mathcal{M}(K) = f_s \cdot \mathcal{F}(K|Q, a, m_\beta, \sigma) + (1 - f_s) \cdot \mathcal{B}(K|a, b, \sigma), \qquad (4.2)$$

where $f_s \equiv S/(S + B)$ is the signal fraction, S is the poisson rate of signal events, and B is the poisson rate of background events. The signal and background PDFs
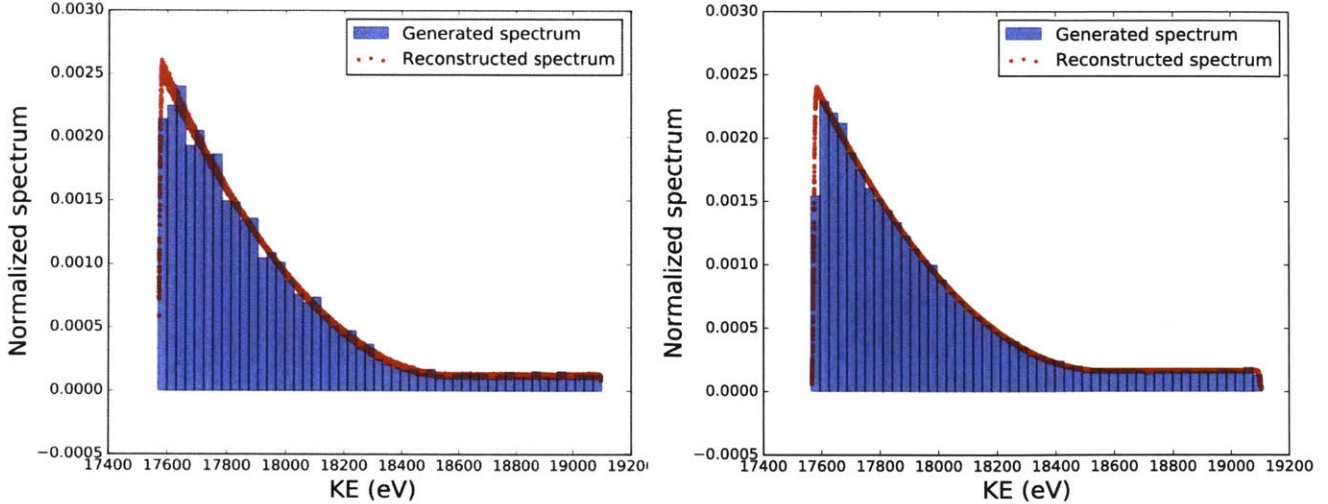
Figure 4-1: Phase II pseudo-data and reconstructed spectrum post-analysis
Left: Generated with a runtime of 3×10^6 s. *Right*: Runtime of 1 year.

are combined in Stan using the `log_mix` function. The total number of events in a spectrum N_{data} is sampled from a poisson distribution with rate $S + B$ before data generation, and the parameters S and B are modeled and extracted using the PDF `poisson`$(S + B)$ during data analysis.

The first steps of this sensitivity analysis, following the procedure laid out in Chapter 2, involved generating data with Eq. 4.2 using inputs sampled from priors. This requires priors to be constructed for the parameters Q_{T_2}, m_β, σ, and f_s (a and b are fixed inputs). However, the Project 8 Collaboration's past calculations and Phase I data supply predictions regarding the signal activity A_s and background activity A_b—as opposed to directly informing a prior on f_s. Here, A_s (A_b) is the number of events per second generated by $\mathcal{F}(K)$ ($\mathcal{B}(K)$) in the energy window [a, Q] ([a, b]). Thus, we can compute $f_s(S, B)$ from A_s and A_b using the equations

$$S = \Delta t \cdot A_s \cdot (Q - a) \qquad\qquad B = \Delta t \cdot A_b \cdot (b - a), \qquad (4.3)$$

where Δt is the experimental runtime for a given data set. The relationship between A_s, A_b and f_s is also incorporated into the analysis model, so that the signal fraction posterior is informed by priors on the signal and background activities.

I perform this analysis for two runtimes—$\Delta t = 3 \times 10^6$ s (about one month), and $\Delta t = 3.154 \times 10^7$ s (one year)—to provide a sense of the effect of the total event number N_{data} on the precision with which Q_{T_2} and other parameters can be resolved. For both runtimes, I consider data within a kinetic energy window [$a = 17573.24$ eV, $b = 19100.00$ eV]. The window extends 1 keV below the value of Q_{T_2} theoretically computed by Bodine, Parno and Robertson in [15]. To permit a reasonably time-efficient analysis process, I histogram the data with 300 bins for

Parameter	Phase II prior	Information contained in prior
Endpoint Q_{T_2} (Analysis)	**norm(18573.24 eV, 15 eV)**	(Extrapolated endpoint$-m_e$) from [15].
Energy Resolution σ	$\boldsymbol{\gamma(13.30,\ 3.961\,\mathbf{eV^{-1}})}$	5% probability mass falls below 2 eV, 5% falls above 5 eV [2, 24].
Neutrino Mass m_β	$\boldsymbol{\gamma(1.042,\ 1.538\,\mathbf{eV^{-1}})}$	1% probability mass below 8 meV [8]), 5% above 2 eV [9, 10].
Background Activity A_b	**lognorm(-16.12, 1.400)**	5% probability mass below 10^{-8} events /eV/s, 5% above 10^{-6} events/eV/s [2].
Signal Activity A_s	$\boldsymbol{\gamma(36,\ 24000\ \mathbf{keV \cdot s})}$	Mean: 4500 events/(3×10^6 s) in last keV [24]. $\sqrt{\text{variance}}$: 750 events/(3×10^6 s).

Table 4.1: Phase II priors for pre-generation sampling and analysis. Prior functions are defined in Appendix A.

both runtimes.[1] Therefore, during data analysis, I model each data point $N_i(K_i)$ as a value sampled from a poisson distribution with spectral rate $\mathcal{M}(K_i)$ (see Eq. 4.2).

4.2 Priors

In this analysis, five parameters require priors based on external information: Q_{T_2}, m_β, σ, A_s and A_b. Before a spectrum is generated, each of these parameters is sampled from its prior, determining the inputs to the generator. By sampling from these priors repeatedly, creating an *ensemble* of model configurations, I can approach an analysis that accounts for the full range of possible data sets, given statistical and systematic error. I directly pass Q_{T_2}, m_β and σ to the generator, as they appear in Eq. 4.2. On the other hand, A_s and A_b are used to compute f_s and N_{data}, which I then input to the generator. The prior distributions used for pre-generation sampling are also included in the analyzer model.

Some key considerations guided my choice of priors. In particular, I selected distributions with boundary conditions that reflected physical limits on parameters. Because each of the five parameters of interest must be positive, I often used γ distributions, which are lower bounded at zero. I employed a gaussian instead of a γ prior for the endpoint, because the distribution is localized far enough from zero that the negative portion of the gaussian tail is negligible. The background activity is also an exception; I chose a lognormal prior to account for the fact that the range of possible values spans multiple orders of magnitude (see Table 4.1). I discuss the reasoning behind each prior in more detail, below.

- The prior on the **energy resolution** (σ) was constructed so that 5% of its probability mass would fall below 2 eV, and 5% would fall above 5 eV. This is consistent with Project 8's Phase I energy resolution measurements of 3.3 eV

---

[1]I performed one test ensemble analysis with 1000 bins for Δt =1 year. It resulted in a coverage within 1% of the corresponding ensemble analysis with 300 bins.

and 3.6 eV, for two different sets of spectral lines [2]. Current predictions for the Phase II resolution (2.6-3.0 eV) also fall near the center of the prior [24].

- The **neutrino mass** (m_β) prior was constructed so that 1% of its probability mass would fall below 8 meV, reflecting the relatively hard lower bound from mass splitting measurements [8]; this bound is not strict only because of uncertainties on those measurements. In addition, 5% of the prior mass falls above 2 eV, to account for the limit by the Mainz and Troitsk experiments [9, 10]).

- 90% of the probability mass of the **background activity** (A_b) prior falls in the range $[10^{-8}, 10^{-6}]$ events/eV/s. This reflects the small number of background events seen previously by Project 8 within some cyclotron frequency window, and it is centered well above the minimum A_b expected due to the emission of delta electrons caused by cosmic rays passing through the source [2, 24].

- The mean of the **signal activity** (A_s) prior was calculated based on Project 8's current expectation of detecting 4500 events/3×10^6 s in the last 1 keV of spectrum [24]. The prior's variance (corresponding to a FWHM of 750 events/3×10^6 s in the last keV) was chosen to roughly account for the uncertainty introduced by possible HT contamination and details of the magnetic trap configuration.

- In a typical Bayesian sensitivity analysis, all available information about the main parameter of interest—in this case, the **endpoint**—would be incorporated into the prior on that parameter. However, Phase II is expected to approximately confirm pre-existing knowledge of Q_{T_2}, not improve upon that knowledge, so the prior used for analysis must be wider than the true best-guess prior. (Otherwise, the Q_{T_2} posteriors would be unaffected by the data.) Thus, while the endpoint is known to $\approx$1eV [15], the analysis prior has a standard deviation of 15 eV, to guide Stan's Markov chains to the desired region without strongly informing the results. For a realistic sensitivity analysis, the endpoint values used to generate spectra should be sampled from a prior with a standard deviation Q_σ^{in} of $\approx$1 eV. However, the model's ability to distinguish between different Q_{T_2} values can be seen more clearly when a broader pre-sampling prior is used. Thus, for each runtime, I performed two sets of pseudo-experments: one realistic set with Q_σ^{in}=1 eV, and one illustrative set with Q_σ^{in}=15 eV.

4.3 Endpoint Sensitivity Results

Fifty pseudo-experiments were performed for each of the four ensemble analyses (with Q_σ^{in}=1 eV or 15 eV, and Δt=3 $\times$ 10^6 s or 1 year). Qualitatively, the posteriors produced by analyzing each spectrum appear to agree with the data, as seen in Figure 4-1. There, "reconstructed spectra" were produced by first sampling from posteriors, then computing spectral points with Eq 4.2, given those sampled values.

Furthermore, the data strongly informed posteriors on Q_{T_2}, as expected. The informative capacity of the experiments is quantified by posterior shrinkage values
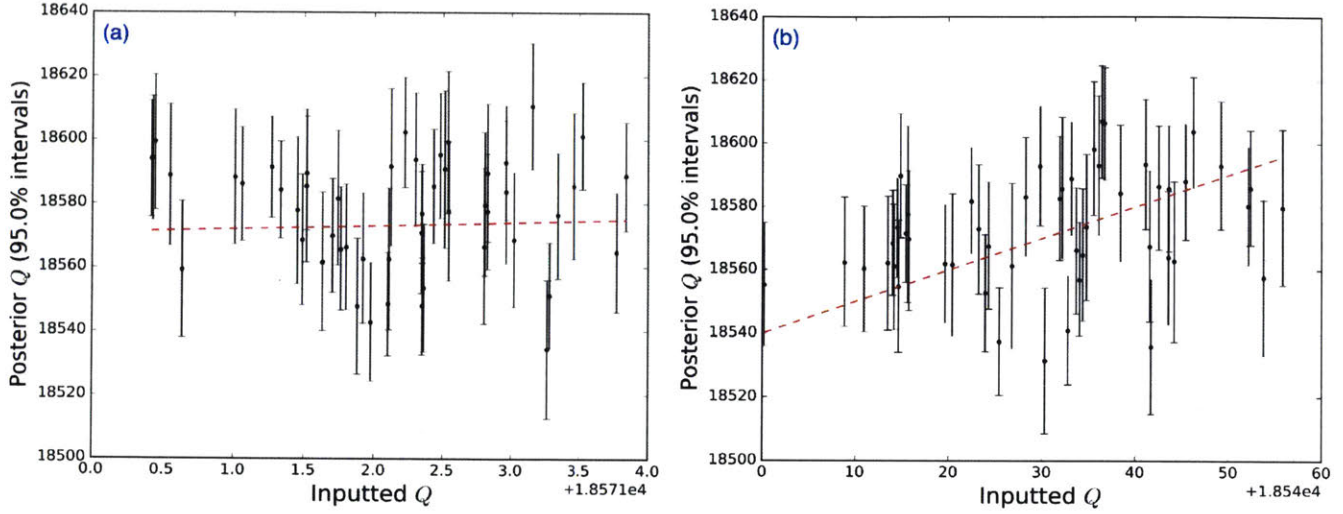
Figure 4-2: Expected Q_{T_2} means and 95% credible intervals after 3×10^6 s of runtime *Left*: $Q_\sigma^{in}=1$ eV, *right*: $Q_\sigma^{in}=15$ eV. The red dashed lines indicate where posterior means would fall if they corresponded exactly with true (inputted) values.
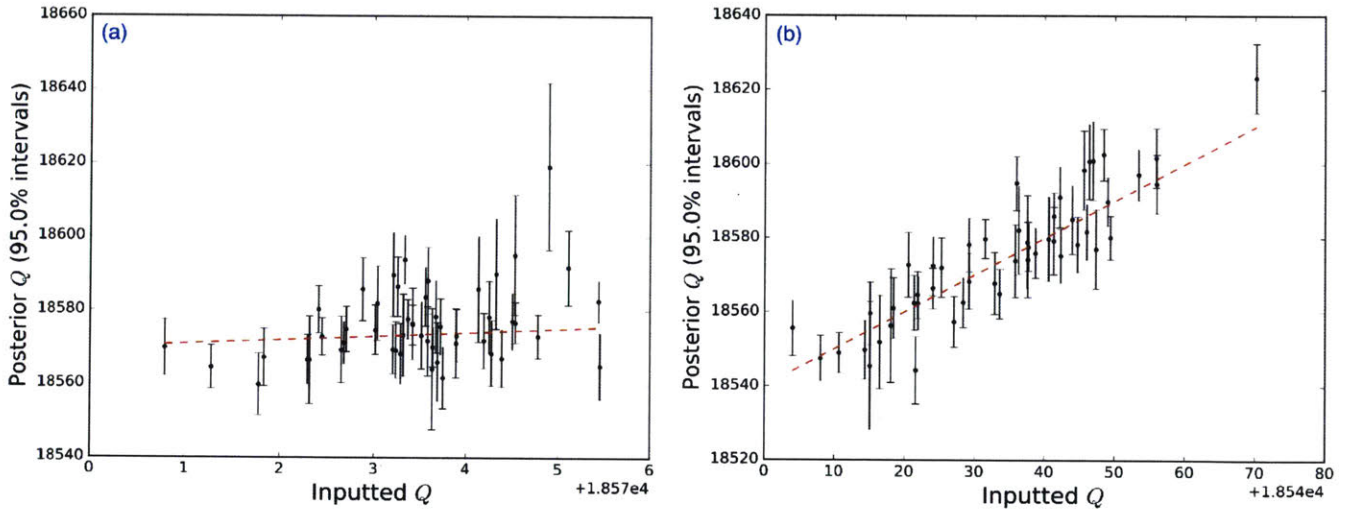
Figure 4-3: Expected Q_{T_2} means and 95% credible intervals after 1 year of runtime *Left*: $Q_\sigma^{in}=1$ eV, *right*: $Q_\sigma^{in}=15$ eV.

for Q_{T_2} — all relatively close to 1, as seen in Figure 4-4. In addition, Figures 4-2b and 4-3b reinforce the notion that the data and spectral model generally affect Q_{T_2} posteriors as expected, since they show that higher inputted values correspond to higher posterior means.

Figures 4-2 and 4-3 show 95% posterior credible intervals (C.I.s) on Q_{T_2}. The lower (upper) bound of each of these intervals was estimated by computing the endpoint value below which 2.5% (97.5%) of the points in the Stan posterior array fall. For each data analysis run, the effective sample size n_{eff} was $\approx 18{,}000$, large enough that 95% C.I.s could be reliably estimated. For each of the four ensemble analyses, posterior

Figure 4-4: Z-score vs. shrinkage plots for the molecular tritium endpoint
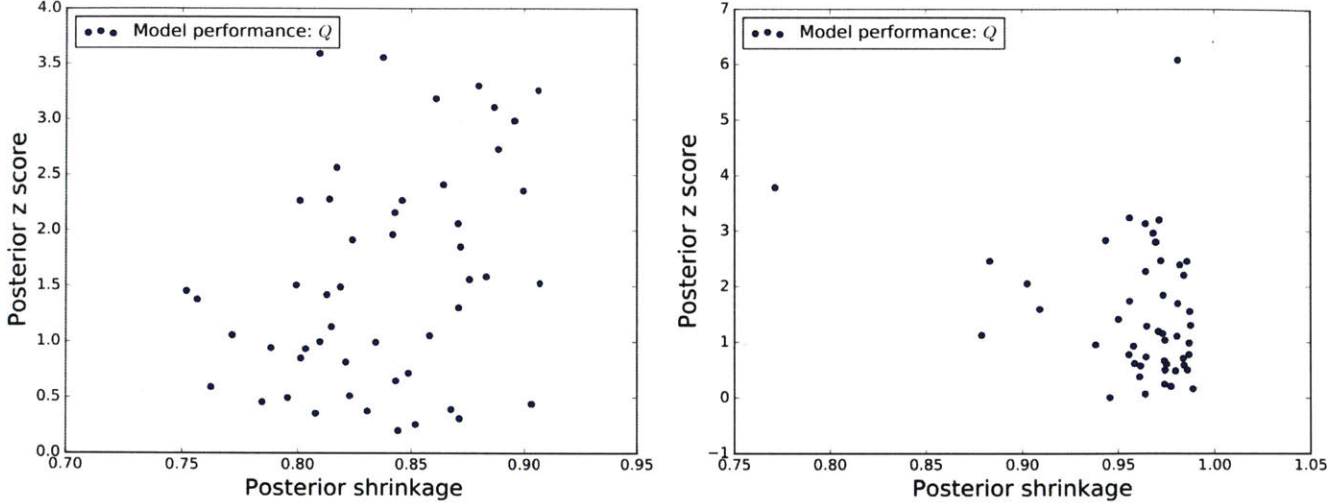The z-score indicates the number of posterior deviations of difference between the true value and posterior mean. Shrinkage indicates how informative the data is. Here, $Q_\sigma^{in}=1$ eV. For the *left* plot, $\Delta t = 3 \times 10^6$ s, and for the *right* plot, $\Delta t = 1$ year.

coverages were computed as described in Chapter 2 for both 90 and 95% credible intervals on each parameter of interest, including Q_{T_2} (see Tables 4.3, 4.4).

	Sensitivity (eV), 3×10^6 s	Sensitivity (eV), 1 year
Standard dev.	10.1	4.1
90% C.I.	33.6	13.7
95% C.I.	39.9	17.7

Table 4.2: Sensitivity to Q_{T_2}: median posterior credible intervals, standard deviations
Results reported for the "realistic" pre-generation sampling prior, $Q_\sigma^{in} =1$ eV.

Sensitivity results for Q_{T_2} are summarized in Table 4.2. Each "sensitivity" is a median of the 50 standard deviations or C.I.s inferred during one ensemble analysis.[2] I find that, using the spectral model $\mathcal{M}$, Q_{T_2} can be resolved after one year of runtime within a ~ 13.7 eV window (90% C.I.) with 62% coverage. In other words, the true value of Q_{T_2} would fall within the reported interval 62% of the time. Alternatively, the endpoint can be resolved within a ~ 16.3 eV window (95% C.I.) with 68% coverage. Those predicted sensitivities correspond to a median posterior standard deviation of 4.1 eV after one year.

Endpoint coverage results were only minimally affected by the width of the pre-sampling endpoint prior and by the runtime. While this sensitivity analysis suggests that Phase II data can enable an endpoint determination with reasonable precision,

---

[2]Average and median energy windows (standard deviations or C.I.s) differ by < 1.5 eV.

Δt	Q_σ^{in} (eV)	Q_{T_2} coverage	σ coverage	m coverage	f_s coverage
3×10^6 s	1.0	**62%**	76%	90%	100%
	15	62 %	84 %	86%	98%
1 year	1.0	**62%**	70%	94%	94%
	15	60%	60%	98%	100%

Table 4.3: Coverage of 90% credible intervals for Phase II
Coverage results reported for runtimes Δt and endpoint standard deviations Q_σ^{in} on the prior used for pre-generation sampling.

Δt	Q_σ^{in} (eV)	Q_{T_2} coverage	σ coverage	m coverage	f_s coverage
3×10^6 s	1.0	**68%**	80%	94%	100%
	15	74 %	90%	88%	100%
1 year	1.0	**68%**	74%	100%	94%
	15	68%	68%	98%	100%

Table 4.4: Coverage of 95% credible intervals for Phase II
Coverage results reported for runtimes Δt and endpoint standard deviations Q_σ^{in} on the prior used for pre-generation sampling.

that determination is likely to be inaccurate about a third of the time. It is possible that this coverage could be improved by including more or different priors, adapting how the spectral model $\mathcal{M}$ is implemented in Stan, or making changes to $\mathcal{M}$ itself.

4.4 Evaluating Model Performance

Several checks and metrics can be considered when evaluating the performance of the Phase II model. First, for each generation or analysis Stan run, one can diagnose whether the MCMC Markov Chain Monte Carlo algorithm appeared to converge properly, as described in section 2.2.1. For all parameters of interest in the generator and analyzer, $\hat{R}$ was < 1.1, N_{eff}/N was < 0.001, and the E-BFMI diagnostic indicated no pathological behavior. A small number of generator iterations ($<0.1\%$) did consistently diverge; those divergences were caused by some aspect of the signal and background PDFs. However, it is unlikely that the generator divergences substantially affected coverage results, because, when $\mathcal{S}$ and $\mathcal{B}$ were replaced with other PDFs that did not yield divergences, the analyzer was able to recover input parameters at similar frequencies.

It is also worth examining the posteriors on parameters other than Q_{T_2}, to assess whether the data and model are informing inferences as expected. We expect posterior means to track with input values; in other words, the points in Figures 4-5 and 4-6 should fall near the red dashed line. For both f_s and σ, input value and posterior means are highly correlated. Note that the true ("inputted") values of f_s were estimated from the generated data by examining an effectively pure background region to compute the overall background fraction. This calculation was performed

to account for the discrepancy between the pre-sampled signal fraction and the f_s, cause by a poisson uncertainty inherent in the data generation process.

The distribution of posterior z-scores and shrinkage values for each parameter helps to diagnose pathologies of Bayesian inference [19]. Most Q_{T_2} posteriors had the desired combination of low z-score and high shrinkage, characteristic of an accurate and informative model (see Figure 4-4). However, some endpoint outcomes overfit to the data (high z-score, high shrinkage). The model overfitted A_s, A_b and σ to the data, as well, while m_β outcomes were weakly identified but accurate, as expected.[3] The model's tendency to display an overfitting pathology indicates that it may be possible to revise the model or adjust its priors in order to improve its accuracy. In considering how the model might be adapted, there is no parameter in the model that obviously requires attention, as the model's ability to recover Q_{T_2} was not correlated with any particular region of parameter space (see Figure 4-7).
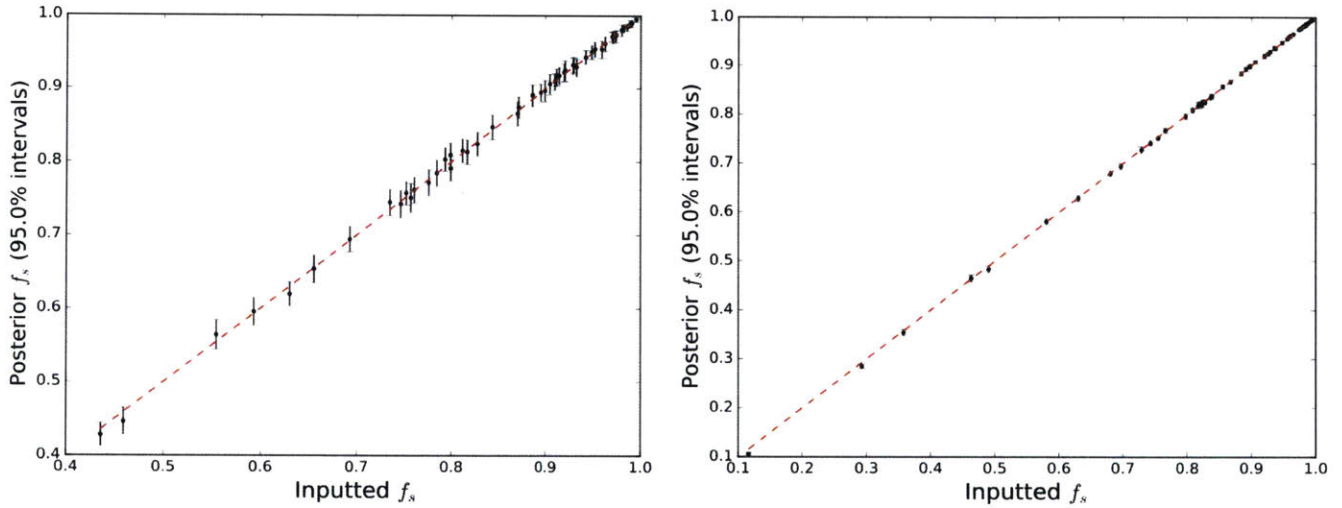

Figure 4-5: Expected f_s means and 95% credible intervals (Phase II) with $Q_\sigma^{in}=1$ eV *Left*: $\Delta t = 3 \times 10^6$ s. *Right*: $\Delta t = 1$ year.

---

[3]While the model overfitted A_s and A_b, it demonstrated ideal behavior with respect to f_s (if the "true" signal fractions are computed from pseudo-data post-generation). That suggests that overfitting of A_s and A_b is likely caused by a poisson counting error inherent in the data generation process, and/or by the anti-correlation between S and B.
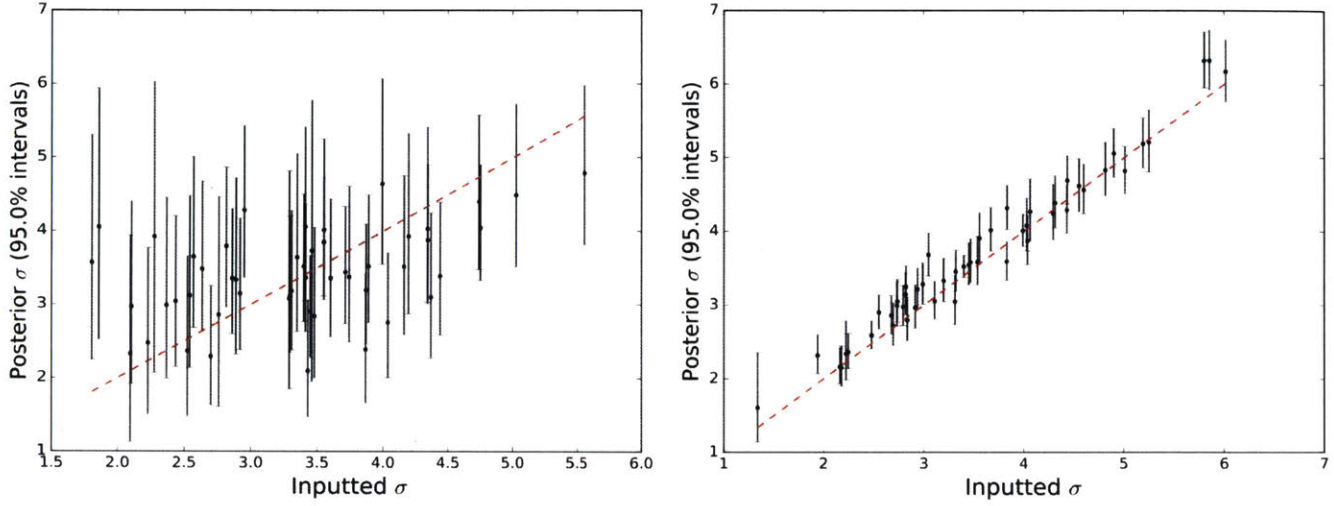
Figure 4-6: Expected σ means and 95% credible intervals (Phase II) with $Q_\sigma^{\mathrm{in}}=1$ eV
Left: $\Delta t = 3 \times 10^6$ s. *Right*: $\Delta t = 1$ year.

Figure 4-7: Relationship between Phase II model coverage of Q_{T_2}, f_s and σ
"Recovered" indicates that the parameter's true (inputted) value fell within a 90%
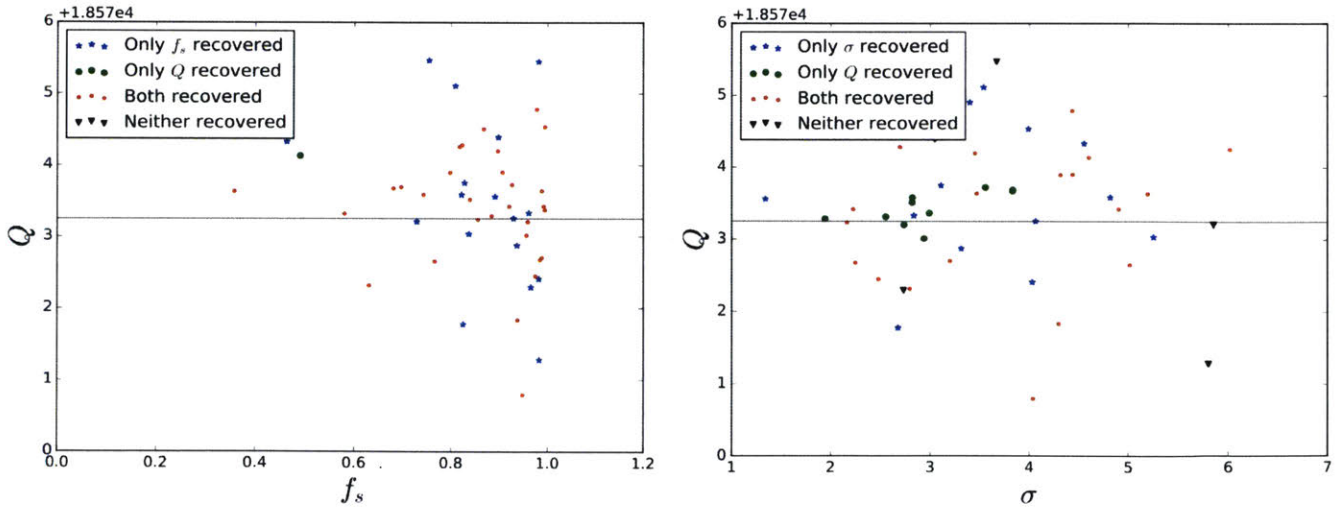credible interval. $Q_\sigma^{\mathrm{in}}=1$ eV. *Left*: $\Delta t = 3 \times 10^6$ s. *Right*: $\Delta t = 1$ year.

Chapter 5

Sensitivity to the Neutrino Mass During Phase IV

The goal of this Phase IV sensitivity analysis is to assess how precisely Project 8 can expect to resolve the electron neutrino mass m_β, and with what coverage. I also extract posteriors and report results for other parameters of interest.

5.1 The Model

The Phase IV β spectrum model largely resembles the Phase II model described in Chapter 4, with a few distinctions:

1. The priors differ, to reflect the difference in experimental conditions.

2. Phase IV data is analyzed within a much narrower kinetic energy window of $[a = 18562.25 \text{ eV}, \ b = 18564.25 \text{ eV}]$ — centered around the atomic tritium mean endpoint $Q_T = 18563.25 \text{ eV}$ [15].

3. In the Phase IV analysis, the endpoint is modeled as a normal distribution with mean Q_T and standard deviation δQ. This reflects the fact that the endpoint varies on an event-to-event basis because of the final state distribution—that is, the variation in energy states of the decay product.

A fuller Phase IV analysis would exploit a model that incorporates multiple neutrino mass eigenstates, as described in Section 6.2.

5.2 Priors

The reasoning underlying the choice of each Phase IV prior is as follows:

- The **endpoint** Q_T: The mean and error on the mean were calculated in [15].
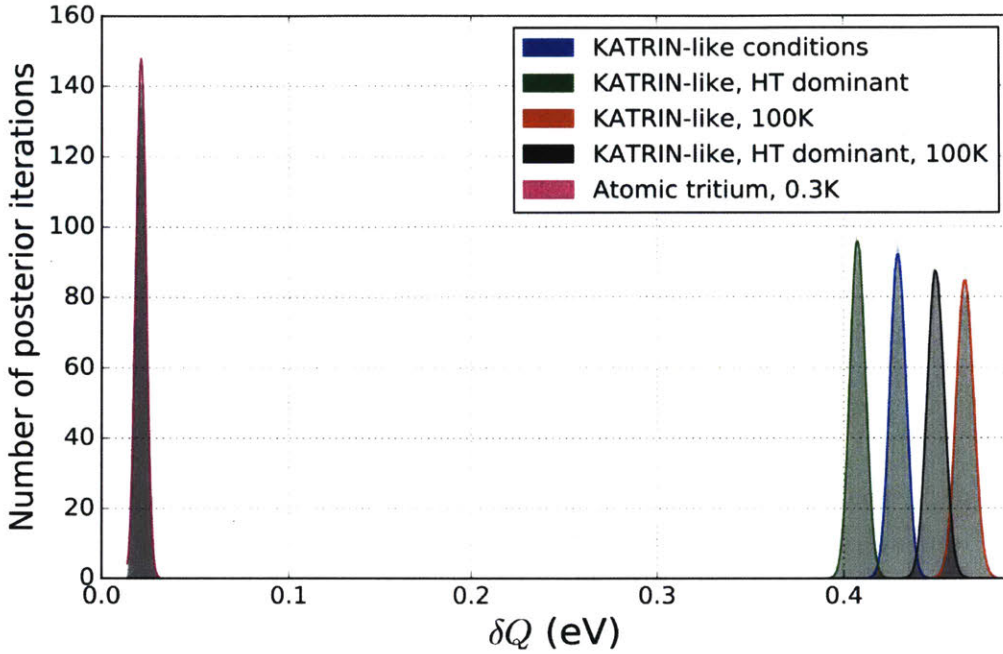
Figure 5-1: Posteriors of the endpoint spread δQ under varying conditions Posteriors inferred using a Stan model of the endpoint given final states. "KATRIN-like" refers to the conditions specified in Table XV of [15]. A gaussian fit to the posterior for an atomic tritium source was used as a prior on δQ for the Phase IV spectrum analysis.

- The **endpoint spread** δQ: I devised a Stan model that extracted a posterior for the mean expected endpoint spread due to final states ($\mu_{\delta Q}$), as well as the uncertainty on that spread ($\sigma_{\delta Q}$), under variable conditions (i.e. temperature and source composition). Normal distributions were fitted to the $\mu_{\delta Q}$ and $\sigma_{\delta Q}$ posteriors for a trap at 0.3 Kelvin with negligible T_2 contamination. The mean ($\sqrt{\text{variance}}$) of the γ distribution prior on δQ was taken to be the mean of the $\mu_{\delta Q}$ ($\sigma_{\delta Q}$) posterior from the final states model. Phase IV will likely operate at a temperature below 0.3 K, making this prior somewhat conservative.

- The **neutrino mass** m_β: The Phase II and Phase IV priors are identical.

- The **energy resolution** σ: To resolve the mass hierarchy using only the value of m_β, $\sigma \lesssim 50$ meV is needed (see Figure 1-2). (For a hierarchy analysis that accounts for the full spectral shape, this requirement may not be as stringent.) Thus, Project 8's target resolution is 40 meV. A magnetic field uncertainty of $\Delta B/B \approx 10^{-7}$ is required to achieve this, if that represents the limiting systematic [2]. As this field uncertainty will be difficult to achieve, the γ-distribution prior on σ has a mean of 50 meV and a variance of $[10 \text{ meV}]^2$.

- The **signal activity** A_s: The expected Phase IV signal activity was computed

Parameter	Phase IV prior	Information contained in prior
Endpoint Q_T	**norm(18563.25 eV, 0.3 eV)**	(Extrapolated endpoint$-m_e$) and error on the mean from [15].
Endpoint Spread δQ_T	$\gamma(59.82, 2868\,\text{eV}^{-1})$	Mean: 0.0209 eV; $\sqrt{\text{variance}}$: 0.0027 eV. From Stan final states model based on [15].
Energy Resolution σ	$\gamma(25, 500\,\text{eV}^{-1})$	Mean: 50 meV; $\sqrt{\text{variance}}$: 10 meV. From planned B-field uncertainty [2].
Neutrino Mass m_β	$\gamma(1.042, 1.538\,\text{eV}^{-1})$	1% probability mass below 8 meV [8]), 5% above 2 eV [9, 10].
Background Activity A_b	**lognorm(-27.63, 1.400)**	5% probability mass below 10^{-8} events /eV/s, 5% above 10^{-6} events/eV/s [24].
Signal Activity A_s	**lognorm(-5.634, 1.400)**	5% probability mass below 0.000357 events eV/s, 5% above 0.0357 events/eV/s [1, 2].

Table 5.1: Phase IV priors for pre-generation sampling and analysis Prior functions are defined in Appendix A.

from the half life of tritium ($\tau_{1/2}$) and the fraction of events in the last 1 eV of the β-spectrum (f), as well as estimated Phase IV quantities: the effective volume (V_{eff}) and the atomic density of the source (ρ). Following the approach in [1], the A_s prior should center around

$$\mu_{A_s} = \eta \cdot V_{\text{eff}} \cdot \frac{\ln(2)}{\tau_{1/2}} \cdot f = 10^{18}\,\text{atoms/m}^3 \cdot 10\text{m}^3 \cdot \frac{\ln(2)}{12.3\,\text{years}} \cdot 2 \times 10^{-13} \tag{5.1}$$

$\to \approx 112,700$ events in one year.

The values chosen for V_{eff} and ρ correspond to the predictions in [2]. The Phase IV model included a lognormal prior on A_s, constructed so that 90% of its probability mass falls within one order of magnitude of A_{s_μ} (the distribution's "soft bounds" differ by two orders of magnitude). This large range reflects the fact that predictions of V_{eff} and ρ are relatively speculative, at this stage.

- **The background activity** A_b: The near elimination of almost all sources of background is needed for Phase IV. Project 8's goal is for the dominant source of background to be cosmic rays passing through the source gas. Since the expected background from cosmic rays is around 10^{-12}/eV/s, I constructed a lognormal prior on A_b with 90% of its probability mass falling between 10^{-13} and 10^{-11}/eV/s.

5.3   Neutrino Mass Sensitivity Results

Nine pseudo-experiments were performed, enough to provide a very preliminary indication of Phase IV sensitivity. This analysis should be conducted on a larger scale

in the future to better account for the range of possible model configurations. Each experiment had a runtime of $\Delta t = 1$ year.

In general, the data strongly informed posteriors on m_β; all mass posteriors demonstrate high shrinkage. Moreover, the m_β outcomes appear to follow the expected trend-line on which posterior means correspond with true values (the red dashed line in Figure 5-2). The left plot in Figure 5-2 shows 90% posterior C.I.s on m_β. The effective sample size n_{eff} was always greater than 3000, large enough that 90% C.I.s could be reliably estimated. However, sometimes, n_{eff} was < 6000, so 95% C.I.s are not reported. Posterior coverages were computed as described in Chapter 2.
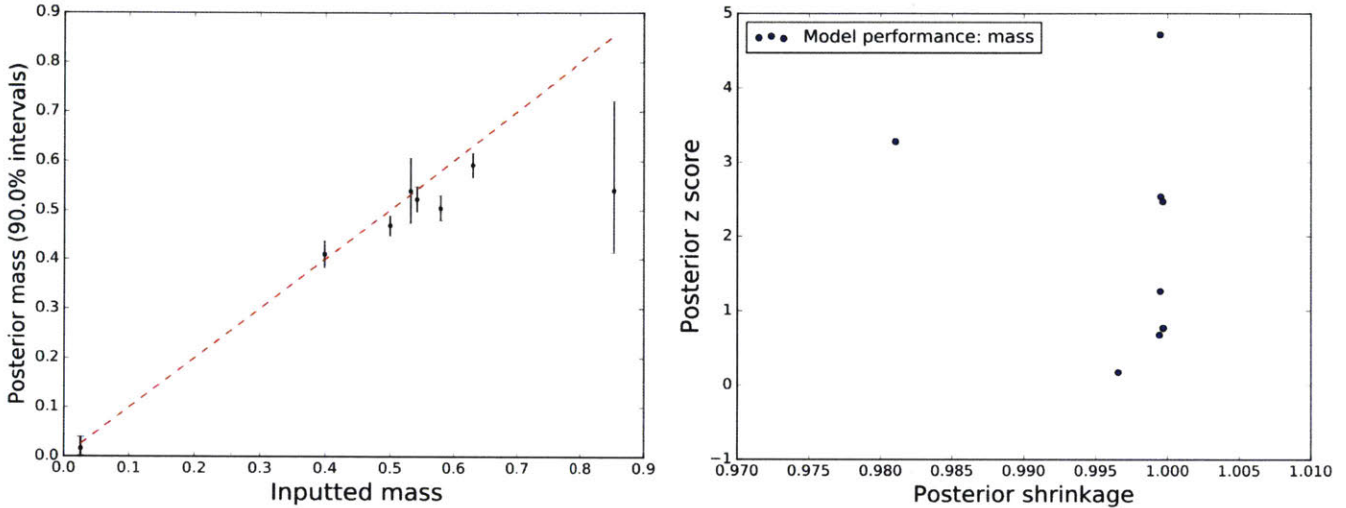

Figure 5-2: *Left*: expected m_β means and 90% C.I.s; *right*: shrinkage vs. z-score

I find that the electron neutrino mass can be resolved after one year of runtime within a median 90% credible window of 0.051 eV with 56% coverage.[1] In other words, the "true value" of m_β would fall within the reported interval 56% of the time. That predicted sensitivity corresponds to a median posterior standard deviation of 0.016 eV after one year.

While this sensitivity analysis suggests that Phase IV will have a neutrino mass sensitivity near that expected from analytical predictions [1], an m_β result with 90% credibility is likely to be inaccurate over 40% of the time. As in the case of the Phase II analysis, this coverage could potentially be improved by modifying priors and the spectral model $\mathcal{M}$.

Stan convergence diagnostics showed similar behavior to diagnostics for the Phase II analysis, as described in Section 4.4. As expected, posteriors means track with input values for σ, Q_T and δQ_T (see Figure 5-3). Signal fractions were never successfully recovered, likely because the Markov chains failed to explore the space corresponding

---

[1]That window is the median of the C.I. windows inferred from the nine experiments. The mean window is somewhat larger, at 0.085 eV. This discrepancy likely reflects both the small number of experiments in this analysis and the wide A_s prior, producing data with widely varying statistics underscoring the importance of achieving a high enough signal activity.

to f_s values very close to 1. Only the mass and signal activity posteriors showed signs of overfitting; still, it would be worth re-parameterizing the model in an effort to improve its accuracy.
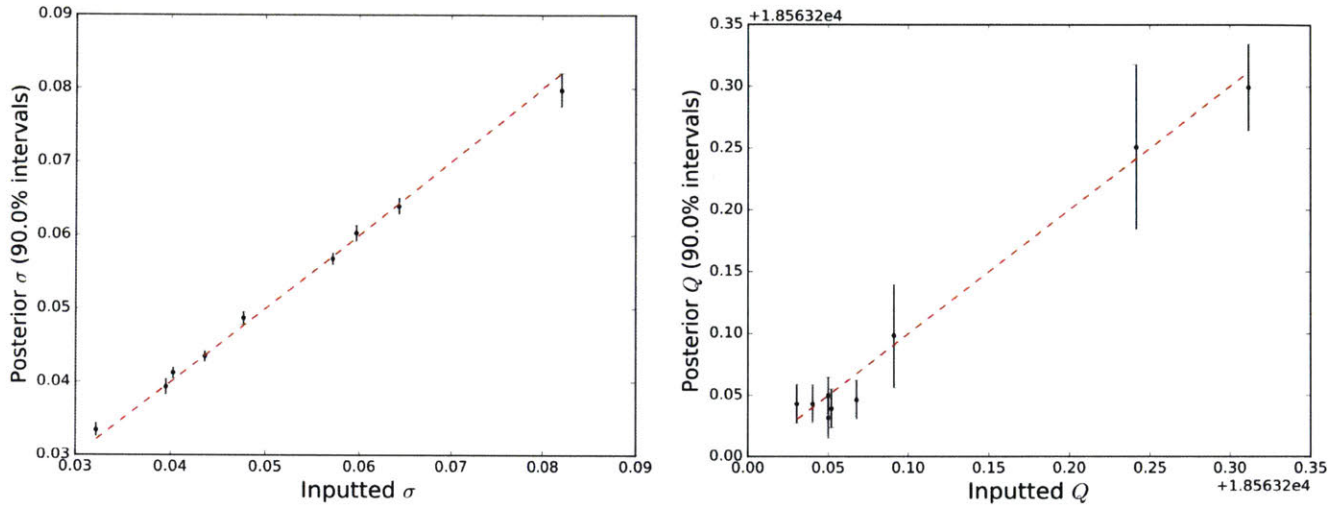

Figure 5-3: Expected σ (*left*) and Q_T (*right*) Phase IV means and 90% C.I.s

Chapter 6

Discussion and Conclusions

6.1　Discussion of One-Neutrino Sensitivity Analysis

In this thesis, I presented a new simplified model of the tritium β-decay spectrum measured with a finite energy resolution. This model is analytic and normalizable [12]; it is therefore amenable to being formulated as a probability density function. I also presented results of a Bayesian analysis of the Project 8 Neutrino Mass Experiment's sensitivity to the T_2 endpoint and electron neutrino mass. I find that, using one year of Project 8 Phase II data, the T_2 endpoint can be resolved within a ~ 13.7 eV window (90% C.I.) with 62% coverage, corresponding to a ~ 4.1 eV posterior standard deviation. Preliminarily, using one year of Phase IV data, the electron neutrino mass can be resolved within a ~ 0.051 eV window (90% C.I.) with 56% coverage. The neutrino mass sensitivity result suggests that Project 8 will be able to probe near the energy at which one can distinguish between mass orderings on the basis of an m_β measurement. While 0.051 eV just exceeds the minimum allowed inverted hierarchy mass, it may be possible to resolve the hierarchy at that sensitivity given a multi-neutrino mass analysis that accounts for the full spectral shape (see Section 6.2).

These results indicate that certain future steps should be taken to advance this sensitivity analysis. First, it is worth re-parameterizing or otherwise adapting the spectral model and priors in an effort to improve coverage. In addition, the Phase IV inference and calibration procedure should be performed for limits on m_β, instead of credible intervals. The limits and credible intervals should converge at some neutrino mass value; this value would serve as an estimate of the mass above which it should be possible to report a measurement instead of a limit. Finally, the Phase II and IV analyses should be repeated using pseudo-data generated with a more exact spectral model, to more thoroughly test the validity of the approximations made in Chapter 3.

6.2 A Bayesian Approach to Assessing Sensitivity to the Mass Hierarchy

6.2.1 Returning to a Two Neutrino Model

The next stage of this sensitivity analysis will involve predicting the precision and accuracy with which Project 8 can expect to resolve the mass hierarchy. Chapter 3 detailed how one could infer a posterior on the normal hierarchy probability P_N, using a spectral model $\mathcal{M}$, spectral data and external mixing information from reactor experiments. To estimate the hierarchy probabilities, one might think that instead of modeling P_N, it makes sense to determine the frequency with which Stan's Markov chains explore typical sets corresponding to each mass ordering. However, the latter approach is ill advised because of a MCMC *multimodality problem*: each chain gets "stuck" exploring one typical set. Thus, it would require infinite time for the probability that Markov chains converge to a solution to reflect the probability that that solution describes the data generating process. In the future, a technique called Adiabatic Monte Carlo technique should address the multimodality problem in Stan from a computation-focused perspective, precluding the need to avoid multimodality when modeling systems [25].

Once an experimenter infers a posterior on P_N from real spectral data, the next step will be to decide what action a to take, given that posterior. Should the experimenter report that the mass ordering is normal? That it is inverted? Should he or she report nothing, until more data can be collected? An analysis of sensitivity to the mass hierarchy involves estimating the utility or loss associated with each of these actions for a range of possible model configurations.

6.2.2 Calibrating the Consequences of Reporting a Hierarchy Result

	Report normal	Report inverted	No claim
Truth: normal	$\mathcal{U}_{NH}^{\mu}$	$\mathcal{U}_{IH}^{\mu}$	$\mathcal{U}_{\text{none}}^{\mu}$
Truth: inverted	$\mathcal{U}_{NH}^{\mu'}$	$\mathcal{U}_{IH}^{\mu'}$	$\mathcal{U}_{\text{none}}^{\mu'}$

Table 6.1: Average expected utilities of claims regarding the mass hierarchy

This approach is an adaptation of the method for calibrating discovery claims detailed in [19]. Say that one can only claim an inverted (normal) hierarchy determination if most of the posterior mass of P_N falls below (above) some value b_l (b_h). If most of the mass falls between b_l and b_h, one should claim to be unable to resolve the hierarchy. It is possible to calibrate the consequences of those actions by calculating the utilities $\mathcal{U}_{NH}$, $\mathcal{U}_{IH}$ and $\mathcal{U}_{\text{none}}$ associated with each claim. These utilities are
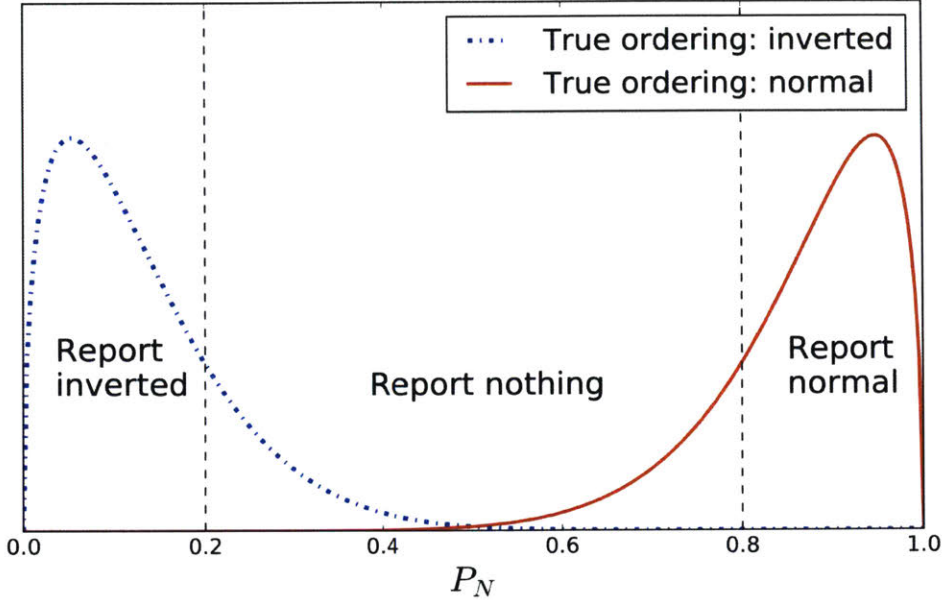
Figure 6-1: Hypothetical P_N posteriors, given normal and inverted true orderings The black dashed lines indicate probability cut-offs (b_l and b_h) that dictate what result is claimed. The locations of the black lines can be adjusted, affecting the relative utilities of different actions.

computed from posterior probabilities P_N:

$$\mathcal{U}_{NH} = \Theta(P_N - b_h) \qquad \mathcal{U}_{IH} = \Theta(b_l - P_N)$$
$$\mathcal{U}_{\text{none}} = \Theta(P_N - b_l)\Theta(b_h - P_N). \qquad (6.1)$$

For a particular data set (β spectrum), one can calculate mean utilities $\mathcal{U}^{\mu}_{NH}$, $\mathcal{U}^{\mu}_{IH}$ and $\mathcal{U}^{\mu}_{\text{none}}$ by taking averages of the posterior arrays obtained using Eq. 6.1. These hypothetical mean values are organized in Table 6.2.2. They quantify the trade-off between the certainty with which an experimenter can claim a mass hierarchy determination and how often that result can be claimed in the first place. Hence, by changing b_l and b_h, one can adjust the relative magnitudes of average utilities. For example, if one moves b_l and b_h closer to 0 and 1, respectively, the average utilities of reporting normal and inverted hierarchy determinations increase, while the average utility of reporting nothing decreases. However, that stricter reporting requirement means that the hierarchy can be resolved less often. Thus, an experimenter can decide how stringent to make his or her hierarchy reporting requirements by considering average utilities. By repeating this process for a large number of model configurations sampled from priors, one can calibrate an experiment's predicted sensitivity to the mass hierarchy.

In conclusion, the study described in this thesis successfully applies model-based sensitivity and calibration procedures to predict the precision and accuracy with which the Project 8 Experiment can expect to measure the molecular tritium end-

43

point and electron neutrino mass. This analysis both accounts for a range of model configurations given parameter uncertainties and corroborates analytical predictions of Project 8's sensitivity. Furthermore, this thesis lays out a procedure for extending model-based sensitivity methods to the neutrino mass hierarchy problem.

Appendix A

Functional Forms of Prior Distributions

The prior distributions used in both the Phase II and Phase IV sensitivity analyses are defined as follows. Each PDF below is implemented via a Stan function that outputs the log of the probability density of a parameter y [20].

1. **Normal distribution**

$$\text{norm}(\mu, \sigma) \equiv \text{norm}(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right)$$

2. **Gamma distribution**

$$\gamma(\alpha, \beta) \equiv \gamma(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}\exp(-\beta y)$$

$$\text{where } \Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx \tag{A.1}$$

3. **Lognormal distribution**

$$\text{lognorm}(\mu, \sigma) \equiv \text{lognorm}(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma y}\exp\left(-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2\right)$$

Bibliography

[1] Doe et al. Project 8: Determining neutrino mass from tritium beta decay using a frequency-based method. *Snowmass Conference Proceedings, arXiv:1309.7093 [nucl-ex]*, 2013.

[2] A. Ashtari Esfahani et al. Determining the neutrino mass with cyclotron radiation emission spectroscopy—project 8. *J. Phys. G: Nucl. Part. Phys*, 44(054004), 2017.

[3] J. A. Formaggio. Direct neutrino mass measurements after planck. *Physics of the Dark Universe*, 4:75–80, 2014.

[4] K. Abe et al. Solar neutrino results in super-kamiokande-iii. *Phys. Rev. D*, 83(052010), 2011.

[5] B. Aharmim et al. Low energy threshold analysis of the phase i and phase ii data sets of the sudbury neutrino observatory. *Phys. Rev. C*, 81(055504), 2010.

[6] G. Mitsuka et al. Study of non-standard neutrino interactions with atmospheric neutrino data in super-kamiokande i and ii. *Phys. Rev. D*, 84(113008), 2011.

[7] F. An et al. Observation of electron?antineutrino disappearance at daya bay. *Phys. Rev. Lett.*, 108(171803), 2012.

[8] C. Patrignani et al. (Particle Data Group). Neutrino mass, mixing, and oscillations. *Chin. Phys. C*, 40(100001), 2016.

[9] C. Kraus et al. Final results from phase ii of the mainz neutrino mass search in tritium β decay. *Eur. Phys. J. C*, 40(447), 2005.

[10] V. Aseev et al. Upper limit on the electron antineutrino mass from the troitsk experiment. *Phys. Rev. D*, 84, 112003 2011.

[11] P. Vogel and A. Piepke. Neutrinoless double-beta decay. *Particle Data Group*, August 2017.

[12] J. Formaggio. Internal note, July 2017.

[13] L. Bornschein et al. on behalf of KATRIN. Status of the karlsruhe tritium neutrino mass experiment katrin. *Fusion Science and Technology*, 71(4):485–490, 2017.

[14] J. Formaggio B. Monreal. Relativistic cyclotron radiation detection of tritium decay electrons as a new technique for measuring the neutrino mass. *Phys. Rev. D*, 80(051301), 2009.

[15] L. Bodine et al. Assessment of molecular effects on neutrino mass measurements from tritium β decay. *Phys. Rev. C*, 91(035505), 2015.

[16] W. Pettus (Project 8). Overview of project 8 and progress towards tritium operation. *XV International Conference on Topics in Astroparticle and Underground Physics Proceedings*, 2017.

[17] B. Carpenter et al. Stan: A probabilistic programming language. *J. Stat. Softw.*, 76:1–32, 2017.

[18] Stan Development Team 2015. Pystan: the python interface to stan, version 2.9.0, 2015.

[19] Betancourt M. Calibrating model-based inferences and decisions. *arXiv:1803.08393v1*, March 2018.

[20] Stan Development Team. Stan modeling language, version 2.17.0: User's guide and reference manual, September 2017.

[21] J. Formaggio M. Guigue J. Johnston B. LaRoque N. Oblath and T. Weiss. morpho, 2018.

[22] M. Betancourt. A conceptual introduction to hamiltonian monte carlo. *arXiv:1701.02434v1 [stat.ME]*, January 2017.

[23] J. Howell. Stan best practices, October 2017.

[24] H. Robertson. Microtritium experiment (internal paper), April 2018.

[25] M. Betancourt. Adiabatic monte carlo. *arXiv:1405.3489v5 [stat.ME]*, August 2015.