Statistical Issues in Long-Baseline Neutrino Experiments

Alex Himmel

PhyStat-v, CERN
January 23rd, 2018
What is a Long-Baseline Neutrino Experiment?

• A beam of neutrinos (primarily $\nu_\mu$) is produced by an accelerator.
• The neutrinos are allowed to travel and oscillate over a long distance.
• Detectors at the near and far end observe the neutrinos before and after oscillations.
• While energy and distance vary, the ratio of $L/E$ is chosen to maximize “atmospheric” oscillations.

Alex Himmel
How to study oscillations: Disappearance

\[ P(\nu_\mu \rightarrow \nu_\mu) \approx 1 - \left( \sin^2(2\theta_{13}) \sin^2(\theta_{23}) + \cos^4(\theta_{13}) \sin^2(2\theta_{23}) \right) \sin^2 \left( \frac{\Delta m^2 L}{4E} \right) \]

Sub-dominant term due to small $\theta_{13}$

One “dip” due to the fixed baseline.

How to study oscillations: Disappearance

Compare the measured spectrum to the unoscillated prediction...

But there are large uncertainties from the neutrino flux and cross-section.

\[
\sin^2 2\theta_{23}
\]

\[
\Delta m^2
\]
How to study oscillations: Disappearance

NOvA Simulation

Compare the measured spectrum to the unoscillated prediction...

Control those systematics with measurements at a Near Detector.

\[ \sin^2 2\theta_{23} \]

\[ \Delta m^2 \]
How to study oscillations: Appearance

\[ P(\nu_\mu \to \nu_e) \approx \left| \sqrt{P_{\text{atm}}} e^{-i(\Delta_{32} + \delta_{CP})} + \sqrt{P_{\text{sol}}} \right|^2 \]

\[ \approx P_{\text{atm}} + P_{\text{sol}} + 2\sqrt{P_{\text{atm}}P_{\text{sol}}} (\cos \Delta_{32} \cos \delta_{CP} \mp \sin \Delta_{32} \sin \delta_{CP}) \]

\[ \sqrt{P_{\text{atm}}} = \sin(\theta_{23}) \sin(2\theta_{13}) \frac{\sin(\Delta_{31} - aL)}{\Delta_{31} - aL} \Delta_{31} \]

- Depends some on every oscillation parameter.
- **Benefit**: can answer more questions.
- **Drawback**: degeneracies make things difficult.

\[ [R(\theta_{23}) \cdot R(\theta_{13}, \delta_{CP}) \cdot R(\theta_{12})] \]

\[ \Delta m_{32}^2 \to O(10^{-3}\text{eV}^2) \]

\[ \Delta m_{21}^2 \to O(10^{-5}\text{eV}^2) \]
How to study oscillations: Appearance

Still start from the **unoscillated** $\nu_\mu$ prediction based on ND measurements.

\[ \nu_\mu \rightarrow \nu_e \text{ oscillations are sub-dominant (a few %)} \]

Need ND measurements to constrain both **signal** and background.

NOvA Simulation

Alex Himmel
Open questions in neutrino oscillations

1. Do neutrino oscillations violate $CP$ symmetry directly via $\delta_{CP}$?

$$R(\theta_{23}) \cdot R(\theta_{13}, [\delta_{CP}] \cdot R(\theta_{12})$$

2. Is the mass hierarchy “normal” or “inverted”?

3. What is the “octant” of $\theta_{23}$?
   – Or is the mixing “maximal” (e.g. $\theta_{23} = 45^\circ$)?
1. Is the mass hierarchy “normal” or “inverted?”
2. Do neutrino oscillations violate $CP$ symmetry?
3. What is the “octant” of $\theta_{23}$?
1. Is the mass hierarchy “normal” or “inverted?”
2. Do neutrino oscillations violate $CP$ symmetry?
3. What is the “octant” of $\theta_{23}?$

CP-violation through $\delta$ creates opposite effects in neutrinos and antineutrinos.
1. Is the mass hierarchy “normal” or “inverted?”
2. Do neutrino oscillations violate $CP$ symmetry?
3. What is the “octant” of $\theta_{23}$?

CP-violation through $\delta$ creates opposite effects in neutrinos and antineutrinos.

$$P\left(\bar{\nu}_\mu \rightarrow \bar{\nu}_e\right) \%$$

$$P\left(\nu_\mu \rightarrow \nu_e\right) \%$$

- $\delta=0, \delta=\pi/2$
- $\delta=\pi, \delta=3\pi/2$
1. Is the mass hierarchy “normal” or “inverted?"
2. Do neutrino oscillations violate $CP$ symmetry?
3. What is the “octant” of $\theta_{23}$?

Matter effects also introduce opposite neutrino-antineutrino effects.
1. Is the mass hierarchy “normal” or “inverted?”
2. Do neutrino oscillations violate $CP$ symmetry?
3. What is the “octant” of $\theta_{23}$?

![Diagram showing the octant of $\theta_{23}$ and neutrino oscillations]
Here I will focus on three:

1. How to handle systematic uncertainties
2. How to handle challenging likelihood surfaces
3. How to say something about a binary choice like the mass hierarchy.
Systematics in a 2-detector experiment

- NOvA has similar near and far detectors:
  - Allows more direct “prediction”, connecting bins at the ND directly to bins at the FD.
  - Big worry: systematics which have different effects in the two detectors (acceptance) or which do not cancel in the analysis (absolute energy scale).
    - Also potentially “hidden” model dependency

- T2K has near and far detectors with different materials and designs:
  - Allows a broad suite of ND measurements, but
  - A complete suite of model parameters are needed to “predict” the FD from the ND.
  - Big worry: need to use the “right” model to extrapolate, how to quantify “model choice” uncertainty.

- See posters by L. Kolupaeva (NOvA) and F. Bench (T2K) for more details.
Difficult Likelihoods: Systematics

- They create non-trivial effects on the spectrum.
  - Both (most?) experiments: effect of each systematic on each analysis bin parameterized as a spline.

- Many systematics are non-linear (and sometimes non-Gaussian)
  - Most extreme example are “alternate models”: $0 = \text{model A}, 1 = \text{model B}$ and only valid $[0,1]$
  - For these you may want a uniform (instead of Gaussian) pull distribution.
Difficult Likelihoods: Systematics

• Where the experiments differ is in how these systematics are handled in the fit.

• T2K
  – Marginalizes everything through brute force for both Frequentist and Bayesian analyses.
  – Slow, but better handles non-linear/non-Gaussian systematics.
    • By integrating over all regions, not just best fit, capture regions where the profiled point does not predict the rest of the distribution.

• NOvA
  – Always uses profiling
  – Faster, but does not capture non-Gaussian well.
  – Mitigate inaccuracy by making conservative choices
    • Doable since dominated by statistical error in all parameters.
Difficult Likelihoods: Boundaries

The obvious: oscillation probability cannot be > 1.

When fitting for $\sin^2(2\theta_{13})$, this boundary is apparent since $\sin^2$ cannot be > 1 where (2-flavor) disappearance is maximal.

When fitting for $\sin^2(\theta_{23})$, there is still a boundary at maximal disappearance, but it is less apparent.
There are 2 parameters that are “binary” choices:
- Mass Hierarchy
- Octant

There are degeneracies between them and $\delta_{\text{CP}}$.

$\delta_{\text{CP}}$ is cyclical creating effective “boundaries” in the likelihood.
- While the parameter can keep varying, we run into limits in how much it can change the observable.
• David van Dyk summarized why these properties are a problem at the last PhyStat-nu, and they haven’t gone away.
What to do about it: Frequentist Approaches

• Feldman-Cousins can be used to give sufficient coverage.
  – Ensures at least correct coverage with nuisance parameters.
  – Used in both NOvA and T2K (and many other experiments)

• Pseudo-experiments are generated assuming each relevant value of $\theta$.

• Replace the standard $\chi^2$ with an empirical distribution, $F(x|\theta)$:

$$F(x|\theta) = \text{Fraction of } N \text{ experiments where } [\chi^2(\text{fixed } \theta) - \chi^2(\text{best fit}) = x]$$

• A point $\theta$ is inside the (1-$\alpha$) confidence interval if less than (1-$\alpha$) experiments are more extreme than the data.
  – i.e. if the integral of $F(x|\theta)$ up to the observed $\Delta \chi^2$ at $\theta$ is $< (1-\alpha)$.

• $F(x|\theta)$ can...
Feldman-Cousins Pseudo-experiments

- The art is in how to throw the pseudo-experiments.
- The parameter(s) of interest, $\theta$, is fixed and used for all experiments at a given test point.
- The question is all the other parameters.
- T2K and NOvA use two different variants of a “posterior predictive” method.
Feldman-Cousins Pseudo-experiments: T2K

• For $\theta_{13}$ and systematics:
  - Draw from prior distributions (PDG or output of ND fit)

• For $\sin^2\theta_{23}$ and $\Delta m^2$:
  - Generate an Asimov dataset at best fit values, and construct a likelihood for this simulated dataset.
  - Convert the likelihood to a PDF, and draw values for the experiments from that distribution.

Pseudo-Experiments
Sample from these prior/posterior distributions
Feldman-Cousins Pseudo-experiments: NOvA

- Fit the data and extract parameters with all possible values of $\theta$.
- When generating experiments, always use the best fit to the nuisance parameters from the fit to data for each $\theta$.
  - Minimizes over-coverage of all methods we examined while still never under-covering.
- Tested coverage with a method from Berger and Boos for handling $p$-values with unknown nuisance parameters.
  - Tested coverage at a variety of choices of oscillation nuisance parameters within 3$\sigma$
    - Reducing quoted significance by a very small amount
  - In all cases, the other choices of nuisance parameters produced stronger rejection than the quoted rejection at the nominal profiled values.
  - This is as expected if everything is working correctly since the profiled point should give the widest CIs or lowest significance.

Alex Himmel
Statistical Approach: Feldman-Cousins

- The outcome of the FC procedure are new “critical values” which vary for different values of the parameters of interest.
Issues with Feldman-Cousins

- **Computational cost**
  - NOvA’s FC corrections required 20M CPU-hours on NERSC supercomputers.
  - See poster by S. Calvez for lots more details.

- **Looks like a likelihood, but it isn’t a likelihood.**
  - For example, profiling doesn’t work: different answers for MH and octant rejection depending on which plot you look at.
    - And both are wrong!

---

**Graphs:**

- **Left graph:**
  - Plot of signficance ($\sigma$) versus $\sin^2\theta_{23}$.
  - Blue line represents normal hierarchy.
  - Red line represents inverted hierarchy.
  - Significant levels: $1.1\sigma$.

- **Right graph:**
  - Plot of significance ($\sigma$) versus $\delta_{CP}$.
  - Dotted blue line for NH Lower octant.
  - Solid blue line for NH Upper octant.
  - Dashed red line for IH Lower octant.
  - Solid red line for IH Upper octant.
  - Significant levels: $1.2\sigma$. 

---

**References:**

- NOvA Preliminary
What to do about it: Bayesian Approach

• Calculate the “posterior probability,” the probability distribution of the parameters given the data:

\[
P(\theta | \text{data}) = \frac{P(\text{data} | \theta) P(\theta)}{P(\text{data})}
\]

Easy to calculate

Must define prior

Constant (can be neglected)

• From \( P(\theta | \text{data}) \), can easily see best fit point, credible intervals, and relative probabilities (Bayes factors)

• Problem 1: the \( \theta \) space is very large
  – >400 parameters in T2K ND+FD fit

• Problem 2: How to define the prior, \( P(\theta) \)
Bayesian: Many-dimensional Integral

- The solution: *Metropolis-Hasting* algorithm
  - A Markov Chain MC whose steps sample the posterior distribution via a random walk.

- Start at a point in the space, $M_0$
- Randomly choose a new point to step to, $M_1$
- Decide whether to step based on $r$
  - Where $q$ is a gaussian with width $\sigma$ tuned to maximize efficiency.

$$r = \frac{P(M_1|\text{data})q(M_0|M_1)}{P(M_0|\text{data})q(M_1|M_0)}$$
  - Have probability $r$ to take the step.

- After many steps, the distribution of steps maps the posterior.
Bayesian: Choice of Prior

Flat in $\delta$

Flat in $\sin(\delta)$

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\sin(\delta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1$\sigma$</td>
<td>-2.39 to -1.13, -1 to -0.68 or -0.91</td>
</tr>
<tr>
<td>2$\sigma$</td>
<td>-2.95 to -0.50, -1 to -0.19 or -0.48</td>
</tr>
<tr>
<td>3$\sigma$</td>
<td>2.80 to 0.13, -1 to 0.33 or 0.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\sin(\delta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1$\sigma$ -1 to -0.79</td>
</tr>
<tr>
<td>2$\sigma$ -1 to -0.32</td>
</tr>
<tr>
<td>3$\sigma$ -1 to 0.25</td>
</tr>
</tbody>
</table>
FC for Mass Hierarchy in NOvA

- Deciding if any individual point $\theta_0$ is outside a CI is equivalent to a hypothesis test where $H_0$ is $\theta = \theta_0$.
  - Same FC technique used for setting CI’s can be used for this hypothesis test.

- Since our best fit is in the NH, we want to know how strongly we reject the IH.
  - So $H_0$ is IH and we generate pseudo-experiments at our best fit in the IH.
  - Follow the FC procedure with:
    
    $\chi^2(\text{test point}) - \chi^2(\text{best fit}) \rightarrow \chi^2(\text{IH}) - \chi^2(\text{best fit})$

  - If an experiment has a best fit in the IH, then the difference is 0.
  - This pile-up at 0 behaves like a physical boundary: it increases significance.

Limiting Case: No sensitivity

- Half of experiments in each hierarchy and $\Delta\chi^2 = 0$
- $p = 0.5$
- 50% for either NH or IH
- All “prior”
Bayesian for Mass Hierarchy in T2K

- Binary parameters are no problem: simply integrate the posterior within each choice.
  - This is my favorite feature of doing a Bayesian analysis.

<table>
<thead>
<tr>
<th></th>
<th>$\sin^2 \theta_{23} &lt; 0.5$</th>
<th>$\sin^2 \theta_{23} &gt; 0.5$</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0.184</td>
<td>0.705</td>
<td>0.889</td>
</tr>
<tr>
<td>Inverted</td>
<td>0.021</td>
<td>0.090</td>
<td>0.111</td>
</tr>
<tr>
<td>Sum</td>
<td>0.205</td>
<td>0.795</td>
<td>1</td>
</tr>
</tbody>
</table>

- Bayes factor of 8 preferring the Normal Hierarchy...
  - ...but most physicists have no instinct for Bayes factors.
Frequentist → Bayesian Mass Hierarchy in T2K

• A similar Bayesian approach is used by the likelihood analyses which usually produce Frequentist CIs.

• The likelihood in each MH is computed by assuming an equal “prior” probability for each MH and marginalizing over all other parameters.

\[ P(\text{NH}) = \frac{0.5 \mathcal{L}_{\text{NH}}}{0.5 \mathcal{L}_{\text{NH}} + 0.5 \mathcal{L}_{\text{IH}}} = \frac{\mathcal{L}_{\text{NH}}}{\mathcal{L}_{\text{NH}} + \mathcal{L}_{\text{IH}}} \]

• The marginal likelihood is computed by numerical integration.
  – The likelihood value is the average over throws of the nuisance parameters.

• This method gives consistent results with the Bayesian analysis.
Conclusions

• The nature of the oscillation probabilities makes quantifying discovery inherently challenging:
  – Physics parameters are variously bounded, cyclical, degenerate, binary

• Low statistics plus important systematics further complicate things.

• The quintessential problem: Mass Hierarchy
  – NOvA – Feldman-Cousins
  – T2K – Bayesian
  – SK – CLS (next talk)

• Discussion:
  – Is one of these MH approaches “right?”
  – How would we know?
  – Any way to compare across methods?
• Use the ND $\nu_\mu$ sample to predict the FD $\nu_\mu$ sample.
• Use the ND $\nu_\mu$ sample to predict the FD $\nu_\mu$ sample.
• Use the ND $\nu_\mu$ sample to predict the FD $\nu_e$ signal.
• Use the ND $\nu_\mu$ sample to predict the FD $\nu_\mu$ sample.
• Use the ND $\nu_\mu$ sample to predict the FD $\nu_e$ signal.
• Use the ND $\nu_e$-like sample to predict the FD $\nu_e$ backgrounds.
Predicting the SK Spectrum

Flux Model

Cross Section Model

Flux Constraints
NA61/SHINE, etc.

Xsec Constraints
MiniBooNE, etc.
Predicting the SK Spectrum

 Flux Model

 Cross Section Model

 Flux Constraints

 Xsec Constraints

 Fit to ND280 $\nu_\mu$ CC Data

 Number of events

 Data/MC

 Muon momentum [MeV/c]

 MC NEUT nominal
 MC Post-fit
 Data

 Alex Himmelf
Predicting the SK Spectrum

Flux Model

Cross Section

Flux Constraints

Xsec Constraints

Fit to ND280 $\nu_\mu$ CC Data

Number of events

Data/MC

Muon momentum [MeV/c]

Flux and Cross Section Parameters
Reduced errors, new correlations

Alex Himmel
Predicting the SK Spectrum

Fit to ND280 $\nu_\mu$ CC Data

SK Prediction

Unoscillated
Oscillated

Flux and Cross Section Parameters
Reduced errors, new correlations