Efficient Neutrino Oscillation Parameter Inference with Gaussian Process

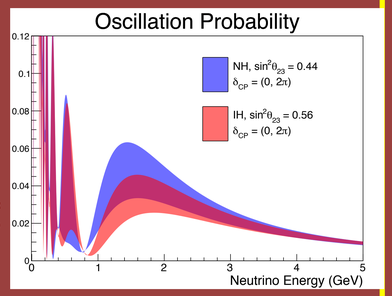
Lingge Li, Nitish Nayak, Jianming Bian, Pierre Baldi
University of California, Irvine

1. Neutrino Oscillations

- Neutrino oscillation proves that neutrinos have mass and that the neutrino mass eigenstates are not their flavor eigenstates. For 3-generation SM neutrinos, the transformation of mass eigenstates into favor eigenstates is parameterized by the PMNS matrix:

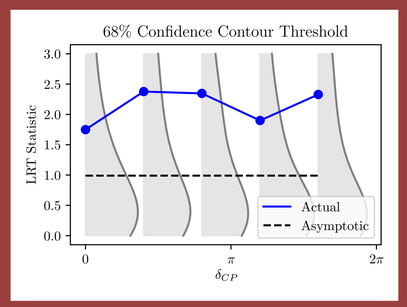
$$U_{PMNS} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{23} & \sin\theta_{23} \\ 0 & -\sin\theta_{23} & \cos\theta_{23} \end{pmatrix} \begin{pmatrix} \cos\theta_{13} & 0 & \sin\theta_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -\sin\theta_{13}e^{i\delta} & 0 & \cos\theta_{13} \end{pmatrix} \begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} & 0 \\ -\sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Neutrino experiments measure oscillation probabilities vs. neutrino energy at given propagation distance to infer parameters in U_{PMNS} and neutrino mass-squared splitting Δm^2_{32} and Δm^2_{21}
- Long-baseline (LBL) neutrino oscillations can solve the CP phase δ, the mass hierarchy Δm^2_{32}>0 (NH) or <0 (IH), and the octant of θ_{23}



2. Statistical Issues and Feldman-Cousins (FC)

- Neutrino oscillation parameters are typically measured via Maximum Likelihood Estimation (MLE) using the underlying PMNS model and comparing it to observation such as energy spectrum
- Due to the low statistics and physical boundaries, confidence intervals are hard to find as Likelihood Ratios (LR) = -2log[$L(H_0)/L(H_1)$] are not asymptotic χ^2-distributions.



Feldman-Cousins approach:

- Explicitly simulate LR distributions with pseudo-experiments at each point in parameter space and compute p-value
- A grid search over the entire parameter space with many toy MC - time consuming

3. Gaussian Process to Optimize Confidence Interval Search

- A Gaussian Process (GP) is a collection of random variables, any finite number of which have joint Gaussian distributions. Each draw from a GP distribution is a function with values *f*(x)
- A GP can be completely specified by its mean function μ*(x)* and its kernel function *k(x,x')*

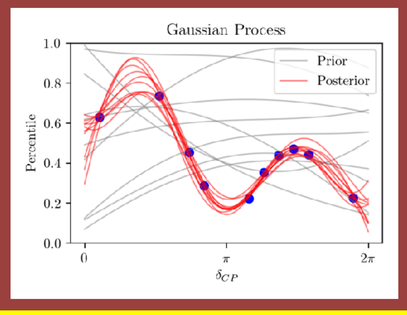
$$\begin{pmatrix} f(x) \\ f(x') \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(x) \\ \mu(x') \end{bmatrix}, \begin{bmatrix} k(x,x) & k(x,x') \\ k(x,x') & k(x',x') \end{bmatrix} \right)$$

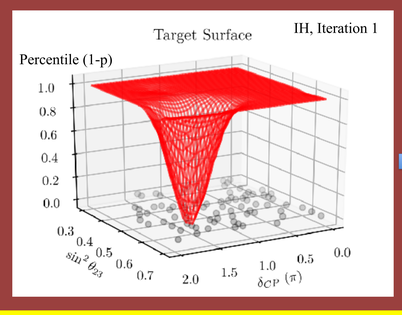
- A GP describes a distribution over functions. Use Bayes' rule, GP function distributions can be predicted based on points with known function values *f(x)*

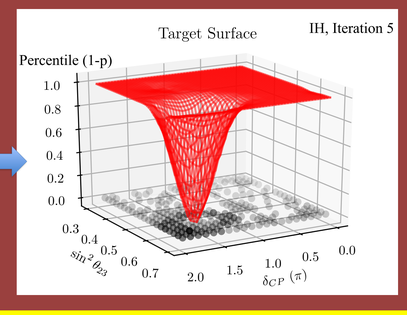
$$f(x')|f(x) \sim \mathcal{N}\left(\frac{k(x,x')}{k(x,x)} f(x), k(x',x') - \frac{k(x,x')^2}{k(x,x)}\right)$$

- Update GP p-value functions iteratively based on points where p-values have been calculated by a small number of pseudo-experiments. Use GP mean p-value to identify new points lie on boundary of interval to generate new pseudo-experiments.

- A GP is uniquely characterized by its kernel function *k*. A common choice of GP kernal *k* is the squared exponential radial basis function (RBF), $k(x_1, x_2) = \exp(-[(x_1-x_2)/l]^2)$ with a length-scale *l*. This kernel tells us that GP results at nearby points are highly influenced by observations at a given point while further out, they aren't.
- In each iteration, kernel hyperparameter *l* is learned via maximizing the likelihood of current set of observations marginalized over the function distribution *f(x)*
- Note for each point in parameter space we still use frequentist approach to determine p-value

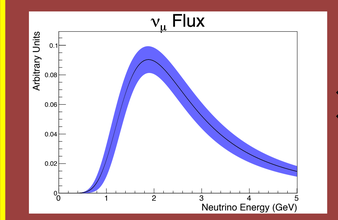




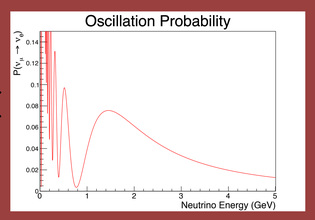


4. Toy Experiment

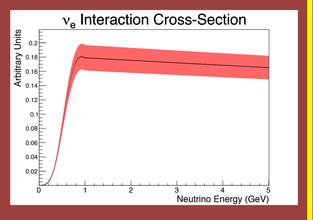
- Pseudo-experiments modelled on NOvA: Baseline L = 810km with NOvA flux peaking at 2GeV; Oscillation parameters similar to 2018 best estimate from NOvA (θ_{23} = 0.56, Δm^2_{32} = 2.44 x 10^{-3} eV^2, δ_{CP} = 1.5π).
- Neutrino energy spectrum simulated by multiplying toy shapes for flux, cross-section and oscillation probability
- Appearance $\nu_\mu \rightarrow \nu_e$ and disappearance $\nu_\mu \rightarrow \nu_\mu$ are simulated



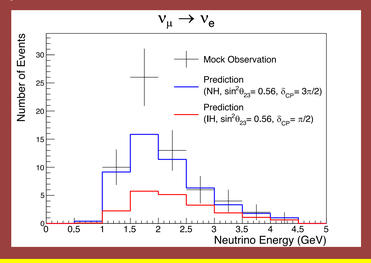
×

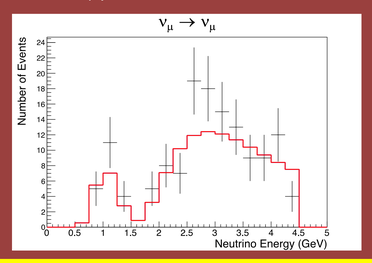


×



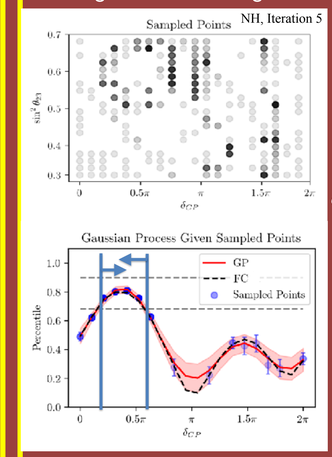
- Toy data generated from Poisson variations at some chosen oscillation parameters
- Best-fit found by minimizing negative log-likelihood over energy bins
- Fitting parameters includes oscillation and nuisance (flux and xsec errors) parameters





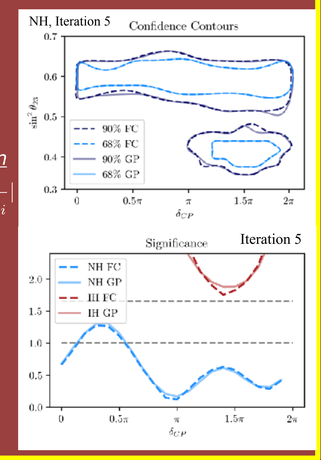
5. Results

- Use acquisition function to determine where to throw pseudo-experiments in parameter space (θ) : $\hat{q}(\theta)$ is GP mean, $\sigma_{\hat{q}(\theta)}$ is GP variance, α_i are confidence levels 68% and 90%
- Finding FC contour edges and δ_{CP} rejection regions with much less experiments



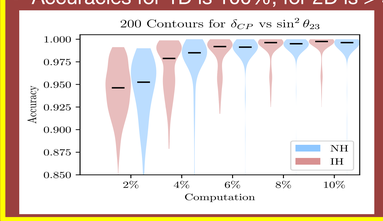
acquisition function

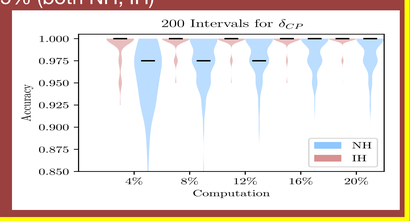
$$a(\theta) = \sum_{\alpha_i} \left| \frac{\sigma_{\hat{q}(\theta)}}{\hat{q}(\theta) - \alpha_i} \right|$$



6. Accuracy and Speed

- GP method converges to the FC value 10x faster for 2D case and 5x for 1D case
- Estimate classification accuracy of all grid points, taking FC result as truth
- Accuracies for 1D is 100%, for 2D is > 99.5% (both NH, IH)





DOI: 10.1103/PhysRevD.101.012001

Lingge Li, linggel@uci.edu
Nitish Nayak, nayakb@uci.edu
Jianming Bian, bianjm@uci.edu
Pierre Baldi, pfbaldi@ics.uci.edu