Efficient Neutrino Oscillation Parameter Inference with Gaussian Process

Lingge Li, Nitish Nayak, Jianming Bian, Pierre Baldi
University of California, Irvine
Irvine, CA 92617
linggel@uci.edu

Introduction

Neutrinos are tiny sub-atomic particles that carry no electrical charge and interact with matter only through the weak nuclear force, which makes them extremely hard to detect. There are three distinct types of neutrinos, called "flavors": $(\nu_e, \nu_\mu, \nu_\tau)$, each of which can "oscillate" into the other with a detectable probability. Many experiments (Abe and others 2015) (Adamson and others 2016) have been set-up to measure the parameters governing the oscillation probabilities accurately, with implications for the fundamental structure of the universe. Very often, this involves inferences from tiny samples of data which have complicated dependencies on multiple oscillation parameters simultaneously. This is typically carried out using the unified approach of Feldman and Cousins (Feldman and Cousins 1998) which is very computationally expensive, on the order of tens of millions of CPU hours. In this work, we propose an iterative method using Gaussian Process to efficiently find a confidence contour for the oscillation parameters and show that it produces the same results at a fraction of the computation cost. To our knowledge, the most similar existing work is using a Gaussian Process surrogate in the approximate Bayesian computation framework (Meeds and Welling 2014) but it may not achieve desired frequentist coverage.

Oscillation Parameter Inference

The probability of $\nu_\mu \to \nu_e$ oscillations, $P(\nu_\mu \to \nu_e)$ is determined by $\theta = (\Delta m^2_{32}, \sin^2\theta_{23}, \delta_{CP})$. To infer θ, a typical neutrino oscillation experiment sends a beam of ν_μ neutrinos into a detector and observes a handful of oscillated ν_e neutrinos from their interactions with the detector. The observed neutrinos are binned by their energy as the oscillation probability is a function of energy. For each energy bin $i \in I$, there is an observed neutrino count x_i and an expected count λ_i, which naturally gives rise to a *Poisson* distribution. However, for a given θ, the expectation λ is also influenced by the underlying model of the experiment, such as the beam configuration and the physics of neutrino interactions. Each of these have their own associated uncertainties, δ. Since the relationship between λ and (θ, δ) is not available analytically, λ is deduced from Monte Carlo. Denote the implicit mapping between λ and (θ, δ) by v. The log-likelihood of (θ, δ) is given by:

$$\log L(\theta, \delta) = \sum_{i \in I} \log Pois(x_i; v(\theta, \delta)_i) - \frac{1}{2}\delta^2 \quad (1)$$

where $-\frac{1}{2}\delta^2$ is a penalty term for systematic error. The best fit parameters $(\hat{\theta}, \hat{\delta})$ can be obtained by maximizing the log-likelihood but a confidence contour is also needed to quantify the statistical uncertainty in $\hat{\theta}$. This can be done by an inverted likelihood ratio test (LRT) as a particular value θ_0 should only be included in the $1-\alpha$ confidence contour if we fail to reject the null $\theta = \theta_0$ at the α level. However, the asymptotic χ^2 distribution of the likelihood ratio statistic D is unreliable when the sample size of observed neutrinos is small. Moreover, the distribution of D can vary drastically as a function of θ due to the complexity of v. Therefore, for a given θ_0, Monte Carlo experiments are used to simulate the LRT critical value $c(\theta_0)$. Since the parameter space is bounded, this is done in a grid for a large number of θ values. Known as the Feldman-Cousins method in high energy physics, this LRT based approach has high statistical power by the Neyman-Pearson lemma but exhausting the grid is computationally inefficient.

Gaussian Process Iterative Method

With applications to hyper-parameter tuning (Snoek, Larochelle, and Adams 2012), Bayesian optimization can be used to find the optimum of any black-box function f that is expensive to evaluate. Bayesian optimization is an iterative procedure. In each iteration, a number of points are evaluated to update an approximation of f. Then based on the approximation, the points in the next iteration are proposed by an acquisition function a. The approximation model is usually a zero-mean Gaussian process $\mathcal{GP}(0, \kappa(\cdot,\cdot))$. A $\mathcal{GP}$ assumes any finite collection of points to be jointly Normal with covariance matrix Σ, which is parametrized by a kernel function κ that defines pairwise covariance. Given observed data, a $\mathcal{GP}$ model can be fitted by optimizing parameters of the kernel. The acquisition function a aims to balance between "exploration," reducing approximation uncertainty, and "exploitation," reaching the optimum.

In our context, the expensive black-box function is the mapping of LRT critical values $c(\theta)$. What we want to find is

the confidence contour, the set of points where the observed likelihood ratio statistic D_X is equal to the critical value. The acquisition function is given by

$$a(\theta) = |\frac{\hat{c}(\theta) - D_X}{\sigma_{\hat{c}(\theta)}}|^{-1} \quad (2)$$

where $\hat{c}(\theta)$ is the $\mathcal{GP}$ approximated critical value at θ and $\sigma_{\hat{c}(\theta)}$ is the $\mathcal{GP}$ standard deviation at θ. Iteratively, the $\mathcal{GP}$ model will seek those points for which it is unsure about being within the confidence contour. The other points would be either included or rejected with some certainty. Figure 1 illustrates one iteration of the proposed method. With more iterations, the certainty will increase so that the approximated confidence contour converges to the one produced by a full grid search.

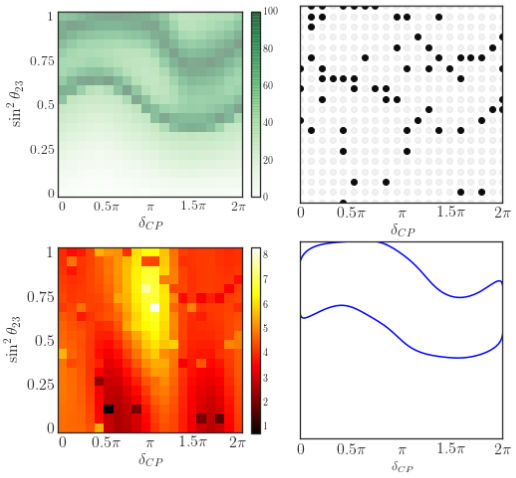

Figure 1: Top (from left to right): percentile of acquisition function and points explored on the grid. Bottom: approximated critical value and confidence contour at the 68% level.

Results

We set up a toy experiment with an oscillated ν_e signal binned uniformly between $0.5 - 4.5$ GeV and no backgrounds. Toy uncertainties for the ν_μ beam configuration and the ν_e interaction probability are included in the nuisance parameters, δ. More details are given in the supplement. With Δm^2_{32} treated as a nuisance parameter as well, the confidence contour of interest is of $\sin^2\theta_{23}, \delta_{CP}$ in the range of $(0, 1) \times (0, 2\pi)$.

400 points on an evenly-spaced grid are used to find the "true" confidence contour using standard Feldman-Cousins method, with 1000 Monte Carlo experiments performed at each point. For the proposed method, 20 points are proposed in each iteration. For evaluation, we compare the proposed method and standard Feldman-Cousins method by calculating the percentage of overlap between the respective confidence contours (set of points)

$$1 - \frac{1}{n}\sum_{i=1}^{n}(I_{FC}(\theta_i) - I_{GP}(\theta_i)) \quad (3)$$

where I_{FC} and I_{GP} are indicator functions for whether θ_i is included in the confidence contour. The same comparison is performed with typical post-hoc smoothing (of confidence contour) as well. As shown in Figure 2, on 200 simulated data sets, the proposed method produces the same results as standard Feldman-Cousins method using 30% of the computation budget. In high energy physics and other fields, there are a wide range of similar inference problems and the proposed method can be applied to them as well.

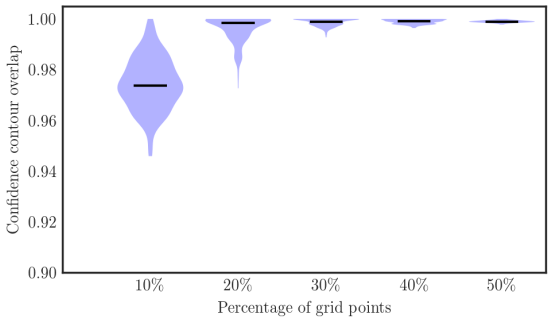

Figure 2: Distributions of contour overlap when different percentages of points on the grid are used. The red horizontal lines indicate the median.

References

Abe, K., et al. 2015. Measurements of neutrino oscillation in appearance and disappearance channels by the T2K experiment with 6.6E20 protons on target. *Physical Review D* 91(7):072010.

Adamson, P., et al. 2016. First measurement of electron neutrino appearance in NOvA. *Physical Review Letters* 116(15):151806.

Feldman, G. J., and Cousins, R. D. 1998. Unified approach to the classical statistical analysis of small signals. *Physical Review D* 57(7):3873.

Meeds, E., and Welling, M. 2014. Gps-abc: Gaussian process surrogate approximate bayesian computation. *arXiv preprint arXiv:1401.2838*.

Snoek, J.; Larochelle, H.; and Adams, R. P. 2012. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, 2951–2959.