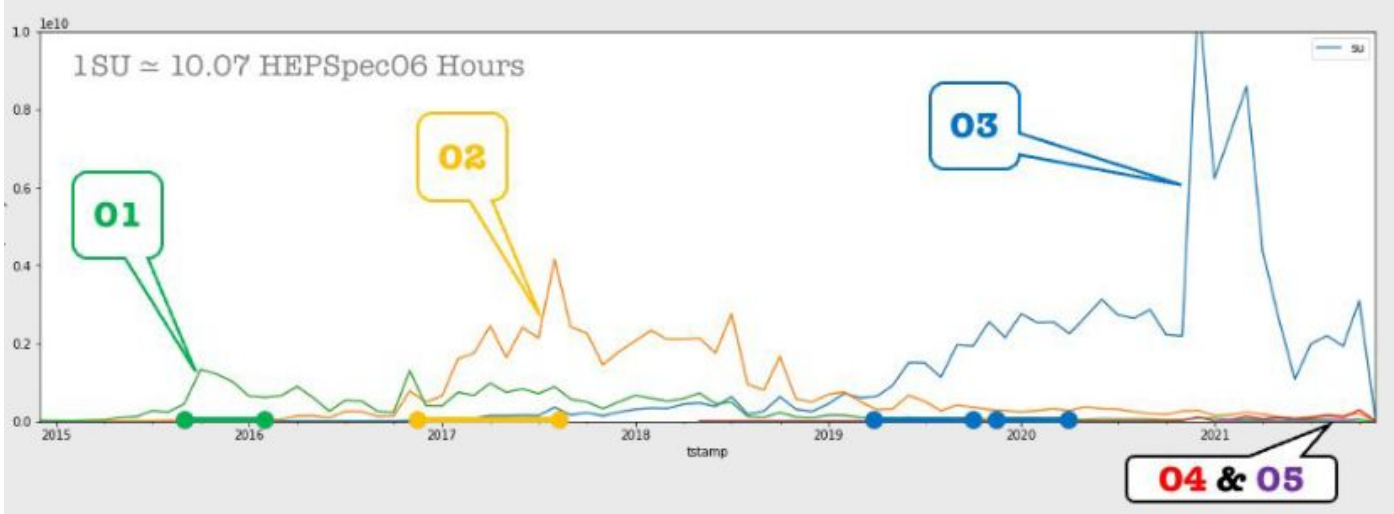
Parameter Estimation of Un-modeled Gravitational-wave Signals using Likelihood-free Inference

Deep Chatterjee, Ethan Marx
W. Benoit, A. Gunny, E. Katsavounidis, P. Harris, M. Coughlin,E. Moreno, R. Raikman, K. Govorkova

subMIT IAP Workshop
Feb 2, 2024

Motivation



Credits
LIGO-T2200209
(Circa 2022)

- Computing cost for searching, inferring parameters, cataloging is increasing with every LVK observing run.
- This is true for most other experiments.
- *Real-time inference is important that ever.*

Parameter Estimation

In experimental physics/astronomy:

- We have a model
  - In the case of gravitational waves, it is waveforms from general relativity.
- We collect data
  - For GWs, this is the strain.
- We have a search that tells us if a segment of data is interesting
  - Can be archival (offline) or real-time (online)
  - In GWs we have modeled searches based on GR waveforms
  - Or unmodeled searches
- Parameter estimation - measure parameters and confidence intervals
  - Traditionally done using stochastic sampling - MCMC, nested sampling,...
  - Roughly this involves: Propose a point in param space, *generate model*, *compute likelihood*, accept/reject point, propose new point,...repeat,...

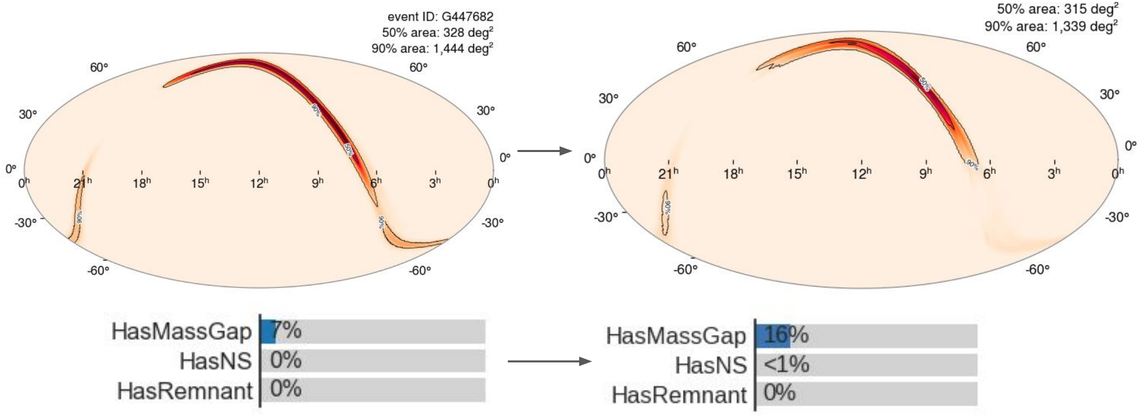
Parameter Estimation

Posterior → $p(\Theta|\mathbf{d}) \propto p(\mathbf{d}|\Theta)\, p(\Theta)$

Likelihood ↑ Prior ↑

- Stochastic sampling involves repeated computation of the likelihood.
- Several methods developed to compute the likelihood efficiently. Ex. ROQ, SVD, Heterodyning. Currently online parameter estimation involves focused ROQ
- Current (O4) parameter estimation times ~ couple of hours
- Comparison to O3: several hours to a day.
  - Update to Sky localization and source properties

S231020ba



Not in the ~second regime yet; May not scale to the next gen. instruments

Likelihood-free Inference

An entirely different approach

We have simulations to generate data

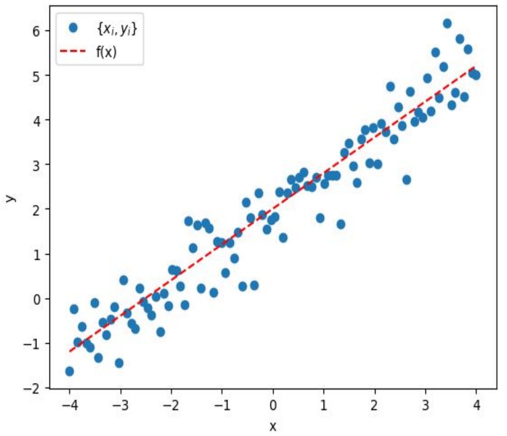
$$\boldsymbol{\Theta}_i \sim p(\boldsymbol{\Theta}) \longrightarrow h(\boldsymbol{\Theta}_\mathbf{i}) + n \rightarrow \mathbf{d}_i$$

Detector noise

We have combinations of $\{\boldsymbol{\Theta}_i, \mathbf{d}_i\}$

In principle we have $p(\boldsymbol{\Theta}, \mathbf{d})$

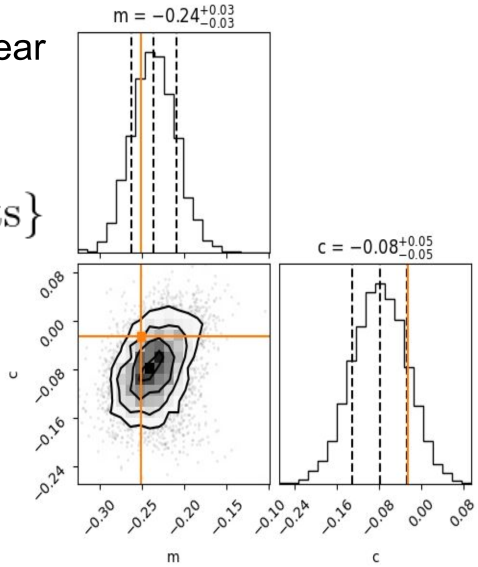
And hence, can get $p(\boldsymbol{\Theta}|\mathbf{d})$ or $p(\mathbf{d}|\boldsymbol{\Theta})$



Simple Example: Linear Reg.

$$\boldsymbol{\Theta} = \{m, c\}$$

$$\mathbf{d} = \{\text{Set of points}\}$$



Challenge: Distributions may be complex

Solution: To find transforms to obtain a simpler distribution.

Normalizing Flows

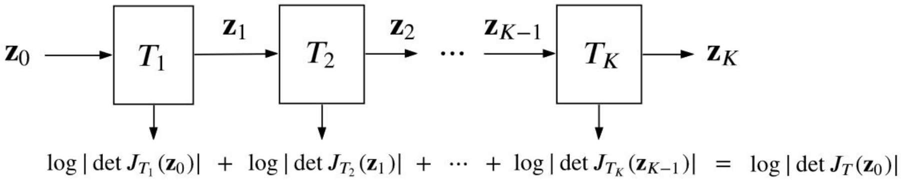
- Have two components:
  - **Simpler, base distribution**, like a normal dist. - easy to sample and density est.
  - **Transform** - A neural network that learns to transform complex variable to base distribution

$$\mathbf{x} = T(\mathbf{u}) \quad \text{where} \quad \mathbf{u} \sim p_{\mathrm{u}}(\mathbf{u}).$$

$$p_{\mathrm{x}}(\mathbf{x}) = p_{\mathrm{u}}(\mathbf{u}) \; |\det J_T(\mathbf{u})|^{-1} \quad \text{where} \quad \mathbf{u} = T^{-1}(\mathbf{x}).$$

- Trained by maximizing likelihood
  - This is easy since base distribution is easy to density estimate from
  - Transform is fast to evaluate*

Details about Normalizing Flow



- Practical details involve how the transform is implemented
  - Affine transforms
  - Splines
  - Integral based
  - …
- Autoregressive property ensures efficient determinant computation
  - Transforms are lower triang.; determinant equal product of diag. elem.

$$p(z_1, z_2, z_3|d) = p(z_1|z_2, z_3, d)\ p(z_2|z_3, d)\ p(z_3|d)$$

- We use a masked autoregressive flow - affine transform between layers

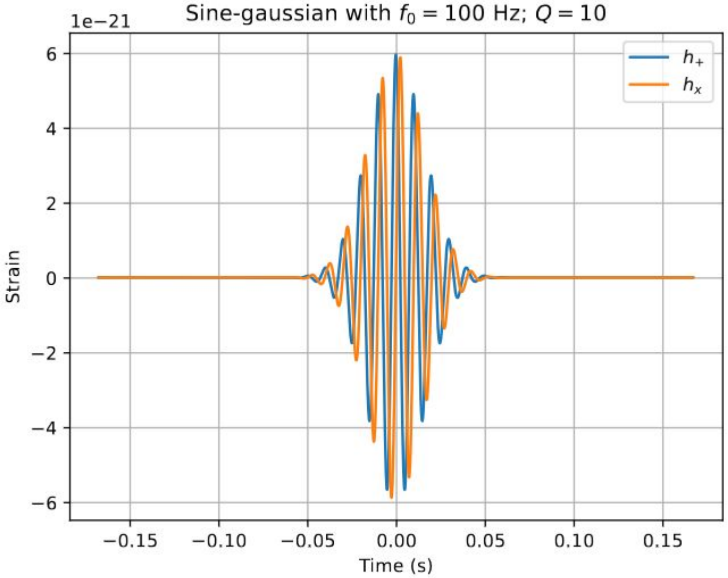
Sine Gaussian Parameter Estimation

- Unmodeled sources; minimal assumptions
- Sine-gaussian forms a wavelet basis
    - Extraction of characteristic amplitude and freq.

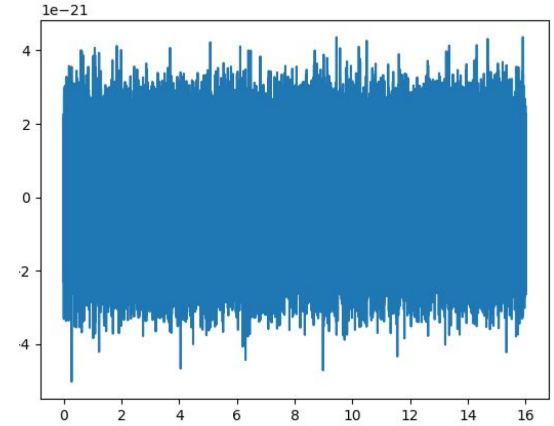
$$h_+(t) \propto cos(2\pi f_0(t - t_0) + \phi_0)e^{(-t-t_0)^2/\tau^2}$$

$$h_\times(t) \propto sin(2\pi f_0(t - t_0) + \phi_0)e^{(-t-t_0)^2/\tau^2}$$

	Parameter	Prior	Limits
intrinsic	Frequency	Uniform	(32, 1024)
	Quality	Uniform	(2, 108)
	hrss	Log Uniform	(1e-23, 1e-19)
	Phase ϕ	Uniform	(0, 2π)
	Eccentricity *e*	Uniform	(0, 1)
extrinsic	Declination δ	Cosine	(0, π)
	Right Ascension	Uniform	(0, 2π)
	Ψ	Uniform	(0, π)

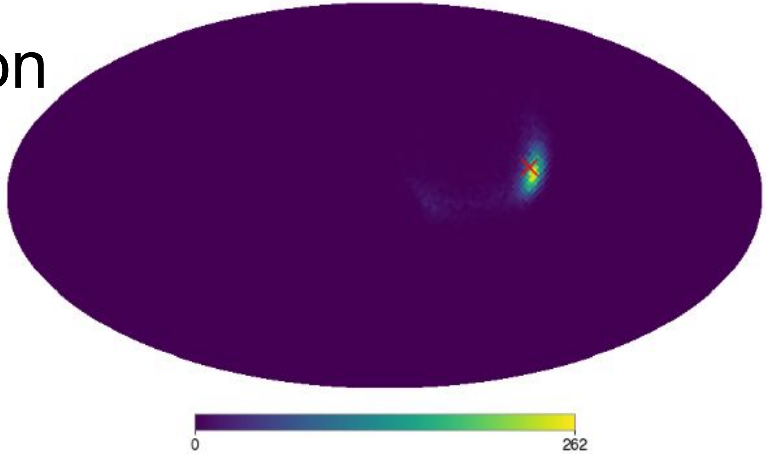


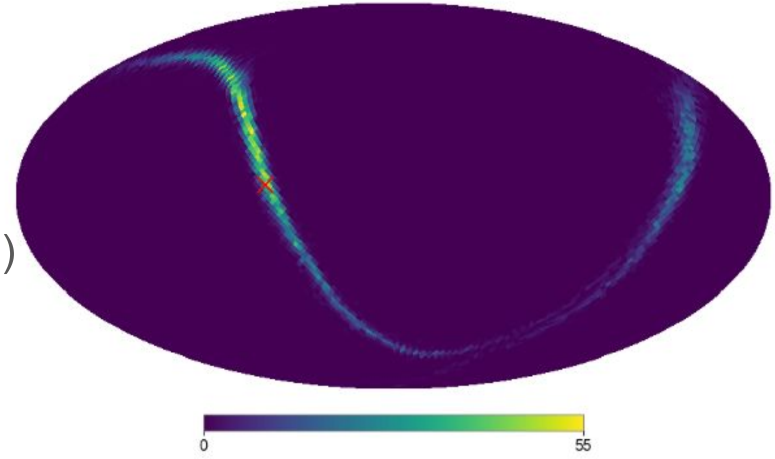
+



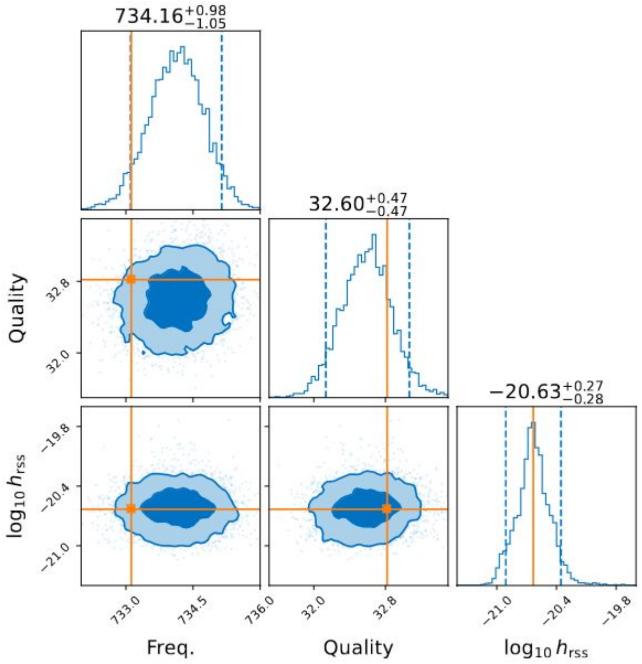
Sine Gaussian Parameter Estimation

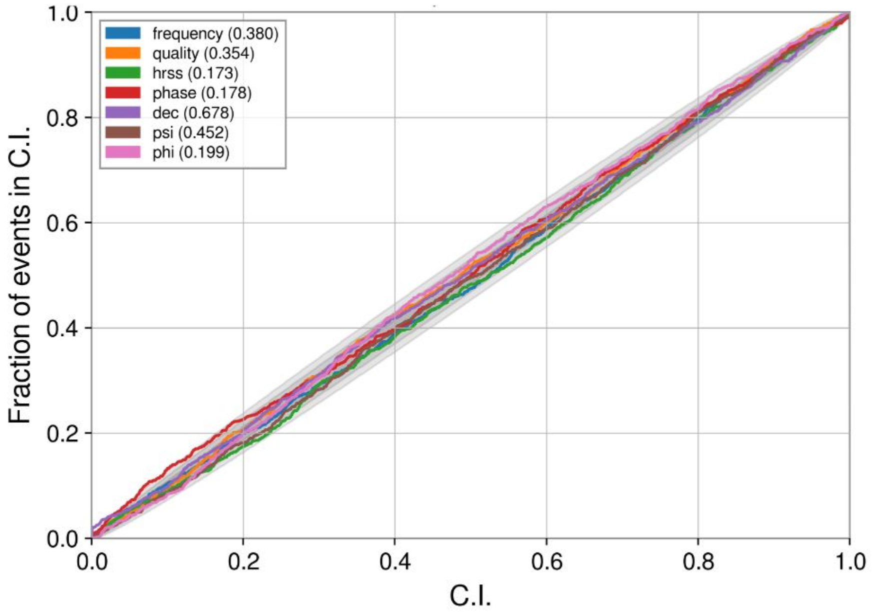
- 7 parameter model: strain amp., *central freq., quality*, coal. phase, polarization, *RA, Dec*.
- Signals generation done via torch implementation of sine-gaussian model.
- Injected into a stretch of real noise instance from GW detectors - 1 sec. duration @ 4KHz
- Multiple detector data represented as different channels.
- Data is Embedded coherently into 128 summary features via learnable ResNet.
- Model ~ 10^5 parameters. Training time ~ O(few hrs) on NVIDIA-A30 (submit-gpu partition)
- Sampling 10^3 injections takes ~ 10 minutes; less than 1 second for each injection. (*Paper in prep.*)





Example Posteriors and PP plots

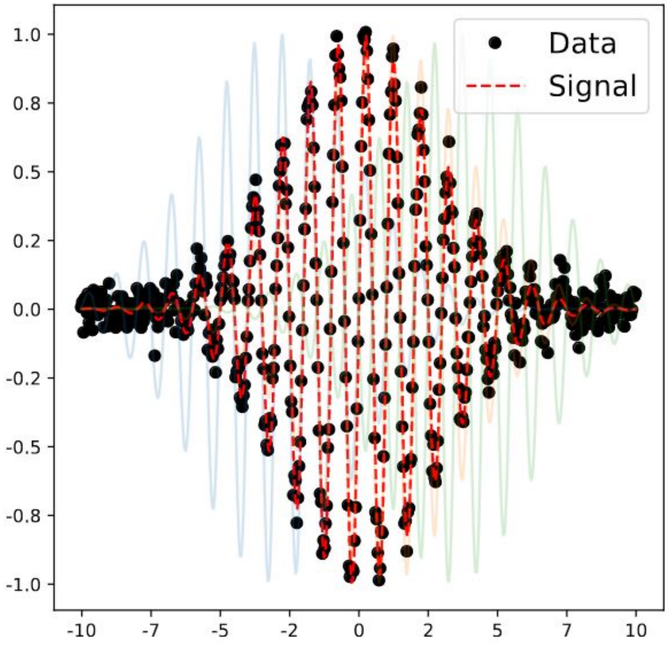




WIP: Extension to Binary BHs

Incorporating Symmetries/Marginalizing nuisance params.

- Some parameters of the problem may not be important
  - Exact time of arrival of a pulse may not be as interesting as properties of the pulse
- Nuisance parameters can be hard to learn, and then marginalize over
  - NN treats same intrinsic properties, but different time of arrivals as *unique data.*
  - Requires large networks
  - More compute time

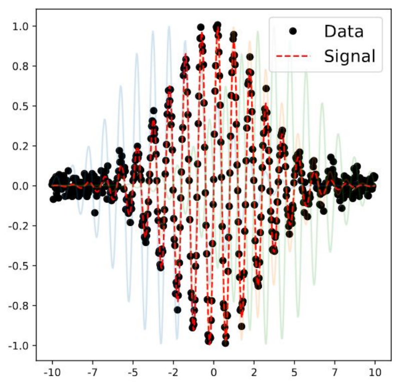


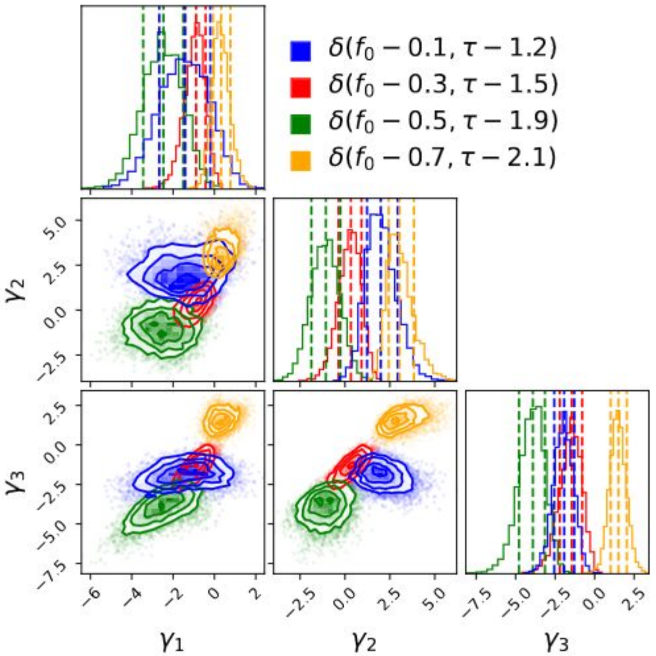
Use of self-supervision

- Marginalize out parameters by joint embedding
    - A batch of data with a fixed reference time of arrival
    - A second batch where the time of arrival is varied
    - Embed the data and use a similarity loss between the batches. We use VICReg loss.

$$\mathcal{L}_{\mathrm{VICReg}}(x, x') = \lambda_1 \ \mathrm{MSE}(x, x') + \lambda_2 \ [\mathrm{Var}(x) + \mathrm{Var}(x')] + \lambda_3 \ [\mathrm{Cov}(x) + \mathrm{Cov}(x')] \, ,$$

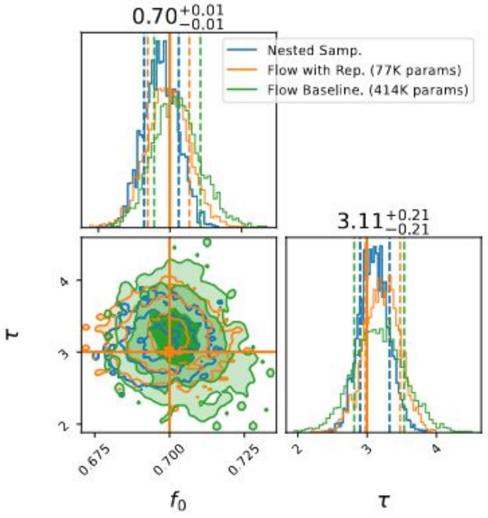
-
    - Use the embedded space as a data summary.
    - Condition parameters on this summary.

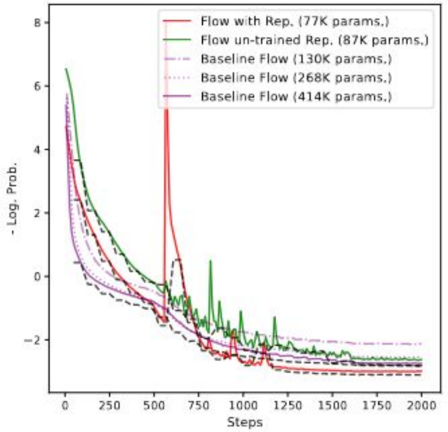




Optimizing LFI using Self-supervision

- Symmetry informed embeddings lead to faster convergence in a smaller number of parameters
- Comparable results as nested sampling consistent with Cramer-Rao bounds.
- Technique used build summary stats marginalizing over parameters we don’t desire.
- Work accepted for NeurIPS ML4PS 2023.
- Technique is broadly applicable across astronomy and physics.





WIP: Application to parameter estimation on lightcurves

Work being done by Malina Desai