Non-Parametric Normalizing Flow Modeling of Binary Black Hole Populations for Unbiased Dark Siren Cosmology

Leonardo Iampieri, Simone Mastrogiovanni

Paris Workshop - 21/05/2025

Istituto Nazionale di Fisica Nucleare

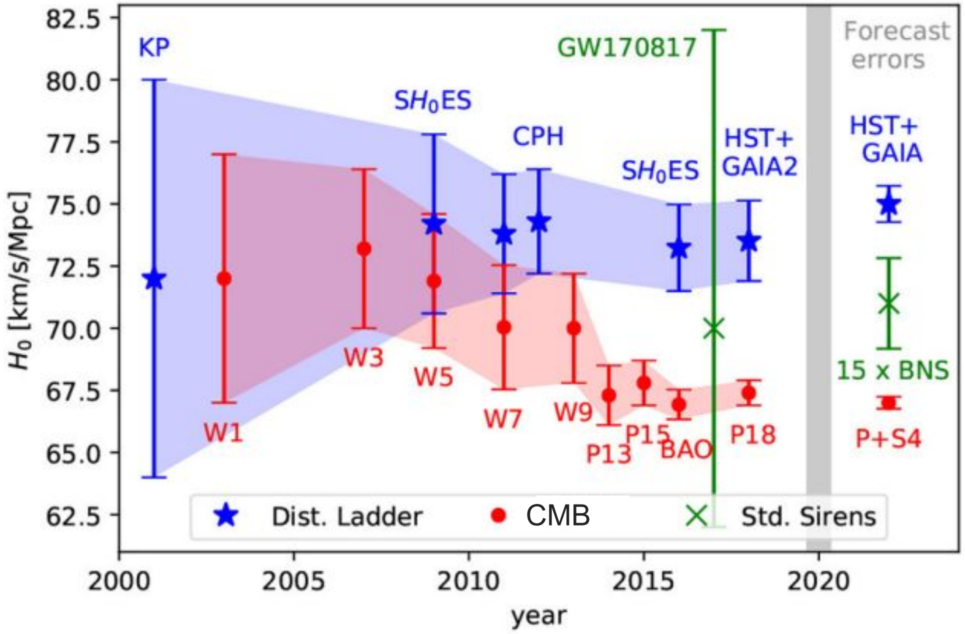
Sapienza Università di Roma

Hubble tension

Discrepancy in the measured values of the **Hubble Constant (H0)**:

- Cosmic Microwave Background (early universe) - Planck experiment.
- Standard candles (late universe) - SH0ES experiment.

- **Discrepancy:** ~5σ → suggests new physics or unrecognized systematics
- **Independent probe needed:** Gravitational Waves (GWs) offer a ladder-free distance measure.



[J. Ezquiaga+ Front. Astron. Space Sci., 21 December 2018]

Gravitational Wave Cosmology

- GW sources are **standard sirens:** we can derive the luminosity distance directly from the GW signal.

- The luminosity distance is given by:

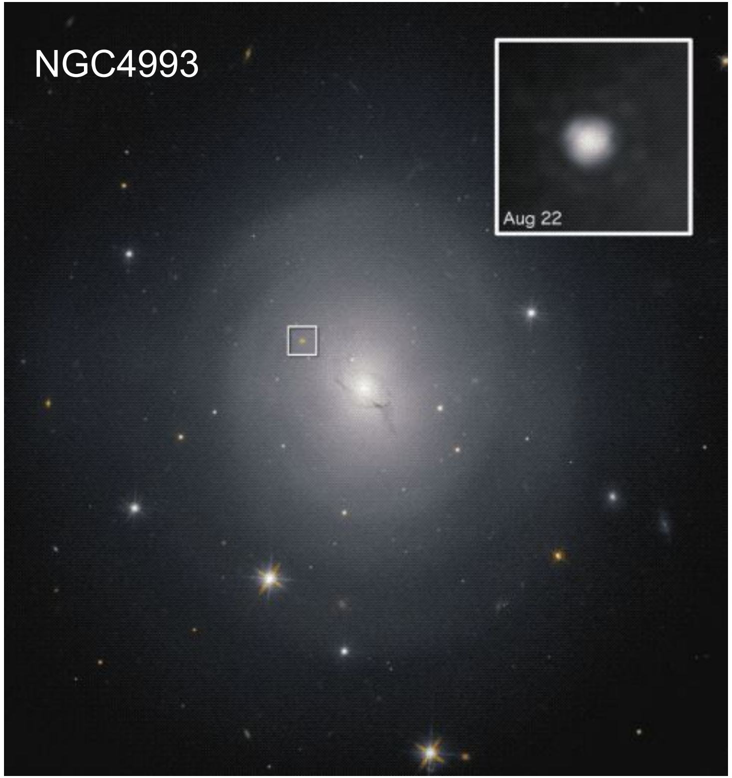
Luminosity distance

$$d_L(z) = \frac{c(1+z)}{H_0} \int_0^z \frac{dz'}{E(z)}$$

Hubble constant

- By identifying the host galaxy, we recover spectroscopic measurement of redshift (example case: **GW170817**).

- In absence of Electromagnetic counterpart, we can infer z statistically from the Binary Black Holes (BBH) population.



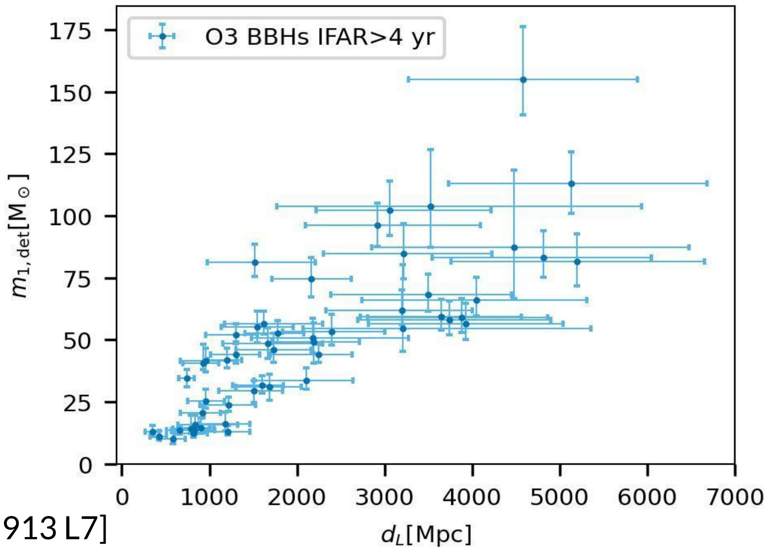
[B. P. Abbott *et al* Nature volume 551 (2017)]

Dark Sirens - Mass Method

- **Core idea:** Infer source redshift z by comparing observed detector-frame BBH masses to the source-frame mass distribution [Mastrogiovanni+, PRD 104 (2021)]

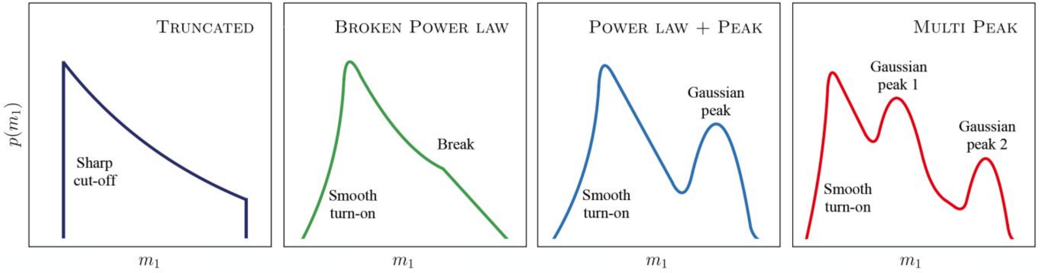
$$m_{1,\mathrm{det}} = m_{1,\mathrm{s}}(1+z)$$

- Inference on z, and therefore H_0, is highly sensitive to the choice of source–mass model
- **Mass-redshift correlation:** The source mass distribution may evolve with redshift



[LVK+ 2021 ApJL 913 L7]

- Parametric Mass-redshift distributions can biases H0 inferences when the assumed functional forms are incorrect [Pierra+, PRD 109 (2024)]



Normalizing Flow

- **Objective**: represent the mass-redshift distribution with a flexible model: a **normalizing flow**.

- Normalizing Flows (NFs) transform a density using a chain of smooth, invertible mappings $f : \mathbb{R}^t \to \mathbb{R}^t$ such that, $\boldsymbol{\theta}_K = f_K \circ \cdots \circ f_2 \circ f_1(\boldsymbol{\theta}_0)$.
- the parameters $\boldsymbol{\theta}_K$ represent the source masses and redshift.

- We parametrize the chain using a set of weights **w**. Exploiting invertibility, we can construct the log-likelihood

$$\log p(\boldsymbol{\theta}_K|\mathbf{w}) = \log p_0(\boldsymbol{\theta}_0) - \sum_{k=1}^{K} \log \det \left| \frac{\partial f_k^{\mathbf{w}}}{\partial \boldsymbol{\theta}_k} \right|$$

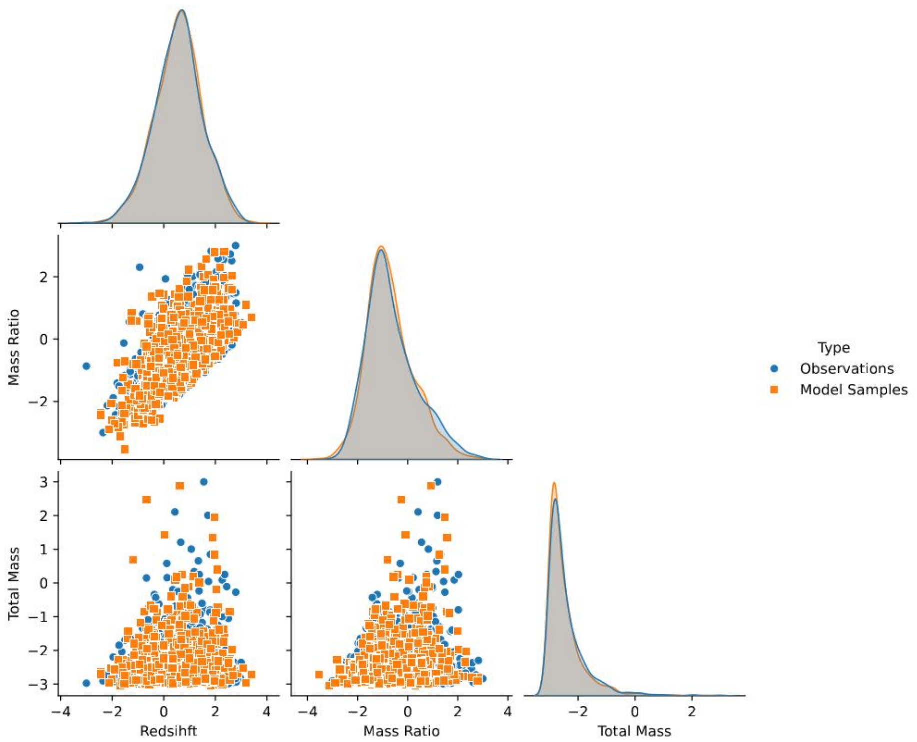
- We optimize the log-likelihood for **w** using stochastic gradient descent by maximizing the posterior distribution on **w**.

Example Scenario

- We generated catalog of 10^3 observations with:
  - m1 and m2 drawn from a Powerlaw+Peak model
  - z drawn from a Madau distribution
  - no measurement uncertainty and selection effects.
- Loss function:

$$\mathcal{L}(w) = -\frac{1}{N}\sum_{i=1}^{N}\log p(\theta_{K,i} \mid w)$$

- NF Architecture: Affine Neural Spline Flow implemented in pytorch [Zuko package]



Hierarchical Bayesian Analysis (HBA)

- **HBA scheme:** the posterior on the weights **w** informed from a catalog of C Gravitational Wave (GW) Observations $\mathcal{D}_C := \{\mathcal{D}_i\}_{i \in C}$ is given by

Number of observed GW events

$$p(w \mid \{\vec{d}_i\}) = \frac{p(w)}{p(\{\vec{d}_i\})} p(\{\vec{d}_i\} \mid w) = \frac{p(w)}{p(\{\vec{d}_i\})} \prod_{i=1}^{N_{\rm obs}} p(\vec{d}_i \mid w)$$

GW Likelihood

Population Model

$$= \frac{p(w)}{p(\{\vec{d}_i\})} \prod_{j=i}^{N_{\rm obs}} \frac{\int p(\vec{d}_i \mid \theta)\, p(\theta \mid w)\, \mathrm{d}\theta}{\int p_{\rm det}(\theta)\, p(\theta \mid w)\, \mathrm{d}\theta}$$

Detection Probability

Hierarchical Likelihood Evaluation - 1

- To evaluate the **numerator** in the **Hierarchical Likelihood** (HL) we generate a set of posterior samples

$$\theta_{i,j} \sim p(\theta \mid \vec{d}_i) = \frac{p(\vec{d}_i \mid \theta)}{\pi_{\mathrm{PE}}(\theta)}$$

- The integral at the numerator is computed through a Monte Carlo integration over the posterior samples

$$\int p(\vec{d}_i \mid \theta)\, p(\theta \mid w)\, \mathrm{d}\theta \approx \frac{1}{N_{s,i}} \sum_{j=1}^{N_{s,i}} \frac{1}{\pi_{\mathrm{PE}}(\theta_{i,j})} p(\theta_{i,j} \mid w) = \frac{1}{N_{s,i}} \sum_{j=1}^{N_{s,i}} w_{i,j}$$

Number of samples

Prior used for Posterior samples

Hierarchical Likelihood Evaluation -2

- The integral at the **denominator** of the HL Expression corrects for selection biases.
- To compute this integral we use **"injections"**, i.e. Monte Carlo simulations of injected and detected events.
- We computed the integral through Monte Carlo integration over detected injections:

Number of detected events

$$I_{\text{norm}} \approx \frac{1}{N_{\text{gen}}} \sum_{i=1}^{N_{\text{det}}} p(\theta_i \mid w) \frac{1}{\pi_{\text{inj}}(\theta_i)} = \frac{1}{N_{\text{gen}}} \sum_{i=1}^{N_{\text{det}}} s_i$$

Number of injections generated (even not detected)

Prior used for injections

Loss Function

- In the presence of noise uncertainty and selection biases, the loss to minimize becomes

Number of observations

Number of detected injections

$$\mathcal{L}(w) = -\sum_{i=1}^{N_{\rm obs}} \log\left[\frac{1}{N_{s,i}}\sum_{j=1}^{N_{s,i}} \frac{p(\theta_{i,j} \mid w)}{\pi_{\rm PE}(\theta_{i,j})}\right] + N_{\rm obs} \log\left[\frac{1}{N_{\rm inj}}\sum_{k=1}^{N_{\rm det}} \frac{p(\theta_k \mid w)}{\pi_{\rm inj}(\theta_k)}\right]$$

Number of PE samples

Prior used for Posterior samples

Number of total injections

Prior used for injections

Example: Selection Bias

Correcting for selection biases is necessary for the NF to recover the true distribution.

Example: we generate 10^4 observation from a 2D Gaussian distribution with detection probability

$$P_{\text{det}}(x, y) = \begin{cases} 1 & x > 0 \\ \frac{1}{2} & x < 0 \end{cases}$$

Training cases:

- Fig. 1 - Uncorrected Loss
- Fig. 2 - Full Loss

Fig. 1

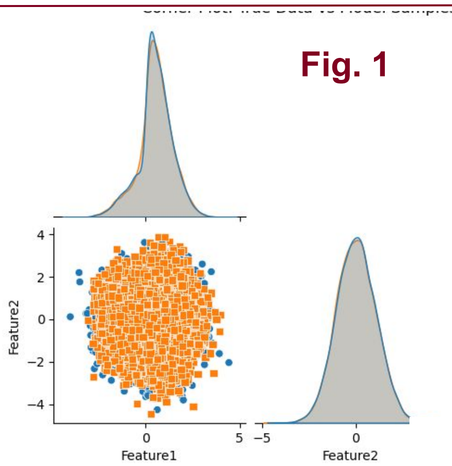
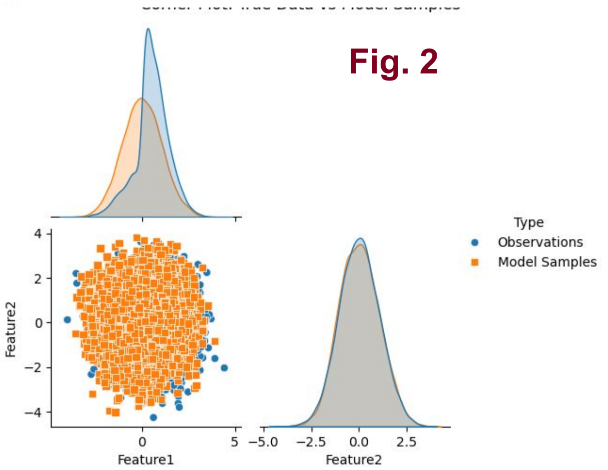


Fig. 2



Next Steps & Conclusions

- Normalizing Flow flexibly capture complex, multimodal distributions without need to set a fixed functional form
- We can define a single hierarchical loss to account for noisy uncertainty and selection biases.

- Next steps will involve:
  - **Realistic Simulations:**
    - Incorporate GW posterior samples and detector biases
    - Valide NF recovery in such scenarios
  - **Real Data Application**
    - Train on O3/O4 BBH Catalogs
    - Compare resulting distribution with current parametric fits