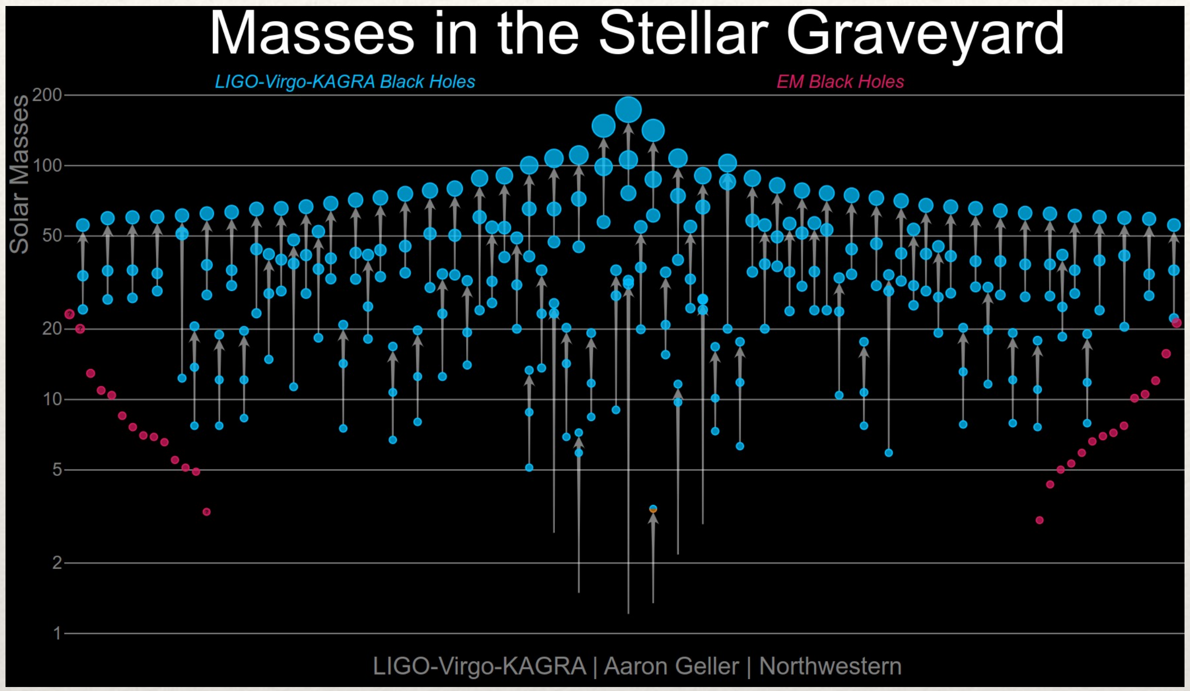
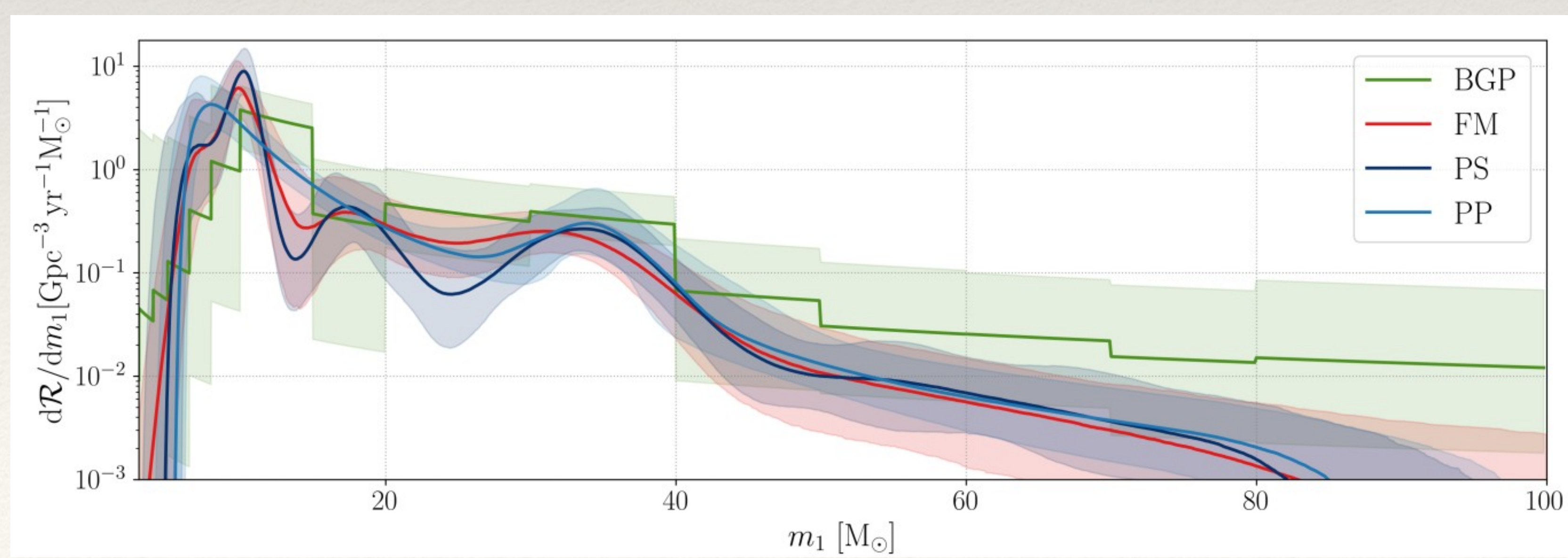
# Methods of population studies

*Rencontre du groupe "Méthodes d'analyse des données", 16/10/2024*

*Alexandre Toubiana* alexandre.toubiana@unimib.it

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA



# Individual events parameter estimation (PE)

- The parameters of a single event, $\theta$, are obtained through Bayes' theorem:

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)\pi_{\mathrm{PE}}(\theta)}{p(d)}$$

$\mathcal{L}(d|\theta)$ : single event likelihood

$\pi_{\mathrm{PE}}(\theta)$ : individual event prior

$p(d)$ : evidence

# Hierarchical Models: intuitive approach

- Consider a prior not only for a single data set, but also for a population of events, e.g., compact binary coalescences.
- The parameters of that prior encode the details of the population and are also of interest. This leads to the notion of a **hierarchical model**.
- In a hierarchical model, the parameters of the prior (termed **hyperparameters**) are regarded as random variables, on which a **hyperprior** is defined.

$$p(d|\Lambda) = \int p(d|\theta)p(\theta|\Lambda)\ \mathrm{d}\theta$$

$$p(\{d\}|\Lambda) = \prod_i p(d_i|\Lambda)$$

- Assumes events are statistically independent

$$p(\Lambda|\{d\}) = \frac{p(\{d\}|\Lambda)p(\Lambda)}{p(\{d\})}$$

# Hierarchical Models: intuitive approach

- No instrument is arbitrarily sensitive and therefore some types of source are easier to see than others. This is important to remember in hierarchical models for populations when we are combining only **detected** events.

- We can correct for this by acknowledging that we only include "detected" events in the analysis and then write down a likelihood for detected events. This must integrate to 1 over all "detected" or "above threshold" data sets.

$$\mathcal{L}(d|\Lambda, \mathrm{obs}) = \frac{p(d|\Lambda)}{p_s(\Lambda)}, \text{ where } p_s(\Lambda) = \int_{d>\mathrm{threshold}} \int p(d|\theta)p(\theta|\Lambda)\ \mathrm{d}\theta \mathrm{d}\Lambda$$

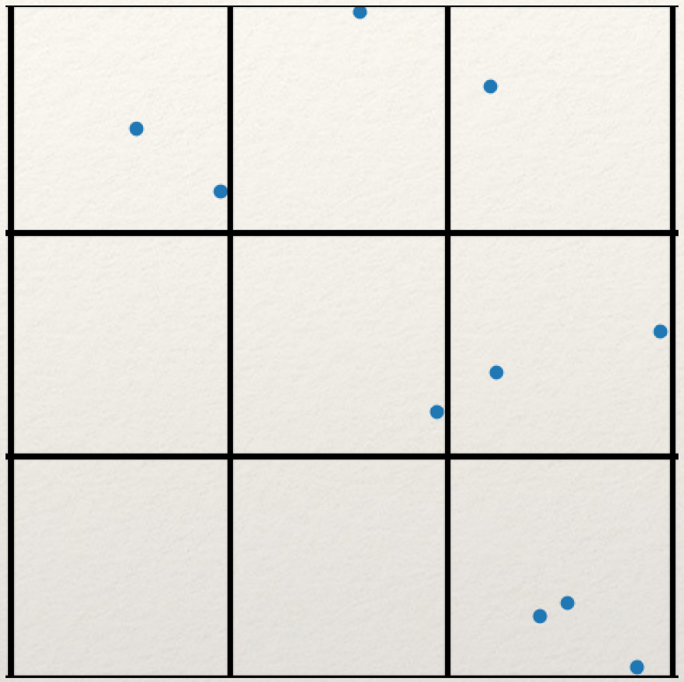
- This framework assumes a priori that the number of detected events contains no information about the parameters of interest.

- Obs: selection effects do not impact the parameter estimation of single events.

# More general approach

Inhomogeneous Poisson process for a given binning scheme:

$$p(\{\theta\}|\Lambda) = \prod_{k=1}^{N_{\rm bins}} \frac{N_k(\Lambda)^{n_k} \exp[-N_k(\Lambda)]}{n_k!}$$



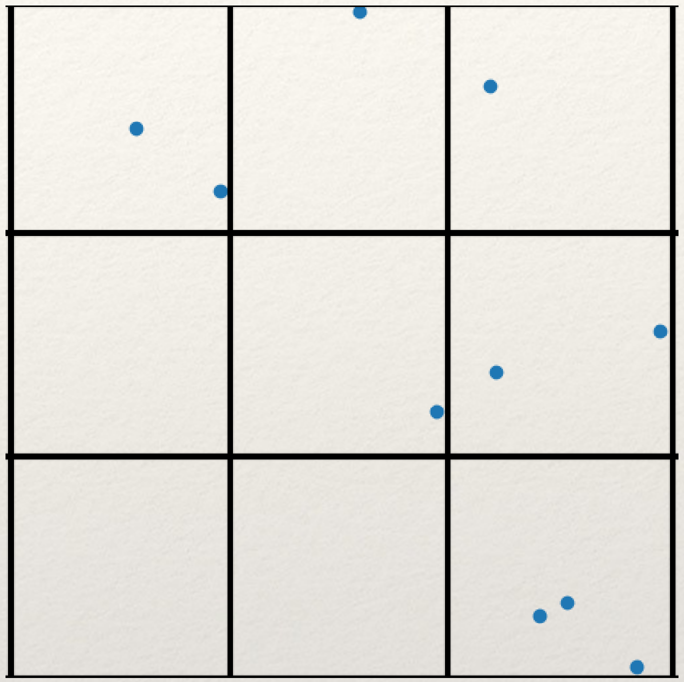
# More general approach

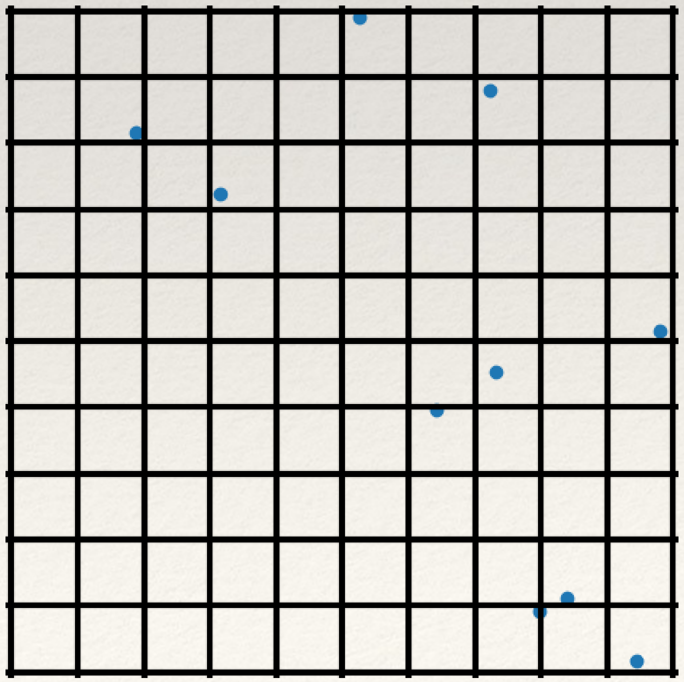
Inhomogeneous Poisson process for a given binning scheme:

$$p(\{\theta\}|\Lambda) = \prod_{k=1}^{N_{\rm bins}} \frac{N_k(\Lambda)^{n_k} \exp[-N_k(\Lambda)]}{n_k!}$$

Make bins infinitesimally small

$$n_k \to 0 \text{ or } 1$$

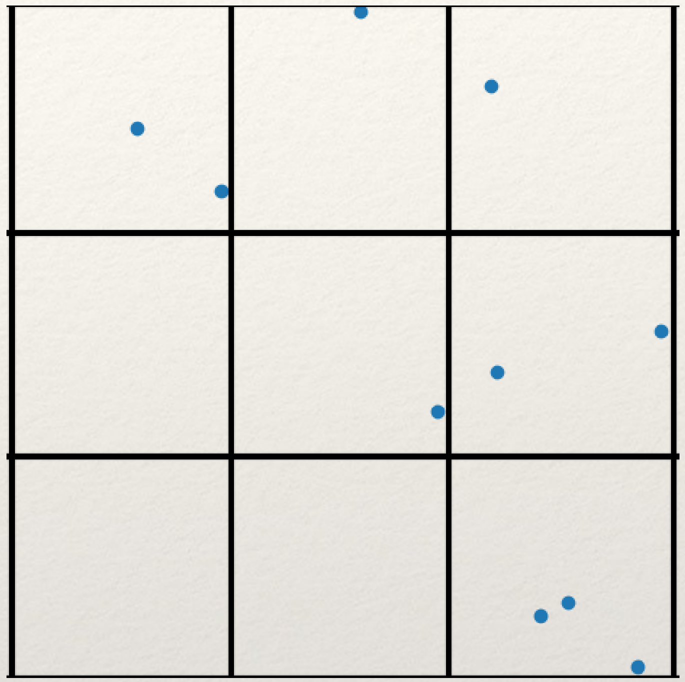




# More general approach

Inhomogeneous Poisson process for a given binning scheme:

$$p(\{\theta\}|\Lambda) = \prod_{k=1}^{N_{\mathrm{bins}}} \frac{N_k(\Lambda)^{n_k} \exp[-N_k(\Lambda)]}{n_k!}$$

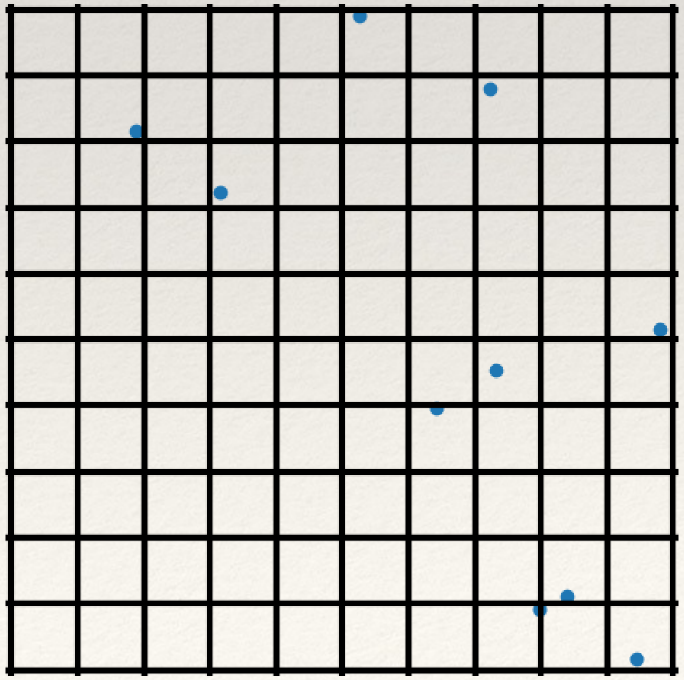


Make bins infinitesimally small

$$n_k \to 0 \text{ or } 1$$

$$p(\{\theta\}|\Lambda) = \exp[-N(\lambda)] \prod_{k=1}^{N_{\mathrm{events}}} \frac{\mathrm{d}N}{\mathrm{d}\theta_k}(\Lambda)$$

$$N(\Lambda) = \int \frac{\mathrm{d}N}{\mathrm{d}\theta}(\Lambda)\mathrm{d}\theta$$



# More general approach

- We write down the joint likelihood for all data/parameter pairs using an infinitesimal inhomogeneous Poisson process, including both **detected** events (indexed by $i$) and **undetected** events (indexed by $j$)

$$p(\{\theta_i\}, \{\theta_j\}, \{d_i\}, \{d_j\}|\Lambda) = \left[\prod_i^{\mathrm{N_{obs}}} \mathcal{L}(d_i|\theta_i)\frac{\mathrm{d}N}{\mathrm{d}\theta_i}(\Lambda)\right]\left[\prod_j^{\mathrm{N_{nobs}}} \mathcal{L}(d_j|\theta_j)\frac{\mathrm{d}N}{\mathrm{d}\theta_j}(\Lambda)\right]\exp[-N(\Lambda)]$$

# More general approach

- We write down the joint likelihood for all data/parameter pairs using an infinitesimal inhomogeneous Poisson process, including both **detected** events (indexed by *i*) and **undetected** events (indexed by *j*)

$$p(\{\theta_i\}, \{\theta_j\}, \{d_i\}, \{d_j\}|\Lambda) = \left[\prod_i^{\mathrm{N_{obs}}} \mathcal{L}(d_i|\theta_i)\frac{\mathrm{d}N}{\mathrm{d}\theta_i}(\Lambda)\right]\left[\prod_j^{\mathrm{N_{nobs}}} \mathcal{L}(d_j|\theta_j)\frac{\mathrm{d}N}{\mathrm{d}\theta_j}(\Lambda)\right]\exp[-N(\Lambda)]$$

- Marginalising over the unobserved data we obtain

$$p(\{\theta_i\}, \{d_i\}) = \left[\prod_i^{N_{\mathrm{obs}}} \mathcal{L}(d_i|\theta_i)\frac{\mathrm{d}N}{\mathrm{d}\theta_i}(\Lambda)\right]\frac{N_{\mathrm{ndet}}^{N_{\mathrm{nobs}}}}{N_{\mathrm{nobs}}!}\exp[-N(\Lambda)]$$

$$N_{\mathrm{ndet}}(\Lambda) = \int_{d<\mathrm{threshold}}\int \mathcal{L}(d|\theta)\frac{\mathrm{d}N}{\mathrm{d}\theta}(\Lambda)\ \mathrm{d}\theta\mathrm{d}d$$

# More general approach

- We write down the joint likelihood for all data/parameter pairs using an infinitesimal inhomogeneous Poisson process, including both **detected** events (indexed by *i*) and **undetected** events (indexed by *j*)

$$p(\{\theta_i\},\{\theta_j\},\{d_i\},\{d_j\}|\Lambda) = \left[\prod_i^{\mathrm{N_{obs}}} \mathcal{L}(d_i|\theta_i)\frac{\mathrm{d}N}{\mathrm{d}\theta_i}(\Lambda)\right]\left[\prod_j^{\mathrm{N_{nobs}}} \mathcal{L}(d_j|\theta_j)\frac{\mathrm{d}N}{\mathrm{d}\theta_j}(\Lambda)\right]\exp[-N(\Lambda)]$$

- Marginalising over the unobserved data we obtain

$$p(\{\theta_i\},\{d_i\}) = \left[\prod_i^{N_{\mathrm{obs}}} \mathcal{L}(d_i|\theta_i)\frac{\mathrm{d}N}{\mathrm{d}\theta_i}(\Lambda)\right]\frac{N_{\mathrm{ndet}}^{N_{\mathrm{nobs}}}}{N_{\mathrm{nobs}}!}\exp[-N(\Lambda)]$$

$$N_{\mathrm{ndet}}(\Lambda) = \int_{d<\mathrm{threshold}}\int \mathcal{L}(d|\theta)\frac{\mathrm{d}N}{\mathrm{d}\theta}(\Lambda)\ \mathrm{d}\theta \mathrm{d}d$$

- Marginalising over the unknown number of unobserved events then gives

$$p(\{\theta_i\},\{d_i\}) = \left[\prod_i^{N_{\mathrm{obs}}} \mathcal{L}(d_i|\theta_i)\frac{\mathrm{d}N}{\mathrm{d}\theta_i}(\Lambda)\right]\exp[-N_{\mathrm{det}}(\Lambda)]$$

# More general approach

- Writing

$$\frac{\mathrm{d}N}{\mathrm{d}\theta} = N(\Lambda)p(\theta|\Lambda)$$

- introducing a scale-invariant prior on the overall rate

$$p(N) \propto \frac{1}{N}$$

- and noting

$$N_{\mathrm{det}}(\Lambda) = \int_{d>\mathrm{threshold}} \int \mathcal{L}(d|\theta)\frac{\mathrm{d}N}{\mathrm{d}\theta}()\ \mathrm{d}\theta \mathrm{d}d$$

- After marginalising over $\{\theta_i\}$, we recover the result with the "intuitive approach".

# Hierarchical Bayesian analysis

- Including rates:

$$p(\Lambda|\{d\}) \propto p(\Lambda) N_{\rm det}(\Lambda)^{N_{\rm obs}} \exp[-N_{\rm det}(\Lambda)] \prod_i^{N_{\rm obs}} \int \mathcal{L}(d_i|\theta) p(\theta|\Lambda) \; {\rm d}\theta$$

- Marginalising over the rate:

$$p(\Lambda|\{d\}) \propto \frac{p(\Lambda)}{p_s(\Lambda)^{N_{\rm obs}}} \prod_i^{N_{\rm obs}} \int \mathcal{L}(d_i|\theta) p(\theta|\Lambda) \; {\rm d}\theta$$

- We often note $p_{\rm pop}(\theta|\Lambda)$ the population prior

Assumes events are statistically independent

# Practical implementation

- Evidences: $$\int \mathcal{L}(d|\theta)p(\theta|\Lambda)\ \mathrm{d}\theta = p(d)\int \frac{p(\theta|d)p_{\mathrm{pop}}(\theta|\Lambda)}{\pi_{\mathrm{PE}}(\theta)}\ \mathrm{d}\theta$$
$$\simeq \frac{p(d)}{N_s}\sum_{\theta_i\sim p(\theta|d)}\frac{p(\theta_i|\Lambda)}{\pi_{PE}(\theta_i)}$$

- Selection: $$p_s(\Lambda) = \int\int_{d>\mathrm{threshold}} \mathcal{L}(d|\theta)p_{\mathrm{pop}}(\theta|\Lambda)\ \mathrm{d}\theta\mathrm{d}d$$
$$= \int p_{\mathrm{det}}(\theta)p_{\mathrm{pop}}(\theta|\Lambda)\ \mathrm{d}\theta$$
$$\simeq \frac{1}{N_{\mathrm{inj}}}\sum_{\theta_i,\mathrm{det}}\frac{p_{\mathrm{pop}}(\theta_i|\Lambda)}{\pi_{\mathrm{inj}}(\theta_i)}$$

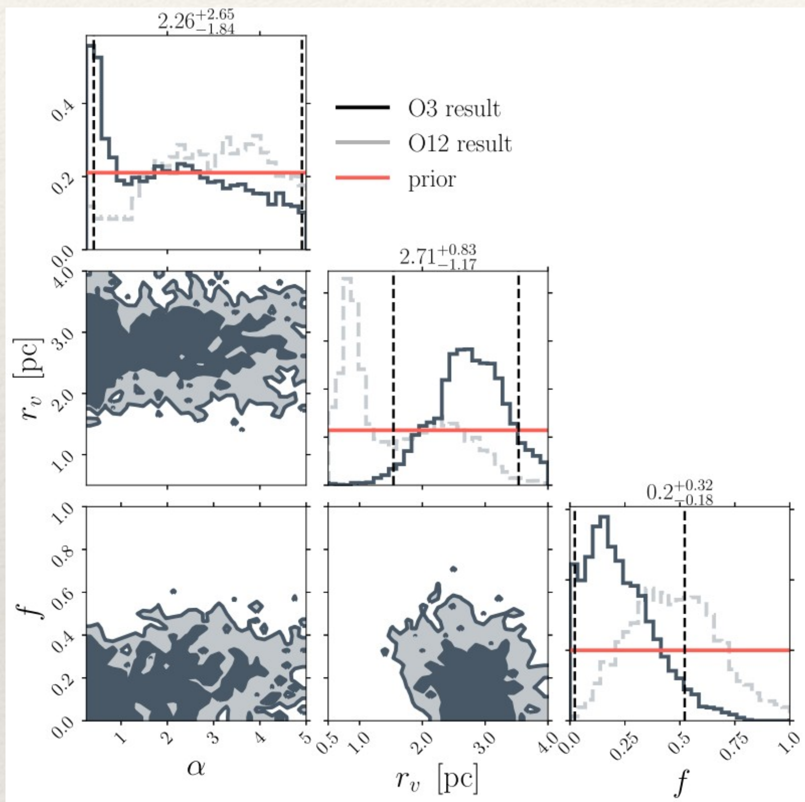
# Population inference

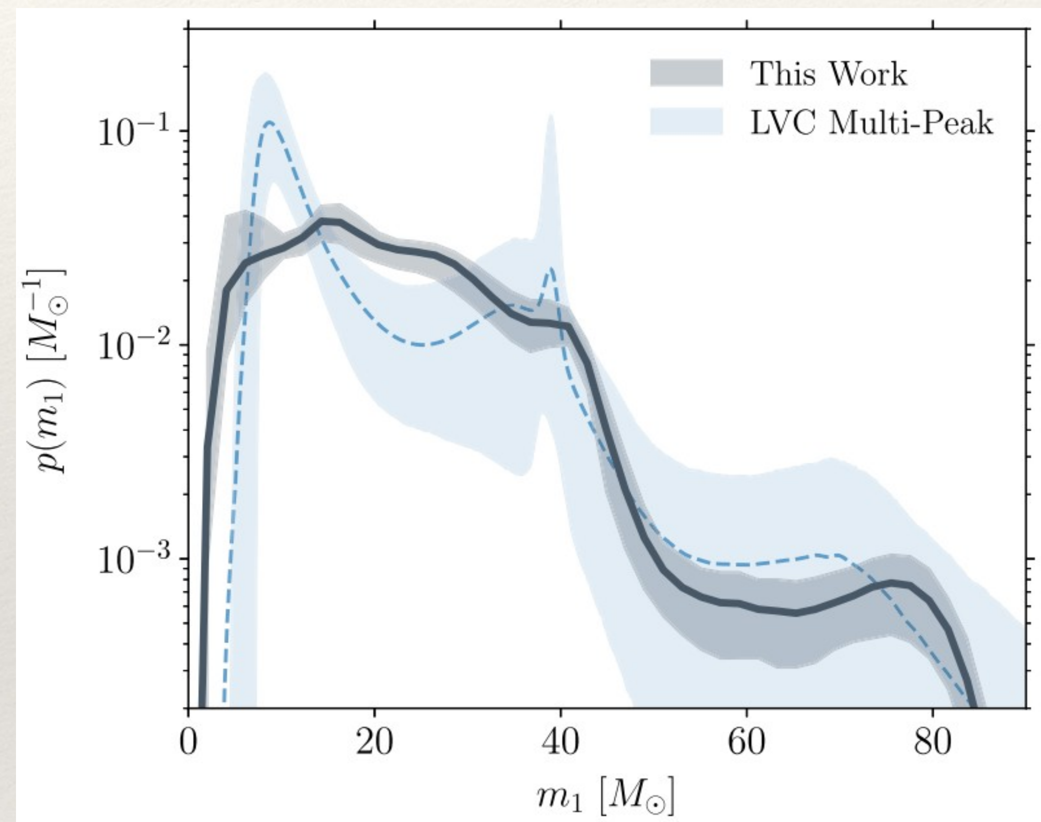
- Assume a form for $p_{\rm pop}(m_1, m_2, \chi_1, \chi_2, z|\Lambda)$
- Can be:
  - astrophysical
  - parametric
  - non-parametric
- Obtain $p(\Lambda|\{d\})$ accounting for selection effects and measurement uncertainty performing a hierarchical Bayesian analysis

# Astrophysical model

- Results on GWTC-2 from Wong et al., PRD 2021

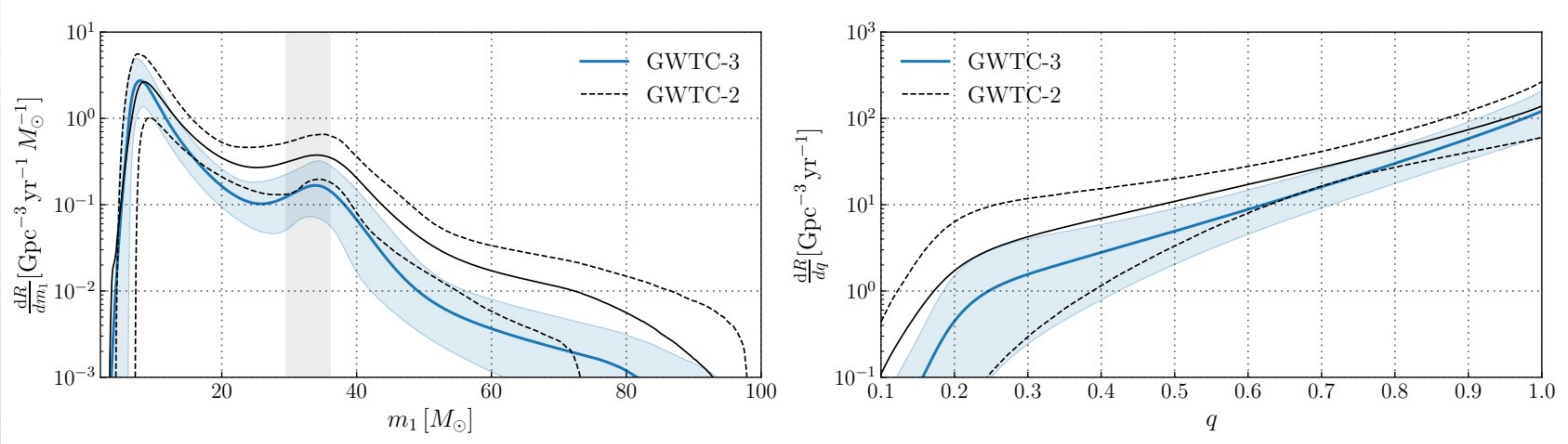
$$p_{\rm pop}(m_1, m_2, z|\alpha, r_v, f) = f p_{\rm iso}(m_1, m_2, z|\alpha) + (1-f) p_{\rm dyn}(m_1, m_2, z|r_v)$$




- See Zevin et al., APJ 2021 for more formation channels et Mould et al., PRD 2022 for results on GWTC-3
- **Pros:**
  - Gives direct information on astrophysical processes
- **Cons:**
  - Huge uncertainty on astro models. We might not include all channels (see Cheng et al., 2307.03129, Raikman et al., 2310.10736)
  - Requires some way to evaluate pdf from samples and to interpolate (see Toubiana et al., PRD 2021 for systematic errors)

# Parametric model

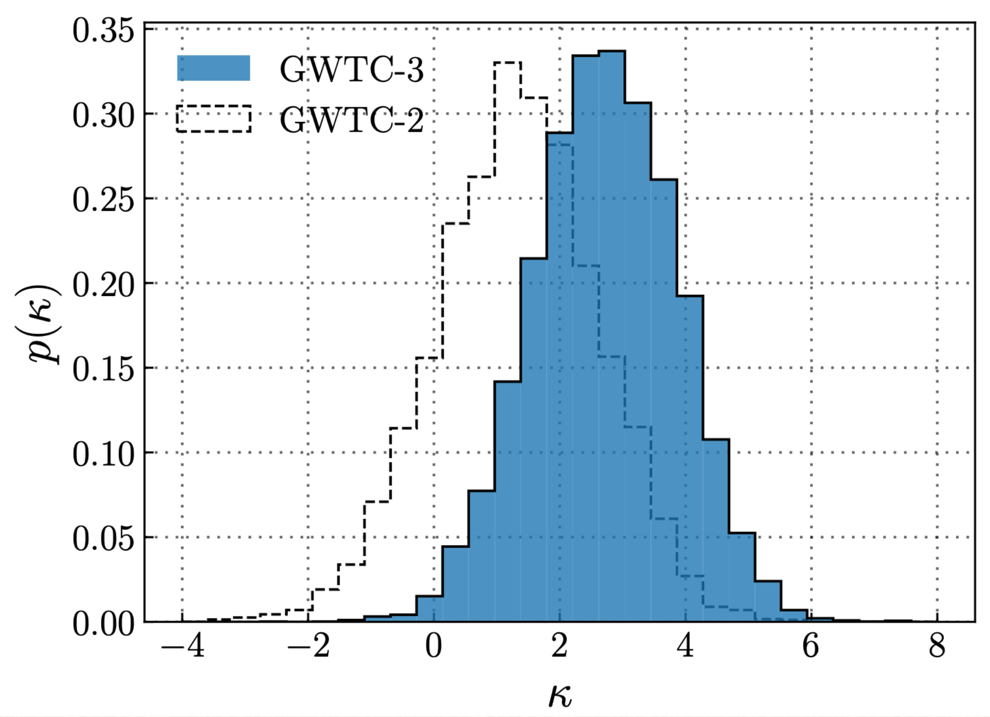



LVK, PRX 2021

$$p_{\rm pop}(m_1|\Lambda) = [\lambda \mathrm{G}(m_1,\mu,\sigma) + (1-\lambda)\mathrm{PL}(m_1, m_{min}, m_{max}, \alpha)]S(m_1, m_{min}, \delta_m)$$

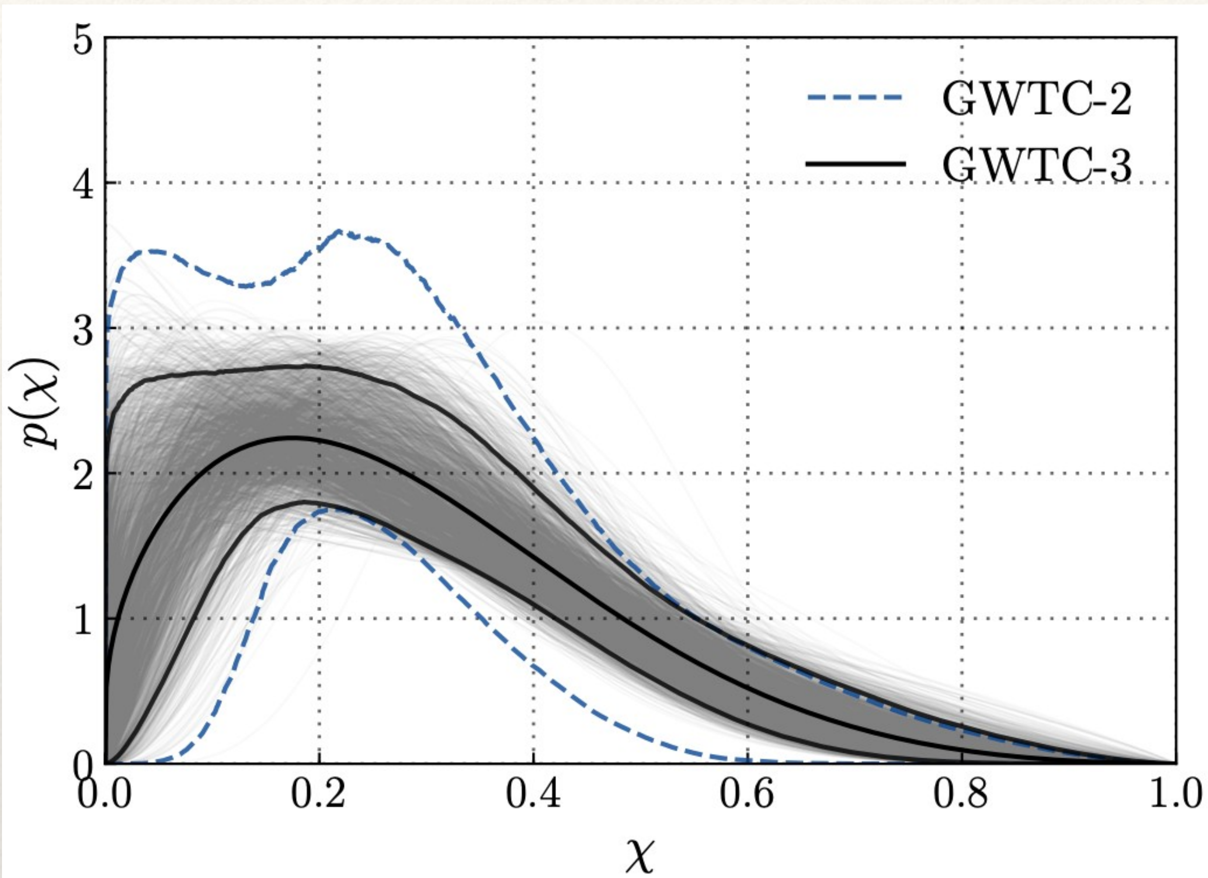
$$p_{\rm pop}(q|m_1,\Lambda) \propto q^{\beta} S(qm_1, m_{min}, \delta_m)$$

$$R(z) = R_0(1+z)^{\kappa}$$




- GWTC-3 shows evidence for a peak at $\sim 35 M_\odot$, a bit low for the pair-instability gap
- Some evidence for for rate evolution, with the rate higher in the past.

# Parametric model

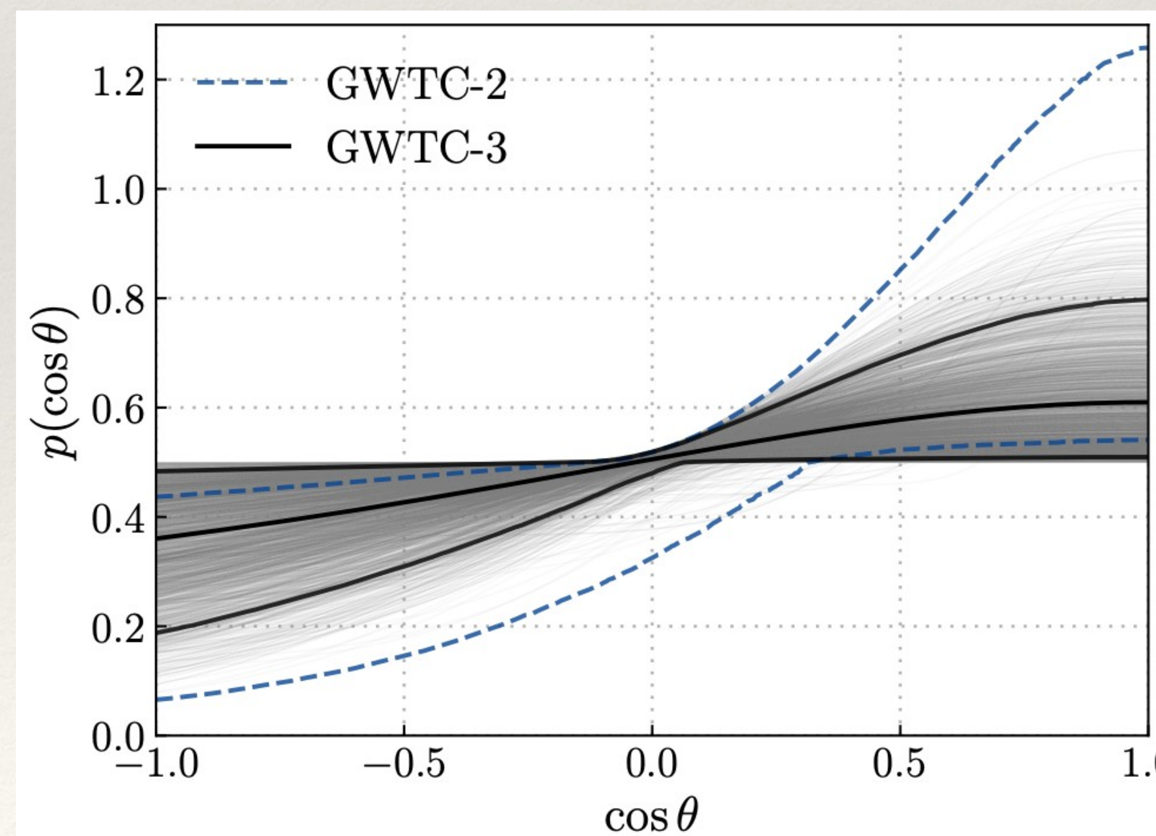



$$p_{\rm pop}(\chi_i|\alpha,\beta) = \frac{\chi_i^{\alpha-1}(1-\chi_1)(\beta-1)}{B(\alpha,\beta)}$$

$$p_{\rm pop}(\cos t_1, \cos t_2|\zeta,\sigma_t) = \frac{1-\zeta}{4}$$

$$+\frac{2\zeta}{\pi\sigma_t^2 \mathrm{erf}(\sqrt{2}/\sigma_t)^2}\exp\left[-\frac{(1-\cos t_1)^2+(1-\cos t_2)^2}{\sigma_t^2}\right]$$

LVK, PRX 2021




- Moderate spins are favoured
- Small preference for aligned spins

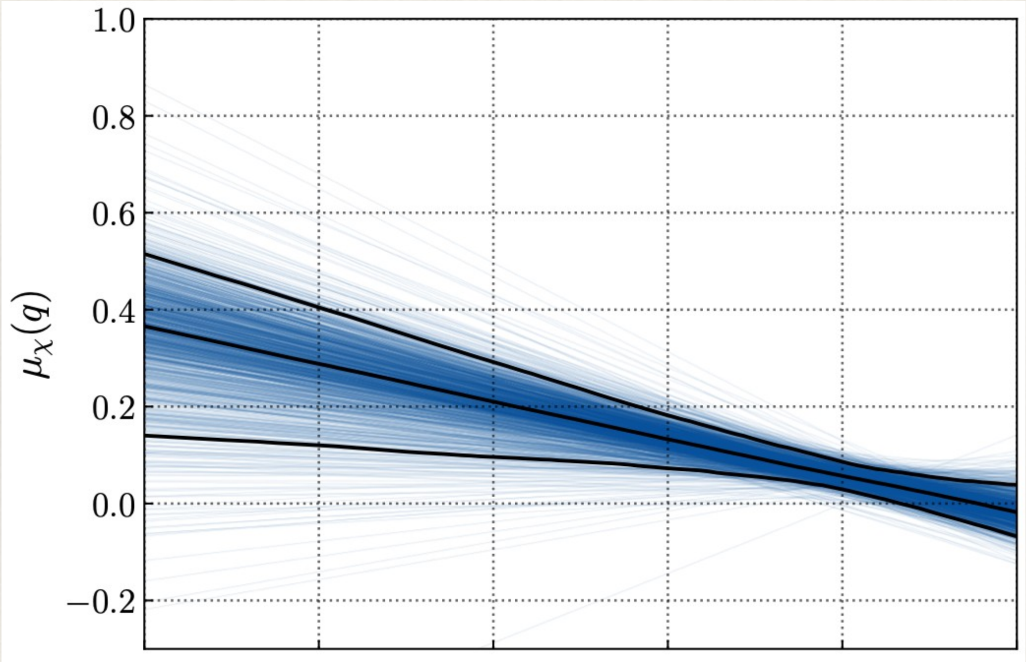
# Parametric model, multi-dimensional

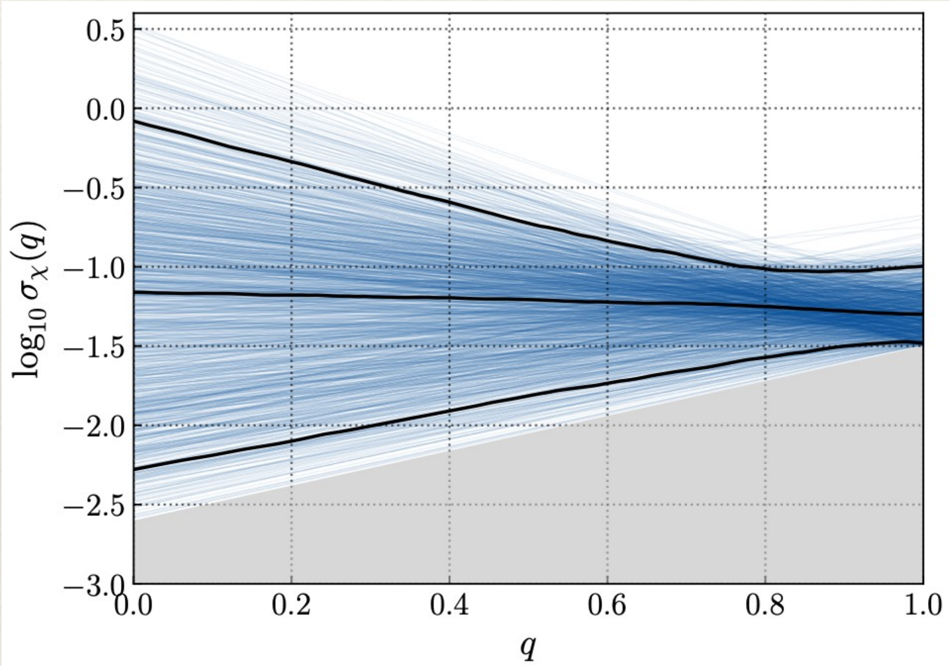
- Probe correlations between parameters by fitting joint distributions, or allowing model parameters to depend on other parameters. In the analysis of GWTC-3, the LVK explored the variation of the spin distribution with mass ratio.

$$p_{\rm pop}(\chi_{\rm eff}|q) \propto \exp\left[-\frac{(\chi_{\rm eff}-\mu(q))}{2\sigma(q)^2}\right]$$

$$\mu(q) = \mu_0 + \alpha(q-1)$$

$$\log_{10}\sigma(q) = \log_{10}(\sigma_0) + \beta(q-1)$$

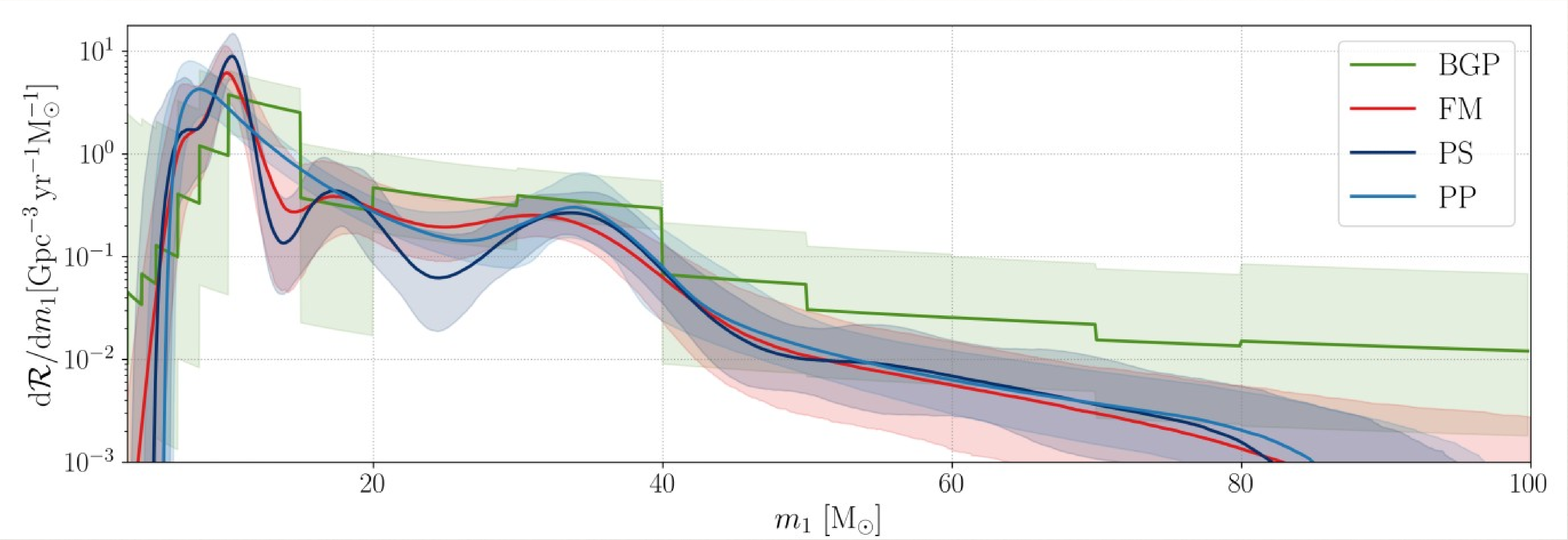



- Also hints of correlation between effetctive spin and redshift (Biscoveanu et al., APJL 2022), mass and redshift (Fishbach et al., APJ 2021), mass and spins (Hoy et al., APJ 2022, non-parametric: Godfrey et al., 2304.01288, Rinaldi et al., 2310.03074)

- **Pros:**
  - Analytic pdfs, easy to evaluate
  - Some astrophysical meaning
- **Cons:**
  - Little flexibility
  - Not so much astrophysical meaning

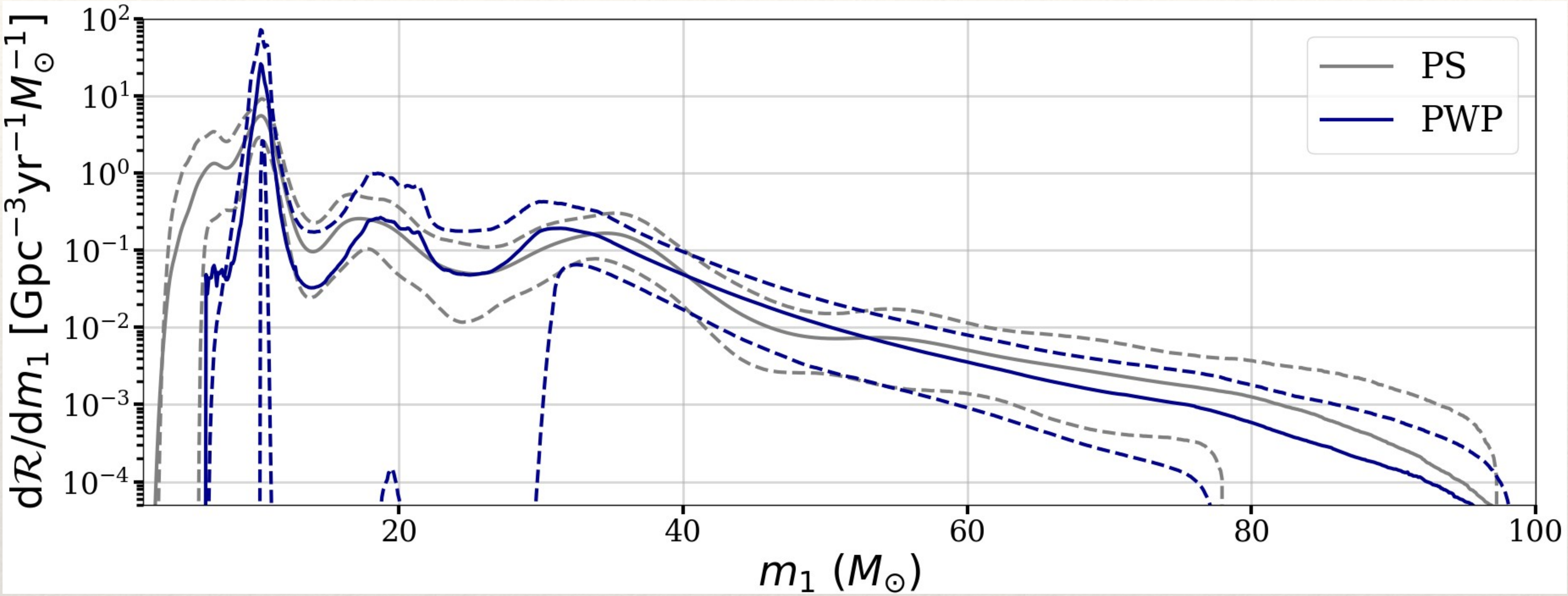
# Non-parametric model




LVK, PRX 2021

- BGP: Binned Gaussian process
- FM: Flexible mixture, total pdf is the sum of elementary functions, here Gaussian for the primary mass, the spins and power-laws for mass ratio
- PS: power-law spline, pdf is power-law times a spline which value at fixed knots is inferred, presence of peak is not imposed

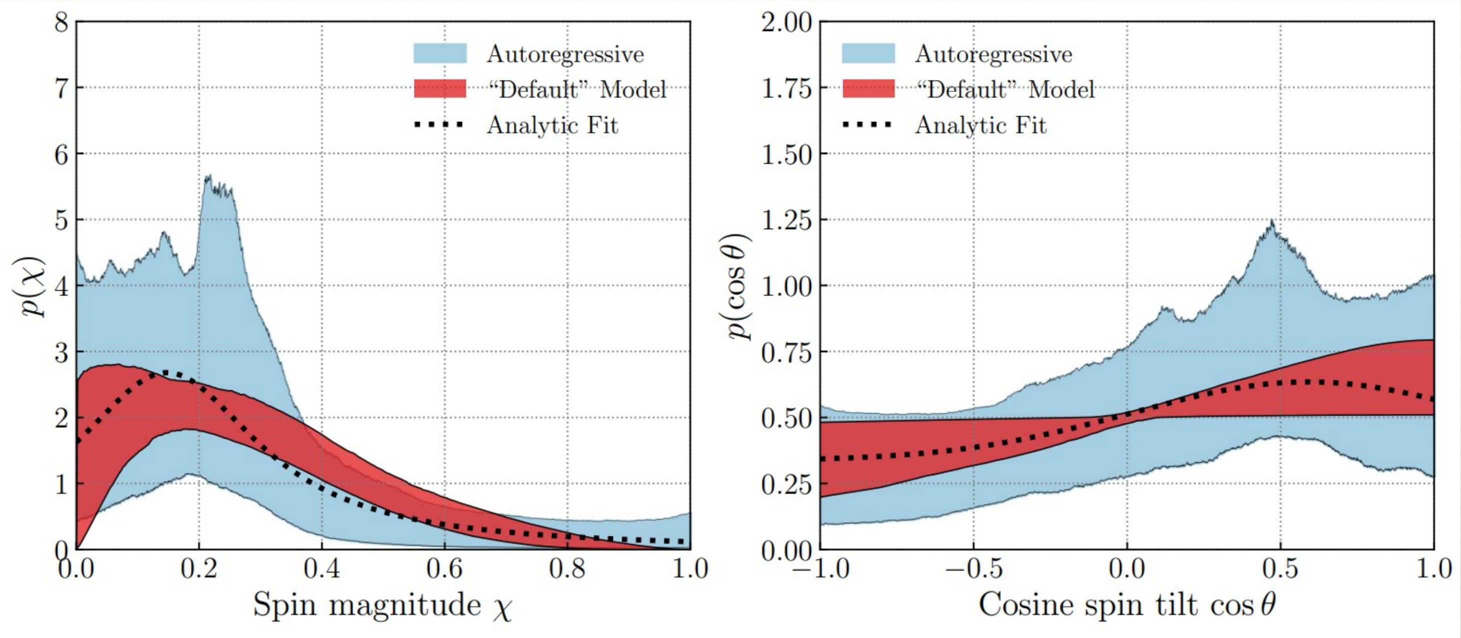
# Non-parametric model



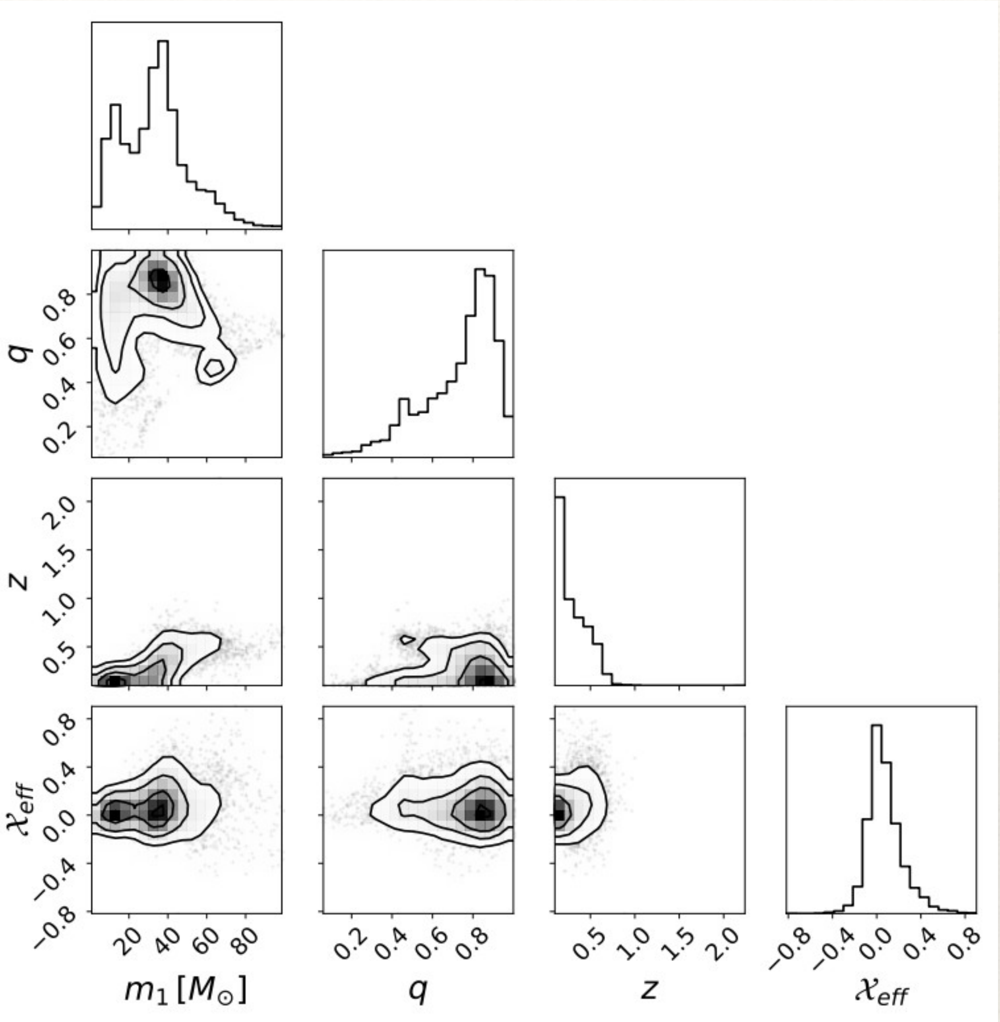

Toubiana et al., MNRAS 2023

- Model the pdf as a piece-wise power-law, vary the position and the number of knots using Reversible Jump MCMC.
- Performing (simplified) mock injections we find a 5% probability of peak at $\sim 20 M_\odot$ to be spurious.

# Non-parametric model




Callister et al., 2302.07289




Ruhe et al., 2211.09008v3

- Note: here selection effects are not included

- **Pros:**
  - Very flexible
  - Requires less a prior knowledge

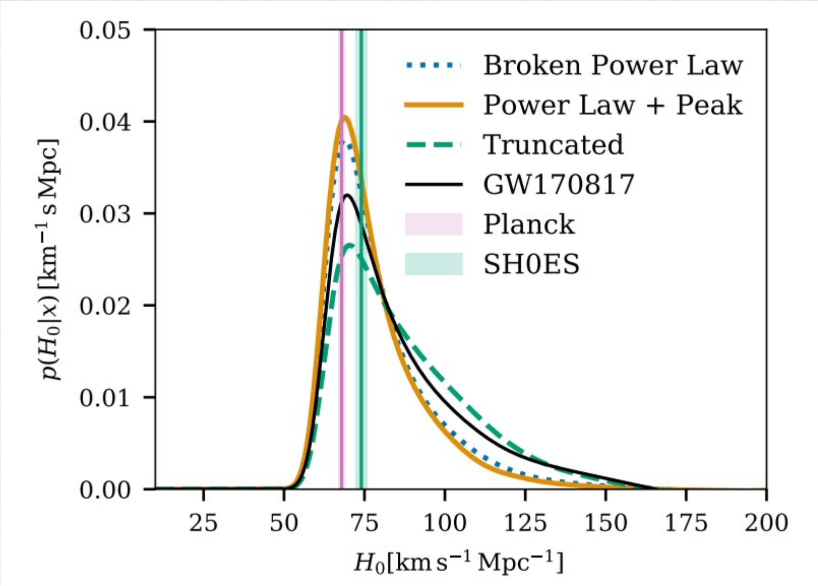
- **Cons:**
  - Parameters have no astrophysical meaning
  - Complexity is a priori arbitrary (might be alleviated using RJMCMC)

# Cosmological inference

- Cosmological parameters enter through the source-frame to detector-frame conversion:

$$m_s = \frac{m_d}{1 + z(\Lambda_{\text{cosmo}})}$$

- **Spectral-sirens**: Measure detector-frame masses while assuming a population model for source-frame ones to obtain cosmological parameters




LVK, APJ 2023

- Can also be used to constrain
  - modified propagation in beyond GR theories (Mancarella, Genoud-Prachex, Maggiore 2024 PRD 2022, Leyde, Mastrogiovanni, Steer, Chassande-Mottin, Karathanasis JCAP 2022)
  - dark energy EoS ( Mangiagli, Caprini, Marsat, Speri, Caldwell, Tamanini 2023)

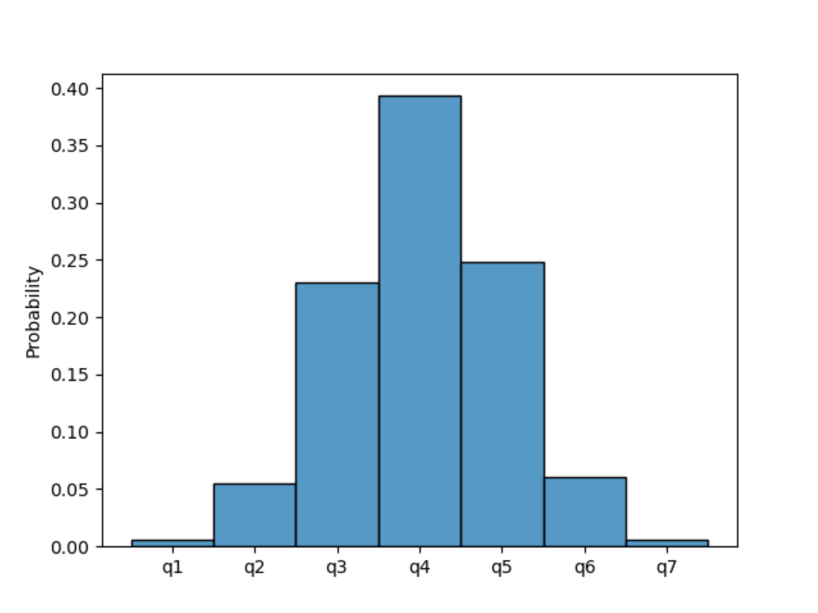
# “Converting fits”

- Ongoing work with S. Rinaldi and J. Gair, also different approach from C. Fabri and D. Gerosa
- Fit one “flexible”model dealing with measurement errors and selection effects and map it a posteriori to any sort of distribution
- Example: first fit a histogram and map it to any population model through a Dirichlet process:

$$p(\{d\}|\Lambda,\alpha) = \int p(\{d\}|q)p(q|\Lambda,\alpha)\ \mathrm{d}q$$

$$p(q_1,...q_n|\Lambda,\alpha) = \frac{\Gamma(\alpha)}{\prod_i^{N_{\mathrm{bin}}}\Gamma(\alpha_i)}\prod_i^{N_{\mathrm{bin}}} q_i^{\alpha_i-1}$$

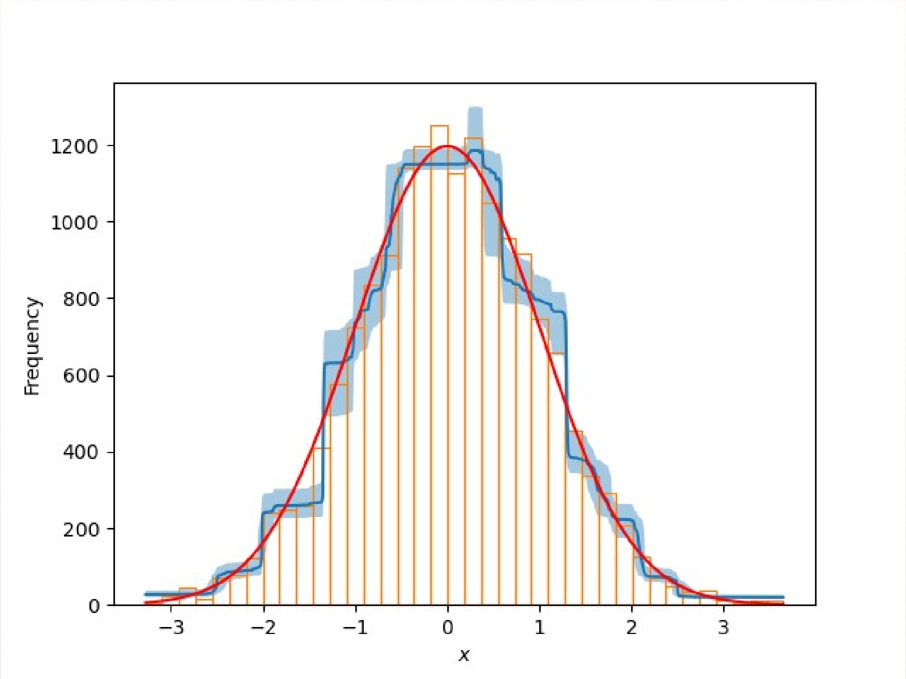
$$\alpha_i = \alpha\int_{\mathrm{bin}_i} p_{\mathrm{pop}}(\theta|\Lambda)\ \mathrm{d}\theta$$




- The concentration parameter $\alpha$ measures the quality of the “conversion”
- If data generated from $p_{\mathrm{pop}}(\theta|\Lambda_0)$, $\Lambda\rightarrow\Lambda_0$ when $N_{\mathrm{bin}}\rightarrow\infty$

# “Converting fits”

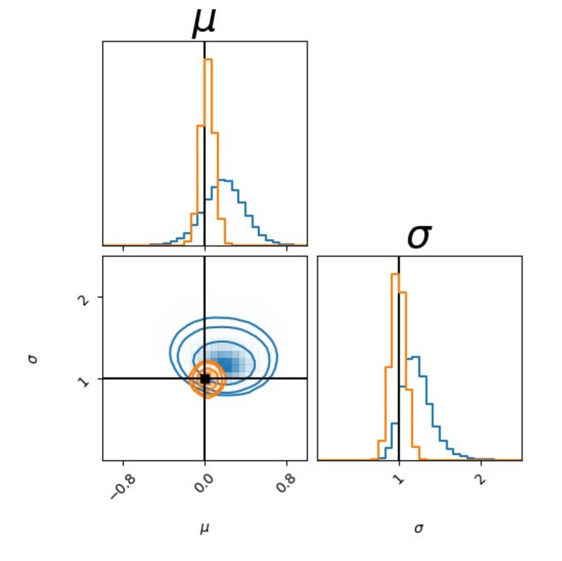
- Use reversible-jump MCMC to get “optimised” binning scheme




- Ongoing work

-

- Compare to standard hierarchical analysis:

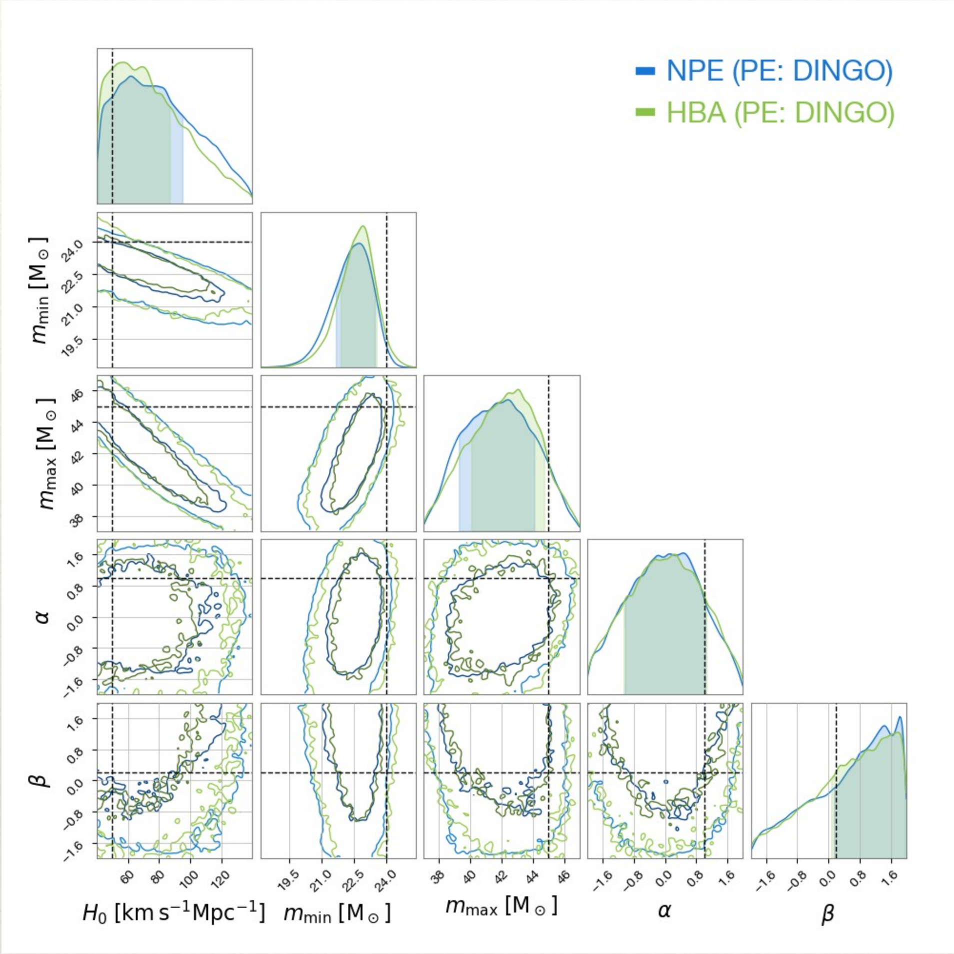


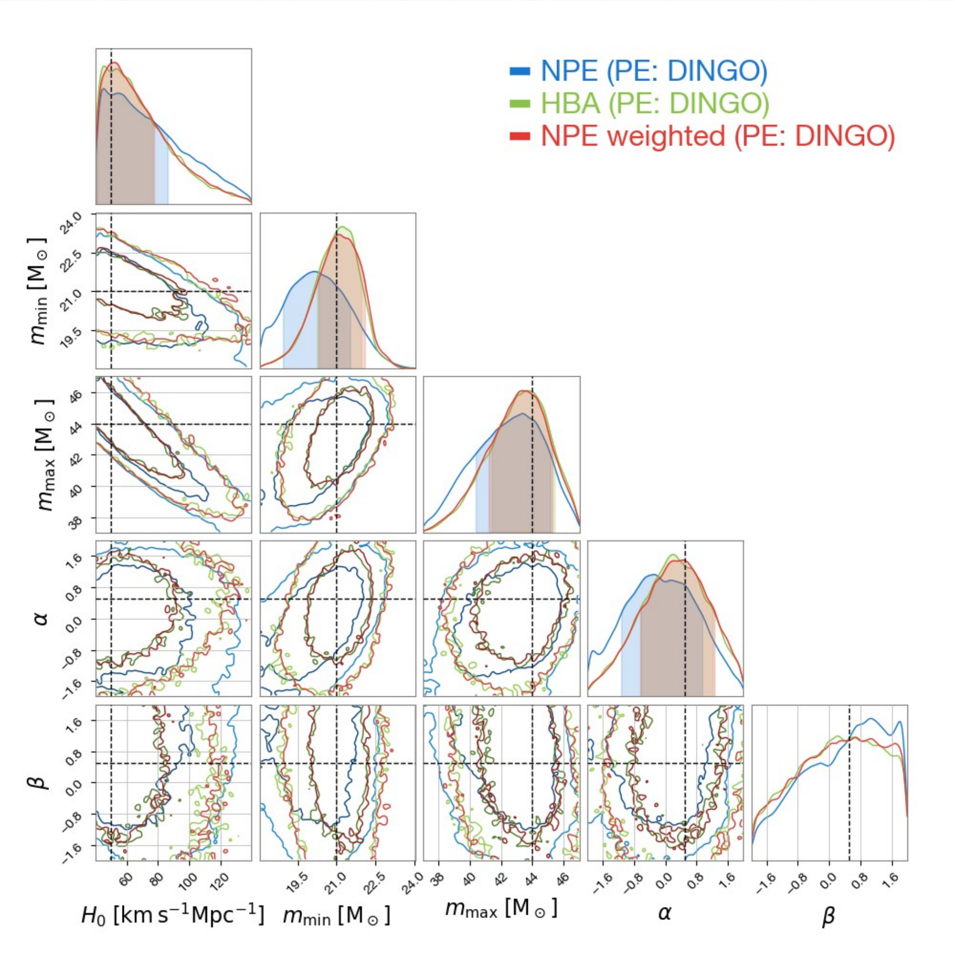

-

# Machine learning methods

- Use normalising flow to interpolate $p_{\rm pop}(\theta|\Lambda)$ from astro simulations
- Use normalising flow to "learn" the posterior on $\Lambda$ (Leyde, Green, Toubiana, Gair PRD 2024) :
  - For $\Lambda \sim p(\Lambda)$, draw events $\{\theta\}$, generate data $\{d\}$, draw samples $\hat{\theta}$
  - From a set of samples from the observed events obtain $p(\Lambda|\{d\})$
-

# Machine learning methods




Leyde, Green, Toubiana, Gair PRD 2024

- Future work:
  - Improve fit
  - Allow for arbitrary number of events
  - Learn directly the strain

# Future challenges

- Computational cost scales (naively) with the number of events:

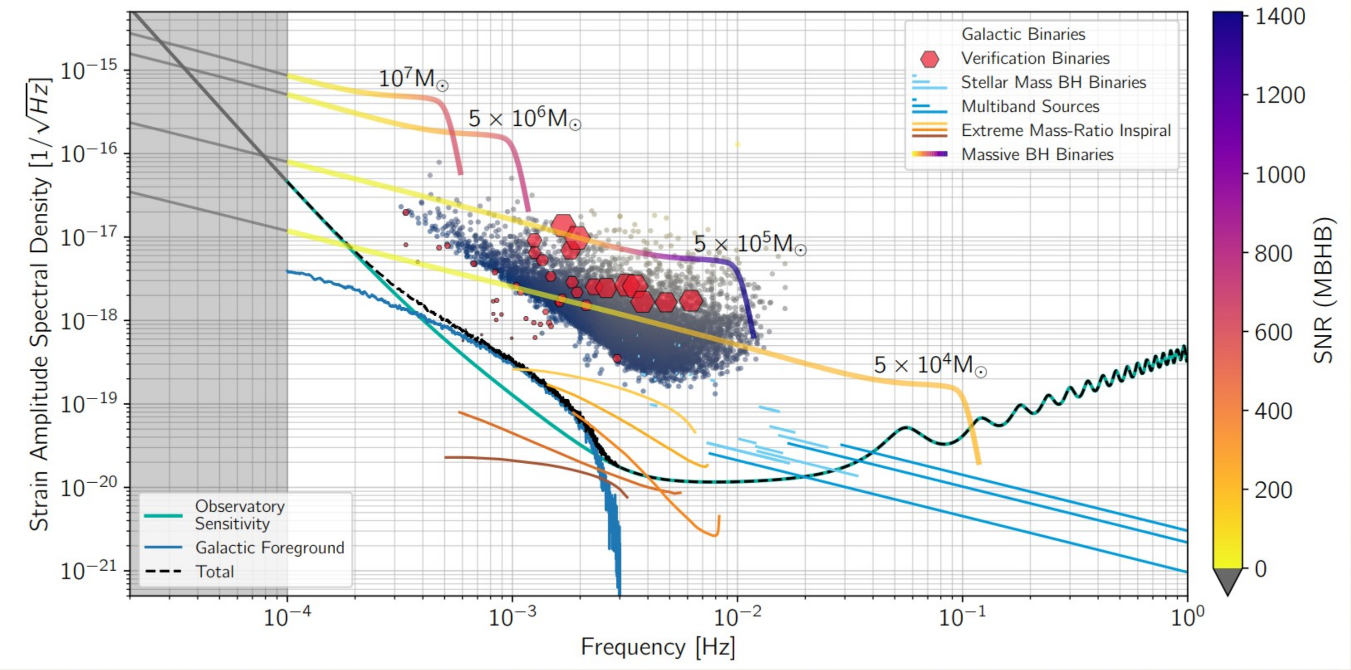
$$\prod_i \int \mathcal{L}(d_i|\theta)p(\theta|\Lambda)\ \mathrm{d}\theta = \prod_i \frac{p(d_i)}{N_s} \sum_{\theta_j \sim p(\theta|d_i)} \frac{p_{\mathrm{pop}}(\theta_j|\Lambda)}{\pi_{\mathrm{PE}}(\theta_j)}$$

- Selection function computation might become inaccurate (see Essick et al., 2204.00461)

$$p_s(\Lambda) = \frac{1}{N_{\mathrm{inj}}} \sum_{\theta_i \sim p_{\mathrm{inj}}(\theta)} \frac{p_{\mathrm{pop}}(\theta_i|\Lambda)}{p_{\mathrm{inj}}(\theta_i)}$$

- Modelling of multi-dimensional distributions
- "Systematics" in the population model might significantly bias the results (posterior predictive checking!)

# LISA challenges




- Events are not independent: "Global Fit":

$$\mathcal{L}(d|\theta) \rightarrow \mathcal{L}(d|\{\theta\}_{\rm res}, \Theta_{\rm stoch})$$

  - Need inverse mapping from foreground to population
  - Selection?

- Toubiana, Gair (on going): create mock "Global fit" to investigate general formalism for population studies with LISA

# Predictive checking

- It is natural to want to test if the assumed model is a good fit to the data. In a Bayesian context this is achieved through **predictive checking**.
- The **prior predictive distribution** is defined by

$$p(d) = \int p(d|\Lambda)p(\Lambda)\mathrm{d}\Lambda$$

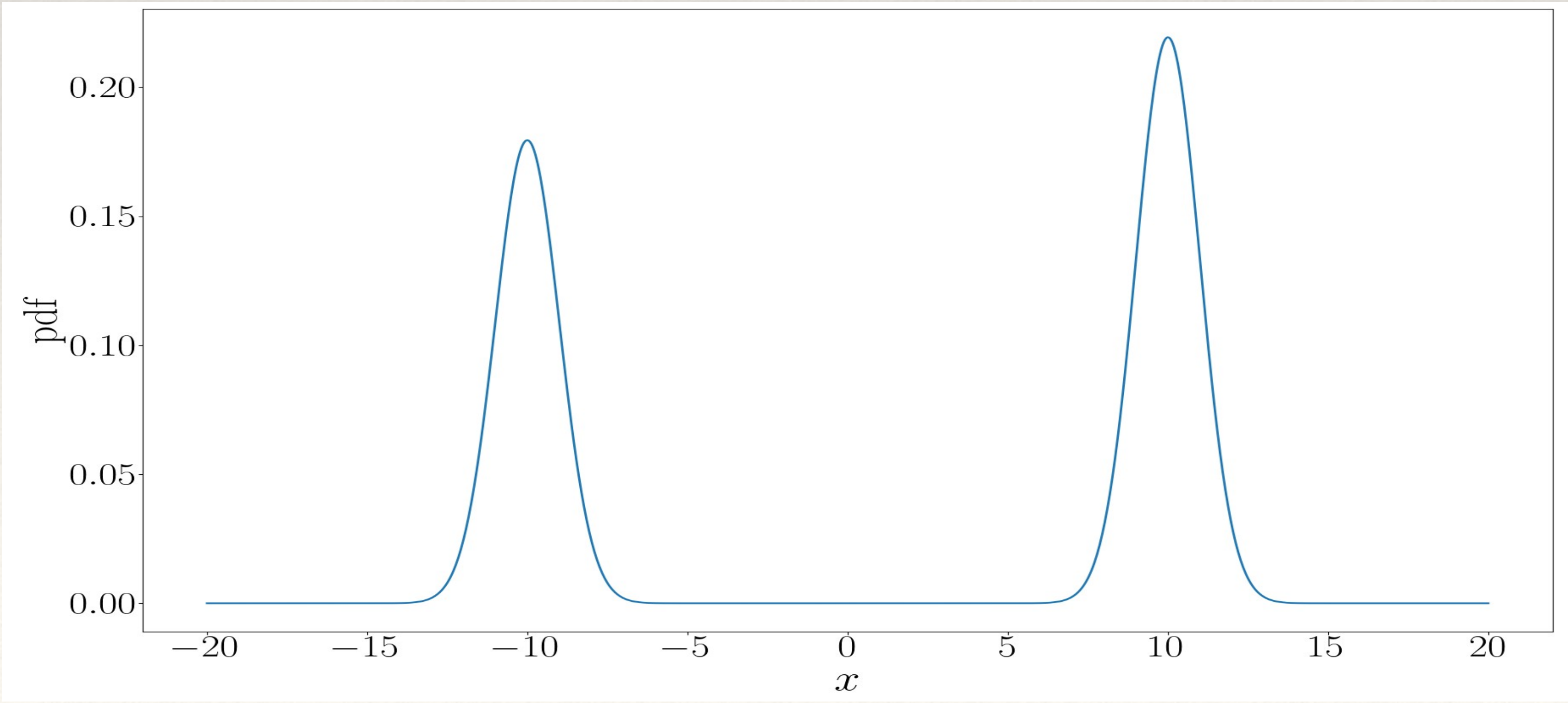
- This is the distribution of observed data sets within the model assumed in the prior. If the observed data is not very consistent with this distribution, the prior parameters might need to be adjusted.
- The **posterior predictive distribution** is defined similarly

$$p(d_{\mathrm{new}}|\{d\}_{\mathrm{old}}) = \int p(d_{\mathrm{new}}|\Lambda)p(\Lambda|\{d\}_{\mathrm{old}})\mathrm{d}\Lambda$$

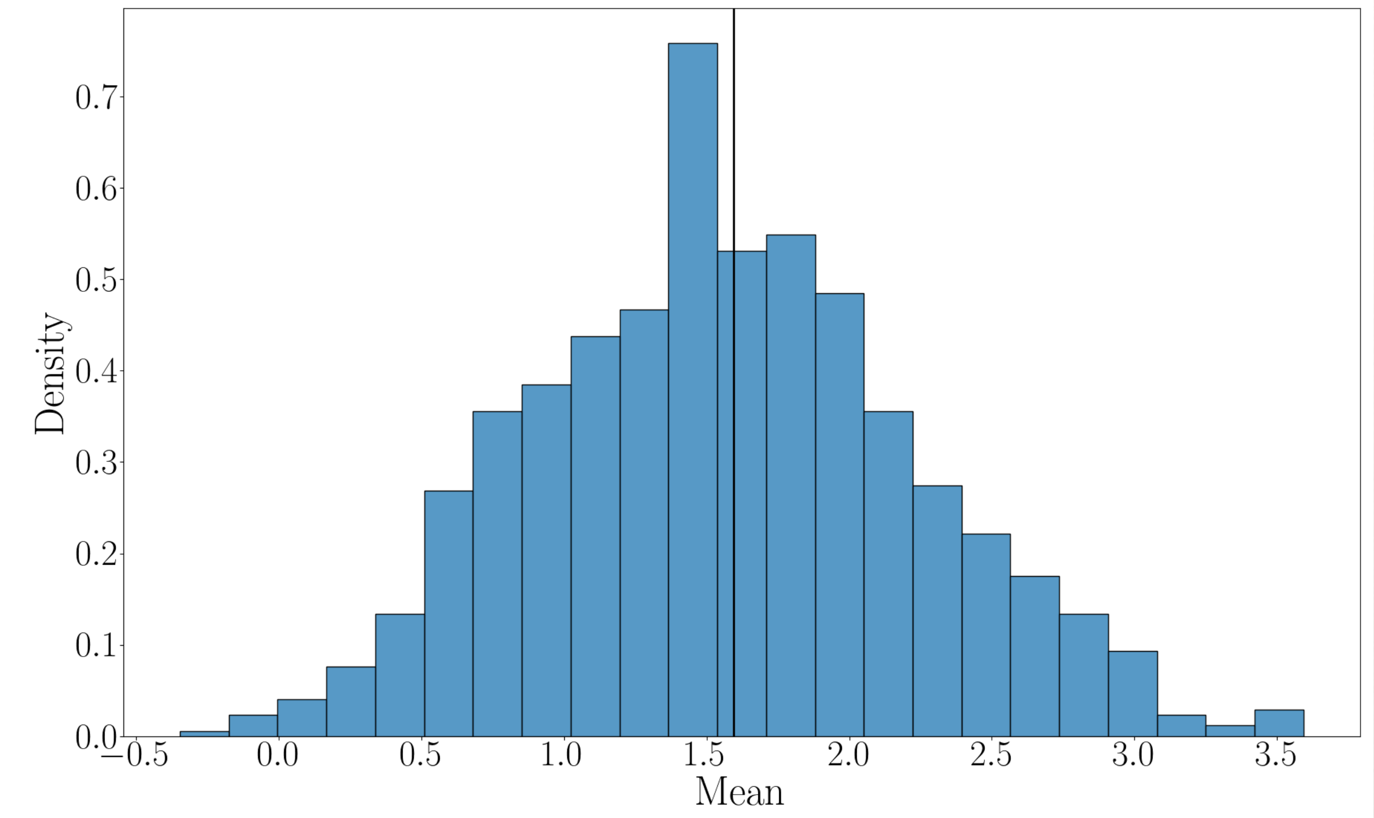
- This is the distribution of new datasets based on the model fitted to the data. The observed data should lie within the body of this distribution if the model is good.

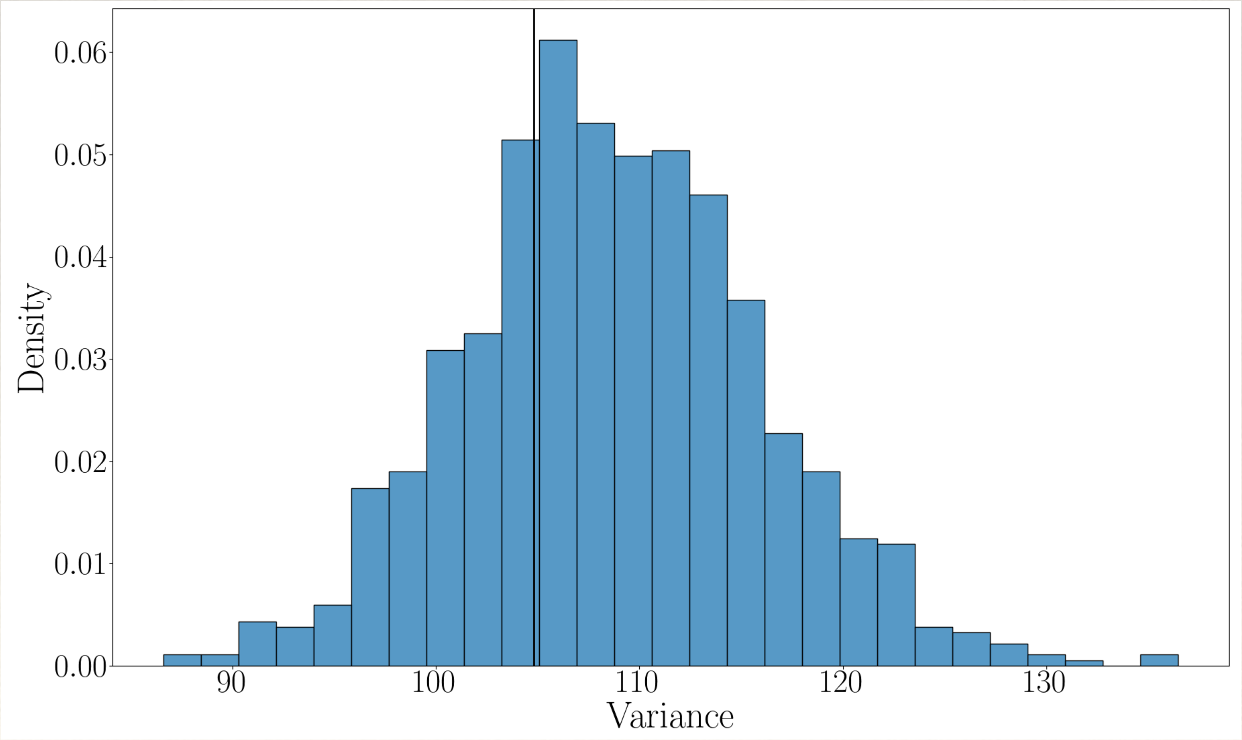
# Example: gaussian fit

- The predictive distribution can be used to compute the distribution of summary quantities. The value of those summary quantities in the observed data can then be compared to these distributions.
- It is better to choose quantities that are somewhat "orthogonal" to what is adjusted to fit the data.
- Example: we try to fit a Gaussian to the following distribution:
- We assume a Gaussian measurement error of 2

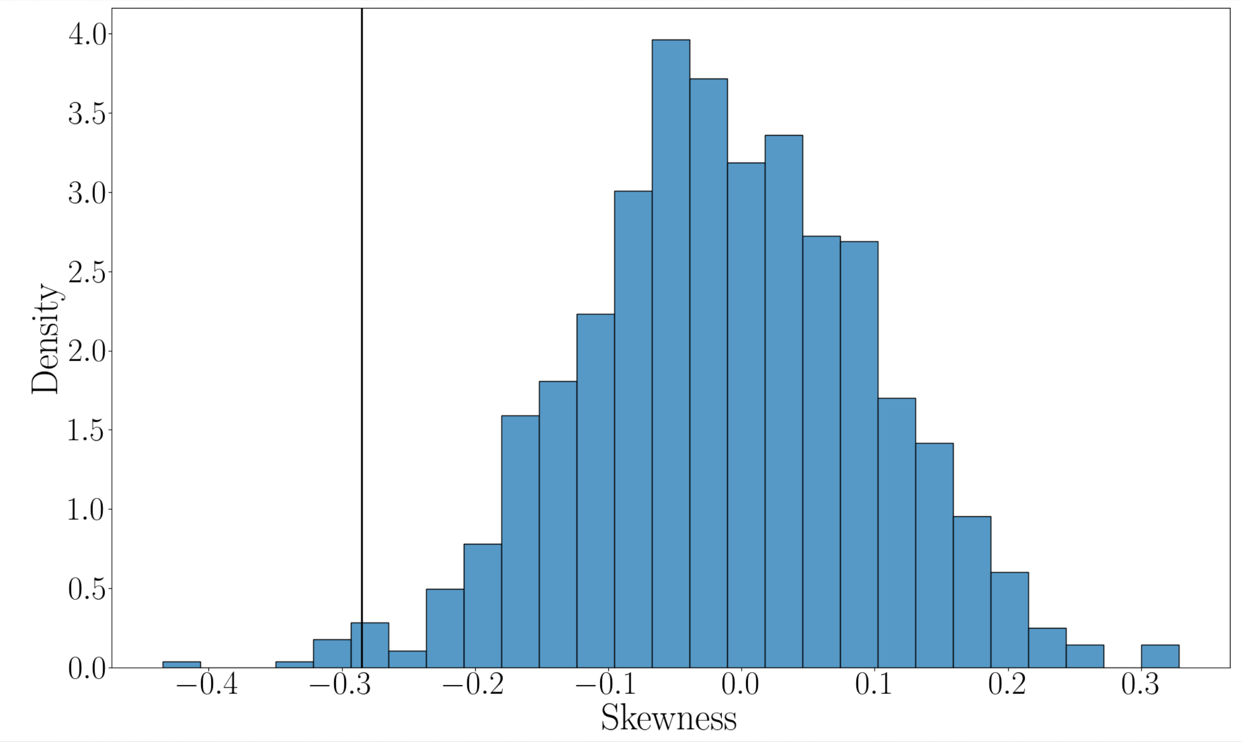


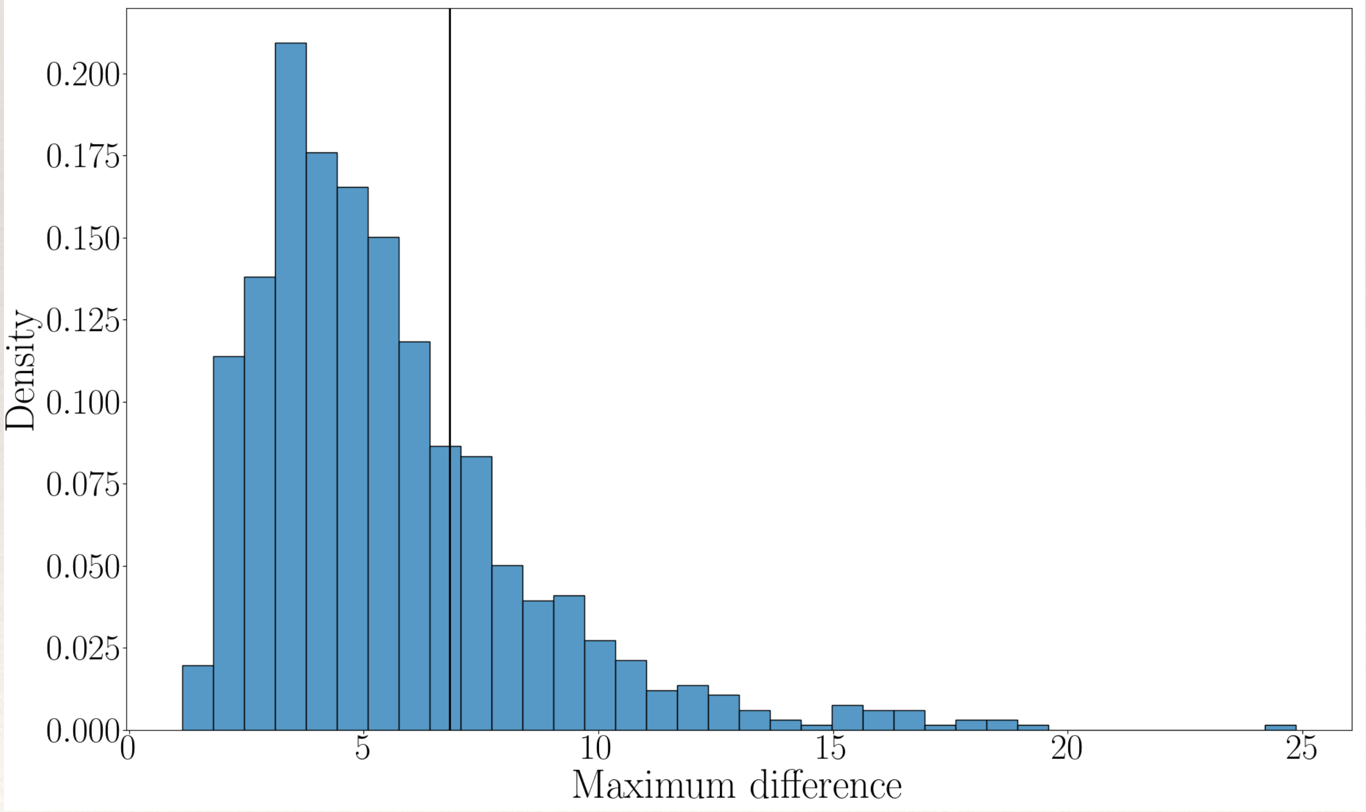
# Example: gaussian fit



# Example: gaussian fit



# Predictive checking

- Posterior predictive checks are “good practice”.
- Can help build intuition how to improve models.
- But are often computationally expensive...

# MCMC

- Starting from a point $\theta$ propose a point $\theta\prime$ with the proposal function $q(\theta \to \theta\prime)$
- Accept with probability $$\frac{p(\theta'|d)q(\theta' \to \theta)}{p(\theta|d)q(\theta \to \theta')}$$
- In any case, record the point

# Reversible jump MCMC

- Vary the dimensionality of parameter space $\Theta = (\theta_1, ..., \theta_n)$
- E.g. global fit for LISA
- Propose to add component through proposal function $q(\theta_{n+1})$
- Accept  with probability

$$\frac{\mathcal{L}(d|\theta_1, ..., \theta_n, \theta_{n+1})\pi_{\mathrm{PE}}(\theta_{n+1})}{\mathcal{L}(d|\theta_1, ..., \theta_n)q(\theta_{n+1})}$$

- In any case, record the point
- Also do jumps in fixed dimension