Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **526,** 3495–3503 (2023) https://doi.org/10.1093/mnras/stad2968
Advance Access publication 2023 September 28

Growing pains: understanding the impact of likelihood uncertainty on hierarchical Bayesian inference for gravitational-wave astronomy

Colm Talbot [1,2]★ and Jacob Golomb[3,4]

[1]*LIGO Laboratory, Massachusetts Institute of Technology, 185 Albany Street, Cambridge, MA 02139, USA*
[2]*Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*
[3]*LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125, USA*
[4]*Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA*

Accepted 2023 September 26. Received 2023 September 26; in original form 2023 June 9

ABSTRACT

Observations of gravitational waves emitted by merging compact binaries have provided tantalizing hints about stellar astrophysics, cosmology, and fundamental physics. However, the physical parameters describing the systems (mass, spin, distance) used to extract these inferences about the Universe are subject to large uncertainties. The most widely used method of performing these analyses requires performing many Monte Carlo integrals to marginalize over the uncertainty in the properties of the individual binaries and the survey selection bias. These Monte Carlo integrals are subject to fundamental statistical uncertainties. Previous treatments of this statistical uncertainty have focused on ensuring that the precision of the inferred inference is unaffected; however, these works have neglected the question of whether sufficient accuracy can also be achieved. In this work, we provide a practical exploration of the impact of uncertainty in our analyses and provide a suggested framework for verifying that astrophysical inferences made with the gravitational-wave transient catalogue are accurate. Applying our framework to models used by the LIGO–Virgo–KAGRA collaboration and in the wider literature, we find that Monte Carlo uncertainty in estimating the survey selection bias is the limiting factor in our ability to probe narrow population models and this will rapidly grow more problematic as the size of the observed population increases.

Key words: gravitational waves – methods: data analysis – methods: statistical.

1 INTRODUCTION

Using data from the first three observing runs of Advanced LIGO (LIGO Scientific Collaboration 2015) and Advanced Virgo (Acernese et al. 2015), ≈70 signals from the merger of compact binary systems have been identified (The LIGO Scientific Collaboration, The Virgo Collaboration & The KAGRA Collaboration 2021c), along with a few tens of less significant additional candidate events (Nitz et al. 2023; Olsen et al. 2022). While individual observations of compact binary mergers provide insights into astrophysics and cosmology, maximizing the physical resolving power using the catalogue of gravitational-wave transients requires analysing the entire population as a hierarchical Bayesian inference problem. Due to computational constraints, these analyses are performed using a multistage process to calculate the population-level likelihood (see e.g. Mandel, Farr & Gair 2019; Thrane & Talbot 2019; The LIGO Scientific Collaboration, The Virgo Collaboration & The KAGRA Collaboration 2021d; Vitale et al. 2022).

First, segments of data that are likely to contain gravitational-wave signals are identified by search pipelines (e.g. Allen et al. 2012). These pipelines are only sensitive to the loudest signals and so the observed sample is biased in favour of nearby high-mass binaries with black hole angular momenta ('spins') aligned with the orbital angular momentum (Campanelli, Lousto & Zlochower 2006). This selection bias is typically accounted for by estimating the fraction of binaries that we expect to observe using simulated 'injection' campaigns.

Next, the strain data from gravitational-wave detectors containing the observed transients are analysed with a fiducial reference model for the population (often referred to as the fiducial prior distribution) in order to obtain samples from the fiducial posterior probability distribution for the parameters (masses, spins, etc.) of each binary. While the fiducial prior distribution impacts the fiducial posterior, it is typically chosen to avoid imprinting astrophysical assumptions on the results. For example, binaries are assumed to be distributed homogeneously and isotropically throughout the Universe. The fiducial model for black hole masses is usually uniform in the mass of each black hole and uniform in spin magnitude and isotropic in direction.

In the final stage, these fiducial samples are importance sampled ('reweighted') using a parametrized model for the underlying population to compute the likelihood for the observed data given population-level parameters (e.g. the maximum allowed black hole mass) marginalized over the per-event parameters. For each model for the underlying population, the fraction of observable binaries is also estimated using importance sampling on the injected signals from the first stage (e.g. Finn & Chernoff 1993; Loredo 2004; Farr et al. 2015).

★ E-mail: colm.talbot@ligo.org

The importance sampling step is an example of using Monte Carlo summation to approximate an integral and as such comes with some intrinsic uncertainty that enters the analysis as a source of systematic error. Typically, this uncertainty is ignored when performing the analysis; however, in recent years several attempts have been made to quantify this uncertainty and theoretically motivated heuristics have been proposed to estimate and (hopefully) mitigate its impact (Farr 2019; Essick & Farr 2022). In this work, we perform a data-driven analysis of the potential systematic uncertainties from our use of Monte Carlo integration. We note that while we apply our formalism to the problem of population inference for gravitational-wave astronomy, it is widely applicable to any context in which an approximate estimator of the true likelihood is used in a Bayesian analysis.

The remainder of this paper is structured as follows. In the next section, we describe how uncertainty appears in our estimate of the population likelihood through Monte Carlo integration and suggest a set of convergence criteria. In Section 3, we analyse a simple toy model to examine the impact of uncertainty on the accuracy of inference. Using this, we establish a threshold beyond which we expect our results to be significantly biased. Following this, we take a range of models previously considered for population analyses and quantify the uncertainty in these results in Section 4. Finally, we provide a closing discussion.

2 UNCERTAINTY IN THE POPULATION LIKELIHOOD APPROXIMATION

The likelihood function typically employed for an analysis of a population of N observed systems with source-dependent selection effects can be written as (see e.g. Mandel et al. 2019; Thrane & Talbot 2019; Vitale et al. 2022, for details)

$$\mathcal{L}(\{d_i|\Lambda) \propto \prod_i^N \frac{\mathcal{L}(d_i|\Lambda)}{P_{\rm det}(\Lambda)}. \tag{1}$$

Here, $\{d_i\}$ are the data containing the observed signals (indexed by i). In the context of gravitational-wave astronomy, these are strain data recorded by gravitational-wave interferometers. The selection function $P_{\rm det}$ is the fraction of all signals that would be observed for a population described by population hyperparameters Λ. We note that this likelihood has been marginalized over the overall rate of events (assuming a uniform-in-log rate prior) and the parameters describing each of the individual systems.

Each of the terms $\mathcal{L}(d_i|\Lambda)$ and $P_{\rm det}(\Lambda)$ is computed by marginalizing over θ, the ≈15 parameters describing the individual binaries, and many more describing the noise properties of the interferometers:

$$\mathcal{L}(d_i|\Lambda) = \int d\theta\, p(d_i, \theta|\Lambda) = \int d\theta \mathcal{L}(d_i|\theta) p(\theta|\Lambda) \tag{2}$$

$$P_{\rm det}(\Lambda) = \int dd \int d\theta\, p(d, \theta|\Lambda)\Theta(\rho(d) - \rho_*) \tag{3}$$

$$= \int dd \int d\theta \mathcal{L}(d|\theta) p(\theta|\Lambda)\Theta(\rho(d) - \rho_*). \tag{4}$$

In both expressions, we have expanded the joint distribution for the observed data and signal parameters into the population model $p(\theta|\Lambda)$ and the likelihood of observing data given single-event parameters $\mathcal{L}(d|\theta)$. The integral over d in the expression for $P_{\rm det}$ is over all of the data collected by the instrument, while the d_i represents the data around the time of a specific observed signal. The final term is a Heaviside step function for the detection statistic (e.g. signal-to-noise ratio or false alarm rate) ρ with threshold ρ_*. In order to minimize the cost of performing the analysis, these integrals are commonly computed using Monte Carlo estimators using some reference set of samples from the fiducial posterior distribution. We denote the estimator of quantity x as $\hat{x}$. As a specific example, the estimator of the log-likelihood (equation 1) is

$$\ln \hat{\mathcal{L}}(\{d_i\}|\Lambda) = \left(\sum_i^N \ln \hat{\mathcal{L}}(d_i|\Lambda)\right) - N \ln \hat{P}_{\rm det}(\Lambda). \tag{5}$$

In practice, these estimates are calculated using Monte Carlo integration:

$$I = \int dx f(x) p(x) \equiv \langle f \rangle_{p(x)}, \tag{6}$$

$$\hat{I} = \frac{1}{M} \sum_{x_j \sim p(x)}^{j=M} f(x_j). \tag{7}$$

Here, $\hat{I}$ is the estimator of the integral I and M is the number of samples in the Monte Carlo integral. We note that $p(x)$ is a normalized probability distribution and $f(x)$ is an arbitrary function of parameters x. Every Monte Carlo has an intrinsic statistical uncertainty

$$\sigma_I^2 = \frac{1}{M}\left[\langle f^2 \rangle_{p(x)} - \langle f \rangle_{p(x)}^2\right] \equiv \frac{1}{M}\bar{\sigma}_I^2. \tag{8}$$

We define the quantity $\bar{\sigma}_I^2$ as the intrinsic variance between the proposal distribution $p(x)$ and the target distribution $f(x)p(x)$. In general, the uncertainty in a Monte Carlo integral will be minimized by choosing $p(x)$ and $f(x)$ to minimize $\bar{\sigma}_I$. For example, for most gravitational-wave population analyses (including this work), we choose

$$f(\theta) \sim \frac{p(\theta|\Lambda)}{p(\theta|\varnothing)}, \quad p(\theta) \sim \mathcal{L}(d|\theta) p(\theta|\varnothing),$$

where $p(\theta|\varnothing)$ is the fiducial prior distribution. However, in some cases it is beneficial to define (e.g. Wysocki, Lange & O'Shaughnessy 2019; Golomb & Talbot 2022b) $f(\theta) \sim \mathcal{L}(d|\theta)$, $p(\theta) \sim p(\theta|\Lambda)$. We also note that the variance scales inversely with the number of samples. A final quantity related to Monte Carlo integrals that we will need is the effective number of independent samples (Kish 1995)

$$n_{\rm eff} = M \frac{\langle f \rangle_{p(x)}^2}{\langle f^2 \rangle_{p(x)}}. \tag{9}$$

In Farr (2019), the author demonstrates that for small values of $n_{\rm eff}$ a Gaussian approximation to the likelihood uncertainty breaks down. In previous works (e.g. Farr 2019; The LIGO Scientific Collaboration et al. 2021d), $n_{\rm eff}$ has been used to assess the convergence of the likelihood estimator and to impose data-dependent cuts on the allowed parameter space. We prefer to work directly with the estimated variance and include $n_{\rm eff}$ here just to compare with previous work.

Since we assume that the reference samples used in each of the Monte Carlo integrals are independent, the variance in the estimate of the population (log-)likelihood is

$$\sigma^2_{\ln \hat{\mathcal{L}}}(\Lambda) = \sum_i^N \sigma^2_{\ln \hat{\mathcal{L}}_i}(\Lambda) + N^2 \sigma^2_{\rm sel}(\Lambda). \tag{10}$$

We note that the contribution to the total variance from the selection function grows quadratically with the population size, as $\mathrm{Var}(Nx) = N^2\mathrm{Var}(x)$.

Assuming the individual observations are independent and identically distributed draws from the underlying population, we recast this

expression in terms of an average per-observation uncertainty $\sigma_{\rm obs}$ to more clearly see the dependence of both terms with the population size

$$\sigma^2_{\ln\hat{\mathcal{L}}}(\Lambda) = N\sigma^2_{\rm obs}(\Lambda) + N^2\sigma^2_{\rm sel}(\Lambda). \tag{11}$$

We have explicitly retained the dependence of this variance on the hyperparameters. We justify the assumption that $\sigma_{\rm obs}$ does not vary with time in Section 4.1.

Since we are predominantly interested in differences in log-likelihood for points with significant posterior support, we need to limit the error in the difference of log-likelihood estimators, $\Delta\ln\hat{\mathcal{L}}$. In general, the errors will not be independent, and so we calculate the variance in this quantity $\sigma^2_{\Delta\ln\hat{\mathcal{L}}}$ as defined in equation (A11) in Essick & Farr (2022). We assume the error in the estimator of the log-likelihood is Gaussian distributed as the contribution to the population log-likelihood from the per-event terms is the sum of N independently and identically distributed estimators and so by the central limit theorem follows a normal distribution and in the high effective-sample size limit the selection function term also follows a normal distribution (Farr 2019) We therefore write $\sigma^2_{\Delta\ln\mathcal{L}} = \sigma^2_{\Delta\ln\hat{\mathcal{L}}}$.

If the uncertainties in the estimators are uncorrelated with Λ, we will have $\sigma^2_{\Delta\ln\hat{\mathcal{L}}} = \sigma^2_{\ln\hat{\mathcal{L}}}$. In Essick & Farr (2022), the authors demonstrate that under certain conditions the variance in likelihood differences in 'local neighbourhoods' avoids the worst-case scaling in equation (11) and rather find that

$$\sigma^2_{\Delta\ln\hat{\mathcal{L}}} = \sigma^2_{\rm obs}(\Lambda) + N\sigma^2_{\rm sel}(\Lambda) \tag{12}$$

for a simple example model due to correlation of the Monte Carlo errors between points with significant posterior support. It is unclear a priori when the local neighbourhood approximation is valid, in this work, we numerically test whether this scaling holds for the specific case of inferring the population properties of merging binary black hole systems.

2.1 Uncertainty as a draw from a Gaussian process

To build an understanding of the impact of uncertainty, we assert that the estimated difference in log-likelihood is a fair draw from the Gaussian process with mean function $\Delta\ln\mathcal{L}$ and (potentially non-stationary) kernel function $\Sigma(\Lambda, \Lambda')$

$$\Delta\ln\hat{\mathcal{L}}(\{d_i\}|\Lambda, \Lambda') \sim \mathcal{GP}(\Delta\ln\mathcal{L}(\{d_i\}|\Lambda, \Lambda'), \Sigma(\Lambda, \Lambda')). \tag{13}$$

Here, $\Sigma(\Lambda, \Lambda') = \sigma^2_{\Delta\ln\mathcal{L}}$ is the $2D$-dimensional covariance matrix, where D is the dimensionality of the population model. In practice, we do not have access to the true kernel function, and so we approximate it using a numerical covariance matrix using the covariance between the likelihood estimator at each pair of points we consider. Specifically, we construct the approximate kernel by numerically calculating

$$\Sigma(\Lambda, \Lambda') = \sigma^2_{\Delta\ln\hat{\mathcal{L}}} \tag{14}$$

following equation (A11) in Essick & Farr (2022). We will use this quantity to estimate the average variance over the posterior for the hyperparameters

$$\left\langle\Delta\ln\hat{\mathcal{L}}\right\rangle \equiv \int d\Lambda \int d\Lambda' p(\Lambda|\{d_i\})p(\Lambda'|\{d_i\})\Sigma(\Lambda, \Lambda') \tag{15}$$

$$\approx \frac{1}{K^2}\sum_{k=1}^{k=K}\sum_{k'=1}^{k'=K}\Sigma(\Lambda_k, \Lambda_{k'}) \tag{16}$$

$$\Lambda_k \sim p(\Lambda|\{d_i\}). \tag{17}$$

We note that this is the average of the covariance matrix weighted by the posterior support.

This is a slightly different statistic than the one considered in Essick & Farr (2022), where the authors replace the integral over Λ' with a fixed value at the mean of the hyperposterior

$$\bar{\Lambda} = \int d\Lambda\, p(\Lambda|\{d_i\})\Lambda.$$

While the simpler expression used in Essick & Farr (2022) likely produces comparable results for posterior distributions with Gaussian uncertainties, for posteriors with more complex shapes, e.g. multimodality or curving degeneracies, the mean of the posterior is not in general representative of points with significant posterior support. In contrast, the full integral over Λ, Λ' ensures that we accurately represent the variance between all pairs of points with posterior support.

3 HOW UNCERTAIN CAN WE BE?

Before turning to real examples, we first motivate an acceptable level of uncertainty in the log-likelihood estimator. Specifically, we want to know a threshold value of $\langle\Delta\ln\hat{\mathcal{L}}\rangle$ above which we expect to see significant biases. To test this, we consider a simple one-dimensional problem where the true posterior distribution is a unit normal distribution. To verify that the threshold is independent of the structure of the covariance matrix, we perform this experiment with four analytical kernel functions: a block-diagonal kernel where each block is fully correlated with a random number of blocks, a Matérn kernel with $\nu = 5/2$ and random correlation length, a completely uncorrelated kernel, and a completely correlated kernel. We find that the result is independent of the kernel choice.

For 4800 iterations, we choose a covariance matrix using one of our kernels with a random value of $\langle\Delta\ln\hat{\mathcal{L}}\rangle$ drawn logarithmically between $[10^{-2}, 20]$. For each covariance matrix, we draw 100 realizations from the covariance matrix $\Sigma(\Lambda, \Lambda')$ to generate biased posterior probability distributions. For each of these realizations, we compute the fraction f of the posterior support below a random point drawn from the true posterior. If there is no bias, f should follow a uniform distribution in [0, 1]. We, therefore, compute a $p_{\rm value}$ comparing the 100 values of f to the uniform distribution. In Fig. 1, we show a two-dimensional histogram of the result of this numerical experiment. We see that when $\langle\Delta\ln\hat{\mathcal{L}}\rangle \lesssim 1$, the $p_{\rm value}$ are uniformly distributed indicating unbiased recovery. However, as the magnitude of the uncertainty rises, the distribution of $p_{\rm value}$ skews heavily towards small $p_{\rm value}$. As a final quantitative test, we compare the distribution of $p_{\rm value}$ in each bin of $\langle\Delta\ln\hat{\mathcal{L}}\rangle$ to compute a combined $p_{\rm value}$. We see that the combined $p_{\rm value}$ is very small for $\langle\Delta\ln\hat{\mathcal{L}}\rangle \gtrsim 1$. We will therefore use $\langle\Delta\ln\hat{\mathcal{L}}\rangle \gtrsim 1$ as our heuristic threshold for significant bias.

4 HOW UNCERTAIN ARE WE?

We now turn to a tangible example of uncertainty in the inference performed on the population of binary black hole mergers observed during the first three observing runs of Advanced LIGO and Advanced Virgo with a false alarm rate of less than one per year. The analyses performed in The LIGO Scientific Collaboration et al. (2021d) imposed cuts on the convergence of the Monte Carlo integrals that implicitly limit the variance in the likelihood to avoid spurious features in the posteriors. All analyses in that work imposed a condition first proposed in Farr (2019) demanding that for the selection function $n_{\rm eff} > 4N$. Some models also enforced the

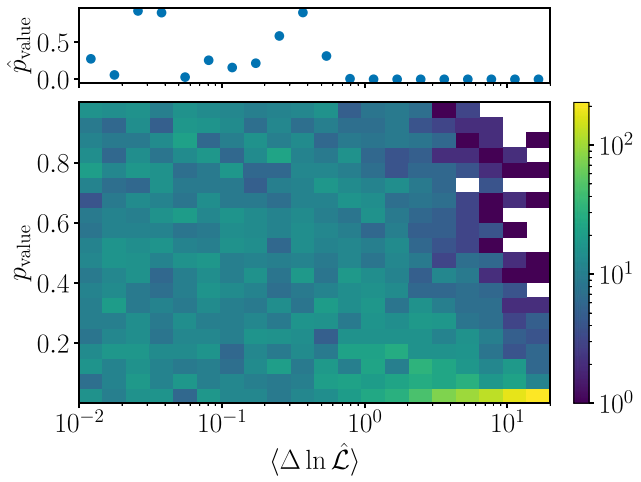

Figure 1. $p_{\rm value}$ versus uncertainty in difference in log-likelihood averaged over the posterior distribution ($\langle \Delta \ln \hat{\mathcal{L}} \rangle$). For unbiased analyses at a given value of $\langle \Delta \ln \hat{\mathcal{L}} \rangle$, we expect $p_{\rm value}$ to follow a uniform distribution in [0, 1]. The upper panel shows a combined $p_{\rm value}$ for all the points in the histogram falling within that range of $\langle \Delta \ln \hat{\mathcal{L}} \rangle$. We note that this is satisfied for $\langle \Delta \ln \hat{\mathcal{L}} \rangle \lesssim 1$; however, when the uncertainty is larger than that value, the analysis is biased on average.

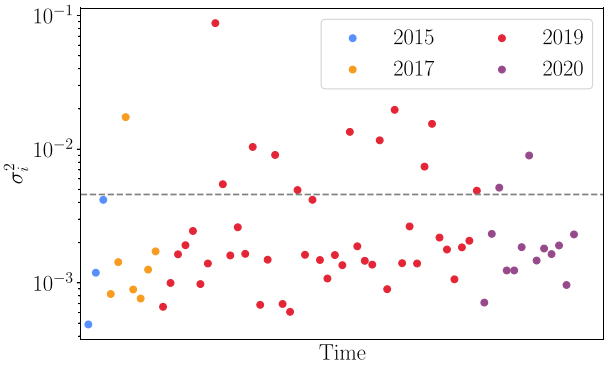

Figure 2. The per-event contribution to the likelihood covariance averaged over the posterior support for our population hyperparameters. We divide the events by the year of the observation, approximately corresponding to different observing runs of Advanced LIGO/Advanced Virgo. We note that there is no obvious trend with time, indicating that we can reliably consider the average uncertainty $\sigma^2_{\rm obs} = \langle \sigma_i^2 \rangle$ (shown by the dashed grey line).

condition that each marginalization over the single event posteriors had $n_{\rm eff} > N$. We consider one of the models that applied both convergence conditions.

We compute the uncertainty in the estimated likelihood for one of the models used in the latest LIGO–Virgo–KAGRA analysis. Specifically, we use the `PowerLaw + Peak` mass model (Talbot & Thrane 2018), `Default` spin model (Talbot & Thrane 2017; Wysocki et al. 2019), and power-law redshift model (Fishbach, Holz & Farr 2018). For our default analysis configuration, we use the same 4278 per-event posterior samples (The LIGO Scientific Collaboration, The Virgo Collaboration & The KAGRA Collaboration 2021a) and injection set (The LIGO Scientific Collaboration, The Virgo Collaboration & The KAGRA Collaboration 2021b) used in the equivalent analysis in The LIGO Scientific Collaboration et al. (2021d) and do not apply any constraints on the convergence of the Monte Carlo integrals.

For all of our analyses, we sample the population posterior using the `nestle` (Barbary 2016) nested sampling package as implemented in `Bilby` (Ashton et al. 2019). We use `GWPopulation` (Talbot et al. 2019) to compute the likelihood function. We use the same prior distributions as in The LIGO Scientific Collaboration et al. (2021d). For each of the posterior samples, we evaluate the uncertainty in each of the 70 Monte Carlo integrals involved (one for each event and the selection function integral).

4.1 Evolution of $\sigma_{\rm obs}$

We begin by testing our assumption that rewriting the total variance in terms of the average per-event variance $\sigma_{\rm obs}$ is reliable. One method in which this could break down is if the average uncertainty changes as the sensitivities of the observatories improve. In Fig. 2, we show the average contribution to the covariance over the posterior for the hyperparameters for each event ordered by observation date. The different colours correspond to events observed in different years. There is no obvious trend over time which validates our approximation of $\sigma^2_{\rm obs} = \langle \sigma_i^2 \rangle$. We show this value with the dashed grey line. The event with the largest contribution to the uncertainty is GW190517, which has masses consistent with the excess at $\sim 35\,{\rm M}_\odot$ and large inferred spins.

4.2 Scaling with the population size

In order to estimate the scaling of the uncertainty with the size of the catalogue, we randomly sample observations from the total catalogue to simulate smaller catalogues and scale the uncertainty on the selection function appropriately. Specifically, we consider catalogues with increments of 5 events from 5 to 65 and all 69 events. For each catalogue size, we sample from the hyperposterior and compute the average variance in the estimated differences of log-likelihood values over the posterior samples $\langle \Delta \ln \hat{\mathcal{L}} \rangle$. We do not apply any of the ad hoc restrictions on Monte Carlo integral convergence proposed in Farr (2019) and The LIGO Scientific Collaboration et al. (2021d) and described above in these analyses. We fit a simple model to the uncertainty coming from the per-event terms and the selection function to obtain fits for the contribution from the individual events and the sensitivity. The model for the total variance is

$$\langle \Delta \ln \hat{\mathcal{L}} \rangle = \sigma^2_{\rm obs} N^a + \sigma^2_{\rm sel} N^b. \tag{18}$$

Here, we emphasize that $\Delta \ln \mathcal{L}$ is proportional to the variance in the estimator and not the standard deviation. We note that equation (11) implies $a = 1$, $b = 2$ while if the assumptions from Essick & Farr (2022) hold we will have $a = 0, b = 1$. We perform this calculation for both the mean variance and the mean covariance over the posterior support.

In Fig. 3, we show the total uncertainty (blue) along with the contributions from the per-event terms (orange) and the selection function (green) as a function of the number of events with the solid curves. The dashed-coloured curves show projections for larger populations based on the analytical fit. The dashed grey lines indicate the number of events in each of the first three gravitational-wave transient catalogues and the grey-shaded region shows a plausible range of observations we may expect after the upcoming fourth gravitational-wave observing run (Petrov et al. 2022; Weizmann Kiendrebeogo et al. 2023). The purple-shaded region shows where, heuristically, we may expect to see noticeable biases, following the criteria developed in Section 3.

We find that in practice, the scaling of the uncertainty lies between the best-case scenario from Farr (2019) and Essick & Farr (2022) and the worst-case scenario in equation (11). Specifically, we find $a =$

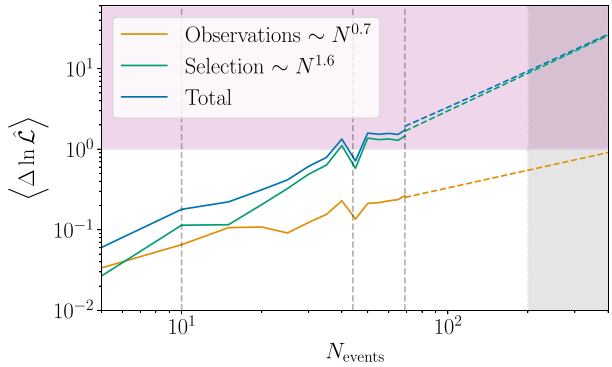

Figure 3. Scaling of the uncertainty in the log-likelihood averaged over the full posterior support with the population size for a simple parametric population model. The dashed vertical lines show the number of confident binary black hole events in the gravitational-wave transient catalogue at the time of publication of GWTC-1 (Abbott et al. 2019a), GWTC-2 (Abbott et al. 2021a), and GWTC-3 (The LIGO Scientific Collaboration et al. 2021c). The grey filled region indicates the projected number of binary black hole observations during the next observing run of the international gravitational-wave detector network (Petrov et al. 2022; Weizmann Kiendrebeogo et al. 2023). The purple shaded region indicates heuristic values for when the uncertainty in the likelihood is likely to cause noticeable bias in the analysis. The solid curves show the empirically obtained uncertainties and the dashed curves are extrapolations based on the power-law fit to the per-event contribution (orange) and the contribution from the selection function (green). The total uncertainty is shown in blue.

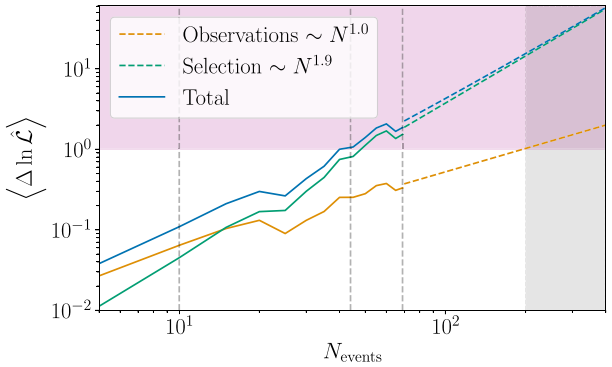

Figure 4. The same a Fig. 3 but with a more flexible model. We note that the same general features are present; however, for this model, the uncertainty grows much more rapidly with population size.

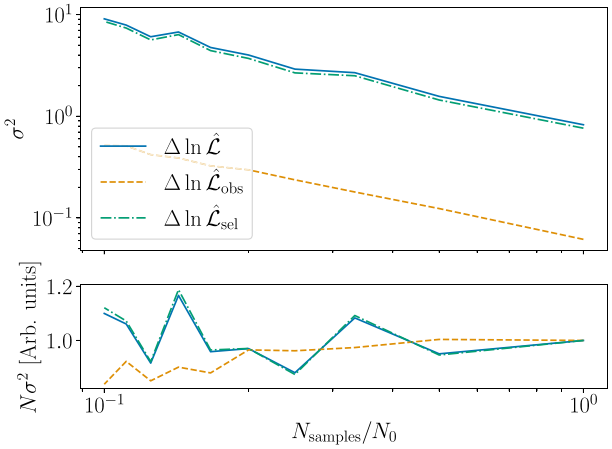

Figure 5. The scaling of the average variance in the log-likelihood with the number of events per Monte Carlo integral. The solid blue, dashed orange, and dash–dotted green curves show the results using the full likelihood, selection function only, and per-observation terms, respectively. In the top panel, we show the variance. In the bottom panel, we show the normalized variance divided the number of samples per integral. As expected, these quantities scale inversely with the number of samples.

0.7, $b = 1.6$. The dominant source of uncertainty is from estimating the selection function when the population is larger than ≈10 events. We note that for populations larger than ≈40 events, the uncertainty is consistently in the purple-shaded region. This is consistent with the fact that ad hoc cuts on the prior space or Monte Carlo convergence were needed to avoid significant biases in Abbott et al. (2021b).

To test if this scaling depends on the functional form used to fit the population, we repeat the above calculation with a more flexible model for the primary mass and spin parameters. Specifically, we take the exponential-spline-modulated power-law mass distribution from Edelman et al. (2022) and the exponential-spline model for black hole spin magnitudes and tilt angles from Golomb & Talbot (2022a). For the mass distribution, we use 10 spline nodes spaced logarithmically between $[2, 100]\,{\rm M}_\odot$ and for the spin parameters we take six nodes equally spaced over the relevant domain. For all spline nodes, our prior on the amplitudes is a unit normal distribution, except for the endpoints for the mass distribution that are fixed to zero.

In Fig. 4, we show the same as Fig. 3 with this more flexible model. We see that the average covariance in both the per-event and selection function terms grows more rapidly in this case than for the simpler model ($a = 1.0$, $b = 1.9$). The more extreme scaling may be due to the greater flexibility of the spline model causing the 'local neighbourhood' assumption of Essick & Farr (2022) to be less appropriate.

4.3 Scaling with the size of Monte Carlo integrals

Having established numerically how the size of the uncertainty in the likelihood estimates varies with the size of the population and configuration settings, we turn to how the number of samples per Monte Carlo integral impacts the uncertainty for this concrete example. To address this, we repeat the uncertainty calculation for the `PowerLaw + Peak` and `Default` configuration above ten times, once using all of the available samples, once with half of the samples, one-third of the samples, etc., down to one-tenth of the samples.

In Fig. 5, we show the mean variance over the posterior distribution as a function of the number of samples per Monte Carlo integral in the upper panel. The solid blue, dashed orange, and dash–dotted green curves show the results using the full likelihood, selection function only, and per-observation terms, respectively. In the lower panel, we show the variance scaled by the number of samples in the integral such that it will be constant if the uncertainty scales linearly with the number of samples. We observe that the variance is consistent with scaling inversely with the number of samples.

4.4 Impact on the inferred astrophysical distributions

To study the impact of the convergence-motivated prior cuts and bias in likelihood estimates, we consider four analysis configurations:

(i) *LVK*. The first configuration is the same one used in the LIGO–Virgo–KAGRA analysis in The LIGO Scientific Collaboration et al. (2021d). This analysis used $\sim 4 \times 10^4$ found injections to estimate

Table 1. Hyperparameters for the injection sets used in each of the analysis configurations we consider as described in Section 4. We additionally list the average variance in the difference between estimated likelihood values.

	$N_{\text{injections}}$	α	m_{max}	m_{min}	δ_{m}	μ_{m}	σ_{m}	λ	$\langle \Delta \ln \hat{\mathcal{L}} \rangle$
LVK	4×10^4	2	100	2	0	–	–	0	0.63
No convergence	4×10^4	2	100	2	0	–	–	0	5.06
Tailored	4×10^4	3.5	105	3	6	33	5	0.04	1.24
More injections	8×10^5	1	100	2	0	–	–	0	0.50
No injections	0	–	–	–	–	–	–	–	0.42

the selection function, and 4278 fiducial posterior samples were used for each event. Specifically, we use the posterior samples released in The LIGO Scientific Collaboration et al. (2021a) and the set of sensitivity injections that combine injections covering the first three observing runs of Advanced LIGO/Advanced Virgo (The LIGO Scientific Collaboration et al. 2021b). For this configuration, there is the prior cut on n_{eff} for each of the Monte Carlo integrals as described at the beginning of this section.

(ii) *No convergence.* The second configuration repeats the analysis from The LIGO Scientific Collaboration et al. (2021d) but removes the prior constraints on n_{eff} in each Monte Carlo integral.

(iii) *Tailored injections.* We replace the injection set released by the LVK, we use synthetic injections drawn using a mass distribution that more closely matches the observed distribution. Specifically, we set the mass distribution using the `PowerLaw + Peak` model using the parameters in Table 1. Since the proposal distribution for our Monte Carlo integral more closely matches the target distribution, we expect this injection set to lead to smaller uncertainties with the same number of found injections.

(iv) *More injections.* Rather than using the $\sim 4 \times 10^4$ found injections used in The LIGO Scientific Collaboration et al. (2021d), we use the $\sim 8 \times 10^5$ synthetic found injections used in Golomb & Talbot (2022a) in order to reduce the uncertainty in the estimate of the selection function. While this uses many more injections, we note that the underlying distribution of signals is different than for the *LVK* configuration.

(v) *No injections.* Rather than using the $\sim 4 \times 10^4$ found injections used in The LIGO Scientific Collaboration et al. (2021d), we ignore the impact of selection effects completely. This will reduce the uncertainty in the estimated likelihoods at the cost of only estimating the observed distribution and not the underlying astrophysical distribution.

For both cases where we use synthetic injection sets, we do not repeat the full injection and recovery using a matched-filter search pipeline due to the large associated computational cost. Instead, we threshold the simulated signals on the optimal signal-to-noise ratio of the injected signal in Gaussian noise with PSDs matching the detector sensitivity during O3 rather than the false-alarm rate (The LIGO Scientific Collaboration et al. 2021c; The LIGO Scientific Collaboration, The Virgo Collaboration & The KAGRA Collaboration 2022). We anticipate that this difference between the detection thresholds does not significantly bias the inferred mass and spin distributions (e.g. Abbott et al. 2019b, 2021b; The LIGO Scientific Collaboration et al. 2021d; Golomb & Talbot 2022a; Essick 2023).

In Table 1, we summarize the population hyperparameters describing the mass distribution used for each injection set. Additionally, we show $\langle \Delta \ln \hat{\mathcal{L}} \rangle$ computed over the respective posterior distributions for the hyperparameters. We find that the *no convergence* case clearly surpasses our threshold. The *tailored* injection set reduces the variance by $\approx 4 \times$ by more closely matching the true underlying distribution but is still in the regime where we expect to see some bias. For the other cases $\langle \Delta \ln \hat{\mathcal{L}} \rangle < 1$ and so, we would expect the results to be unimpacted by Monte Carlo convergence.

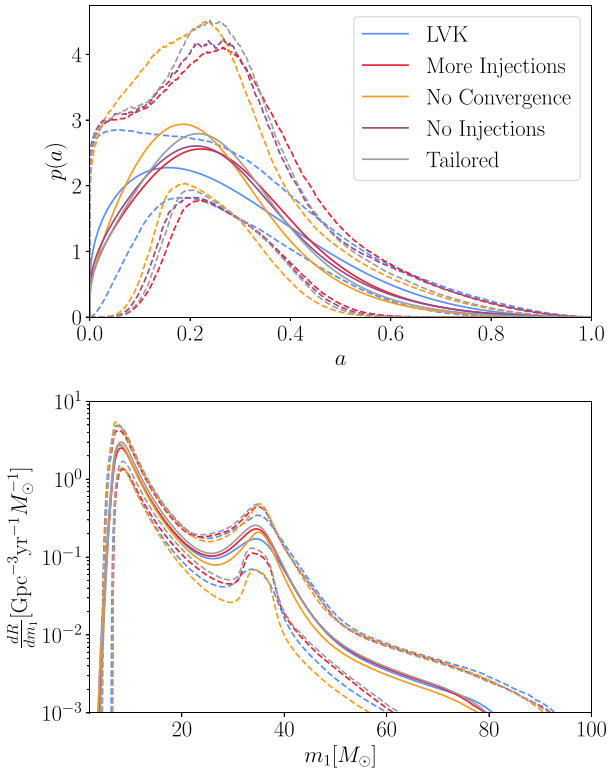

Figure 6. The inferred spin magnitude (top) and primary mass (bottom) distributions for a range of analysis configurations. The solid curves show the mean inferred distribution and the shaded regions show the 90 per cent symmetric credible interval. The blue curves show the results presented in The LIGO Scientific Collaboration et al. (2021d). In orange, we show results obtained using the same input samples but without performing the ad hoc constraints on the number of effective samples per Monte Carlo integral. In red, we show the results when using more found injections to compute the selection function. In purple, we show the results obtained when neglecting the selection function, we note that in this case, we do not show the inferred mass distribution as that is significantly biased by neglecting selection effects. In grey, we show the results obtained using our tailored injection set.

In Fig. 6, we show the inferred distribution for spin magnitude (top panel) and primary mass (lower panel) with our five analysis configurations. We note that the *no injections* configuration is excluded for the primary mass distribution as that distribution is strongly biased by not accounting for selection effects. The solid lines indicate the mean inferred distributions and the dashed curves enclose the 90 per cent uncertainty region. While the uncertainties

of all of the results agree within their error bars there are visible differences between the inferred results. Specifically, we find that for both parameters, the width of the peak at $a \approx 0.2$ and $m_1 \approx 35\,\mathrm{M}_\odot$ are broadest for the result that imposes cuts on the prior based on Monte Carlo convergence (blue) and narrowest for the analysis that has the largest average uncertainties (orange) with the analyses with reduced uncertainty (grey, red, purple) lying in between. This indicates that for commonly used analysis configurations, the inferred shape of features in the distribution of black hole mass and spin are notably impacted by uncertainty in the estimate of the likelihood.

We note that the inferred spin magnitude distributions for the *more injections* and *no injections* configurations are the most consistent. This would be the expected outcome if the impact of the spin magnitude on the selection function is small and the uncertainty in the likelihood estimates is small. We thus infer that the larger injection set is sufficient to remove the bias present when using the found injections released by the LIGO–Virgo–KAGRA collaboration. While the cuts on the number of effective samples in each Monte Carlo integral in the *LVK* configuration control the average uncertainty in the likelihood estimates, the cuts have a visible impact on the inferred distributions.

4.5 Result differences are explainable due to Monte Carlo uncertainty

The posterior predictive distribution (PPD) for the binary parameters is defined as

$$p(\theta|\{d\}) = \int d\Lambda\, p(\theta|\Lambda) p(\Lambda|\{d\}) \approx \frac{1}{N} \sum_{\Lambda_i \sim p(\Lambda|\{d\})}^{N} p(\theta|\Lambda_i). \quad (19)$$

In Fig. 6, the solid curves show the PPD using our different analysis set-ups (solid curves). While the curves are visibly different, we wish to know whether the differences can be explained as the result of statistical fluctuations expected due to the uncertainty in our estimator of the likelihood.

Our aim is to estimate the range of different PPDs we might expect to measure given the PPD with no systematic uncertainty and a covariance $\Sigma(\Lambda, \Lambda')$. In the absence of a ground truth, we take the *no injections* case as our reference analysis as it has the lowest uncertainty estimator of the likelihood and neglect the impact of the per-event integrals as all analyses use the same set of samples for each event.

We begin by taking the samples $\Lambda_i \sim p(\Lambda|\{d\})$ for the reference case. We then construct the covariance matrix by numerically calculating the covariance between the likelihood estimates for every pair of posterior samples. Using this covariance matrix, we generate weights for each of the samples $\delta \sim \mathcal{N}(0, \Sigma(\Lambda, \Lambda'))$. Finally, we compute the PPD using these weights as

$$\hat{p}(\theta|\{d\}) = \frac{\langle \delta(\Lambda_i) p(\theta|\Lambda_i) \rangle_{\Lambda_i \sim p(\Lambda|\{d\})}}{\langle \delta(\Lambda_i) \rangle_{\Lambda_i \sim p(\Lambda|\{d\})}}. \quad (20)$$

By repeating this many times, we can construct the 90 per cent credible interval for the systematic error.

In Fig. 7, we show the same PPDs for the *no injections*, *no convergence*, and *more injections* configurations and the statistical uncertainty for the *no injections* configuration (dotted curves) as in Fig. 6. The estimated systematic uncertainty is shown by the dashed curves. We note that in both cases, the PPDs with our specific realization are entirely consistent with the systematic uncertainty region indicating that the differences in the PPDs can be fully explained by Monte Carlo uncertainty. For the *no convergence* case, the estimated systematic uncertainty is comparable to the statistical uncertainty. One limitation of our method is that the realizations cannot deviate outside the set of samples used for importance sampling and so cannot accurately resolve cases where the systematic uncertainty is larger than the statistical uncertainty in the posterior.

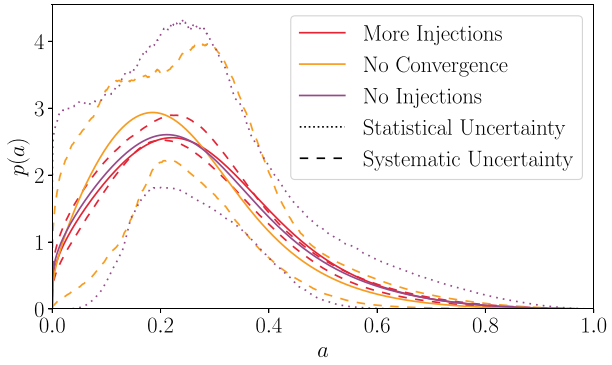

Figure 7. Comparison of statistical and systematic uncertainties in our inference of the distribution of black hole spin magnitude a. The solid curves show the PPD for three of the analysis configurations described in Section 4. The dotted curves show the 5th and 95th percentiles of our statistical uncertainty for the lowest variance analysis (*no injections*). The orange and green dashed curves show the 5th and 95th percentiles of the additional systematic uncertainty from estimating the selection function. We note that for the *more injections* case the systematic uncertainty is much smaller than the statistical. However, for the *no convergence* case the systematic uncertainty is comparable to the statistical.

5 CONCLUSIONS

Often when performing Bayesian inference, we cannot calculate the true likelihood function, but rather a computationally tractable approximation. For example, the use of Monte Carlo integration to approximate marginal likelihoods is widespread in population inference in gravitational-wave astronomy and beyond. However, often, the uncertainty associated with these finite numerical integrals is neglected. We specifically examine the requirement of performing unbiased population inference on binary black holes with Monte Carlo integrals used to marginalize over the parameters of the individual sources. Previous work has claimed that as the size of the population increases, keeping the allowed uncertainty in each marginal likelihood constant (e.g. the number of samples used in each Monte Carlo integral does not have to increase with the population size) is sufficient for precise inference of the population parameters (Essick & Farr 2022).

In this work, through a series of numerical experiments, we demonstrated that for models widely used to characterize the population of merging black hole binaries, this scaling is insufficient and the actual scaling depends on the functional form chosen to fit the distribution. Failing to use a larger number of samples per Monte Carlo integral will result in an increasingly significant bias in the recovery of the population as the number of observations grows. We recommend that the calculations described in this work be routinely performed for any population analysis to identify cases where the inference may be impacted by Monte Carlo uncertainty. We provide scripts to evaluate this in the accompanying code release.

By considering a model routinely employed to characterize the distribution of masses and spins of merging compact binaries, we found that the uncertainty in the likelihoods estimated as part of population inference on the gravitational-wave transient catalogue

is sufficient to lead to noticeable bias with the current size of the gravitational-wave transient catalogue. Additionally, by examining the impact of the specific choice of input samples and convergence requirements, we observed changes in the width of features in the distribution of black hole masses and spin magnitudes. While the differences observed here are within the statistical uncertainties, more significant biases have been observed when using more flexible models, e.g. appendix B of Golomb & Talbot (2022a).

The results presented in this work are somewhat in conflict with the results from Essick & Farr (2022). One difference between this work and theirs is that in Essick & Farr (2022) the authors only consider population models where the uncertainties on each measurement are smaller than the width of the population. In contrast, in many of the models considered here, including the models for the black hole mass and spin, the individual measurements are broader than the underlying population model. The spin magnitudes of individual black holes are very poorly measured, and so the individual posterior distributions are inevitably broader than the population for the majority of systems. For black hole masses, one might think that the total population model is broader than individual measurements; very few black holes are consistent with masses ranging from 5 to 80 $M_\odot$. However, the relevant quantity is not the whole domain of the model, but rather change in the population model over the individual event posterior support. For events intersecting the Gaussian peak at $\sim$35 $M_\odot$, the uncertainty in the mass is almost always larger than the preferred width of 1–5 $M_\odot$. We defer detailed investigations into whether this is a relevant difference to future work.

In the next observing runs, we can conservatively expect the size of the observed population to double or triple (Petrov et al. 2022). With a population of this size, we can expect that if we continue to use the same number of samples per Monte Carlo integral, the variance in the log-likelihood will reach $\sim$4–10 and we will be in danger of making severely biased inferences. In order to avoid this, we will either need to use dramatically more samples in our Monte Carlo integrals or consider novel approaches.

There are a number of questions posed by our results that should be explored in future work. In Section 4, we found an approximate scaling for the growth of the uncertainty with the population size; developing a theoretical understanding of this scaling may prove instructive in developing improved methods to deal with large populations. Ensuring accurate estimation of the population likelihood is an increasingly complex task as the population size increases, and so we will require increasingly sophisticated methods.

As shown in Section 4, a simple method to reduce the uncertainty in Monte Carlo integrals is to reduce the divergence between the initial model and the target model. Fortunately, as the size of the population grows, we can use our existing knowledge to generate initial models that well approximate the true distribution, e.g. by drawing our injections to determine the survey sensitivity by our best estimate of the true population. Additionally, one can recast the Monte Carlo integral using continuous representations of the per-event likelihoods in order to minimize the uncertainty, e.g. Wysocki et al. (2019) and Golomb & Talbot (2022b). Finally, one can limit the analysis to only consider slowly varying source models, e.g. by imposing smoothing priors on the population model (Callister & Farr 2023; Edelman, Farr & Doctor 2023). However, this can lead to missing any sharp features in the distribution.

Each of these improvements is likely to fail eventually, and new methods will be needed. One possibility is to remove the Monte Carlo integral to determine the selection function and instead directly model the observed distribution of compact binaries. If desired, the astrophysical distribution can then be obtained as a post-processing stage using continuous estimates of the selection function such as those in e.g. Veske et al. (2021) and Talbot & Thrane (2022). Similar approaches have been proposed for analyses of online polling data (e.g. Elliott & Valliant 2017; Liu, Scholtus & De Waal 2022). Since the contribution of the uncertainty from estimating the selection function grows most rapidly with population size, this will significantly alleviate bias in the inferred distribution.

While we considered uncertainties in the likelihood function used for gravitational-wave population inference, our analysis holds for any problem where there are parameter-dependent biases in calculating likelihoods. For example, when characterizing individual compact binary coalescences, there are a number of sources of bias in the likelihood function, including waveform systematics (Pürrer & Haster 2020), detector calibration uncertainty (Payne et al. 2020; Vitale et al. 2021), and likelihood acceleration methods (Smith et al. 2016; Leslie, Dai & Pratten 2021; Morisaki 2021). While the specific results in Section 4 will not be relevant to these cases, the general expressions in Sections 2.1 and 3 are relevant.

ACKNOWLEDGEMENTS

We thank Reed Essick and Will Farr for multiple discussions. We thank Sylvia Biscoveanu, Tom Callister, Jack Heinzel, Eric Thrane, and Salvatore Vitale for useful conversations and comments. JG acknowledges funding from National Science Foundation (NSF) grants 2207758 and PHY-1764464. CT was supported by an MIT Kavli Institute (MKI) Kavli Fellowship. This material is based upon work supported by NSF's LIGO Laboratory, which is a major facility fully funded by the National Science Foundation. This work used computational resources provided by the Caltech LIGO Lab and supported by NSF grants PHY-0757058 and PHY-0823459. This work made use of the following softwares: NUMPY (Oliphant 2006; Harris et al. 2020), CUPY (Okuta et al. 2017), NESTLE (Barbary 2016), BILBY (Ashton et al. 2019), GWPOPULATION (Talbot et al. 2019), and GWPOPULATION_PIPE (Talbot 2021).

DATA AVAILABILITY

This work used publicly available data produced by the LIGO–Virgo–KAGRA collaboration (The LIGO Scientific Collaboration et al. 2021a,b). We additionally used a larger synthetic injection set that we will make available on request. Scripts and `Jupyter` notebooks required to reproduce this analysis are available at github.com/ColmTalbot/monte-carlo-uncertainty-scaling.

REFERENCES

Abbott B. P. et al., 2019a, Phys. Rev. X, 9, 031040
Abbott B. P. et al., 2019b, ApJ, 882, L24
Abbott R. et al., 2021a, Phys. Rev. X, 11, 021053
Abbott R. et al., 2021b, ApJ, 913, L7
Abbott R. et al., The LIGO Scientific Collaboration, The Virgo Collaboration, The KAGRA Collaboration, 2021d, Phys. Rev. X 13, 011048, Published 29 March 2023
Acernese F. et al., 2015, Class. Quantum Gravity, 32, 024001
Allen B., Anderson W. G., Brady P. R., Brown D. A., Creighton J. D. E., 2012, Phys. Rev. D, 85, 122006
Ashton G. et al., 2019, ApJS, 241, 27
Barbary K., 2016, Nestle, Jan 29, 2019, https://github.com/kbarbary/nestle
Callister T. A., Farr W. M., 2023, preprint (arXiv:2302.07289)
Campanelli M., Lousto C. O., Zlochower Y., 2006, Phys. Rev. D, 74, 041501
Edelman B., Doctor Z., Godfrey J., Farr B., 2022, ApJ, 924, 101
Edelman B., Farr B., Doctor Z., 2023, ApJ, 946, 16

Elliott M. R., Valliant R., 2017, Stat. Sci., 32, 249
Essick R., 2023, preprint (arXiv:2307.02765)
Essick R., Farr W., 2022, preprint (arXiv:2204.00461)
Farr W. M., 2019, Res. Notes Am. Astron. Soc., 3, 66
Farr W. M., Gair J. R., Mandel I., Cutler C., 2015, Phys. Rev. D, 91, 023005
Finn L. S., Chernoff D. F., 1993, Phys. Rev. D, 47, 2198
Fishbach M., Holz D. E., Farr W. M., 2018, ApJ, 863, L41
Golomb J., Talbot C., 2022a, preprint (arXiv:2210.12287)
Golomb J., Talbot C., 2022b, ApJ, 926, 79
Harris C. R. et al., 2020, Nature, 585, 357
Kish L., 1995, Survey Sampling, 3rd edn. Wiley-Interscience, Oxford
Leslie N., Dai L., Pratten G., 2021, Phys. Rev. D, 104, 123030
LIGO Scientific Collaboration et al., 2015, Class. Quantum Gravity, 32, 074001
Liu A.-C., Scholtus S., De Waal T., 2022, J. Surv. Stat. Meth., p. smac029
Loredo T. J., 2004, in bayesian inference and maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. Am. Inst. Phys., New York, p. 195
Mandel I., Farr W. M., Gair J. R., 2019, MNRAS, 486, 1086
Morisaki S., 2021, Phys. Rev. D, 104, 044062
Nitz A. H. et al., 2023, ApJ, 946, 59
Okuta R., Unno Y., Nishino D., Hido S., Loomis C., 2017, in Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS), http://learningsys.org/nips17/assets/papers/paper_16.pdf
Oliphant T. E., 2006, A Guide to NumPy, Vol. 1. Trelgol Publishing, USA
Olsen S., Venumadhav T., Mushkin J., Roulet J., Zackay B., Zaldarriaga M., 2022, Phys. Rev. D, 106, 043009
Payne E., Talbot C., Lasky P. D., Thrane E., Kissel J. S., 2020, Phys. Rev. D, 102, 122004
Petrov P. et al., 2022, ApJ, 924, 54
Pürrer M., Haster C.-J., 2020, Phys. Rev. Res., 2, 023151
Smith R., Field S. E., Blackburn K., Haster C.-J., Pürrer M., Raymond V., Schmidt P., 2016, Phys. Rev. D, 94, 044031
Talbot C., 2021, GWPopulation pipe, 2022-12-13, https://git.ligo.org/RatesAndPopulations/gwpopulation_pipe
Talbot C., Smith R., Thrane E., Poole G. B., 2019, Phys. Rev. D, 100, 043030
Talbot C., Thrane E., 2017, Phys. Rev. D, 96, 023012
Talbot C., Thrane E., 2018, ApJ, 856, 173
Talbot C., Thrane E., 2022, ApJ, 927, 76
The LIGO Scientific Collaboration, The Virgo Collaboration, The KAGRA Collaboration, 2021a, GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run – Parameter Estimation Data Release, https://doi.org/10.5281/zenodo.5546663
The LIGO Scientific Collaboration, The Virgo Collaboration, The KAGRA Collaboration, 2021b, GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run – O3 Search Sensitivity Estimates. https://doi.org/10.5281/zenodo.5546676
The LIGO Scientific Collaboration, The Virgo Collaboration, The KAGRA Collaboration, 2021c, preprint (arXiv:2111.03606)
The LIGO Scientific Collaboration, The Virgo Collaboration, The KAGRA Collaboration, 2022, GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run – Data Behind the Figures, https://doi.org/10.5281/zenodo.6368595
Thrane E., Talbot C., 2019, Publ. Astron. Soc. Aust., 36, e010
Veske D., Bartos I., Márka Z., Márka S., 2021, ApJ, 922, 258
Vitale S., Gerosa D., Farr W. M., Taylor S. R., 2022, in Bambi C., ed., Handbook of Gravitational Wave Astronomy. Springer Nature, London, p. 45
Vitale S., Haster C.-J., Sun L., Farr B., Goetz E., Kissel J., Cahillane C., 2021, Phys. Rev. D, 103, 063016
Weizmann Kiendrebeogo R. et al., 2023, preprint (arXiv:2306.09234)
Wysocki D., Lange J., O'Shaughnessy R., 2019, Phys. Rev. D, 100, 043012

This paper has been typeset from a TEX/LATEX file prepared by the author.