A Fisher matrix for gravitational-wave population inference

Jonathan R. Gair,[1][★] Andrea Antonelli [2] and Riccardo Barbieri[1]

[1]*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, Potsdam-Golm 14476, Germany*
[2]*Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA*

ABSTRACT

We derive a Fisher matrix for the parameters characterizing a population of gravitational-wave events. This provides a guide to the precision with which population parameters can be estimated with multiple observations, which becomes increasingly accurate as the number of events and the signal-to-noise ratio of the sampled events increase. The formalism takes into account individual event measurement uncertainties and selection effects, and can be applied to arbitrary population models. We illustrate the framework with two examples: an analytical calculation of the Fisher matrix for the mean and variance of a Gaussian model describing a population affected by selection effects, and an estimation of the precision with which the slope of a power-law distribution of supermassive black hole masses can be measured using extreme-mass-ratio inspiral observations. We compare the Fisher predictions to results from Monte Carlo analyses, finding very good agreement.

Key words: gravitational waves – methods: analytical.

1 INTRODUCTION

Population analyses aim at inferring the parameters that describe the distribution of the properties of a set of observed events drawn from a common population. In the context of gravitational-wave (GW) astrophysics, such analyses have been carried out for the 90 coalescing compact-object binaries that have so far been observed by ground-based gravitational wave detectors, and reported in the third gravitational wave transient catalogue, GWTC-3 (Abbott et al. 2021a, b, c). Together with simulation-based studies (Taylor & Gerosa 2018), these population analyses aimed at understanding the astrophysical processes that lead to the formation of the binaries (Abbott et al. 2019; Rodriguez et al. 2020), their evolution (Fishbach et al. 2021; Mould et al. 2022) and at measuring the current parameters describing their population (Vitale, Biscoveanu & Talbot 2022). Furthermore, population analyses are also used to constrain cosmic expansion history by estimating parameters like the Hubble constant (Mastrogiovanni et al. 2021; Abbott et al. 2021d; Mancarella, Genoud-Prachex & Maggiore 2022; Mukherjee et al. 2022).

Given a set of observed events, the usual approach to estimate distribution parameters is to complete a Bayesian hierarchical analysis using techniques such as Markov Chain Monte Carlo (MCMC). While these are the most reliable way to obtain posterior samples from actual data, they are typically computationally expensive and so it can become impractical to use these approaches to make forecasts for future observations that include surveys over parameter space. However, such surveys are crucial for scoping out the science cases of future detectors, such as the Einstein Telescope (Punturo et al. 2010), Cosmic Explorer (Reitze et al. 2019), and the spaceborne

LISA mission (Amaro-Seoane et al. 2017), all of which are expected to detect thousands of sources from multiple populations. For explorations of this nature, one can trade off accuracy in the estimates of parameter-measurement precision for computational speed by using approximations that are valid in the limit of high signal-to-noise ratio (SNR). In the context of *source* parameters for *individual* signals, the Fisher matrix is commonly used to cheaply assess the measurement precision of a parameter (Vallisneri 2008). Within the linear-signal approximation, valid for high SNR sources, the inverse of the Fisher matrix is an approximation to the covariance matrix and therefore the width of the likelihood function. Under the assumption of flat priors, it also approximates the shape of the Bayesian posterior probability distribution we would expect to obtain in an MCMC analyses. For a parameter set $\vec{\lambda}$, the Fisher matrix can be written in general terms as the expectation value over the data generating process of derivatives of the log likelihood $p(\mathbf{d}|\vec{\lambda})$,

$$(\Gamma_\lambda)_{ij} = \mathbb{E}\left[-\frac{\partial^2 \ln p(\mathbf{d}|\vec{\lambda})}{\partial \lambda^i \partial \lambda^j} \right]. \tag{1}$$

While this provides a guide to measurement uncertainties for individual events, the Fisher matrix does not directly provide an indication of how well the properties of the population can be inferred when those events are subsequently combined in a hierarchical model. In this paper, we address this shortcoming by deriving a Fisher Matrix for the population parameters assuming Gaussian noise and using the likelihood for population inference in the presence of selection effects from Mandel, Farr & Gair (2019). The expression we obtain is valid for high SNRs and small biases in the individual events' parameters. We illustrate our formalism with two examples. First, we consider a 'Gaussian–Gaussian' case, in which both noise and the data generation processes are normally distributed, and check our expressions against the direct calculation of the Fisher matrix as

★ E-mail: jonathan.gair@aei.mpg.de

an expectation value over data realizations. We also perform MCMC analyses with and without selection effects, as a cross-check to verify our results. Secondly, we consider the more astrophysically relevant case of a power-law distributed population (while still assuming Gaussian noise) and again validate our results against MCMC analyses with and without selection effects. We generally find an excellent agreement between the Fisher and MCMC estimates, while confirming results in Gair, Tang & Volonteri (2010) for the latter scenario.

The paper is organized as follows. In Section 2, we describe our population Fisher Matrix formalism, highlighting the main assumptions and steps to obtain the result. A derivation of corrections to this formula and their scalings is found in Appendix B. In Section 3, we consider the Gaussian population model, checking our formula against a direct calculation of the Fisher Matrix and an MCMC analysis. In Section 4, we consider the case of inference of a power-law massive black hole mass distribution using extreme-mass-ratio inspiral (EMRI) observations, once again comparing the result against MCMC. Finally, in Section 5, we discuss our results and prospects for future work. The framework we develop here could be applied in a wide variety of contexts. The focus on gravitational wave detectors and the choice of the examples provided here are driven purely by the authors' areas of expertise. The results can be fully reproduced with codes made publicly available at https://github.com/aantonelli94/PopFisher.

2 THE FISHER MATRIX FOR POPULATION DISTRIBUTIONS

The standard model used to represent the data stream, $\mathbf{d}$, of a gravitational wave detector is as a linear combination of a signal, $\mathbf{h}(\vec{\theta})$, dependent on some parameters $\vec{\theta}$, and noise, $\mathbf{n}$, that is usually assumed to be a realization of a stationary and Gaussian stochastic process described by a power spectral density $S_h(f)$,

$$\mathbf{d} = \mathbf{h}(\vec{\theta}) + \mathbf{n}, \qquad \langle \tilde{n}^*(f)\tilde{n}(f') \rangle = S_h(f)\delta(f - f'). \tag{2}$$

In this model, the likelihood is

$$p(\mathbf{d}|\vec{\theta}) \propto \exp\left[-\frac{1}{2}\left(\mathbf{d} - \mathbf{h}(\vec{\theta})|\mathbf{d} - \mathbf{h}(\vec{\theta})\right)\right],$$

where $(\mathbf{a}|\mathbf{b}) = 4\mathrm{Re}\displaystyle\int_0^\infty \frac{\tilde{a}^*(f)\tilde{b}(f)}{S_h(f)}\,df. \tag{3}$

To understand the precision with which gravitational wave observations can determine the parameters of a source, it is common to compute the Fisher information matrix, defined by

$$(\Gamma_\theta)_{ij} = \mathbb{E}\left[\frac{\partial \ln p(\mathbf{d}|\vec{\theta})}{\partial \theta^i}\frac{\partial \ln p(\mathbf{d}|\vec{\theta})}{\partial \theta^j}\right], \tag{4}$$

where the expectation value is taken over realizations of the data drawn from the data generating process, $\mathbf{d}$. For the gravitational wave detector likelihood in equation (3), the Fisher information matrix can be seen to reduce to

$$(\Gamma_\theta)_{ij} = \left(\frac{\partial \mathbf{h}}{\partial \theta^i}\bigg|\frac{\partial \mathbf{h}}{\partial \theta^j}\right), \tag{5}$$

where we are using the inner product introduced in equation (3). The Fisher matrix provides a leading order approximation to the shape of the likelihood and hence also the Bayesian posterior when using priors that are approximately flat over the support of the likelihood. It becomes an increasingly good guide to the precision of parameter estimation as the SNR with which the source is observed increases.

In population inference, we are no longer primarily interested in the parameters of the individual events, but in the parameters that characterize the population from which the individual events are drawn. We assume that we have some population model, $p(\vec{\theta}|\vec{\lambda})$, that describes the probability distribution of the parameters, $\vec{\theta}$, of individual events drawn randomly from a population characterized by parameters, $\vec{\lambda}$. We want to infer the parameters of the population by combining the information from many observed events. For a given choice of population parameters, the distribution of observed data sets is characterized by

$$p(\mathbf{d}|\vec{\lambda}) = \frac{p_{\mathrm{full}}(\mathbf{d}|\vec{\lambda})}{P_{\mathrm{det}}(\vec{\lambda})} \tag{6}$$

where $p_{\mathrm{full}}(\mathbf{d}|\vec{\lambda}) = \int p(\mathbf{d}|\vec{\theta})p(\vec{\theta}|\vec{\lambda})\,d\vec{\theta}$

$P_{\mathrm{det}}(\vec{\lambda}) = \int P_{\mathrm{det}}(\vec{\theta})p(\vec{\theta}|\vec{\lambda})\,d\vec{\theta}$

$P_{\mathrm{det}}(\vec{\theta}) = \int_{\mathbf{d}>\mathrm{thresh}} p(\mathbf{d}|\vec{\theta})\,d\mathbf{d}. \tag{7}$

Here and elsewhere, we will use lowercase $p(x)$ to denote probability density functions, which have units of $1/x$, and uppercase $P(x)$ to denote cumulative density functions, which are dimensionless. This expression accounts for the fact that not all events that occur in the Universe are detected. Detection is a property of the observed data, $\mathbf{d}$, and the last integral is over all data sets that would pass the threshold to be counted as a detected event and hence included in the population inference. The normalization term, $P_{\mathrm{det}}(\vec{\lambda})$, depends only on the population parameters and represents the fraction of events in the Universe that are detectable. We refer the reader to Mandel et al. 2019 for further details. This form of the likelihood assumes that the number of events observed in a fixed time period does not convey any information about the population parameters. However, the precision with which the population parameters are estimated asymptotically is independent of that assumption. This is discussed in more detail in Appendix A.

Equation (1) is the equivalent of equation (4) for this population likelihood, and so it should give a guide to the precision with which the population parameters can be measured. Note that the two forms of the expression are slightly different, but it is straightforward to show that the two results are equivalent by integrating by parts and using conservation of probability. This will be a good guide for a 'high SNR', which for populations means a large number of observed events. The fact that equation (1) is a good approximation to the precision of population inference can be seen as follows. In a general population inference problem, we have observed a set of events, indexed by i, with corresponding data sets $\{\mathbf{d}_i\}$. The posterior distribution on the population parameters from this set of events can be found from Bayes' theorem and takes the form

$$p(\vec{\lambda}|\{\mathbf{d}_i\}) \propto \pi(\vec{\lambda})\prod_{i=1}^n p(\mathbf{d}_i|\vec{\lambda}), \tag{8}$$

where n is the total number of events observed, $\pi(\vec{\lambda})$ is the prior on the population parameters, and $p(\mathbf{d}_i|\vec{\lambda})$ is the likelihood of the population parameters $\vec{\lambda}$ for data set $\mathbf{d}_i$. The log posterior is

$$\ln p(\vec{\lambda}|\{\mathbf{d}_i\}) \propto \ln \pi(\vec{\lambda}) + \sum_{i=1}^n \ln p(\mathbf{d}_i|\vec{\lambda}). \tag{9}$$

The latter quantity is a sum of independent random variables (assuming that all observations are independent). In the limit that $n \to \infty$, we can use the central limit theorem to deduce

$$\frac{1}{n}\sum_{i=1}^n \ln p(\mathbf{d}_i|\vec{\lambda}) \sim N\left(\mu(\vec{\lambda}|\vec{\lambda}_t), \frac{\sigma^2(\vec{\lambda}|\vec{\lambda}_t)}{n}\right), \tag{10}$$

where

$$\mu(\vec{\lambda}|\vec{\lambda}_t) = \mathbb{E}\left[\ln p(\mathbf{d}|\vec{\lambda})\right], \quad \sigma^2(\vec{\lambda}|\vec{\lambda}_t) = \mathbb{E}\left[(\ln p(\mathbf{d}|\vec{\lambda}) - \mu)^2\right] \quad (11)$$

and the expectation value is taken over the data generating process, which we assume to be consistent with the likelihood we are using, evaluated for the true values of the population parameters $\vec{\lambda}_t$.

Since

$$\frac{\partial \mu}{\partial \lambda^i} = \int p(\mathbf{d}|\vec{\lambda}_t) \frac{1}{p(\mathbf{d}|\vec{\lambda})} \frac{\partial p(\mathbf{d}|\vec{\lambda})}{\partial \lambda^i} \, d\mathbf{d}$$

$$\Rightarrow \quad \frac{\partial \mu}{\partial \lambda^i}\bigg|_{\vec{\lambda}=\vec{\lambda}_t} = \int \frac{\partial p(\mathbf{d}|\vec{\lambda})}{\partial \lambda^i} \, d\mathbf{d} = \frac{\partial}{\partial \lambda^i} \int p(\mathbf{d}|\vec{\lambda}) \, d\mathbf{d}$$

$$= \frac{\partial}{\partial \lambda^i}(1) = 0, \quad (12)$$

we deduce that the population likelihood is peaked at the true parameters asymptotically (this is not true on average for a finite number of observations, as discussed in Appendix B). Note that this happens by virtue of the assumed consistency between the likelihood and the data-generating process, and would not be the case if the likelihood was only an approximation to the true population. As $n \to \infty$, the log posterior converges to the function $n\mu(\vec{\lambda})$, and so the posterior becomes increasingly concentrated around $\vec{\lambda}_t$. Expanding the function $\mu(\vec{\lambda}|\vec{\lambda}_t)$ near $\vec{\lambda}_t$ we have

$$\mu(\vec{\lambda}|\vec{\lambda}_t) = \mu\left(\vec{\lambda}_t|\vec{\lambda}_t\right)$$
$$+ \frac{1}{2}\left(\frac{d^2\mu}{d\lambda^i d\lambda^j}\right)_{\vec{\lambda}_t} (\lambda^i - \lambda_t^i)(\lambda^j - \lambda_t^j) + \cdots. \quad (13)$$

We deduce that the asymptotic covariance matrix is Γ_λ^{-1}/n, where

$$(\Gamma_\lambda)_{ij} = -\left(\frac{d^2\mu}{d\lambda^i d\lambda^j}\right)_{\vec{\lambda}_t} = \mathbb{E}\left[-\frac{\partial^2 \ln p(\mathbf{d}|\vec{\lambda})}{\partial \lambda^i \partial \lambda^j}\right]. \quad (14)$$

In the last equality, we use equation (12). This justifies the use of equation (1) to characterize the precision of parameter estimation in the limit $n \to \infty$. It becomes increasingly reliable as $n \to \infty$, as corrections to this formula scale like $n^{-1/2}$ relative to leading order. This is justified in more detail in Appendix B.

This result can be evaluated at various levels of approximation. The full asymptotic posterior is described by the function $\mu(\vec{\lambda}|\vec{\lambda}_t)$, which can be evaluated through Monte Carlo integration. This is computationally expensive as it requires evaluation over different choices of $\vec{\lambda}$ and $\vec{\lambda}_t$. The next level of approximation is to evaluate equation (1) directly. This makes a linear signal approximation in the population parameters, but no approximation to the evaluation of $p(\mathbf{d}|\vec{\lambda})$. This is less complex because evaluation is only needed in the vicinity of $\vec{\lambda}_t$. A final level of approximation is to simplify $p(\mathbf{d}|\vec{\lambda})$ by using the linear signal approximation for the individual event parameters as well. This is the approach we will now describe.

We consider a single observation of a source with parameters $\vec{\theta}_0$, and data $\mathbf{d} = \mathbf{h}(\vec{\theta}_0) + \mathbf{n}$. Taking the expectation value over the true data distribution then reduces to taking the expectation value over the distribution of the noise $\mathbf{n}$ and the distribution of the parameters $\vec{\theta}_0$, which is $p(\vec{\theta}_0|\vec{\lambda}_t)$. Under the linear signal approximation, we expand

$$\mathbf{h}(\vec{\theta}) = \mathbf{h}(\vec{\theta}_0) + \frac{\partial \mathbf{h}}{\partial \theta^i} \Delta\theta^i \quad (15)$$

where $\Delta\theta^i = \theta^i - \theta_0^i$. The gravitational wave likelihood can then be written

$$\tilde{p}(\{\mathbf{d}\}|\vec{\theta}) \propto \exp\left[-\frac{1}{2}(\mathbf{d} - \mathbf{h}(\vec{\theta})|\mathbf{d} - \mathbf{h}(\vec{\theta}))\right]$$
$$\approx \exp\left[-\frac{1}{2}(\mathbf{n}|\mathbf{n}) + N_i \Delta\theta^i - \frac{1}{2}(\Gamma_\theta)_{ij}\Delta\theta^i \Delta\theta^j\right]$$

where $N_i = \left(\frac{\partial \mathbf{h}}{\partial \theta^i}\bigg|\mathbf{n}\right)$ (16)

and $(\Gamma_\theta)_{ij}$ is the single source Fisher matrix defined in equation (5). This is to be evaluated at θ_0 and therefore has a dependence on those parameters. We similarly expand the source prior term

$$\ln p(\vec{\theta}|\vec{\lambda}) = \ln p(\vec{\theta}_0|\vec{\lambda}) + P_i\Delta\theta^i - \frac{1}{2}H_{ij}\Delta\theta^i \Delta\theta^j + \cdots$$

where $P_i = \frac{\partial \ln p(\vec{\theta}|\vec{\lambda})}{\partial \theta^i}, \qquad H_{ij} = -\frac{\partial^2 \ln p(\vec{\theta}|\vec{\lambda})}{\partial \theta^i \partial \theta^j},$ (17)

in which the derivatives are evaluated at the parameter space point θ_0. Substituting the preceding two expressions into equation (7) and integrating over $\vec{\theta}$, which is equivalent to integrating over $\Delta\vec{\theta}$ in the linear signal approximation, we obtain

$$\tilde{p}(\{\mathbf{d}\}|\vec{\lambda}) \approx \frac{\exp[-(\mathbf{n}|\mathbf{n})/2]}{P_{\det}(\vec{\lambda})} \int d\Delta\vec{\theta} \left[p(\vec{\theta}_0|\vec{\lambda})\right.$$
$$\exp\left\{-\frac{1}{2}(\Gamma_{ij} + H_{ij})(\Delta\theta^i - \Delta\theta_{\mathrm{bf}}^i)(\Delta\theta^j - \Delta\theta_{\mathrm{bf}}^j)\right.$$
$$\left.\left.+ \frac{1}{2}(N_i + P_i)(\Gamma + H)_{ij}^{-1}(N_j + P_j)\right\}\right]$$
$$= (2\pi)^{N/2} \frac{p(\vec{\theta}_0|\vec{\lambda}) \exp[-(\mathbf{n}|\mathbf{n})/2]}{P_{\det}(\vec{\lambda})\sqrt{\det(\Gamma + H)}}$$
$$\times \exp\left[\frac{1}{2}(N_i + P_i)(\Gamma + H)_{ij}^{-1}(N_j + P_j)\right], \quad (18)$$

where we have written

$$\Delta\theta_{\mathrm{bf}}^i = (\Gamma + H)_{ik}^{-1}(N_k + P_k).$$

This is the point at which the likelihood is maximized and hence is the 'best-fit' point in parameter space. We can now evaluate the population Fisher matrix using the expression

$$-(\Gamma_\lambda)_{ij} = \int \left(\frac{\partial^2 \ln p(\mathbf{d}|\vec{\lambda})}{\partial \lambda^i \partial \lambda^j}\right)_{\vec{\lambda}_t} p(\mathbf{d}|\vec{\lambda}_t) d\mathbf{d}$$
$$= \int \int \left(\frac{\partial^2 \ln p(\mathbf{d}|\vec{\lambda})}{\partial \lambda^i \partial \lambda^j}\right)_{\vec{\lambda}_t} p(\vec{\theta}_0|\vec{\lambda}_t) p(\mathbf{n}|\vec{\theta}_0) d\mathbf{n} d\vec{\theta}_0. \quad (19)$$

In the above, the integral over the noise distribution is conditioned on $\vec{\theta}_0$ because of selection effects. This integral is over all noise realizations that ensure $\mathbf{d} = \mathbf{h}(\vec{\theta}_0) + \mathbf{n}$ is above the detection threshold. Substituting equation (18) into the above, we obtain a sequence of terms. To simplify these, we carry out the integral over the noise, $\mathbf{n}$. The only terms in equation (18) that depend on $\mathbf{n}$ are N_i and the prefactor $\exp[-(\mathbf{n}|\mathbf{n})/2]$. The latter enters $\ln \tilde{p}$ additively and has no dependence on the population parameters, so it does not contribute to the final result. The former term also has no explicit dependence on the population parameters, but it appears multiplied by terms that do. There are thus three distinct types of term that appear in the argument of the integral – terms that have no explicit dependence on $\mathbf{n}$, terms that are linear in N_i, and terms that are quadratic in N_i. We

define these integrals as follows

$$P_{\text{det}}(\vec{\theta}_0) = \int p(\mathbf{n}|\vec{\theta}_0)\mathrm{d}\mathbf{n}$$

$$D_i \equiv \int \left(\mathbf{n} \left| \frac{\partial \mathbf{h}}{\partial \theta^i} \right. \right) p(\mathbf{n})\mathrm{d}\mathbf{n} = \frac{\partial P_{\text{det}}(\vec{\theta}_0)}{\partial \theta^i} ,$$

$$D_{ij} \equiv \int \left(\mathbf{n} \left| \frac{\partial \mathbf{h}}{\partial \theta^i} \right. \right) \left(\mathbf{n} \left| \frac{\partial \mathbf{h}}{\partial \theta^j} \right. \right) p(\mathbf{n})\mathrm{d}\mathbf{n}. \tag{20}$$

Using these expressions to carry out the integrals over $\mathbf{n}$, we obtain the final result

$$(\Gamma_\lambda)_{ij} \equiv (\Gamma_{\text{I}})_{ij} + (\Gamma_{\text{II}})_{ij} + (\Gamma_{\text{III}})_{ij} + (\Gamma_{\text{IV}})_{ij} + (\Gamma_{\text{V}})_{ij}, \tag{21}$$

with

$$(\Gamma_{\text{I}})_{ij} = -\int \frac{\partial^2 \ln(p(\vec{\theta}_0|\vec{\lambda})/P_{\text{det}}(\vec{\lambda}))}{\partial \lambda^i \partial \lambda^j} \frac{P_{\text{det}}(\vec{\theta}_0)}{P_{\text{det}}(\vec{\lambda})} p(\vec{\theta}_0|\vec{\lambda})\mathrm{d}\vec{\theta}_0,$$

$$(\Gamma_{\text{II}})_{ij} = \frac{1}{2}\int \frac{\partial^2 \ln \det(\Gamma+H)}{\partial \lambda^i \partial \lambda^j} \frac{P_{\text{det}}(\vec{\theta}_0)}{P_{\text{det}}(\vec{\lambda})} p(\vec{\theta}_0|\vec{\lambda})\mathrm{d}\vec{\theta}_0,$$

$$(\Gamma_{\text{III}})_{ij} = -\frac{1}{2}\int \frac{\partial^2}{\partial \lambda^i \partial \lambda^j} \left[(\Gamma+H)_{kl}^{-1} \right] D_{kl} \frac{p(\vec{\theta}_0|\vec{\lambda})}{P_{\text{det}}(\vec{\lambda})} \mathrm{d}\vec{\theta}_0,$$

$$(\Gamma_{\text{IV}})_{ij} = -\int \frac{\partial^2}{\partial \lambda^i \partial \lambda^j} \left[P_k(\Gamma+H)_{kl}^{-1} \right] D_l \frac{p(\vec{\theta}_0|\vec{\lambda})}{P_{\text{det}}(\vec{\lambda})} \mathrm{d}\vec{\theta}_0,$$

$$(\Gamma_{\text{V}})_{ij} = -\frac{1}{2}\int \frac{\partial^2}{\partial \lambda^i \partial \lambda^j} \left[P_k(\Gamma+H)_{kl}^{-1} P_l \right] \frac{P_{\text{det}}(\vec{\theta}_0)}{P_{\text{det}}(\vec{\lambda})} p(\vec{\theta}_0|\vec{\lambda})\mathrm{d}\vec{\theta}_0.$$

This is an approximate expression for the population Fisher matrix which can be used to estimate the precision with which observations will be able to determine the population parameters. In deriving the above expressions, we have made use of the standard form of the likelihood for the gravitational wave detection problem, which permits some simplifications. In Appendix C, we describe how the result is changed for a generic likelihood, $p(\mathbf{d}|\theta_0)$.

We note that when measurement errors for the source parameters are small, only the first of these terms is required. This limit corresponds to $\Gamma \to \infty$, so that $\Gamma + H \approx \Gamma$ and $(\Gamma + H)^{-1} \to 0$. In this limit, it is clear that $(\Gamma_{\text{III}})_{ij}$, $(\Gamma_{\text{IV}})_{ij}$, and $(\Gamma_{\text{V}})_{ij}$ immediately vanish. The matrix $(\Gamma_{\text{II}})_{ij}$ also vanishes because Γ does not depend on the population parameters λ. Therefore, we expect $(\Gamma_{\text{I}})_{ij}$ to dominate and provide a good approximation to the population Fisher matrix. This will be true whenever individual measurement errors are small relative to the scale over which the population parameters change the source parameter distribution. The approximation holds in the three examples we describe below, but this will not always be the case.

2.1 Validity of approximations

To derive the population Fisher matrix, we have made two approximations. First, we have used expression (14) to define the population Fisher matrix. Corrections to this expression are derived in Appendix B and are shown to scale with inverse powers of the number of observed events, n. This assumption will therefore always be valid once we have made sufficiently many observations, and this is the limit in which we want to use this result. The second approximation was to use the linear signal approximation to represent the posteriors for individual events in equation (16) and equation (17). This approximation will not necessarily be valid in all circumstances, or across the whole of parameter space. In Vallisneri 2008, a criterion is provided for the validity of the individual event Fisher matrix at

$\vec{\theta}_0$

$$\frac{1}{2}\left(\Delta\mathbf{h}_{\text{LSA}}(\vec{\Delta\theta}) | \Delta\mathbf{h}_{\text{LSA}}(\vec{\Delta\theta}) \right) \ll 1$$

$$\forall \vec{\Delta\theta}_{1\sigma} : \Gamma_{ij}\Delta\theta_{1\sigma}^i \Delta\theta_{1\sigma}^j = 1$$

where $\quad \Delta\mathbf{h}_{\text{LSA}}(\vec{\Delta\theta}) = \Delta\theta_{1\sigma}^i \frac{\partial \mathbf{h}}{\partial \theta^i} - \left[\mathbf{h}(\vec{\theta}_0 + \vec{\Delta\theta}_{1\sigma}) - h(\vec{\theta}_0) \right]. \tag{22}$

If this criterion holds throughout the parameter space of observed events, then the approximations used to derive the population Fisher matrix will definitely be valid. However, this condition is more stringent than is strictly required since the population Fisher matrix is determined by derivatives with respect to the population parameters of the average of the individual event Fisher matrix over the parameter space.

An alternative criterion can be obtained by identifying the next higher order terms in equation (16) and equation (17). These contribute a multiplicative correction to the integral (18) of the form

$$\exp\left[\left(\Delta_{ijk} + T_{ijk} \right) \Delta\theta^i \Delta\theta^j \Delta\theta^k + N_{ij}\Delta\theta^i \Delta\theta^j \right],$$

where

$$T_{ijk} = \frac{1}{6}\frac{\partial^3 \ln p(\vec{\theta}|\vec{\lambda})}{\partial \theta^i \partial \theta^j \partial \theta^k} \qquad \Delta_{ijk} = \frac{1}{2}\left(\frac{\partial^2 \mathbf{h}}{\partial \theta^i \partial \theta^j} \left| \frac{\partial \mathbf{h}}{\partial \theta^k} \right. \right)$$

$$N_{ij} = \left(\frac{\partial^2 \mathbf{h}}{\partial \theta^i \partial \theta^j} \left| \mathbf{n} \right. \right). \tag{23}$$

Approximating the exponential as $\exp(x) \approx 1 + x$, these terms contribute additively to the integral over $\Delta\theta$ an amount δI. The contribution to $\ln p(\mathbf{d}|\vec{\lambda})$ is then an additive $\ln(1 + \delta I/I_0) \approx \delta I/I_0$, where I_0 is the value of the leading order integral. We deduce that the next order correction to the population Fisher matrix is

$$\Gamma_{\text{VI}} = \int \int \left(\frac{\partial^2(\delta I/I_0)}{\partial \lambda^i \partial \lambda^j} \right)_{\vec{\lambda}_t} p(\vec{\theta}_0|\vec{\lambda}_t)p(\mathbf{n}|\vec{\theta}_0)\mathrm{d}\mathbf{n}\mathrm{d}\vec{\theta}_0,$$

$$\frac{\delta I}{I_0} = \left(\Delta_{ijk} + T_{ijk} \right) \left(3\Delta\theta_{\text{bf}}^i(\Gamma+H)_{jk}^{-1} + \Delta\theta_{\text{bf}}^i \Delta\theta_{\text{bf}}^j \Delta\theta_{\text{bf}}^k \right)$$

$$+ N_{ij}\left(\Delta\theta_{\text{bf}}^i \Delta\theta_{\text{bf}}^j + (\Gamma+H)_{ij}^{-1} \right). \tag{24}$$

This expression can be simplified further, but as we will not use it elsewhere in this paper, we will leave it in this form, but we will make a few observations

(i) This expression can be used to assess the validity of the approximations used to build the population Fisher matrix. If the predicted errors computed including this correction are similar to those computed without then we can trust the population Fisher matrix.

(ii) The correction depends on derivatives with respect to the population parameters. If higher order corrections are only significant for parameters that are weakly coupled to those described by the population model, then this correction is still likely to be small, and the predictions of the population Fisher matrix are likely to be trustworthy.

(iii) In the examples discussed later, the dominant contribution to the Fisher matrix comes from Γ_{I}, which is independent of the individual event uncertainties. Thus, even if Γ_{VI} is of comparable size to Γ_{V}, it might still be negligible relative to Γ_{I}. In that case, Γ_{I} can continue to be used to estimate the population parameter uncertainties.

(iv) In general, the Fisher Matrix approximation will be better for events of higher SNR, and so the population Fisher matrix will tend to be a better approximation if we use a higher threshold for including events in the analysis. By adjusting the detection threshold

to be high enough that the individual events are well characterized by the Fisher matrix, we will be able to obtain a reliable estimate from the population Fisher matrix for any analysis. This will provide a conservative estimate to the precision that could be achieved using all events.

In summary, when using the population Fisher matrix to scope out the potential of future GW observations, it is important to monitor the validity of the approximations by using equation (24). In some contexts, the individual event Fisher matrix will be valid throughout the parameter space of observable events and so the full population Fisher matrix can be used directly. For example, in the context of extreme-mass-ratio inspirals, on which the GW-like example in Section 4.1 is based, it is expected that an SNR of at least 20 will be required for the confident detection of an event in the data (Babak et al. 2017), and so the Fisher matrix is likely to be a good approximation for all observed events. In other contexts, the individual event Fisher matrix might be a poor approximation for some parameters, but if it is valid for the parameters for which the population model has been written down, and these are weakly correlated with the other parameters, then the population Fisher matrix is still likely to provide a good estimate of measurement precision. There will be situations in which the approximation will not be valid, but even there, the population Fisher matrix might be accurate if it is dominated by the measurement-error independent part, Γ_I. In any scenario, it can be used to obtain a conservative estimate of accuracy by raising the detection threshold sufficiently. It will thus always provide a valuable tool for quickly scoping out the potential of future observations without the need for expensive simulations.

3 ILLUSTRATION I: A GAUSSIAN–GAUSSIAN MODEL

We will now consider several examples, which will demonstrate that the population Fisher matrix works and show how to compute it in practice. The first application of equation (21) we will consider is to a 'Gaussian–Gaussian' model in which both observations and noise are normally distributed. We simplify the setting by assuming a waveform dependent on a single parameter θ. The distribution of the parameter is

$$p(\theta|\vec{\lambda}) = \mathcal{N}(\mu, \Sigma^2) = \frac{1}{\sqrt{2\pi\,\Sigma^2}} \exp\left[-\frac{(\theta-\mu)^2}{2\Sigma^2}\right], \qquad (25)$$

with population parameters (henceforth, hyperparameters) $\vec{\lambda} = \{\mu, \Sigma^2\}$. Noise is also a Gaussian with zero mean and variance σ. Since the data stream is a sum of Gaussians, it is modelled by $\mathcal{N}(\mu, \sigma^2 + \Sigma^2)$, with mean and variance given by the sums of individual means and variances,

$$p(\boldsymbol{d}|\vec{\lambda}) = \frac{1}{\sqrt{2\pi(\sigma^2 + \Sigma^2)}} \exp\left[-\frac{(\boldsymbol{d}-\mu)^2}{2(\sigma^2+\Sigma^2)}\right]. \qquad (26)$$

Given the implicit simple choice for the signals, the Fisher matrix of source parameters reduces to

$$\Gamma_\theta = \left(\frac{\partial \boldsymbol{h}}{\partial\theta}\bigg|\frac{\partial \boldsymbol{h}}{\partial\theta}\right) = \frac{1}{\sigma^2}\left|\frac{\partial \boldsymbol{h}}{\partial\theta}\right|^2 = \frac{1}{\sigma^2}, \qquad (27)$$

while from equation (25), we have that P and H in (17) are

$$P = -\frac{(\theta-\mu)}{\Sigma^2} \quad \text{and} \quad H = \frac{1}{\Sigma^2}\,. \qquad (28)$$

The example reported in this section does not have an immediate analogy in GW astrophysics, but it can be thought of as a more

general application of the population Fisher matrix. The advantage of choosing such a simple setting is that the matrix entries can be directly integrated as expectation values over data realizations. In the presence of selection effects, the integrals to be solved are

$$
\begin{aligned}
(\Gamma_\lambda)_{ij} &= \mathbb{E}\left(-\frac{\partial^2}{\partial\lambda^i\partial\lambda^j}\left[\ln\left(\frac{p(\boldsymbol{d}|\theta,\vec{\lambda})}{P_\mathrm{det}(\vec{\lambda})}\right)\right]\right) \\
&= -\int_{\boldsymbol{d}_\mathrm{th}}^\infty \frac{\partial^2}{\partial\lambda^i\partial\lambda^j}\left[\ln\left(\frac{p(\boldsymbol{d}|\vec{\lambda})}{P_\mathrm{det}(\vec{\lambda})}\right)\right]\frac{p(\boldsymbol{d}|\vec{\lambda})}{P_\mathrm{det}(\vec{\lambda})}d\boldsymbol{d}.
\end{aligned} \qquad (29)
$$

We only select realizations of the data $\boldsymbol{d} > \boldsymbol{d}_\mathrm{th}$ that are above a certain threshold. The predictions for the various components of the population Fisher matrix are[1]

$$
(\Gamma_\lambda)_{\mu\mu} = \frac{\partial^2\ln P_\mathrm{det}(\vec{\lambda})}{\partial\mu^2} + \frac{1}{(\sigma^2+\Sigma^2)},
$$

$$
(\Gamma_\lambda)_{\mu\Sigma^2} = \frac{\partial^2\ln P_\mathrm{det}(\vec{\lambda})}{\partial\mu\partial\Sigma^2} + \frac{1}{(\sigma^2+\Sigma^2)}\frac{p(\boldsymbol{d}_\mathrm{th}|\vec{\lambda})}{P_\mathrm{det}(\vec{\lambda})},
$$

$$
\begin{aligned}
(\Gamma_\lambda)_{\Sigma^2\Sigma^2} = {} & \frac{\partial^2\ln P_\mathrm{det}(\vec{\lambda})}{(\partial\Sigma^2)^2} + \frac{1}{2(\sigma^2+\Sigma^2)^2} \\
& + \frac{(\boldsymbol{d}_\mathrm{th}-\mu)}{(\sigma^2+\Sigma^2)^2}\frac{p(\boldsymbol{d}_\mathrm{th}|\vec{\lambda})}{P_\mathrm{det}(\vec{\lambda})},
\end{aligned} \qquad (30)
$$

where we have used equations (26) and (29), the fact that the integral is normalized through

$$P_\mathrm{det}(\vec{\lambda}) = \int_{\boldsymbol{d}_\mathrm{th}}^\infty p(\boldsymbol{d}|\vec{\lambda})d\boldsymbol{d} = \frac{1}{2}\mathrm{erfc}\left(\frac{(\boldsymbol{d}_\mathrm{th}-\mu)}{\sqrt{2(\sigma^2+\Sigma^2)}}\right), \qquad (31)$$

and the definition

$$p(\boldsymbol{d}_\mathrm{th}|\vec{\lambda}) \equiv \frac{1}{\sqrt{2\pi(\sigma^2+\Sigma^2)}}\exp\left[-\frac{(\boldsymbol{d}_\mathrm{th}-\mu)^2}{2(\sigma^2+\Sigma^2)}\right]. \qquad (32)$$

Finally, from equation (31), it follows that

$$
\begin{aligned}
\frac{\partial^2\ln p_\mathrm{det}(\lambda)}{(\partial\mu)^2} &= \left(\frac{\boldsymbol{d}_\mathrm{th}-\mu}{\sigma^2+\Sigma^2}\right)\frac{p(\boldsymbol{d}_\mathrm{th}|\lambda)}{p_\mathrm{det}(\lambda)} - \frac{p(\boldsymbol{d}_\mathrm{th}|\lambda)^2}{p_\mathrm{det}(\lambda)^2}, \\
\frac{\partial^2\ln p_\mathrm{det}(\lambda)}{\partial\mu\partial\Sigma^2} &= \frac{p(\boldsymbol{d}_\mathrm{th}|\lambda)}{p_\mathrm{det}(\lambda)}\left[\frac{(\boldsymbol{d}_\mathrm{th}-\mu)^2}{2(\sigma^2+\Sigma^2)^2} - \frac{1}{2(\sigma^2+\Sigma^2)}\right] \\
&\quad - \frac{p(\boldsymbol{d}_\mathrm{th}|\lambda)^2}{2p_\mathrm{det}(\lambda)^2}\left(\frac{\boldsymbol{d}_\mathrm{th}-\mu}{\sigma^2+\Sigma^2}\right), \\
\frac{\partial^2\ln p_\mathrm{det}(\lambda)}{(\partial\Sigma^2)^2} &= \frac{p(\boldsymbol{d}_\mathrm{th}|\lambda)}{p_\mathrm{det}(\lambda)}\left[\frac{(\boldsymbol{d}_\mathrm{th}-\mu)^3}{4(\sigma^2+\Sigma^2)^3} - \frac{3(\boldsymbol{d}_\mathrm{th}-\mu)}{4(\sigma^2+\Sigma^2)^2}\right] \\
&\quad - \frac{p(\boldsymbol{d}_\mathrm{th}|\lambda)^2}{4p_\mathrm{det}(\lambda)^2}\left(\frac{\boldsymbol{d}_\mathrm{th}-\mu}{\sigma^2+\Sigma^2}\right)^2.
\end{aligned} \qquad (33)
$$

3.1 Solution from the population Fisher matrix

While in this simple setting, the direct evaluation of the Fisher Matrix as expectation value is much simpler, we wish to now evaluate it using equation (21) as an important sanity check of that general formula. For $(\Gamma_\lambda)_{\mu\mu}$, we notice that the $(\Gamma_\mathrm{II})_{\mu\mu}$, $(\Gamma_\mathrm{III})_{\mu\mu}$, and $(\Gamma_\mathrm{IV})_{\mu\mu}$ vanish. The second and third terms vanish because $\Gamma + H$ does not depend on μ, while the fourth term vanishes because P is only linear in μ and

---

[1]For the last integral, it is useful to notice that

$$\int_{\boldsymbol{d}_\mathrm{th}}^\infty (\boldsymbol{d}-\mu)^2 p(\boldsymbol{d}|\vec{\lambda})d\boldsymbol{d} = (\sigma^2+\Sigma^2)\left[1 + (\boldsymbol{d}_\mathrm{th}-\mu)\frac{p(\boldsymbol{d}_\mathrm{th}|\vec{\lambda})}{P_\mathrm{det}(\vec{\lambda})}\right],$$

which follows from $(\boldsymbol{d}-\mu)p(\boldsymbol{d}|\vec{\lambda}) = -(\sigma^2+\Sigma^2)\partial p(\boldsymbol{d}|\vec{\lambda})/\partial\boldsymbol{d}$, integrating by parts, and using equation(31).

two derivatives with respect to it are needed for $(\Gamma_\lambda)_{\mu\mu}$. The only terms contributing are therefore $(\Gamma_I)_{\mu\mu}$ and $(\Gamma_V)_{\mu\mu}$, which lead to

$$
\begin{aligned}
(\Gamma_\lambda)_{\mu\mu} &= -\int \frac{\partial^2 \ln(p(\theta|\vec{\lambda})/P_{\det}(\vec{\lambda}))}{\partial\mu^2} \frac{P_{\det}(\theta)}{P_{\det}(\vec{\lambda})} p(\theta|\vec{\lambda})d\theta \\
&\quad -\frac{1}{2}\int \frac{\partial^2}{\partial\mu^2}\left[P^2(\Gamma+H)^{-1}\right]\frac{P_{\det}(\theta)}{P_{\det}(\vec{\lambda})}p(\theta|\vec{\lambda})d\theta \\
&= \frac{\partial^2 \ln P_{\det}(\vec{\lambda})}{\partial\mu^2} + \frac{1}{(\sigma^2+\Sigma^2)},
\end{aligned} \tag{34}
$$

where we have used P, Γ, and H given in equations (25), (27), and (28), as well as the normalization

$$
\int \frac{P_{\det}(\theta)}{P_{\det}(\vec{\lambda})}p(\theta|\vec{\lambda})d\theta = 1. \tag{35}
$$

The result matches the prediction of equation (30) as expected.

In the case of $(\Gamma_\lambda)_{\mu\Sigma^2}$, the second and third terms, $(\Gamma_{II})_{\mu\Sigma^2}$ and $(\Gamma_{III})_{\mu\Sigma^2}$, vanish for the same reason as above, but now the fourth term $(\Gamma_{IV})_{\mu\Sigma^2}$ does contribute. Overall, we have that

$$
\begin{aligned}
(\Gamma_\lambda)_{\mu\Sigma^2} &= -\int \frac{\partial^2 \ln(p(\theta|\vec{\lambda})/P_{\det}(\vec{\lambda}))}{\partial\mu\partial\Sigma^2} \frac{P_{\det}(\theta)}{P_{\det}(\vec{\lambda})} p(\theta|\vec{\lambda})d\theta \\
&\quad -\int \frac{\partial^2}{\partial\mu\partial\Sigma^2}\left[P(\Gamma+H)^{-1}\right]D_l \frac{P_{\det}(\theta)}{P_{\det}(\vec{\lambda})}p(\theta|\vec{\lambda})d\theta \\
&\quad -\frac{1}{2}\int \frac{\partial^2}{\partial\mu\partial\Sigma^2}\left[P^2(\Gamma+H)^{-1}\right]\frac{P_{\det}(\theta)}{P_{\det}(\vec{\lambda})}p(\theta|\vec{\lambda})d\theta \\
&= \frac{\partial^2 \ln P_{\det}(\vec{\lambda})}{\partial\mu\partial\Sigma^2} + \int \frac{(\theta-\mu)}{\Sigma^4}\frac{P_{\det}(\theta)}{P_{\det}(\vec{\lambda})}p(\theta|\vec{\lambda})d\theta \\
&\quad + \int\left[\frac{\sigma^2}{(\sigma^2+\Sigma^2)^2}\right]\frac{\partial P_{\det}(\theta)}{\partial\theta}\frac{p(\theta|\vec{\lambda})}{P_{\det}(\vec{\lambda})}d\theta \\
&\quad -\frac{\sigma^2(\sigma^2+2\Sigma^2)}{\Sigma^4(\sigma^2+\Sigma^2)^2}\frac{1}{P_{\det}(\vec{\lambda})}\int(\theta-\mu)P_{\det}(\theta)p(\theta|\vec{\lambda})d\theta,
\end{aligned}
$$

where we have used the definition of D_i in equation (20) and (35). We also notice that, from the definition

$$
\begin{aligned}
P_{\det}(\theta) &= \int_{d_{th}}^{\infty} p(d|\theta)dd \\
&= \int_{d_{th}}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(d-\theta)^2}{2\sigma^2}\right]dd = \frac{1}{2}\text{erfc}\left(\frac{d_{th}-\theta}{\sqrt{2\sigma^2}}\right),
\end{aligned} \tag{36}
$$

it follows that

$$
\begin{aligned}
\frac{\partial P_{\det}(\theta)}{\partial\theta} &= \int_{d_{th}}^{\infty}\frac{(d-\theta)}{\sigma^2}p(d|\theta)dd \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(d_{th}-\theta)^2}{2\sigma^2}\right] \equiv p(d_{th}|\theta),
\end{aligned} \tag{37}
$$

which can be rearranged to give

$$
\begin{aligned}
\int(\theta-\mu)\frac{\partial P_{\det}(\theta)}{\partial\theta}p(\theta|\vec{\lambda})d\theta \\
= \Sigma^2\frac{\partial}{\partial\mu}\int p(d_{th}|\theta)p(\theta|\vec{\lambda})d\theta \\
= \Sigma^2\frac{\partial p(d_{th}|\vec{\lambda})}{\partial\mu} = \frac{\Sigma^2}{(\sigma^2+\Sigma^2)}(d_{th}-\mu)p(d_{th}|\vec{\lambda}).
\end{aligned} \tag{38}
$$

From equation (31), it also follows that $\partial P_{\det}(\vec{\lambda})/\partial\mu = p(d_{th}|\vec{\lambda})$. Using this constraint, and the fact that

$$
\begin{aligned}
\frac{\partial p(d_{th}|\vec{\lambda})}{\partial\mu} &= \int \frac{\partial}{\partial\mu}\left[P_{\det}(\theta)p(\theta|\vec{\lambda})\right]d\theta \\
&= \frac{1}{\Sigma^2}\int(\theta-\mu)P_{\det}(\theta)p(\theta|\vec{\lambda})d\theta,
\end{aligned} \tag{39}
$$

we immediately get

$$
\int(\theta-\mu)P_{\det}(\theta)p(\theta|\vec{\lambda})d\theta = \Sigma^2 p(d_{th}|\vec{\lambda}), \tag{40}
$$

and, together with (35) and (38), that

$$
(\Gamma_\lambda)_{\mu\Sigma^2} = \frac{\partial^2 \ln P_{\det}(\vec{\lambda})}{\partial\mu\partial\Sigma^2} + \frac{1}{(\sigma^2+\Sigma^2)}\frac{p(d_{th}|\vec{\lambda})}{P_{\det}(\vec{\lambda})}. \tag{41}
$$

This agrees with equation (30) as expected.

Finally, we consider the case of $(\Gamma_\lambda)_{\Sigma^2\Sigma^2}$, in which no term in (21) vanishes. The first term can be rearranged to give[2]

$$
\begin{aligned}
(\Gamma_I)_{\Sigma^2\Sigma^2} &= \frac{\partial^2 \ln P_{\det}(\vec{\lambda})}{(\partial\Sigma^2)^2} - \frac{1}{2\Sigma^4} \\
&\quad + \frac{1}{\Sigma^6}\int(\theta-\mu)^2\frac{P_{\det}(\theta)}{P_{\det}(\vec{\lambda})}p(\theta|\vec{\lambda})d\theta \\
&= \frac{\partial^2 \ln P_{\det}(\vec{\lambda})}{(\partial\Sigma^2)^2} + \frac{1}{2\Sigma^4} + \frac{(d_{th}-\mu)}{\Sigma^2(\sigma^2+\Sigma^2)}\frac{p(d_{th}|\vec{\lambda})}{P_{\det}(\vec{\lambda})}.
\end{aligned} \tag{43}
$$

Using equation (35), the second term $(\Gamma_{II})_{\Sigma^2\Sigma^2}$ is easily found to be

$$
(\Gamma_{II})_{\Sigma^2\Sigma^2} = \frac{\sigma^2}{2\Sigma^4}\frac{(\sigma^2+2\Sigma^2)}{(\sigma^2+\Sigma^2)^2}. \tag{44}
$$

The third term $(\Gamma_{III})_{\Sigma^2\Sigma^2}$ contains D_{ij} from equation(20), which can be rearranged to give

$$
\begin{aligned}
D_{\theta\theta} &= \frac{1}{\sqrt{2\pi}\sigma^3}\int_{d_{th}}^{\infty}(d-\theta)^2\exp\left[-\frac{(d-\theta)^2}{2\sigma^2}\right]dd \\
&= \frac{(d_{th}-\theta)}{\sqrt{2\pi}\sigma^3}\exp\left[-\frac{(d_{th}-\theta)^2}{2\sigma^2}\right] + \frac{1}{2\sigma^2}\text{erfc}\left(\frac{d_{th}-\theta}{\sqrt{2\sigma^2}}\right).
\end{aligned}
$$

The complementary error function can be replaced by $P_{\det}(\theta)$ using (36). With this in mind, we have

$$
(\Gamma_{III})_{\Sigma^2\Sigma^2} = \frac{\sigma^4}{(\sigma^2+\Sigma^2)^4}(d_{th}-\mu)\frac{p(d_{th}|\vec{\lambda})}{P_{\det}(\vec{\lambda})} + \frac{\sigma^2}{(\sigma^2+\Sigma^2)^3}. \tag{45}
$$

The results follow from rearranging equations (37) and (32) into

$$
\begin{aligned}
\int(d_{th}-\theta)\exp&\left[-\frac{(d_{th}-\mu)^2}{2\sigma^2}\right]\frac{p(\theta|\vec{\lambda})}{P_{\det}(\vec{\lambda})}d\theta \\
&= -\sigma^3\sqrt{2\pi}\int\frac{\partial p(d_{th}|\theta)}{\partial d_{th}}\frac{p(\theta|\vec{\lambda})}{P_{\det}(\vec{\lambda})}d\theta \\
&= -\frac{\sigma^3\sqrt{2\pi}}{P_{\det}(\vec{\lambda})}\int\frac{\partial}{\partial d_{th}}\left[p(d_{th}|\theta)p(\theta|\vec{\lambda})\right]d\theta \\
&= \frac{\sigma^3\sqrt{2\pi}}{(\sigma^2+\Sigma^2)}(d_{th}-\mu)\frac{p(d_{th}|\vec{\lambda})}{P_{\det}(\vec{\lambda})}.
\end{aligned}
$$

---

[2]The integral over θ can be evaluated by parts to obtain

$$
\int(\theta-\mu)^2 \frac{P_{\det}(\theta)}{P_{\det}(\vec{\lambda})}p(\theta|\vec{\lambda})d\theta = \Sigma^2 + \frac{(d_{th}-\mu)\Sigma^4}{(\sigma^2+\Sigma^2)}\frac{p(d_{th}|\vec{\lambda})}{P_{\det}(\vec{\lambda})}, \tag{42}
$$

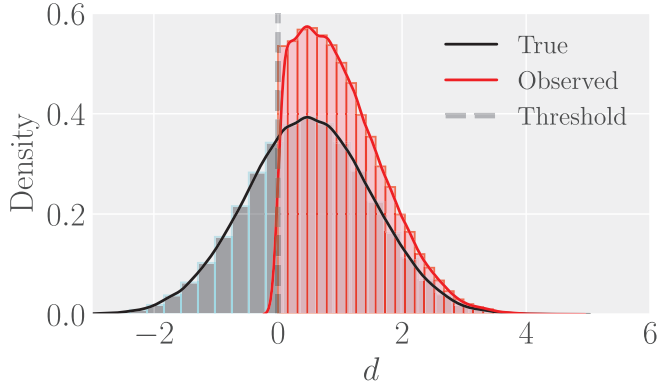using (31) and (40) with vanishing boundary terms.

Figure 1. Normally-distributed data with (red) and without (black) selection effects.

The fourth term $(\Gamma_{\text{IV}})_{\Sigma^2\Sigma^2}$ is found using (40), and reads

$$(\Gamma_{\text{IV}})_{\Sigma^2\Sigma^2} - \int \frac{\partial^2}{(\partial \Sigma^2)^2} \left[P(\Gamma + H)^{-1} \right] D_\theta \frac{P_{\text{det}}(\theta)}{P_{\text{det}}(\vec{\lambda})} p(\theta|\vec{\lambda}) d\theta$$
$$= \frac{2\sigma^2 \Sigma^2}{(\sigma^2 + \Sigma^2)^4} (\boldsymbol{d}_{\text{th}} - \mu) \frac{p(\boldsymbol{d}_{\text{th}}|\vec{\lambda})}{P_{\text{det}}(\vec{\lambda})}. \quad (46)$$

Finally, the final term $(\Gamma_{\text{V}})_{\Sigma^2\Sigma^2}$ is found through (42) to be

$$(\Gamma_{\text{V}})_{\Sigma^2\Sigma^2} = -\frac{1}{2} \int \frac{\partial^2}{(\partial \Sigma^2)^2} \left[P^2(\Gamma + H)^{-1} \right] \frac{P_{\text{det}}(\theta)}{P_{\text{det}}(\vec{\lambda})} p(\theta|\vec{\lambda}) d\theta$$
$$= -\frac{\sigma^2}{\Sigma^4} \frac{(\sigma^4 + 3\sigma^2\Sigma^2 + 3\Sigma^4)}{(\sigma^2 + \Sigma^2)^3}$$
$$\left[1 + \frac{\Sigma^2}{(\sigma^2 + \Sigma^2)} (\boldsymbol{d}_{\text{th}} - \mu) \frac{p(\boldsymbol{d}_{\text{th}}|\vec{\lambda})}{P_{\text{det}}(\vec{\lambda})} \right]. \quad (47)$$

Adding up (43) to (47), we find

$$\Gamma_{\Sigma^2\Sigma^2} = \frac{\partial^2 \ln P_{\text{det}}(\vec{\lambda})}{(\partial \Sigma^2)^2} + \frac{1}{2(\sigma^2 + \Sigma^2)^2} + \frac{(\boldsymbol{d}_{\text{th}} - \mu)}{(\sigma^2 + \Sigma^2)^2} \frac{p(\boldsymbol{d}_{\text{th}}|\vec{\lambda})}{P_{\text{det}}(\vec{\lambda})}, (48)$$

which matches (30) as expected. This concludes the analytic check. Equation (21) can be used to reproduce the predictions (30) obtained from the more general definition of the Fisher matrix as an expectation value over data realizations.

3.2 MCMC analysis

The Fisher predictions for the Gaussian–Gaussian model can be compared with MCMC simulations as a further check of the formalism. While this example is arguably textbook material, see for instance section (6) in Vitale et al. 2020, we report a few details below for completeness. The results of the present section can be fully reproduced with the codes accompanying this paper. We simulate synthetic data including $N_{\text{tot}} = 10^5$ observations from the observation model (26), choosing true mean $\mu_{\text{tr}} = 0.5$, true variance $\Sigma_{\text{tr}}^2 = 1.0$. and noise variance $\sigma = 0.1$. The latter two indicate that each event is taken with a high SNR. We then apply an arbitrary cutoff, imposing that only positive data are observed. That is, $\boldsymbol{d}_{\text{th}} = 0$, resulting in around $N_{\text{det}} \sim 70\,000$ detected events. Fig. 1 shows the total (black) and detected (red) populations under our specified assumptions. We perform the MCMC analyses in both cases in which we do and do not have selection effects using EMCEE (Foreman-Mackey et al. 2013). As log likelihood, we take the sum of individual log likelihoods,

$$\log p(d|\lambda) = -N \log p_{\text{det}}(\lambda) + \sum_{i}^{N} \log p(d_i|\lambda), \quad (49)$$
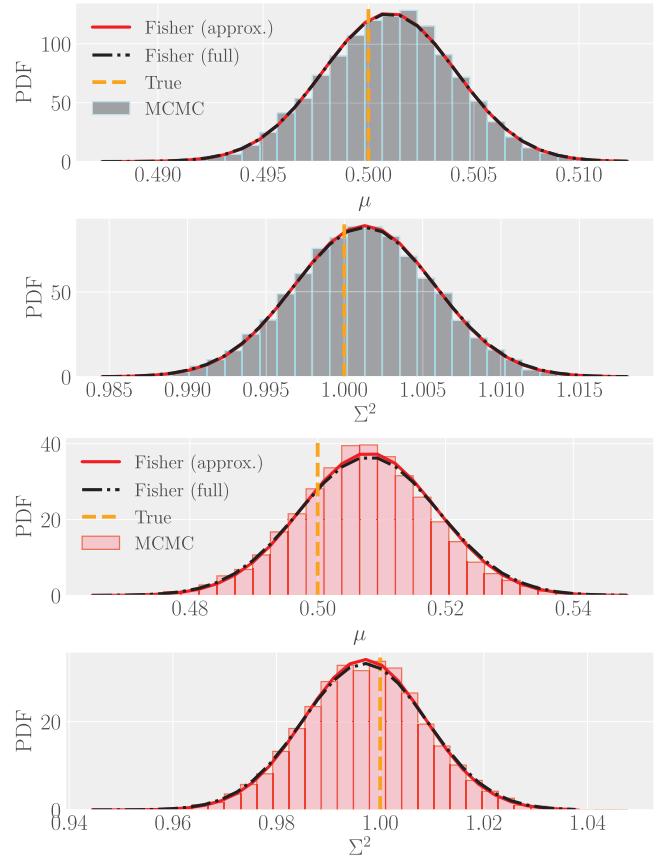

Figure 2. Fisher and MCMC predictions for the toy Gaussian case with (red) and without (black) selection effects.

where $N = N_{\text{tot}}$ in the case in which we do not include selection effects, and $N = N_{\text{det}}$ in the case in which we do. We choose flat hyperpriors over a very broad range that includes the true values. The selection function is defined and integrated as in equation (31), and is non-trivial only in the latter case. The MCMC posteriors for μ and Σ^2, in both the cases considered, can be found in Fig. 2.

We can then compare the predictions from the population Fisher matrix with what we obtain numerically. To this end, we invert the matrix

$$(\Gamma_\lambda)_{ij} = \begin{pmatrix} \Gamma_{\mu\mu} & \Gamma_{\mu\Sigma^2} \\ \Gamma_{\mu\Sigma^2} & \Gamma_{\Sigma^2\Sigma^2} \end{pmatrix} \quad (50)$$

with entries given in equation (30) and below. The errors are normalized by the number of events,

$$\Delta\mu = \sqrt{(\Gamma_\lambda^{-1})_{\mu\mu}/N}, \quad \Delta\Sigma^2 = \sqrt{(\Gamma_\lambda^{-1})_{\Sigma^2\Sigma^2}/N}.$$

In Fig. 2, the Fisher predictions are shown (in black) to reproduce the widths from the MCMC runs.

The same widths can be well approximated by the inverse of the matrix

$$(\Gamma_\lambda)_{ij} = \begin{pmatrix} (\Gamma_{\text{I}})_{\mu\mu} & (\Gamma_{\text{I}})_{\mu\Sigma^2} \\ (\Gamma_{\text{I}})_{\mu\Sigma^2} & (\Gamma_{\text{I}})_{\Sigma^2\Sigma^2}, \end{pmatrix} \quad (51)$$

in which only the first terms in equation (21) are retained. The predictions are shown in red in Fig. 2, and they overlap well with the full Fisher matrix predictions. As discussed earlier, this corresponds to taking the limit of equation (50) in which $\sigma = 0$, i.e. the parameters of the individual events are measured perfectly. The reason that this is

a good approximation here is because we have chosen $0.1 = \sigma \ll \Sigma = 1$ for this example. We would expect the other terms contributing to equation (21) to become increasingly important as the measurement errors become larger. If we consider for simplicity the case without selection effects, we see that the ratio of the uncertainties in the population parameters computed using only Γ_I to that computed using the full Fisher matrix, equation (50), are

$$\frac{\Delta \mu_I}{\Delta \mu_{full}} = \frac{1}{1 + \sigma^2/\Sigma^2} \qquad \frac{\Delta \Sigma_I^2}{\Delta \Sigma_{full}^2} = \frac{1}{1 + (\sigma^2/\Sigma^2)^2}. \qquad (52)$$

For $\sigma^2 \ll \Sigma^2$, these ratios are approximately 1, as expected, but as $\sigma^2/\Sigma^2 \to \infty$, both ratios tend to 0, implying that Γ_I would significantly overestimate the precision with which the population parameters can be determined. So, it is not always possible to use Γ_I to estimate the population parameter uncertainties. However, in many applications, individual events are constrained to a small region of the much larger parameter space of the population, and so equation (51) will often be a good approximation to the full Fisher matrix. This includes the GW-like illustrations we will consider in the next section.

4 ILLUSTRATION II: AN EXAMPLE FROM GRAVITATIONAL-WAVE ASTROPHYSICS

The spaceborne LISA mission is expected to detect extreme-mass-ratio inspirals (EMRIs), namely binary systems in which one compact object, typically a stellar remnant, has a mass that is much smaller than the companion, typically a supermassive black hole (SMBH) in the centre of a galaxy (Amaro-Seoane et al. 2007; Barack 2009; Babak et al. 2017). Inference of the parameters that characterize EMRI systems is expected to provide accurate constraints on the theory of gravity (Gair et al. 2013), as well as an insight into the astrophysical population of and stellar environments surrounding SMBHs (Barausse, Cardoso & Pani 2014). LISA might also detect a foreground generated by individually unresolved EMRIs, which would porvide information about the properties of the population of these systems (Gair et al. 2010, 2011; Sesana et al. 2011; Bonetti & Sesana 2020). The use of LISA observations of EMRIs to provide measurements of the BH mass function in the range probed by LISA has previously been investigated in Gair et al. 2010, henceforth 'GTV'. GTV assumed that the mass function was described by a power law $p(\theta|\lambda) \equiv p(M|\lambda) \propto M^{\alpha - 1}$, with a true value that is close to flat in the log of the masses, i.e. $\alpha \approx 0$. GTV explored the ability of LISA to constrain the parameters of this mass function, using MCMC techniques to carry out hierarchical analyses on an extensive set of populations of simulated events. They found that with 10(1000) events, the spectral index α could be constrained at a level of precision $\Delta \alpha = 0.3(0.03)$. Here, we will use the population Fisher Matrix formalism described above to predict the precision with which a set of EMRI observations might be able to constrain a power-law mass function. We do not expect to get exactly the same answer here, as the two analyses make a few different simplifying assumptions. In GTV, the raw data was taken to be counts of events in a binned analysis, provided by point estimates of the parameters. Selection effects were included in the rate of events in each bin, by accounting for the length of time, a source with the given parameters would be observable. This ignores the fact that the time remaining to plunge is constrained by the gravitational wave data. In this analysis, we again approximate the observation process, assuming that the data can be reduced to a measurement of a single parameter, but we handle selection effects more carefully. GTV's results was computed ignoring measurement uncertainties in the model used in the analysis, although they did demonstrate consistency between

results obtained on simulated data with and without measurement uncertainties. Here, we will include measurement uncertainties, but we will approximate these as Gaussian. We will see that despite these differences in assumptions, the population Fisher matrix is able to predict qualitatively the results observed in GTV without the need for costly computational sampling of many posterior distributions. For a more direct comparison with numerical results, we also perform our own MCMC analysis, under identical assumptions to those used to compute the population Fisher matrix, and find very good agreement. All the results in this section can be reproduced with the codes made available with this publication.

4.1 Simple scenario: mass measurement only

In the first scenario we consider, we will assume that events in the population are characterized by a single parameter, the mass, which we measure with our detector with a Gaussian uncertainty that has a fixed variance, independent of the parameters of the source. We note that this choice of a constant variance is made for convenience, but is not required by the formalism. If errors vary form event to event, these are characterized by a $\vec{\theta}_0$ dependence of Γ, which just changes the integrands of the various components of the population Fisher matrix. Due to the integration over $\vec{\theta}_0$, the population Fisher matrix effectively depends on the average measurement precision over the population. In the final example, described in Section 4.2, we will consider a case in which the errors vary from event to event. We draw $N = 100$ masses from a power-law distribution,

$$p(M|\alpha) = \frac{\alpha}{M_{max}^\alpha - M_{min}^\alpha} M^{\alpha - 1}, \qquad (53)$$

with maximal and minimal observable masses $M_{min} = 10^4 M_\odot$ and $M_{max} = 10^7 M_\odot$ and a true value for the spectral index that is exactly flat in the log of the masses, $\alpha = 0$. The observed data is assumed to be a point estimate of the log of the mass, which is equal to the true value plus a normally-distributed uncertainty that has a variance $\sigma = 0.1$. An arbitrary hard cutoff $\mathbf{d}_{th}$ corresponding to masses $M \sim 5 \times 10^5 M_\odot$ is imposed, and only events with observed values above this threshold are included in the analysis. This introduces a large selection effect, and leads to only $N_{det} = 39$ of the original 100 sources being observed. The true (underlying) and observed populations of events are represented in the top panel of Fig. 4.

The Fisher-matrix prediction can again be obtained from equation (21), this time with $\theta = M$ and $\lambda = \alpha$. For sufficiently simple models, the integrals are analytically tractable. With the power-law distribution considered here, and in the presence of selection effects, this is already not possible. Because of this, we obtain the Fisher prediction in a semi-analytical fashion by solving the integrals with Monte-Carlo methods, i.e. by generating a sufficiently large set of N_s samples, $\{M_i\}$, from a distribution $p(M|\alpha)$, we can approximate the integral of an arbitrary function $X(M)$ via

$$\int X(M) p(M|\alpha) dM \approx \frac{1}{N_s} \sum_i X(M_i). \qquad (54)$$

In this case, we can draw samples from the distribution (53) directly using the method of inversion. The various terms entering the arguments of equation (21) – D_i, D_{ij}, P_i, and H_{ij} – can be computed analytically. A MATHEMATICA notebook that solves for the arguments of the Fisher-matrix integrals can be found in the accompanying codes. Following this procedure, we find that the Fisher matrix predicts an error $\Delta \alpha = \sqrt{(N_{det} \Gamma_\alpha)^{-1}} \approx 0.19$. When rescaled to 10
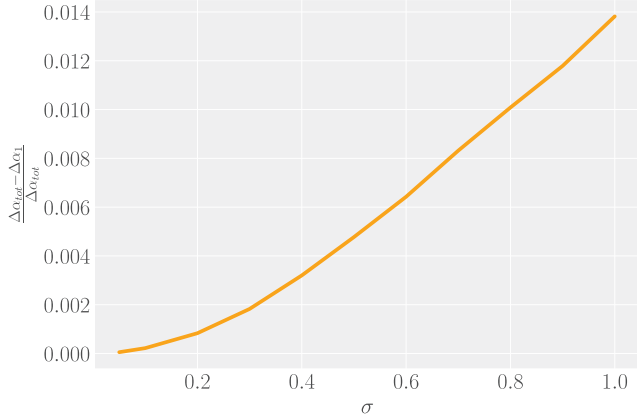
Figure 3. Fractional error in the estimate of the precision on the population slope, α, obtained from using only the first term in the Fisher matrix, Γ_{I}, instead of the full Fisher matrix, as a function of the assumed uncertainty on the measurement of individual event parameters, σ.

observations by multiplying by $\sqrt{39/10}$, the inferred error is $\Delta\alpha \approx 0.37$, which is in good agreement with what was obtained in GTV, despite the differences in the assumptions used in each case.

As in the Gaussian–Gaussian example, we find that the dominant contribution comes from the first term of the Fisher matrix, Γ_{I}. In this example, this term can be directly computed

$$\Gamma_\alpha \approx N_{\mathrm{tot}} \left(\frac{1}{\alpha^2} - \frac{M_{\max}^\alpha M_{\min}^\alpha (\ln M_{\max} - \ln M_{\min})^2}{(M_{\max}^\alpha - M_{\min}^\alpha)^2} \right). \qquad (55)$$

As argued above, the dominance of this term is driven by the assumed precision of measurement on the individual events. If the noise in the individual measurements, σ, is increased, the other terms make a larger contribution, although always a sub-dominant contribution for the range of values we have tried. This is illustrated in Fig. 3, which shows how the error in the prediction for the precision on the slope from using only the first term of the Fisher matrix varies with σ. Even for $\sigma = 1$, the fractional error from this approximation is only 0.014. This fact suggests that, when individual events are expected to be characterized with a precision better than the typical lengthscale over which the population prior varies, keeping only the first term in the sum will provide a good estimate for the expected precision of population inference, regardless of the population model chosen. This observation and the fact that the first term is typically relatively easy to evaluate, could help to reduce the complexity of using our formalism. One obvious application would be to forecast studies for future detectors, where this formalism for the precision of population inference can nicely complement estimates for the precision of individual event parameter inference, estimated with state-of-the-art Fisher codes (Borhanian 2021; Harms et al. 2022).

We validate the Fisher predictions with an MCMC analysis for the same data set. The MCMC set-up is similar to the one used in the first illustrative example. The likelihood is modified, and the selection function is no longer known analytically, but is approximated by a Monte Carlo integral, $p_{\mathrm{det}}(\alpha) = (1/N_s) \sum \mathrm{erfc}[(\boldsymbol{d}_{\mathrm{th}} - M_i)/\sqrt{2\sigma^2}]/2$, with $\{M_i\}$ drawn from the power-law distribution $p(M|\alpha)$. The posterior, KDE and 2σ percentile for the estimate of α are shown in the bottom panel of Fig.(4). These are compared with the Fisher predictions above, and once again show very good agreement. We have also repeated the calculation in the absence of selection effects, $\boldsymbol{d}_{\mathrm{th}} \to -\infty$, and find a similar

level of agreement between the Fisher predictions and the MCMC analysis.

We expect that the accuracy of the population Fisher matrix prediction should improve as the number of observations included in the analysis, N_{tot}, increases. We assess this by comparing the result of MCMC analyses of data sets with increasing numbers of observations to the population Fisher matrix. We do this first for the case without selection effects, so that we can make use of the analytical prediction given above. These results are shown in Fig. 5 as N_{tot} is varied from 2 to 30 events. For each N_{tot}, we repeat the MCMC analysis several times to allow us to estimate the variance in the posterior width between different runs. This variance is larger when there are fewer events, as expected, and so more MCMC analyses were performed for lower N_{tot}'s to ensure the variance was accurately characterized. We find that for $N_{\mathrm{tot}} \gtrsim 10$, the (simplified) Fisher widths agree well with the numerical simulations. For $N_{\mathrm{tot}} < 10$, the differences are progressively more pronounced, but the variance in the MCMC widths also increases, and the Fisher matrix prediction is usually within the range spanned by the MCMC runs. The agreement becomes worse for very small numbers of observations, consistent with the expectation that this is an approximation valid in the limit of large N_{tot}. Finally, we check whether a similar level of agreement is seen in the case when the observations are subject to selection effects. For this, we obtain the population Fisher matrix prediction (dashed orange line) by rescaling the $\Delta\alpha$ prediction in Fig. 4 by a factor $N_{\mathrm{det}}^{-1/2}$. These results are also shown in Fig. 4. We see a similar trend – the population Fisher matrix is very accurate for $N_{\mathrm{tot}} > 10$, but the accuracy diminishes for very small numbers of observations, as expected.

4.2 More realistic scenario: SNR distribution in the population

As a final example, we will now make the previous scenario slightly more realistic by adding an additional property to each source, the SNR. This example demonstrates how to compute the population Fisher matrix in a more realistic setting in which the measurement uncertainties depend on the source parameters, and with a more realistic model of selection effects.

For this example, we assume that individual events are characterized by two parameters – a mass, M, drawn from the same power-law population used in the previous example, and an SNR, ρ. We assume that ρ scales with the inverse of distance and that distances are uniform in Euclidean volume, so we have $p(\rho) \propto \rho^{-4}$. We additionally assume that the SNR of a source at a particular distance is proportional to the mass. These assumptions are encoded in the SNR distribution

$$P(\rho < P) = \begin{cases} 1 - \left(\frac{M}{d_{\max} P}\right)^3 & P > \frac{M}{d_{\max}} \\ 0 & \text{otherwise} \end{cases}. \qquad (56)$$

The parameter $d_{\max}$ represents a maximum distance for sources in the population and sets a lower limit on the SNR distribution which avoids divergences. In practice, we choose $d_{\max} \gg M_{\max}/\rho_{\mathrm{th}}$, where ρ_{th} is the SNR threshold for detection, so that the exact choice of $d_{\max}$ does not influence parameter estimation. We note that the assumption that the SNR distribution is biased toward higher values for higher mass systems is not a particularly good model for EMRIs. It would be more appropriate for massive black hole binary systems, but even then the shape of the LISA sensitivity curve is such that this would only apply in a certain range of masses. We choose to make this assumption since we want to demonstrate that the population Fisher matrix works even when there are more
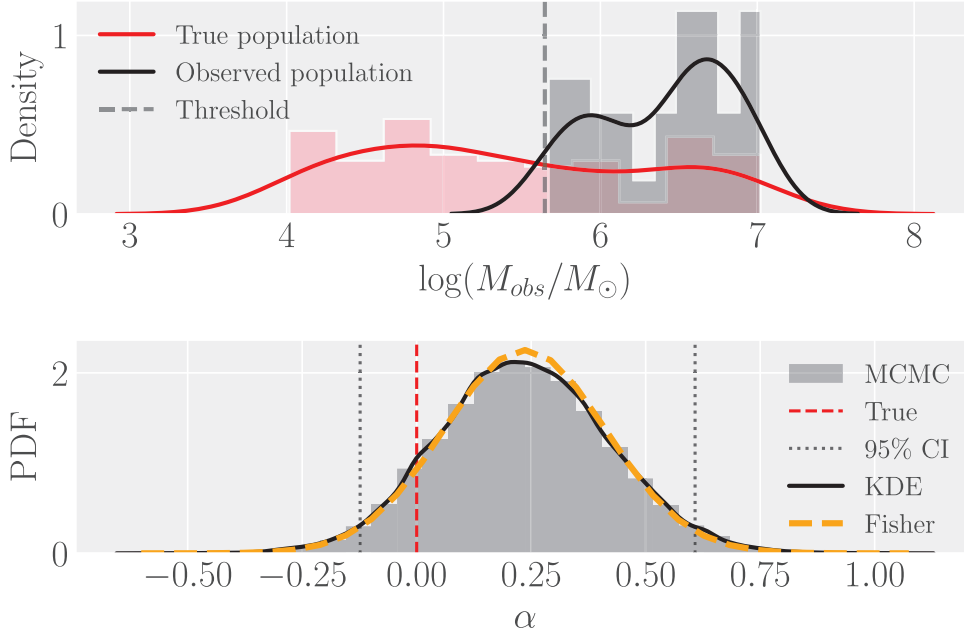
Figure 4. Top panel: distribution of masses for the underlying true population of Section 4 (red) and for the observed population (dark gray). The true population is composed of $N = 100$ events drawn from a power-law model that is flat in the logarithm of the masses. The threshold has been arbitrarily set at $\sim 5 \times 10^5 \, M_\odot$, leading to 39 events actually being observed. Bottom panel: MCMC posterior distribution for the spectral index describing the mass distribution. The histogram and KDE are compared against the Fisher estimate obtained as described in the text, demonstrating very good agreement between the two.
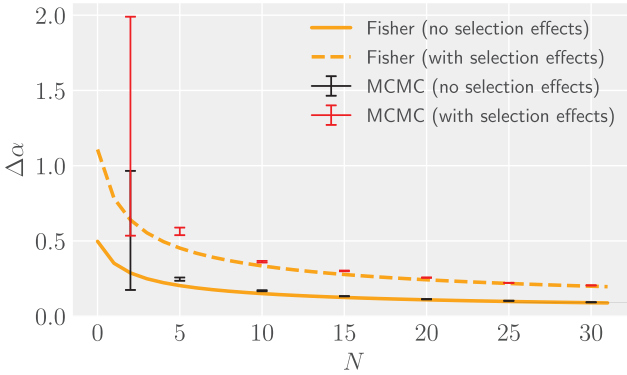

Figure 5. Fisher predictions for the width of the spectral index α as a function of the number of observed events N, compared to the range of measured uncertainties obtained over a set of MCMC runs (red and black points with error bars). The Fisher matrix prediction is approximated with equation (55). The agreement between the Fisher matrix predictions and the MCMC results is very good, especially if the number of events is increased beyond $N \sim 10$.

complicated interactions between the source parameters, including parameters whose distribution is independent of the population parameters.

We assume that a GW observation consists of a noisy measurement, $\hat{\rho}$, of ρ, and a noisy measurement, $\rho \hat{M}$, of ρM, so that the GW likelihood is

$$p(\mathbf{d} = (\hat{\rho}, \rho \hat{M})|\vec{\theta}) = \frac{1}{2\pi \sigma_\rho \sigma_M} \exp\left[-\frac{(\hat{\rho}-\rho)^2}{2\sigma_\rho^2}\right]$$
$$\times \exp\left[-\frac{(\rho \hat{M} - \rho M)^2}{2\sigma_M^2}\right]. \quad (57)$$

We fix $\sigma_\rho = 1$, which follows from the definition of ρ. We assume that selection is based on $\hat{\rho}$ only, with events with $\hat{\rho} > \rho_{th}$ being deemed detectable. We use $\rho_{th} = 10$ in this example. With this likelihood,

the individual source Fisher matrix is

$$\Gamma_{ij} = \begin{pmatrix} 1 + M^2/\sigma_M^2 & \rho M/\sigma_M^2 \\ \rho M/\sigma_M^2 & \rho^2/\sigma_M^2 \end{pmatrix}. \quad (58)$$

We see that the measurement uncertainties vary from event to event, with uncertainties in mass scaling like $1/\rho$, as desired [3]

We simulate observations of a population of events with true slope $\alpha = 0$ and $N_{obs} = 499$ observed events, by drawing $500/P_{det}(\alpha) \approx 28370$ trial systems from the underlying population. We analyse these events using MCMC and compare to the predictions of the population Fisher matrix. Details of how the latter is calculated can be found in Appendix D. For this example, in the limit $\sigma_M \to 0$, the matrices Γ_X for $X = II, \cdots, V$ do not vanish, because we have fixed $\sigma_\rho = 1$. However, for a reasonable choice of $\sigma_M = 10 \, M_\odot$, we find that the contributions from these matrices are again subdominant to the contribution from Γ_I, making only a $\sim 5\%$ change to the prediction for the uncertainty on α. This is true for a wide range of choices of σ_M up to at least $10^4 \, M_\odot$. A comparison of the MCMC results and the population Fisher matrix prediction is shown in Fig. 6, demonstrating once again very precise agreement.

This example has demonstrated that the population Fisher matrix gives accurate predictions even when using a more complicated model that includes non-trivial selection effects and heteroscedastic measurement errors. This model is a more realistic representation to a GW observation scenario, but the result cannot be directly

---

[3] We could have achieved the same result by assuming that the GW data comprises a measurement of ρ and of M, with independent Gaussian errors with variances $\sigma_\rho = 1$ and σ_M/ρ, respectively. However, this model cannot be put into the standard GW likelihood form, equation (3), which assumes the noise variances are parameter independent. This alternative form of the model can be analysed using the generalized formalism described in Appendix C, but we wanted the model to be of the standard GW form. The analysed model is equivalent to setting $\mathbf{h} = (\rho, \rho M)$ and $S_h(f) = (\sigma_\rho^2, \sigma_M^2)$ in equation (3).
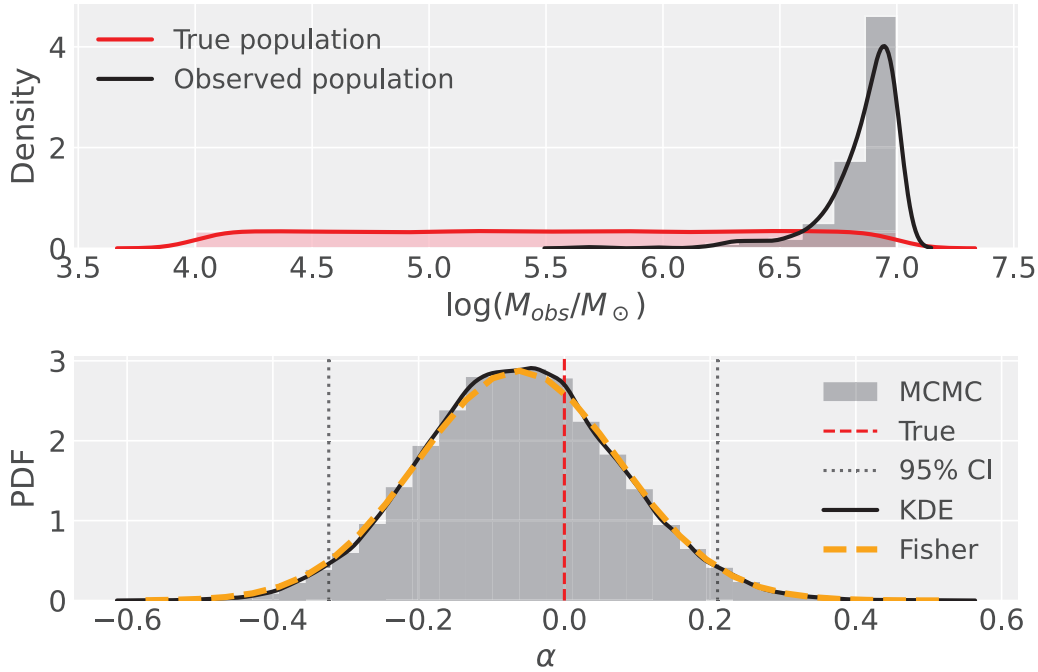
Figure 6. As Fig. 4 but for the slightly more realistic GW-like population model. Top panel: distribution of masses for the underlying true population of Section 4 (red) and for the observed population (dark gray). Bottom panel: MCMC posterior distribution for the spectral index describing the mass distribution. The histogram and KDE are compared against the Fisher estimate obtained as described in the text.

compared to the results presented in GTV, because the assumed SNR distribution is different. Here, we have assumed that sources are distributed uniformly in Euclidean space, following an ρ^{-4} distribution, with SNRs additionally increased in proportion to M. In GTV, sources were distributed based on a computation of SNR that included the impact of other parameters, in particular time to coalescence. Nearby sources generate enough SNR that they can be observed several years before merger, which enhances the rate of nearby events and partially compensates for the fact that there are a larger number of systems further away. For this reason, the fact that we find a distribution that is approximately a factor of 3 broader than GTV is not a cause of concern. Indeed, it is remarkable that the agreement was so close for the simpler example considered in Section 4.1.

We conclude this section by using the population Fisher matrix to explore the impact of the detection threshold, ρ_{th}, on the precision of inference of the population parameters for this simple model. There are two effects of changing the threshold. One effect is that the population Fisher matrix changes. This represents the average uncertainty over detected events and so the elements of Γ_λ tend to become smaller as the threshold is decreased, corresponding to a worse constraint per event. This is because lower SNR events tend to provide less precise parameter estimates. The second effect is that the detection probability, $P_{det}(\vec{\lambda} = \{\alpha\})$, changes, increasing as ρ_{th} decreases. It is therefore useful to consider the quantity $\Gamma_{\alpha\alpha}^{-1}/\sqrt{P_{det}(\alpha)}$, where $\Gamma_{\alpha\alpha}^{-1}$ is the diagonal element of the inverse of the population Fisher matrix. This quantity is a measure of the precision on α that could be obtained in a fixed amount of observation time. We show this quantity, relative to its value for the reference threshold, $\rho_{th} = 10$, in Fig. 7. We see that the precision improves as the threshold is lowered, indicating that the increase in the number of events outweighs the decrease in the average precision per event. In practice, there will be some limit to how much we can lower the threshold, beyond which we can no longer confidently identify
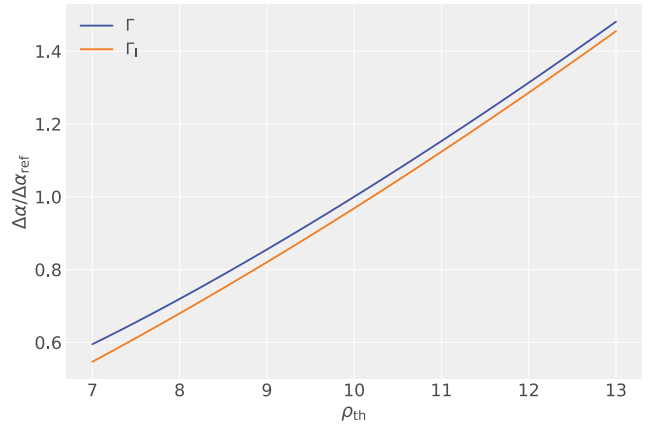

Figure 7. Precision of measurement of the slope of the black hole population distribution obtained in a fixed observation time, as a function of the threshold needed for detection, ρ_{th}. The blue curve shows results using the full population Fisher matrix, while the yellow curve shows results based on Γ_I only. Both curves are expressed as ratios relative to the precision estimated from the full Fisher matrix with $\rho_{th} = 10$.

events, or run into limitations on computational power, but within those constraints these results suggest we should lower the threshold as much as possible. In general, it is at low thresholds that the approximation that the individual events can be well represented by the Fisher matrix will become less valid. This is not fully captured here because of the simplified assumption of the Gaussian likelihood. Moreover, a lot of the trend is captured by Γ_I, which is independent of that approximation. The yellow line in Fig. 7 shows the precision estimated from Γ_I alone, again expressed relative to the value estimated from the full Fisher matrix with $\rho_{th} = 10$. We see that the trend is similar. There is a slightly bigger difference between

the full and approximate Fisher matrices for the lowest values of ρ_{th}, but for all thresholds considered the approximate Fisher matrix gives a good indication of the achievable precision.

The trend in Fig. 7 is specific to the simple model considered here and the behaviour will be different in other contexts. However, this exercise illustrates the usefulness of using the population Fisher matrix to quickly assess the impact of different assumptions on the accuracy of inference. We note, however, that when using it to assess the contribution from low-SNR events, it is important to check the accuracy of the approximation in that regime, as discussed in Section 2.1, to ensure that the conclusions are robust.

5 CONCLUSIONS

The Fisher information matrix is a valuable tool for estimating the precision attainable in parameter inference, especially in contexts where the cost of doing full posterior estimation via Bayesian sampling is highly expensive (Vallisneri 2008). The Fisher matrix has been widely used in GW analyses to make forecasts for the precision with which the *source* parameters describing individual GW signals can be estimated by current and future detectors. In this paper, we have extended the Fisher matrix concept to the estimation of the parameters characterizing the *population* from which a set of observed sources is drawn. Our result was derived from equation (1), which is the most general definition for the population Fisher matrix. We obtained equation (21), which is valid under the assumption that individual events are observed with high enough SNR that the measurement uncertainties can be well approximated by the linear signal approximation. We also identified the part of population Fisher matrix that is independent of measurement uncertainties, given by $\Gamma_{\rm I}$, which is even simpler to evaluate and provides a good approximation when the individual event measurement uncertainties are much smaller than the scale on which the population varies. We have tested this result both analytically and against numerical Monte Carlo results for a reference Gaussian model (Section 3) and for a more GW-like scenario (Section 4), in which, we have used GW events to estimate the slope of a power-law population. We find that equation (21) is generally in very good agreement with the numerical results, for a sufficiently large number of observations. In this case, sufficiently large was only $O(10)$. Results for the power-law population case can be compared to previous results in the literature (Gair et al. 2010), and are found to be in very good agreement, despite very different assumptions. We conclude that we can reproduce the results of extensive sets of computationally expensive MCMC simulations much more cheaply, while also correctly including selection effects. In addition, we found that in the GW-like example the measurement-error-independent part of the population Fisher matrix, $(\Gamma_{\rm I})_{ij}$ in equation (21), is sufficient to accurately reproduce the precision estimated from the full Fisher matrix. This is because the noise-induced uncertainty in the parameter measurements of each individual event, σ, is sufficiently smaller than the scale on which the population model varies, that measurement errors are essentially ignorable. This result could be useful to further reduce the computational cost of computing the population Fisher matrix in other contexts.

We note that these results are based on the approximation that individual event measurements are well represented by the individual event Fisher matrix. There will be contexts in which this is not true, but measurement uncertainties are important so $\Gamma_{\rm I}$ is not dominant. We provided a criterion in Section 2.1 that can be used to evaluate the validity of the approximation. When the approximation is not valid, equation (21) should still provide a rough estimate of the precision of inference, or the threshold can be increased such that

the approximations are valid and a conservative estimate of precision obtained in this way.

The results presented in this paper represent the first attempt at describing population inference within a Fisher formalism for generic population models, and with a likelihood that takes into account selection effects in the way of Mandel et al. (2019). The formalism developed here can be used to obtain forecasts for the precision of population analyses with future ground-based and spaceborne detectors, which are expected to detect many thousands (or even millions) of signals. Obtaining such population inference forecasts in specific contexts of relevance to current and future observations is one possible future direction for the present project. Finally, it would be interesting to generalize the results of Cutler & Vallisneri (2007) and Antonelli, Burke & Gair (2021) to assess inference biases on population parameters from waveform modelling errors or confusion noise.

DATA AVAILABILITY STATEMENT

The results in this paper use NUMPY (Harris et al. 2020), MATPLOTLIB (Hunter 2007), SEABORN (Waskom 2021), EMCEE (Foreman-Mackey et al. 2013), and ARVIZ (Kumar et al. 2019). The results can be fully reproduced with codes made publicly available at https://github.com/aantonelli94/PopFisher.

REFERENCES

Abbott B. P. et al., 2019, Phys. Rev. X, 9, 031040
Abbott R. et al., 2021a, preprint (arXiv:2108.01045)
Abbott R. et al., 2021b, preprint (arXiv:2111.03606)
Abbott R. et al., 2021c, preprint (arXiv:2111.03634)
Abbott R. et al., 2021d, preprint (arXiv:2111.03604)
Amaro-Seoane P., Gair J. R., Freitag M., Coleman Miller M., Mandel I., Cutler C. J., Babak S., 2007, Class. Quant. Grav., 24, R113
Amaro-Seoane P. et al., 2017, preprint (arXiv:1702.00786)
Antonelli A., Burke O., Gair J. R., 2021, MNRAS, 507, 5069
Babak S. et al., 2017, Phys. Rev. D, 95, 103012
Barack L., 2009, Class. Quant. Grav., 26, 213001
Barausse E., Cardoso V., Pani P., 2014, Phys. Rev. D, 89, 104059
Bonetti M., Sesana A., 2020, Phys. Rev. D, 102, 103023
Borhanian S., 2021, Class. Quant. Grav., 38, 175014
Cutler C., Vallisneri M., 2007, Phys. Rev. D, 76, 104018
Fishbach M. et al., 2021, AstrophyJ., 912, 98
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Publ. Astron. Soc. Pac., 125, 306
Gair J. R., Tang C., Volonteri M., 2010, Phys. Rev. D, 81, 104014
Gair J. R., Sesana A., Berti E., Volonteri M., 2011, Class. Quant. Grav., 28, 094018
Gair J. R., Vallisneri M., Larson S. L., Baker J. G., 2013, Living Rev. Rel., 16, 7
Harms J. et al., 2022, preprint (arXiv:2205.02499)
Harris C. R. et al., 2020, Nature, 585, 357
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Kumar R., Carroll C., Hartikainen A., Martin O., 2019, J. Open Source Softw., 4, 1143
Mancarella M., Genoud-Prachex E., Maggiore M., 2022, Phys. Rev. D, 105, 064030
Mandel I., Farr W. M., Gair J. R., 2019, MNRAS, 486, 1086
Mastrogiovanni S. et al., 2021, Phys. Rev. D, 104, 062009
Mould M., Gerosa D., Broekgaarden F. S., Steinle N., 2022, MNRAS, 517, 2738

Mukherjee S., Krolewski A., Wandelt B. D., Silk J., 2022, preprint (arXiv:2203.03643)

Punturo M. et al., 2010, Class. Quant. Grav., 27, 194002

Reitze D. et al., 2019, Bull. Am. Astron. Soc., 51, 035

Rodriguez C. L. et al., 2020, ApJ, 896, L10

Sesana A., Gair J., Berti E., Volonteri M., 2011, Phys. Rev. D, 83, 044036

Taylor S. R., Gerosa D., 2018, Phys. Rev. D, 98, 083017

Vallisneri M., 2008, Phys. Rev. D, 77, 042001

Vitale S., Gerosa D., Farr W. M., Taylor S. R., 2021, in Bambi C., Katsanevas S., Kokkotas K. D., eds, Handbook of Gravitational Wave Astronomy. Springer, Singapore

Vitale S., Biscoveanu S., Talbot C., 2022, Astronomy & Astrophysics, 668, L2

Waskom M. L., 2021, J. Open Source Softw., 6, 3021

APPENDIX A: DEALING WITH RATES

The likelihood in equation (6) assumes that the number of events that are detected conveys no information about the population. Relaxing this assumption the joint likelihood takes the alternative form

$$p(\{\mathbf{d}_i\}|\vec{\lambda}, R) \propto \left[\prod_{i=1}^{n} p_{\text{full}}(\mathbf{d}_i|\vec{\lambda}) \right] R^n \exp[-R P_{\text{det}}(\vec{\lambda})], \tag{A1}$$

in which R is the rate of events occurring in the Universe over the total time data has been collected, and all other terms are as before. The derivation of this expression can be found in Mandel et al. (2019). Imposing an (improper) scale-invariant prior on the total rate, $p(R) \propto 1/R$, and marginalizing over R, we obtain the form of the joint likelihood used in equation (6) and (8).

We denote the rate-dependent terms by

$$p_{\text{rate}}(n|R, \vec{\lambda}) \equiv R^n \exp[-R P_{\text{det}}(\vec{\lambda})]. \tag{A2}$$

The contribution of these terms to the joint log likelihood is

$$\ln p_{\text{rate}}(n|R, \vec{\lambda}) = n \ln R - R P_{\text{det}}(\vec{\lambda}), \tag{A3}$$

which is maximized when $R = n/P_{\text{det}}(\vec{\lambda})$. Expanding about this maximum-likelihood point we can write $R = n/P_{\text{det}} + \delta R$ and obtain

$$\begin{aligned}\ln p_{\text{rate}}(n|R, \vec{\lambda}) &= n \ln n - n \ln P_{\text{det}}(\vec{\lambda}) + n \ln(1 + \delta R P_{\text{det}}/n) \\ &\quad - n - \delta R P_{\text{det}} \\ &= (n \ln n - n) - n \ln P_{\text{det}}(\vec{\lambda}) - \frac{\delta R^2 P_{\text{det}}^2}{2n} + \cdots . \end{aligned} \tag{A4}$$

The second term here, $-n \ln P_{\text{det}}(\vec{\lambda})$, is what is needed to change $p_{\text{full}}(\mathbf{d}_i|\vec{\lambda})$ into $p(\mathbf{d}_i|\vec{\lambda})$ in the product term in equation (A1), reducing that to the form analysed in the main body of the paper. We deduce that the asymptotic Fisher matrix for the joint estimation of R and $\vec{\lambda}$ is block diagonal

$$\Gamma = \begin{pmatrix} \Gamma_\lambda & 0 \\ 0 & \Gamma_R \end{pmatrix} \tag{A5}$$

with the shape parameter block, Γ_λ, as before and the rate precision given by the inverse of $\Gamma_R = P_{\text{det}}^2/2n$, where this result can be obtained directly from the coefficient of δR^2 in equation (A4). We conclude that the precision with which the shape parameters can be determined does not depend on which particular form of the likelihood is being used. This makes sense since we know that the two forms are equivalent for a particular choice of rate prior, and we expect results to be asymptotically independent of the initial prior choice. We note also that the precision with which the rate of observed events, $R P_{\text{det}}$, can be measured is $\sqrt{n}$, consistent with the expected uncertainty in the estimation of the rate of a Poisson process.

To conclude this section, we note that in the above we have been assuming that the rate parameter R is an additional parameter of the

model, separate to the parameters $\vec{\lambda}$ that characterize the shape of the population distribution. If instead, both R and $\vec{\lambda}$ are functions of another set of population parameters, $\vec{\mu}$, we can use the linear signal approximation to change variables and obtain the usual result that the Fisher matrix for the $\vec{\mu}$ parameters is

$$(\Gamma_\mu)_{cd} = (\Gamma_\lambda)_{ab} \frac{\partial \lambda^a}{\partial \mu^c} \frac{\partial \lambda^b}{\partial \mu^d} + \Gamma_R \frac{\partial R}{\partial \mu^c} \frac{\partial R}{\partial \mu^d}. \tag{A6}$$

In this case, the result is different to what would be obtained by transforming the Fisher matrix that ignores the rate, which would be the first term only. This reflects the fact that if R also depends on $\vec{\mu}$, the measurement of R provides additional information that can help to improve the estimation of the $\vec{\mu}$ parameters.

APPENDIX B: THE ASYMPTOTIC BEHAVIOUR OF THE FISHER MATRIX

In this section, we will rederive the expression for the Fisher matrix given in the main body of the paper by directly expanding the posterior distribution. In doing so, we will derive the form and scaling of the leading corrections to the Fisher matrix approximation. We will proceed by computing the posterior mode, mean, and variance in the limit $n \gg 1$. These are all random variables, since they depend on the particular realization of the data that is being analysed, and so we can characterize them by their expectation value and variance. We will show how to compute the first two terms in a large-n expansion of both the mean and variance for all three posterior summary statistics, and give the result explicitly for the posterior mode. We note that similar results for corrections to the individual event Fisher matrix were given in Vallisneri (2008), but those relied on the assumption of a Gaussian likelihood which permits simplifications. The expansion presented here is valid for any population-level likelihood, $p(\mathbf{d}|\vec{\lambda})$. We write

$$\hat{\mu}(n, \{\mathbf{d}_i\}, \vec{\lambda}) = -\frac{1}{n} \sum_{i=1}^{n} \ln p(\mathbf{d}_i|\vec{\lambda}). \tag{B1}$$

such that the posterior distribution is proportional to $\exp[-n\hat{\mu}(n, \{\mathbf{d}_i\}, \vec{\lambda}) + \ln \pi(\vec{\lambda})]$. We use $\vec{\beta}(n, \{\mathbf{d}_i\}, \vec{\lambda}_t)$ to denote the solution to

$$U_i + U_i^\pi + (V_{ij} + V_{ij}^\pi)\beta^j = 0 \tag{B2}$$

where

$$U_i = \left(\frac{\partial \hat{\mu}}{\partial \lambda^i} \right)_{|\vec{\lambda}_t}, \qquad U_i^\pi = -\frac{1}{n} \left(\frac{\partial \ln \pi}{\partial \lambda^i} \right)_{|\vec{\lambda}_t}$$

$$V_{ij} = \left(\frac{\partial^2 \hat{\mu}}{\partial \lambda^i \partial \lambda^j} \right)_{|\vec{\lambda}_t}, \qquad V_{ij}^\pi = -\frac{1}{n} \left(\frac{\partial^2 \ln \pi}{\partial \lambda^i \partial \lambda^j} \right)_{|\vec{\lambda}_t}. \tag{B3}$$

The quantity $\hat{\mu}$, and its derivatives, are averages of a set of independent identically distributed (IID) random variables and so have predictable scalings. The expectation value is $O(1)$, covariances are $O(1/n)$, three and four point functions are $O(1/n^2)$, and so on. In this case $\mathbb{E}[\mathbb{U}] = 0$, as shown in equation (12) in the main body of the paper. This facilitates obtaining a solution for $\vec{\beta}$ perturbatively

$$\vec{\beta} = \vec{\beta}_{\frac{1}{2}} + \vec{\beta}_1 + \vec{\beta}_{\frac{3}{2}} + \cdots$$

where $\vec{\beta}_{\frac{1}{2}} = (\mathbb{E}[\mathbf{V}])^{-1} \mathbf{U}$

$$\vec{\beta}_1 = (\mathbb{E}[\mathbf{V}])^{-1} \left[(\mathbf{V} - \mathbb{E}[\mathbf{V}])\vec{\beta}_{\frac{1}{2}} + \mathbf{U}^\pi \right]$$

$$\vec{\beta}_{\frac{3}{2}} = (\mathbb{E}[\mathbf{V}])^{-1} \left[(\mathbf{V} - \mathbb{E}[\mathbf{V}])\vec{\beta}_1 + \mathbf{V}^\pi \vec{\beta}_{\frac{1}{2}} \right], \tag{B4}$$

in which $\vec{\beta}_k \sim n^{-k}$.

The $\vec{\beta}_k$'s are random variables, but we can compute their means

$$\mathbb{E}\left[\vec{\beta}_{\frac{1}{2}}\right] = 0$$

$$\mathbb{E}\left[\beta_1^i\right] = (\mathbb{E}[\mathbf{V}])_{ij}^{-1}\left[(\mathbb{E}[\mathbf{V}])_{kl}^{-1}\Sigma_{jkl}^{VU} + \mathbf{U}_j^\pi\right]$$

$$\mathbb{E}\left[\beta_{\frac{3}{2}}^i\right] = (\mathbb{E}[\mathbf{V}])_{ij}^{-1}(\mathbb{E}[\mathbf{V}])_{kl}^{-1}(\mathbb{E}[\mathbf{V}])_{mn}^{-1}\Sigma_{jklmn}^{VVU}$$

$$\Sigma_{ijk}^{VU} = \mathbb{E}\left[(V_{ij} - \mathbb{E}[V_{ij}])U_k\right]$$

$$= \frac{1}{n}\int l_{,ij}(\mathbf{d})l_{,k}(\mathbf{d})p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$

$$\Sigma_{ijklm}^{VVU} = \mathbb{E}\left[(V_{ij} - \mathbb{E}[V_{ij}])(V_{kl} - \mathbb{E}[V_{kl}])U_m\right]$$

$$= \frac{1}{n^2}\int l_{,ij}(\mathbf{d})l_{,kl}(\mathbf{d})l_{,m}(\mathbf{d})\, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$

$$- \mathbb{E}[V_{ij}]\Sigma_{klm}^{VU} - \mathbb{E}[V_{kl}]\Sigma_{ijm}^{VU} \qquad (B5)$$

where we are using the notation $l_{,i}(\mathbf{d})$ to denote the derivative $\partial \ln p(\mathbf{d}|\vec{\lambda})/\partial\lambda^i$ evaluated at $\vec{\lambda} = \vec{\lambda}_t$. Additional indices after the comma indicate further partial derivatives as usual. The fact that the first expectation value vanishes is why we have continued the expansion to three terms, allowing us to obtain the first two terms in an expansion of the mean.

We can also compute their covariances, using the usual notation $\mathrm{cov}(a^i, b^j) = \mathbb{E}[(a^i - \mathbb{E}[a^i])(b^j - \mathbb{E}[b^j])]$.

$$\mathrm{cov}(\beta_{\frac{1}{2}}^i, \beta_{\frac{1}{2}}^j) = (\mathbb{E}[\mathbf{V}])_{ik}^{-1}(\mathbb{E}[\mathbf{V}])_{jl}^{-1}\Sigma_{kl}^{UU}$$

$$\mathrm{cov}(\beta_{\frac{1}{2}}^i, \beta_1^j) = (\mathbb{E}[\mathbf{V}])_{im}^{-1}(\mathbb{E}[\mathbf{V}])_{jk}^{-1}(\mathbb{E}[\mathbf{V}])_{ln}^{-1}\Sigma_{klmn}^{VUU}$$

$$\mathrm{cov}(\beta_1^i, \beta_1^j) = (\mathbb{E}[\mathbf{V}])_{ik}^{-1}(\mathbb{E}[\mathbf{V}])_{lm}^{-1}(\mathbb{E}[\mathbf{V}])_{jp}^{-1}(\mathbb{E}[\mathbf{V}])_{qr}^{-1}\Sigma_{klpqmr}^{VVUU}$$

$$- (\mathbb{E}[\mathbf{V}])_{ik}^{-1}(\mathbb{E}[\mathbf{V}])_{lm}^{-1}\Sigma_{klm}^{VU}\Sigma_{pqr}^{VU}$$

$$\mathrm{cov}(\beta_{\frac{3}{2}}^i, \beta_{\frac{1}{2}}^j) = (\mathbb{E}[\mathbf{V}])_{ik}^{-1}(\mathbb{E}[\mathbf{V}])_{jq}^{-1}(\mathbb{E}[\mathbf{V}])_{lm}^{-1}$$

$$\times \left[(\mathbb{E}[\mathbf{V}])_{np}^{-1}\Sigma_{klmnpq}^{VVUU} + \mathbf{U}_m^\pi\Sigma_{klq}^{VU} + \mathbf{V}_{kl}^\pi\Sigma_{klmq}^{VUU}\right]$$

$$\Sigma_{ij}^{UU} = \mathbb{E}\left[U_iU_j\right]$$

$$= \frac{1}{n}\int l_{,i}(\mathbf{d})l_{,j}(\mathbf{d})\, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$

$$\Sigma_{ijkl}^{VUU} = \mathbb{E}\left[(V_{ij} - \mathbb{E}[V_{ij}])U_kU_l\right]$$

$$= \frac{1}{n^2}\int l_{,ij}(\mathbf{d})l_{,k}(\mathbf{d})l_{,l}(\mathbf{d})\, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$

$$- \mathbb{E}[V_{ij}]\Sigma_{kl}^{UU}$$

$$\Sigma_{ijklmn}^{VVUU} = \mathbb{E}\left[(V_{ij} - \mathbb{E}[V_{ij}])(V_{kl} - \mathbb{E}[V_{kl}])U_mU_n\right]$$

$$= \frac{n(n-1)}{n^4}\left[\int l_{,ij}(\mathbf{d})l_{,kl}(\mathbf{d})\, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}\right.$$

$$\times \int l_{,m}(\mathbf{d})l_{,n}(\mathbf{d})\, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$

$$+ \int l_{,ij}(\mathbf{d})l_{,m}(\mathbf{d})\, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$

$$\times \int l_{,kl}(\mathbf{d})l_{,n}(\mathbf{d})\, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$

$$\left. + m \leftrightarrow n\right]$$

$$+ \frac{1}{n^3}\int l_{,ij}(\mathbf{d})l_{,kl}(\mathbf{d})l_{,m}(\mathbf{d})l_{,n}(\mathbf{d})\, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$

$$- \mathbb{E}[V_{ij}]\Sigma_{klmn}^{VUU} - \mathbb{E}[V_{kl}]\Sigma_{ijmn}^{VUU}$$

$$- \mathbb{E}[V_{ij}]\mathbb{E}[V_{kl}]\Sigma_{mn}^{UU} \qquad (B6)$$

All four of these terms are needed to obtain the first two terms in the covariance of $\vec{\beta}$. We need only retain terms of $O(1/n^2)$ in the final

expression, which means the last term in Σ_{ijklmn}^{VVUU} can be ignored and $n(n-1)/n^4$ replaced by n^{-2}. The final result, keeping the first two orders is

$$\mathrm{cov}\left(\beta^i, \beta^j\right) = \mathrm{cov}\left(\beta_{\frac{1}{2}}^i, \beta_{\frac{1}{2}}^j\right) + \mathrm{cov}\left(\beta_{\frac{1}{2}}^i, \beta_1^j\right)$$

$$+ \mathrm{cov}\left(\beta_{\frac{1}{2}}^j, \beta_1^i\right) + \mathrm{cov}\left(\beta_1^i, \beta_1^j\right)$$

$$+ \mathrm{cov}\left(\beta_{\frac{3}{2}}^i, \beta_{\frac{1}{2}}^j\right) + \mathrm{cov}\left(\beta_{\frac{3}{2}}^j, \beta_{\frac{1}{2}}^i\right) \qquad (B7)$$

The posterior mode is at $\vec{\lambda} = \vec{\lambda}_t + \vec{\beta} + \vec{\delta}$. The random variable $\vec{\delta}$ obeys the equation

$$0 = \delta^j V_{ij} + \frac{1}{2}(\beta^j + \delta^j)(\beta^k + \delta^k)W_{ijk} + V_{ij}^\pi\delta^j$$

$$+ \frac{1}{2}(\beta^j + \delta^j)(\beta^k + \delta^k)W_{ijk}^\pi$$

$$W_{ijk} = \left(\frac{\partial^3\hat{\mu}}{\partial\lambda^i\partial\lambda^j\partial\lambda^k}\right)_{|\vec{\lambda}_t}$$

$$W_{ijk}^\pi = -\frac{1}{n}\left(\frac{\partial^3\ln\pi}{\partial\lambda^i\partial\lambda^j\partial\lambda^k}\right)_{|\vec{\lambda}_t}. \qquad (B8)$$

We can find a perturbative solution as we did for $\vec{\beta}$

$$\vec{\delta} = \vec{\delta}_1 + \vec{\delta}_{\frac{3}{2}} + \cdots$$

where $\delta_1^i = \frac{1}{2}(\mathbb{E}[\mathbf{V}])_{ij}^{-1}\mathbb{E}[W_{jkl}]\beta_{\frac{1}{2}}^k\beta_{\frac{1}{2}}^l$

$$\delta_{\frac{3}{2}}^i = (\mathbb{E}[\mathbf{V}])_{ij}^{-1}\left[\beta_{\frac{1}{2}}^k\beta_1^l\mathbb{E}[W_{jkl}] + \beta_{\frac{1}{2}}^k\delta_1^l\mathbb{E}[W_{jkl}]\right.$$

$$+ \left(V_{jk} - \mathbb{E}[V_{jk}]\right)\delta_1^k$$

$$\left. + \frac{1}{2}(W_{jkl} - \mathbb{E}[W_{jkl}])\beta_{\frac{1}{2}}^k\beta_{\frac{1}{2}}^l\right] \qquad (B9)$$

The means and relevant covariances are

$$\mathbb{E}\left[\delta_1^i\right] = \frac{1}{2}(\mathbb{E}[\mathbf{V}])_{ij}^{-1}\mathbb{E}[W_{jkl}]\,\mathrm{cov}\left(\beta_{\frac{1}{2}}^k, \beta_{\frac{1}{2}}^l\right)$$

$$\mathbb{E}\left[\delta_{\frac{3}{2}}^i\right] = (\mathbb{E}[\mathbf{V}])_{ij}^{-1}\left[\mathbb{E}[W_{jkl}]\,\mathrm{cov}\left(\beta_{\frac{1}{2}}^k, \beta_1^l\right)\right.$$

$$+ \mathbb{E}[W_{jkl}]\,\mathrm{cov}\left(\beta_{\frac{1}{2}}^k, \delta_1^l\right)$$

$$+ \frac{1}{2}(\mathbb{E}[\mathbf{V}])_{kn}^{-1}\mathbb{E}[W_{npq}](\mathbb{E}[\mathbf{V}])_{pl}^{-1}(\mathbb{E}[\mathbf{V}])_{qm}^{-1}\Sigma_{jklm}^{VUU}$$

$$\left. + \frac{1}{2}(\mathbb{E}[\mathbf{V}])_{km}^{-1}(\mathbb{E}[\mathbf{V}])_{ln}^{-1}\Sigma_{jklmn}^{WUU}\right]$$

$$\mathrm{cov}\left(\delta_1^i, \delta_1^j\right) = \frac{1}{2}(\mathbb{E}[\mathbf{V}])_{ik}^{-1}(\mathbb{E}[\mathbf{V}])_{jl}^{-1}(\mathbb{E}[\mathbf{V}])_{mx}^{-1}(\mathbb{E}[\mathbf{V}])_{ny}^{-1}$$

$$(\mathbb{E}[\mathbf{V}])_{pr}^{-1}(\mathbb{E}[\mathbf{V}])_{qs}^{-1}\mathbb{E}[W_{kmn}]\mathbb{E}[W_{lpq}]\Sigma_{xyrs}^{UUUU}$$

$$- \mathbb{E}[\delta_1^i]\mathbb{E}[\delta_1^i]$$

$$\mathrm{cov}\left(\delta_1^i, \beta_{\frac{1}{2}}^j\right) = \frac{1}{2}(\mathbb{E}[\mathbf{V}])_{ik}^{-1}(\mathbb{E}[\mathbf{V}])_{jp}^{-1}(\mathbb{E}[\mathbf{V}])_{lq}^{-1}(\mathbb{E}[\mathbf{V}])_{mr}^{-1}$$

$$\mathbb{E}[W_{klm}]\Sigma_{pqr}^{UUU}$$

$$\mathrm{cov}\left(\delta_1^i, \beta_1^j\right) = \frac{1}{2}(\mathbb{E}[\mathbf{V}])_{ik}^{-1}(\mathbb{E}[\mathbf{V}])_{lp}^{-1}(\mathbb{E}[\mathbf{V}])_{mq}^{-1}(\mathbb{E}[\mathbf{V}])_{jr}^{-1}$$

$$\mathbb{E}[W_{klm}]\left((\mathbb{E}[\mathbf{V}])_{sx}^{-1}\Sigma_{rsxpq}^{VUUU} + U_r^\pi\Sigma_{pq}^{UU}\right)$$

$$- \mathbb{E}[\delta_1^i]\mathbb{E}[\beta_1^j]$$

$$\text{cov}\left(\delta_{\frac{3}{2}}^i, \beta_{\frac{1}{2}}^j\right) = (\mathbb{E}[\mathbf{V}])_{ik}^{-1} (\mathbb{E}[\mathbf{V}])_{jp}^{-1} \left[\mathbb{E}[W_{klm}] (\mathbb{E}[\mathbf{V}])_{lq}^{-1}\right.$$
$$\left((\mathbb{E}[\mathbf{V}])_{mx}^{-1} (\mathbb{E}[\mathbf{V}])_{yz}^{-1} \Sigma_{xyzpq}^{VUUU} + U_x^\pi \Sigma_{pq}^{UU}\right)$$
$$+ \frac{1}{2}\mathbb{E}[W_{klm}]\mathbb{E}[W_{xyz}] (\mathbb{E}[\mathbf{V}])_{lq}^{-1} (\mathbb{E}[\mathbf{V}])_{mx}^{-1}$$
$$(\mathbb{E}[\mathbf{V}])_{yu}^{-1} (\mathbb{E}[\mathbf{V}])_{zv}^{-1} \Sigma_{pquv}^{UUUU}$$
$$+ \frac{1}{2} (\mathbb{E}[\mathbf{V}])_{lq}^{-1} (\mathbb{E}[\mathbf{V}])_{xr}^{-1} (\mathbb{E}[\mathbf{V}])_{ys}^{-1}$$
$$\mathbb{E}[W_{qxy}]\Sigma_{klrsp}^{VUUU}$$
$$\left. + \frac{1}{2} (\mathbb{E}[\mathbf{V}])_{ln}^{-1} (\mathbb{E}[\mathbf{V}])_{mq}^{-1} \Sigma_{klmnpq}^{WUUU}\right] \tag{B10}$$

where

$$\Sigma_{ijklm}^{WUU} = \mathbb{E}\left[(W_{ijk} - \mathbb{E}[W_{ijk}])U_l U_m\right]$$
$$= \frac{1}{n^2} \int l_{,ijk}(\mathbf{d})l_{,l}(\mathbf{d})l_{,m}(\mathbf{d}) \, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$
$$- \mathbb{E}[W_{ijk}]\Sigma_{lm}^{UU}$$
$$\Sigma_{ijk}^{UUU} = \mathbb{E}\left[U_i U_j U_k\right]$$
$$= \frac{1}{n^2} \int l_{,i}(\mathbf{d})l_{,j}(\mathbf{d})l_{,k}(\mathbf{d}) \, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$
$$\Sigma_{ijkl}^{UUUU} = \mathbb{E}\left[U_i U_j U_k U_l\right]$$
$$= \frac{n(n-1)}{n^4}\left[\int l_{,i}(\mathbf{d})l_{,j}(\mathbf{d}) \, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}\right.$$
$$\left.\times \int l_{,k}(\mathbf{d})l_{,l}(\mathbf{d}) \, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}\right]$$
$$+ \frac{1}{n^3} \int l_{,i}(\mathbf{d})l_{,j}(\mathbf{d})l_{,k}(\mathbf{d})l_{,l}(\mathbf{d}) \, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$
$$\Sigma_{ijklm}^{VUUU} = \mathbb{E}\left[(V_{ij} - \mathbb{E}[V_{ij}])U_k U_l U_m\right]$$
$$= \frac{n(n-1)}{n^4}\left[\int l_{,ij}(\mathbf{d})l_{,k}(\mathbf{d}) \, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}\right.$$
$$\times \int l_{,l}(\mathbf{d})l_{,m}(\mathbf{d}) \, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$
$$\left. + k \leftrightarrow l + k \leftrightarrow m\right]$$
$$+ \frac{1}{n^3} \int l_{,ij}(\mathbf{d})l_{,k}(\mathbf{d})l_{,l}(\mathbf{d})l_{,m}(\mathbf{d}) \, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$
$$- \mathbb{E}[V_{ij}]\Sigma_{klm}^{UUU}$$
$$\Sigma_{ijklmn}^{WUUU} = \mathbb{E}\left[(W_{ijk} - \mathbb{E}[W_{ijk}])U_l U_m U_n\right]$$
$$= \frac{n(n-1)}{n^4}\left[\int l_{,ijk}(\mathbf{d})l_{,l}(\mathbf{d}) \, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}\right.$$
$$\times \int l_{,m}(\mathbf{d})l_{,n}(\mathbf{d}) \, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$
$$\left. + l \leftrightarrow m + l \leftrightarrow n\right]$$
$$+ \frac{1}{n^3} \int l_{,ijk}(\mathbf{d})l_{,l}(\mathbf{d})l_{,m}(\mathbf{d})l_{,n}(\mathbf{d}) \, p(\mathbf{d}|\vec{\lambda}_t)\mathrm{d}\mathbf{d}$$
$$- \mathbb{E}[W_{ijk}]\Sigma_{lmn}^{UUU}. \tag{B11}$$

From these we can construct the leading order covariances

$$\text{cov}\left(\beta^i, \delta^j\right) = \text{cov}\left(\beta_{\frac{1}{2}}^i, \delta_1^j\right) + \text{cov}\left(\beta_1^i, \delta_1^j\right) + \text{cov}\left(\beta_{\frac{1}{2}}^i, \delta_{\frac{3}{2}}^j\right)$$
$$\text{cov}\left(\delta^i, \delta^j\right) = \text{cov}\left(\delta_1^i, \delta_1^j\right). \tag{B12}$$

Putting these together, we can obtain the mean and covariance of the posterior mode, $\hat{\lambda}^i - \lambda_t^i = \beta^i + \delta^i$, expressed as a deviation from the

true population parameters.

$$\mathbb{E}[\hat{\lambda}^i - \lambda_t^i] = \mathbb{E}[\beta^i] + \mathbb{E}[\delta^i]$$
$$\text{cov}\left(\hat{\lambda}^i - \lambda_t^i, \hat{\lambda}^j - \lambda_t^j\right) = \text{cov}\left(\beta^i, \beta^j\right) + \text{cov}\left(\delta^i, \delta^j\right)$$
$$+ \text{cov}\left(\beta^i, \delta^j\right) + \text{cov}\left(\beta^j, \delta^i\right) \tag{B13}$$

We now turn our attention to the posterior mean and variance. These are averages over the posterior. The definition of $\vec{\beta}$ was motivated to ensure the leading terms of the log posterior can be written as a quadratic in $\vec{\lambda} - \vec{\beta} - \vec{\lambda}_t$. Denoting $\bar{B} = \vec{\lambda}_t + \vec{\beta}$, averages of a function $f(\vec{\lambda})$ over the posterior then take the form

$$\langle f \rangle = \frac{\int f(\vec{\lambda})g(\vec{\lambda}) \exp\left[-\frac{n}{2}(\lambda^i - B^i)(V_{ij} + V_{ij}^\pi)(\lambda^j - B^j)\right] \mathrm{d}\vec{\lambda}}{\int g(\vec{\lambda}) \exp\left[-\frac{n}{2}(\lambda^i - B^i)(V_{ij} + V_{ij}^\pi)(\lambda^j - B^j)\right] \mathrm{d}\vec{\lambda}}$$
$$\ln g(\vec{\lambda}) = -\frac{n}{6}W_{ijk}(\lambda^i - \lambda_t^i)(\lambda^j - \lambda_t^j)(\lambda^k - \lambda_t^k)$$
$$- \frac{n}{24}X_{ijkl}(\lambda^i - \lambda_t^i)(\lambda^j - \lambda_t^j)(\lambda^k - \lambda_t^k)(\lambda^l - \lambda_t^l)$$
$$X_{ijkl} = \left(\frac{\partial^4\hat{\mu}}{\partial\lambda^i\partial\lambda^j\partial\lambda^k\partial\lambda^l}\right)_{|\vec{\lambda}_t} \tag{B14}$$

There are further corrections in $g(\vec{\lambda})$ from higher derivates in the expansion, and from the prior terms, W_{ijk}^π, X_{ijkl}^π, etc. However, the contributions from the included terms can be seen to be $1/n$ down relative to the leading terms in the integral, and these other corrections are at least $1/n^{\frac{3}{2}}$ down from leading and are hence sub-dominant.

Integrals of these form are standard and we will make use of the following results

$$I_0(\Gamma) = \int \exp\left[-\frac{1}{2}\mathbf{x}^T\Gamma^{-1}\mathbf{x}\right] \mathrm{d}\mathbf{x} = (2\pi)^{\frac{N}{2}}\sqrt{|\Gamma|} \tag{B15}$$

$$I_{ij}(\Gamma) = \int x_i x_j \exp\left[-\frac{1}{2}\mathbf{x}^T\Gamma^{-1}\mathbf{x}\right] \mathrm{d}\mathbf{x}$$
$$= (2\pi)^{\frac{N}{2}}\sqrt{|\Gamma|}\Gamma_{ij} \tag{B16}$$

$$I_{ijkl}(\Gamma) = \int x_i x_j x_k x_l \exp\left[-\frac{1}{2}\mathbf{x}^T\Gamma^{-1}\mathbf{x}\right] \mathrm{d}\mathbf{x}$$
$$= (2\pi)^{\frac{N}{2}}\sqrt{|\Gamma|}\left(\Gamma_{ij}\Gamma_{kl} + \Gamma_{ik}\Gamma_{jl} + \Gamma_{il}\Gamma_{jk}\right) \tag{B17}$$

$$I_{ijklmn}(\Gamma) = \int x_i x_j x_k x_l x_m x_n \exp\left[-\frac{1}{2}\mathbf{x}^T\Gamma^{-1}\mathbf{x}\right] \mathrm{d}\mathbf{x}$$
$$= (2\pi)^{\frac{N}{2}}\sqrt{|\Gamma|}\left(\Gamma_{ij}\Gamma_{kl}\Gamma_{mn} + \Gamma_{ij}\Gamma_{km}\Gamma_{ln}\right.$$
$$+ \Gamma_{ij}\Gamma_{kn}\Gamma_{lm} + \Gamma_{ik}\Gamma_{jl}\Gamma_{mn} + \Gamma_{ik}\Gamma_{jm}\Gamma_{ln}$$
$$+ \Gamma_{ik}\Gamma_{jn}\Gamma_{lm} + \Gamma_{il}\Gamma_{jk}\Gamma_{mn} + \Gamma_{il}\Gamma_{jm}\Gamma_{kn}$$
$$+ \Gamma_{il}\Gamma_{jn}\Gamma_{km} + \Gamma_{im}\Gamma_{jk}\Gamma_{ln} + \Gamma_{im}\Gamma_{jl}\Gamma_{kn}$$
$$+ \Gamma_{im}\Gamma_{jn}\Gamma_{kl} + \Gamma_{in}\Gamma_{jk}\Gamma_{lm} + \Gamma_{in}\Gamma_{jl}\Gamma_{km}$$
$$\left. + \Gamma_{in}\Gamma_{jm}\Gamma_{kl}\right) \tag{B18}$$

where $|\Gamma|$ denotes the determinant of Γ. We will also use the notation $\tilde{I}_{ij}(\Gamma) \equiv I_{ij}(\Gamma)/I_0(\Gamma)$ and similarly for other terms. In this case, the covariance matrix $\Gamma = (\mathbf{V} + \mathbf{V}^\pi)^{-1}/n$. Every additional factor of Γ therefore introduces an extra negative power of n. This allows us to identify the dominant terms. To evaluate the above expressions, we need to be able to compute Γ, which can be done perturbatively by noting

$$n\Gamma(\mathbb{E}[\mathbf{V}] + (\mathbf{V} - \mathbb{E}[\mathbf{V}]) + \mathbf{V}^\pi) = \mathbf{I} \tag{B19}$$

from which

$$\Gamma = \frac{1}{n}\left(\Gamma_0 + \Gamma_{\frac{1}{2}} + \Gamma_1 + \cdots\right)$$
$$\Gamma_0 = (\mathbb{E}[\mathbf{V}])^{-1}$$
$$\Gamma_{\frac{1}{2}} = -(\mathbb{E}[\mathbf{V}])^{-1}(\mathbf{V} - \mathbb{E}[\mathbf{V}])$$
$$\Gamma_1 = -(\mathbb{E}[\mathbf{V}])^{-1}\left[\Gamma_{\frac{1}{2}}(\mathbf{V} - \mathbb{E}[\mathbf{V}]) + \mathbf{V}^\pi\right]. \tag{B20}$$

To obtain the posterior mean, we first compute

$$\langle g \rangle = \int g(\vec{\lambda}) \exp\left[-\frac{n}{2}(\lambda^i - B^i)(V_{ij} + V_{ij}^\pi)(\lambda^j - B^j)\right] d\vec{\lambda}$$
$$= I_0(\Gamma) - \frac{n}{6}W_{ijk}\left[\beta^i I_{jk}(\Gamma) + \beta^j I_{kl}(\Gamma) + \beta^l I_{ij}(\Gamma)\right]$$
$$- \frac{n}{6}W_{ijk}\beta^i\beta^j\beta^k I_0(\Gamma) + \frac{n^2}{72}W_{ijk}W_{lmn}I_{ijklmn}(\Gamma)$$
$$- \frac{n}{24}X_{ijkl}I_{ijkl}(\Gamma). \tag{B21}$$

Using similar notation to before we can write

$$\langle g \rangle = I_0(\Gamma)\left(1 + g_{\frac{1}{2}} + g_1 + \cdots\right)$$
$$g_{\frac{1}{2}} = -\frac{1}{6}\mathbb{E}[W_{ijk}]\left(\beta^i_{\frac{1}{2}}(\Gamma_0)_{jk} + \beta^j_{\frac{1}{2}}(\Gamma_0)_{kl} + \beta^l_{\frac{1}{2}}(\Gamma_0)_{ij}\right)$$
$$- \frac{n}{6}\mathbb{E}[W_{ijk}]\beta^i_{\frac{1}{2}}\beta^j_{\frac{1}{2}}\beta^k_{\frac{1}{2}}$$
$$g_1 = -\frac{1}{6}(W_{ijk} - \mathbb{E}[W_{ijk}])$$
$$\times \left(\beta^i_{\frac{1}{2}}(\Gamma_0)_{jk} + \beta^j_{\frac{1}{2}}(\Gamma_0)_{kl} + \beta^l_{\frac{1}{2}}(\Gamma_0)_{ij}\right)$$
$$- \frac{1}{6}\mathbb{E}[W_{ijk}]\left(\beta^i_{\frac{1}{2}}(\Gamma_{\frac{1}{2}})_{jk} + \beta^j_{\frac{1}{2}}(\Gamma_{\frac{1}{2}})_{kl} + \beta^l_{\frac{1}{2}}(\Gamma_{\frac{1}{2}})_{ij}\right.$$
$$\left. + \beta^i_1(\Gamma_0)_{jk} + \beta^j_1(\Gamma_0)_{kl} + \beta^l_1(\Gamma_0)_{ij}\right)$$
$$- \frac{n}{6}(W_{ijk} - \mathbb{E}[W_{ijk}])\beta^i_{\frac{1}{2}}\beta^j_{\frac{1}{2}}\beta^k_{\frac{1}{2}}$$
$$- \frac{n}{6}\mathbb{E}[W_{ijk}]\left(\beta^i_1\beta^j_{\frac{1}{2}}\beta^k_{\frac{1}{2}} + \beta^i_{\frac{1}{2}}\beta^j_1\beta^k_{\frac{1}{2}} + \beta^i_{\frac{1}{2}}\beta^j_{\frac{1}{2}}\beta^k_1\right)$$
$$+ \frac{1}{72}\mathbb{E}[W_{ijk}]\mathbb{E}[W_{lmn}]\frac{1}{n}\tilde{I}_{ijklmn}(\Gamma_0)$$
$$- \frac{1}{24}\frac{1}{n}\mathbb{E}[X_{ijkl}]\tilde{I}_{ijkl}(\Gamma_0) \tag{B22}$$

Now, we compute the posterior mean, expressed as a distance from the true population parameters

$$\bar{\lambda}^i - \lambda^i_t \equiv \langle(\lambda^i - \lambda^i_t)\rangle = \beta^i + \langle(\lambda^i - B^i)\rangle$$
$$= \beta^i + \frac{1}{\langle g \rangle}\left[-\frac{n}{6}W_{jkl}\left(I_{ijkl}(\Gamma) + 3\beta^j\beta^k I_{il}(\Gamma)\right)\right.$$
$$- \frac{n}{6}X_{jklm}\left(\beta^j I_{iklm}(\Gamma) + \beta^j\beta^k\beta^l I_{im}(\Gamma)\right)$$
$$+ \frac{n^2}{36}W_{jkl}W_{mnp}\left(3\beta^j I_{iklmnp}(\Gamma)\right.$$
$$+ \beta^j\beta^k\beta^l I_{imnp}(\Gamma) + 3\beta^j\beta^k\beta^m I_{ilnp}(\Gamma)$$
$$\left.\left. + 3\beta^j\beta^k\beta^l\beta^m\beta^n I_{ip}(\Gamma)\right)\right]. \tag{B23}$$

We note that the leading order correction in the bracketed term is $1/n$. To obtain the posterior mean to the same order as β, we therefore

only need to retain terms up to $g_{\frac{1}{2}}$ in $\langle g \rangle$. Specifically, we can write

$$\bar{\lambda}^i - \lambda^i_t = \bar{\lambda}^i_{\frac{1}{2}} + \bar{\lambda}^i_1 + \bar{\lambda}^i_{\frac{3}{2}} + \cdots$$
$$\bar{\lambda}^i_{\frac{1}{2}} = \beta^i_{\frac{1}{2}}$$
$$\bar{\lambda}^i_1 = \beta^i_1 - \frac{1}{6}\mathbb{E}[W_{jkl}]\left(\frac{1}{n}\tilde{I}_{ijkl}(\Gamma_0) + 3\beta^j_{\frac{1}{2}}\beta^k_{\frac{1}{2}}(\Gamma_0)_{il}\right)$$
$$\bar{\lambda}^i_{\frac{3}{2}} = \beta^i_{\frac{3}{2}} + \left(\frac{g_{\frac{1}{2}}}{6}\mathbb{E}[W_{jkl}] - \frac{1}{6}(W_{jkl} - \mathbb{E}[W_{jkl}])\right)$$
$$\times \left(\frac{1}{n}\tilde{I}_{ijkl}(\Gamma_0) + 3\beta^j_{\frac{1}{2}}\beta^k_{\frac{1}{2}}(\Gamma_0)_{il}\right)$$
$$- \frac{1}{6}\mathbb{E}[W_{jkl}]\left(\frac{2}{n}\left((\Gamma_0)_{ij}(\Gamma_{\frac{1}{2}})_{kl} + (\Gamma_0)_{ik}(\Gamma_{\frac{1}{2}})_{jl}\right)\right.$$
$$+ (\Gamma_0)_{il}(\Gamma_{\frac{1}{2}})_{jk}$$
$$+ 3(\beta^j_1\beta^k_{\frac{1}{2}} + \beta^j_{\frac{1}{2}}\beta^k_1)(\Gamma_0)_{il}$$
$$\left. + 3\beta^j_{\frac{1}{2}}\beta^k_{\frac{1}{2}}(\Gamma_{\frac{1}{2}})_{il}\right)$$
$$- \frac{1}{6}\mathbb{E}[X_{jklm}]\left(\beta^j_{\frac{1}{2}}\frac{1}{n}\tilde{I}_{iklm} + \beta^j_{\frac{1}{2}}\beta^k_{\frac{1}{2}}\beta^l_{\frac{1}{2}}(\Gamma_0)_{im}\right)$$
$$+ \frac{1}{36}\mathbb{E}[W_{jkl}]\mathbb{E}[W_{mnp}]\left(\frac{3}{n}\beta^j_{\frac{1}{2}}\tilde{I}_{iklmnp}(\Gamma_0)\right.$$
$$+ \beta^j_{\frac{1}{2}}\beta^k_{\frac{1}{2}}\beta^l_{\frac{1}{2}}\tilde{I}_{imnp}(\Gamma_0) + 3\beta^j_{\frac{1}{2}}\beta^k_{\frac{1}{2}}\beta^m_{\frac{1}{2}}\tilde{I}_{ilnp}(\Gamma_0)$$
$$\left. + 3n\beta^j_{\frac{1}{2}}\beta^k_{\frac{1}{2}}\beta^l_{\frac{1}{2}}\beta^m_{\frac{1}{2}}\beta^n_{\frac{1}{2}}(\Gamma_0)_{ip}\right) \tag{B24}$$

Finally, we consider the posterior covariance

$$\hat{\Gamma}_{ij} \equiv \langle(\lambda^i - \bar{\lambda}^i)(\lambda^j - \bar{\lambda}^j)\rangle$$
$$= \langle(\lambda^i - B^i)(\lambda^j - B^j)\rangle$$
$$+ (B^i - \bar{\lambda}^i)\langle(\lambda^j - B^j)\rangle + i \leftrightarrow j$$
$$+ (B^i - \bar{\lambda}^i)(B^j - \bar{\lambda}^j). \tag{B25}$$

The term on the first line has a leading order dependence of $1/n$, plus corrections of $1/n^{\frac{3}{2}}$. The terms on the second and third lines are $O(1/n^2)$ and so are sub-dominant. We deduce

$$\hat{\Gamma}_{ij} = \frac{1}{\langle g \rangle}\left[I_{ij}(\Gamma) - \frac{n}{6}W_{klm}\left(3\beta^k I_{ijlm}(\Gamma) + \beta^k\beta^l\beta^m I_{ij}(\Gamma)\right)\right] \tag{B26}$$

and expand

$$\hat{\Gamma}_{ij} = (\hat{\Gamma}_1)_{ij} + \left(\hat{\Gamma}_{\frac{3}{2}}\right)_{ij} + \cdots$$
$$(\hat{\Gamma}_1)_{ij} = \frac{1}{n}(\Gamma_0)_{ij}$$
$$\left(\hat{\Gamma}_{\frac{3}{2}}\right)_{ij} = \frac{1}{n}\left(\Gamma_{\frac{1}{2}}\right)_{ij} - \frac{g_{\frac{1}{2}}}{n}(\Gamma_0)_{ij}$$
$$- \frac{1}{6}\mathbb{E}[W_{klm}]\left(\frac{3}{n}\beta^k_{\frac{1}{2}}\tilde{I}_{ijlm}(\Gamma_0) + \beta^k_{\frac{1}{2}}\beta^l_{\frac{1}{2}}\beta^m_{\frac{1}{2}}(\Gamma_0)_{ij}\right) \tag{B27}$$

Using the preceding expressions, we could now compute the mean and variance of the posterior mean and covariance as we did for the shift in the posterior mode. However, this calculation is very similar to the calculations carried out above and is tedious, so we leave it out. Instead, we note a number of features.

(i) The leading order posterior covariance is constant and equal to $(\mathbb{E}[\mathbf{V}])^{-1}/n$, which is the expression we used to derive the population Fisher matrix in the body of the paper.

(ii) The leading order difference between either the posterior mode or mean and the true parameter value has expectation value that scales like $1/n$ and variance that also scales like $1/n$. Thus, the posterior

bias is noise dominated, i.e. fluctuations due to the particular random realization of the population that was observed dominate over the fixed bias. If multiple sets of observations of n events were repeated and averaged, then this bias would eventually be significant. In practice, we would never do this since it weakens the precision of inference. This means that computing corrections to the posterior mean is unnecessary.

(iii) Similarly, the expectation value of $\left(\hat{\Gamma}_{\frac{3}{2}}\right)_{ij}$ is zero, so the leading order correction to the expected value of the posterior covariance scales like n^{-2}, while the variance in the posterior covariance scales like n^{-3}. So, fluctuations in the posterior covariance due to randomness in the observed population are larger than the size of corrections from the finite number of observations.

APPENDIX C: GENERALIZATION TO OTHER LIKELIHOODS

Expression 21 was derived for the GW likelihood defined by equation (3), but it can be extended to more general likelihoods, $p(\mathbf{d}|\vec{\theta})$. The likelihood enters the result through the definitions of Γ, N_i, D_i, and D_{ij}. For a more general likelihood, we have

$$\Gamma_{ij} = -\frac{\partial^2 \ln p(\mathbf{d}|\vec{\theta})}{\partial \theta^i \partial \theta^j}$$
$$N_i = \frac{\partial \ln p(\mathbf{d}|\vec{\theta})}{\partial \theta^i}, \tag{C1}$$

where derivatives are evaluated at $\vec{\theta}_0$. In the gravitational wave case, these become

$$\Gamma_{ij} = \left(\frac{\partial \mathbf{h}}{\partial \theta^i}\bigg|\frac{\partial \mathbf{h}}{\partial \theta^j}\right) - \left(\mathbf{d} - \mathbf{h}(\vec{\theta}_0)\bigg|\frac{\partial^2 \mathbf{h}}{\partial \theta^i \partial \theta^j}\right)$$
$$N_i = \left(\mathbf{d} - \mathbf{h}(\vec{\theta}_0)\bigg|\frac{\partial^2 \mathbf{h}}{\partial \theta^i \partial \theta^j}\right). \tag{C2}$$

Dropping the second term in the expressions for Γ on the grounds that it is smaller by a factor of ρ^{-1} than the first, we recover the expressions used in the earlier derivation. In particular, we note that with this simplification, Γ_{ij} does not depend on $\mathbf{d}$ and hence, we can take the terms that depend on Γ outside of the integral over data, simplifying the final form of the population Fisher matrix. For a more general likelihood, we can not assume this is the case, but have $\Gamma_{ij}(\mathbf{d}, \vec{\theta}_0)$ and $N_i(\mathbf{d}, \vec{\theta}_0)$. The first contributions to the population likelihood, Γ_I takes the same form as before, but the other contributions are modified to

$$(\Gamma_{II})_{ij} = \frac{1}{2} \int \int \frac{\partial^2 \ln \det(\Gamma + H)}{\partial \lambda^i \partial \lambda^j}$$
$$\times p(\mathbf{d}|\vec{\theta}_0)\frac{p(\vec{\theta}_0|\vec{\lambda})}{P_{\det}(\vec{\lambda})}\mathrm{d}\mathbf{d}\mathrm{d}\vec{\theta}_0,$$

$$(\Gamma_{III})_{ij} = -\frac{1}{2} \int \int \frac{\partial^2}{\partial \lambda^i \partial \lambda^j}\left[(\Gamma + H)_{kl}^{-1}\right] N_k N_l$$
$$\times p(\mathbf{d}|\vec{\theta}_0)\frac{p(\vec{\theta}_0|\vec{\lambda})}{P_{\det}(\vec{\lambda})}\mathrm{d}\mathbf{d}\mathrm{d}\vec{\theta}_0,$$

$$(\Gamma_{IV})_{ij} = -\int \int \frac{\partial^2}{\partial \lambda^i \partial \lambda^j}\left[P_k(\Gamma + H)_{kl}^{-1}\right] N_l$$
$$\times p(\mathbf{d}|\vec{\theta}_0)\frac{p(\vec{\theta}_0|\vec{\lambda})}{P_{\det}(\vec{\lambda})}\mathrm{d}\mathbf{d}\mathrm{d}\vec{\theta}_0,$$

$$(\Gamma_{V})_{ij} = -\frac{1}{2} \int \int \frac{\partial^2}{\partial \lambda^i \partial \lambda^j}\left[P_k(\Gamma + H)_{kl}^{-1} P_l\right]$$
$$\times p(\mathbf{d}|\vec{\theta}_0)\frac{p(\vec{\theta}_0|\vec{\lambda})}{P_{\det}(\vec{\lambda})}\mathrm{d}\mathbf{d}\mathrm{d}\vec{\theta}_0,$$

where the integrals over $\mathbf{d}$ are over detectable data sets. A further simplification can be obtained if we assume that the variance of Γ_{ij} over realizations of the data, $p(\mathbf{d}|\vec{\theta}_0)$, is small. This allows us to use the fact that the expectation value of any function, $f(X, Y)$, of two random variables X and Y, can be expanded

$$\mathbb{E}[f(X, Y)] = f(\mathbb{E}[X], \mathbb{E}[Y]) + \frac{1}{2}\mathrm{Var}(X)\left(\frac{\partial^2 f}{\partial X^2}\right)_{(\mathbb{E}(X), \mathbb{E}(Y))}$$
$$+ \mathrm{Cov}(X, Y)\left(\frac{\partial^2 f}{\partial X \partial Y}\right)_{(\mathbb{E}(X), \mathbb{E}(Y))}$$
$$+ \frac{1}{2}\mathrm{Var}(Y)\left(\frac{\partial^2 f}{\partial Y^2}\right)_{(\mathbb{E}(X), \mathbb{E}(Y))} + \cdots \tag{C3}$$

Ignoring all but the leading term allows us to replace Γ by its expectation value. With this additional assumption, the population Fisher matrix for the general case takes the sane form as before, with the substitutions $\Gamma_{ij} \to \bar{\Gamma}_{ij}$, $D_i \to \bar{D}_i$ and $D_{ij} \to \bar{D}_{ij}$, where

$$\bar{\Gamma}_{ij} = \int \left[-\frac{\partial^2 \ln p(\mathbf{d}|\vec{\theta})}{\partial \theta^i \partial \theta^j}\right] p(\mathbf{d}|\vec{\theta}_0)\mathrm{d}\mathbf{d}$$
$$\bar{D}_i = \int \left[\frac{\partial \ln p(\mathbf{d}|\vec{\theta})}{\partial \theta^i}\right] p(\mathbf{d}|\vec{\theta}_0)\mathrm{d}\mathbf{d} = \frac{\partial P_{\det}(\vec{\theta})}{\partial \theta^i}$$
$$\bar{D}_{ij} = \int \left[\frac{\partial \ln p(\mathbf{d}|\vec{\theta})}{\partial \theta^i}\frac{\partial \ln p(\mathbf{d}|\vec{\theta})}{\partial \theta^j}\right] p(\mathbf{d}|\vec{\theta}_0)\mathrm{d}\mathbf{d} \tag{C4}$$

APPENDIX D: CALCULATION OF THE FISHER MATRIX FOR THE POWER-LAW POPULATION WITH SNR DISTRIBUTION

Here, we provide a guide to computing the Fisher matrix for the more realistic GW-like example described in Section 4.2. The source parameters are $\vec{\theta} = (\rho, M)$ and the population parameter is $\vec{\lambda} = (\alpha)$.

The full population model is

$$p(\vec{\theta}|\vec{\lambda}) = \frac{3M^3}{d_{\max}^3}\frac{1}{\rho^4}\frac{\alpha M^{\alpha-1}}{M_{\max}^\alpha - M_{\min}^\alpha}$$
$$= \frac{3\alpha}{d_{\max}^3(M_{\max}^\alpha - M_{\min}^\alpha)}\frac{1}{\rho^4}M^{2+\alpha} \tag{D1}$$

from which we can deduce

$$H_{ij} = \begin{pmatrix} -\frac{4}{\rho^2} & 0 \\ 0 & \frac{(2+\alpha)}{M^2} \end{pmatrix}. \tag{D2}$$

The single event Fisher matrix, Γ, was given in equation (58). The detection probability is

$$P_{\det}(\vec{\theta}) = \frac{1}{2}\mathrm{erfc}\left[-\frac{(\rho_{\mathrm{th}} - \rho)}{\sqrt{2}}\right]. \tag{D3}$$

The determinant of $\Gamma + H$ is

$$\det(\Gamma + H) = \frac{rho^2}{\sigma_M^2} - \frac{4}{\sigma_M^2} + \frac{(2+\alpha)}{M^2} - \frac{4(2+\alpha)}{\rho^2 M^2} + \frac{(2+\alpha)}{\sigma_M^2} \tag{D4}$$

which has first derivative

$$\frac{\partial}{\partial \alpha}\det(\Gamma + H) = \frac{1}{M^2} + \frac{1}{\sigma_M^2} - \frac{4}{\rho^2 M^2} \tag{D5}$$

and the second derivative vanishes. The integrals required for the matrices Γ_{III}, Γ_{IV}, and Γ_V all take the form

$$\Gamma_X = -\int \frac{\partial^2}{\partial \alpha^2}\left(\frac{A_X}{\det(\Gamma + H)}\right)\frac{p(\vec{\theta}_0|\alpha)}{P_{\det}(\alpha)}\mathrm{d}\vec{\theta}_0, \tag{D6}$$

where

$$A_{III} = \frac{1}{2}\left(\frac{\rho^2}{\sigma_M^2} + \frac{(2+\alpha)}{M^2}\right)\frac{(\rho_{th}-\rho)}{\sqrt{2\pi}}\exp\left[-\frac{(\rho_{th}-\rho)^2}{2}\right]$$
$$+ \left(\frac{\rho^2}{\sigma_M^2} - \frac{(2+\alpha)}{2\sigma_M^2} + \frac{(2+\alpha)}{2M^2}\right)P_{det}(\theta_0)$$

$$A_{IV} = -\left(\frac{(6+\alpha)\rho}{\sigma_M^2} + \frac{4(2+\alpha)}{\rho M^2}\right)\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(\rho_{th}-\rho)^2}{2}\right]$$

$$A_V = \left(\frac{8}{\sigma_M^2} - \frac{8(2+\alpha)}{\rho^2 M^2} + \frac{4(2+\alpha)}{\sigma_M^2} - \frac{2(a+\alpha)^2}{\rho^2 M^2}\right.$$
$$\left. + \frac{(2+\alpha)^2}{2M^2} + \frac{(2+\alpha)2}{2\sigma_M^2}\right)P_{det}(\vec{\theta}_0). \tag{D7}$$

These terms have first derivatives

$$\frac{\partial A_{III}}{\partial \alpha} = \frac{(\rho_{th}-\rho)}{\sqrt{2\pi}M^2}\exp\left[-\frac{(\rho_{th}-\rho)^2}{2}\right]$$
$$+ \frac{1}{2}\left(\frac{1}{\sigma_M^2} + \frac{1}{M^2}\right)P_{det}(\vec{\theta}_0) \tag{D8}$$

$$\frac{\partial A_{IV}}{\partial \alpha} = -\left(\frac{4}{\rho^2 M^2} + \frac{\rho}{\sigma_M^2}\right)\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(\rho_{th}-\rho)^2}{2}\right]$$

$$\frac{\partial A_V}{\partial \alpha} = \left(\frac{(2+\alpha)}{M^2} + \frac{(6+\alpha)}{\sigma_M^2} - \frac{4\alpha}{\rho^2 M^2}\right)P_{det}(\vec{\theta}_0) \tag{D9}$$

and all second derivatives vanish except

$$\frac{\partial^2 A_V}{\partial \alpha^2} = \left(-\frac{4}{\rho^2 M^2} + \frac{1}{M^2} + \frac{1}{\sigma_M^2}\right)P_{det}(\vec{\theta}_0). \tag{D10}$$

These expressions allow all of the integrands that determine the different parts of the population Fisher matrix to be evaluated. The final stage of computing the Fisher matrix is to carry out the integrals over the population distribution $p(\vec{\theta}_0|\vec{\lambda})$. This must be done numerically, but it is facilitated by doing a coordinate transformation

$$v(\vec{\theta}_0) = \frac{M^\alpha - M_{min}^\alpha}{M_{max}^\alpha - M_{min}^\alpha} \tag{D11}$$

$$u(\vec{\theta}_0) = 1 - \left(\frac{M}{d_{max}\rho}\right)^3, \tag{D12}$$

which reduces the population integral

$$\int p(\vec{\theta}_0|\lambda)d\vec{\theta}_0 \to \int du dv. \tag{D13}$$

Further computational efficiencies can be obtained by restricting the range of u considered for each v so that only SNRs $\rho > \rho_{th} - 5$ are included. Codes to compute the Fisher matrix using this procedure are available at https://github.com/aantonelli94/PopFisher.

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