Enlighten – Research publications by members of the University of Glasgow http://eprints.gla.ac.uk

Avoiding selection bias in gravitational wave astronomy

C Messenger[1] and J Veitch[2]

[1] School of Physics and Astronomy, Cardiff University, Queen's Buildings, The Parade, Cardif, Wales CF24 3AA, UK
[2] Nikhef, Science Park 105, Amsterdam 1098-XG, Netherlands
E-mail: chris.messenger@astro.cf.ac.uk and johnv@nikhef.nl

Abstract. When searching for gravitational waves in the data from ground-based gravitational wave detectors, it is common to use a detection threshold to reduce the number of background events which are unlikely to be the signals of interest. However, imposing such a threshold will also discard some real signals with low amplitude, which can potentially bias any inferences drawn from the population of detected signals. We show how this selection bias is naturally avoided by using the full information from the search, considering both the selected data and our ignorance of the data that are thrown away, and considering all relevant signal and noise models. This approach produces unbiased estimates of parameters even in the presence of false alarms and incomplete data. This can be seen as an extension of previous methods into the high false rate regime where we are able to show that the quality of parameter inference can be optimized by lowering thresholds and increasing the false alarm rate.

IOP Institute of Physics   Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Contents

1. Introduction

Future generations of ground-based gravitational wave (GW) detectors, such as the under-construction Advanced LIGO [1], Advanced Virgo [2] and Kagra [3] or the proposed Einstein Telescope [4], are expected to detect a multitude of GW sources in the coming years [5]. The analysis of these first detected signals will help to answer questions about the rate of binary coalescence [6] and the astrophysical distribution of neutron star and black hole masses [7]. As the number of observed sources increases, it will be possible to use populations of GW sources to perform cosmological parameter estimation [8–11] and to test general relativity [12–14]. Previous searches for GWs have used a hierarchical pipeline based on a detection statistic, with a threshold applied to this quantity to reduce the amount of data to be processed by later stages of the pipeline [15, 16]. Because of the very low amplitude of the expected signals relative to the detector noise level, this process eliminates signals which fail to reach the threshold as well as background noise, amounting to a selection of the loudest events from all the true signals. Any analysis which attempts to draw inferences from a population of signals or events which have been chosen in this way is vulnerable to *selection bias* if the population of detected signals does not match that of the underlying sources. We draw a careful distinction between selection bias and biases induced by an incorrect application of prior information [17, 18], as the terms 'Malmquist bias' [19–21] or 'Lutz–Kelker bias' [22] are sometimes used to describe both types of bias.

Many areas of astrophysics are subject to such effects, either by intentional selection or thresholding to reduce the background while performing data analysis, or by limited sensitivity of the instrument (e.g. a flux limited survey [23]). Examples from the literature include searches for Type Ia supernovae (SNIa) [24], gravitational waves (GWs) [16], extra-solar planets [25] and high energy neutrinos [26, 27]; and galaxy cluster surveys [28, 29]. Often the aim is to use the resulting set of observations to infer global parameters of interest, e.g. quantities that characterize the population distribution. In the case of SNIa for example, one may aim to infer cosmological parameters from luminosity distance and redshift measurements. In cosmology, the traditional approach has been to produce unbiased *estimators* of individual event parameters [30–34] which are then used as input for inference of global quantities. Thresholding has been used to reduce the number of false alarms from a background population to a tolerable level, for example thresholding on radio pulse width is applied to data within the LOFAR NuMoon project [35]. Adrián-Martínez *et al* [36] apply a threshold on a fitting parameter to optimize the separation of cosmic neutrinos and background for the ANTARES neutrino experiment. In GW searches [37], thresholds are set on the signal-to-noise-ratio (SNR) of an

event to reduce the volume of data to be processed before additional cuts are applied to minimize the false alarm rate due to detector noise (which is another example of selection). Candidates from future GW searches will be used as input for inferences on global parameters, such as the parameters of the binary neutron-star (BNS) mass distribution.

It has been previously shown [38], that to avoid bias in the analysis of astronomical data in a signal-dominated environment (i.e. no false alarms), one should account for both the detections *and* the false dismissals within the experiment. In this paper we provide a prescription for completely avoiding selection bias by extending that approach beyond the signal dominated regime. To do this we must consider the cases of detection and dismissal for all other relevant models, most importantly the noise model. Our approach accounts for the data which are thrown away by the application of the threshold, making no assumption that they contain only noise, nor do we assume that detected events are produced only by true signals. Without this extension, future global inferences from GW signals would be limited to the high-signal regime of [38], which discards all events that do not meet the threshold, resulting in sub-optimal performance. Our method will therefore be useful in extracting the maximum information from a thresholded search, and is also applicable to many other sensitivity-limited analyses. We demonstrate this with an example, where lowering the threshold increases the precision of parameter estimation for the BNS distribution and rate, which remains unbiased despite the presence of false alarms.

2. General formulation

Consider an experiment in which many measurements are made, each producing data which may contain potential detections, and which are then compared with a detection threshold. We refer to any measurements which pass the threshold as 'triggers', since in most practical cases we will include one or more noise-only models alongside the true signal model. Any measurements which do not pass the threshold are discarded. These non-trigger cases also carry information about the source population, and one should not assume the lack of a trigger implies the lack of true signal, or vice versa.

We divide the experiment into statistically independent multiple measurements, and denote the outcome of each as D. In practice this might mean dividing a search into multiple time periods. The threshold is set on the measured quantity x, which we will call the 'detection statistic'. In general x may be a maximum-likelihood statistic or simply a direct measurement of a physical parameter value but in all circumstances will be subject to noise.

If the statistic passes a threshold $x > x_{th}$ a trigger is produced, and the value of x is recorded; we denote this case D^+, and $D = (D^+, x)$. If on the other hand $x \leqslant x_{th}$, no trigger is produced and no value for x is recorded and $D = D^-$. We define $\{\mathcal{H}\}$ as our set of plausible models and must include at least two models with one describing the presence and the other the absence of a signal. In general the method extends to many alternative signal models. When specifying a set of models in this way we allow that any model may be true, and that the trigger has a non-zero probability under all of the models. For any single model $\mathcal{H}$ the likelihood function defining the distribution of the statistic x is expressed as $p(x|\mathcal{H}, \theta_{\mathcal{H}}, \gamma, I)$ where $\theta_{\mathcal{H}}$ are the parameters unique to this model, and γ are the *global* parameters which are common to all trials (following standard Bayesian notation we use I throughout to represent additional information regarding the experiment, and the global parameters γ). Upon performing the single trial, we are interested in inferring the parameters γ. In the case of a statistic above threshold,

D^+, we can write the posterior distribution on $\boldsymbol{\gamma}$ as

$$p(\boldsymbol{\gamma}|D^+, x, I) = \sum_{\{\mathcal{H}\}} p(\boldsymbol{\gamma}, \mathcal{H}|D^+, x, I)$$

$$= \frac{p(\boldsymbol{\gamma}|I)}{P(x|I)} \overbrace{\sum_{\{\mathcal{H}\}} p(x|\boldsymbol{\gamma}, \mathcal{H}, I) P(\mathcal{H}|\boldsymbol{\gamma}, I)}^{\text{trigger likelihood } p(D^+, x|\boldsymbol{\gamma}, I)}, \tag{1}$$

where $P(\mathcal{H}|\boldsymbol{\gamma}, I)$ represents the prior probability on the model $\mathcal{H}$ conditional on the global parameters. Here we are summing the trigger likelihoods over the set of possible models $\{\mathcal{H}\}$, and we do not assume that the model priors are independent of the global parameters, which will be important when discussing signal abundance and rates. The range of possible models can include those where more than one signal is present in the data. We have used the fact that the probability of a trigger conditional on the value x is strictly equal to $H(x - x_{\text{th}})$ (where $H(x)$ is the Heaviside step function) which is equal to unity for $x > x_{\text{th}}$. If the model has local parameters $\boldsymbol{\theta}_{\mathcal{H}}$, we calculate the factor $p(x|\boldsymbol{\gamma}, \mathcal{H}, I)$, i.e. the likelihood of the data conditional on the model and the global parameters, by marginalization:

$$p(x|\boldsymbol{\gamma}, \mathcal{H}, I) = \int_{\Theta_{\mathcal{H}}} d\boldsymbol{\theta}_{\mathcal{H}} \, p(x|\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\gamma}, \mathcal{H}, I) p(\boldsymbol{\theta}_{\mathcal{H}}|\boldsymbol{\gamma}, \mathcal{H}, I), \tag{2}$$

where $\Theta_{\mathcal{H}}$ is the local parameter space associated with the local parameter $\theta_{\mathcal{H}}$ and the model $\mathcal{H}$. Note that we do not modify this likelihood function to reflect our knowledge of the threshold.

In the alternative case D^- where we discarded the specific value of x, we do still know that $x \leqslant x_{\text{th}}$. For a non-trigger we can therefore write the posterior probability on $\boldsymbol{\gamma}$ (analogous to equation (1)) as

$$p(\boldsymbol{\gamma}|D^-, I) = \frac{p(\boldsymbol{\gamma}|I)}{P(D^-|I)} \overbrace{\sum_{\{\mathcal{H}\}} P(D^-|\boldsymbol{\gamma}, \mathcal{H}, I) P(\mathcal{H}|\boldsymbol{\gamma}, I)}^{\text{non-trigger likelihood } p(D^-|\boldsymbol{\gamma}, I)}. \tag{3}$$

As the value of x is unknown the posterior on $\boldsymbol{\gamma}$ must be marginalized over both the local parameters $\boldsymbol{\theta}_{\mathcal{H}}$ and the unknown data x in the range $x \leqslant x_{\text{th}}$. Therefore only the information D^- appears in the likelihood of a non-trigger conditional on the model and the global parameters:

$$P(D^-|\boldsymbol{\gamma}, \mathcal{H}, I) = \int_{\Theta_{\mathcal{H}}} d\boldsymbol{\theta}_{\mathcal{H}} \int_{x \leqslant x_{\text{th}}} dx \, p(x|\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\gamma}, \mathcal{H}, I) p(\boldsymbol{\theta}_{\mathcal{H}}|\boldsymbol{\gamma}, \mathcal{H}, I).$$
$$\tag{4}$$

The quantity derived in equation (4) is related to what is known as the *selection function* or *detection efficiency*, defined as the probability of a trigger given the true signal parameters for a given signal model. Here it appears in a modified form: the probability of a *non-trigger* given a set of global parameters and a model. It is only in this term that the detection threshold x_{th} actually appears. In practice this quantity (just like the detection efficiency) can be difficult to estimate in some cases.

2.1. An ensemble of trials

Using equations (1) and (3) together with Bayes' theorem we are able to isolate the single trial likelihood functions (identified by the over-braces) in both the trigger and non-trigger cases,

$P(D^+, x|\boldsymbol{\gamma}, I)$ and $p(D^-|\boldsymbol{\gamma}, I)$ respectively. We can now also consider analysing an ensemble of N trials denoted as $\{D\}$, where we index each D_i using $i \in (1, N)$. The posterior probability for the global parameters after N independent trials are performed is simply

$$p(\boldsymbol{\gamma}|\{D\}, I) = \frac{p(\boldsymbol{\gamma}|I)}{P(\{D\}|I)} \prod_{i=1}^{N} p(D_i|\boldsymbol{\gamma}, I). \tag{5}$$

We can further decompose this expression into trigger D^+ and non-trigger D^- factors,

$$p(\boldsymbol{\gamma}|\{D\}, I) = \frac{p(\boldsymbol{\gamma}|I)}{P(\{D\}|I)} \prod_{j=1}^{n} p(D^+, x_j|\boldsymbol{\gamma}, I_j) \prod_{k=1}^{N-n} P(D^-|\boldsymbol{\gamma}, I_k), \tag{6}$$

where j and k index the n triggers and the $N - n$ non-triggers respectively, and I_j, I_k allow for different background information (such as time-varying detector sensitivity) in each measurement.

We now focus on the common case where there are only two possible models: the presence ($\mathcal{H}^+$) or absence ($\mathcal{H}^-$) of a single signal per trial. We can reach this case in practice by considering a trial to be a period of time where the probability of >1 signal is arbitrarily small. Equation (6) becomes

$$p(\boldsymbol{\gamma}|\{D\}, I) = \frac{p(\boldsymbol{\gamma}|I)}{P(\{D\}|I)}$$

$$\times \prod_{j=1}^{n} \left[p(D^+, x_j|\boldsymbol{\gamma}, \mathcal{H}^+, I_j) P(\mathcal{H}^+|\boldsymbol{\gamma}, I_j) + p(D^+, x_j|\boldsymbol{\gamma}, \mathcal{H}^-, I_j) P(\mathcal{H}^-|\boldsymbol{\gamma}, I_j) \right]$$

$$\times \prod_{k=1}^{N-n} \left[P(D^-|\boldsymbol{\gamma}, \mathcal{H}^+, I_k) P(\mathcal{H}^+|\boldsymbol{\gamma}, I_k) + P(D^-|\boldsymbol{\gamma}, \mathcal{H}^-, I_k) P(\mathcal{H}^-|\boldsymbol{\gamma}, I_k) \right]. \tag{7}$$

To relate this to more standard statistical terms, consider the key components which make up the expression, each of which are in reference to single trials:

- $p(D^+, x_j|\boldsymbol{\gamma}, \mathcal{H}^+, I_j)$ is the likelihood of producing a trigger with statistic value x_j when a signal is truly present.
- $p(D^+, x_j|\boldsymbol{\gamma}, \mathcal{H}^-, I_j)$ is the likelihood of producing a trigger with statistic value x_j from only the background distribution.
- $P(D^-|\boldsymbol{\gamma}, \mathcal{H}^+, I_k)$ is the likelihood that a true signal does not produce a trigger, i.e. the *false dismissal likelihood* (or $1 -$ detection efficiency).
- $P(D^-|\boldsymbol{\gamma}, \mathcal{H}^-, I_k)$ is the likelihood that no trigger is produced when no signal is present, i.e. the *true dismissal likelihood*.

The abundance or true rate of signal events is important in the two model case, when the two models are for a single signal and background. As the two models give a complete description of the data, we can write $P(\mathcal{H}^+|\boldsymbol{\gamma}, I) = 1 - P(\mathcal{H}^-|\boldsymbol{\gamma}, I)$. The actual rate of signals, or a quantity from which it can be derived, must be included as one of the global parameters $\boldsymbol{\gamma}$, and if it is not known then it can be estimated from the ensemble of trials. This makes the prior probabilities for the models dependent on the global parameters $\boldsymbol{\gamma}$, so the joint posterior distribution for the rate will not in general be separable from the rest of global parameters. Alternatively, if the true rate of signals is known, it can be substituted into these expressions

instead. However, using incorrect assumptions about the rate parameter will therefore bias inference on other global parameters.

The key idea of our approach (in terms best suited to the two model problem) is to use the true and false dismissal probabilities to incorporate the information from the absence of triggers. This requires that these quantities be either calculated or measured. Without these, any observations which can produce non-triggers risk being either biased or sub-optimal and may require ad hoc corrections. It is only by acknowledging and accounting for ignorance that we can derive the correct results. Equation (7) contains all the parts necessary to infer global parameters without bias, making full use of the information that is kept, and accounting for that which is deliberately (or necessarily) discarded.

3. A gravitational wave example

Searches for GWs signals from compact binary coalescences employ a threshold when selecting which triggers to keep, and so astrophysical statements drawn from these observations will be subject to selection effects [37]. To show how our method can be used to avoid bias, we will consider a toy model in which a GW detector searches for BNSs, and we want to infer the global parameters: the event rate per unit co-moving volume R, and the mean μ_m and standard deviation σ_m of the component mass distribution. We assume that the population of component masses is governed by a Gaussian distribution, and that a single trial constitutes an all-sky search over the duration $\Delta t = 60$ s. For each trial, the data comprises a measurement of the SNR ρ and estimates of the component masses $\mathbf{m} = (m_1, m_2)$ of any event that exceeds the SNR threshold ρ_{th}. In reality, the SNR (our detection statistic) is generated through matched filtering of the data. This process includes both analytical maximization of the SNR over one or more of the local nuisance parameters e.g. phase at coalescence, and numerical maximization over a bank of templates on the component mass space. For simplicity, we assume the component mass estimates to be noise free exact measurements although in practice there will be some uncertainty in their values. In addition we ignore the complications associated with the template bank search over the component mass space noting that in practice, maximization of a detection statistic over a bank of templates would act to modify the distribution of the detection statistic for both the signal and noise models. Although our model is simple and lacks the detail of real observations, it will serve to show the salient points of the method.

We suppose that events arise through either a signal model $\mathcal{H}^+$ or through a noise model $\mathcal{H}^-$. According to the signal model we define the expected (optimal) SNR-squared of a source as

$$\rho_{opt}^2 = \frac{5\pi^{-4/3}}{96} \frac{\mathcal{M}^{5/3}}{d_L^2} \int_{20}^{1500} \frac{f^{-7/3}}{S_h(f)} df, \tag{8}$$

where the chirp-mass $\mathcal{M} = M\eta^{3/5}$ with the total mass $M = m_1 + m_2$ and the symmetric mass ratio $\eta = m_1 m_2 / M^2$. The luminosity distance is given by d_L and the detector noise spectral density is denoted by $S_h(f)$ for which we use the anticipated Advanced LIGO noise curve [39]. Under the signal model, we assume that the measured SNR-squared is drawn from a non-central χ^2 distribution with two degrees of freedom and a non-centrality parameter ρ_{opt}^2. The component masses are generated from the prior distributions governed by the global parameters μ_m, σ_m and are measured without noise. Under the noise model, we assume the measured SNR-squared is drawn from a central χ^2 distribution with two degrees of freedom and should be

interpreted as arising from detector noise. All data recorded above the SNR threshold due to the noise model will have corresponding component mass measurements drawn from a uniform distribution of masses within our prior search range defined as $\mathcal{S}_m \in (0.9M_\odot, 1.8M_\odot)$. Hence, a noise model trigger will be associated with any possible component mass configuration with equal probability, unlike the signal model events that follow the astrophysical prior distribution. In practice the distribution of inferred masses caused by the noise background will follow its own distribution which can be modelled or measured.

We assume a uniform distribution of sources in volume giving a prior distribution on the luminosity distance of $p(d_L|\boldsymbol{\gamma}, I) = 3d_L^2/(d_L^{max})^3$. We have used $d_L^{max} = 600$ Mpc, chosen to be beyond the distance at which BNS systems will be detectable, and ignored cosmological effects since the reach of Advanced LIGO is $z \sim 0.05$ for BNSs. We simulate 1 year of trials with a true astrophysical event rate (per unit volume per unit observation time) of $R = 10^{-7}$ Mpc^{-3} yr^{-1} with component mass distribution parameters $\mu_m = 1.2M_\odot$ and $\sigma_m = 0.1M_\odot$.

First consider the probability of obtaining a non-trigger per trial under the assumption of the signal model $\mathcal{H}^+$,

$$P(D^-|\boldsymbol{\gamma}, \mathcal{H}^+, I) = \int_0^{\rho_{th}^2} dx \int_{\mathcal{S}_m^2} d\mathbf{m}' \, p(x, \mathbf{m}'|\boldsymbol{\gamma}, \mathcal{H}^+, I)$$
$$= \int_0^{\rho_{th}^2} dx \, \frac{1}{2} \exp\left(-\frac{1}{2}\left[x + \rho_{opt}^2(\mathbf{m})\right]\right) I_0\left(x\rho_{opt}(\mathbf{m})\right), \qquad (9)$$

where I_0 is a modified Bessel function of the first kind and $x, \mathbf{m}'$ represents the unknown true squared SNR and unknown true component masses over which we marginalize. The complementary quantity, the probability density associated with detection for the signal model, for the jth trial is

$$p(\rho_j^2, \mathbf{m}_j|\boldsymbol{\gamma}, \mathcal{H}^+, I) = \int_{\mathcal{S}_m^2} d\mathbf{m}' \, p(\rho_j^2, \mathbf{m}_j, \mathbf{m}'|\boldsymbol{\gamma}, \mathcal{H}^+, I)$$
$$= \frac{1}{2} \exp\left\{-\frac{1}{2}\left(\rho_j^2 + \rho_{opt}^2(\mathbf{m}_j)\right)\right\} I_0\left(x\rho_{opt}(\mathbf{m}_j)\right), \qquad (10)$$

where we have used $p(\mathbf{m}|\mathbf{m}', \boldsymbol{\gamma}, \mathcal{H}^+, I) = \delta(m_1 - m_1')\delta(m_2 - m_2')$ since the mass measurements are noise free in this example. For non-trigger and trigger cases under the noise model we have

$$P(D^-|\boldsymbol{\gamma}, \mathcal{H}^-, I) = \int_0^{\rho_{th}^2} dx \int_{\mathcal{S}_m^2} d\mathbf{m}' \, p(x, \mathbf{m}'|\boldsymbol{\gamma}, \mathcal{H}^-, I)$$
$$= 1 - \exp\left(-\rho_{th}^2/2\right), \qquad (11)$$

$$p(\rho_j^2, \mathbf{m}_j|\boldsymbol{\gamma}, \mathcal{H}^-, I) = \int_{\mathcal{S}_m^2} d\mathbf{m}' \, p(\rho_j^2, \mathbf{m}_j, \mathbf{m}'|\boldsymbol{\gamma}, \mathcal{H}^-, I)$$
$$= \frac{1}{2\Delta m^2} \exp\left(-\rho_{th}^2/2\right), \qquad (12)$$

where $\Delta m = 0.9M_\odot$ is the prior range of component masses and we have used the noise model component mass prior distribution $p(\mathbf{m}'|\boldsymbol{\gamma}, \mathcal{H}^-, I) = 1/\Delta m^2$. This is also equal to the noise model component mass likelihood $p(\mathbf{m}|\mathbf{m}', \boldsymbol{\gamma}, \mathcal{H}^-, I)$. We define the model priors
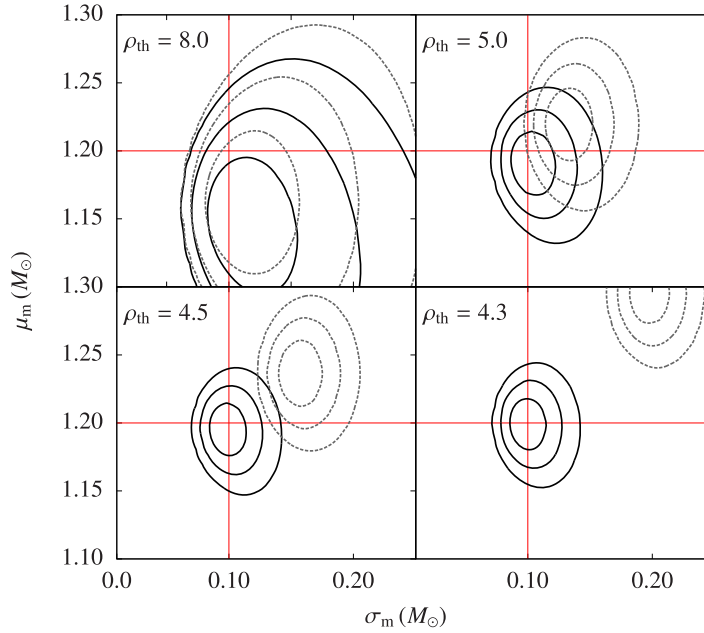
Figure 1. Joint probability contours on the combined global parameters μ_{m} and σ_{m} for four different threshold values $\rho_{\mathrm{th}} = 8, 5, 4.5, 4.3$. There were a total of 525 600 experiments of which 87 contained signals. For $\rho_{\mathrm{th}} = 8$ there were six triggers and all were signals. For $\rho_{\mathrm{th}} = 5$ there were 29 triggers of which 25 were signals and 4 were noise events. For $\rho_{\mathrm{th}} = 4.5$ there were 48 triggers of which 37 were signals and 11 were noise events, and finally, for $\rho_{\mathrm{th}} = 4.3$ there were 78 triggers of which 40 were signals and 38 were noise events. The true values of the mean and standard deviation of the mass are indicated by the solid red lines. Contours enclose 68, 95 and 99.7% of the probability. The black contours correspond to the application of our method accounting for both triggers and non-triggers and both the signal and noise models. The dashed grey contours represent results from an analysis accounting for triggers-only and assuming a signal model only.

assuming that Δt is sufficiently small that within a volume V, $P(\mathcal{H}^{+}|\boldsymbol{\gamma}, I) = RV\Delta t \ll 1$, so $P(\mathcal{H}^{-}|\boldsymbol{\gamma}, I) = 1 - RV\Delta t$.

In figure 1 we show the joint posterior probability on the global parameters μ_{m} and σ_{m} for a set of different SNR threshold values. The model can be made more complicated by including any number of extra parameters, global or local, e.g. cosmological parameters or the nuisance parameters of individual sources. As the threshold is decreased we increase the number of triggers due to signal events but also increase the number of false alarms (triggers due to noise events). For the case where $\rho_{\mathrm{th}} = 8$ the number of false alarms is zero but for $\rho_{\mathrm{th}} = 4.3$ the number of false alarms is nearly equal to the number of true alarms (see figure caption for details). The solid contours in figure 1 show the results of applying our method to the data where it can be seen that decreasing the threshold and correctly accounting for the increase in true and false alarms improves parameter estimation accuracy whilst remaining consistent with the correct parameter values.
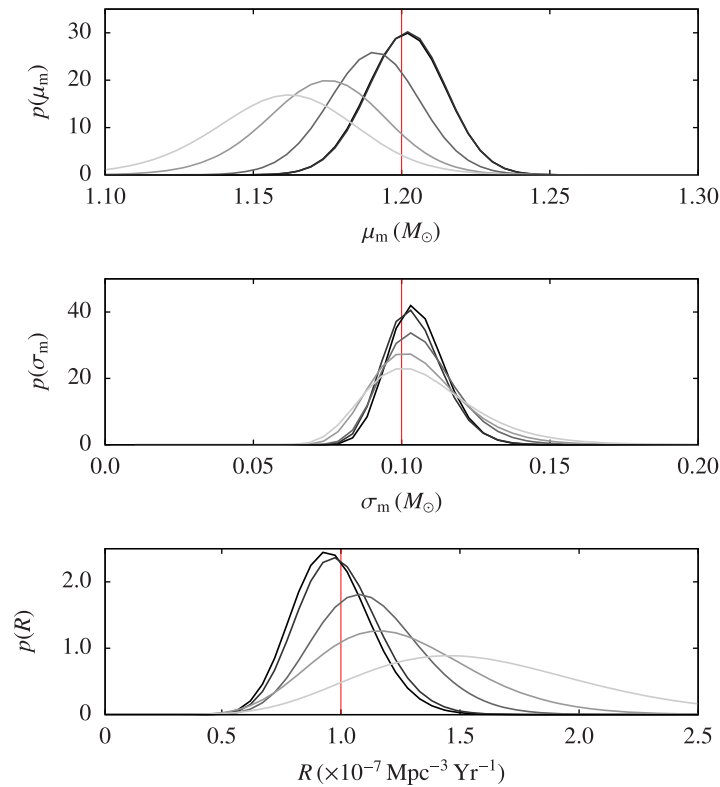
Figure 2. Marginalized posterior probability distributions on the global mass parameters $\mu_{\rm m}$, $\sigma_{\rm m}$ and the rate R for a range of different threshold values $\rho_{\rm th} = 7, 6, 5, 4, 3$. The fraction of noise triggers above threshold associated with these values are $0, 0, 0.138, 0.80, 0.988$ respectively. There were a total of $525\,600$ experiments of which 87 contained signals and the curves go from light grey to black as the threshold decreases. The true values of the global parameters are indicated by the solid vertical red lines.

An additional result is included in dashed grey lines showing the biases that can occur when failing to incorporate all relevant information into an analysis. This incorrect result is calculated by evaluating equation (10) for each trigger and therefore ignores the information contained within both the model prior and the non-triggers. Although it performs as expected when the threshold is high enough to allow no false alarms, the subsequent global parameter estimates become strongly biased for both mass parameters as the threshold decreases and the false alarm rate increases. This analysis assumes that the wider distribution of masses caused by the noise model reflects the real signals, leading to an overestimate of the mass standard deviation. The mean mass is also overestimated since the average mass measurement due to the noise model is equal to the centre of the mass prior range which in this case is higher than the true value. The specific behaviour depends on the nature of the signal, background, and the (incorrect) assumptions that are made, but this is illustrative of the approach implicitly taken by those studies which assume that a given set of GW triggers will all be due to the signal model.

We also show the precision of the global parameter estimates as a function of the threshold value. This is seen in figure 2 where the marginalized posterior distributions on the

global parameters including the rate R are plotted for a range of threshold values. We see that the posterior distributions become narrower as the threshold is decreased where more signals *and* noise events are recorded above threshold. We stress that this improvement in the precision of global parameter estimation is unbiased by the presence of false alarms. This is only true when triggers, non-triggers, signal model, and the noise model are correctly incorporated in the construction of the global parameter posteriors. It is also notable that the reduction in posterior width does not continue linearly with the reduction in threshold or with the corresponding increase in false alarms. In fact, further decreasing the threshold yields diminishing improvements, and the resulting distribution asymptotically approaches the optimal one which would be obtained by removing the threshold altogether. In that case the presence of noise still limits the precision attainable even though all the data is analysed and none discarded. This indicates that there exists a sensible choice of threshold value at which additional computation expense in processing triggers from lower thresholds returns little information. It also implies that operating at thresholds corresponding to high false alarm values will yield the most precise parameter estimates.

4. Discussion

We have derived a general formalism for analysing data which have been subjected to a threshold, which accounts for both the knowledge of and the ignorance of the actual result of an experiment. This formalism was applied to a simplified real-world example problem from the field of GW data analysis which will be encountered when sufficient detections have been made. The method succeeds for any number of false alarms and reduces to known formulae in limiting cases, such as zero false alarm rate [38]. We find that knowledge of the true rate or abundance of signals represents crucial information when making inferences with an ensemble of trials, and that it can therefore be inferred from them. Prior knowledge of signal and noise models must be incorporated, but measured values of efficiency and false alarm rates can be used when they are difficult to derive analytically.

Our work has implications for research which uses sets of GW signals to probe astrophysical, cosmological, or gravitational parameters, of the kind which have already been suggested in the literature [8–14]. We have shown that the best global parameter estimates will come from using the lowest threshold allowed by computational limitations, which will include using events which were (probably) generated by the background noise.

Our method could be applied in many other noise-limited fields of research, ranging from particle physics to cosmology.

IOP Institute of Physics Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

References

[1] Harry G M (the Scientific LIGO Collaboration) 2010 Advanced LIGO: the next generation of gravitational wave detectors *Class. Quantum Grav.* **27** 084006

[2] The Virgo Collaboration 2009 Advanced Virgo baseline design *Virgo Technical Report* VIR-0027A-09, Virgo (https://pub3.ego-gw.it/itf/tds/)

[3] Kuroda K (on behalf of the LCGT Collaboration) 2010 Status of LCGT *Class. Quantum Grav.* **27** 084004

[4] Aoudia S *et al* 2010 Einstein gravitational wave telescope conceptual design study *Virgo Technical Report* ET-0106A-10, Virgo (https://pub3.ego-gw.it/itf/tds/)

[5] Abadie J *et al* 2010 Predictions for the rates of compact binary coalescences observable by ground-based gravitational-wave detectors *Class. Quantum Grav.* **27** 173001

[6] Biswas R, Brady P R, Creighton J D E and Fairhurst S 2009 The loudest event statistic: general formulation, properties and applications *Class. Quantum Grav.* **26** 175009

[7] Mandel I 2010 Parameter estimation on gravitational waves from multiple coalescing binaries *Phys. Rev.* D **81** 084029

[8] Del Pozzo W 2012 Inference of cosmological parameters from gravitational waves: applications to second generation interferometers *Phys. Rev.* D **86** 043011

[9] Taylor S R, Gair J R and Mandel I 2012 Cosmology using advanced gravitational-wave detectors alone *Phys. Rev.* D **85** 023535

[10] Messenger C and Read J 2012 Measuring a cosmological distance–redshift relationship using only gravitational wave observations of binary neutron star coalescences *Phys. Rev. Lett.* **108** 091101

[11] Taylor S R and Gair J R 2012 Cosmology with the lights off: standard sirens in the Einstein telescope era *Phys. Rev.* D **86** 023502

[12] Del Pozzo W, Veitch J and Vecchio A 2011 Testing general relativity using Bayesian model selection: applications to observations of gravitational waves from compact binary systems *Phys. Rev.* D **83** 082002

[13] Li T G F, Del Pozzo W, Vitale S, Van Den Broeck C, Agathos M, Veitch J, Grover K, Sidery T, Sturani R and Vecchio A 2012 Towards a generic test of the strong field dynamics of general relativity using compact binary coalescence *Phys. Rev.* D **85** 082003

[14] Li T G F, Del Pozzo W, Vitale S, Van Den Broeck C, Agathos M, Veitch J, Grover K, Sidery T, Sturani R and Vecchio A 2012 Towards a generic test of the strong field dynamics of general relativity using compact binary coalescence: further investigations *J. Phys.: Conf. Ser.* **363** 012028

[15] Babak S *et al* 2013 Searching for gravitational waves from binary coalescence *Phys. Rev.* D **87** 024033

[16] Abbott B *et al* 2008 Search for gravitational waves from binary inspirals in s3 and s4 LIGO data *Phys. Rev.* D **77** 062002

[17] Hendry M A and Simmons J F L 1990 Bias of cosmological distance estimators *Astron. Astrophys.* **237** 275–82

[18] Loredo T J and Hendry M A 2010 *Bayesian Multilevel Modelling of Cosmological Populations* (Cambridge: Cambridge University Press) p 245

[19] Eddington A S 1913 On a formula for correcting statistics for the effects of a known error of observation *Mon. Not. R. Astron. Soc.* **73** 359–60

[20] Malmquist K G 1920 *Medd. Lund. Astron. Obs. Ser.* 2, No. 22

[21] Sir Eddington A S 1940 The correction of statistics for accidental error *Mon. Not. R. Astron. Soc.* **100** 354

[22] Lutz T E and Kelker D H 1973 On the use of trigonometric parallaxes for the calibration of luminosity systems: theory *Publ. Astron. Soc. Pac.* **85** 573

[23] Loredo T J 2007 Analyzing data from astronomical surveys: issues and directions *Statistical Challenges in Modern Astronomy IV (ASP Conf. Ser.* vol 371) p 121

[24] Hicken M, Michael Wood-Vasey W, Blondin S, Challis P, Jha S, Kelly P L, Rest A and Kirshner R P 2009 Improved dark energy constraints from 100 new CfA supernova type Ia light curves *Astrophys. J.* **700** 1097–140

[25] Udry S and Santos N C 2007 Statistical properties of exoplanets *Annu. Rev. Astron. Astrophys.* **45** 397–439

[26] Ageron M *et al* 2011 ANTARES: the first undersea neutrino telescope *Nucl. Instrum. Methods Phys. Res.* A **656** 11–38

[27] Abbasi R *et al* 2010 Calibration and characterization of the IceCube photomultiplier tube *Nucl. Instrum. Methods Phys. Res.* A **618** 139–52

[28] Allen S W, Evrard A E and Mantz A B 2011 Cosmological parameters from observations of galaxy clusters *Annu. Rev. Astron. Astrophys.* **49** 409–70

[29] Cunha C 2009 Cross-calibration of cluster mass observables *Phys. Rev.* D **79** 063009

[30] Landy S D and Szalay A S 1992 A general analytical solution to the problem of Malmquist bias due to lognormal distance errors *Astrophys. J.* **391** 494–501

[31] Mantz A, Allen S W, Rapetti D and Ebeling H 2010 The observed growth of massive galaxy clusters: I. Statistical methods and cosmological constraints *Mon. Not. R. Astron. Soc.* **406** 1759–72

[32] Hendry M A and Simmons J F L 1994 Optimal galaxy distance estimators *Astrophys. J.* **435** 515–27

[33] Willick J A and Strauss M A 1998 Maximum likelihood comparison of Tully–Fisher and redshift data: II. Results from an expanded sample *Astrophys. J.* **507** 64–83

[34] Lynden-Bell D, Faber S M, Burstein D, Davies R L, Dressler A, Terlevich R J and Wegner G 1988 Spectroscopy and photometry of elliptical galaxies: V. Galaxy streaming toward the new supergalactic center *Astrophys. J.* **326** 19–49

[35] Mevius M, Buitink S, Falcke H, Hörandel J, James C W, McFadden R, Scholten O, Singh K, Stappers B and Ter Veen S 2012 Detecting ultra high energy neutrinos with LOFAR *Nucl. Instrum. Methods Phys. Res.* A **662** 26

[36] Adrián-Martínez S *et al* 2012 Search for neutrino emission from Gamma-ray flaring Blazars with the ANTARES Telescope *Astropart. Phys.* **36** 204–10

[37] Abadie J *et al* 2010 All-sky search for gravitational-wave bursts in the first joint LIGO-GEO-Virgo run *Phys. Rev.* D **81** 102001

[38] Kelly B C 2007 Some aspects of measurement error in linear regression of astronomical data *Astrophys. J.* **665** 1489–506

[39] Cutler C and Thorne K S 2002 *An Overview of Gravitational-wave Sources* (*General Relativity and Gravitation*) ed N T Bishop and S Maharaj (Hackensack, NJ: World Scientific) pp 72–112