PBH notes for Sumedha

(Dated: August 27, 2024)

I. PARAMETER ESTIMATION WITH A SINGLE SIGNAL

In typical parameter estimation scenarios, you consider four elements only: The model evidence, the posterior, the prior, and the likelihood.

The model evidence answers the question ”how likely is it that this model is correct,” and it is quantified typically through

$$\mathcal{Z} = \int \mathcal{L}(\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}\,, \tag{1}$$

where $\mathcal{L}(\boldsymbol{\theta})$ is the gravitational-wave likelihood, which answers ”how likely is there a gravitational wave in the data after the measurement”, $p(\boldsymbol{\theta})$ is the prior, which answers ”what do we know about the gravitational-wave/black hole population parameters before the measurement,” and $\boldsymbol{\theta}$ are the gravitational-wave parameters. BILBY gives the evidence $\mathcal{Z}$ after the run.

The likelihood includes both the detector noise and the gravitational-wave signal, and therefore does not depend on the population properties of binary black holes. So it is unchanged between population models (e.g. primordial black hole population vs stellar black hole population). So you don't need to care about changing it.

However, the prior $p(\boldsymbol{\theta})$ does change between the models, since it describes the prior knowledge (what we know about the gravitational-wave parameters before the measurement).

Therefore, if there were two priors ($p_A(\boldsymbol{\theta})$ and $p_B(\boldsymbol{\theta})$), then we could write their respective evidences as follows

$$\mathcal{Z}_A = \int \mathcal{L}(\boldsymbol{\theta}) p_A(\boldsymbol{\theta}) d\boldsymbol{\theta}\,, \tag{2}$$

$$\mathcal{Z}_B = \int \mathcal{L}(\boldsymbol{\theta}) p_B(\boldsymbol{\theta}) d\boldsymbol{\theta}\,, \tag{3}$$

The ”Bayes factor” then answers which of these priors is the correct one

$$\mathcal{B}^B_A = \frac{\mathcal{Z}_B}{\mathcal{Z}_A}\,. \tag{4}$$

If it is a large number, then the prior 2 is correct. If it is a small number, then prior 1 is correct. Alternatively, we could say that the model 2 is correct/incorrect based on the Bayes factor.

In this example, the $p_B(\boldsymbol{\theta})$ could be the primordial black hole prior, and the $p_A(\boldsymbol{\theta})$ could be the stellar black hole prior.

A. Solving the evidence with bilby

Let us assume that the BILBY run has used some prior $p(\boldsymbol{\theta})$, and we want to compare if a given gravitational-wave is more likely from a primordial black hole population with prior $p_B(\boldsymbol{\theta})$, or from a stellar black hole population with prior $p_A(\boldsymbol{\theta})$.

The corresponding evidences for the primordial/stellar models are:

$$\mathcal{Z}_A = \int \mathcal{L}(\boldsymbol{\theta}) p_A(\boldsymbol{\theta}) d\boldsymbol{\theta}\,, \tag{5}$$

$$\mathcal{Z}_B = \int \mathcal{L}(\boldsymbol{\theta}) p_B(\boldsymbol{\theta}) d\boldsymbol{\theta}\,. \tag{6}$$

However, the issue with evaluating these evidence is that the BILBY run only evaluated

$$\mathcal{Z} = \int \mathcal{L}(\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}\,, \tag{7}$$

with prior $p(\boldsymbol{\theta})$, which is different from either of the models we're interested in (i.e. $p(\boldsymbol{\theta}) \neq p_A(\boldsymbol{\theta}) \neq p_A(\boldsymbol{\theta})$).

So the reason you do all of the trickery (as in below and as in Eric Thrane's paper) is not because of any theoretical reason, but due to computational reasons (you could re-run BILBY for all runs with different priors to get $\mathcal{Z}_A$ and $\mathcal{Z}_B$ and waste a lot of computational resources but get the same result; here we need to do only one run).

This means that we need to change the form of the evidence to compute it:

$$\begin{aligned}
\mathcal{Z}_B &= \int \mathcal{L}(\boldsymbol{\theta}) p_B(\boldsymbol{\theta}) d\boldsymbol{\theta}\,, \\
&= \int \mathcal{L}(\boldsymbol{\theta}) p(\boldsymbol{\theta}) \frac{p_B(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} d\boldsymbol{\theta}\,, \\
&= \mathcal{Z} \int p(\boldsymbol{\theta}|d) \frac{p_B(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} d\boldsymbol{\theta}\,, \\
&= \mathcal{Z} \left\langle \frac{p_B(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} \right\rangle_{p(\boldsymbol{\theta}|d)}\,,
\end{aligned} \tag{8}$$

where in the second-to-last line we used the Bayes theorem:

$$p(\boldsymbol{\theta}|d) = \frac{\mathcal{L}(\boldsymbol{\theta}) p(\boldsymbol{\theta})}{\mathcal{Z}}\,, \tag{9}$$

to compute the **posterior** of the measurement (what we know about the gravitational-wave parameters after the measurement) which includes the gravitational-wave detecor data d. The posterior $p(\boldsymbol{\theta}|d)$ is given by the BILBY run.

In the last line we solve the integral by *importance sampling* the posterior.

B. Example

Let us assume that the only difference between the three priors is in their mass estimate (in reality you would use whatever population model you have):

$$
\begin{aligned}
p(m_A) &= \text{constant}\,,\\
p_A(m_1) &= C_A m_1^{-2.35}\,,\\
p_B(m_1) &= C_B m_1^{-1}\,,
\end{aligned}
\tag{10}
$$

i.e., have different mass slopes but all else is the same. The $C_A = 1/\int p_A(m_1)dm_1$ and $C_B = 1/\int p_B(m_1)dm_1$ are *normalization constants* (the integral of all probability functions must add to 1).

Then, the respective evidences can be solved by

$$
\begin{aligned}
\mathcal{Z}_B &= \mathcal{Z}\left\langle \frac{p_B(\boldsymbol{\theta})}{p(\boldsymbol{\theta})}\right\rangle_{p(\boldsymbol{\theta}|d)}\,,\\
&= \mathcal{Z}\left\langle \frac{C_B m_1^{-1}}{\text{constant}}\right\rangle_{p(m_1|d)}
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
\mathcal{Z}_A &= \mathcal{Z}\left\langle \frac{p_A(\boldsymbol{\theta})}{p(\boldsymbol{\theta})}\right\rangle_{p(\boldsymbol{\theta}|d)}\,,\\
&= \mathcal{Z}\left\langle \frac{C_A m_1^{-2.35}}{\text{constant}}\right\rangle_{p(m_1|d)}
\end{aligned}
\tag{12}
$$

Both the evidence $\mathcal{Z}$ and the posterior $p(m_1|d)$ is given by BILBY.

Our main question is: which model is better at describing the data? Primordial or stellar black holes? To answer this, we can divide the evidences to compute the Bayes factor

$$
\begin{aligned}
\mathcal{B}_A^B &= \frac{\mathcal{Z}_B}{\mathcal{Z}_A}\,,\\
&= \left\langle \frac{C_B m_1^{-1}}{\text{constant}}\right\rangle_{p(m_1|d)} \Bigg/ \left\langle \frac{C_A m_1^{-2.35}}{\text{constant}}\right\rangle_{p(m_1|d)}
\end{aligned}
\tag{13}
$$

If, e.g., the $\mathcal{B}_A^B = 5$, then it is five times more likely that the signal came from primordial black holes, and not from stellar-mass black holes.

II. COMBINING THE EVIDENCES

Let us then assume that we have detect N gravitational-wave events with data d_i, and not only one. The evidence for each of the event is

$$
\mathcal{Z}^i = \int \mathcal{L}_i(\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}\,,
\tag{14}
$$

The *combined* evidence is the product of the individual evidences

$$
\mathcal{Z} = \prod_{i=1}^{N} \mathcal{Z}^i\,.
\tag{15}
$$

A. Solving the evidence with bilby

Let us assume that the BILBY run has used some prior $p(\boldsymbol{\theta})$ for a set of N events. We want to compare if the population of gravitational wave events is more likely from a primordial black hole population with prior $p_B(\boldsymbol{\theta})$, or from a stellar black hole population with prior $p_A(\boldsymbol{\theta})$.

We already know the individual evidences, so we can simply compute the combined evidence by using the already derived formulas

$$
\mathcal{Z}_A = \prod_{i=1}^{N} \mathcal{Z}^i \left\langle \frac{C_A m_1^{-2.35}}{\text{constant}}\right\rangle_{p(m_1|d_i)}\,,
\tag{16}
$$

$$
\mathcal{Z}_B = \prod_{i=1}^{N} \mathcal{Z}^i \left\langle \frac{C_B m_1^{-1}}{\text{constant}}\right\rangle_{p(m_1|d_i)}\,,
\tag{17}
$$

where the only change (with respect to individual evidences) is that we are now importance sampling the posterior of the ith event $p(m_1|d_i)$ and take the product of all individual evidences $\mathcal{Z}_i$.

The combined Bayes factor that answers which model is correct is

$$
\mathcal{B}_A^B = \prod_{i=1}^{N}\left[\left\langle \frac{C_B m_1^{-1}}{\text{constant}}\right\rangle_{p(m_1|d_i)} \Bigg/ \left\langle \frac{C_A m_1^{-2.35}}{\text{constant}}\right\rangle_{p(m_1|d_i)}\right]
\tag{18}
$$

If, e.g., the $\mathcal{B}_A^B = 5$, then it is five times more likely that the signal came from primordial black holes, and not from stellar-mass black holes.

III. EXAMPLE: ADDING "POPULATION PARAMETERS"

If there are population parameters (fancy name: hyperparameters) in the model, then we need to re-compute the evidence in a slightly different way. A population parameter describes the population.

For example, "fraction of primordial black holes f" is a population parameters, which describes how many primordial black hole binaries there are in the Universe relative to stellar black hole binaries.

We know that stellar black holes exist, so in this hypothetical example what we want to compare is: "do the GWs come from stellar black holes + primordial black holes, or only primordial black holes?"

The priors become:

$$
\begin{aligned}
p(m_A) &= \text{constant}\,,\\
p_A(m_1) &= C_A m_1^{-2.35}\,,\\
p_B(m_1|f) &= C_B(f m_1^{-1} + (1-f)m_1^{-2.35})\,,
\end{aligned}
\tag{19}
$$

where now the primordial black hole model (model B) has a component for the primordial black holes (f), and a component for the stellar black holes ($1-f$). We call $p_B(m_1|f)$ the prior *for given fraction/population parameter* f.

However, we don't know what the fraction f is. So we typically do not want to consider any analysis for "given f", but for "all possible f". To get the latter, we need to marginalize over all possible fractions, and we need to have a "guess" for what the fraction should be.

An agnostic guess might be that f can be distributed uniformly anywhere between $[0, 1]$:

$$p(f) = \text{constant}\,. \tag{20}$$

Perhaps a better guess would be that f is more likely lower than larger, so we could alternatively guess that

$$p(f) \propto (1+f)^{-1}\,. \tag{21}$$

Then, the evidence becomes (same as in Eric Thrane's paper)

$$\mathcal{Z}_B = \prod_{i=1}^{N} \int \mathcal{L}_i(\boldsymbol{\theta}) p_B(\boldsymbol{\theta}|f) p(f) d\boldsymbol{\theta} df\,. \tag{22}$$

$$\mathcal{Z}_B = \prod_{i=1}^{N} \int \mathcal{L}_i(\boldsymbol{\theta}) p(\boldsymbol{\theta}) \frac{p_B(\boldsymbol{\theta}|f)}{p(\boldsymbol{\theta})} p(f) d\boldsymbol{\theta} df\,. \tag{23}$$

$$\mathcal{Z}_B = \prod_{i=1}^{N} \int \frac{p_B(\boldsymbol{\theta}|f)}{p(\boldsymbol{\theta})} p(\boldsymbol{\theta}|d) p(f) d\boldsymbol{\theta} df\,. \tag{24}$$

$$\mathcal{Z}_B = \prod_{i=1}^{N} \int \frac{p_B(\boldsymbol{\theta}|f)}{p(\boldsymbol{\theta})} g(\boldsymbol{\theta}, f) d\boldsymbol{\theta} df\,. \tag{25}$$

After a bit of algebra, we arrive at the Bayes factor :

$$\mathcal{B}_A^B = \prod_{i=1}^{N} \left\langle \frac{C_B(f m_1^{-1} + (1-f) m_1^{-2.35})}{\text{constant}} \right\rangle_{p(m_1|d_i), p(f)} \Bigg/ \left\langle \frac{C_A m_1^{-2.35}}{\text{constant}} \right\rangle_{p(m_1|d_i)}, \tag{26}$$

where we now take the average over the posterior $p(m_1|d_i)$ *and* the distribution of $p(f)$ samples.

Whether or not we need "population parameters" depends on how much we know about the model. The more tuneable (unknown) parameters there are, the more population parameters we typically need. If the model is fixed (assumed to have no unknown parameters), we do not need population parameters and can compute everything as in the previous two sections. Personally, I do not know if there are any population parameters for primordial black holes.

I would go through this whole process step-by-step: Do section I for a single bilby run. Once that works, do Section II for all bilby runs. Once that works, do section III for all bilby runs.

IV. SELECTION EFFECTS

Technically, we could also include selection effects in the analysis. They come into play when there is a large chunk of data, and stem from the fact that some events are more detectable than others. To include selection effects, we can include the $\alpha(f)$ term in the evidence:

$$\mathcal{Z}_B = \prod_{i=1}^{N} \int \frac{1}{\alpha_B(f)} \mathcal{L}_i(\boldsymbol{\theta}) p_B(\boldsymbol{\theta}|f) p(f) d\boldsymbol{\theta} df\,, \tag{27}$$

where

$$\alpha_B(f) = \int p(\det|\vec{\theta}) p(\vec{\theta}|f, \mathcal{H}_B) d\vec{\theta}, \tag{28}$$

with $p(\det|\vec{\theta})$ being the detection probability for parameters $\vec{\theta}$. The $p(\det|\vec{\theta})$ can be retrieved by using the GWDET package `https://github.com/dgerosa/gwdet`.

Similarly, the evidence for the A hypothesis without hyperparameters becomes

$$\mathcal{Z}_A = \prod_{i=1}^{N} \frac{1}{\alpha_A} \int \mathcal{L}_i(\boldsymbol{\theta}) p_B(\boldsymbol{\theta}|f) p(f) d\boldsymbol{\theta} df\,, \tag{29}$$

where

$$\alpha_A = \int p(\det|\vec{\theta}) p(\vec{\theta}|\mathcal{H}_A) d\vec{\theta}. \tag{30}$$

The Bayes factor is thus modified so that

$$\mathcal{B}_A^B = \prod_{i=1}^{N} \left\langle \frac{1}{\alpha_B(f)} \frac{C_B(f m_1^{-1} + (1-f) m_1^{-2.35})}{\text{constant}} \right\rangle_{p(m_1|d_i), p(f)} \Bigg/ \left\langle \frac{1}{\alpha_A} \frac{C_A m_1^{-2.35}}{\text{constant}} \right\rangle_{p(m_1|d_i)}, \tag{31}$$

V.