

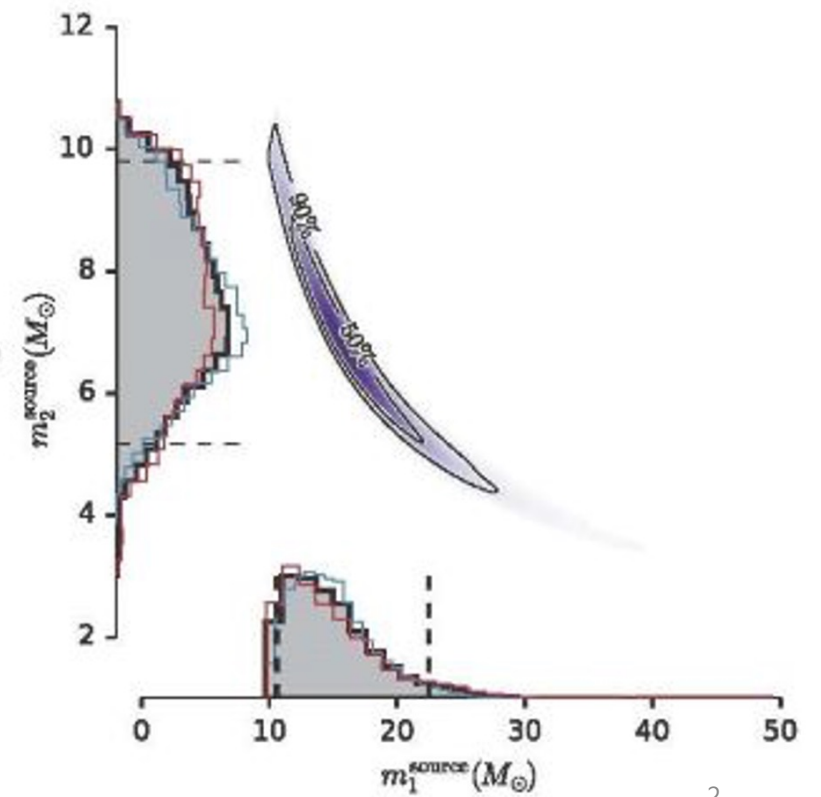
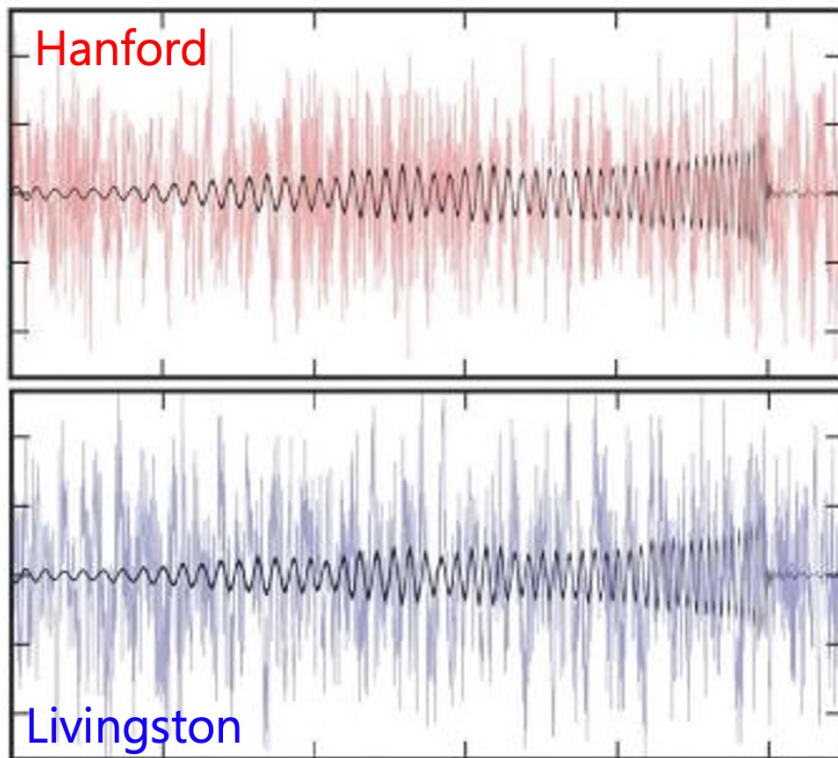
# 중력파 모수추정

JeongCho Kim(SNU)

2024 GWNR Winter School

2024.01.31

# Parameter Estimation (PE)



# Frequentist vs Bayesian

## Frequentist Statistics

- all about probability in the long run
- frequentist analyses base their results only on the data they collect

## Bayesian Statistics

- describes a method to update probabilities based on data and past knowledge
- parameters and hypotheses are seen as probability distributions and the data as fixed

The prior is one of the key differences between frequentist and Bayesian inference

# Bayesian inference

Bayesian inference is a way of making statistical inferences in which the statistician assigns prior distribution to the distributions that could generate the data.

After the data is observed, Bayes' rule is used to update the prior, that is, posterior distribution assigned to the possible data generating distributions.

- We observe some data (a sample), that we collect in a vector  $x$ .
- We regard  $x$  as the realization of a random vector  $X$ .
- We do not know the probability distribution of  $X$ .
- We define a statistical model, that is a set  $\theta$  of probability distributions that could have generated the data.
- We parametrize the model, that is, we put the elements of  $\theta$  in correspondence with a set of real vectors called parameters
- we use the sample and the statistical model to make an inference about the unknown data generating distribution

# Bayes' rule

It is a rule for computing conditional probabilities.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Let be two events, A and B.

Denote it's probabilities by P(A) and P(B).

Suppose the both  $P(A) > 0$  and  $P(B)$ .

# Bayes' rule – proof

conditional probability formula :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(B|A) = \frac{P(A \cap B)}{P(A)}$$

The second formula re-arranged as follows :

$$P(B|A)P(A) = P(A \cap B)$$

It is plugged into the first formula, we obtain :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

# Bayes theorem

Initial Understanding + New Observation = Updated Understanding

D : the observed data  
 $\theta$  : the model parameters

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(\theta)P(D|\theta) d\theta}$$

Posterior = Likelihood \* Prior

# The likelihood

The first building block of a parametric Bayesian model is the likelihood  $p(D|\theta)$

Suppose that the sample is a vector of  $n$  independent and identically distributed draws  $x_1, \dots, x_n$  from a normal distribution.

$$x = [x_1 \cdots x_n]$$

The mean ( $\mu$ ) of the distribution is unknown, while its variance  $\sigma^2$  is known. These are the two parameters of the model.

The Probability density function of a generic draw  $x_i$  is  $P(x_i|\mu) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2} \frac{(x_i-\mu)^2}{\sigma^2}\right)$ ,

where we use the notation  $P(x_i|\mu)$  to highlight the fact that  $\mu$  is unknown and the density of  $x_i$  depends on this unknown parameter.

Because the observations  $x_1 \cdots x_n$  are independent, we can write the likelihood as

$$P(x|\mu) = \prod_{i=1}^n P(x_i|\mu) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right),$$



# The likelihood

$p(D|\theta)$  : The probability that the gravitational wave generated by the model parameter  $\theta$  produces the signal  $D$ .

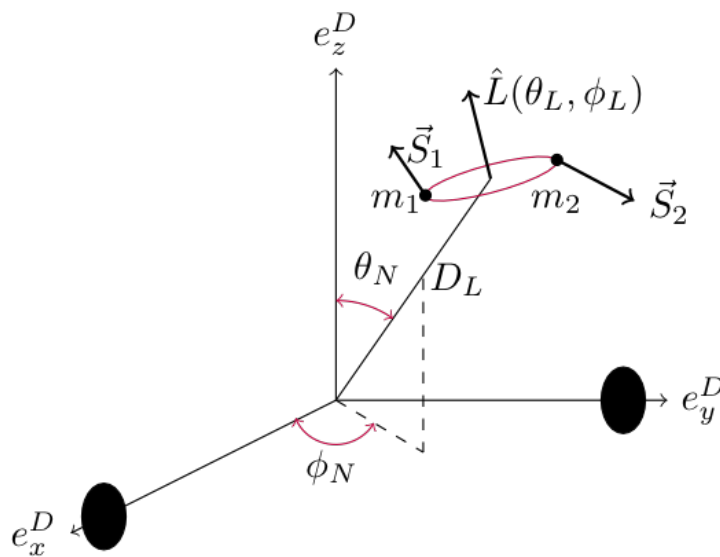
$$P(D|\theta) = \prod_{i=1}^n P(d_i|\theta) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (d_i - \theta)^2\right) \propto e^{-\langle D-h|D-h \rangle/2}$$

$$\langle a|b \rangle \equiv 4\Re \int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S(f)} df$$

$$\langle a|b \rangle = 2\Re \int_{-\infty}^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S(|f|)} df = \int_{-\infty}^\infty \frac{\tilde{a}(f)\tilde{b}^*(f) + \tilde{a}^*(f)\tilde{b}(f)}{S(|f|)} df$$

# The prior

The prior is the subjective probability density assigned to the parameter  $\theta$ .



$$p(h') = \delta(h' - h'(\vec{\theta}))p(\vec{\theta})$$

priors on the system parameters

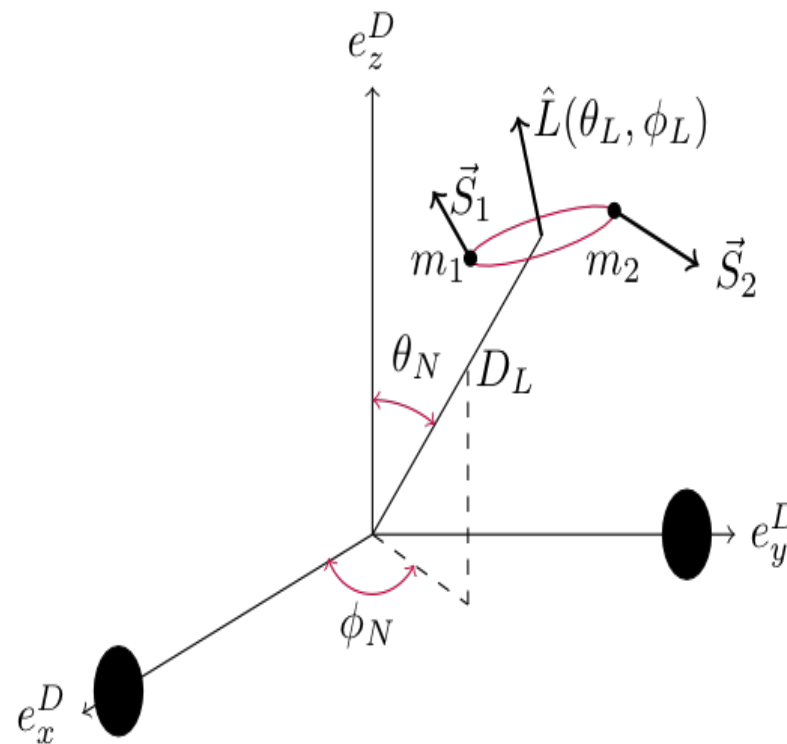
What we know about the parameters of the system *before* obtaining the data

# Intrinsic & extrinsic parameters

$m_1$ $m_2$	Uniform in some range
----------------	--------------------------

$\vec{S}_2$ $\vec{S}_1$	Uniform in direction and magnitude in $[0, m_i^2]$
----------------------------	-------------------------------------------------------------

$\lambda_1$ $\lambda_2$	Uniform in (0,5000)
----------------------------	------------------------



$D_L$	Uniform in volume
-------	----------------------

$\theta_N$ $\phi_N$	Uniform in the sky
------------------------	-----------------------

$\theta_L$ $\phi_L$	Uniform in direction
------------------------	-------------------------

# The posterior

The posterior gives the probability density that a model describes the data.

$P(\theta|D)$  : The posterior probability is the conditional probability  $P(\theta|D)$  , calculated after receiving the information that the event D has happened.

The marginal distribution  $f(x)$  is derived from the prior and the likelihood.

We first derive the joint distribution  $f(x, \theta) = f(x|\theta)f(\theta)$  and then we marginalize it to obtain the posterior.

In the continuous case, the marginal is computed by integration as  $f(x) = \int f(x, \theta)d\theta$ .

In the discrete case, it is derived by calculating a sum as  $f(x) = \sum_{\theta} f(x, \theta)$ .

→ there are numerical methods that allow us to draw MCMC samples from the posterior distribution.

# Probability density function (PDF)

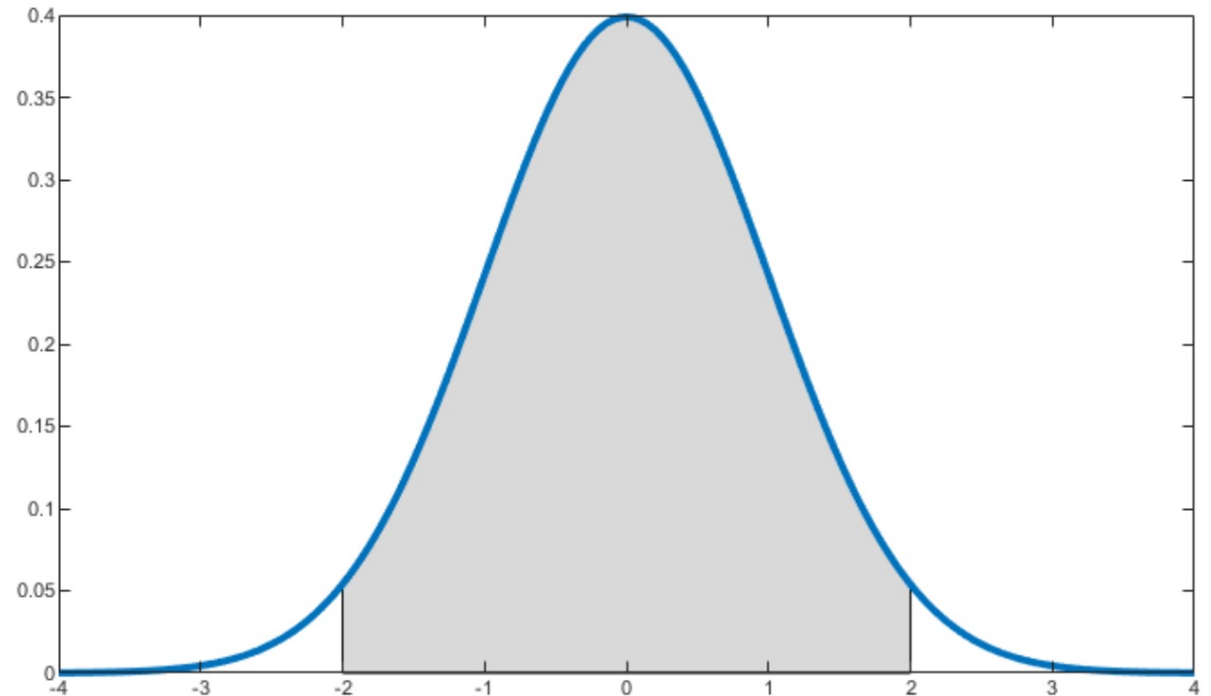
The probability density function is a function that completely characterizes the distribution of a continuous random variable.

The PDF of a continuous random variable  $X$  is a function  $PDF_X: \mathbb{R} \rightarrow [0, \infty)$  such that

$$P(X \in [a, b]) = \int_a^b PDF_X(x) dx$$

for any interval  $[a, b] \subseteq \mathbb{R}$ .

The set of values  $x$  for which  $PDF_X(x) > 0$  is called the support of  $X$ .



# Markov Chain Monte Carlo (MCMC)

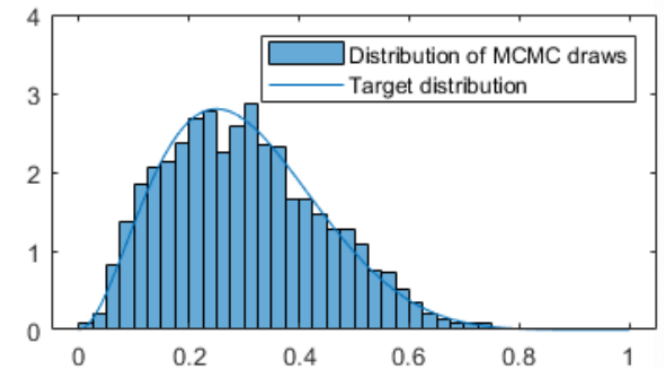
Markov Chain Monte Carlo (MCMC) methods are very powerful Monte Carlo methods that are often used in Bayesian inference.

The classical Monte Carlo methods rely on computer-generated samples made up of independent observations.

MCMC methods are used to generate sequences of dependent observations. These sequences are Markov chains.

- MCMC methods work like standard Monte Carlo methods, although with a twist: the computer-generated draws  $x_1, \dots, x_2$  are not independent, but they are serially correlated.
- They are the realizations of  $T$  random variables  $X, \dots, X_T$  that form a Markov Chain.

# Markov property



A random sequence  $(X_t)$  is a Markov chain if and only if, given the current value  $X_t$ , the future observation  $X_{t+n}$  are conditionally independent of the past values  $X_{t-k}$ , for any positive integers  $k$  and  $n$ .

$$P(X_{t+n} = x | X_t, X_{t-1}, \dots, X_{t-k}) = P(X_{t+n} = x | X_t)$$

This property, known as the Markov property, The probability distribution of the future values of the chain depends only on its current value  $X_t$ , regardless of how the value was reached regardless of the path followed by the chain in the past

Although this is not true in general of any Markov chain, the chains generated by MCMC method have the following property :

Two variables  $X_t$  and  $X_{t+n}$  are not independent, but they become closer and closer to being independent as  $n$  increases.

This property implies that  $f(X_{t+n} | X_t)$  converges to  $f(X_{t+n})$  as  $n$  becomes large.

# Bayes theorem

Initial Understanding + New Observation = Updated Understanding

D : the observed data  
 $\theta$  : the model parameters

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(\theta)P(D|\theta) d\theta}$$



# The evidence

Normalization factor for parameter estimation

Important for model selection

Compare competing models, for example  
'GW170817 was a BNS' vs 'GW170817 was a BBH'

$$\mathcal{O}_{ij} = \frac{p(M_i|d)}{p(M_j|d)}$$

$$= \frac{p(M_i)p(d|M_i)}{p(M_j)p(d|M_j)}$$

Bayes Factor  
Likelihood of the models  
Ratio of the evidences

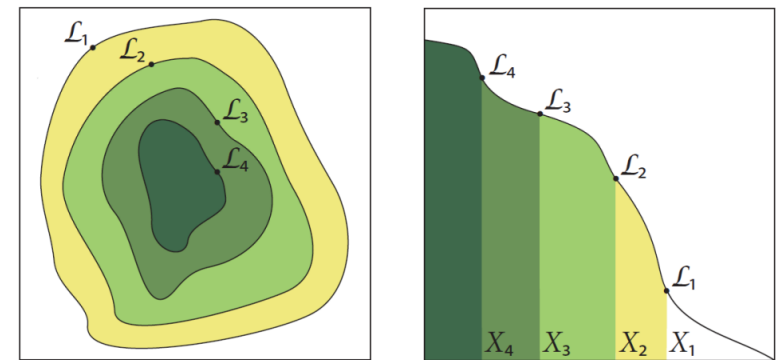
# Nested Sampling

Nested sampling estimates directly how the likelihood function relates to prior mass. The evidence (alternatively the marginal likelihood, marginal density of the data, or the prior predictive) is immediately obtained by summation.

$$Z = \text{evidence} = \int L dX$$

$L = L(\theta)$  is the likelihood function.

$dX = \pi(\theta)d\theta$  is the element of prior mass and  $\theta$  represents the unknown parameters.



Feroz et al. (2013)

## Setting the priors

The model describing the black hole merger depends on 15 parameters. Running a sampler with this many parameters can take half a day to run. To save some time, I'll fix all but 4 parameters: the chirp mass, the mass ratio, the phase, and the geocent time.

The chirp mass and mass ratio are two parameters which depend only on the mass of the two black holes that caused the gravitational waves. When these two parameters are derived, we can use them to calculate the masses of the black holes. Since we don't know much about the system parameters that we're allowing to vary, uniform priors are most appropriate:

```
prior = bilby.core.prior.PriorDict()
# uniform priors for variable parameters
prior['chirp_mass'] = Uniform(name='chirp_mass', minimum=10.0, maximum=100.0)
prior['mass_ratio'] = Uniform(name='mass_ratio', minimum=0.5, maximum=1)
prior['phase'] = Uniform(name="phase", minimum=0, maximum=2*np.pi)
prior['geocent_time'] = Uniform(name="geocent_time", minimum=event_start_time-0.1,
                                maximum=event_start_time+0.1)

# fixed values for all other parameters
prior['a_1'] = 0.0
prior['a_2'] = 0.0
prior['tilt_1'] = 0.0
prior['tilt_2'] = 0.0
prior['phi_12'] = 0.0
prior['phi_jl'] = 0.0
prior['dec'] = -1.2232
prior['ra'] = 2.19432
prior['theta_jn'] = 1.89694
prior['psi'] = 0.532268
prior['luminosity_distance'] = 412.066
```

## Setting the likelihood

Bilby comes with in-built likelihood functions. To use them, first a waveform has to be generated using the properties of the data, like duration and sampling frequency. We also need to combine the interferometers into a list, so that both will be used at once when it comes to sampling.

```
interferometers = [H1, L1]

waveform_arguments = dict(waveform_approximant='IMRPhenomPv2',
                          reference_frequency=50., minimum_frequency=20.)

# set sampling frequency of data
sampling_frequency = 2048.

# generate waveforms
waveform_generator = bilby.gw.WaveformGenerator(
    duration=duration, sampling_frequency=sampling_frequency,
    frequency_domain_source_model=bilby.gw.source.lal_binary_black_hole,
    parameter_conversion=bilby.gw.conversion.convert_to_lal_binary_black_hole_pa
    waveform_arguments=waveform_arguments)

likelihood = bilby.gw.GravitationalWaveTransient(
    interferometers=interferometers, waveform_generator=waveform_generator)
```

## Sampling the data

Now that the likelihood and prior functions are set up, the sampler can now be implemented to derive some of the properties of the binary black hole system. This process is intensive, and takes a long time (1 hour 15 minutes) to run on my machine.

The sampler I chose to use is "dynesty", which uses nested sampling to estimate the model parameters. For a more in-depth work through of dynesty, you can read through an example of [gamma-ray spectroscopy using dynesty](#) on this site.

Using bilby, we can run the sampler. First, we have to set the number of live points, and the stopping criterion as hyperparameters. For neatness, I've removed the log that bilby outputs whilst running. Instead, I'll just show the summary that is outputted by bilby once the sampler has finished iterating. This summary contains information on the log evidence, and the Bayes factor of the model.

```
nlive = 1000          # live points
stop = 0.1           # stopping criterion
method = "unif"      # method of sampling
sampler = "dynesty"  # sampler to use

result = bilby.run_sampler(
    likelihood, prior, sampler=sampler, outdir=outdir, label=label,
    conversion_function=bilby.gw.conversion.generate_all_bbh_parameters,
    sample=method, nlive=nlive, dlogz=stop)
```

# GW - Q.4

File Edit View Run Kernel Tabs Settings Help

OPEN TABS [Close All](#)

- Untitled1.ipynb
- nrgw20@olaf-c197:~

KERNELS [Shut Down All](#)

File Edit View Run Kernel Tabs Settings Help

OPEN TABS [Close All](#)

- Untitled1.ipynb
- nrgw20@olaf-c197:~

KERNELS [Shut Down All](#)

TERMINALS [Shut Down All](#)

- terminals/1

```
(base) [nrgw20@olaf-c197 ~]$ ls
H1_GW_NR2024_PROBLEM2.gwf  jhub_9226181.log  jhub_9226190.log  jhub_9251470.log  Untitled1.ipynb
jhub_9226144.log          jhub_9226188.log  jhub_9226208.log  pub               Untitled.ipynb
(base) [nrgw20@olaf-c197 ~]$

(base) [nrgw20@olaf-c197 ~]$ ls
H1_GW_NR2024_PROBLEM2.gwf  jhub_9226181.log  jhub_9226190.log  jhub_9251470.log  Untitled1.ipynb
jhub_9226144.log          jhub_9226188.log  jhub_9226208.log  pub               Untitled.ipynb
(base) [nrgw20@olaf-c197 ~]$ conda deactivate
[nrgw20@olaf-c197 ~]$
```

```
$conda deactivate
$conda env list
$conda activate igwn-py39
```

```
$nohup python *.py
```

```
$tail -f nohup.out
```

```
(igwn-py39) [nrgw20@olaf-c197 ~]$ nohup: ignoring input and appending output to 'nohup.out'
```

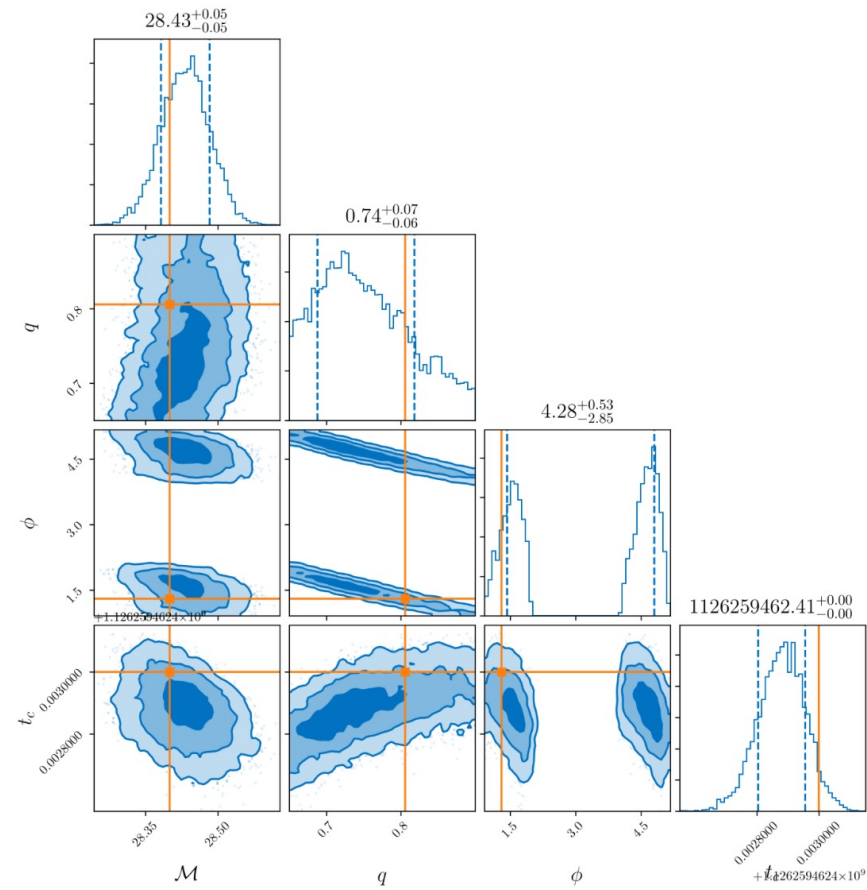
```
(igwn-py39) [nrgw20@olaf-c197 ~]$ tail -f nohup.out
```



## Plotting the posterior

Eventually, the sampling will be complete. Bilby creates a few useful plots of the posteriors, along with saving samples of the posteriors for us to further analyse. Lets start by plotting a corner plot, which nicely shows the posteriors for all the parameters, as well as contour plots showing how one parameter varies with any other.

```
result.plot_corner()
```





# GW - Q.4

```
Mc = result_short.posterior["chirp_mass"].values
```

We can then get some useful quantities such as the 90% credible interval

```
lower_bound = np.quantile(Mc, 0.05)
upper_bound = np.quantile(Mc, 0.95)
median = np.quantile(Mc, 0.5)
print("Mc = {} with a 90% C.I = {} -> {}".format(median, lower_bound, upper_bound))
```

## [문제4] 주어진 중력파 자료파일에서 bilby

주어진 중력파 자료 파일(H1\_GW\_NR2024\_PR

L1\_GW\_NR2024\_PROBLEM3.gwf)에 채널명 각각 H1:NR2024\_WINTER, L1:NR2024\_WI

으로 저장된 중력파 신호 파일에서 신호가 갖고 있는 질량  $m_1$ ,  $m_2$ ,  $a_{1z}$ ,  $a_{2z}$ , 그리고

$m_2$ 에 대한 표준편차를 계산하여 제출한다. 문제를 풀기 위해서는 <https://github.com>

[odw/odw-2023](https://github.com/odw/odw-2023)를 참조한다. 특히 bilby를 활용한 모수추정을 설명한 [odw-](https://github.com/odw/odw-2023)

[2023/Tutorials/Day\\_3/Tuto\\_3.2\\_Parameter\\_estimation\\_for\\_compact\\_object\\_mergers.i](https://github.com/odw/odw-2023/Tutorials/Day_3/Tuto_3.2_Parameter_estimation_for_compact_object_mergers.i)

[at main · gw-odw/odw-2023 · GitHub](https://github.com/odw/odw-2023) 를 참고한다. 참고 사이트와의 차이는 우리는

어 있는 gwf 파일을 사용하고 psd도 `pycbc.psd.AdvDesignSensitivityT1800044` 를 사

는 점이다. 그리고 신호의 도착시간은 문제 3에서 찾은 값을 사용한다. 우리가 모수추

값은  $m_1, m_2, a_{1z}, a_{2z}$  이지만 실제로 모수 추정을 위한 posterior sample을 생성할 때

리 D까지 같이 해야 한다. 생성된 posterior sample을 이용하여  $m_1, m_2$ 에 대한 corne

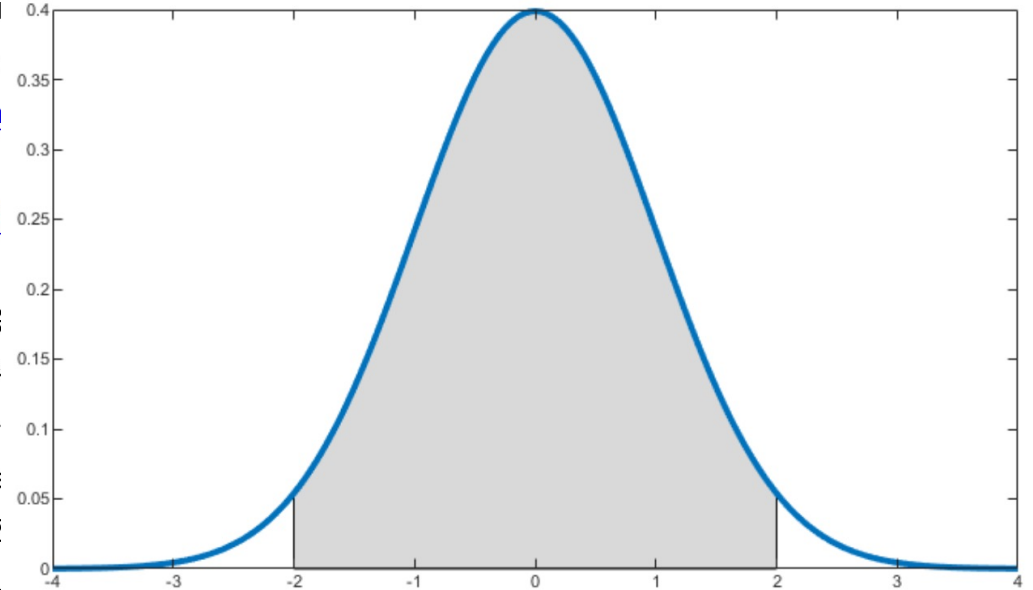
를 그리면  $m_1, m_2$ 의 값과 표준 편차를 알 수 있다. 상한과 하한이 다를 경우에는 하한

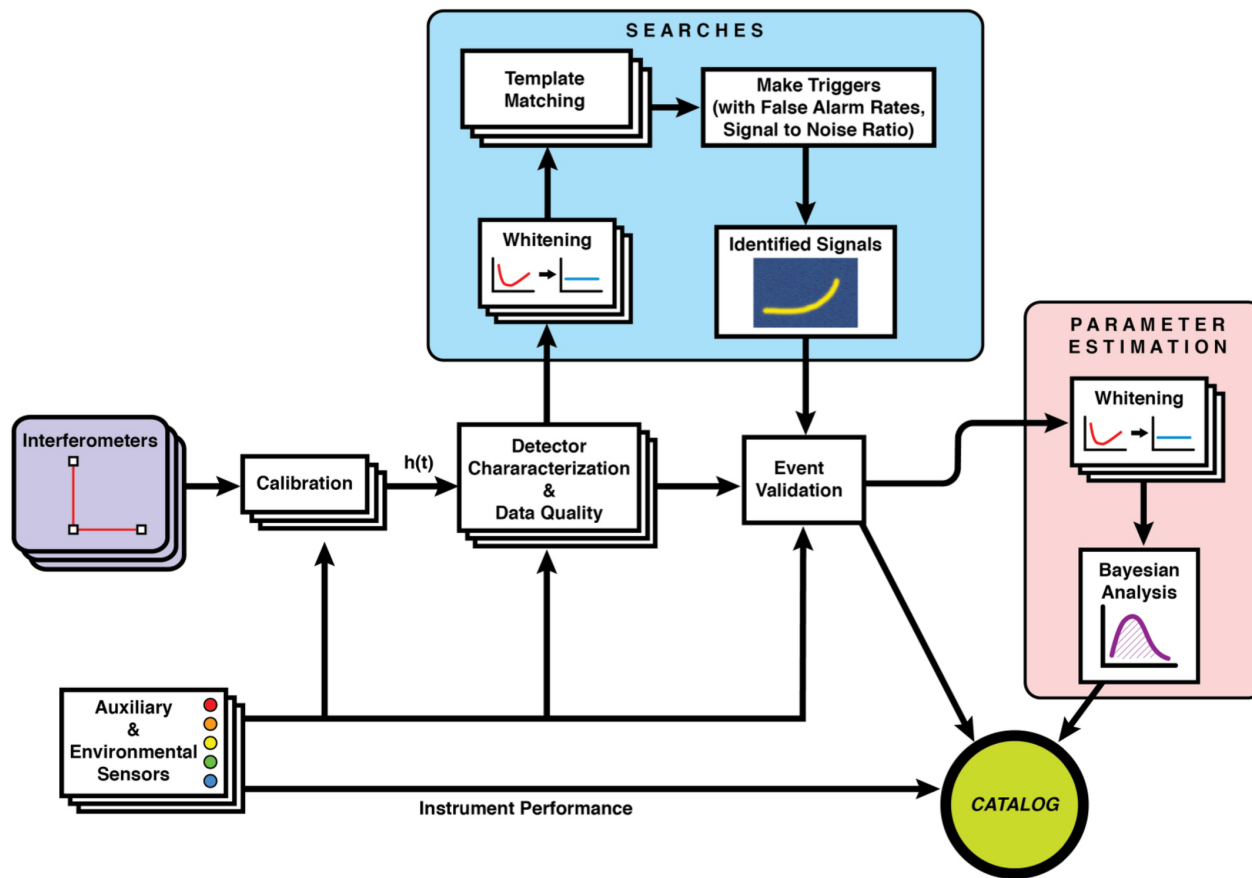
제출한다.  $a_{1z}, a_{2z}$ 의 값은  $a_1, a_2$  에 대한 corner plot을 그려서 그 값을 알 수 있다.

각 값마다 5점씩 부여한다. **표준편차의 값은 실제로 90% confidence interval을 계산한 후에**

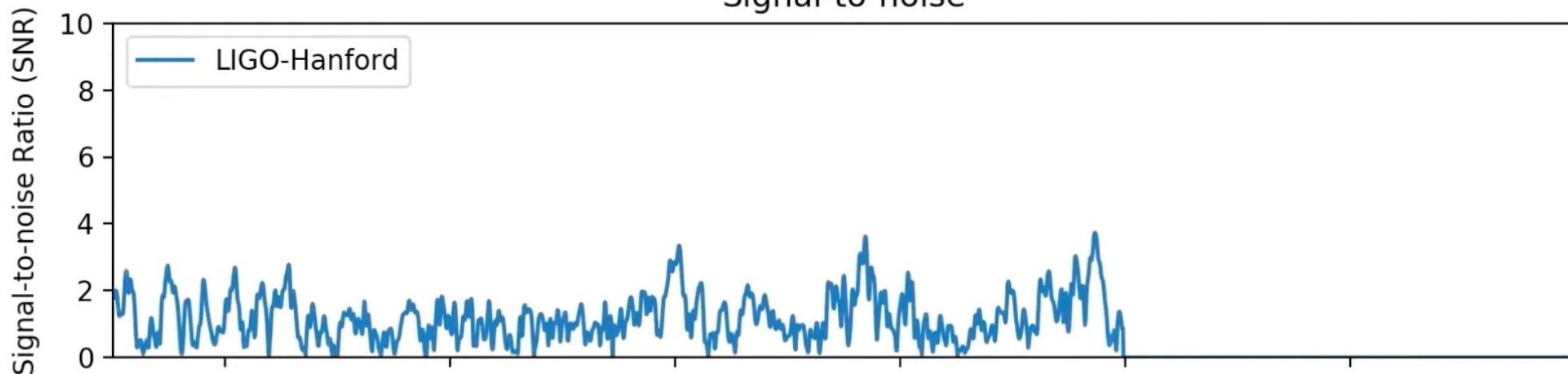
**중간 값에서 하한 값을 뺀 값을 제출한다. 계산한 posterior sampler에서 중간값(median)과**

**90% confidence interval을 계산하는 예제는 위에 언급한 사이트를 참고하시오.**





Signal-to-noise



Data from the LIGO Hanford Observatory (whitened and bandpassed)

