Understanding gravitational-wave sources with Bayesian inference

Or: how I learned to stop worrying and love counting

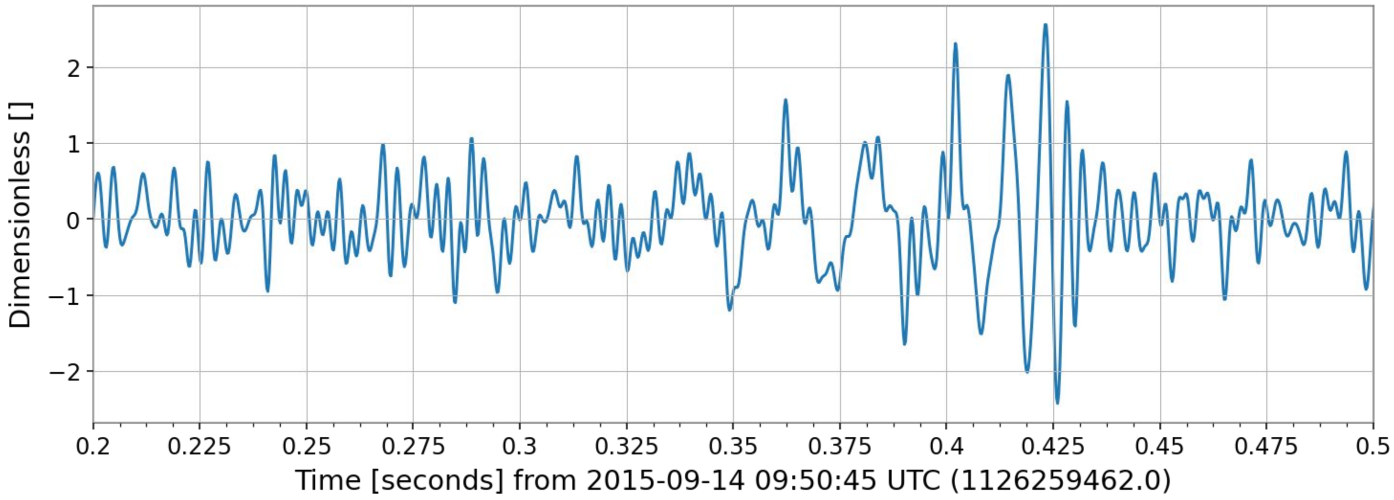
Rory Smith
Monash University
Adelaide Winter School July 20th 2022

Overview

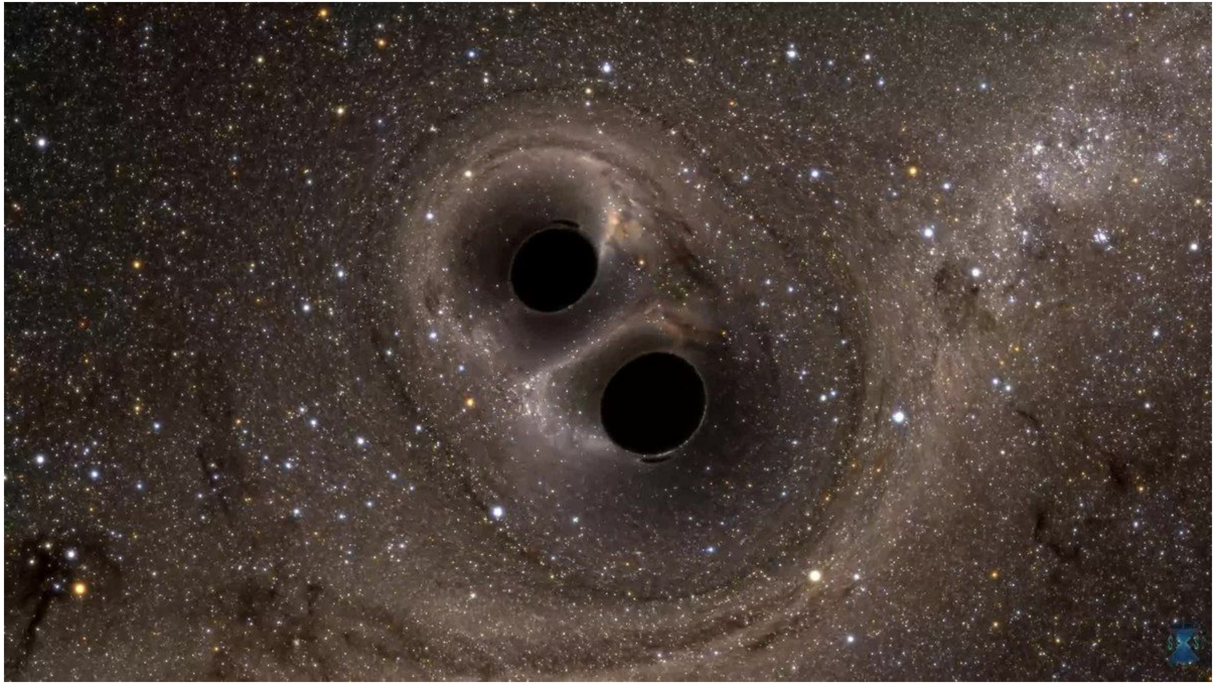
- Building intuition for astrophysical measurements as probability distributions
    - Measuring binary black hole masses in a toy universe by counting repeated trials
    - Understanding the link between counting and Bayesian inference/computation

Gravitational-wave astronomy

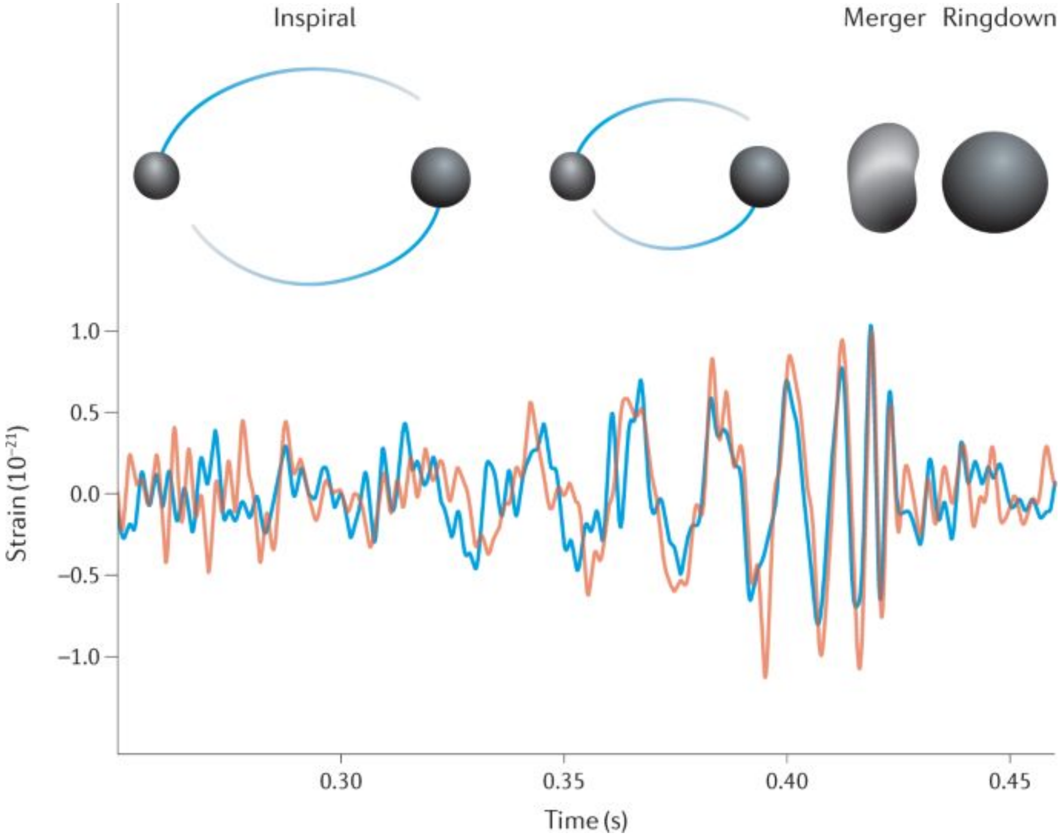
GW150914

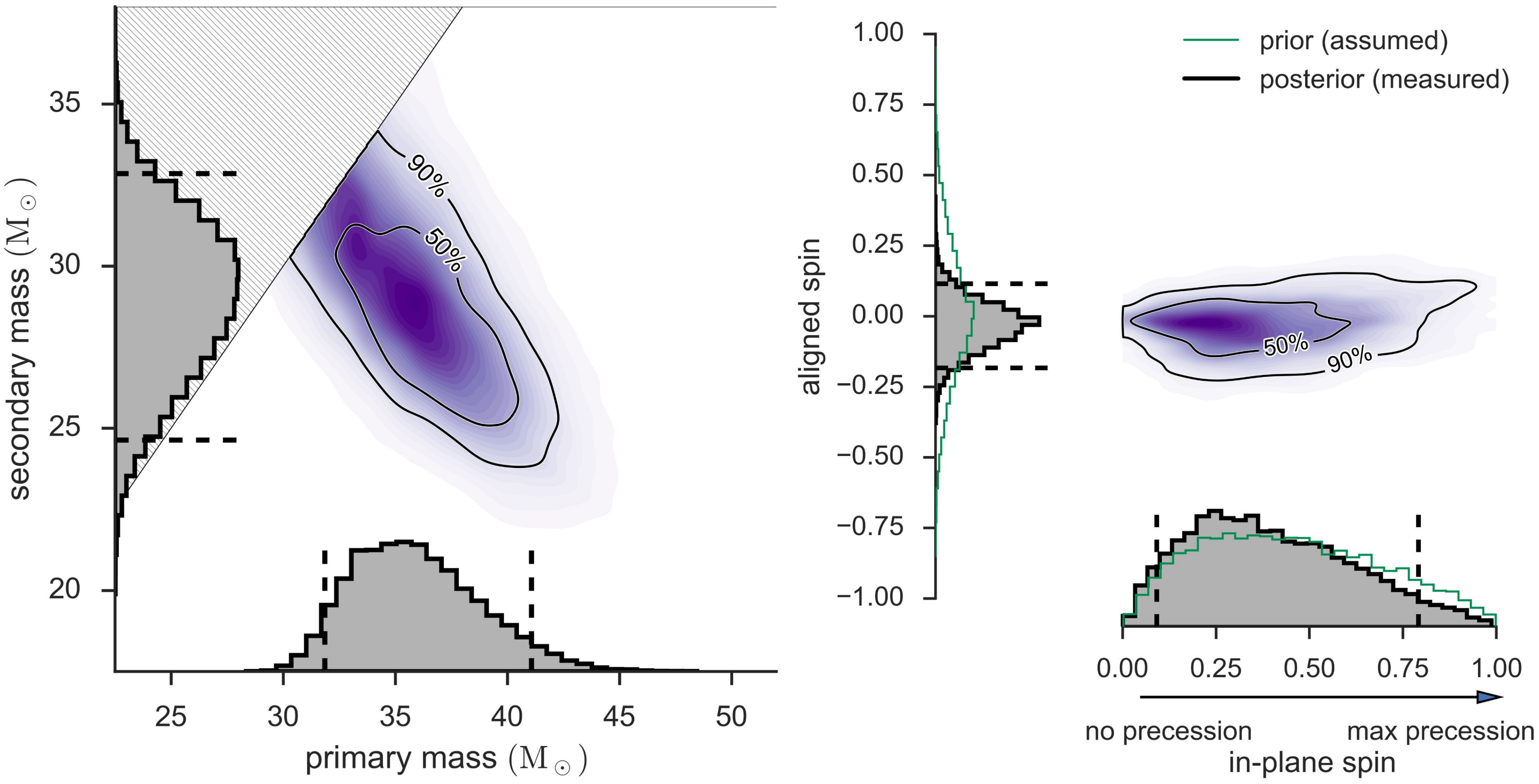


Binary black holes exist!



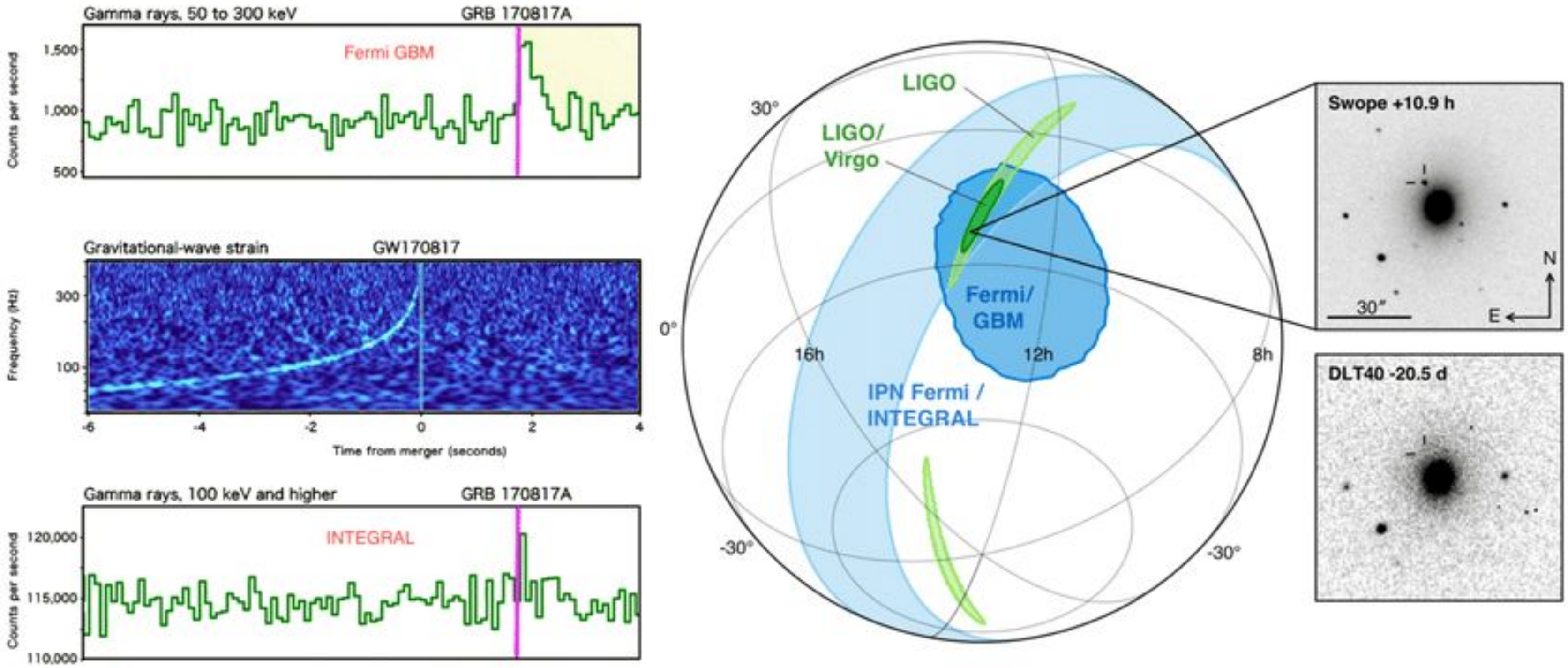
Gravitational-wave astronomy



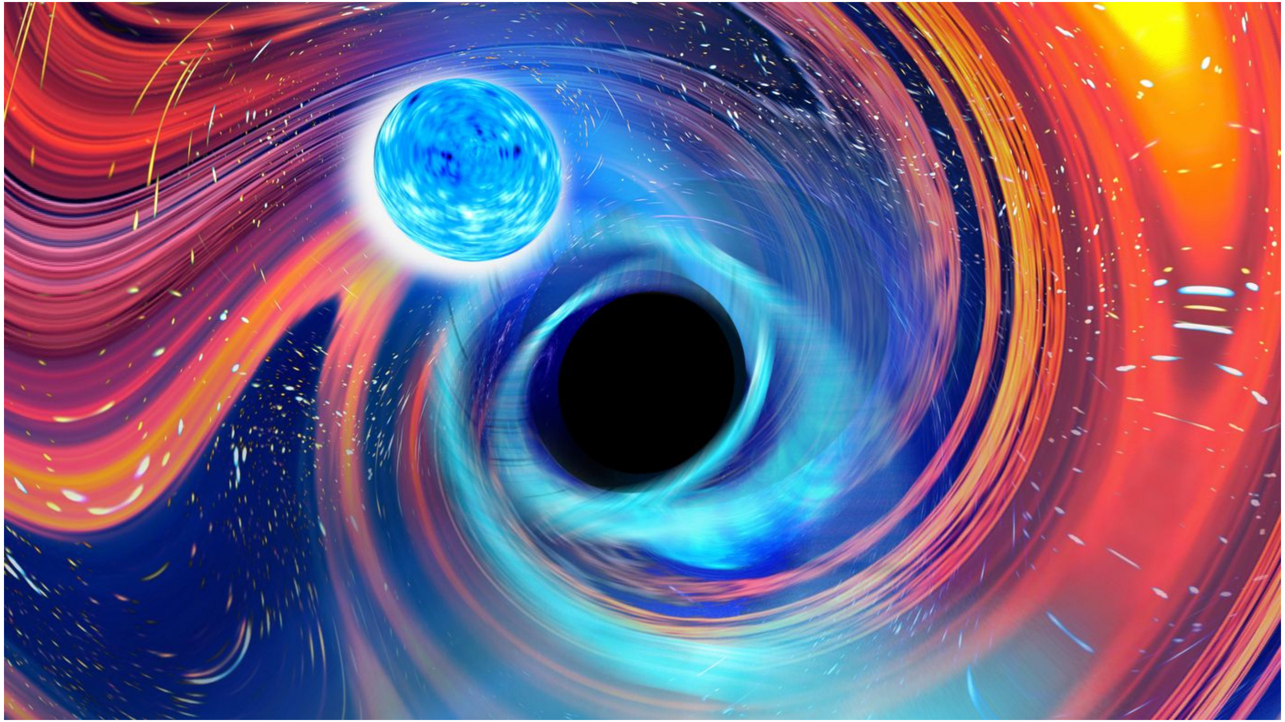


secondary mass $(\mathrm{M}_\odot)$
35
30
25
20
90%
50%
25
30
35
40
45
50
primary mass $(\mathrm{M}_\odot)$
prior (assumed)
posterior (measured)
aligned spin
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
50%
90%
0.00
0.25
0.50
0.75
1.00
no precession
max precession
in-plane spin

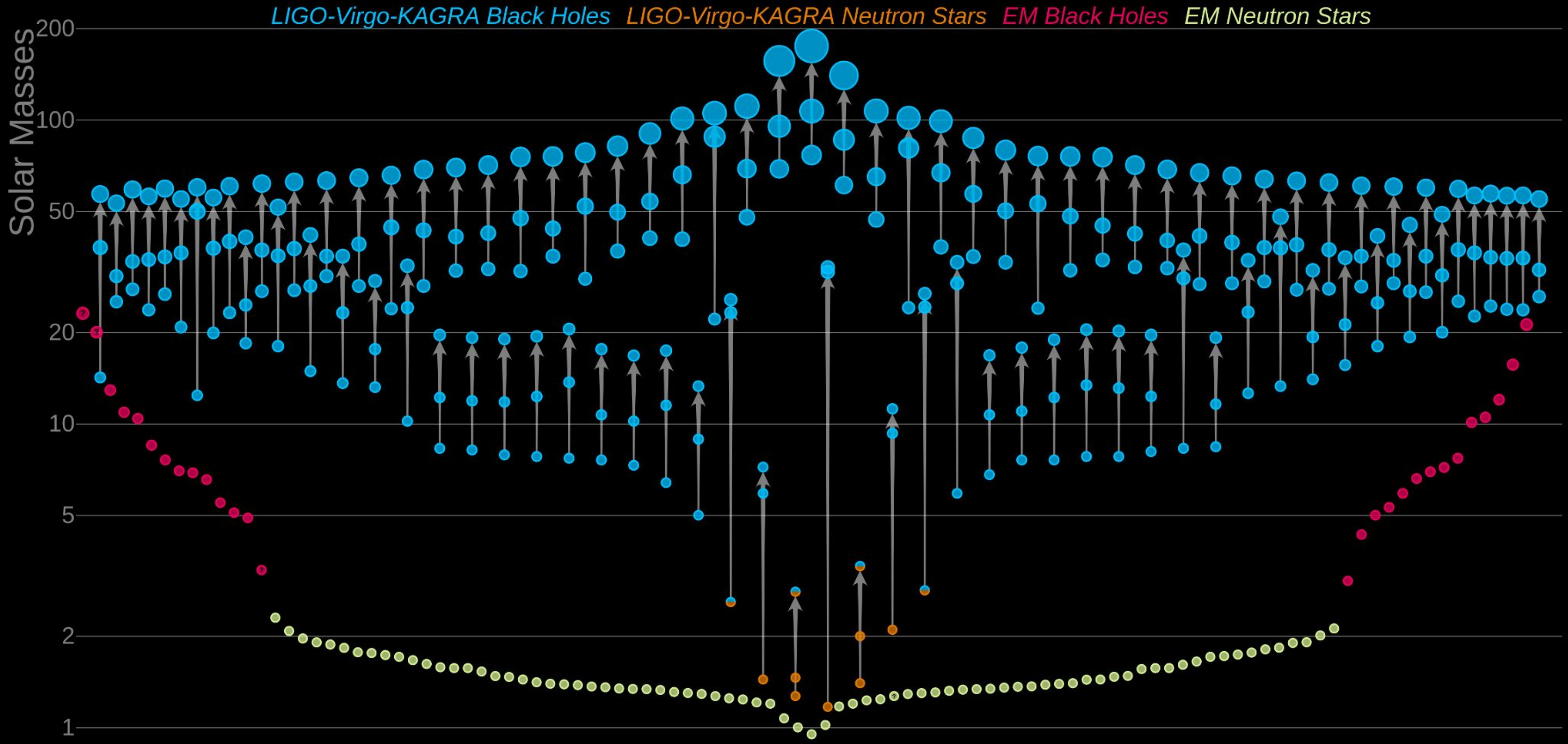
Binary neutron star mergers happen!



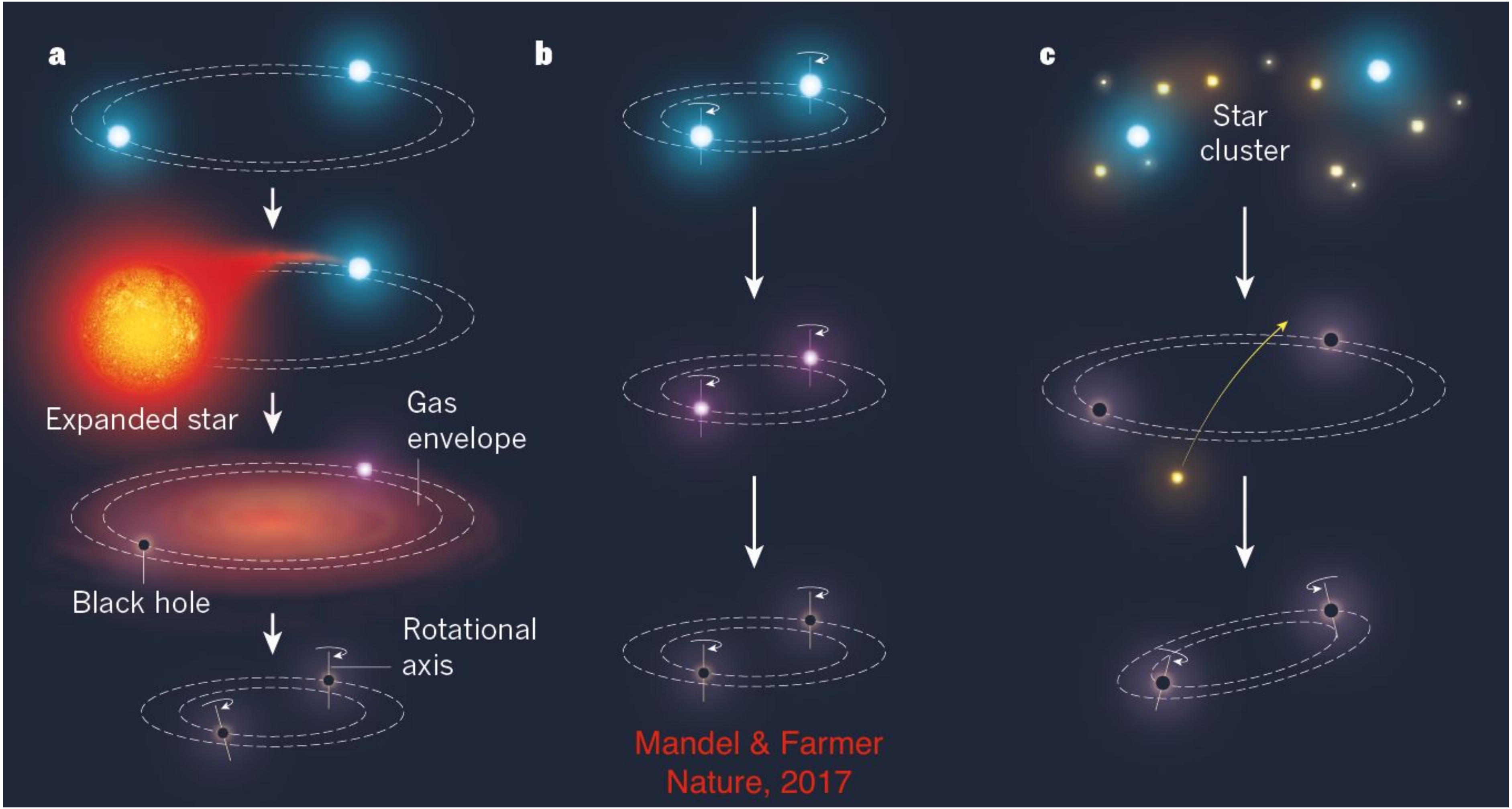
Neutron stars and black holes



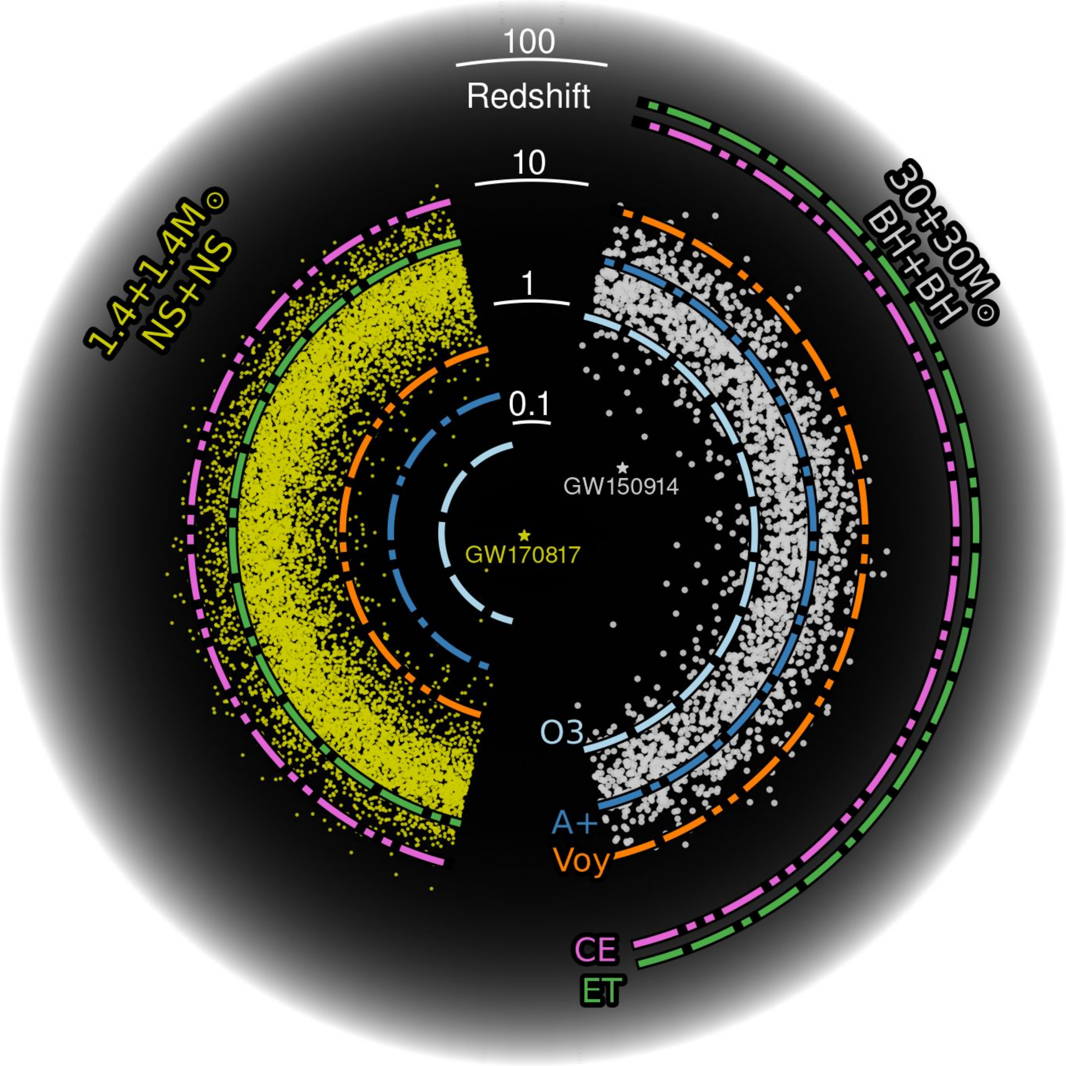
Masses in the Stellar Graveyard

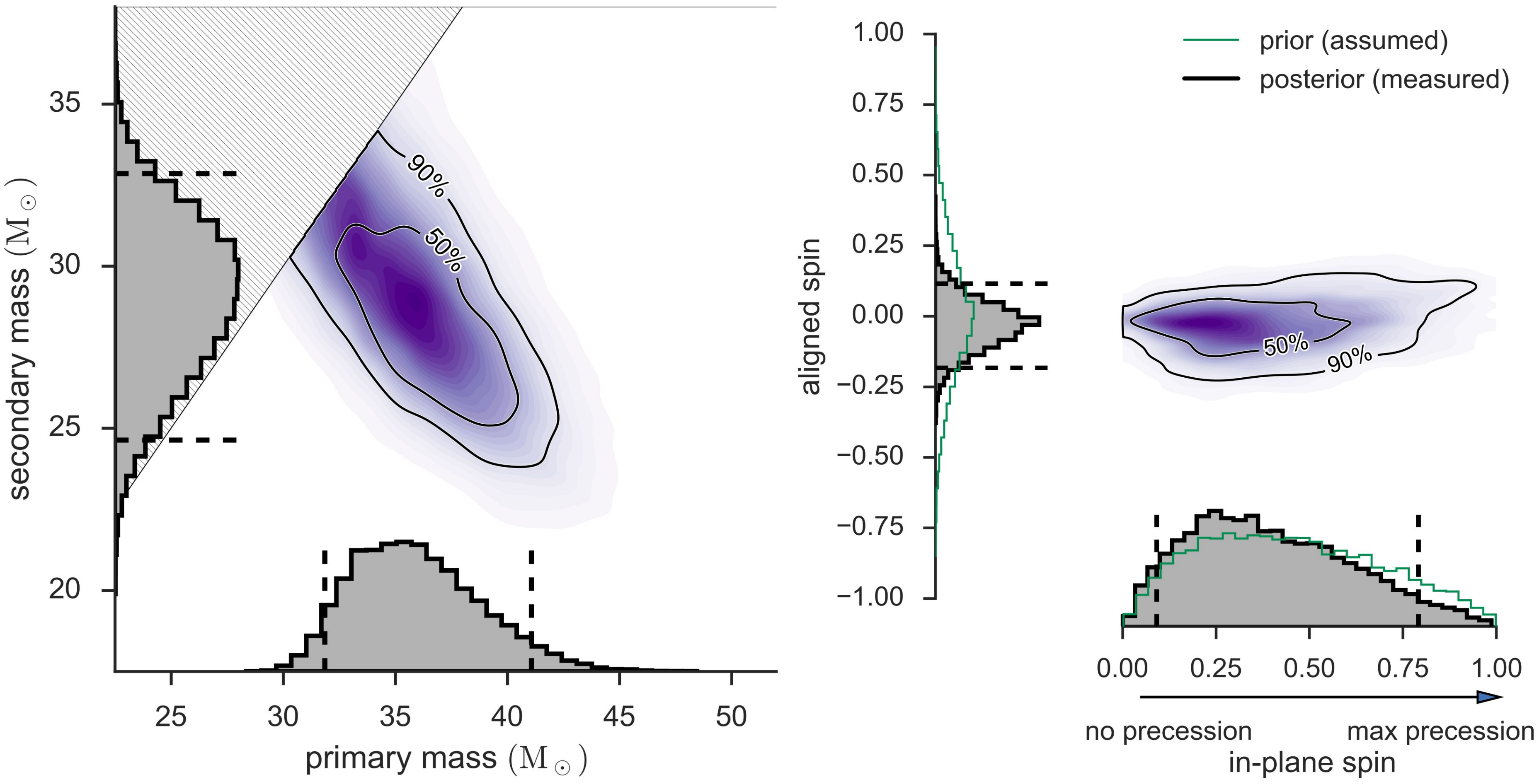


LIGO-Virgo-KAGRA | Aaron Geller | Northwestern


a
b
c
Expanded star
Gas envelope
Black hole
Rotational axis
Star cluster
Mandel & Farmer
Nature, 2017

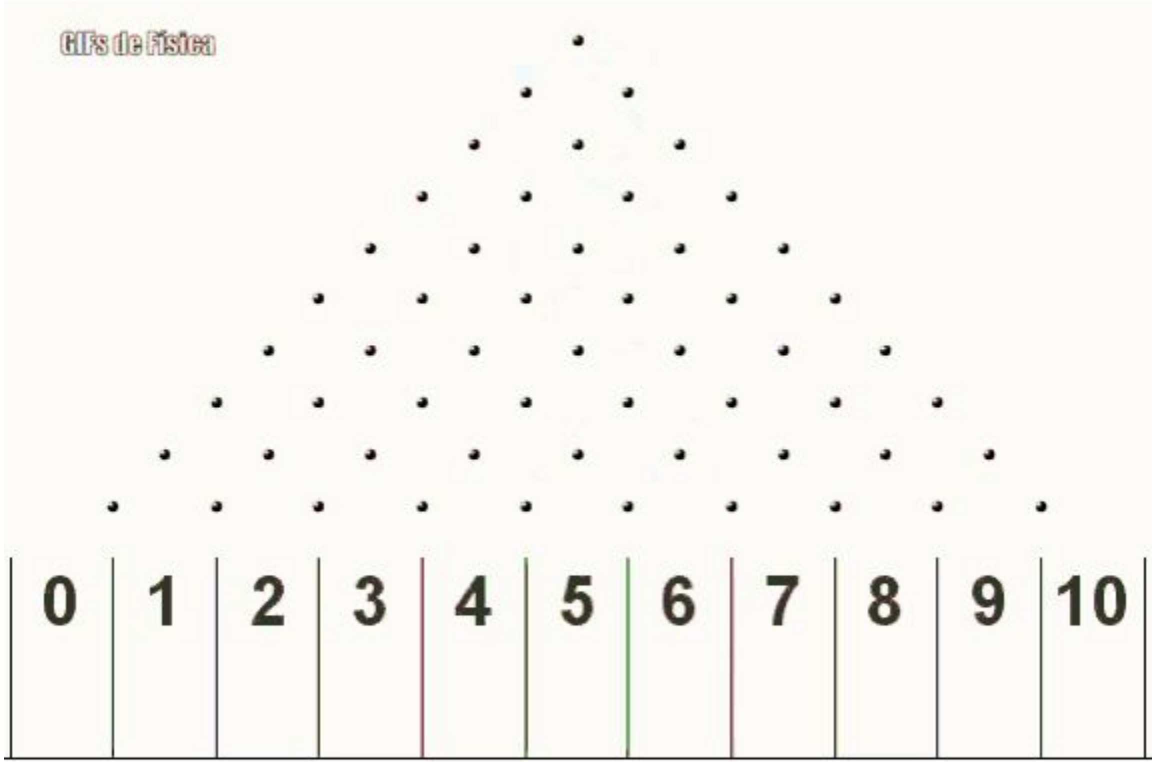
Next Gen


100
Redshift
10
1
0.1
1.4+1.4M⊙
NS+NS
30+30M⊙
BH+BH
GW150914
GW170817
O3
A+
Voy
CE
ET



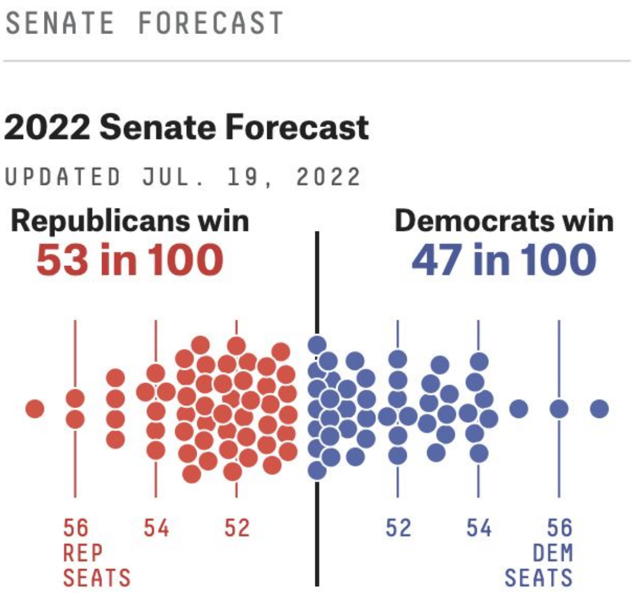
secondary mass $(\mathrm{M}_\odot)$
35
30
25
20
90%
50%
25
30
35
40
45
50
primary mass $(\mathrm{M}_\odot)$
prior (assumed)
posterior (measured)
aligned spin
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
50%
90%
0.00
0.25
0.50
0.75
1.00
no precession
max precession
in-plane spin

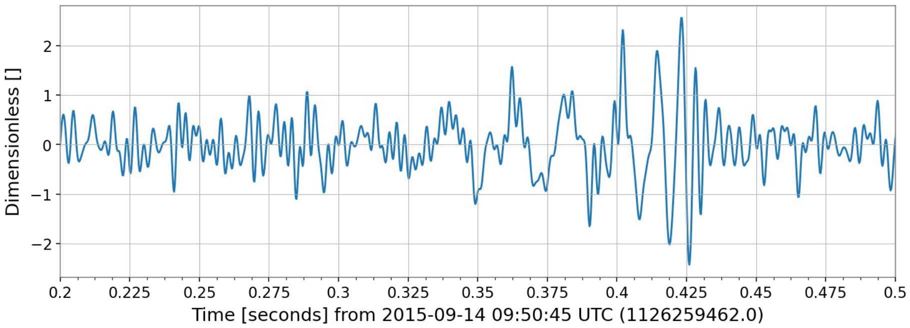
Probability and counting



The “why” of it all

- Real-world events are often not repeatable

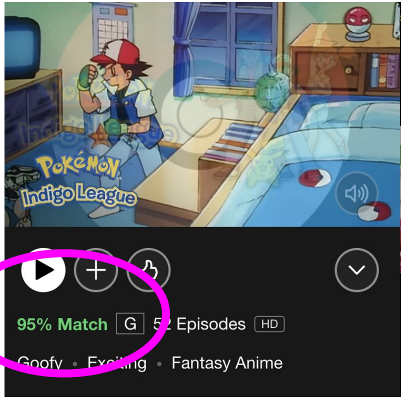


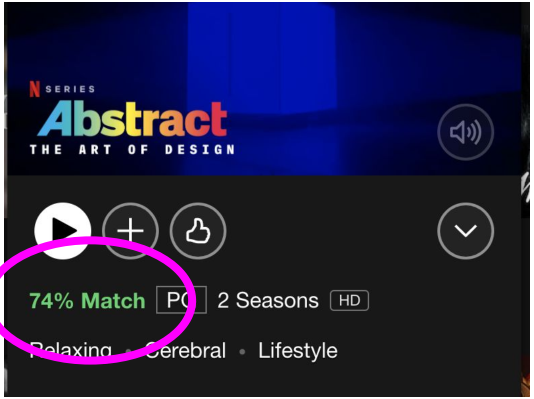


Credit: **FiveThirtyEight**

The “why” of it all

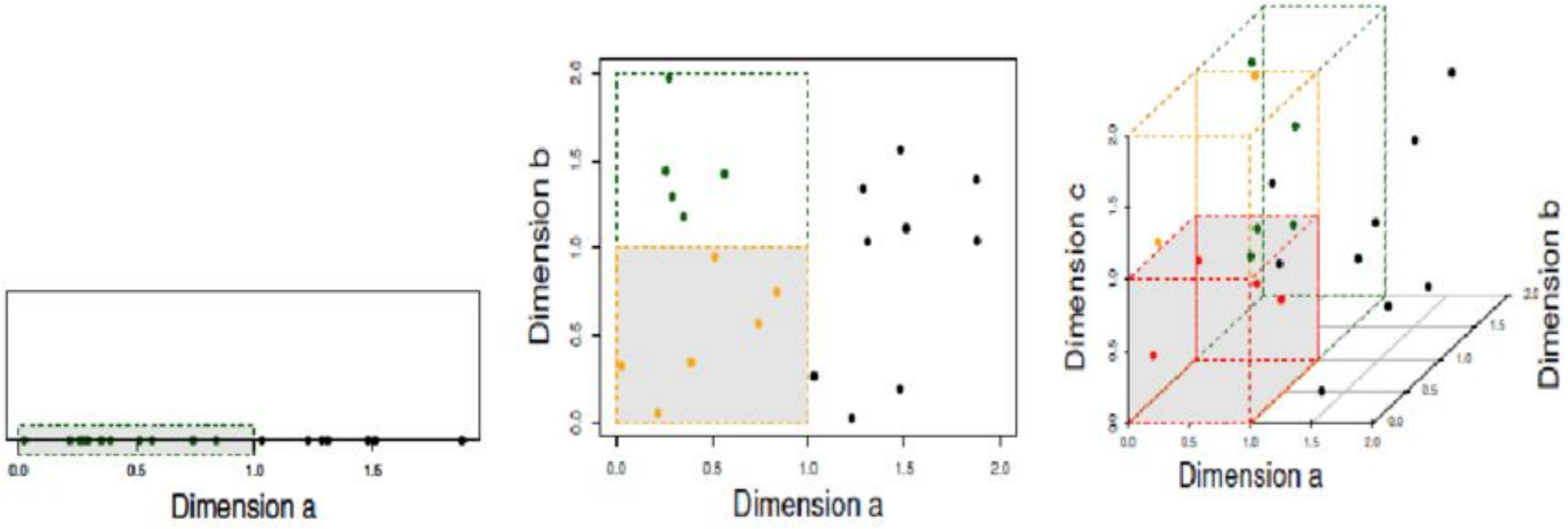
- Data can be sparse or degenerate, but we still want to make predictions





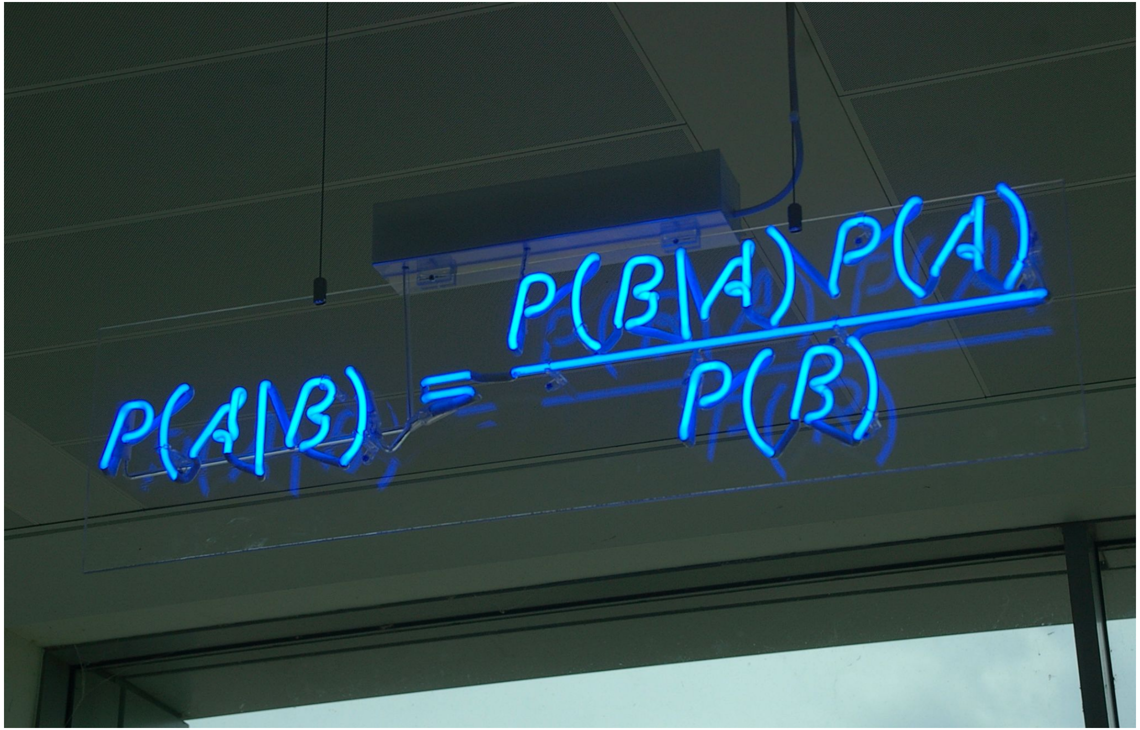
The “why” of it all

- *Curse of dimensionality:* Data can be so high-dimensional that we can never get an adequate sampling



Credit: medium

Bayes theorem



Probability: Counting and computation

Bayesian inference is a way **computing** probability densities of things given data and models of uncertainty, e.g.,

In:

- Experimental Data
- Model of a gravitational wave signal
- Model of detector noise

Out: probabilities of things like

    - Hypotheses: signal or noise? GR or not-GR? etc…
    - Source properties of binary black holes given noisy data and uncertainty about the population

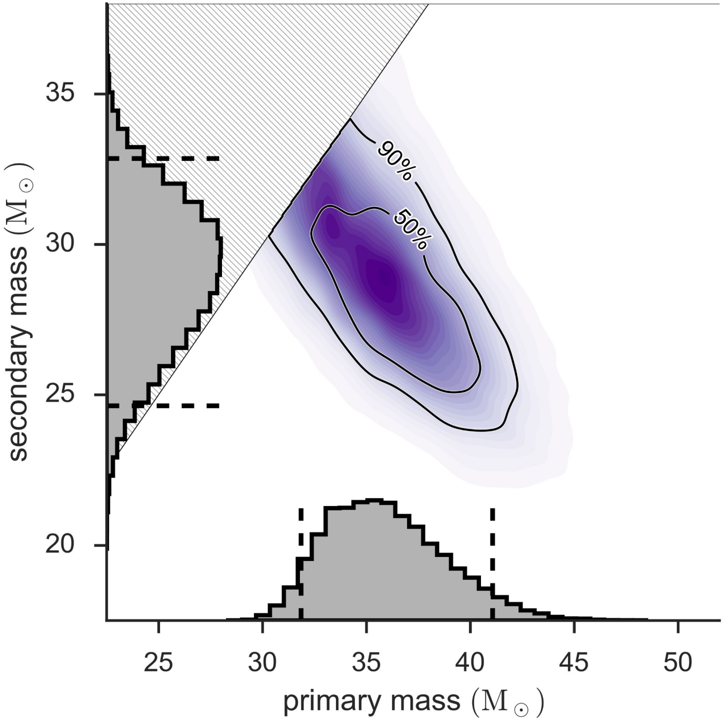
Does not necessarily involve counting the frequency of outcomes of an experiment/trial

Bayes theorem in gravitational-wave astronomy

We are looking at a *posterior probability density*

$$p(m_1, m_2 | d_{\mathrm{Hanford}}, d_{\mathrm{Livingston}})$$

It was computed using Bayes theorem



Building intuition

- Starting from mathematical definitions of Bayes theorem can be complex, unintuitive and time-consuming

Building intuition

- Let’s imagine we could measure BH masses without Bayesian inference/computation

- We will use this example to draw the link between counting repeated mergers and Bayesian inference – the two approaches should agree!

Weighing black holes in a simulated toy Universe

How could we estimate the properties of a binary black hole given gravitational-wave detector data?

Create millions of binary black hole mergers and measure the GW signals in a noisy detector

Weighing black holes in a simulated toy Universe

How could we estimate the properties of a binary black hole given gravitational-wave detector data?

Create millions of binary black hole mergers and measure the GW signals in a noisy detector

- 1: Simulate many noisy gravitational-wave observations to get the relationship between source properties and data
- 2: Given some new data, infer the probability of the black hole's mass by *counting*

Use example to **build intuition** for posteriors etc… in GW astronomy

- 3: Repeat (2) using Bayesian inference, i.e., computation instead of counting

Weighing black holes in a simulated toy Universe

How could we estimate the properties of a binary black hole given gravitational-wave detector data?

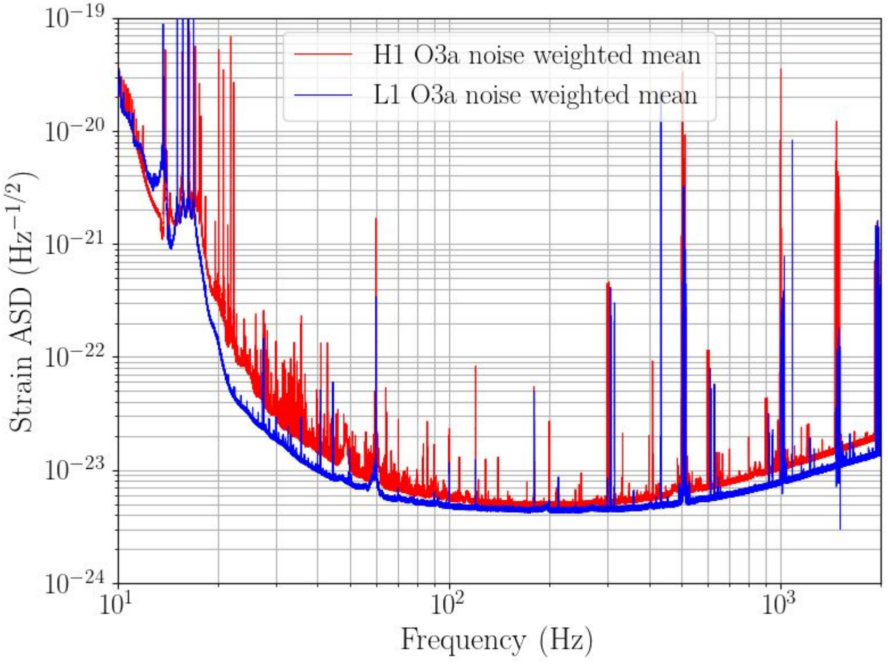
- First, recall what a GW detector like LIGO/Virgo measure

Weighing black holes in a simulated toy Universe

How could we estimate the properties of a binary black hole given gravitational-wave detector data?

- The detectors also record random noise with well known statistical properties

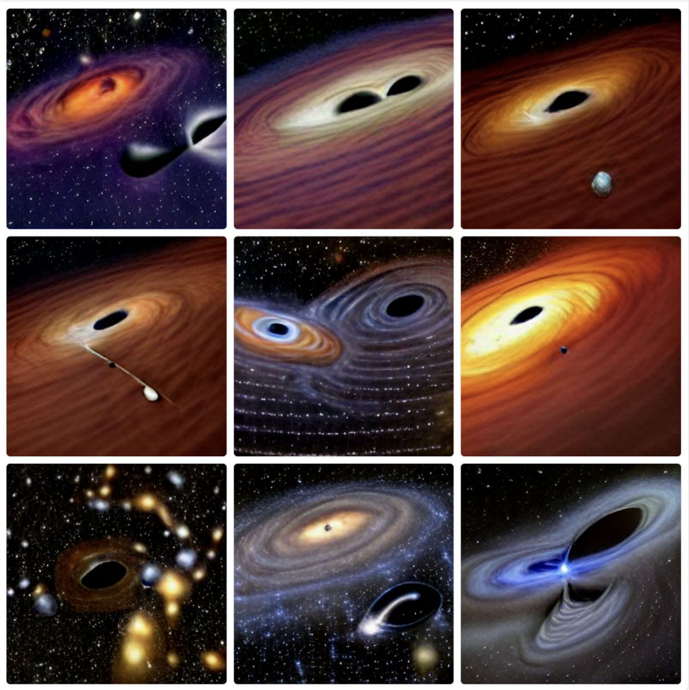
- The noise *amplitude spectral density* (**ASD**) is proportional to the standard deviation of the noise
  [Lecture_1_Intro_to_LIGO_data_and_signals_adelaide.ipynb]()



Weighing black holes in a simulated toy Universe

How could we estimate the properties of a binary black hole given gravitational-wave detector data?

- We also need signals!
- Assume signals are thrown at us randomly with total mass drawn from some distribution that we can freely choose (we're the deity in our toy Universe!)



DALL·E mini

Weighing black holes in a simulated toy Universe

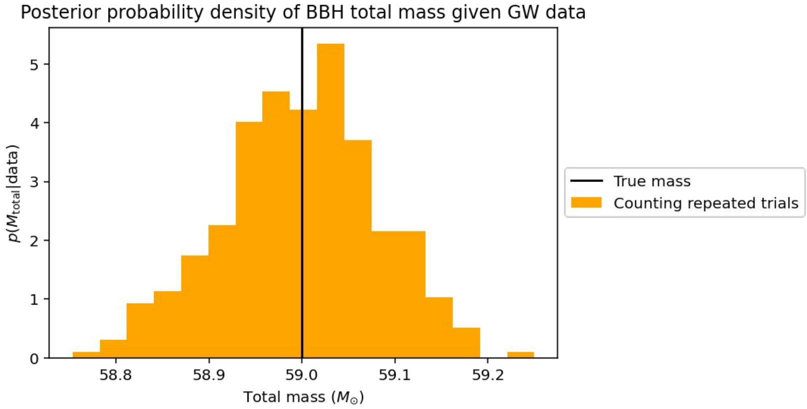
Follow along here:

https://github.com/rory-smith/Lectures-on-inference-in-GW-astronomy/blob/master/GW_inference_in_a_toy_universe_probability_as_counting_and_computation.ipynb

Weighing black holes in a simulated toy Universe

- Clearly we cannot do this experiment in our Universe!
  - In order to obtain a measurement like this for real GW events, we need to resort to **Bayesian Computation!**

- The goal in Bayesian inference is to compute the numerical value of the probability density (y-axis) using our knowledge of the **uncertainty** of our data



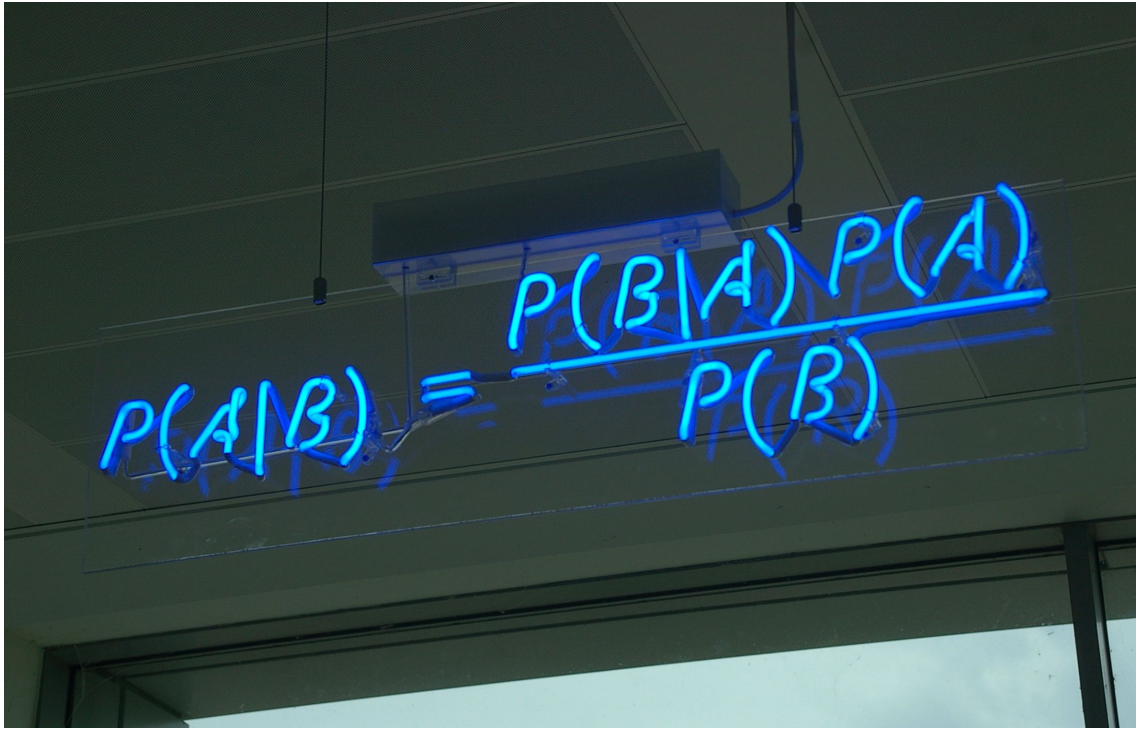
Weighing black holes in the real Universe

- We want the left hand side

$$p(M_{\text{total}}|\text{data})$$

which, in the real Universe, is unobtainable by events

- What does the right hand side look like?



Weighing black holes in a simulated toy Universe

Question

- What are the sources of uncertainty in both the toy and Real Universe?

Weighing black holes in a simulated toy Universe

- What are the sources of uncertainty in our toy Universe?
  - Detector noise
  - Distribution of BBH total mass

Weighing black holes in the real Universe

"Likelihood function" (probability density of the data given a BBH signal with M_{total})

"Prior probability density" (The probability that the Universe produces a BBH with a particular M_{total})

$$p(M_{\rm total}|\mathrm{data}) = \frac{\mathcal{L}(\mathrm{data}|M_{\rm total})\,\pi(M_{\rm total})}{\mathcal{Z}(\mathrm{data})}$$

"Posterior probability density"

"Evidence" (the probability density of obtaining the strain data)

Weighing black holes in the real Universe

Okay…so what?

Weighing black holes in the real Universe

$$p(M_{\mathrm{total}}|\mathrm{data}) = \frac{\mathcal{L}(\mathrm{data}|M_{\mathrm{total}})\,\pi(M_{\mathrm{total}})}{\mathcal{Z}(\mathrm{data})}$$

It turns out we can compute everything on the right hand side!

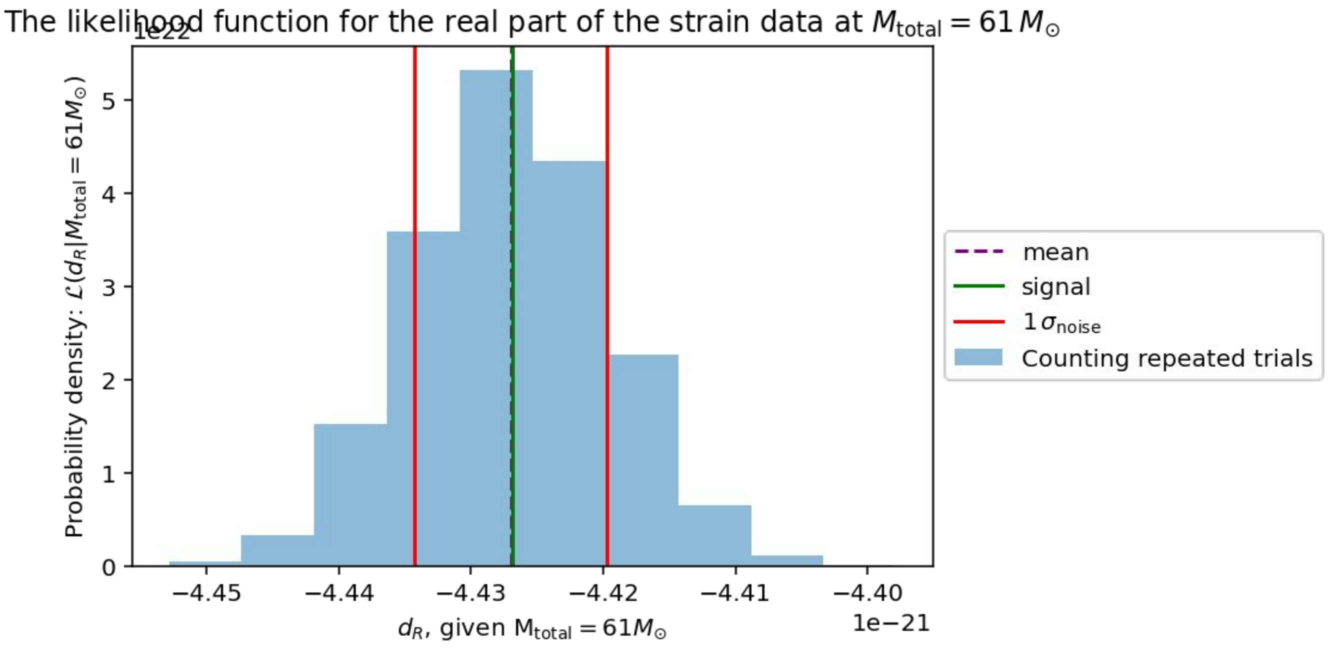
- The likelihood “only” depends on the properties of noise and the deterministic signal evaluated at M_{total}
- The prior is uncertain (we ultimately want to measure it) but we can make assumptions based on, e.g., astrophysics or other considerations
- The evidence can be found by integrating likelihood x prior (more on that later)

Weighing black holes in the real Universe

$$p(M_{\rm total}|{\rm data}) = \frac{\mathcal{L}({\rm data}|M_{\rm total})\,\pi(M_{\rm total})}{\mathcal{Z}({\rm data})}$$

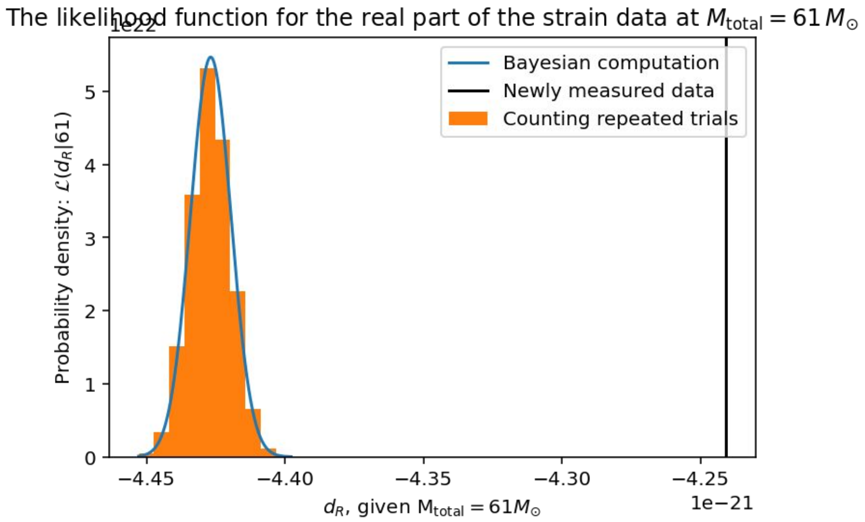
1. Let's start by plotting and figuring out how to compute the likelihood

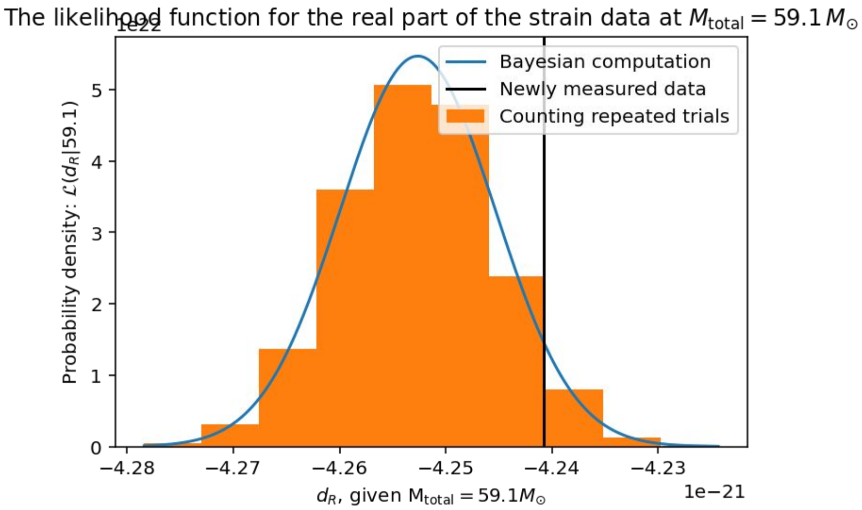
The likelihood function



The likelihood function

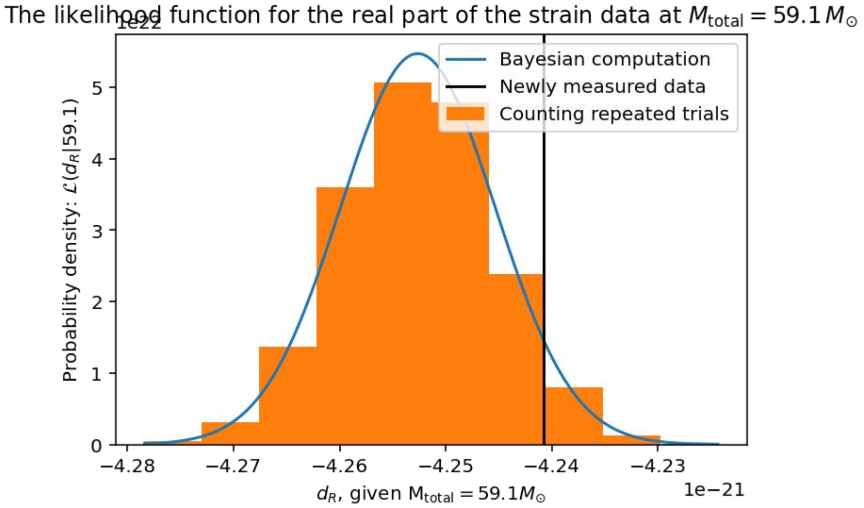
$$\mathcal{L}(\text{data}|M_{\text{total}}) = \mathcal{N}(\text{data}, \mu = h(M_{\text{total}}), \sigma^2 \propto \text{PSD})$$





The likelihood function

$$\mathcal{L}(\text{data}|M_{\text{total}}) = \mathcal{N}(\text{data}, \mu = h(M_{\text{total}}), \sigma^2 \propto \text{PSD})$$



We have accurate models for GW signals and PSDs!

Weighing black holes in the real Universe

$$p(M_{\rm total}|{\rm data}) = \frac{\mathcal{L}({\rm data}|M_{\rm total})\,\pi(M_{\rm total})}{\mathcal{Z}({\rm data})}$$

It turns out we can compute everything on the right hand side!

- ~~The likelihood "only" depends on the properties of noise and the deterministic signal evaluated at~~ $\cancel{M_{\rm total}}$
- The prior is uncertain (we ultimately want to measure it) but we can make assumptions based on, e.g., astrophysics or other considerations
- The evidence can be found by integrating likelihood x prior (more on that later)

A note on priors

In our toy Universe, we know how BBH masses are distributed (as well having perfect knowledge of their other properties)

- Here, we will just use our perfect knowledge of the prior

- In practice, we have to make *assumptions* about the priors of all the binary's properties! **See sessions by Shanika and Simon!**
    - Learning the prior distribution is all about astrophysics

Weighing black holes in the real Universe

$$p(M_{\rm total}|{\rm data}) = \frac{\mathcal{L}({\rm data}|M_{\rm total})\,\pi(M_{\rm total})}{\mathcal{Z}({\rm data})}$$

It turns out we can compute everything on the right hand side!

- ~~The likelihood “only” depends on the properties of noise and the deterministic signal evaluated at~~ ~~$M_{\rm total}$~~
- ~~The prior is uncertain (we ultimately want to measure it) but we can make assumptions based on, e.g., astrophysics or other considerations~~
- The evidence can be found by integrating likelihood x prior (more on that later)

The Evidence

$$p(M_{\rm total}|{\rm data}) = \frac{\mathcal{L}({\rm data}|M_{\rm total})\,\pi(M_{\rm total})}{\mathcal{Z}({\rm data})}$$

…is more than just a normalization!

- Note that the posterior probability is normalized to 1 when integrated over masses. So the evidence is just

$$\mathcal{Z}({\rm data}) = \int dM_{\rm total}\mathcal{L}({\rm data}|M_{\rm total})\,\pi(M_{\rm total})$$

Once we compute likelihood x prior, we can go back and compute the evidence

The Evidence

…is more than just a normalization!

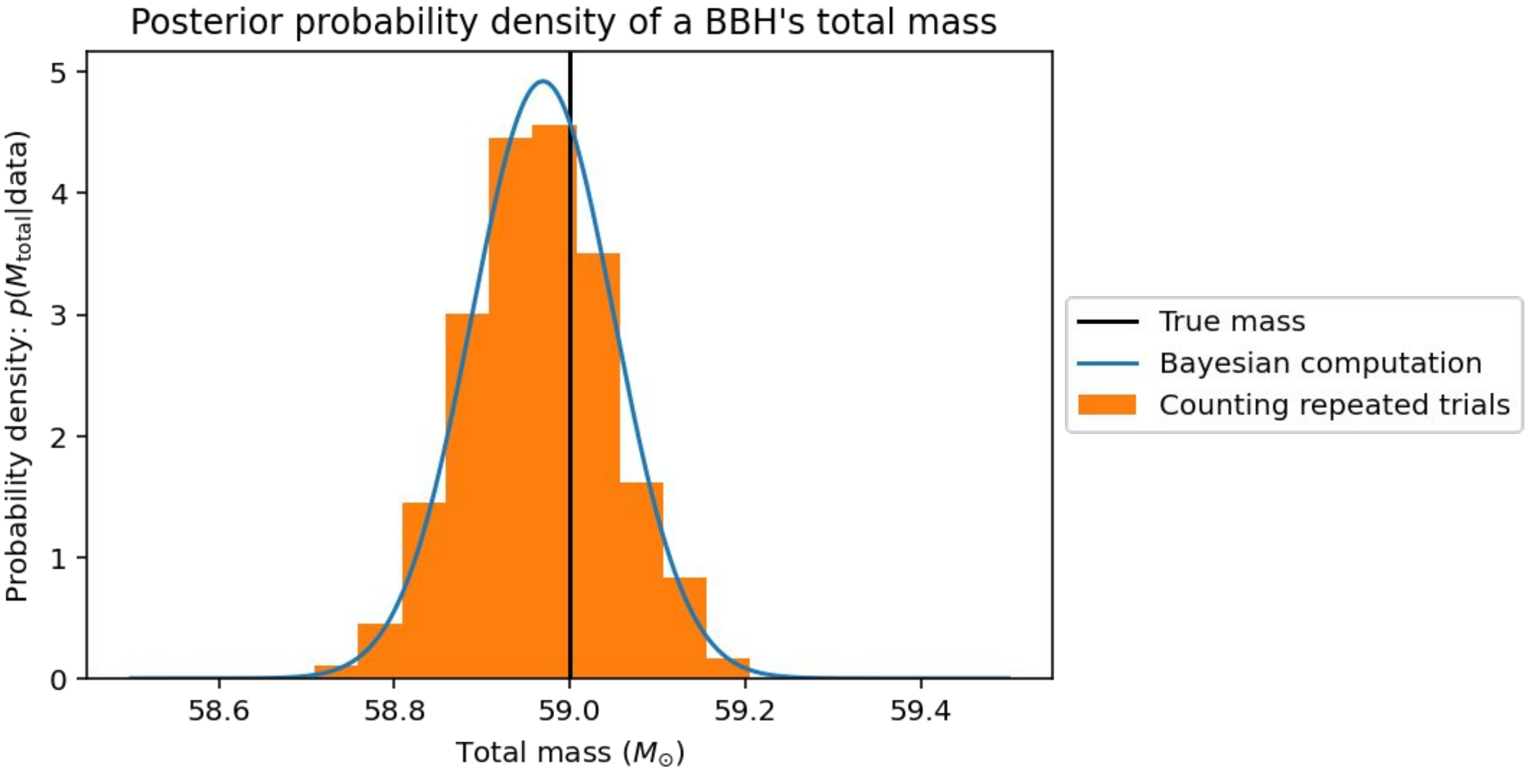
- It is a cornerstone of **Hypothesis Testing** (more on this later)

Putting it all together

- **Goal**: show that by using our knowledge of uncertainty in the data, we can compute the numerical value of the posterior probability (using Bayes theorem) without having to conduct an experiment where we create and collide millions of binary black holes…This is good news for astrophysicists and the environment.

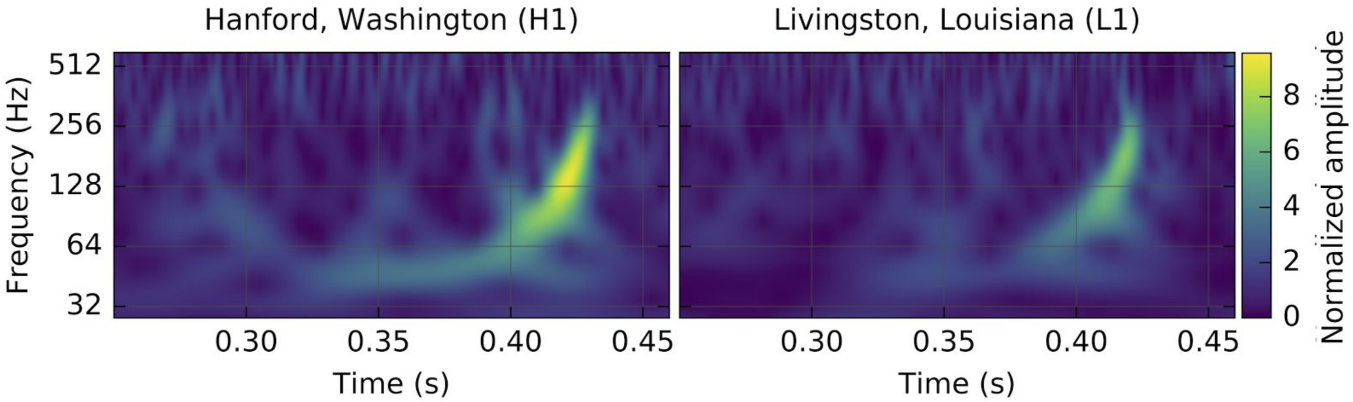
- The Bayesian computation should, however, agree with our perfectly good and intuitive definition of probability as counting repeated trials

Weighing black holes in the real Universe



Some comments: analysing real data

- Real data has a spectrum of Fourier frequencies



This does not change any of the stuff we've covered, beyond needing to compute probabilities of a vector of data rather than a single number

Some comments: parameters

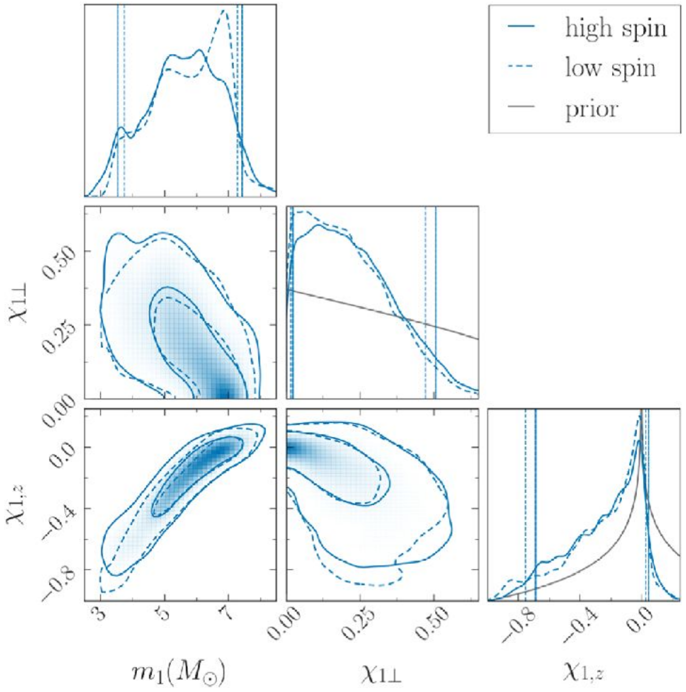
- We have only looked at a single parameter. Binary black holes have at least 15, and binary neutron stars at least 17.

Exercise: explore the literature to familiarise yourself with all of these parameters

Hint: a binary black hole system has two masses, six spin components, orbital inclination w.r.t. our line of sight, luminosity distance, sky location, polarization, and time and phase at coalescence.

Some comments: parameters

- We have only looked at a single parameter. Binary black holes have at least 15, and binary neutron stars at least 17.

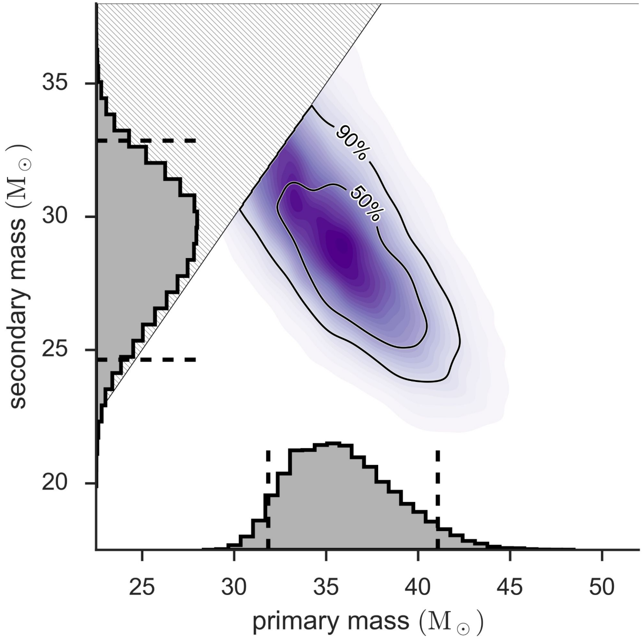


Often you'll see **"corner plots"** like this.

They are a way to visualise low-dimensional projections of a 15+ dimensional(!) probability distribution

Some comments: sampling parameters

- We have only looked at a single parameter. Binary black holes have at least 15, and binary neutron stars at least 17.



Cannot plot the posterior in 15+ dimensions (curse of dimensionality)

Use ***"sampling" algorithms*** to pick samples from the posterior distribution with the right probability

Some comments: priors

- In our toy model, we assumed we know the prior perfectly. In practice, we make assumptions that we know are incorrect
  - You'll often see people talk about "uninformative" priors
  - This is to cast a wide net
- This will bias posterior PDFs!
  - However, with multiple observations we can actually measure the prior and "undo" the choice of bad priors **(See Shanika and Simon's presentations)**

B. Choice of priors

We analyze coherently 8 s of data with a uniform prior on t_c of width of ± 0.1 s, centered on the time reported by the online analysis [1,84], and a uniform prior in $[0, 2\pi]$ for ϕ_c. We consider the frequency region between 20 Hz, below which the sensitivity of the instruments significantly degrades (see panel (b) of Fig. 3 in Ref. [1]), and 1024 Hz, a safe value for the highest frequency contribution to radiation from binaries in the mass range considered here.

Given the lack of any additional astrophysical constraints on the source at hand, our prior choices on the parameters are uninformative. We assume sources uniformly distributed in volume and isotropically oriented. We use uniform priors in $m_{1,2} \in [10, 80] M_\odot$, with the constraint that $m_2 \leq m_1$. We use a uniform prior in the spin magnitudes $a_{1,2} \in [0, 1]$. For angles subject to change due to precession effects we give values at a reference GW frequency $f_{\mathrm{ref}} = 20$ Hz. We use isotropic priors on the spin orientation for the precessing model. For the nonprecessing model, the prior on the spin magnitudes may be interpreted as the dimensionless spin projection onto $\boldsymbol{L}$ having a uniform distribution $[-1, 1]$. This range includes binaries where the two spins are strongly antialigned relative to one another. Many such antialigned-spin comparable-mass systems are unstable to large-angle precession well before

Properties of the Binary Black Hole Merger GW150914, LVC, PRL 2016

Some comments: hypotheses

- One important thing we haven't touched on is explicitly talking about hypotheses for the data and source populations.
  - Partly, this has been to try to streamline the intro. However, it's important to be aware that in practice you'll usually see Bayes theorem written like this:

$$p(\theta|d,\mathcal{H}) = \frac{\mathcal{L}(d|\theta,\mathcal{H})\pi(\theta|\mathcal{H})}{\mathcal{Z}(d|\mathcal{H})}$$

This explicitly tells us that our measurements are contingent on a particular hypothesis H, and using a different hypothesis will change our results, e.g., we could pick different priors, use different signal models (with beyond GR effects)

Some comments: hypothesis testing

“The evidence is more than just a normalization”

$$\text{“Bayes Factor”} = \frac{\mathcal{Z}(d|\mathcal{H}_1)}{\mathcal{Z}(d|\mathcal{H}_2)}$$

The literal meaning of the evidence is how likely the measured data are given a model of the uncertainty in the data, i.e., an hypothesis

Bayes factors are proportional to **betting odds** (see, e.g., Veitch and Vecchio 2010 “Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network”)

Some comments: hypothesis testing

$$\text{“Bayes Factor”} = \frac{\mathcal{Z}(d|\mathcal{H}_1)}{\mathcal{Z}(d|\mathcal{H}_2)}$$

Simple example: are the data more likely to contain a signal or noise?

Some comments: hypothesis testing

$$\text{“Bayes Factor”} = \frac{\mathcal{Z}(d|\mathcal{H}_1)}{\mathcal{Z}(d|\mathcal{H}_2)}$$

Simple example: are the data more likely to contain a signal or noise?

$$B = e^{169047}$$

So, very unlikely to be pure noise. For fun, see what happens when you add a very low-amplitude sine way into noisy data. It’s amazing how knowledge of noise statistics can detect a signal that’s impossible to see by eye

Some comments: hypothesis testing

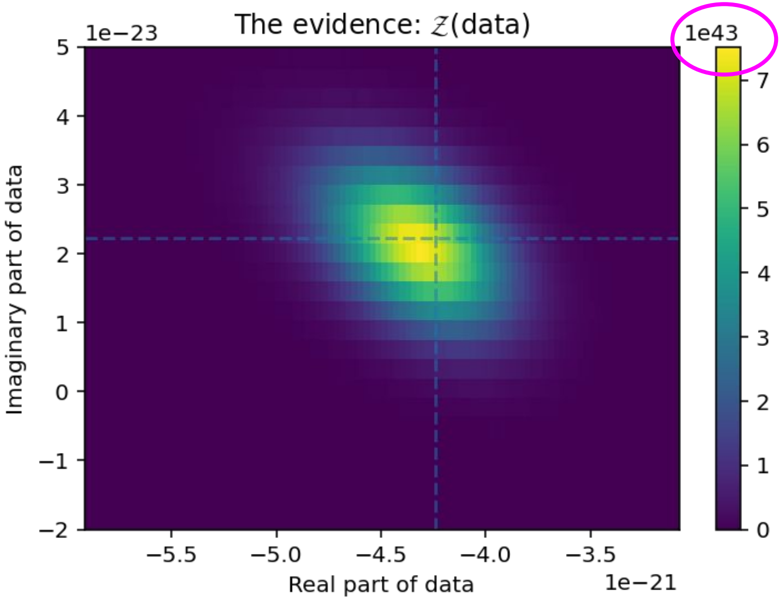
$$\text{“Bayes Factor”} = \frac{\mathcal{Z}(d|\mathcal{H}_1)}{\mathcal{Z}(d|\mathcal{H}_2)}$$

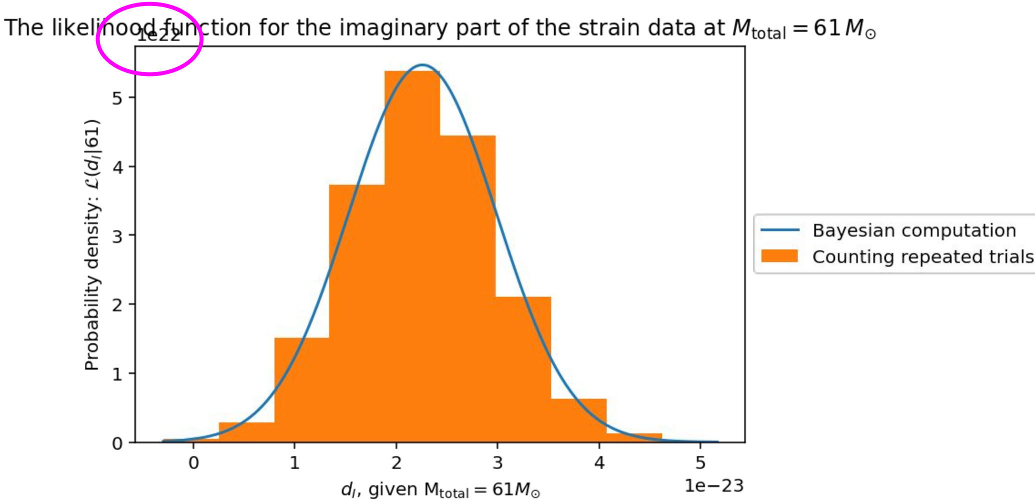
Not-so-simple examples:

- Is GR violated?
- Do black holes obey the no hair theorem?
- Is there a mass gap in the mass spectrum of black holes and neutron stars?
- Do neutron stars contain free quarks?

Some comments: numbers large and small

- The numerical values of the likelihoods, evidences and bayes factors may look absurd, but it’s a consequence of two factors

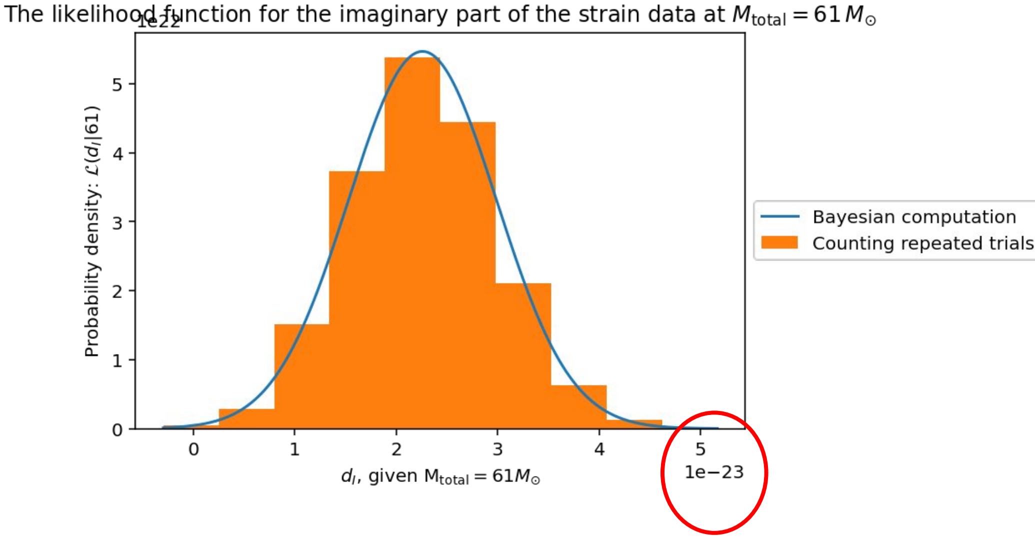




Some comments: numbers large and small

- The numerical values of the likelihoods, evidences and bayes factors may look absurd, but it's a consequence of two factors

1. The scale of the data is tiny! When you integrate the probability density of the data over all values you get 1. So the height of the probability density has to be roughly the inverse of the x-scale



Some comments: numbers large and small

- The numerical values of the likelihoods, evidences and bayes factors may look absurd, but it's a consequence of two factors

2. The dimensionality of the data plays a role when analysing real data, though it's not obvious here…

Exercise: draw a million random numbers (roughly the size of the data containing GW170817) from a Gaussian distribution with mean 0 and variance 1. Compute the (ln!) probability density of obtaining that sequence, and explain qualitatively why the *probability* (not density) of obtaining a sequence within a small neighbourhood should be absurdly small

Hint: would you expect to ever get that sequence again?