Panel: Model selection with gravitational wave observations

Strong Gravity and Binary dynamics, Oxford, MS
2017-02-27

Panel:
Richard O’Shaughnessy
Salvatore Vitale
Chris Pankow
Simon Stevenson
Rules of thumb

- Distributions more useful than rates
  - Scale tricky (IMF, all past SFR, distribution of conditions, many channels)
  - Functions can encode an infinite number of parameters [e.g., ROS PRD 2013]
  - At least one distribution (chirp mass) is easy to measure

- Robust observables are tricky
  - Cluster formation: strong spin misalignment = $\chi_{\text{eff}} < 0$
  - Mass gaps
    - (e.g., pair instability SN)

Dominik et al (2015: 1405.7016)

Rodriguez et al (2016: 1609.05916)
Do we need more proof of concept calculations?

- Several for discrete model selection or ad-hoc mixtures, but…
- When do we / how do we measure real parameters?

**Scope Of The Project**

- Pop-Synth simulations predict different component masses and merger time-delay of BBH starting from ZAMS
- Different realizations of unknown model parameters lead to significantly different distributions of the component mass pairs
- We demonstrate the proof-of-concept analysis with a very large suite (~ 1,200) of realistic pop-synth simulations with a few key phenomenological model parameters
- Estimated suitable mass-parameters ({m1, m2} or {Mc, q}) using either fully Bayesian (e.g., Lal-Inference) effective-Fisher method (e.g., with rapid-PE)

**Some Initial Results**

- We computed model evidences (un-normalized though!) of every single pop-synth model with the observed data set.
- Computed the Odds-ratios of various models having different values of pop-synth model parameters
- Looked for any plausible correlation of model evidences as the model parameters are varied.
- Found that winds parameters is very strongly correlated for GW150914 event (not surprising!!)
- A more detailed study is in progress.

This indicates, BBH events observed with LIGO/Virgo detectors can be extremely useful to accomplish constraint on the values of pop-synth model parameters.

Mukherjee, ROS et al (in prep) [based on ROS et al 2008,2010]
Reconstructing and reporting the observations

- Density estimation

Figure 6. Mean density inferred across mass space from mock observations using a binned distribution model with a Gaussian process prior for $N = 10, 20, 40$ (top row, left to right) and $80, 160, 400$ (bottom row, left to right) observations.

Figure 7. Water-filling clustering on the mean estimates of the population fraction in each bin, as inferred from 400 mock observations. Distinct NS-NS, NS-BH, and BH-BH clusters clearly appear around 40 – 80 observations, consistently with the estimated requirement of $\sim 60$ observations made by Mandel et al. (2015).

In order to identify specific clusters, we use a water-filling algorithm on the mean estimates of the population density in each bin (see, e.g., Nielsen & Nock 2008; Van & Pham-Gia 2010; Applegate et al. 2011, for other proposed approaches to distributional clustering). We gradually flood the posterior landscape until only three clusters stand above the water level over the $m_1 > m_2$ half of the plane. Clusters here are defined as sets of bins such that all elements of a cluster are connected through shared edges, but such connections do not exist between distinct clusters. Some of the posterior ends up in the under-water bins; the clustering is deemed successful only when under-water bins account for no more than a few percent of the posterior. This happens starting with $N = 80$ for the plots in Figure 6. As an example, Figure 7 shows the results of applying the water-filling clustering strategy to the distribution inferred from $N = 400$ observations (mirrored across $m_1 = m_2$ for plotting). In this case, the NS-NS, NS-BH, and BH-BH sub-populations contain 23%, 25%, and 51% of the population, respectively, while less than 2% of the posterior is under-water.

In general, the appropriate number of clusters does not need to be assumed in advance, but should be chosen from the data during the water-filling stage. Specifically, the amount of water used for flooding can be optimised against the flooded area. Flooding should continue only while the flooded area grows rapidly with a modest increase in the posterior volume (the amount of water used for flooding), with the remaining above-water areas identified as clusters.

We can obtain estimates of the statistical uncertainty on the inferred posterior fraction in each cluster by taking advantage of the full PDFs on the fractional mass distribution within each bin. We use the cluster boundaries provided by the water-filling clustering algorithm and compute the posterior on the total mass density within each cluster.
How many events to distinguish populations?

- KL divergence: unambiguous way to compute average information gain per event
  \[ D_{KL}(p|q) \equiv \int dx p \ln p/q \]

  - Standard tool in probability and statistics
  - Arbitrary dimensions / # of observables. Coordinate-system independent
  - Includes measurement error, selection bias (=apply to observed distribution)

  \[ \langle \hat{L}(X)/n \rangle = \langle \frac{\ln p(n|\mu)}{n} \rangle - [H_{p_*} + D_{KL}(p_*|p)] \]

  - Trivial to use for for toy models (e.g., power laws, gaussians, …)

- Hard part:
  - Evaluating & exploring the model space with sufficient accuracy
  - KL divergence is infinitely sensitive to gaps / exclusions, which are always decisive
  - As written, distinguishes two models (=points in hyperparameter space), not family
• Questions from Richard & audience
  
  • Systematics: The approximants are approximate. How do you build confidence in the result given uncertainty in strong merger?
    • What about NR (higher modes)? Precession? Uncertain high PN terms (tides?)
  
  • Calibration errors: How can we test GR or measure EOS in future instruments, given systematic amplitude and phase errors?
  
  • Dependence on parameters: What if tides / modified GR effects depend sensitively on nature of binary? How do we stack them?
  
  • Prior: past infinity or in band?
TIGER - caveats

- Odds in favor of modGR **not** necessarily equivalent to "GR is wrong"
- Could be that waveform model is inappropriate to start with
- Something weird with the data or calibration
- Unaccounted (GR) physics
  - E.g. non-linear NS tides (Essick+ 2016)
- Priors on GR parameters (?)
- Most of these effects shown to be under control in Agathos+ 2013
Measuring the mixture fraction

Posterior distribution for mixture parameter

100 NSBH
200 BBH

True underlying fraction of aligned sources

S. Vitale, Aug 26 2016
Caveats – To dos

• Assumed what I called "aligned” is what the universe calls aligned – should include possible prior mismatch
• Can extend the model so that they also take into account mass ratios, eccentricity, or anything else that might be useful to distinguish
• Can include more than 2 models
Chris Pankow

• Questions from Richard and audience
  • Does reweighting posteriors work?

  • How do we deal with selection bias of real searches against interesting things (e.g., precessing; modified GR; …)
• Questions from Richard and audience
  
  • Joint constraints: How can you do multi-observation constraints with an interpolated model? Interpolate all observations?

  • [Technical] How does interpolation work safely and with high contrast? Basis functions for log(rate)?

  • [Technical] Are you also interpolating observable universe (selection bias-selected) or full universe (including distribution of conditions and z)
Distinguishing a discrete model set straightforward

but this is driven by large rate differences. Rate is highly degenerate with other factors…
Distinguishing a discrete model set straightforward

- Mass distributions alone are more similar, given measurement error

**O2-scale, as before**

**O2-scale, no rate info**
Bayesian Model Selection

- GW PE: (mostly) straightforward application of Bayes’ Law — posterior distribution on binary parameters derived from (mostly uninformative, but astrophysically motivated priors) and influenced through the data + waveform model through the likelihood ratio
  
- Obtain a set of samples of physical parameters of interest: chirp mass ($M_c$), mass ratio ($q$), spin orientations and magnitudes ($s_1$, $s_2$), and at some point probably eccentricity (not addressed here)

- Question: **Given a set of plausible astrophysical formation channels, how do we select a model resembling nature as well as quantify any parameters of that model?**

- Need to map $\{M_c, q, s_1, s_2\}$ to mass/spin spectrums, progenitor metallicity, SN kick prescriptions, evolutionary pathways, etc…
Bayesian Hierarchical Modeling

- Foreman-Mackey, et al. 2014 lays out the foundation

- convert $p(\text{mod}|\text{obs}) \rightarrow p(\text{mod}|\text{PE})$

$$p(\{h_i\} | \beta) = \prod_i p(h_i) \int \frac{p(\theta | h_i) p(\theta | \beta)}{p(\theta)} d\theta$$

- Integral over model parameters ($\beta$) can be evaluated via importance sampling using parameter estimation ($\theta_k$) samples

$$\rightarrow p(\beta | \{h_i\}) \propto \prod_i \frac{1}{N} \sum_k \frac{p(\theta_k | \beta)}{p(\theta_k)} p(\beta)$$

- Recasts the problem as a “higher level” parameterization with no dependence on original data $\{h_i\}$
Beyond Two Parameter Models

• Are kick *direction* prescriptions *(isotropic / polar)* measurable at the level of mass spectrums?

  • **Spoilers:** *No*. Most mass spectrums are degenerate, and spins (Stevenson, et al. 2017, Rodriguez, et al. 2016) are required

\[ N_{\text{obs}} = 10000 \]
Stevenson
Richard

- Slides from KITP talk, 2016
Familiar statistical challenge

• Inference via Poisson likelihood + bayes

\[ L(\Lambda) = e^{-\mu} \frac{\mu^n}{n!} \prod_k \int d\lambda_k p(d_k|\lambda_k) p(\lambda_k|\Lambda) \]

• Same likelihood for nonparametric, parametric, and physical models
  • \( \mu \) expected n (selection bias)
  • \( p(d_k|\lambda_k) \) measurements and error
  • \( p(\lambda_k|\Lambda) \) binary parameter distribution, given model parameters

• Informal approaches: weighted histograms (=gaussian mixture models)

Ivezic et al, *Statistics, data mining, and machine learning in astronomy*
Gregory and Loredo (discrete photon light curves)

ROS PRD 2013
Hogg and Bovy
W. Farr, LIGO LIGO-T1600562; Mandel, Farr, Gair LIGO-P1600187
ROS LIGO T1600208
Confronting theory with observations

A function has infinitely many degrees of freedom


Abbott et al. O1 BBH (1606.04856)
Distributions vary significantly...

**Fig. 6.** Compact BH-NS binaries visible by advanced LIGO: Properties of the BH-NS binaries detectable by a single advanced LIGO instrument, scaled in proportion to their detection probability. Different color and line styles indicate results for different binary evolution models: high BH kicks (blue); delayed SN (green); our standard model (black); and optimistic CE (red). The top and bottom panels show the distribution of birth time $t_{\text{birth}}$, and chirp mass, respectively. Though our simulations use a discrete array of metallicity bins, to guide the eye, their relative contributions have been joined by solid lines; this histogram makes no correction for the density of metallicity bins.

**Fig. 7.** BH-BH binaries visible by advanced LIGO: Properties of BH-BH binaries detectable by a single advanced LIGO instrument, scaled in proportion to their detection probability. Different color and line styles indicate results for different binary evolution models: high BH kicks (blue); delayed SN (green); our standard model (black); and optimistic CE (red). Unlike NS-NS binaries, the detected population of BH-BH binaries was preferentially formed in the early universe over a wide range of metallicities. Many detectable BH-BH binaries have high chirp mass and form at significantly subsolar metallicities.

$$M_c = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}$$

Detected distribution

Dominik et al (2015: 1405.7016)
Distributions vary significantly...

\[ \mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \]

(Detected distribution)  Dominik et al (2015: 1405.7016)
...and for physical reasons, like pair instability

![Graph showing chirp-mass distribution and metallicity dependence.](image)
...or multiple mergers and single star evolution

Nuclear Star Clusters
Antonini and Rasio 2016

Fig. 6.—Same as Figure 5 but for $Z=0$. NSCs and GCs obtained from our models. NSCs are defined here as clusters with masses in the range $5 \times 10^5$ to $5 \times 10^7 M_\odot$, while GCs have masses in the range $10^5$ to $10^7 M_\odot$. We assumed that all clusters formed 12 Gyr ago regardless of their mass and consider two values of metallicities, $Z=0.01$ and $0.25 Z_\odot$. In order to obtain the mean rate of mergers we weighted the number of mergers from each of the cluster models by a cluster initial mass function (CIMF). For GCs we assume a power law CIMF: $dM/dN \propto M^{-2}$ (e.g., Bik et al. 2003). For NSCs the initial mass function is largely unknown. Here we take the IMF of NSCs directly from the mass distribution of NSCs at $z=2$ from the galaxy formation models of Antonini et al. (2015b) (their Figure 10). These models produce a mass distribution at $z=0$ that is consistent with the observed NSC mass distribution from Georgiev et al. (2016). We note that here we might be underestimating the number of massive mergers from NSCs occurring at low redshift because we have assumed that these systems are as old as Galactic GCs. In fact, while most NSCs appear to be dominated by old stellar components they are also known to have a complex star formation history and to contain young stellar populations which can produce high mass mergers also at later times (we will come back to this point below). It is also possible that a large fraction of the NSC stars accumulated gradually in time by infalling globular clusters that decayed to the center through dynamical friction. If this process is the main mechanism for NSC formation, then NSCs and GCs will comprise similar stellar populations (Antonini 2014).

Table 1 shows that our models predict a few thousands BH mergers per NSC over 12 Gyr of evolution. This expectation also appears to be consistent with previous estimates (Portegies Zwart & McMillan 2000; Miller & Lauburg 2009). In addition, NSCs produce between 50 to $\approx 500$ BH mergers with high mass $>50 M_\odot$ at $z<0.3$ depending on the BH spin magnitudes and assumed metallicities distribution of the underlying stellar population. Our GC models produce only a few mergers per cluster within $z<0.3$ and total mass $>50 M_\odot$. These massive binaries are found to form only in the most massive GCs ($M_{cl}\gtrsim10^6 M_\odot$). The number of massive mergers at low redshift is also sensitive to the spin magnitude distribution we assume. For high spin models, a smaller number of BHs are retained in the clusters compared to the uniform spin models. Consequently, high spin models produce fewer high mass BH mergers at low redshift compared to models that assume low spins. However, in either spin models a number of inspiraling BH binaries with mass $\lesssim50 M_\odot$ is found to merge at low redshift. Finally, Table 1 gives the number of BH mergers that are retained inside the cluster. Between 10 and 20 percent of high mass ($>50 M_\odot$) mergers occurring in NSCs at $z<1$ are retained inside.
...that may be observationally accessible soon

![Graph showing merger rate density as a function of total redshifted binary mass](image)
but this is driven by large rate differences. Rate is highly degenerate with other factors…

Stevenson, Ohme, Fairhurst (1504.07802), based on Dominik et al 2012
See also Miyamoto et al, GWPAW 2016; Dhani, Muverjee et al 2016 (LVC meeting)
Distinguishing a discrete model set straightforward

- Mass distributions alone are more similar, given measurement error

**O2-scale, as before**

**O2-scale, no rate info**
Beyond the mass distribution: Power of spin

- High mass binaries may be strictly and positively aligned (fallback)
- Low spins required for GW150914...possible? [Kushnir et al]
  - Tells us something about how massive stars evolve? About tides?
  - Or favors dynamics?

---

**FIG. 7.** Beyond the mass distribution: Power of spin

- Marginalized log likelihoods for various iron masses, showing a peak around 0.8. Points with $\log(M_1 + M_2) > 8$ are shown in black, and with $\log(M_1 + M_2) < 8$. Each point represents a different starting configuration, where $Z_\odot$, $Z_\odot/10$, and $Z_\odot/50$ are included. Points with $Z_\odot/50$ may retain enough angular momentum in the core to avoid a supernova kick. The effects of geometry are expected, given both our self-imposed restrictions and the magnitude and direction of either BHs spin cannot be significantly constrained by our method.

- Our inability to determine the most likely transverse spin of the nonprecessing prediction, nor independent of rotations of precessing spins are consistent with the GW150914: many configurations have different initial masses and spins.

- Not all precessing simulations with suitable initial spins about the initial orbital angular momentum by the amount of available angular momentum, we expect many of the configurations have not within 10 of the peak; see the right panel of Fig. 8) The effects of geometry. For example, the lack of apparent modulation in the signal references between the results reported here and comparable parts of the posterior in the bottom right quadrant of Figure 7 in the plane perpendicular to the angular momentum axis.

- Our inability to determine the most likely transverse spin of the nonprecessing prediction, nor independent of rotations of precessing spins are consistent with the GW150914: many configurations have not within 10 of the peak; see the right panel of Fig. 8) The effects of geometry. For example, the lack of apparent modulation in the signal references between the results reported here and comparable parts of the posterior in the bottom right quadrant of Figure 7 in the plane perpendicular to the angular momentum axis.

- Our inability to determine the most likely transverse spin of the nonprecessing prediction, nor independent of rotations of precessing spins are consistent with the GW150914: many configurations have not within 10 of the peak; see the right panel of Fig. 8) The effects of geometry. For example, the lack of apparent modulation in the signal references between the results reported here and comparable parts of the posterior in the bottom right quadrant of Figure 7 in the plane perpendicular to the angular momentum axis.

---

**FIG. 8.** Angular-momentum profiles at core helium depletion for the non-rotating (Schwarzschild) and critically rotating (Kerr) black hole are also included. That said, comparisons conducted under similar limitations form models, but it also adopts a slightly different starting orbital angular momentum by the amount of available angular momentum, we expect many of the configurations have not within 10 of the peak; see the right panel of Fig. 8) The effects of geometry. For example, the lack of apparent modulation in the signal references between the results reported here and comparable parts of the posterior in the bottom right quadrant of Figure 7 in the plane perpendicular to the angular momentum axis.

- Our inability to determine the most likely transverse spin of the nonprecessing prediction, nor independent of rotations of precessing spins are consistent with the GW150914: many configurations have not within 10 of the peak; see the right panel of Fig. 8) The effects of geometry. For example, the lack of apparent modulation in the signal references between the results reported here and comparable parts of the posterior in the bottom right quadrant of Figure 7 in the plane perpendicular to the angular momentum axis.

---

**FIG. 9.** Marginalized log likelihoods for various iron masses, showing a peak around 0.8. Points with $\log(M_1 + M_2) > 8$ are shown in black, and with $\log(M_1 + M_2) < 8$. Each point represents a different starting configuration, where $Z_\odot$, $Z_\odot/10$, and $Z_\odot/50$ are included. Points with $Z_\odot/50$ may retain enough angular momentum in the core to avoid a supernova kick. The effects of geometry are expected, given both our self-imposed restrictions and the magnitude and direction of either BHs spin cannot be significantly constrained by our method.

- Our inability to determine the most likely transverse spin of the nonprecessing prediction, nor independent of rotations of precessing spins are consistent with the GW150914: many configurations have not within 10 of the peak; see the right panel of Fig. 8) The effects of geometry. For example, the lack of apparent modulation in the signal references between the results reported here and comparable parts of the posterior in the bottom right quadrant of Figure 7 in the plane perpendicular to the angular momentum axis.

---

**FIG. 10.** Angular-momentum profiles at core helium depletion for the non-rotating (Schwarzschild) and critically rotating (Kerr) black hole are also included. That said, comparisons conducted under similar limitations form models, but it also adopts a slightly different starting orbital angular momentum by the amount of available angular momentum, we expect many of the configurations have not within 10 of the peak; see the right panel of Fig. 8) The effects of geometry. For example, the lack of apparent modulation in the signal references between the results reported here and comparable parts of the posterior in the bottom right quadrant of Figure 7 in the plane perpendicular to the angular momentum axis.

- Our inability to determine the most likely transverse spin of the nonprecessing prediction, nor independent of rotations of precessing spins are consistent with the GW150914: many configurations have not within 10 of the peak; see the right panel of Fig. 8) The effects of geometry. For example, the lack of apparent modulation in the signal references between the results reported here and comparable parts of the posterior in the bottom right quadrant of Figure 7 in the plane perpendicular to the angular momentum axis.

---

**Marchant et al A&A 2016 (1601.03718)**