Compact binary coalescence:

Bayesian model selection and parameter estimation

Chris Van Den Broeck

Nikhef – National Institute for Subatomic Physics
Amsterdam, The Netherlands

SUSSP73: Gravitational Wave Astronomy
23 July – 5 August 2017, University of St Andrews, UK

Bayesian inference

Aim: use available data to

- Construct probability density distributions for parameters associated with these hypotheses
  → Parameter estimation
- Evaluate which out of several hypotheses is the most likely
  → Model selection

Do this while making explicit all extraneous assumptions

Inductive logic

- Propositions (i.e. statements, events) denoted by uppercase letters, e.g. A, B, C, …, X
- Boolean algebra:
  - Conjuction: *A and B are both true*

    $AB \quad \text{or} \quad A \wedge B$

  - Disjunction: *At least one of A or B is true*

    $A + B \quad \text{or} \quad A \vee B$

  - Negation: *A is false*

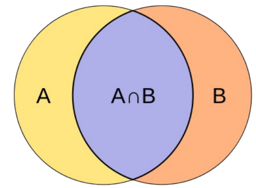
    $\bar{A} \quad \text{or} \quad \neg A$

  - Implication: *From A follows B*

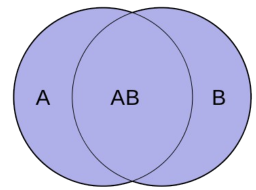
    $A \rightarrow B \quad \text{or} \quad A \Rightarrow B$

Probabilities for propositions

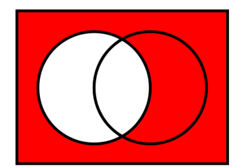
- Useful to view statements as sets which are subsets of a "Universe"
  - Conjunction: intersection of sets
    $AB \quad \text{or} \quad A \wedge B$



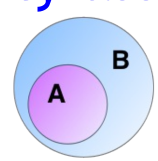
  - Disjunction: union of sets
    $A + B \quad \text{or} \quad A \vee B$



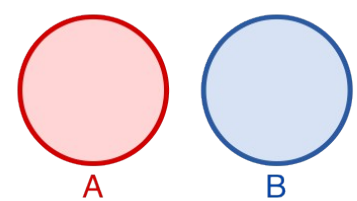
  - Negation: complement within Universe
    $\bar{A} \quad \text{or} \quad \neg A$



- Each of these sets have a probability associated with them
  - If A ⊂ B then $p(A) \leq p(B)$



  - If A and B disjoint then $p(A \cup B) = p(A) + p(B)$



    - The Universe has probability 1, so that e.g.
      $p(A) + p(\bar{A}) = 1$

Conditional probability

- *Conditional probability*: $p(A|B) \equiv \frac{p(A \cap B)}{p(B)}$
  - *Product rule*:

    $$p(A, B) = p(A \cap B) = p(A|B)\, p(B)$$

- It is customary to explicitly denote probabilities being conditional on "all background information we have": $p(A|I)$, $p(B|I)$, ...
  - All essential formulae unaffected, e.g. product rule:

    $$p(A, B|I) = p(A|B, I) P(B|I)$$

- From the product rule follows Bayes' theorem:

$$\boxed{p(A|B, I) = \frac{p(B|A, I)\, p(A|I)}{p(B|I)}}$$

Marginalization

- Note that for any A and B,

$$A \cap B \quad \text{and} \quad A \cap \bar{B}$$

are disjoint sets whose union is A, so

$$p(A|I) = p(A, B|I) + p(A, \bar{B}|I)$$

- Consider sets $\{B_k\}$ such that
  - They are disjoint: $B_k \cap B_l = \emptyset, \quad k \neq l$
  - They are exhaustive: $\cup_k B_k$ is the Universe, or $\sum_k p(B_k|I) = 1$

Then

$$\boxed{p(A|I) = \sum_k p(A, B_k|I)}$$

Marginalization rule

Marginalization over a continuous variable

- Consider the proposition

  "The continuous variable *x* has the value α"

  Then not necessarily a well-defined meaning of probability $p(x = \alpha | I)$

- Instead assign probabilities to finite intervals:

$$p(x_1 \leq x \leq x_2 | I) = \int_{x_1}^{x_2} \mathrm{pdf}(x) dx$$

  where "pdf" is the *probability density function*

  – Exhaustiveness written as

$$\int_{x_{\min}}^{x_{\max}} \mathrm{pdf}(x) dx = 1$$

- Marginalization for continuous variables:

$$\boxed{p(A) = \int_{x_{\min}}^{x_{\max}} \mathrm{pdf}(A, x) dx}$$

Parameter estimation

- Experiment performed, data *d* collected
- Parameter θ being measured
- Consider a model *H* that allows to calculate probability of getting data d if parameter θ is known ("*generative model*")
  - Can calculate the *likelihood* $p(d|\theta, H, I)$
- What is wanted is instead posterior probability of θ, $p(\theta|d, H, I)$
- Use Bayes' theorem:

$$p(\theta|d, H, I) = \frac{p(d|\theta, H, I)p(\theta|H, I)}{p(d|H, I)}$$

  - "*Prior*" $p(\theta|H, I)$ is our knowledge of θ before experiment
  - "Evidence" $p(d|H, I)$ doesn't depend on θ, ignore for now

$$\boxed{p(\theta|d, H, I) \propto p(d|\theta, H, I)p(\theta|H, I)}$$

Parameter estimation

$$\boxed{p(\theta|d,H,I) \propto p(d|\theta,H,I)p(\theta|H,I)}$$

- Posterior is likelihood weighted by prior

  Conclusions drawn based on:

  - Information available before experiment
  - Experimental data obtained

- Can extend to more parameters: joint posterior $p(\theta_1,\ldots,\theta_N|d,H,I)$
- If we want posterior distribution just for variable θ_1, $p(\theta_1|d,H,I)$,

  then we marginalize:

$$p(\theta_1|d,H,I) = \int_{\theta_2^{\min}}^{\theta_2^{\max}} \ldots \int_{\theta_N^{\min}}^{\theta_N^{\max}} p(\theta_1,\ldots,\theta_N|d,H,I)\, d\theta_2 \ldots d\theta_N$$

- Mean of a 1D posterior:

$$\mu = E[\theta] = \int_{\theta^{\min}}^{\theta^{\max}} \theta\, p(\theta|d,H,I)\, d\theta$$

- Variance of a 1D posterior:

$$\sigma^2 = E\left[(\theta-\mu)^2\right] = \int_{\theta^{\min}}^{\theta^{\max}} (\theta-\mu)^2\, p(\theta|d,H,I)\, d\theta$$

- Means for *N* variables:

$$\mu_i = E[\theta_i] = \int_{\theta_1^{\min}}^{\theta_1^{\max}} \ldots \int_{\theta_N^{\min}}^{\theta_N^{\max}} \theta_i\, p(\theta_1,\ldots,\theta_N|d,H,I)\, d\theta_1 \ldots d\theta_N$$

- Covariance matrix:

$$\Sigma_{ij} \equiv E\left[(\theta_i-\mu_i)(\theta_j-\mu_j)\right] = \int_{\theta_1^{\min}}^{\theta_1^{\max}} \ldots \int_{\theta_N^{\min}}^{\theta_N^{\max}} (\theta_i-\mu_i)(\theta_j-\mu_j)\, p(\theta_1,\ldots,\theta_N|d,H,I)\, d\theta_1 \ldots d\theta_N$$

- *Confidence interval* is the smallest interval within whose limits a fraction γ of the posterior is contained:

$$\gamma = \int_{\theta^{\mathrm{lo}}}^{\theta^{\mathrm{hi}}} p(\theta|d,H,I)d\theta$$

where $\theta^{\mathrm{hi}} - \theta^{\mathrm{lo}}$ is minimal

- In most literature γ is taken to be 0.68 or 0.95, roughly corresponding to 1-sigma and 2-sigma intervals of Gaussian distribution

- Multi-dimensional confidence intervals:

$$\gamma_{\theta_1} = \int_{\theta_1^{\mathrm{lo}}}^{\theta_1^{\mathrm{hi}}} p(\theta_1|d,H,I)d\theta_1$$
$$= \int_{\theta_1^{\mathrm{lo}}}^{\theta_1^{\mathrm{hi}}} \int_{\theta_2^{\mathrm{min}}}^{\theta_2^{\mathrm{max}}} \ldots \int_{\theta_N^{\mathrm{min}}}^{\theta_N^{\mathrm{max}}} p(\theta_1,\ldots,\theta_N|d,H,I)d\theta_1\ldots d\theta_N$$

Hypothesis testing

- Estimating parameters is possible if generative model known
- If we want to compare possible generative models, e.g. X, Y: calculate posterior probabilities $p(X|d,I)$ and $p(Y|d,I)$
- Bayes' theorem:

$$p(X|d,I) = \frac{p(d|X,I)p(X|I)}{p(d|I)}$$

- Compute *odds ratio*

$$O_Y^X \equiv \frac{p(X|d,I)}{p(Y|d,I)}$$
$$= \frac{p(d|X,I)}{p(d|Y,I)} \frac{p(X|I)}{p(Y|I)}$$

where factors of $p(d|I)$ have canceled out

- $p(X|I)/p(Y|I)$ ratio of prior odds
- $p(d|X,I)/p(d|Y,I)$ ratio of evidences, or *Bayes factor* $B_Y^X = \frac{p(d|X,I)}{p(d|Y,I)}$

Hypothesis testing

- Hypotheses usually have parameters associated with them
- Bayes theorem relating posterior to likelihood:

$$p(\theta|d,H,I) = \frac{p(d|\theta,H,I)p(\theta|H,I)}{p(d|H,I)}$$

or

$$p(\theta|d,H,I)p(d|H,I) = p(d|\theta,H,I)p(\theta|H,I)$$

- Marginalize both sides over parameter(s):

$$\int p(\theta|d,H,I)p(d|H,I)d\theta = \int p(d|\theta,H,I)p(\theta|H,I)d\theta$$

Note that $p(d|H,I)$ independent of parameter(s), and posterior $p(\theta|d,H,I)$ normalized by definition, hence left hand side:

$$\int p(\theta|d,H,I)p(d|H,I)d\theta = p(d|H,I)\int p(\theta|d,H,I)d\theta = p(d|H,I)$$

Therefore evidence is given by $\boxed{p(d|H,I) = \int p(d|\theta,H,I)p(\theta|H,I)d\theta}$

Hypothesis testing

- Odds ratio

$$O_Y^X \equiv \frac{p(X|d,I)}{p(Y|d,I)}$$
$$= \frac{p(d|X,I)}{p(d|Y,I)} \frac{p(X|I)}{p(Y|I)}$$

Bayes factor

$$B_Y^X = \frac{p(d|X,I)}{p(d|Y,I)}$$

Marginalized evidences e.g. $p(d|X,I) = \int p(d|\theta,X,I)p(\theta|X,I)d\theta$

- Hypotheses can have arbitrary number of free parameters
  - Does model that fits data the best give the highest evidence?
  - If so, model with more parameters would give highest evidence even if incorrect!

Occam's razor

- For simplicity, compare two generative hypotheses:
  - *X* has no free parameters
  - *Y* has one free parameter, λ

Will Y automatically be favored over X?

- Odds ratio $O_Y^X = \dfrac{p(d|X,I)}{p(d|Y,I)} \dfrac{p(X|I)}{p(Y|I)}$
- Evidence for *X* is straightforward, but for *Y*:

$$p(d|Y,I) = \int p(d|\lambda,Y,I)p(\lambda|Y,I)d\lambda$$

Assume flat prior for $\lambda \in [\lambda_{\text{min}}, \lambda_{\text{max}}]$:

$$p(\lambda|Y,I) = \frac{1}{\lambda_{\text{max}} - \lambda_{\text{min}}}, \quad \text{for } \lambda_{\text{min}} \leq \lambda \leq \lambda_{\text{max}}$$

Occam's razor

- Evidence for *Y*:

$$p(d|Y,I) = \int p(d|\lambda,Y,I)p(\lambda|Y,I)d\lambda$$

- Flat prior:

$$p(\lambda|Y,I) = \frac{1}{\lambda_{\max} - \lambda_{\min}}, \quad \text{for } \lambda_{\min} \le \lambda \le \lambda_{\max}$$

- For definiteness, assume likelihood of the form

$$p(d|\lambda,Y,I) = p(d|\lambda_0,Y,I)\exp\left[-\frac{(\lambda-\lambda_0)^2}{2\sigma_\lambda^2}\right]$$

- Evidence for *Y*:

$$\begin{aligned}
p(d|Y,I) &= \int p(d|\lambda,Y,I)p(\lambda|Y,I)d\lambda \\
&= \int \frac{1}{\lambda_{\max} - \lambda_{\min}} p(d|\lambda_0,Y,I)\exp\left[-\frac{(\lambda-\lambda_0)^2}{2\sigma_\lambda^2}\right] d\lambda \\
&= \frac{p(d|\lambda_0,Y,I)}{\lambda_{\max} - \lambda_{\min}} \int \exp\left[-\frac{(\lambda-\lambda_0)^2}{2\sigma_\lambda^2}\right] d\lambda \\
&= p(d|\lambda_0,Y,I)\frac{\sigma_\lambda\sqrt{2\pi}}{\lambda_{\max} - \lambda_{\min}}.
\end{aligned}$$

Occam's razor

- Evidence for *Y*:

$$p(d|Y,I) = p(d|\lambda_0, Y, I)\frac{\sigma_\lambda\sqrt{2\pi}}{\lambda_{\max} - \lambda_{\min}}$$

- Hence odds ratio becomes:

$$O_Y^X = \frac{p(X|I)}{p(Y|I)}\frac{p(d|X,I)}{p(d|\lambda_0,Y,I)}\frac{\lambda_{\max} - \lambda_{\min}}{\sigma_\lambda\sqrt{2\pi}}$$

where

- $p(X|I)/p(Y|I)$ ratio of prior odds; can be set to 1 in this example
- $p(d|X,I)/p(d|\lambda_0,Y,I)$ just compares best fits; will usually be < 1
- $(\lambda_{\max} - \lambda_{\min})/(\sigma_\lambda\sqrt{2\pi})$ *penalizes Y if experimental uncertainty on* λ *much smaller than prior range*
  - Will tend to be the case if λ not needed!

Occam's Razor:

"It is vain to do with more what can be done with fewer"

Nested sampling

- Parameter estimation requires computing the posterior density distribution from likelihood and prior using Bayes' theorem:

$$p(\boldsymbol{\theta}|d,H,I) = \frac{p(d|\boldsymbol{\theta},H,I)p(\boldsymbol{\theta}|H,I)}{p(d|H,I)}$$

- Often the parameter space has high dimensionality (e.g. 15 for quasi-circular binary inspiral), making it computationally challenging to map out the likelihood
- Similarly calculation of evidence integral over high-dimensional space:

$$\begin{aligned} p(d|H,I) &= \int d^N\boldsymbol{\theta}\; p(d|\boldsymbol{\theta},H,I)p(\boldsymbol{\theta}|H,I) \\ &= \int d^N\boldsymbol{\theta}\; L(\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \end{aligned}$$

- Efficient way of obtaining both: *nested sampling*

Nested sampling: basic idea

$$p(d|H,I) = \int d^N\boldsymbol{\theta}\; p(d|\boldsymbol{\theta},H,I)p(\boldsymbol{\theta}|H,I)$$
$$= \int d^N\boldsymbol{\theta}\; L(\boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

- Nested sampling computes the evidence by rewriting the above integration in terms of a single scalar called *prior mass X*
- "Fraction of volume with likelihood greater than λ" Mathematically:

$$X(\lambda) \equiv \int\int \cdots \int_{L(\boldsymbol{\theta})>\lambda} \pi(\boldsymbol{\theta}) d^N\boldsymbol{\theta}$$

Element of prior mass: $dX = \pi(\boldsymbol{\theta})d^N\boldsymbol{\theta}$

- Since prior is normalized, $X \in [0,1]$
  - Lower bound $X = 0$:

    surface within which no higher likelihood; $\lambda = L_{\max}$
  - Upper bound $X = 1$:

    surface within which all points higher likelihood; $\lambda = L_{\min}$

Nested sampling: basic idea

$$p(d|H,I) = \int d^N\boldsymbol{\theta}\; p(d|\boldsymbol{\theta},H,I)p(\boldsymbol{\theta}|H,I)$$
$$= \int d^N\boldsymbol{\theta}\; L(\boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

- Rewrite as

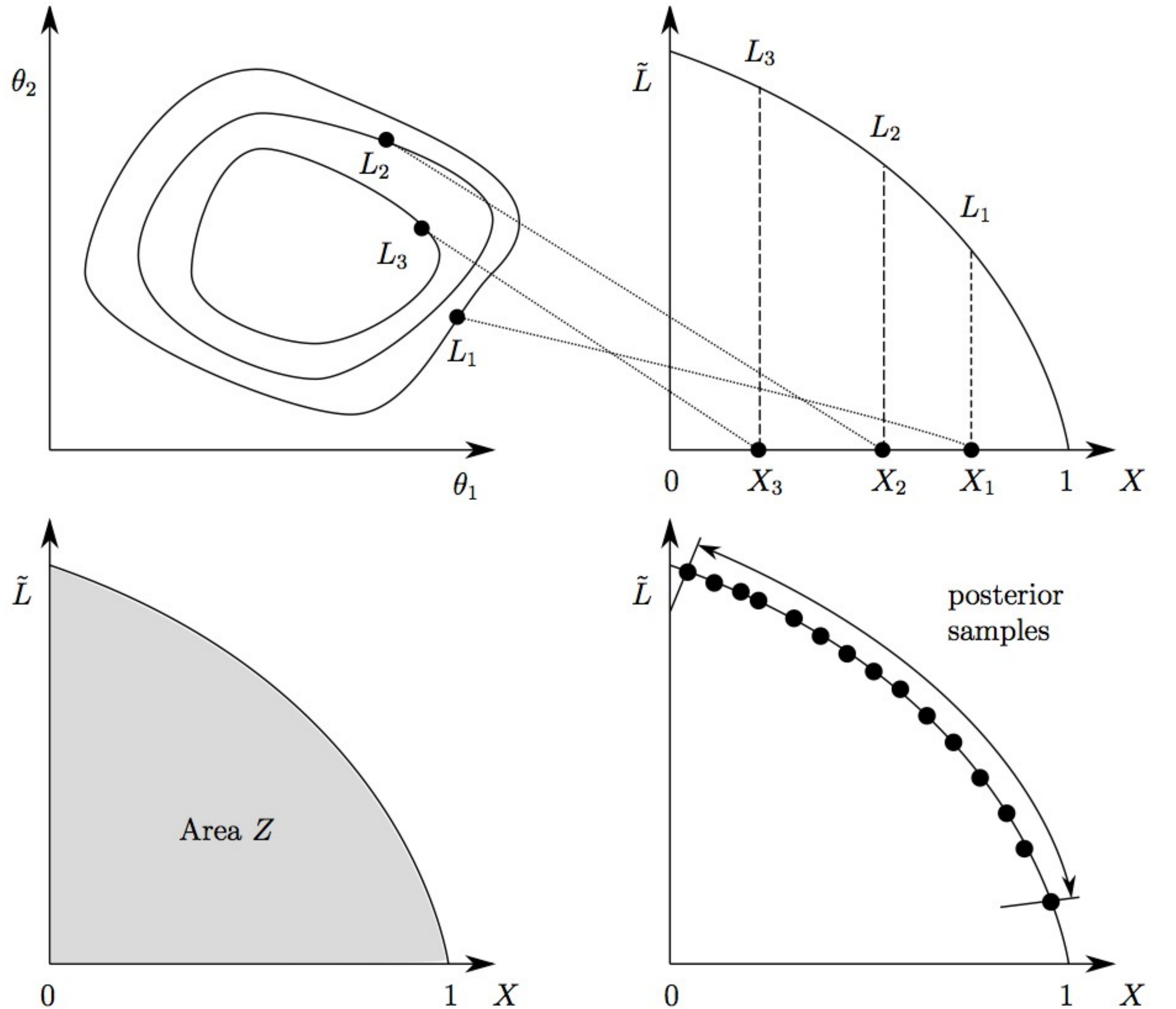
$$Z = \int\int\cdots\int L(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d^N\boldsymbol{\theta}$$
$$= \int \tilde{L}(X)dX.$$

- Posterior obtained trivially from

$$\tilde{P}(X) = \frac{\tilde{L}(X)}{Z}$$

- Idea behind nested sampling: construct the function $\tilde{L}(X)$ by progressively finding locations in parameter space with higher likelihood and associated progressively smaller prior mass
  - Then use above formulae for evidence, posterior

Nested sampling: schematically



Nested sampling: the algorithm

- Drop *M* samples across parameter space, sampled from the prior
  These are called "live points"
  - Each has likelihood associated with it
  - Associated with volume s.t. likelihood lowest at the surface
  - Uniformly sampled in prior mass between 0 and 1
- Discard live point with lowest likelihood L_0, i.e. highest prior mass X_0
  - Replace by new live point, sampled from the prior, which has higher likelihood
  - New point with lowest likelihood L_1 must have $X_1 < X_0$
  - *Statistically* assign value for X_1
- Repeat the step above

Nested sampling: the algorithm

- Having discarded the old lowest-likelihood point with prior mass X_0, how do we statistically assign a prior mass X_1 to the new lowest-likelihood point?
- Probability that the surface with highest prior mass is at $X = \chi$ is joint probability that none of the samples have prior mass $> \chi$

$$P(X_i < \chi) = \prod_{i=1}^{M} \int_0^{\chi} dX_i = \prod_{i=1}^{M} \chi = \chi^M$$

- Probability *density* that highest of *M* samples has prior mass χ

$$P(\chi, M) = M\chi^{M-1}$$

- Define *shrinkage ratio* between new and old highest prior mass:

$$t = X_1/X_0$$

  This has same probability density:

$$P(t, M) = Mt^{M-1}$$

- Hence we assign X_1 by drawing a shrinkage ratio from the above distribution

Nested sampling

- At first step: set $X = 1$
- At k^{th} iteration: live point with largest prior mass has

$$X_k = \prod_{j=1}^{k} t_k$$

- Recall distribution of shrinkage ratios:

$$P(t, M) = M t^{M-1}$$

Mean and standard deviation of $\log(t)$:

$$\log t = (-1 \pm 1)/M$$

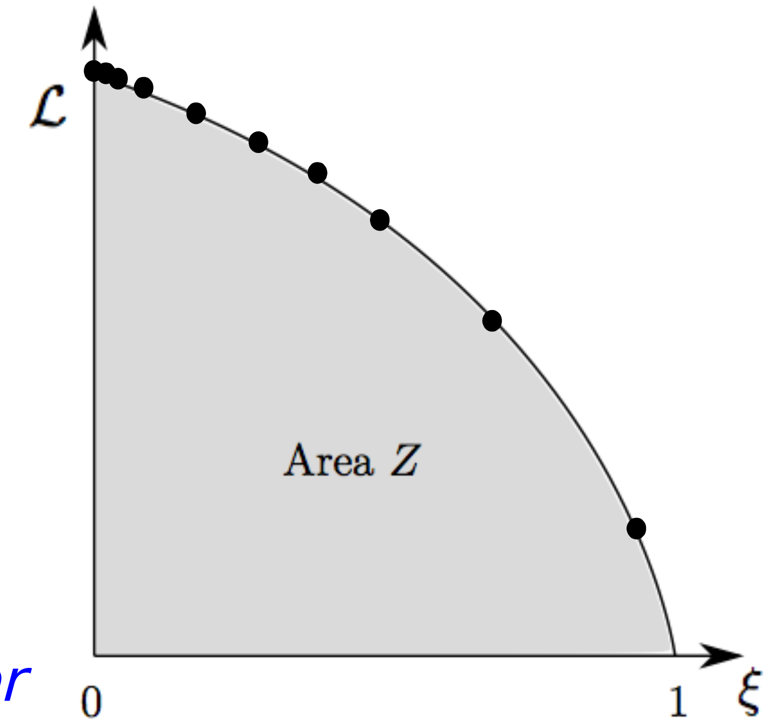
- Hence $\log(X_k)$ has mean and stdev

$$\log X_k = (k \pm \sqrt{k})/M$$

Hence mean values go like

$$X_k = \exp(-k/M)$$

- *Very quickly reaches prior mass where likelihood is largest*
- *Errors decrease exponentially*
- *Larger number of live points is better*



Nested sampling: termination condition

- No obvious choice for ending the sampling process
  - Use practical guidelines

- Estimate *information* as function of evidence and likelihood:

$$\mathcal{H} = \int P(X) \ln\left(P(X)\right) dX$$
$$\approx \sum_k \frac{L_k}{Z} \ln \frac{L_k}{Z} \Delta X_k,$$

Terminate when $X = e^{-\mathcal{H}}$

- Or, can estimate amount of evidence yet to be accumulated and compare with evidence already accumulated

Terminate when $L_{\max} X_{\mathrm{cur}} < \alpha Z_{\mathrm{cur}}$ where α is user-specified

Nested sampling: accuracy

- Take termination condition

$$X = e^{-\mathcal{H}}$$

- Means go like

$$X_k = \exp(-k/M)$$

"Terminate when count *k* exceeds $M\mathcal{H}$"

- Evidence:

$$Z = \int \tilde{L}(X)dX \approx \sum_k L_k \Delta X_k$$

Recall

$$\log X_k = (k \pm \sqrt{k})/M$$

Hence uncertainty on the evidence:

$$\Delta \log Z = \sqrt{\frac{\mathcal{H}}{M}}$$

- In gravitational-wave applications, with a few thousand live points this is typically $O(10^{-1})$ whereas for detectable signal $\log Z = O(10^2)$

Application to gravitational waves

- Compute evidence for hypothesis that there is a signal in the data, $\mathcal{H}_S$:

$$p(d|\mathcal{H}_S, I) = Z = \int \tilde{L}(X)\, dX \approx \sum_k L_k \Delta X_k$$

- Compute posterior density function for signal parameters, $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta}|d, \mathcal{H}_S, I) \approx \frac{L_k}{Z} \Delta X_k$$

- In the case of a coalescing binary (black holes and/or neutron stars):

$$\boldsymbol{\theta} = \{t_c, \varphi_c, m_1, m_2, \vec{S}_1, \vec{S}_2, \theta, \phi, \iota, \psi, D\}$$

- Posterior density for a given parameter, e.g. m_1:

  - Use some smooth interpolation of the above posterior density
  - Marginalize over all other parameters

$$p(\theta_1|d, H, I) = \int_{\theta_2^{\min}}^{\theta_2^{\max}} \ldots \int_{\theta_N^{\min}}^{\theta_N^{\max}} p(\theta_1, \ldots, \theta_N|d, H, I)\, d\theta_2 \ldots d\theta_N$$

Gravitational-wave parameter estimation

- Parameter space is 15-dimensional:

$$\vec{\theta} = \{m_1, m_2, \vec{S}_1, \vec{S}_2, \alpha, \delta, \iota, \psi, d_{\rm L}, t_c, \varphi_c\}$$

- Different detectors *D* have different response to signals:

$$\tilde{h}^{(D)}(f) = \left[F_+^{(D)}\tilde{h}_+(f) + F_\times^{(D)}\tilde{h}_\times(f)\right] e^{-2\pi i f \Delta t^{(D)}}$$

where $F_+^{(D)}(\alpha, \delta, \psi, t_0)$ and $F_\times^{(D)}(\alpha, \delta, \psi, t_0)$ antenna pattern functions at geocentric arrival time t_0 while $\Delta t^{(D)}(\alpha, \delta, t_0)$ differences between arrival times at geocenter and at detectors

- Different noise realizations in different detectors:

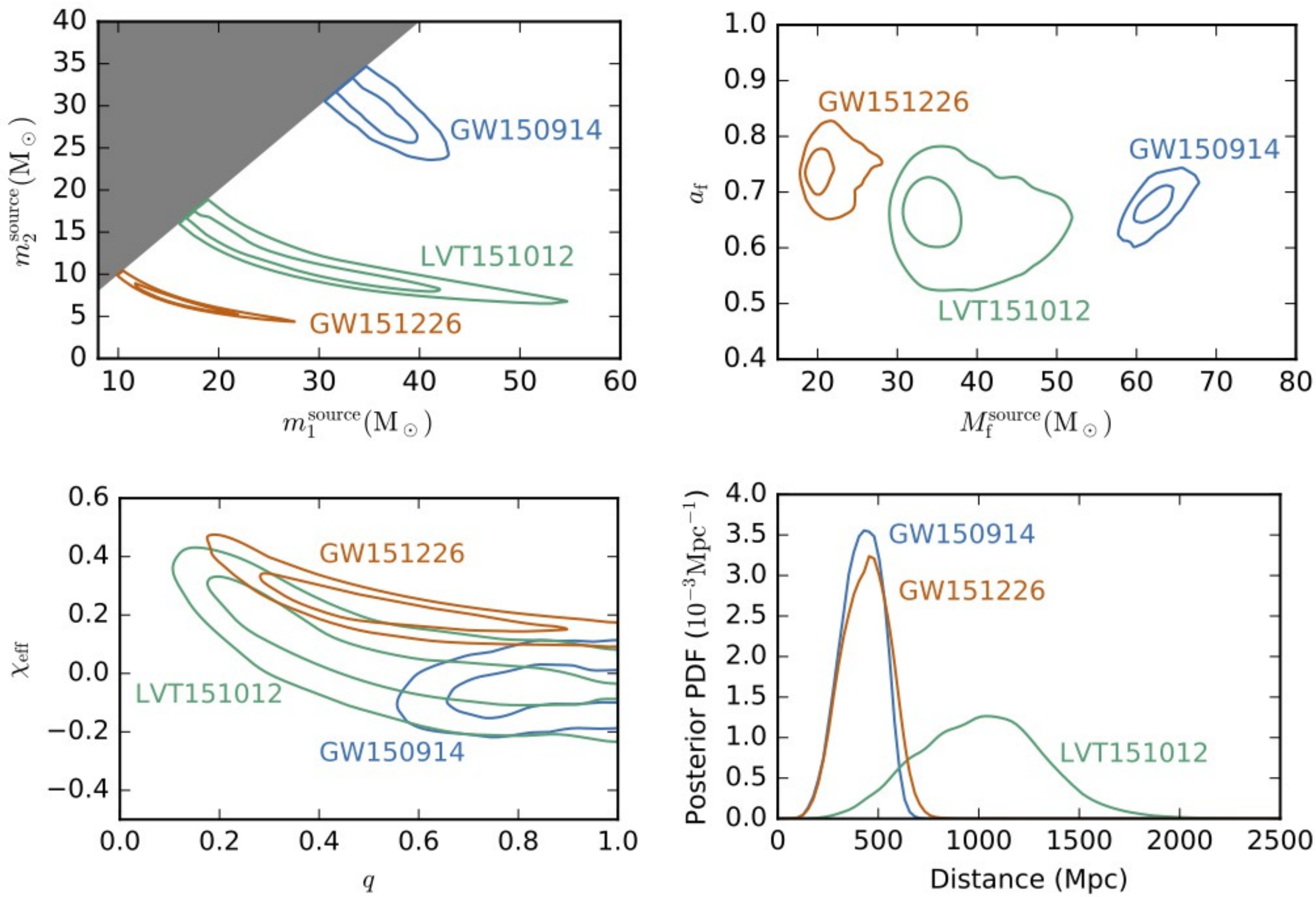
$$\tilde{d}^{(D)}(f) = \tilde{h}^{(D)}(f) + \tilde{n}^{(D)}(f)$$

- Different noise power spectral densities:

$$\langle \tilde{n}^{(D)}(f)\, \tilde{n}^{(D')^*}(f') \rangle = \frac{1}{2}\delta(f - f')\delta_{DD'} S^{(D)}(f)$$

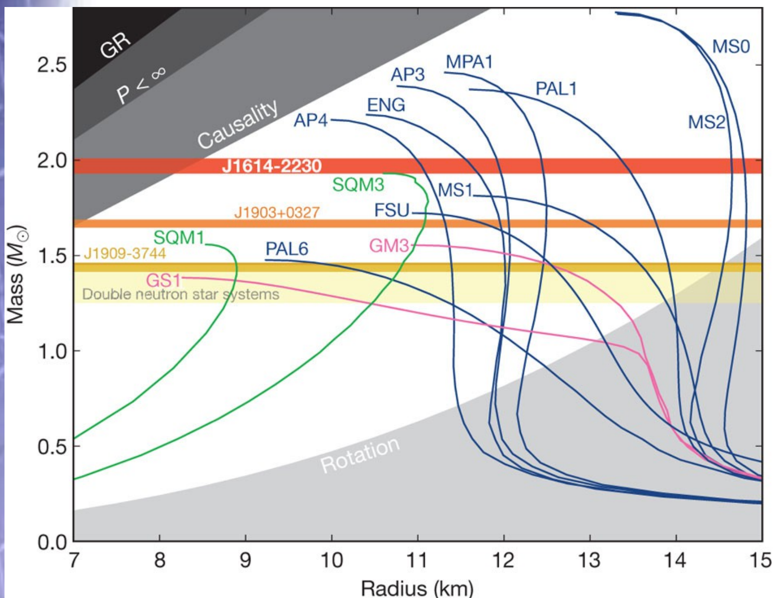
- Joint likelihood: $p(\vec{d}|\vec{\theta}, \mathcal{H}, I) = \prod_{(D)} p(\vec{d}^{(D)}|\vec{\theta}, \mathcal{H}, I)$

Masses, spins, distances

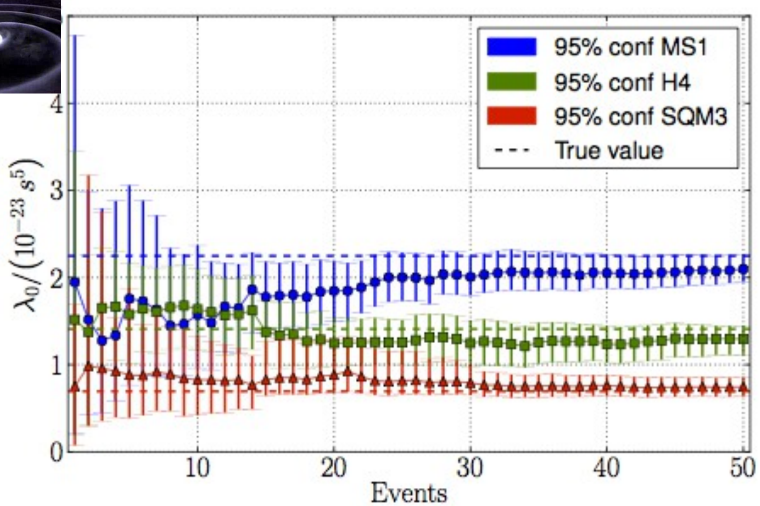


https://arxiv.org/abs/1606.04856

Binary neutron star coalescences



Demorest et al., Nature **467**, 1081 (2010)

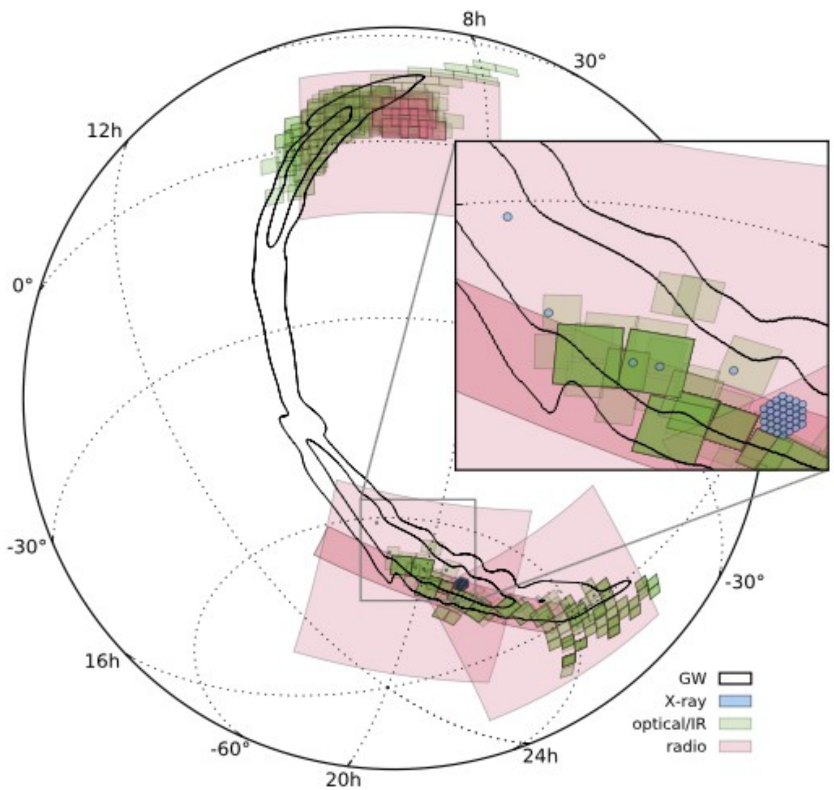


Del Pozzo et al., Phys. Lett. **111**, 071101 (2013)

- Internal structure of neutron stars not well understood: major open problem in astrophysics
- Large uncertainty in equation of state:
  - Pressure as function of density
  - Mass as function of radius
  - Tidal deformability as function of mass
- Tidal deformability leaves imprint on gravitational wave signal
  - After few tens of detections, distinguish between stiff, moderate, and soft equation of state
  - Also information in merger itself, though hard to extract (high frequency)

Detecting binary neutron stars

- Would be helpful to see electromagnetic counterpart
- Sky map for GW150914 was sent to astronomers, and they looked (though no EM emission expected from binary black holes!)

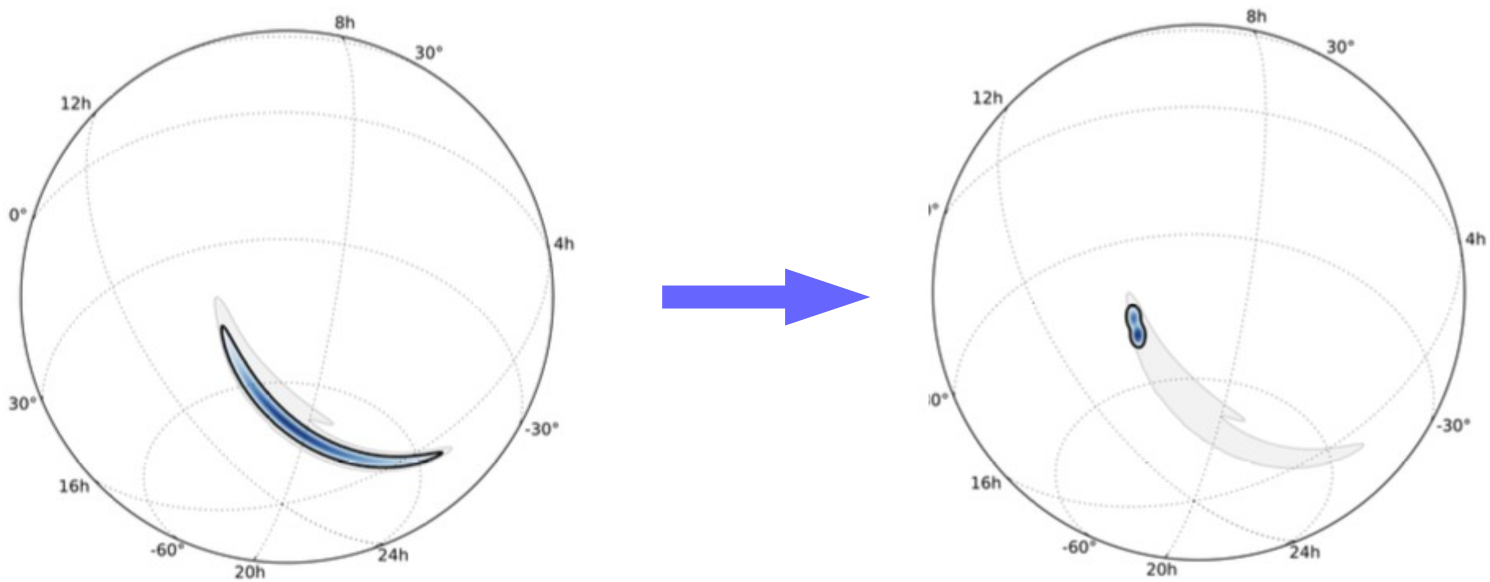


Footprints of Tiled Observations

Group	Area (deg^2)	Contained probability (%) cWB[a]	LIB[b]	LALInf[c]
Swift	2	0.6	0.8	0.1
DES	94	32.1	13.4	6.6
INAF	93	28.7	9.5	6.1
J-GEM	24	0.0	1.2	0.4
MASTER	167	9.3	3.3	6.0
Pan-STARRS	355	27.9	22.9	8.8
SkyMapper	34	9.1	7.9	1.7
TZAC	29	15.1	3.5	1.6
ZTF	140	3.1	2.9	0.9
(total optical)	759	76.5	46.8	23.9
LOFAR-TKSP	103	26.6	1.3	0.5
MWA	2615	97.8	71.8	59.0
VAST	304	25.3	1.7	6.3
(total radio)	2623	97.8	71.8	59.0
(total)	2730	97.8	76.8	62.1

Detecting binary neutron stars

- What if we had seen binary neutron star coalescence as loud as GW150914?
- With **Advanced Virgo** included, 90% confidence sky error box would be reduced from ~180 deg^2 to ~10 deg^2



LIGO Hanford + LIGO Livingston

LIGO Hanford + LIGO Livingston

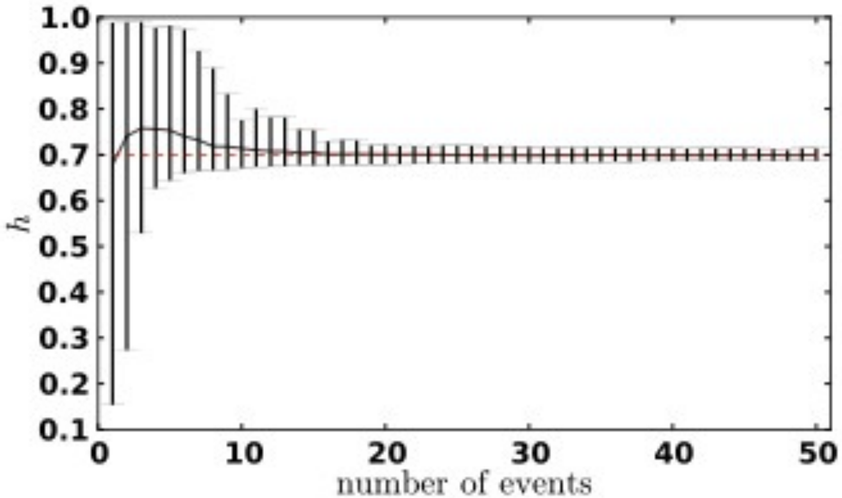
+ Advanced Virgo

Cosmography

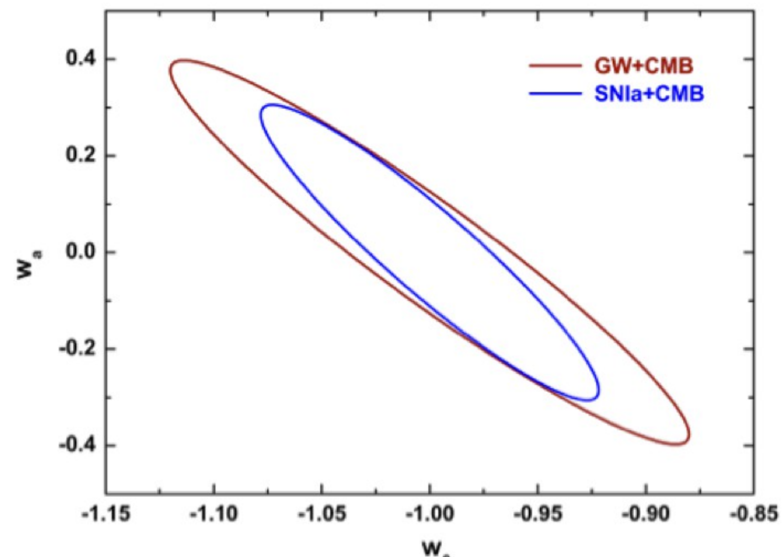
- Distance to source can be obtained from gravitational wave signal itself, without need for calibration against other types of sources:

$$\mathcal{A}(t) \propto \frac{1}{D_L} \mathcal{M}_{\rm obs}^{5/3} \mathcal{F}(\theta, \phi, \iota, \psi)\, F^{2/3}(t)$$

- If redshift can also be obtained, the probe relationship between luminosity distance and redshift, $D_L(z)$
  - Measurement of cosmological parameters $\vec{\Omega} = (H_0, \Omega_M, \Omega_{DE}, \Omega_k, w)$
  - With 2nd generation detectors: only Hubble constant
  - With 3rd generation detectors: dark energy equation of state $P = w\,\rho$



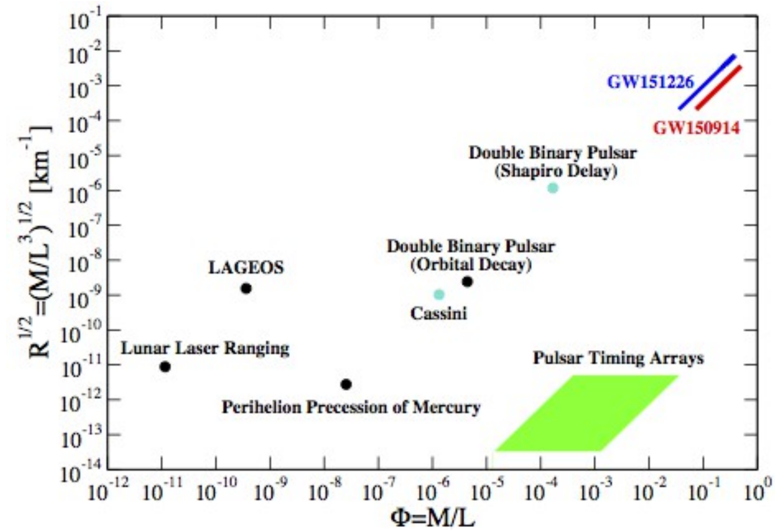
Del Pozzo, PRD **86**, 043011 (2012)

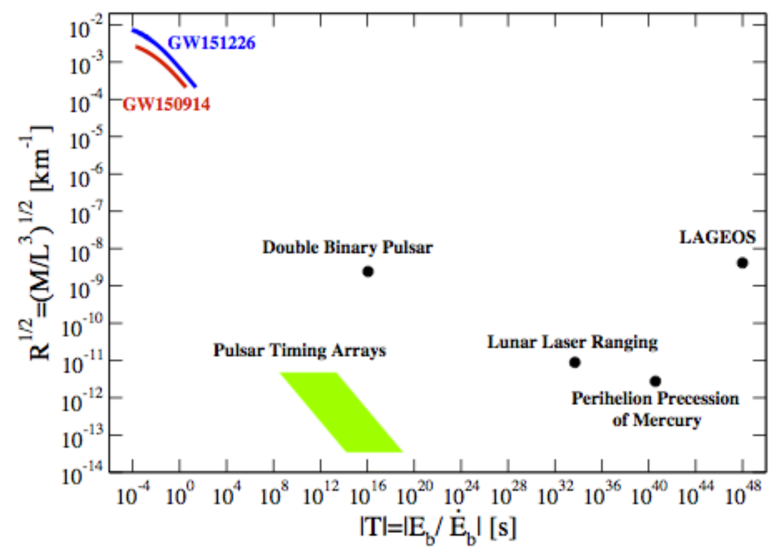


Zhao, Van Den Broeck, Baskaran, Li, PRD **83**, 023005 (2011)

First access to the strong-field dynamics of spacetime

- Before the direct detection of gravitational waves:
  - Solar system tests: weak-field; dynamics of spacetime itself not being probed
  - Binary neutron stars: relatively weak-field test of spacetime dynamics
  - Cosmology: dark matter and dark energy may signal GR breakdown
- Direct detection of GW from binary black hole mergers:
  - Genuinely strong-field dynamics
  - (Presumed) pure spacetime events





Yunes, Yagi, Pretorius, Phys. Rev. D **94**, 084002 (2016)