Bayesian methods in the search for gravitational waves

Reinhard Prix

Albert-Einstein-Institut Hannover

Bayes forum

Garching, Oct 7 2016

Statistics as applied Probability Theory

Probability Theory: extends deductive logic to situations of *incomplete information* (☞ "Inference") [Jaynes, Cox]

Logical propositions, e.g.

A = "There is a signal in this data"

$A(h_0, f)$ = "The signal has amplitude h_0 and frequency f"

$P(A|I) \equiv$ quantifies plausibility of A being true *given* I

I = relevant background knowledge and assumptions

☞ quantifies an observer's state of knowledge about A

☞ not a property of the observed system! ("Mind projection fallacy")

(Cox 1946, 1961, Jaynes) Requiring 3 conditions for $P(A|I)$:
(i) $P \in \mathbb{R}$, (ii) consistency, (iii) agreement with "common sense"
one can *derive* unique laws of probability (up to gauge):

The Three Laws

1. $P(A|I) \in [0,1]$ $\begin{cases} P(A|I) = 1 & \Leftrightarrow (A|I) \text{ certainly true} \\ P(A|I) = 0 & \Leftrightarrow (A|I) \text{ certainly false} \end{cases}$
2. $P(A|I) + P(\text{not } A|I) = 1$
3. $P(A \text{ and } B|I) = P(A|B,\, I)\; P(B|I)$

☞ Bayes' theorem: $P(A|B,I) = P(B|A,I)\, \frac{P(A|I)}{P(B|I)}$

☞ Sum rule: $P(A \text{ or } B|I) = P(A|I) + P(B|I) - P(A \text{ and } B|I)$

Bayesian data analysis

Q: We observe data $\mathbf{x}$, what can we learn from it?

Formulate "question" as a proposition A and *compute* $P(A|\mathbf{x}, I)$

The 'standard' GW hypotheses

$\mathcal{H}_{\mathrm{G}}$: data is pure Gaussian noise: $\mathbf{x}(t) = \mathbf{n}(t)$

$\mathcal{H}_{\mathrm{S}}$: data is *signal* + Gaussian noise: $\mathbf{x}(t) = \mathbf{n}(t) + \mathbf{h}(t; \theta)$

- signal parameters, e.g. θ = { masses, spins, position ... }
- Data from several detectors: $\mathbf{x} = \{x^{\mathrm{H1}}, x^{\mathrm{L1}}, \ldots\}$
- Gaussian noise pdf: $P(\mathbf{n}|\mathcal{H}_{\mathrm{G}}) = \kappa\, e^{-\frac{1}{2}(\mathbf{n}|\mathbf{n})}$
  - ☞ "matched-filter" scalar product $(x|y) = \int \frac{\tilde{x}(f)\,\tilde{y}^*(f)}{S_{\mathrm{n}}(f)}\, df$
  - ☞ *assumes known* (i.e. estimated) noise PSDs $S_{\mathrm{n}}(f)$ (alternative: marginalize)

Bayes factor I

Q1: Given data $\mathbf{x}$, what can we learn about $\mathcal{H}_{\mathrm{G}}$ and $\mathcal{H}_{\mathrm{S}}$?

Two possibilities:

1. Complete set of hypotheses: directly compute $P\left(\mathcal{H}_{\mathrm{S}}|\mathbf{x}, I\right)$
2. Alternative: relative probabilities ("odds"):

$$O_{\mathrm{S/G}}(\mathbf{x}) \equiv \underbrace{\frac{P\left(\mathcal{H}_{\mathrm{S}}|\mathbf{x}\right)}{P\left(\mathcal{H}_{\mathrm{G}}|\mathbf{x}\right)}}_{\text{Posterior odds}} = \underbrace{\frac{P\left(\mathbf{x}|\mathcal{H}_{\mathrm{S}}\right)}{P\left(\mathbf{x}|\mathcal{H}_{\mathrm{G}}\right)}}_{\text{Bayes factor } B_{\mathrm{S/G}}} \times \underbrace{\frac{P\left(\mathcal{H}_{\mathrm{S}}\right)}{P\left(\mathcal{H}_{\mathrm{G}}\right)}}_{\text{prior odds}},$$

☞ $B_{\mathrm{S/G}}(\mathbf{x})$ "updates" our knowledge about $\mathcal{H}_{\mathrm{S}}/\mathcal{H}_{\mathrm{G}}$

Bayes factor II

Q1': How to deal with unknown signal parameters θ ?

Likelihood ratio (function):

$$\mathcal{L}(\mathbf{x};\theta) \equiv \frac{P(\mathbf{x}|\mathcal{H}_{\mathrm{S}},\theta)}{P(\mathbf{x}|\mathcal{H}_{\mathrm{G}})} = \exp\left[(\mathbf{x}|\mathbf{h}(\theta)) - \frac{1}{2}(\mathbf{h}(\theta)|\mathbf{h}(\theta))\right]$$

Laws of probability ☞ "marginalize":

$$B_{\mathrm{S/G}}(\mathbf{x}) = \int \mathcal{L}(\mathbf{x};\theta)\, P(\theta|\mathcal{H}_{\mathrm{S}})\, d\theta$$

"Orthodox" maximum-likelihood (ML) approach:

$$\mathcal{L}_{\mathrm{ML}}(\mathbf{x}) = \max_{\theta} \mathcal{L}(\mathbf{x};\theta)$$

Bayesian parameter estimation

Q2: What can we learn about signal parameters θ ?

☞ directly compute posterior probability $P(\theta|\mathbf{x}, \mathcal{H}_S)$

$$\underbrace{P(\theta|\mathbf{x}, \mathcal{H}_S)}_{\text{posterior}} \propto \underbrace{\mathcal{L}(\mathbf{x};\theta)}_{\text{likelihood function}} \times \underbrace{P(\theta|\mathcal{H}_S)}_{\text{prior}}$$

Q2': What can we learn about a *subset* of parameters λ ?

$\theta = \{\mathcal{A}, \lambda\}$ ☞ "marginalize" over "uninteresting" parameters $\mathcal{A}$:

$$P(\lambda|\mathcal{H}_S, \mathbf{x}) = \int P(\mathcal{A}, \lambda|\mathcal{H}_S)\, d\mathcal{A} \propto \int \mathcal{L}(\mathbf{x};\theta)\, P(\theta|\mathcal{H}_S)\, d\mathcal{A}$$

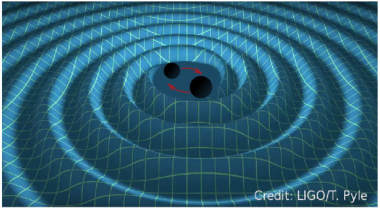
Summary: Bayesian data analysis – strengths and weaknesses

Bayesian probability is the "perfect *machine*" for data analysis, but the difficulty lies in

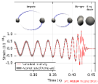
- choosing the "right" inputs:
  hypotheses $\mathcal{H}_i$, priors $P(\theta|\mathcal{H})$, ...
  ☞ What do we (really) know?
  ☞ How to quantify/formalize it?
- evaluation: can write down "optimal answer", but may be
  - impossible to compute
  - much slower than an efficient "ad-hoc" statistic
  - not more detection power than empirical/ad-hoc approaches

☞ use wisely ...

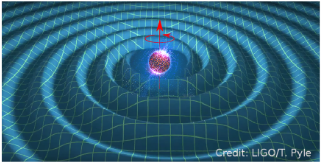
Compact Binary Coalescence (CBC)



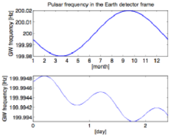
- sources: inspirals of compact objects (NSs, BHs)
- strong ($h_0 \sim 10^{-21}$) & short $\sim \mathcal{O}(\mathrm{s})$
- approximate waveforms from GR



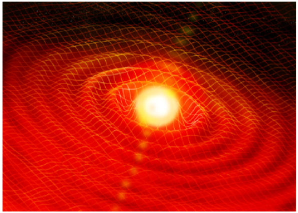
Continuous Waves (CW)



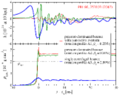
- sources: rotating, non-axisymmetric neutron stars
- weak ($h_0 \lesssim 10^{-25}$)
- long-lasting (days – years): integrate to gain $\mathrm{SNR} \propto \sqrt{T}$
- quasi-periodic, sinusoidal waveform
- signal phase- and amplitude- modulated
- parameter-space resolution (number of templates) grows $\mathcal{N} \propto T^n$ with $n \gtrsim 5$
  $\Longrightarrow$ sensitivity limited by finite computational power
  ☞ semi-coherent methods. . .



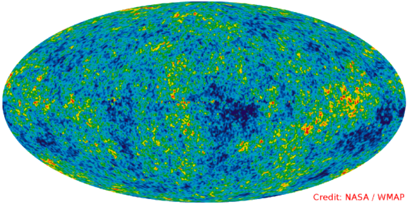
'Unmodelled' bursts



- sources: all CBC sources + supernovae, GRBs, ..
- strong ($h_0 \sim \mathcal{O}(10^{-21})$)
- short $\sim \mathcal{O}(\mathrm{s})$
- minimal assumptions on waveform



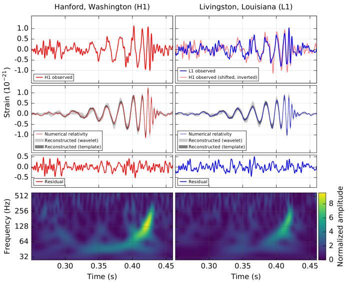
Stochastic gravitational waves



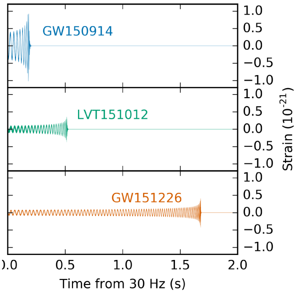
Credit: NASA / WMAP

- sources: cosmological (big bang) or “background” of BBHs
- weak, long-lasting, all directions, all frequencies, power-spectrum
- looking for correlated GW signals between detectors

CBC: Detection/Discovery



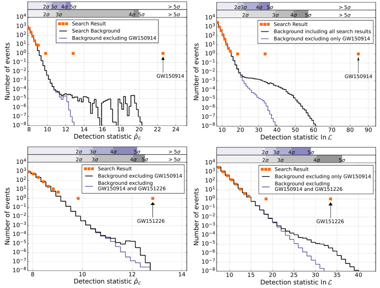
LVC, PRL**116**, 061102 (2016)



LVC, arXiv:1606.04856

Highly empirical/non-Bayesian:

- 2 detection pipelines (PyCBC, GstLAL)
- per-detector matched-filter SNR $\rho_{\mathrm{H1,L1}}$
- “goodness-of-fit” re-weighting (e.g. χ^2) ☞ $\hat{\rho}_{\mathrm{H1,L1}}$
- keep coincident “triggers” ($\hat{\rho} >$threshold) within 15 ms
- combined ranking statistic $\hat{\rho}^2 = \hat{\rho}^2_{\mathrm{H1}} + \hat{\rho}^2_{\mathrm{L1}}$
- What is the noise distribution / “background” ? ☞ time-slides / interpolated detector trigger distribution
- p-value: $P\left(\hat{\rho} \geq \hat{\rho}_{\mathrm{candidate}} \,|\, \text{background}\right)$

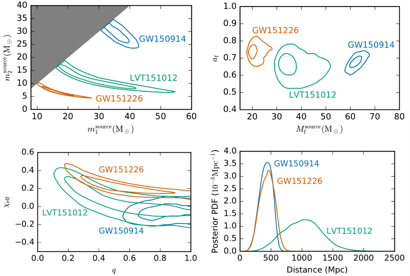


LVC, arXiv:1606.04856

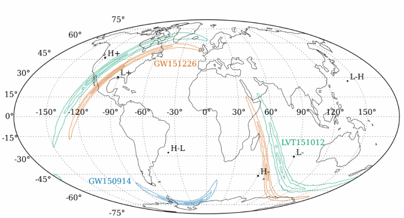
CBC: [fully Bayesian] Parameter estimation

- 15 parameters θ for full signal waveform:
  - 8 intrinsic: masses, spins
  - 7 extrinsic: sky-position, distance, orientation, time and phase
- ☞ Compute $P(\theta|\mathcal{H}_S, \mathbf{x})$: using stochastic samplers
  - Markov Chain Monte Carlo (MCMC)
  - Nested sampling
- Two families of “physical” waveforms (tuned against NR)
- marginalize over calibration uncertainties

☞ real showcase application of Bayesian methods!
☞ Gravitational-wave “astronomy” is fully Bayesian!



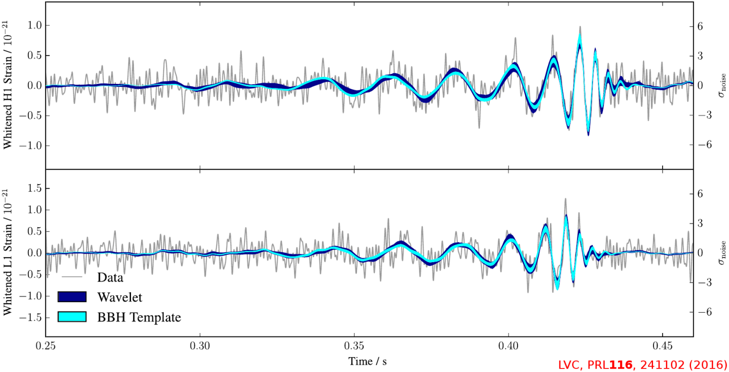
LVC, arXiv:1606.04856



LVC, arXiv:1606.04856

'Unmodelled' reconstruction

- relax assumption about inspiral waveform
- superposition of arbitrary number of sine-Gaussians "wavelets"
- Bayesian ('BayesWave') reconstruction of waveform
- agrees very well ($\sim$ 94%) with best-matching CBC waveform

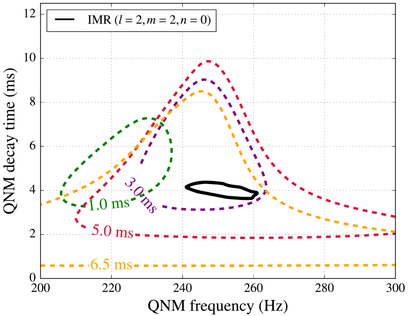


LVC, PRL**116**, 241102 (2016)

GW150914: QNM ringdown

Surprise: GW150914 had a 'visible' ringdown post-merger!

- Bayesian parameter-estimation and evidence for damped sinusoid starting at t_0:
  $h(t) = \mathcal{A}\, e^{-\frac{t-t_0}{\tau}} \cos\left(2\pi f\,(t - t_0) + \phi_0\right)$
  ☞ analytically marginalize $\{\mathcal{A}, \phi_0\}$, search $\{f, \tau\}$ at fixed t_0
- GR/NR: QNM ringdown frequency f expected to be stabilized $\sim 10 - 20M \approx 3.5\text{ms} - 7\text{ms}$ after merger
- posterior estimates of ringdown frequency and damping time consistent with GR prediction
- need ≥ 2 ringdown modes to test Kerr/no-hair theorem

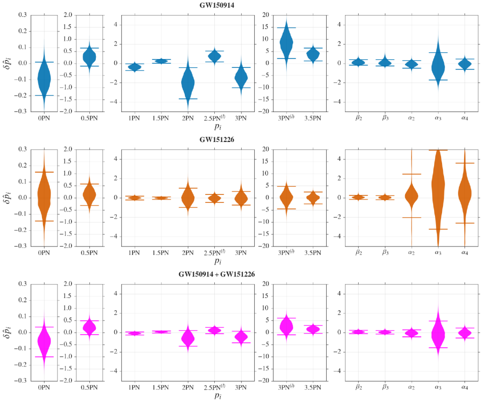


LVC, PRL**116**, 221101 (2016)

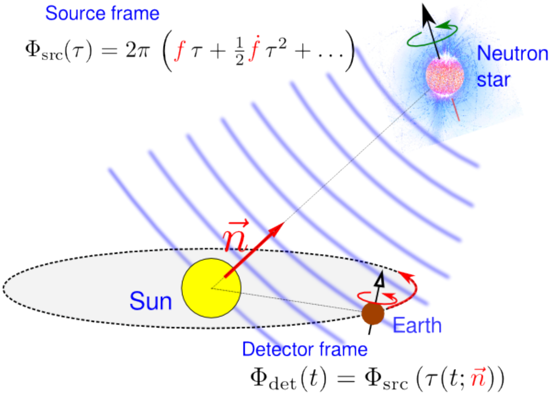
Tests of general relativity

Express GR waveform in terms of post-Newtonian and phenomological (merger+ringdown) coefficients. Test non-zero deviations from GR as “alternative hypothesis”, estimate relative deviations:

Waveform regime	Parameter	f dependence
Early-inspiral regime	$\delta\hat{\varphi}_0$	$f^{-5/3}$
	$\delta\hat{\varphi}_1$	$f^{-4/3}$
	$\delta\hat{\varphi}_2$	f^{-1}
	$\delta\hat{\varphi}_3$	$f^{-2/3}$
	$\delta\hat{\varphi}_4$	$f^{-1/3}$
	$\delta\hat{\varphi}_{5l}$	$\log(f)$
	$\delta\hat{\varphi}_6$	$f^{1/3}$
	$\delta\hat{\varphi}_{6l}$	$f^{1/3}\log(f)$
	$\delta\hat{\varphi}_7$	$f^{2/3}$
Intermediate regime	$\delta\hat{\beta}_2$	$\log f$
	$\delta\hat{\beta}_3$	f^{-3}
Merger-ringdown regime	$\delta\hat{\alpha}_2$	f^{-1}
	$\delta\hat{\alpha}_3$	$f^{3/4}$
	$\delta\hat{\alpha}_4$	$\tan^{-1}(af+b)$



LVC, arXiv:1606.04856

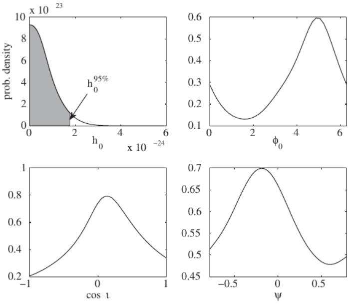


Measured signal strain $h(t; \mathcal{A}, \lambda)$ depends on

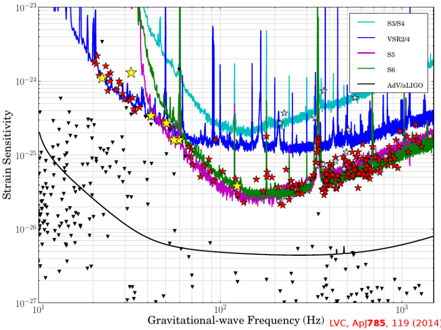
- Amplitude parameters: $\mathcal{A} \equiv \{h_0, \cos\iota, \psi, \phi_0\}$
- Phase-evolution parameters: $\lambda \equiv \{\vec{n}, f, \dot{f}, \ldots\}$

Glasgow Bayesian known-pulsar ULs

- in use since first LIGO science run (S1) [2004]
- Bayesian parameter-estimation pipeline for amplitude parameters $\{h_0, \cos\iota, \psi, \phi_0\}$ for known λ (sky-position, frequency, spindown, ...) [Dupuis, Woan PRD72 (2005)]
- set 95% credible ULs on h_0 from posteriors
- most sensitivity search / ULs on known pulsars



LSC, PRD**76**, 042001 (2007)



LVC, ApJ**785**, 119 (2014)

Frequentist/orthodox approach: optimal statistic?

Simple hypotheses ($\mathcal{A}$ *known*): Neyman-Pearson lemma

"Optimal":= highest detection probability at fixed false-alarm

☞ Likelihood ratio is optimal: $\mathcal{L}(x;\mathcal{A}) \equiv \frac{P(x|\mathcal{H}_\mathrm{S},\mathcal{A})}{P(x|\mathcal{H}_\mathrm{G})}$

Unknown amplitude parameters $\mathcal{A}$ ☞ $\mathcal{F}$-statistic

[Jaranowski, Królak, Schutz, PRD58 (1998)]

change $\mathcal{A}$-coordinates: $\mathcal{A}^\mu = \mathcal{A}^\mu(h_0, \cos\iota, \psi\,\phi_0)$

Likelihood ratio $\mathcal{L}(\mathbf{x};\mathcal{A}) \propto \exp[-\frac{1}{2}\mathcal{A}^\mu\mathcal{M}_{\mu\nu}\mathcal{A}^\nu + \mathcal{A}^\mu\,x_\mu]$

☞ Can analytically maximize $\mathcal{L}(x;\,\mathcal{A})$ over $\mathcal{A}^\mu$:

$$\mathcal{L}_\mathrm{ML}(x) \equiv \max_{\{\mathcal{A}^\mu\}} \mathcal{L}(x;\mathcal{A}^\mu) = e^{\mathcal{F}(x)}$$

- widely-used CW statistics
- efficient (FFT) implementation, no explicit search over $\mathcal{A}$

Bayesian "re-discovery" of the $\mathcal{F}$-statistic

$$B_{\mathrm{S/G}}(x) = \int \mathcal{L}(x; \mathcal{A}) \underbrace{P(\mathcal{A}|\mathcal{H}_{\mathrm{S}})}_{\mathcal{A}-\text{prior}} d^4\mathcal{A}$$

simplest choice: *flat* $\mathcal{A}^\mu$-prior: $P(\mathcal{A}^\mu|\mathcal{H}_{\mathrm{S}}) = \mathrm{const}$

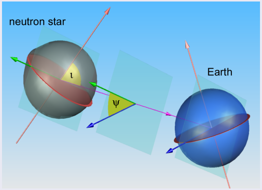
$$\Longrightarrow B_{\mathcal{F}}(x) \propto \int \mathcal{L}(x; \mathcal{A}^\mu)\, d^4\mathcal{A}^\mu \propto e^{\mathcal{F}(x)}$$

☞ ML $\mathcal{F}$-statistic is equivalent to Bayes factor with flat $\mathcal{A}^\mu$-prior!

What is the "right" $\mathcal{A}$-prior?

Ignorance prior in physical coordinates $\{h_0, \cos\iota, \psi, \phi_0\}$:

- initial phase ☞ uniform in ϕ_0
- NS orientations equally likely isotropic ☞ uniform in $\{\cos\iota, \psi\}$
- h_0: astrophysical prior or simplicity $\propto \{h_0{}^{-4},\ h_0{}^{-1},\ \mathrm{const}\}$

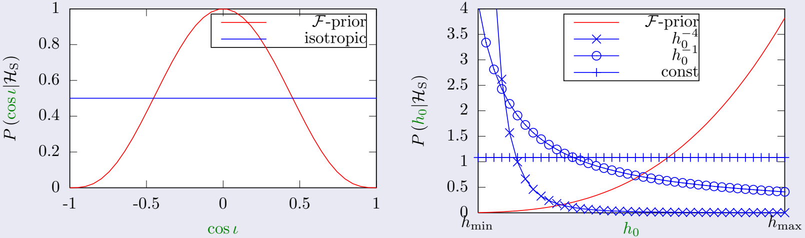


$\mathcal{F}$-statistic prior in physical coordinates:

$$P(\mathcal{A}|\mathcal{H}_S, \text{flat}\{\mathcal{A}^\mu\}) \propto \underbrace{h_0{}^3}_{\text{favors strong signals}} \times \underbrace{(1-\cos\iota^2)^3}_{\text{favors linear polarization}}$$

"unphysical" in $\{h_0, \cos\iota\}$: ✗

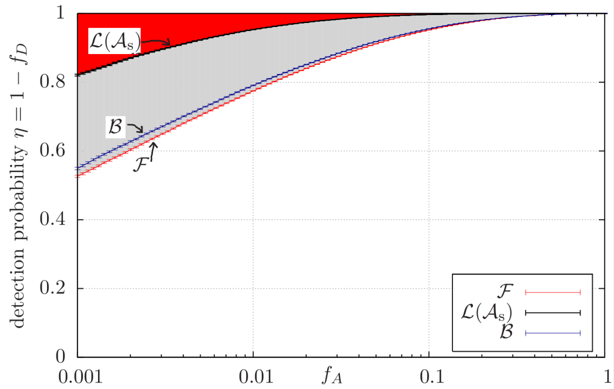
uniform in $\{\psi, \phi_0\}$: ✓



Bayes factor with "physical" $\mathcal{A}$-priors: "$\mathcal{B}$-statistic"

$$\mathcal{B}(x) \propto \int \mathcal{L}(x; \mathcal{A})\; dh_0\, d\cos\iota\, d\psi\, d\phi_0$$

Inject signals with uniform $P(\cos\iota, \psi, \phi_0 | \mathcal{H}_S)$ at fixed SNR=4



- ☞ $\mathcal{F}$-statistic is not N-P "optimal" [Prix, Krishnan, CQG26 (2009)]
- ☞ drawing from priors $\Longrightarrow$ Bayes-factor is N-P optimal!

[A. Searle, arXiv:0804.1161 (2008)]

Summary: $\mathcal{F}$-statistic versus Bayes factor

- classical maximum-likelihood $\mathcal{F}$-statistic can be interpreted as a Bayes factor, but with an *unphysical* implicit prior

  [similar for burst searches: Searle, Sutton, Tinto CQG 26 (2009)]
- physical priors result in *optimal* Bayes factor $\mathcal{B}(x)$, but
  - gains in detection power rather minor
  - computing cost impractical (numerical $\mathcal{A}$-integration)

  ☞ $\mathcal{F}$-statistic is a practical & efficient $\mathcal{B}$ approximation!
- Viewing $e^{\mathcal{F}}$ as a Bayes factor allows for better interpretation and extensions ☞ line-robust statistics

Can we make $\mathcal{F}$ more robust vs "line" artifacts?

Problem with $O_{\mathrm{S/G}}(x) = P(\mathcal{H}_\mathrm{S}|x) / P(\mathcal{H}_\mathrm{G}|x) \propto e^{\mathcal{F}(x)}$

Anything that looks more like $\mathcal{H}_\mathrm{S}$ than Gaussian noise $\mathcal{H}_\mathrm{G}$ can result in large $O_{\mathrm{S/G}}$, regardless of its "goodness-of-fit" to $\mathcal{H}_\mathrm{S}$!

e.g. quasi-monochromatic+stationary detector artifacts ("lines")

Alternative hypothesis $\mathcal{H}_\mathrm{L}$ to capture "lines"

"Zeroth order" simple line model:

$$\mathcal{H}_\mathrm{L} = \text{data } \mathbf{x} \text{ consistent with signal in only one detector}$$
$$= \left[\left(\mathcal{H}_\mathrm{S}^1 \text{ and } \mathcal{H}_\mathrm{G}^2 \right) \text{ or } \left(\mathcal{H}_\mathrm{G}^1 \text{ and } \mathcal{H}_\mathrm{S}^2 \right) \right]$$

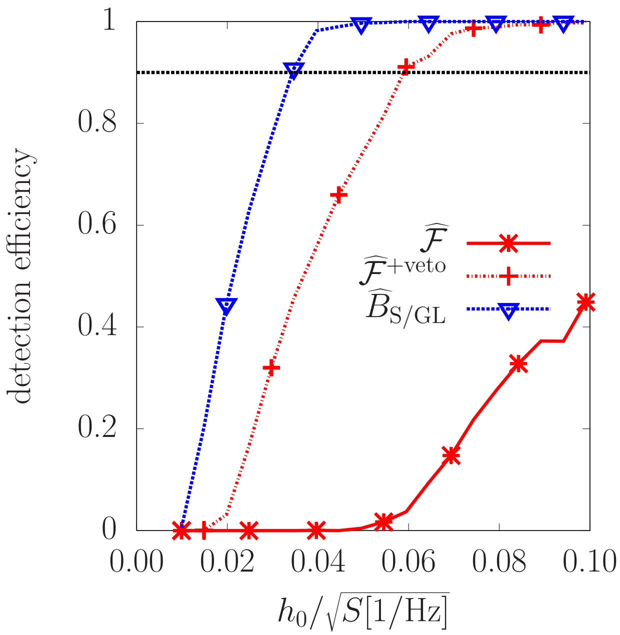
Extended odds: "line-robust" detection statistic

Use simple $\mathcal{F}$-statistic priors $P(\mathcal{A}^\mu|\mathcal{H}_\mathrm{L}) = \text{const}$:

$$O_{\mathrm{S/GL}}(\mathbf{x}) \equiv \frac{P(\mathcal{H}_\mathrm{S}|\mathbf{x})}{P(\mathcal{H}_\mathrm{G} \text{ or } \mathcal{H}_\mathrm{L}|\mathbf{x})} \propto \frac{e^{\mathcal{F}(\mathbf{x})}}{e^{\mathcal{F}_*} + p_\mathrm{L}^1\, e^{\mathcal{F}^1(x^1)} + p_\mathrm{L}^2\, e^{\mathcal{F}^2(x^2)}}$$

[Keitel et al, PRD89 (2014)]

- recent "transient" extensions: [Keitel, PRD93 (2016)]
  - ☞ robust against transient lines (tL): $O_{\mathrm{S/GLtL}}$
  - ☞ sensitive to transient signals (tS): $O_{\mathrm{tS/GLtL}}$
- arbitrary prior cutoff h_{max} leads to a "tuning parameter" $\mathcal{F}_*$
  - ☞ eliminate $\mathcal{F}_*$ by using more physical prior approximation e.g. $P(\mathcal{A}^\mu|\mathcal{H}_\mathrm{S}) \propto e^{-\mathcal{A}^2/2\sigma}$ [work in progress]


detection efficiency
1
0.8
0.6
0.4
0.2
0
0.00
0.02
0.04
0.06
0.08
0.10
$h_0/\sqrt{S[1/\mathrm{Hz}]}$
$\widehat{\mathcal{F}}$
$\widehat{\mathcal{F}}^{+\mathrm{veto}}$
$\widehat{B}_{\mathrm{S/GL}}$

Conclusions

Bayesian methods are gaining ground in GW searches ...

- ✗ Search/detection/"confidence" relies most heavily on empirical/frequentist methods
- ✓ Estimation of signal parameters and astrophysical rates ("GW astronomy") fully Bayesianized (CBC+CW)
- ✓ Various tests of General relativity
- ✓ Bayes factor with alternative hypotheses used in CW searches to be more robust versus detector artifacts ($O_{\mathrm{S/GL}}$, $O_{\mathrm{S/GLtL}}$, $O_{\mathrm{tS/GLtL}}$)

☞ Help us find GWs and join Einstein@Home!
https://einsteinathome.org

einstein@home

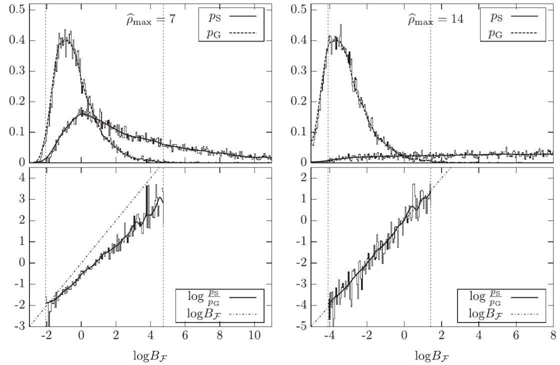
Extra slide

Bayes-factor self-consistency relation

$$B_{S/G} \equiv \frac{P(x|\mathcal{H}_S)}{P(x|\mathcal{H}_G)} = \frac{P(B_{S/G}|\mathcal{H}_S)}{P(B_{S/G}|\mathcal{H}_G)}$$

☞ "Bayes factor predicts its own relative frequencies!"

[Prix, Giampanis, Messenger PRD84 (2011)]



Prix, Giampanis, Messenger, PRD84 023007 (2011)