Bayesian methods in the search for gravitational waves

Reinhard Prix

Albert-Einstein-Institut Hannover

Bayes forum
Garching, Oct 7 2016
Probability Theory: extends deductive logic to situations of *incomplete information* (☞ “Inference”) [Jaynes, Cox]

Logical propositions, e.g.

\[ A = \text{“There is a signal in this data”} \]
\[ A(h_0, f) = \text{“The signal has amplitude } h_0 \text{ and frequency } f \text{”} \]

\[ P(A|I) \equiv \text{quantifies plausibility of } A \text{ being true given } I \]
\[ I = \text{relevant background knowledge and assumptions} \]

☞ quantifies an observer’s state of knowledge about \( A \)
☞ not a property of the observed system! (“Mind projection fallacy”)
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. $P(A|I) \in [0, 1]$\[ 
\begin{align*} 
P(A|I) &= 1 & \iff (A|I) \text{ certainly true} \\
P(A|I) &= 0 & \iff (A|I) \text{ certainly false} 
\end{align*} \]

2. $P(A|I) + P(\text{not } A|I) = 1$

3. $P(A \text{ and } B|I) = P(A|B, I) P(B|I)$

—in Bayes’ theorem: $P(A|B, I) = P(B|A, I) \frac{P(A|I)}{P(B|I)}$

—in Sum rule: $P(A \text{ or } B|I) = P(A|I) + P(B|I) - P(A \text{ and } B|I)$
Q: We observe data $x$, what can we learn from it?

Formulate “question” as a proposition $A$ and compute $P(A|x, I)$

The ‘standard’ GW hypotheses

$\mathcal{H}_G$: data is pure Gaussian noise: $x(t) = n(t)$

$\mathcal{H}_S$: data is signal + Gaussian noise: $x(t) = n(t) + h(t; \theta)$

- signal parameters, e.g. $\theta = \{\text{masses, spins, position} \ldots\}$
- Data from several detectors: $x = \{x^{H1}, x^{L1}, \ldots\}$
- Gaussian noise pdf: $P(n|\mathcal{H}_G) = \kappa e^{-\frac{1}{2}(n|n)}$

⚠️ “matched-filter” scalar product $(x|y) = \int \frac{\tilde{x}(f)\tilde{y}^*(f)}{S_n(f)} df$

⚠️ assumes known (i.e. estimated) noise PSDs $S_n(f)$ (alternative: marginalize)
Q1: Given data $\mathbf{x}$, what can we learn about $\mathcal{H}_G$ and $\mathcal{H}_S$?

Two possibilities:

1. Complete set of hypotheses: directly compute $P(\mathcal{H}_S|\mathbf{x}, I)$

2. Alternative: relative probabilities ("odds"):

$$O_{S/G}(\mathbf{x}) \equiv \frac{P(\mathcal{H}_S|\mathbf{x})}{P(\mathcal{H}_G|\mathbf{x})} = \frac{P(\mathbf{x}|\mathcal{H}_S)}{P(\mathbf{x}|\mathcal{H}_G)} \times \frac{P(\mathcal{H}_S)}{P(\mathcal{H}_G)},$$


Posterior odds

Bayes factor $B_{S/G}$

prior odds

$B_{S/G}(\mathbf{x})$ “updates” our knowledge about $\mathcal{H}_S/\mathcal{H}_G$
Q1': How to deal with unknown signal parameters \( \theta \)?

Likelihood ratio (function):

\[
\mathcal{L}(\mathbf{x}; \theta) \equiv \frac{P(\mathbf{x}|\mathcal{H}_S, \theta)}{P(\mathbf{x}|\mathcal{H}_G)}
= \exp\left[ (\mathbf{x}|\mathbf{h}(\theta)) - \frac{1}{2} (\mathbf{h}(\theta)|\mathbf{h}(\theta)) \right]
\]

Laws of probability “marginalize”:

\[
B_{S/G}(\mathbf{x}) = \int \mathcal{L}(\mathbf{x}; \theta) P(\theta|\mathcal{H}_S) d\theta
\]

“Orthodox” maximum-likelihood (ML) approach:

\[
\mathcal{L}_{ML}(\mathbf{x}) = \max_{\theta} \mathcal{L}(\mathbf{x}; \theta)
\]
**Q2**: What can we learn about signal parameters $\theta$?

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

$$P(\theta|x, \mathcal{H}_S) \propto \mathcal{L}(x; \theta) \times P(\theta|\mathcal{H}_S)$$

**Q2′**: What can we learn about a *subset* of parameters $\lambda$?

$\theta = \{ \mathcal{A}, \lambda \}$ “marginalize” over “uninteresting” parameters $\mathcal{A}$:

$$P(\lambda|\mathcal{H}_S, x) = \int P(\mathcal{A}, \lambda|\mathcal{H}_S) \, d\mathcal{A} \propto \int \mathcal{L}(x; \theta) \, P(\theta|\mathcal{H}_S) \, d\mathcal{A}$$
Bayesian probability is the “perfect machine” for data analysis, but the difficulty lies in

- choosing the “right” inputs: hypotheses $H_i$, priors $P(\theta|H)$, . . .
  - 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 $\sim O(s)$
- 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 $\text{SNR} \propto \sqrt{T}$
- quasi-periodic, sinusoidal waveform
- signal phase- and amplitude- modulated
- parameter-space resolution (number of templates) grows $N \propto T^n$ with $n \gtrsim 5$
- sensitivity limited by finite computational power
- semi-coherent methods...

'Unmodelled' bursts

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

Stochastic gravitational waves

- 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

Highly empirical/non-Bayesian:

- 2 detection pipelines (PyCBC, GstLAL)
- per-detector matched-filter SNR $\rho_{H1,L1}$
- “goodness-of-fit” re-weighting (e.g. $\chi^2$) $\hat{\rho}_{H1,L1}$
- keep coincident “triggers” ($\hat{\rho} >$ threshold) within 15 ms
- combined ranking statistic $\hat{\rho}^2 = \hat{\rho}_{H1}^2 + \hat{\rho}_{L1}^2$

What is the noise distribution / “background”?

- time-slides / interpolated detector trigger distribution
- $p$-value: $P(\hat{\rho} \geq \hat{\rho}_{\text{candidate}} \vert \text{background})$

Bayesian methods in GW searches

LIGO-G1601953-v1 R. Prix
CBC: [fully Bayesian] Parameter estimation

- 15 parameters $\theta$ for full signal waveform:
  - 8 intrinsic: masses, spins
  - 7 extrinsic: sky-position, distance, orientation, time and phase

- Compute $P(\theta | H_S, x)$: 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!
Unmodelled’ reconstruction

- relax assumption about inspiral waveform
- superposition of arbitrary number of sine-Gaussians “wavelets”
- Bayesian (’BayesWave’) reconstruction of waveform
- agrees very well (∼94%) with best-matching CBC waveform

LVC, PRJ116, 241102 (2016)
GW150914: QNM ringdown

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

- Bayesian parameter-estimation and evidence for damped sinusoid starting at $t_0$:
  \[ h(t) = \mathcal{A} e^{-\frac{t-t_0}{\tau}} \cos(2\pi f (t - t_0) + \phi_0) \]
  - analytically marginalize $\{A, \phi_0\}$, 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 $\geq 2$ ringdown modes to test Kerr/no-hair theorem

![Graph showing QNM decay time vs. QNM frequency](image-url)
Tests of general relativity

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

<table>
<thead>
<tr>
<th>Waveform regime</th>
<th>Parameter</th>
<th>$f$ dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early-inspiral regime</td>
<td>$\delta \phi_0$</td>
<td>$f^{-5/3}$</td>
</tr>
<tr>
<td></td>
<td>$\delta \phi_1$</td>
<td>$f^{-4/3}$</td>
</tr>
<tr>
<td></td>
<td>$\delta \phi_2$</td>
<td>$f^{-1}$</td>
</tr>
<tr>
<td></td>
<td>$\delta \phi_3$</td>
<td>$f^{-2/3}$</td>
</tr>
<tr>
<td></td>
<td>$\delta \phi_4$</td>
<td>$f^{-1/3}$</td>
</tr>
<tr>
<td></td>
<td>$\delta \phi_5$</td>
<td>$\log(f)$</td>
</tr>
<tr>
<td></td>
<td>$\delta \phi_6$</td>
<td>$f^{1/3}$</td>
</tr>
<tr>
<td></td>
<td>$\delta \phi_7$</td>
<td>$f^{2/3}$</td>
</tr>
<tr>
<td>Intermediate regime</td>
<td>$\delta \beta_2$</td>
<td>$\log f$</td>
</tr>
<tr>
<td></td>
<td>$\delta \beta_3$</td>
<td>$f^{-3}$</td>
</tr>
<tr>
<td>Merger-ringdown regime</td>
<td>$\delta \alpha_2$</td>
<td>$f^{-1}$</td>
</tr>
<tr>
<td></td>
<td>$\delta \alpha_3$</td>
<td>$f^{3/4}$</td>
</tr>
<tr>
<td></td>
<td>$\delta \alpha_4$</td>
<td>$\tan^{-1}(af + b)$</td>
</tr>
</tbody>
</table>

LVC, arXiv:1606.04856
Continuous gravitational waves (CWs)

Source frame
\[ \Phi_{\text{src}}(\tau) = 2\pi \left( f \tau + \frac{1}{2} \dot{f} \tau^2 + \ldots \right) \]

Detector frame
\[ \Phi_{\text{det}}(t) = \Phi_{\text{src}}(\tau(t; \vec{n})) \]

Measured signal strain \( h(t; \mathcal{A}, \lambda) \) depends on
- **Amplitude parameters:** \( \mathcal{A} \equiv \{ h_0, \cos \nu, \psi, \phi_0 \} \)
- **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 \(\lambda\) (sky-position, frequency, spindown, ...) [Dupuis, Woan PRD72 (2005)]
- set 95\% credible ULs on \(h_0\) from posteriors
- most sensitivity search / ULs on known pulsars
**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: \( \mathcal{L}(x; \mathcal{A}) \equiv \frac{P(x|\mathcal{H}_S, \mathcal{A})}{P(x|\mathcal{H}_G)} \)

**Unknown amplitude parameters \( \mathcal{A} \) \( \cong \) \( \mathcal{F} \)-statistic**

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

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

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

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

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

- widely-used CW statistics
- efficient (FFT) implementation, no explicit search over \( \mathcal{A} \)
Bayesian “re-discovery” of the $\mathcal{F}$-statistic

$$B_{S/G}(x) = \int \mathcal{L}(x; \mathcal{A}) P(\mathcal{A}|\mathcal{H}_S) \, d^4\mathcal{A}$$

simplest choice: flat $\mathcal{A}^\mu$-prior: $P(\mathcal{A}^\mu|\mathcal{H}_S) = \text{const}$

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

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 $\phi_0$
- NS orientations equally likely isotropic uniform in $\{\cos \iota, \psi\}$
- $h_0$: astrophysical prior or simplicity $\propto \{h_0^{-4}, h_0^{-1}, \text{const}\}$
$F$-statistic prior in physical coordinates:

$$P(\mathcal{A}|\mathcal{H}_S, \text{flat}\{\mathcal{A}^\mu\}) \propto h_0^3 \times (1 - \cos \iota^2)^3$$

- favors strong signals
- favors linear polarization

"unphysical" in $\{h_0, \cos \iota\}$: ❌
uniform in $\{\psi, \phi_0\}$: ✔️
Bayes factor with “physical” $A$-priors: “$B$-statistic”

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

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

\[ F \text{-statistic is not N-P “optimal” [Prix, Krishnan, CQG26 (2009)]} \]

\[ \text{drawing from priors} \implies \text{Bayes-factor is N-P optimal!} \]

[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 $O_{S/G}(x) = P(\mathcal{H}_S|x)/P(\mathcal{H}_G|x) \propto e^{\mathcal{F}(x)}$

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

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

Alternative hypothesis $\mathcal{H}_L$ to capture “lines”

“Zeroth order” simple line model:

$$\mathcal{H}_L = \text{data } x \text{ consistent with signal in only one detector}$$

$$= \left[ (\mathcal{H}_S^1 \text{ and } \mathcal{H}_G^2) \text{ or } (\mathcal{H}_G^1 \text{ and } \mathcal{H}_S^2) \right]$$
Extended odds: “line-robust” detection statistic

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

$$O_{S/GL}(x) \equiv \frac{P(\mathcal{H}_S|x)}{P(\mathcal{H}_G \text{ or } \mathcal{H}_L|x)} \propto \frac{e^{\mathcal{F}(x)}}{e^{\mathcal{F}_*} + p_1^L e^{\mathcal{F}_1(x^1)} + p_2^L e^{\mathcal{F}_2(x^2)}}$$

[Keitel et al, PRD89 (2014)]

- recent “transient” extensions: [Keitel, PRD93 (2016)]
  - robust against transient lines ($t_L$): $O_{S/GLtL}$
  - sensitive to transient signals ($t_S$): $O_{tS/GLtL}$

- arbitrary prior cutoff $h_{\text{max}}$ leads to a “tuning parameter” $\mathcal{F}_*$
  - eliminate $\mathcal{F}_*$ by using more physical prior approximation

  e.g. $P(\mathcal{A}^\mu|\mathcal{H}_S) \propto e^{-\mathcal{A}^2/2\sigma}$ [work in progress]
Bayesian methods in GW searches
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 ($O_{S/GL}$, $O_{S/GLtL}$, $O_{tS/GLtL}$)

Help us find GWs and join Einstein@Home! [Link to Einstein@Home]

https://einsteinathome.org
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)]