Probability and statistics for GW inference §7.4

Tuesday: Matched-filtering approach:

For detector signal $s(t) = h_t(t) + n(t)$ we find optimal SNR

$$\rho = \frac{S}{N} = \frac{\langle s \rangle}{\sqrt{\langle n^2 \rangle}} \quad \text{with filter } \tilde{K}(f) = \frac{\tilde{h}(f)}{S_n(f)} \text{ and the}$$

Wiener inner product $(h|g) = 4\,\mathrm{Re} \int_0^{+\infty} \frac{\tilde{h}(f)\tilde{g}^*(f)}{S_n(f)}\, df$

$\rightsquigarrow \rho = \frac{(s|h)}{\sqrt{(h|h)}}$ $\rightsquigarrow$ maximize over template bank

Result of approach: A list of "events" with $\rho > \rho_{thr}$

- What can we conclude from this? What can we learn from this?
- When do we claim detection?

1) Frequentist and Bayesian approaches

S – sample/even space of all possible outcomes

$A, B, \ldots \subset S$ – events with associated probability $P(A) \in [0,1]$

Kolmogorov axioms:

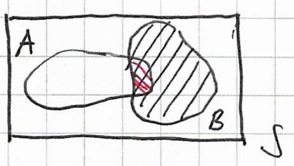
(i) $\forall A \subset S$: $P(A) \geq 0$

(ii) For disjoint subsets $\underbrace{A \cap B}_{\text{intersection}} = \emptyset$: $P(A \cup B) = P(A) + P(B)$

(iii) Normalization: $P(S) = 1$

Conditional probability: $P(A|B) := \frac{P(A \cap B)}{P(B)}$

Probability of A happening if we know B has already occurred



Bayes' theorem: $P(A|B) \cancel{P(B)} = \frac{P(B|A)\,P(A)}{P(B)}$ flip conditional probabilities

$$= \frac{P(B|A)\,P(A)}{\sum_i P(B|C_i)\,P(C_i)}$$ for disjoint subsets $\bigsqcup_i C_i = S$

Proof: $P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$ &

$\sum_i P(B|C_i)P(C_i) = \sum_i P(B \cap C_i) = P(\bigsqcup_i B \cap C_i)$

$= P(B \cap S) = P(B)$

Nomenclature: $\overset{\text{posterior}}{P(\text{hypothesis}|\text{data})} = \dfrac{\overset{\text{likelihood}}{P(\text{data}|\text{hypothesis})}\ \overset{\text{prior}}{P(\text{hypothesis})}}{\underset{\text{normalization}}{P(\text{data})}}$

Typically: $S = \mathbb{R}^n$, $P(A) = \int_A dx^n\, p(x)$ ← pdf

	Frequentist	Bayesian
Development	• old: 1930's – 1970's	• new: widespread use since 1990's
Usage	• medicine, social sciences, experimental/particle physics	• astrophysics, economics
GW	• search pipelines, rates	• parameter estimation (PE)
Philosophy	• Events are the outcomes of repeatable experiments; probabilities are the frequencies of occurence ⇓ Posterior is not a probability	• Likelihoods, ~~and~~ priors and posteriors quantify a "degree of belief" for events and are real objects ⇓ Posterior can be used as a probability
In particular	• P(data|hypothesis) ✓	• P(data|hypothesis) ✓
	• P(hypothesis|data) ✗ (= $\mathcal{B}$(hypothesis – true hypothesis) "likelihood")	• P(hypothesis|data) ✓ We will see the implications and limitations later on
Assumptions	• PDF type known	• PDF type known + prior assumed
Conclusion	• Exact, but hard on a theoretical level	• Approximate, but straight-forward on a computer

Answered question same: Given data from an experiment, how well can we constrain the real parameters?

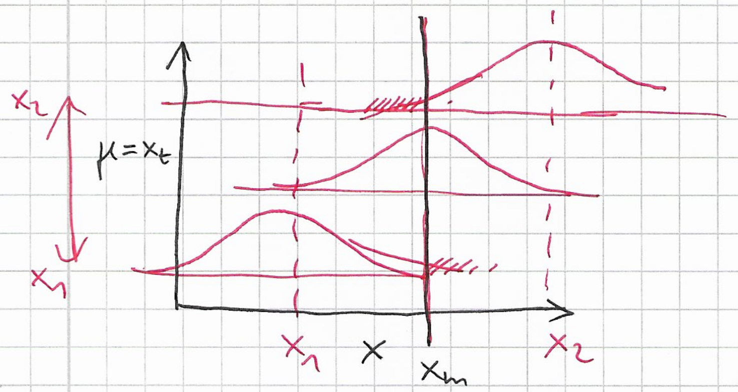
Confidence interval

$x \sim p(x|x_t)$, construct $[x_1, x_2]$ s.t. $x_t \in [x_1, x_2]$

in e.g. $\underbrace{90\%}_{\text{confidence level C.L.}}$ of repeats → "coverage"

Example: $x \sim N(\mu = x_t, \sigma = 1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_t)^2}{2\sigma^2}} = p(x|x_t)$

→ Frequentist: Neumann 1937 construction



Argument:

If $x_t < x_1$ / $> x_2$, x was a fluctuation that happens in $\leq 5\%$ of cases

Generalization: If multiple draws or the series, data can be summarized by a statistic

- Draw repeated data $d_i \sim p(d|\theta)$ can be calculated
- Statistic $\hat{d} = \hat{d}(d_1, \dots, d_N) \sim p(\hat{d}|\theta)$

↳ Repeat approach on observed statistic

Ex: Sample mean $\hat{x} = \sum_{i=1}^{N} \frac{x_i}{N} \sim N(x_t, \frac{\sigma^2}{N})$ ↪ construct conf. intervals

Estimator is a statistic $\hat{\theta}$ that is

- consistent: $\lim_{N\to\infty} \hat{\theta} = \theta_t$
  ⇓
- unbiased: $E_N[\hat{\theta}] = \theta_t$
- efficient: $Var[\hat{\theta}]$ ~~is~~ minimal compared to other statistics
- robust: Insensitive to small departures from assumed pdf

→ Bayesianist: Construct posterior and use it as a PDF:

$$p(x_t|x_m) \cancel{=} \frac{p(x_m|x_t)\,p(x_t)}{p(x_m)} = \mathcal{N} e^{-\frac{(x-x_t)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_t)^2}{2\sigma^2}}$$

⇝ most likely value: $x_t = x_m$, confidence interval

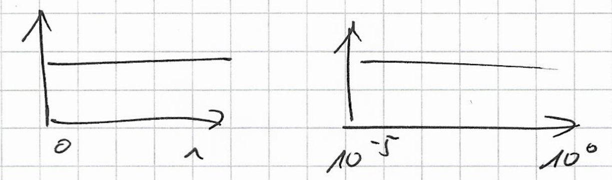
$$\int_{x_1}^{x_2} P(x_t|x_m)\,dx_t = 90\% \quad + \text{ condition (e.g. symmetric)}$$

→ show code

Cousins (1995) "Why isn't every physicist a Bayesian?"

⇝ for the 68% C.L. Poisson distribution the Bayesian analysis doesn't cover anymore, $P(x_t \in C.I.) < 68\%$!

- coverage lost: posterior is not a real probability
- subjectivity introduced: prior should be "uninformative", what does that mean? Eccentricity: $p(e) \propto 1$ or $p(e) \propto \log(e)$?



$\rightsquigarrow$ changes result!

+ works for unrepeatable experiments
+ — " — high dimensionality of the problem or semi-analytical likelihoods
+ gives the full posterior → see multi-modalities & correlations visually ~~(???)~~ (table reference)

2) Parameter estimation

- Frequentist: - compare data to templates: $\hat{\rho} > \rho_{thr}$

pycbc, gstlal
- coincidences in multiple detectors
- SNR deposited in all freq. ranges

$\rightsquigarrow$ FAR $< \frac{1}{6\ \text{month}} = 6.34 \cdot 10^{-8}$ Hz $\leftrightarrow$ C.I.

Stationary Gaussian noise: $s = h_t + n_0 \leftrightarrow n_0 = s - h_t$

$$p(n_0) = \mathcal{N} \exp\left(-\frac{1}{2} \int_0^{+\infty} df \, \frac{|\tilde{n}_0(f)|^2}{S_n(f)}\right)$$

$$= \mathcal{N} \exp\left(-\frac{1}{2} (n_0 | n_0)\right)$$

$$\Rightarrow p(s|\theta_t) = \mathcal{N} \exp\left(-\frac{1}{2} (s - h_t | s - h_t)\right)$$

- Bayesian: Posterior: $p(\theta_t | s) \propto \exp\left(-\frac{1}{2}(s - h_t | s - h_t)\right) p(\theta_t)$

$$\propto \exp\left(-\frac{1}{2}\left[(s|s) - 2(s|h_t) + (h_t|h_t)\right]\right) \times p(\theta_t)$$

$$\propto \exp\left((s|h_t) - \frac{1}{2}(h_t|h_t)\right) p(\theta_t)$$

$\rightsquigarrow$ 15D function $\Rightarrow$

a) Estimators for θ_t

b) Look at the posterior in all its glory

1) Maximum Likelihood estimator (MLE) (still Frequentist)

Assume $p(\theta_t)$ as flat:

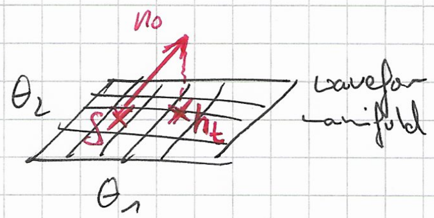
$$\hat{\theta}_{MLE} = \underset{\theta}{\mathrm{argmax}}\; p(\theta|s) = \underset{\theta}{\mathrm{argmax}}\; p(s|\theta)$$

$$= \underset{\theta}{\mathrm{argmax}}\; \log p(s|\theta) = \underset{\theta}{\mathrm{argmax}} \left[\underbrace{(s|h(\theta)) - \tfrac{1}{2}(h(\theta)|h(\theta))}_{=g(\theta)} \right]$$

$$\rightsquigarrow \quad 0 = \frac{\partial}{\partial \theta_i} g(\theta) = (\partial_i h(\theta)|s) - (\partial_i h(\theta)|h(\theta))$$

→ Mostly numerically, although some parameters analytically, see Alessandra

Interpretation: Maximize $g(\theta)$ ↔ Minimize $(s-h|s-h) = \|s-h\|^2$



Inter-connectedness: $\hat{\theta}_{MLE}$ maximizes $\hat{\rho} = \frac{(s|h)}{\sqrt{(h|h)}}$ (best template)

Seperate $h = a \cdot h_a$:

$$0 = \frac{\partial}{\partial a} g = (h_a|s) - a(h_a|h_a) \Rightarrow a = \frac{(h_a|s)}{(h_a|h_a)}$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\mathrm{argmax}} \left[(s|h_a)\frac{(h_a|s)}{(h_a|h_a)} - \frac{1}{2}\frac{(h_a|s)^2}{(h_a|h_a)^2}(h_a|h_a) \right]$$

$$= \underset{\theta}{\mathrm{argmax}} \left[\frac{(h_a|s)^2}{(h_a|h_a)} = \hat{\rho}^2 \right]$$

2) Bayes estimator ⇔ (now Bayesian)

= Posterior expected value:

$$\hat{\theta}_B^{(i)} = \int d\theta^n\, \theta^{(i)}\, p(\theta|s) = E[\theta]_{p(\theta|s)} \quad \text{with variance}$$

$$\Sigma_B^{(ij)} = \int d\theta^n \left[\theta^{(i)} - \theta_B^{(i)}\right]\left[\theta^{(j)} - \theta_B^{(j)}\right] p(\theta|s) = \mathrm{Var}[\theta]$$

✓ Includes prior info

✓ Invariant to marginalizations

✓ Free variance

For large SNR:

Assume $p(\theta) \propto 1$ flat and $\theta = \hat{\theta} + \Delta\theta$

$$p(\theta|s) = \mathcal{N} \exp\left(-\tfrac{1}{2} (s-h|s-h)\right)$$

$$= \mathcal{N}^{(0)} \exp\Big(\underbrace{-\tfrac{1}{2}\, \partial_i \partial_j (s-h|s-h)}\tfrac{1}{2} \Delta\theta_i \Delta\theta_j\Big) \quad \text{Gaussian!}$$

$$= \Gamma_{ij} = (\partial_i \partial_j h | \underbrace{h-s}_{|n| \ll |h|}) + (\partial_i h | \partial_j h)$$

$$\simeq (\partial_i h | \partial_j h)$$

$$= \mathcal{N}^{(0)} \exp\left(-\tfrac{1}{2} \Delta\vec{\theta}^T \underline{\underline{\Sigma}}^{-1} \Delta\vec{\theta}\right) \quad \text{multi-variate Gaussian!}$$

$$\rightsquigarrow \sigma^2_{\theta_i} = \Sigma_{ii} = (\Gamma^{-1})_{ii}$$

→ Very useful prognostic tool to estimate statistical errors for future detectors!

b) State of the art: Statistical sampling with bilby (also ML now)

→ ~~Also just~~ Given posterior, plot it! ⇝ How to visualize 15D function?

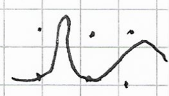
- Prior is typically compact box ✓
- Marginalization: $p(\theta_1, \theta_2)$ can be projected down:

$$p(\theta_1) = \int d\theta_2\, p(\theta_1, \theta_2) \quad \text{still valid pdf in } \theta_1$$

⇝ bilby-gwtc1.github.io/GW150914/

→ How can we get $p(\theta|s)$ and do the integrations?

- Grid? ✗ Might miss modes and no indication for that



  ✗ Number of necessary grid points grows as N^D exponentially
- Statistical sampling: Draw N events from the pdf $p(\theta|s)$ directly
  ✓ Good algorithms explore the prior space globally
  ✓ Marginalization is trivial: If you have vectors $\vec{\theta}$ ⇝ do a histogram in θ_j across all draws

Casino in Monaco, name from the mathematician Stanislaw Ulam inspired by his uncle's gambling habits

Markov Chain Monte Carlo (MCMC)

A randomly sampled sequence of samples $\theta^{(0)}, \theta^{(1)}, \dots$ such that $\theta^{(n+1)}$ depends only on $\theta^{(n)}$ but not on earlier samples through a transition kernel $K(\theta^{(n+1)} | \theta^{(n)})$ conditional probability

Detailed balance: If there exists a stationary probability $P(\theta)$, s.t.

$$\forall \theta, \theta' : \quad K(\theta | \theta') P(\theta') = K(\theta' | \theta) P(\theta)$$

then the Markov Chain converges to draws from the stationary pdf $P(\theta)$:

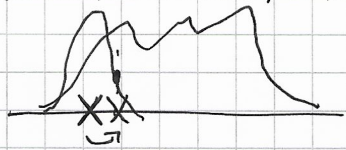
$$\overset{\text{landon}}{pr(\theta)} = \int d\theta' \, K(\theta | \theta') P(\theta')$$

$$= \underbrace{\int d\theta' \, K(\theta' | \theta)}_{=1} P(\theta) = P(\theta) \quad \checkmark$$

(+ ergodicity)

So if we tune our Markov Chain to have $P(\theta) = p(\theta | d)$ posterior, we can sample a Markov Chain "from the posterior"

Metropolis-Hastings algorithm Need a proposal $q(\theta | \theta') \sim N(0, \sigma)$



1. Initialize $\theta^{(0)}$ from the prior $\pi(\theta)$
2. For $n \in \{0, \dots, N-1\}$:
   (a) Propose $\phi \sim q(\theta | \theta^{(n)})$
   (b) Compute the acceptance probability

$$\alpha = \min\left(1, \frac{p(d|\phi) \cdot \pi(\phi)}{p(d|\theta^{(n)}) \cdot \pi(\theta^{(n)})}\right) = \frac{\frac{p(d|\phi) \cdot \pi(\phi)}{\cancel{\pi(d)}}}{\frac{p(d|\theta^{(n)}) \cdot \pi(\theta^{(n)})}{\cancel{\pi(d)}}}\Bigg)$$

   (c) Set $\theta^{(n+1)} = \phi$ with probability α, else set $\theta^{(n+1)} = \theta^{(n)}$

Detailed balance?

Case 1) $p(d|\phi)\,\pi(\phi) > p(d|\theta)\,\pi(\theta)$ (accept definitely)

$$K(\phi|\theta')\,P(\theta') = 1 \cdot q(\phi|\theta') \cdot \cancel{p(\theta|\theta')\,\pi(\theta')}\ \frac{p(d|\theta')\,\pi(\theta')}{p(d)}$$

$$||$$

$$K(\theta'|\phi)\,p(\phi) = \frac{p(d|\theta)\,\pi(\theta)}{\cancel{p(d|\phi)\,\pi(\phi)}} \cdot \underbrace{q(\theta|\phi)}_{\substack{=q(\phi|\theta)\\ \text{symmetric}}} \cdot \frac{\cancel{p(d|\phi)\,\pi(\phi)}}{p(d)} \quad \checkmark$$

Case 2) $p(d|\phi)\,\pi(\phi) < p(d|\theta)\,\pi(\theta)$ (accept with chance)

Other way round ✓

⇝ show .ipynb and if time Dhani et al. (2024) 2404.05811