Data Analysis I

Antoine Petiteau
(APC – Université Paris-Diderot)

School on Gravitational Waves for Cosmology and Astrophysics
Benasque - Spain
5 - 9 June 2017

Introduction

- 4 + 4 lectures to present observation systems and the way we analyse their data
- Observation systems:
  - LIGO / Virgo
  - LISA
  - Pulsar Timing Array
- Analysis
  - Bayesian / Bayesian
  - Sources: binaries, stochastic background

Structure of the lectures

Structure of the lectures

- DA I: Statistic basis for DA: Likelihood, frequentist/Bayesian

Structure of the lectures

- DA I: Statistic basis for DA: Likelihood, frequentist/Bayesian
- GW Obs I: History, response to GW

Structure of the lectures

- DA I: Statistic basis for DA: Likelihood, frequentist/Bayesian
- GW Obs I: History, response to GW
- GW Obs II: LISA: LISAPathfinder, noises, ...

Structure of the lectures

- DA I: Statistic basis for DA: Likelihood, frequentist/Bayesian
- GW Obs I: History, response to GW
- GW Obs II: LISA: LISAPathfinder, noises, ...
- DA II: 3 main classes of signal, parameter estimations, Fisher Matrix

Structure of the lectures

- DA I: Statistic basis for DA: Likelihood, frequentist/Bayesian
- GW Obs I: History, response to GW
- GW Obs II: LISA: LISAPathfinder, noises, ...
- DA II: 3 main classes of signal, parameter estimations, Fisher Matrix
- DA III: LISA DA: Global analaysis, MBHB, stochastic, ...

Structure of the lectures

- DA I: Statistic basis for DA: Likelihood, frequentist/Bayesian
- GW Obs I: History, response to GW
- GW Obs II: LISA: LISAPathfinder, noises, ...
- DA II: 3 main classes of signal, parameter estimations, Fisher Matrix
- DA III: LISA DA: Global analaysis, MBHB, stochastic, ...
- GW Obs III: LIGO

Structure of the lectures

- DA I: Statistic basis for DA: Likelihood, frequentist/Bayesian
- GW Obs I: History, response to GW
- GW Obs II: LISA: LISAPathfinder, noises, ...
- DA II: 3 main classes of signal, parameter estimations, Fisher Matrix
- DA III: LISA DA: Global analaysis, MBHB, stochastic, ...
- GW Obs III: LIGO
- GW Obs IV: PTA

Structure of the lectures

- DA I: Statistic basis for DA: Likelihood, frequentist/Bayesian
- GW Obs I: History, response to GW
- GW Obs II: LISA: LISAPathfinder, noises, ...
- DA II: 3 main classes of signal, parameter estimations, Fisher Matrix
- DA III: LISA DA: Global analaysis, MBHB, stochastic, ...
- GW Obs III: LIGO
- GW Obs IV: PTA
- DA IV: PTA data analysis

Structure of the lectures

- DA I: Statistic basis for DA: Likelihood, frequentist/Bayesian
- GW Obs I: History, response to GW
- GW Obs II: LISA: LISAPathfinder, noises, ...
- DA II: 3 main classes of signal, parameter estimations, Fisher Matrix
- DA III: LISA DA: Global analaysis, MBHB, stochastic, ...
- GW Obs III: LIGO
- GW Obs IV: PTA
- DA IV: PTA data analysis

Overview of Data Analysis I
Statistic basis of DA

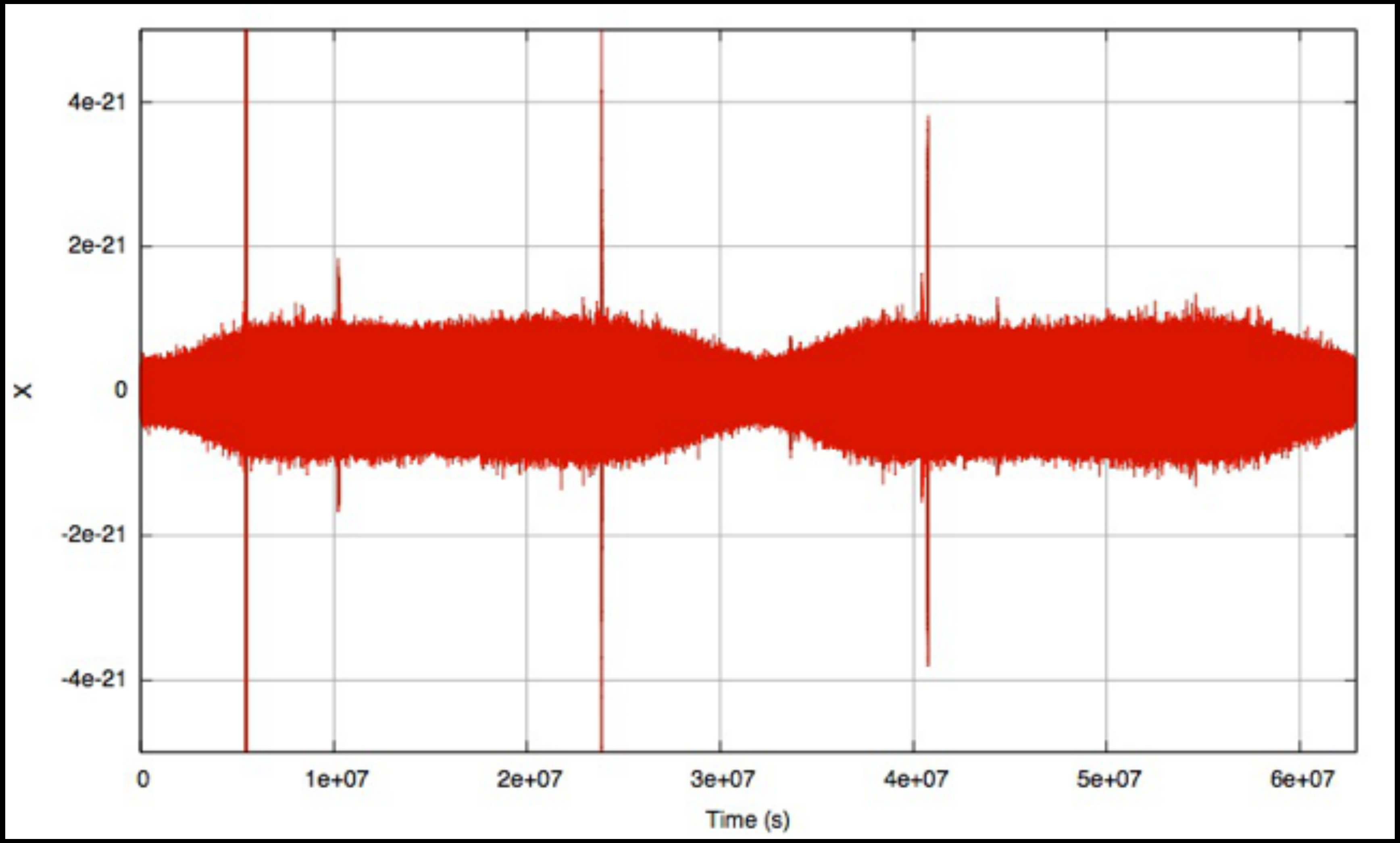
- Extracting signals from noisy data
  - the likelihood function
- Noise weighted inner products, Match-filter, Signal to Noise Ratio and
- Two approaches
  - Frequentist approach
  - Bayesian approach
- Detection statistics and model evidence

References

- Romano & Cornish Liv. Rev. Relativ. (2017) 20:2 “Detection methods for stochastic GW backgrounds: a unified treatment”
- Janarowski & Krolak LRR (2012) 15, 4 “Gravitational-Wave Data Analysis. Formalism and Sample Applications: The Gaussian Case”
- Rover et al.
- Design document of the LISAPathfinder parameter estimation pipeline

Extracting signal from data

- Example of eLISA simulated data (LISACode):
  - about 100 SMBHs,
  - Galactic binaries

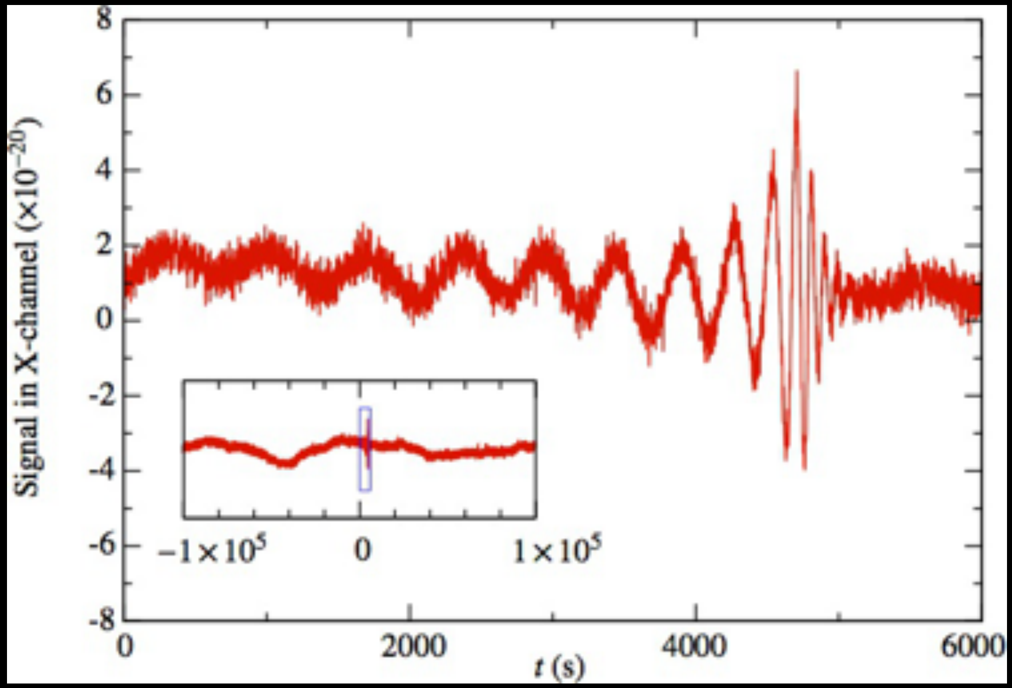


Extracting signal from data

- Example of eLISA simulated data (LISACode):
  - about 100 SMBHs,
  - Galactic binaries
- First « simple data analysis » :

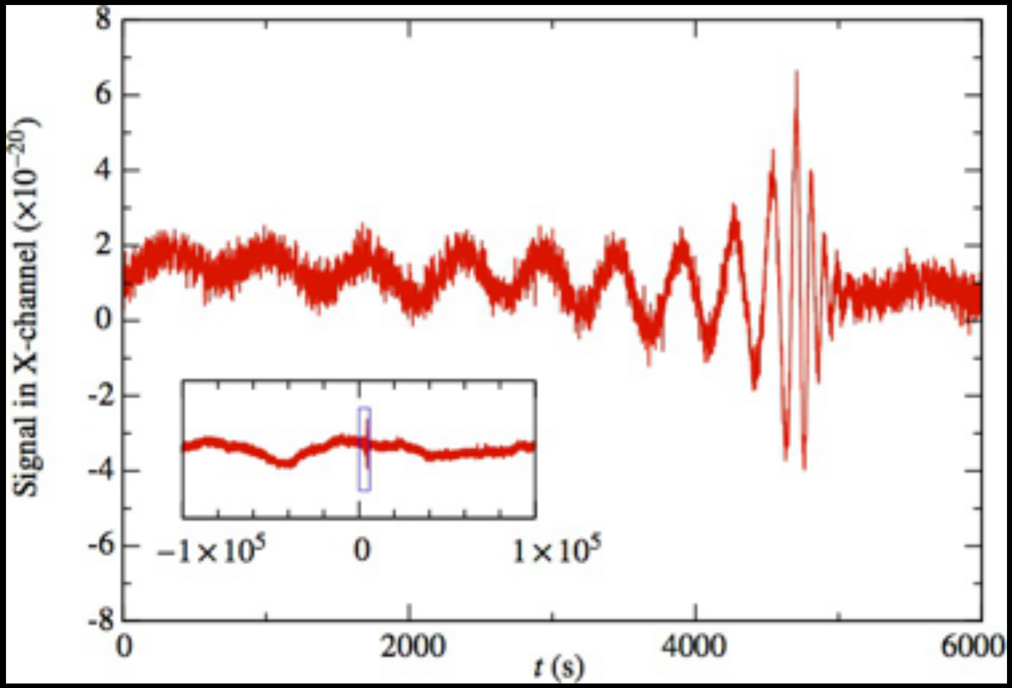
Extracting signal from data

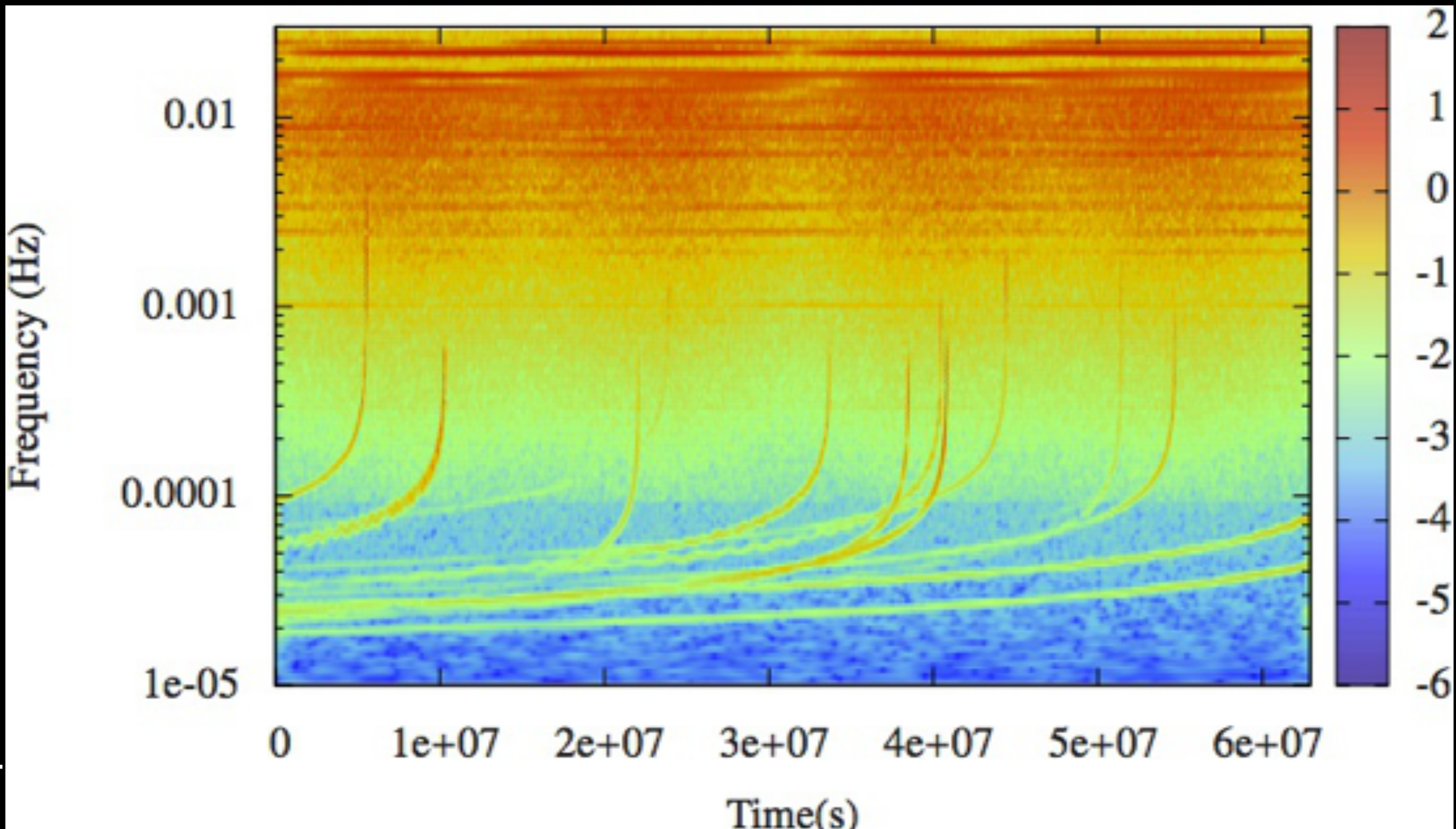
- Example of eLISA simulated data (LISACode):
  - about 100 SMBHs,
  - Galactic binaries
- First « simple data analysis » :
  - low pass + zoom : we can « see » end of waveform for powerful sources



Extracting signal from data

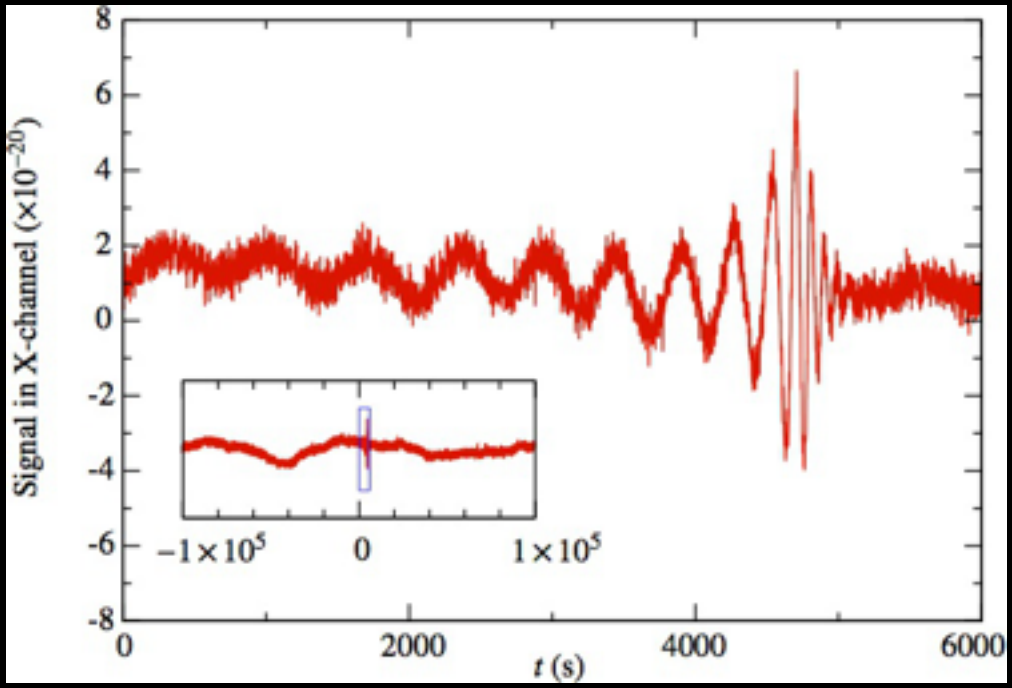
- Example of eLISA simulated data (LISACode):
  - about 100 SMBHs,
  - Galactic binaries
- First « simple data analysis » :
  - low pass + zoom : we can « see » end of waveform for powerful sources
  - wavelet transform :
    - chirps
    - Galactic binaries



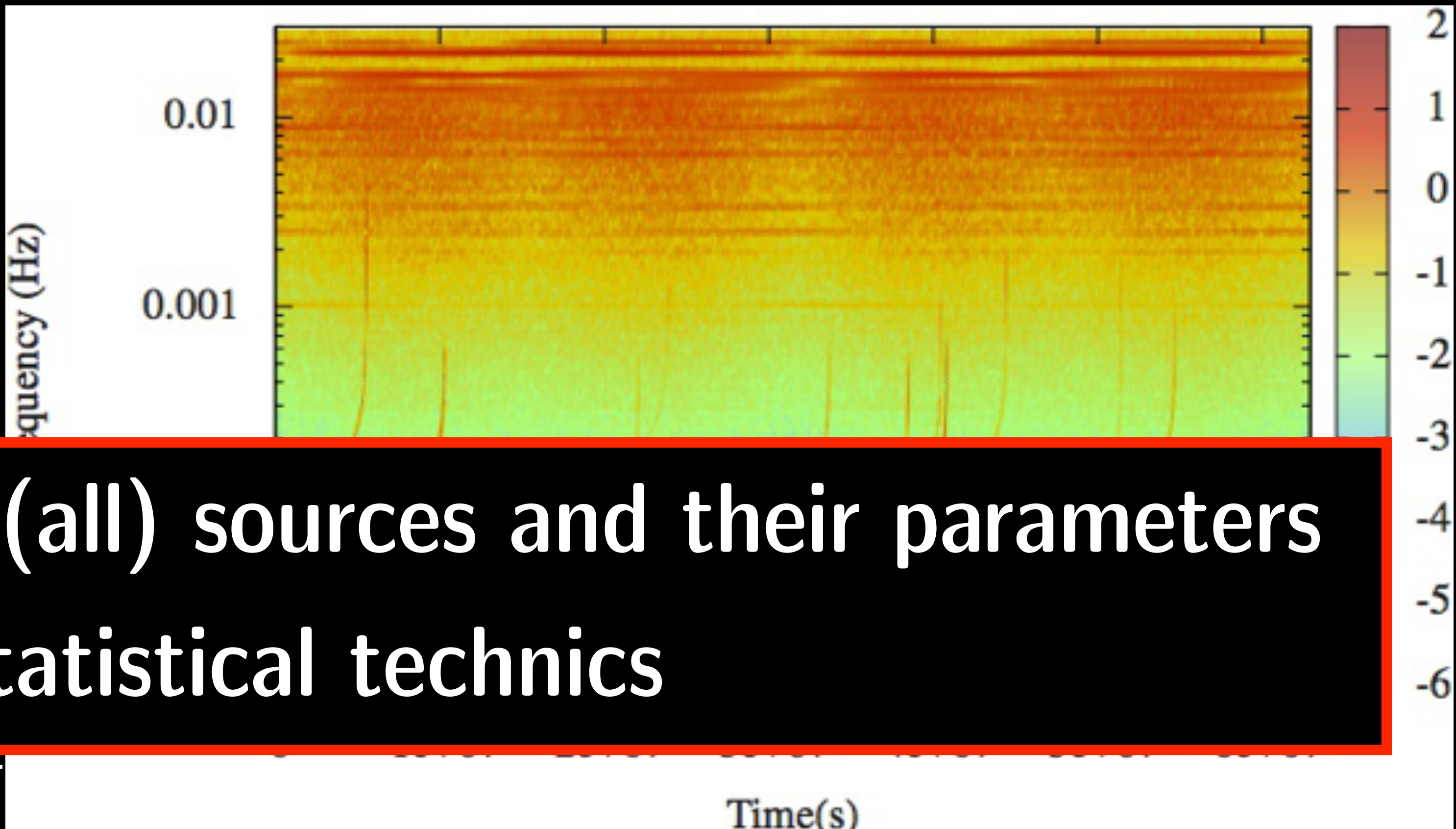


Extracting signal from data

- Example of eLISA simulated data (LISACode):
  - about 100 SMBHs,
  - Galactic binaries



- First « simple data analysis » :
  - low pass + zoom : we can « see » end of waveform for powerful sources
  - wavelet transform :
    - chirps
    - Galactic binaries



- BUT for identifying more (all) sources and their parameters we need more advanced statistical technics

Data, signal and noise

- The time data $d(t)$ contains:
  - $h(t)$: signals that can be characterized by a sets of parameters
    - deterministic / stochastic
    - resolvable or not
  - $n(t)$: noises from
    - instrument itself
    - other sources
- Assumption 1: GW and noise are linearly independent:

$$d(t) = h_{real}(t) + n(t)$$

  - $h(t)$: GW perturbation $h_{ab}(t, \vec{x})$ convolved with instrument response

François Arago Centre

université PARIS DIDEROT PARIS 7

Likelihood

- Goal: find the $h_{model} = h_{real}$
- Likelihood: found by demanding residual compatible with noise distribution $p_n(x)$:
  - The likelihood of observing $d \equiv \{d_1, d_2, \ldots, d_N\}$ where $d_i = d(t_i)$, is given by:

$$p(d(t)|h_{real}(t)) = p_n(r(t)) = p_n(d(t)\text{-}h_{real}(t))$$

- So if $p(d(t)|h_{model}(t))$ is compatible with the noise distribution: $h_{model}(t) = h_{real}(t)$

Likelihood

- Usual case: noise is a multi-variate gaussian distribution:

$$\boxed{p(d|h) = p_n(r) = \frac{1}{\sqrt{det(2\pi C_n)}} e^{-\frac{1}{2}\sum_{i,j} r_i \left(C_n^{-1}\right)_{ij} r_j}}$$

where the correlation matrix is : $C_n = \langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle$

- Generalization for a network of detectors:

$$p(d|h) = \frac{1}{\sqrt{det(2\pi C_n)}} e^{-\frac{1}{2}\sum_{Ii,Jj} r_{Ii} \left(C_n^{-1}\right)_{Ii,Jj} r_{Jj}}$$

where $I,\ J$ labels the detector and $i,\ j$ the discrete time or frequency sample

Likelihood

- **Inner product:**

$$< x|y >= \sum_{Ii,Jj} x_{Ii} \ \left(C_n^{-1}\right)_{Ii,Jj} \ y_{Jj}$$

- **Likelihood:**

$$\mathcal{L} = p(d|h) = \frac{1}{\sqrt{det(2\pi C_n)}} e^{-\frac{1}{2}\langle d-h|d-h\rangle}$$

- **If C_n^{-1} is diagonal with $1/\sigma_i^2$ the inner product is similar to**

$$\chi^2 = \sum_i \left(\frac{d_i - h_i}{\sigma_i}\right)^2 \quad => \quad \mathcal{L} = Ce^{-\frac{1}{2}\langle d-h|d-h\rangle} = Ce^{-\frac{1}{2}\chi^2}$$

Likelihood

- If stationary noise
  - ➡ C_n only depend to $|t_i$ - $t_j|$
  - ➡ C_n ~ diagonal in the Fourier domain (Discrete Fourier Transform) with on the diagonal $S_{n,k}\ T/2$
  - ➡ Inner product: $$< \tilde{x}|\tilde{y} > = 2 \sum_{j=0}^{N/2-1} \Delta f \frac{\tilde{x}_j^* \tilde{h}_j + \tilde{x}_j \tilde{h}_j^*}{S_{n,j}}$$
  - ➡ Continuous limit: $$< \tilde{x}|\tilde{y} > = 2 \int_0^\infty df \frac{\tilde{x}^*(f)\tilde{h}(f) + \tilde{x}(f)\tilde{h}^*(f)}{S_n(f)}$$ $$= 4\ \Re e \int_0^\infty df \frac{\tilde{x}^*(f)\tilde{h}(f)}{S_n(f)}$$
  - ➡ Match-filtering

Likelihood

- If noise C_n is known (i.e. known parameters) and stationary, the factor in front is neglected and we only consider the logarithm of likelihood:

$$\log \mathcal{L} = -\frac{1}{2}\langle d-h|d-h\rangle$$
$$= \langle d|h\rangle - \frac{1}{2}\langle h|h\rangle - \frac{1}{2}\langle d|d\rangle$$

- $<d\,|d>$ is fixed so the relevant term that is usually used is the reduced likelihood:

$$\boxed{\log \mathcal{L}' = \langle d|h\rangle - \frac{1}{2}\langle h|h\rangle}$$

Maximum likelihood

- Considering the signal: $h = A\ h_A$

$$\log \mathcal{L}' = A \langle d|h_A\rangle - \frac{A^2}{2} \langle h_A|h_A\rangle$$

- If the maximum likelihood corresponds to A_{ML}:

$$\left.\frac{\partial \log \mathcal{L}'}{\partial A}\right|_{A_{ML}} = 0 \quad \Rightarrow \quad A_{ML} = \frac{\langle d|h_A\rangle}{\langle h_A|h_A\rangle}$$

- then the maximum likelihood is :

$$max(\log \mathcal{L}') = \frac{\langle d|h_A\rangle^2}{2\langle h_A|h_A\rangle} = \frac{\langle d|h\rangle^2}{2\langle h|h\rangle}$$

Signal to Noise Ratio

- We can define the SNR using the power ratio

$$SNR^2 = \frac{P_{signal}}{P_{noise}}$$

- Average noise power: $P_{noise} = \int_0^\infty df\, S_n(f)$

- Average signal power: $P_{signal} = 2 \int_0^\infty df\, |h(f)|^2$

- Optimal SNR:

$$SNR^2_{opt} = 2 \int_0^\infty df \frac{|h(f)|^2}{S_n(f)} = \langle h|h \rangle$$

SNR vs likelihood

- " Usual " SNR: $$SNR = \frac{\langle d|h\rangle}{\sqrt{\langle h|h\rangle}}$$
- Maximum likelihood: $$max(log\mathcal{L}') = \frac{\langle d|h\rangle^2}{2\,\langle h|h\rangle}$$
- The relation between SNR and maximum likelihood is simply:

$$SNR = \sqrt{2\,max(log\mathcal{L}')}$$

Statistical inference

- Our core tool: the likelihood: $\mathcal{L} = p(d/h)$
- **Likelihood** measures the **probability** of having **data** d **given the hypothesis/model** h .
- How to use it to infer about :
  - detectability of a signal in data ?
  - value of parameters of a signal ?
- Depending where the uncertainty is put, 2 approaches:
  - the frequentist inference
  - the bayesian inference

Frequentist inference

- General ideas:
  - Uncertainties in the data
  - Parameters of the system we want to observe are fixed
  - "long-run relative occurrence of an event in a set of identical experiments"
  - Probability related to the frequencies of events
- Probability of observing the data d given the hypothesis H
- Measured data drawn from an underlying probability distribution $p(d/H)$.

Frequentist inference

- Statistic: function of the data
  - Likelihood or something else
- Required:
  - knowledge of the probability distribution of the statistic (analytic or simulation).
- Problem:
  - Distribution is constructed on non observed data

Frequentist: hypothesis testing

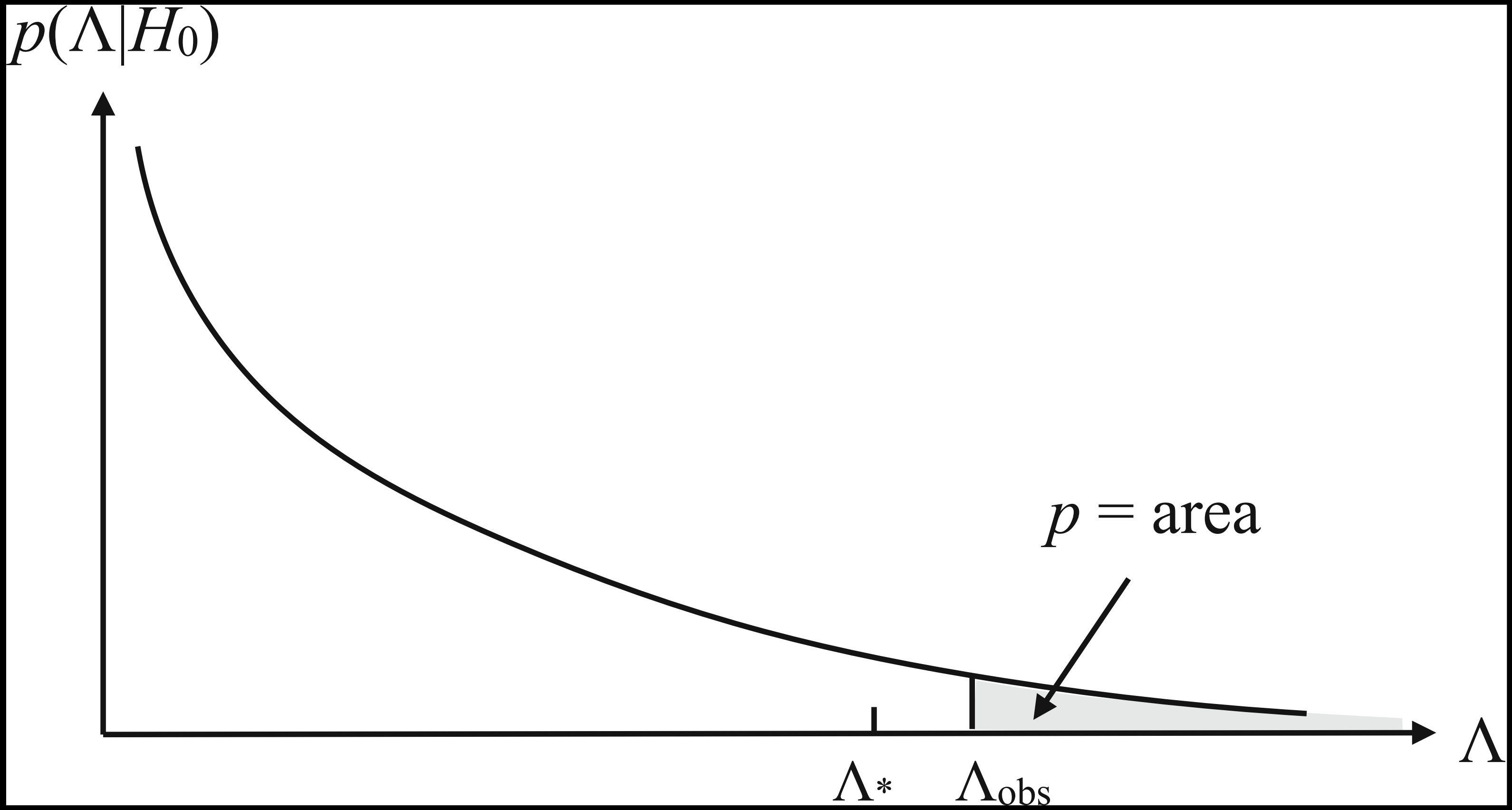
- Hypothesis:
  - H_0 : no signal
  - H_1 : signal: composite hypothesis => several parameter's values
- Argue for H_1 by arguing against H_0
- Construct a statistic $\mathbf{\Lambda}$, called a test or detection statistic
- Calculate $p(\mathbf{\Lambda}/H_0)$: the sampling distribution of $\mathbf{\Lambda}$ under assumption of null hypothesis
- If data distribution different => reject H_0 and accept H_1 at p x 100% level

Frequentist: hypothesis testing

- If data distribution $\neq p(\Lambda|H_0)$ => reject H_0 and accept H_1 at p x 100% level

$$p \equiv prob(\Lambda > \Lambda_{obs}|H_0) \equiv \int_{\Lambda_{obs}}^{\infty} d\Lambda p(\Lambda|H_0)$$

- p-value required to reject the null hypothesis determines a threshold Λ_* .



Romano & Cornish, LRR 2017

François Arago Centre

université PARIS DIDEROT PARIS 7

Frequentist: hypothesis testing

- If data distribution $\neq p(\Lambda|H_0)$ => reject H_0 and accept H_1 at p x 100% level

$$\boxed{p \equiv prob(\Lambda > \Lambda_{obs}|H_0) \equiv \int_{\Lambda_{obs}}^{\infty} d\Lambda p(\Lambda|H_0)}$$

- p-value required to reject the null hypothesis determines a threshold Λ_* .

- Errors: 2 types:
  - FAP: false alarm error: $\Lambda_{obs} > \Lambda_*$: reject H_0 but no signal:

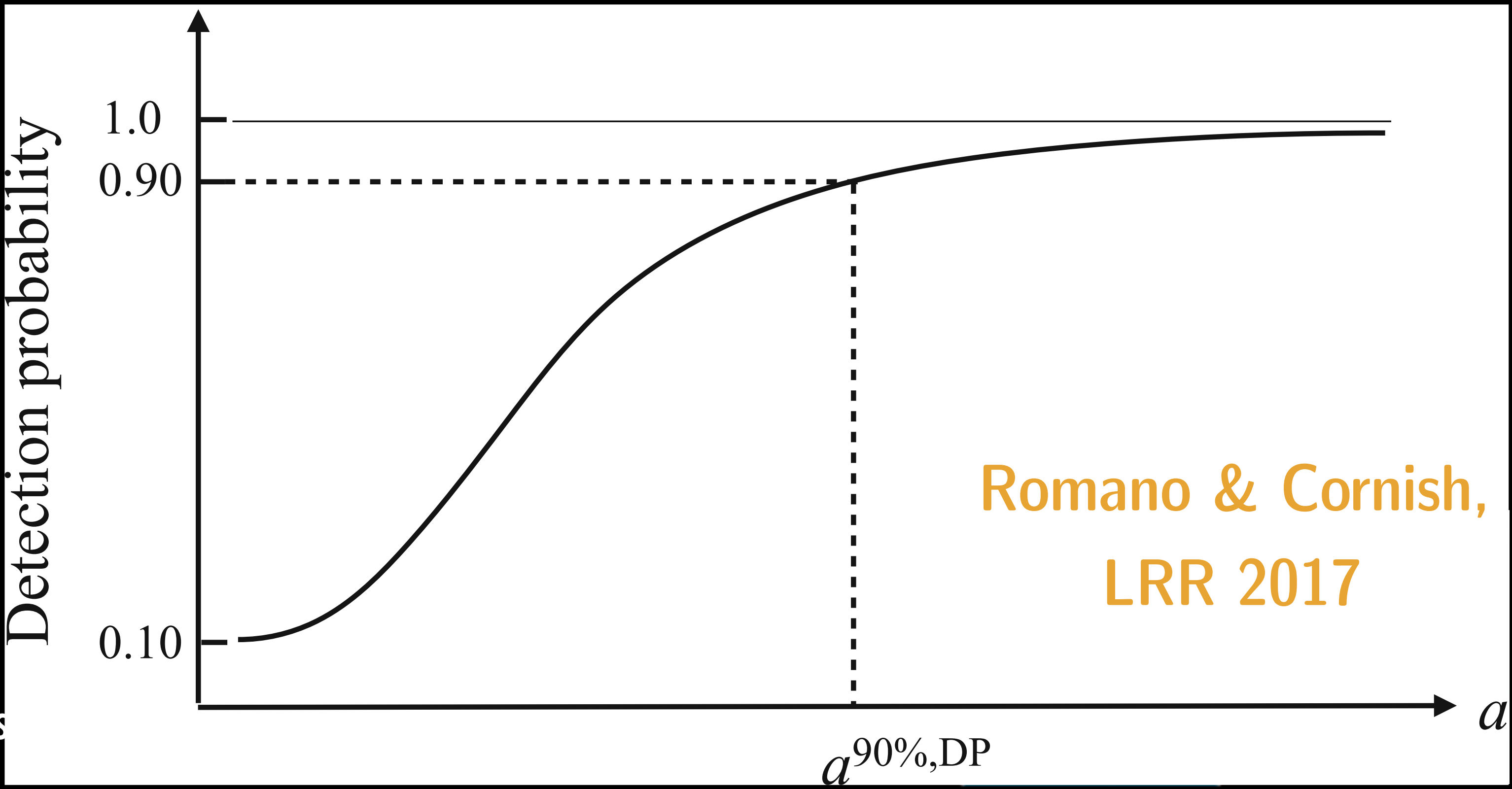
$$\boxed{FAP \equiv \alpha = prob(\Lambda > \Lambda_*|H_0)}$$

  - FDP: false dismissal error: $\Lambda_{obs} < \Lambda_*$: accept H_0 but signal:

$$\boxed{FDP \equiv \beta(a) = prob(\Lambda < \Lambda_*|H_a)}$$

Frequentist: hypothesis testing

- Ideally: FAP and FDP as small as possible … but compete
  - GW: FAP very small
  - Medical: FDP very small
- Newman-Pearson criterion: for fixed FAP, the best statistic is the one minimizing FDP: FAP => $\boldsymbol{\Lambda}_*$.
- Detection probability: $1\text{-}\boldsymbol{\beta}(a) = 1 - prob(\boldsymbol{\Lambda} < \boldsymbol{\Lambda}_*/H_0)$:
  - Independent from data
  - Depends only on
    - sampling distribution
    - FAP



Romano & Cornish, LRR 2017

Frequentist: upper limit

- No detection => set an upper limit (amplitude) based on:
  - Λ_{obs}: observed detection statistic
  - Confidence level
- Example:
  - 90% confidence-level upper-limit $a^{90\%,UL}$

    = minimal value of a for which $\Lambda > \Lambda_{\text{obs}}$ at least 90% of the time

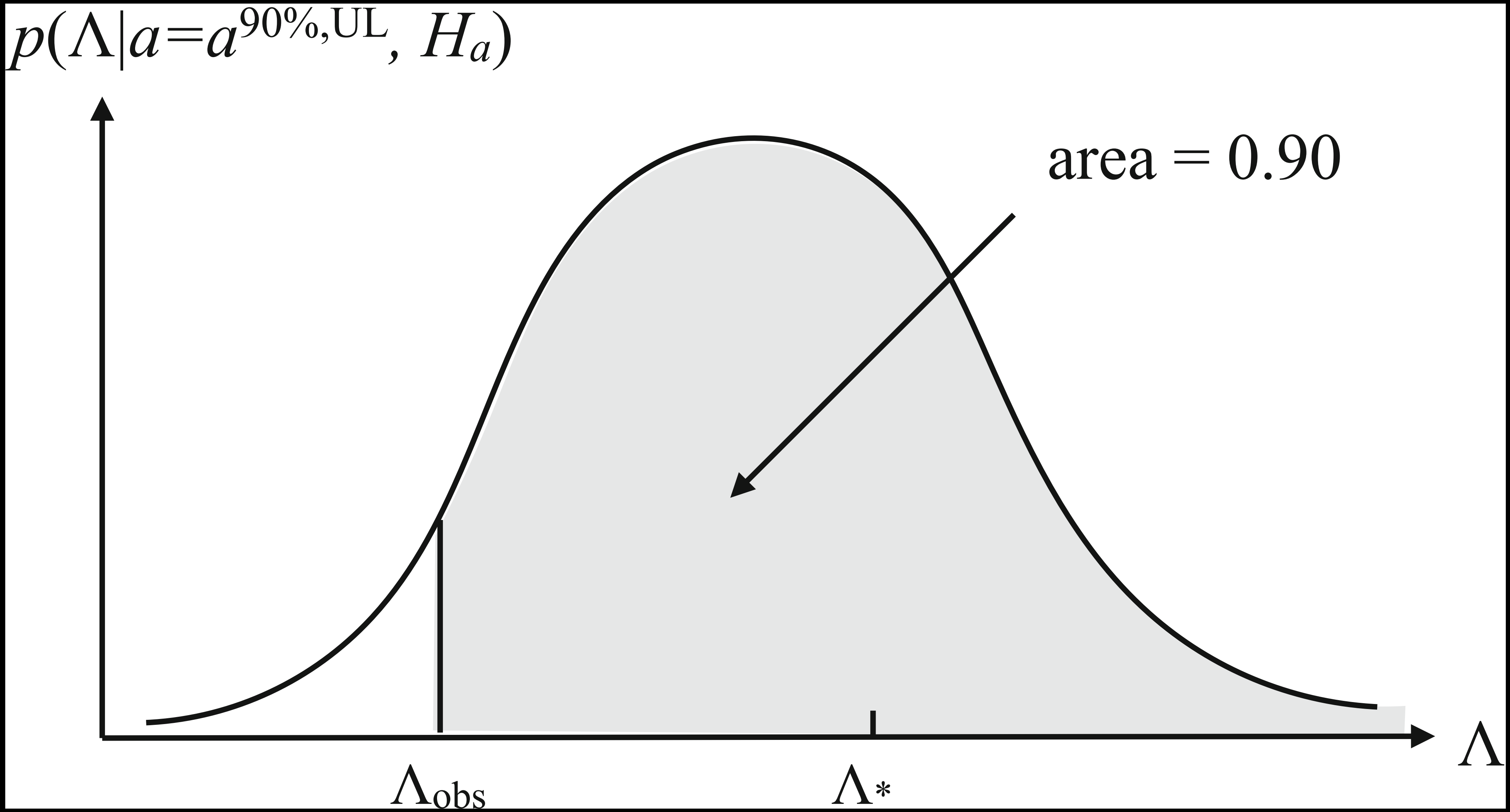
$$prob(\Lambda > \Lambda_{obs} | a \geq a^{90\%,UL}, H_a) \geq 0.9$$

Frequentist: upper limit

- 90% confidence-level upper-limit $a^{90\%,UL}$

  $=$ minimal a for which $\Lambda > \Lambda_{obs}$ at least 90% of the time

$$prob(\Lambda > \Lambda_{obs} | a \geq a^{90\%,UL}, H_a) \geq 0.9$$



Romano & Cornish, LRR 2017

Frequentist: upper limit

- In practice, you can use injection in your data
- Example: GW signal from binary described by amplitude a + other parameters
  1. Test amplitude a
     1.1. Produce fake data by injecting a signal in your data signal with a given a randomizing other parameters
     1.2. Calculate $\boldsymbol{\Lambda}$ for the fake data
     1.3. Repeat 1.1 large number of time
  2. Compute the ratio $N(\boldsymbol{\Lambda} > \boldsymbol{\Lambda}_{obs}) \;/\; N_{total}$
  3. If adjust a and restart from 1. until you get $N(\boldsymbol{\Lambda} > \boldsymbol{\Lambda}_{obs}) / N_{total} =$ confidence level

Frequentist: parameter estimation

- Construct the estimator $\hat{a}$ of the parameter a : it's a statistic that can be maximum likelihood or others estimators.
- Calculate the sampling distribution $p(\hat{a}/a,H_a)$
- Using $p(\hat{a}/a,H_a)$ + a confidence level of 95%, construct the frequentist confidence interval $[\hat{a}\text{-}\boldsymbol{\Delta}, \hat{a}+\boldsymbol{\Delta}]$, such as

$$prob(\hat{a} - \Delta < a < \hat{a} + \Delta) = 0.95$$

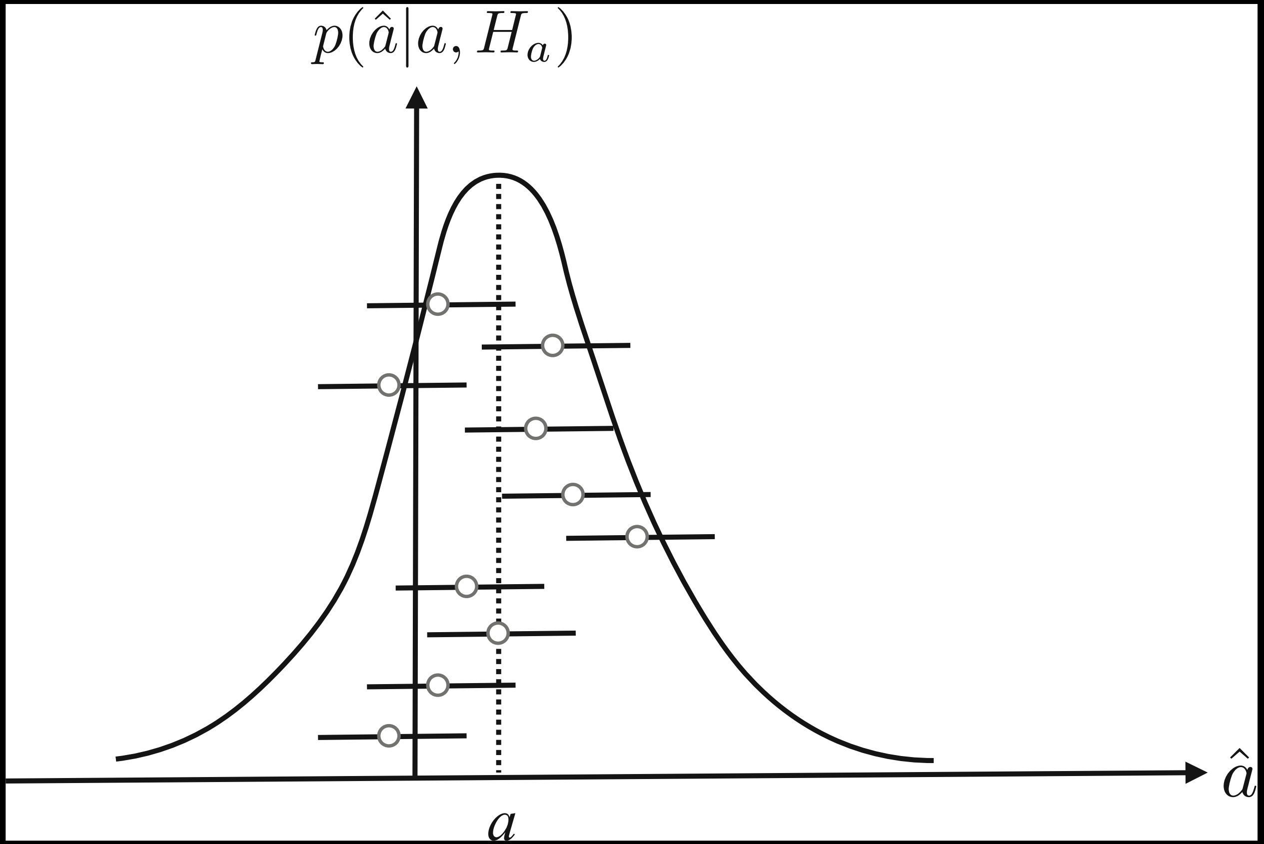
- Interpretation:
  - in a set of many repeated experiments, in 95% of the case the true value of a is in the intervals
  - a not a random variable so its not a probability on a.

Frequentist: parameter estimation

- Using $p(\hat{a}|a,H_a)$ + a frequentist confidence interval of 95%, construct interval $[\hat{a}\text{-}\boldsymbol{\Delta},\hat{a}\text{+}\boldsymbol{\Delta}]$, such as

$$prob(\hat{a}-\Delta < a < \hat{a}+\Delta) = 0.95$$

- Not physical value are allowed.



Romano & Cornish, LRR 2017

Frequentist: summary

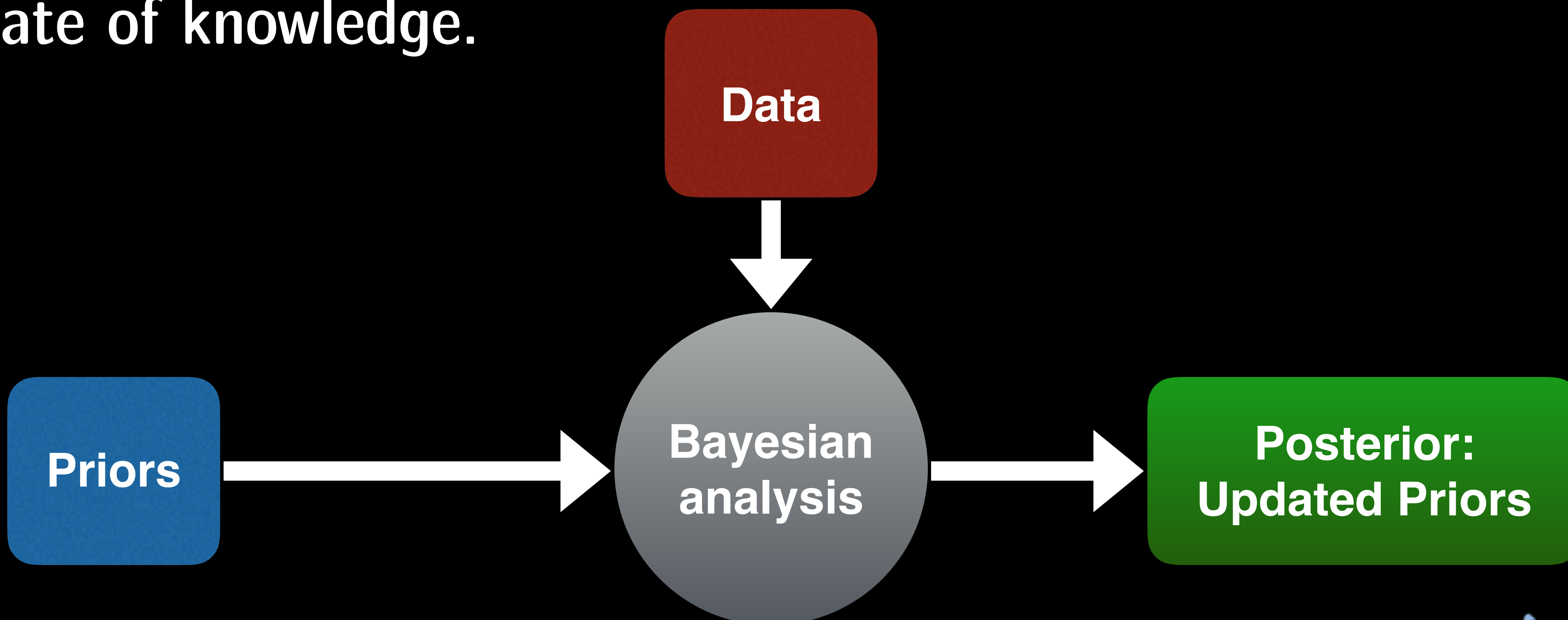
- Uncertainty in the data => define a statistic Λ for the data and define its sampling distribution
- p-value required to reject H_0 => Λ_* :

$$FAP \equiv \alpha = prob(\Lambda > \Lambda_* | H_0)$$

- Detection: compare Λ_{obs} and Λ_*
  - $\Lambda_{obs} > \Lambda_*$ => H_0 rejected => detection
    - Parameter estimation: estimator $\hat{a} \rightarrow$ distribution $p(\hat{a}/a, H_a)$ + confidence level $\rightarrow$ frequentist confidence interval
  - $\Lambda_{obs} < \Lambda_*$ => H_0 accepted => no detection
    - Upper limit: minimal value of a for which $\Lambda > \Lambda_{obs}$ at least CL% of the time, with CL the confidence interval

Bayesian inference

- Data are given
- The uncertainties are on the model / parameters
- Our prior knowledge is updated by what we learn from the data, as measured by the likelihood to give our posterior state of knowledge.



Bayesian inference

- Bayes theorem:

prior

likelihood

$$p(a|d) = \frac{p(a)\ p(d|a)}{p(d)}$$

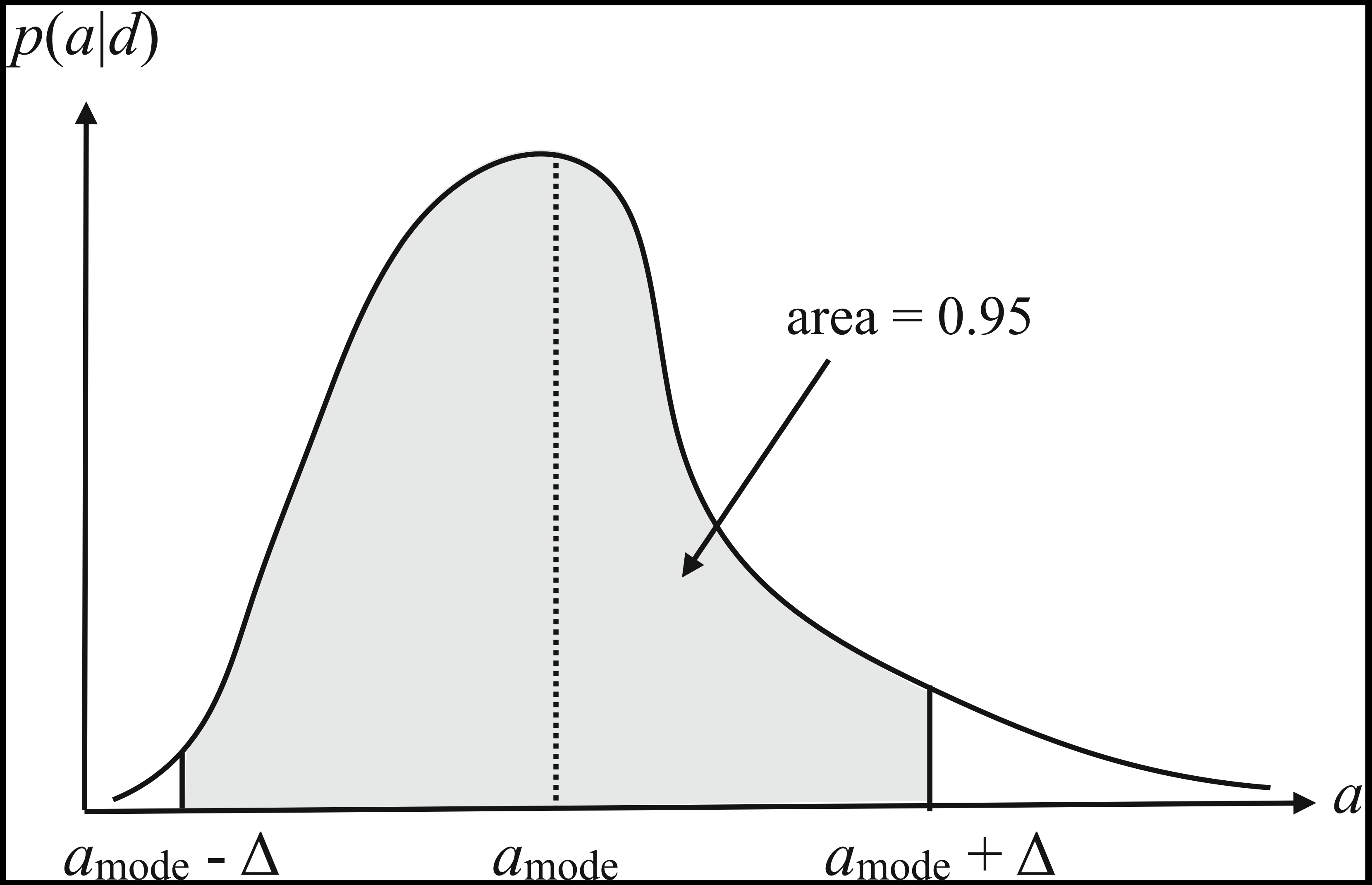
posterior distribution

evidence

- "Everything" about the parameters is in the posterior distribution

Bayesian inference

- Confidence interval = credible interval (degree of belief): area under the posterior between one parameter value and another



Romano & Cornish, LRR 2017

Bayesian: marginalization

- More than one parameter in your model but there are parameters we don't care about => marginalized over them
- Example with 2 parameters a,b:
  - care only about a => marginalized over b
  - $p(a|d) = \int db\, p(a,b|d)$
  - relation between joint probabilities and conditional probabilities

$$p(a,b) = p(a|b)\; p(b)$$

  - => the marginalization over b is simply

$$p(a|d) = \int db\, p(a|b,d)\; p(b)$$

Bayesian: information

- **Information gain:**

$$I = \int da\, p(a|d)\ \log\left(\frac{p(a|d)}{p(a)}\right)$$

- **If there no gain of information from the data, likelihood $\mathcal{L}(a)$ is constant**

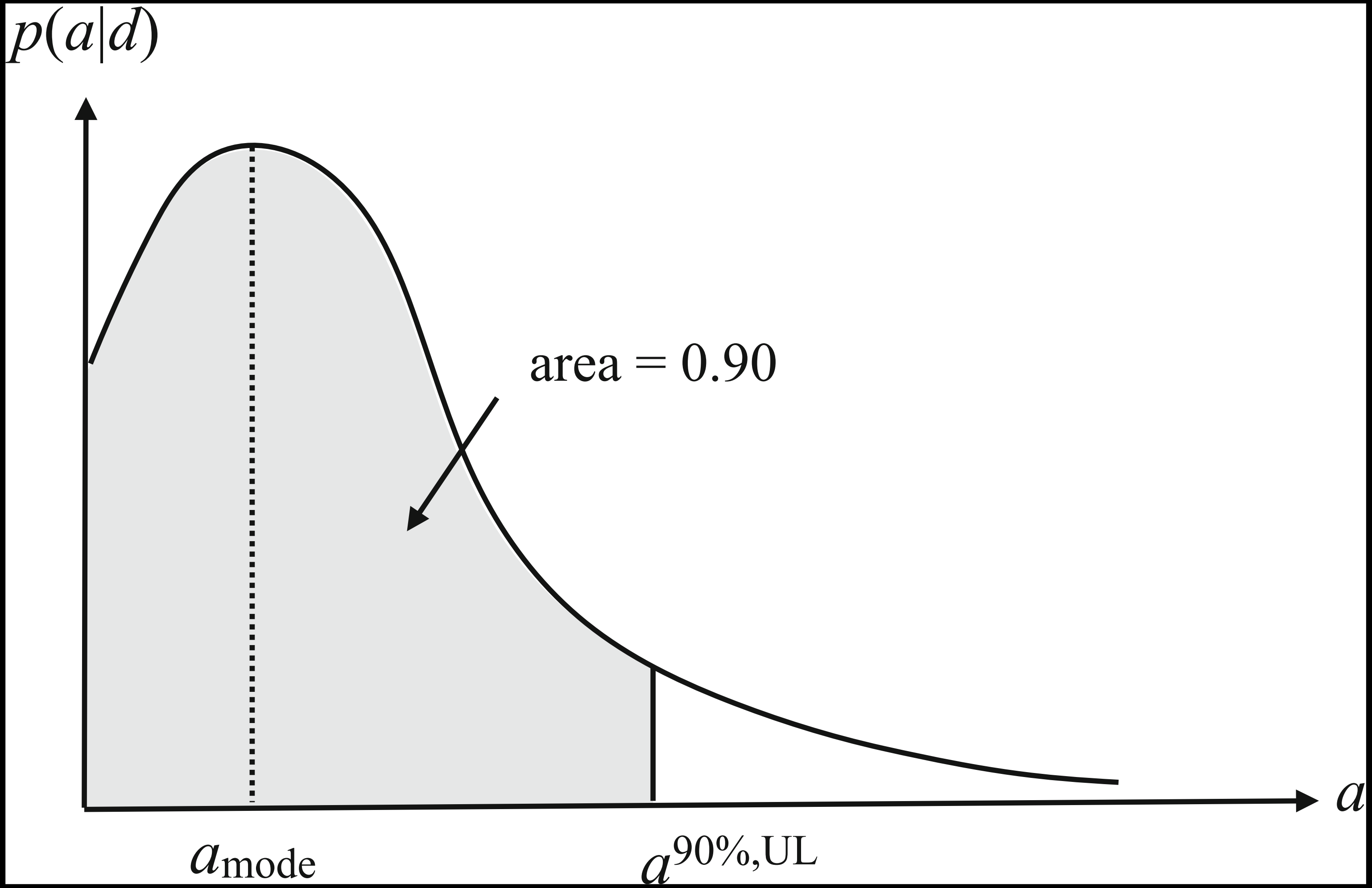
$$=> p(a/d) = p(a)$$

$$=> I = 0$$

Bayesian: upper limit

- If the Bayesian credible interval is compatible with the minimum value for the parameter, we can set un upper limit for a "confidence level":

$$prob(0 < a < a^{UL,90\%}|d) = 0.9$$

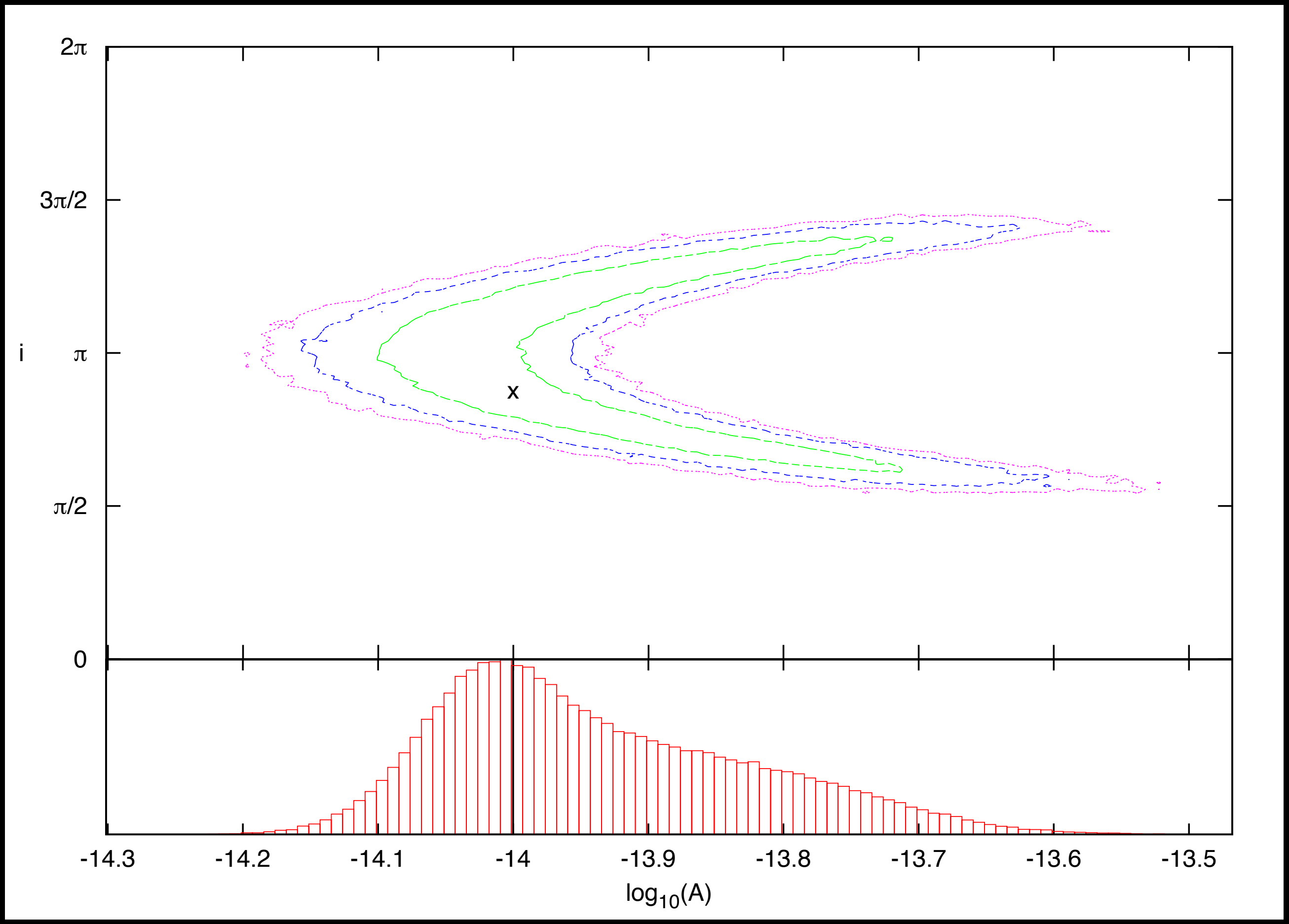


Romano & Cornish, LRR 2017

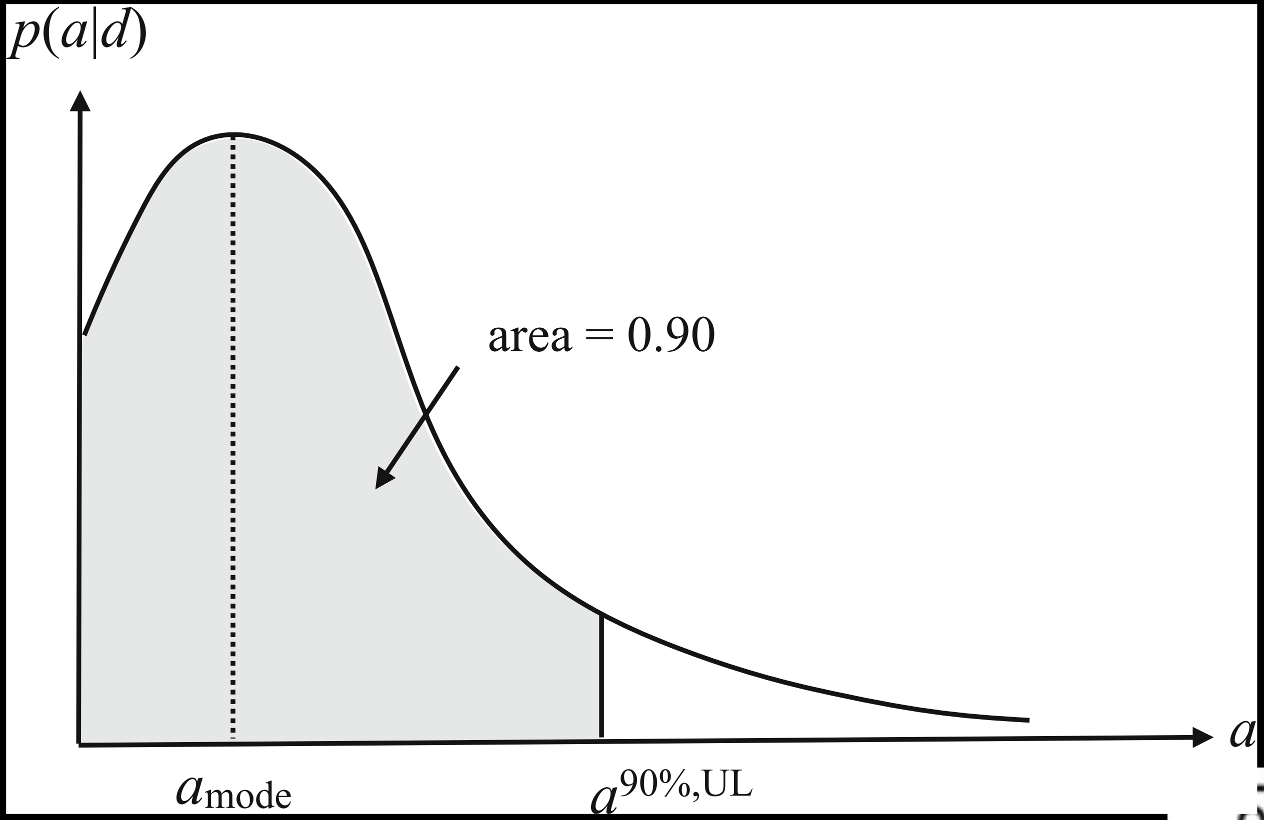
Bayesian: upper limit

- If the parameter is the amplitude
  - Confidence interval exclude 0 => potential detection ...
  - Confidence interval include 0 => result compatible with no detection => upper limit

$$prob(0 < a < a^{UL,90\%}|d) = 0.9$$



Lassus & al.



Romano & Cornish, LRR 2017

Bayesian: model selection

Goal: use Bayes theorem to compare models

- $\mathscr{M}_\alpha$: models $\qquad\qquad$ Θ_α : parameters
- Posterior distribution for given the model :

$$p\left(\theta_\alpha | d, \mathcal{M}_\alpha\right) = \frac{p\left(d | \theta_\alpha \mathcal{M}_\alpha\right)\ p\left(\theta_\alpha | \mathcal{M}_\alpha\right)}{p\left(d | \mathcal{M}_\alpha\right)}$$

- Evidence given a model:

$$p\left(d | \mathcal{M}_\alpha\right) = \int d\theta_\alpha\, p\left(d | \theta_\alpha, \mathcal{M}_\alpha\right) p\left(\theta_\alpha, \mathcal{M}_\alpha\right)$$

Bayesian: model selection

- Posterior probability of models $\mathscr{M}_\alpha$:

$$p\left(\mathcal{M}_\alpha|d\right) = \frac{p\left(d|\mathcal{M}_\alpha\right)p\left(\mathcal{M}_\alpha\right)}{p(d)}$$

- Evidence: sum of all possible model ... but total number unknown => use a subset

$$p(d) = \sum_\alpha p\left(d|\mathcal{M}_\alpha\right)\ p\left(\mathcal{M}_\alpha\right)$$

- Odds ratio between 2 models:

$$\mathcal{O}_{\alpha\beta} = \frac{p\left(\mathcal{M}_\alpha|d\right)}{p\left(\mathcal{M}_\beta|d\right)}$$

Bayesian: model selection

- Odds ratio between 2 models:

$$\mathcal{O}_{\alpha\beta} = \frac{p\left(\mathcal{M}_\alpha | d\right)}{p\left(\mathcal{M}_\beta | d\right)}$$

$$\mathcal{O}_{\alpha\beta} = \boxed{\frac{p\left(d | \mathcal{M}_\alpha\right)}{p\left(d | \mathcal{M}_\beta\right)}} \boxed{\frac{p\left(\mathcal{M}_\alpha\right)}{p\left(\mathcal{M}_\beta\right)}}$$

prior odds ratio

Bayes factor evidence ratio

$$\mathcal{B}_{\alpha\beta} = \frac{p\left(d | \mathcal{M}_\alpha\right)}{p\left(d | \mathcal{M}_\beta\right)}$$

Bayesian: Bayes factor

- Bayes factor: usual tool to compare model, in particular signal versus no signal $\mathscr{M}$
- Problem: interpretation of Bayes factor

$\mathcal{B}_{\alpha\beta}$	$2 \ln \mathcal{B}_{\alpha\beta}(d)$	Evidence for model $\mathscr{M}_\alpha$ relative to $\mathscr{M}_\beta$
< 1	< 0	Negative (supports model $\mathscr{M}_\beta$)
1 - 3	0 - 2	Not worth more than a bare mention
3 - 20	2 - 6	Positive
2 - 150	6 - 10	Strong
>150	>10	Very strong

- Need proper calibration (simulations, ...)

Bayesian in practice

- In practice, we need to sample the parameters space computing likelihood to construct the posterior distribution of parameters.
- Several methods:
  - Monte-Carlo Markov Chain,
  - Metropolis Hasting Markov Chain,
  - Multi-Nest,
  - EMCEE,
  - ...

Bayesian vs frequentist

	Frequentist	Bayesian
Probabilities	Probabilities assigned only to propositions about outcomes of repeatable experiments, not to hypotheses or parameters which have fixed but unknown values	Probabilities can be assigned to hypotheses and parameters since probability is degree of belief in any proposition
Data	Assumes measured data are drawn from an underlying probability distribution, which assumes the truth of a particular hypothesis or model (likelihood function)	Same
Input	Constructs a statistic to estimate a parameter or to decide whether or not to claim a detection	Needs to specify prior degree of belief in a particular hypothesis or parameter
Methods	Calculates the probability distribution of the statistic (sampling distribution)	Uses Bayes' theorem to update the prior degree of belief in light of new data
Results	Constructs confidence intervals and p-values for parameter estimation and hypothesis testing	Constructs posteriors and odds ratios for parameter estimation and hypothesis testing/model comparison

Bayesian vs frequentist: GW obs

- In the past, almost only frequentists
- Now, Bayesian methods become more and more popular
- For all GW observatories, we used the two approaches and hybrid approaches mixing the two.
  - LIGO:
    - methods based on Freq. or Bayesian
  - LISA:
    - mainly Bayesian methods
  - PTA:
    - methods based on Freq. or Bayesian

Bayesian vs frequentist

- But at the end, what we need is computing a large number of likelihood, or equivalent estimators
  - Main computing cost
- Joke about Bayesian inference from a colleague:
  - "That's the beauty of Bayesian inference:
    - likelihood*prior
    - realize that you have no idea how to pick the prior
    - assume flat prior
    - realize is a likelihood computation
  - Now you just computed a likelihood, but you are cool because you did it in a Bayesian way."

Likelihood / noise knowledge

- Depending on the level of knowledge of the noise, different flavors of likelihood. Some examples:
  - Perfectly known (S_n) $=>$ reduced likelihood
  - Known shape components

    $=>$ C_n described using parameters included in the search with model parameters
  - Partially known noise levels, and taking into account heavier tail distribution effects

    $=>$ Student-t [Rover 2011]: each frequency bin follows a multi-variate distribution with ν_j degrees of freedom.

$$\log \mathcal{L} = -\sum_j \frac{\nu_j + 2}{2} \log \left(1 + \frac{1}{\nu_j} \chi^2 \right)$$

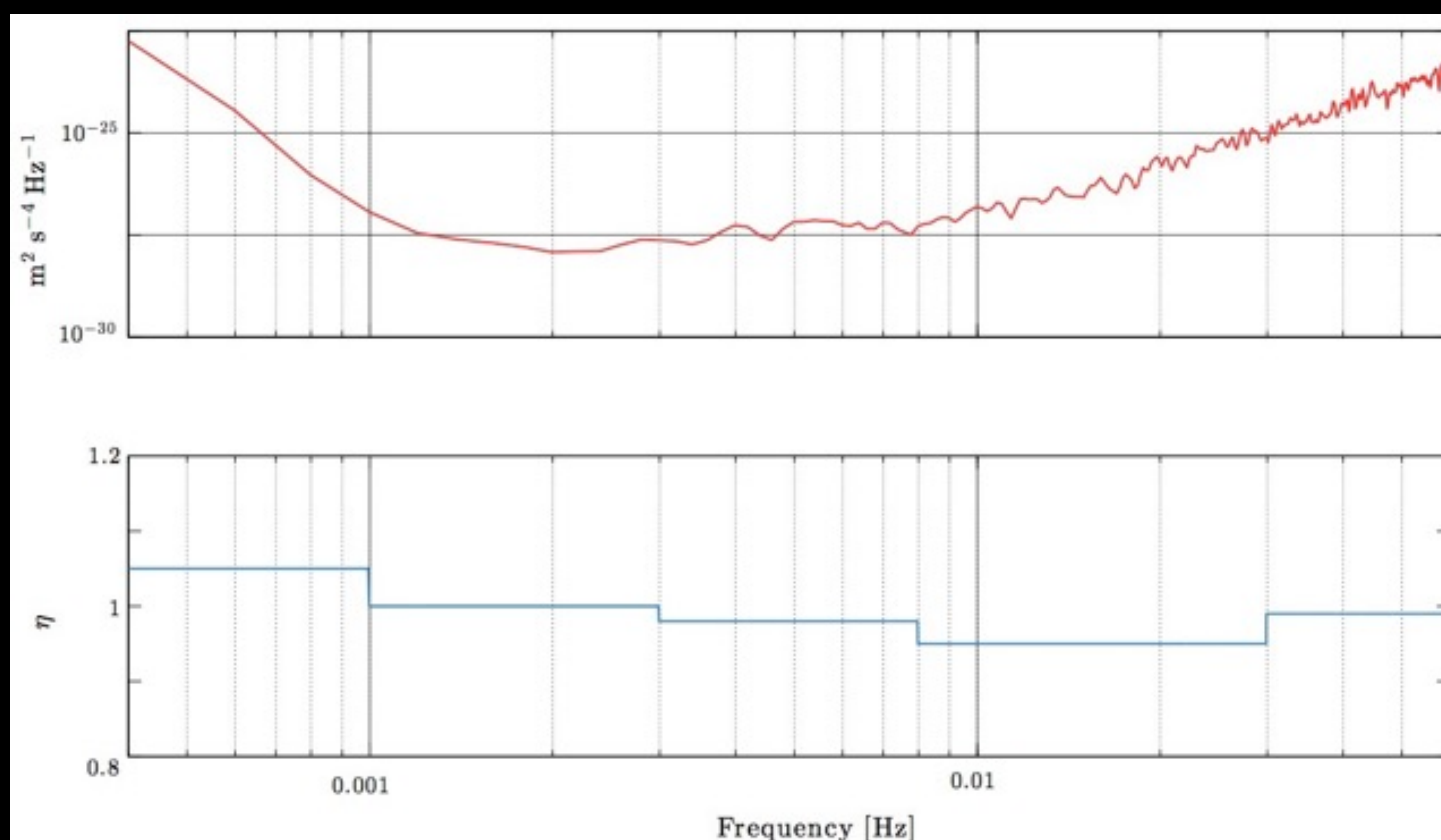
Likelihood / noise knowledge

- **Partially known noise levels but fluctuations of S_n by segment**

  => one parameter per segment

$$S_{n,i} \to \eta_j S_{n,i}, \quad i_j < i \le i_{j+1}$$

$$\log \mathcal{L} = -\frac{1}{2}\left(\chi^2 + N_{j,bins}\sum_j \log \eta_j\right)$$



Thank you