Bayesian inference and ISM studies

some statistical tools to make the most of your observations

“The theory of probabilities is nothing but common sense reduced to calculus.”

Pierre-Simon Laplace, in Théorie Analytique Des Probabilités, 1814

Pierre Palud, APC laboratory

PhD at the interface between data science and ISM

Now PostDoc at the interface between data science and gravitational waves

Top notch statistics for the ISM: why would you invest time in this?

More and more observations

More sensitive instruments

Surveys: large hyperspectral cubes with multiple emission lines

With multiple noise sources and large range of S/N

Resolved and unresolved environments

Mixing different types of environments

Complex astrophysical simulators

"Simple" simulators modelling a physical aspect (RADEX, chemistry, etc.)

Holistic and very complex simulators modelling a specific environment such PDRs, Hii regions, dense cores, shock-dominated regions, etc.

Ex: the Meudon PDD code computes the integrated intensity of ~5400 emission lines, taking into account eg ~3000 chemical reactions

Top notch statistics for the ISM: why would you invest time in this?

More and more observations

More sensitive instruments

Surveys: large hyperspectral cubes with multiple emission lines

With multiple noise sources and large range of S/N

Resolved and unresolved environments

Mixing different types of environments

Statistics (unsupervised)

Complex astrophysical simulators

“Simple” simulators modelling a physical aspect (RADEX, chemistry, etc.)

Holistic and very complex simulators modelling a specific environment such PDRs, Hii regions, dense cores, shock-dominated regions, etc.

Ex: the Meudon PDD code computes the integrated intensity of ~5400 emission lines, taking into account eg ~3000 chemical reactions

Statistics (machine learning, entropy)

Top notch statistics for the ISM: why would you invest time in this?

More and more observations

More sensitive instruments

Surveys: large hyperspectral cubes with multiple emission lines

With multiple noise sources and large range of S/N

Resolved and unresolved environments

Mixing different types of environments

Statistics (unsupervised)

Statistics

(Bayesian inference, model selection, etc.)

Complex astrophysical simulators

“Simple” simulators modelling a physical aspect (RADEX, chemistry, etc.)

Holistic and very complex simulators modelling a specific environment such PDRs, Hii regions, dense cores, shock-dominated regions, etc.

Ex: the Meudon PDD code computes the integrated intensity of ~5400 emission lines, taking into account eg ~3000 chemical reactions

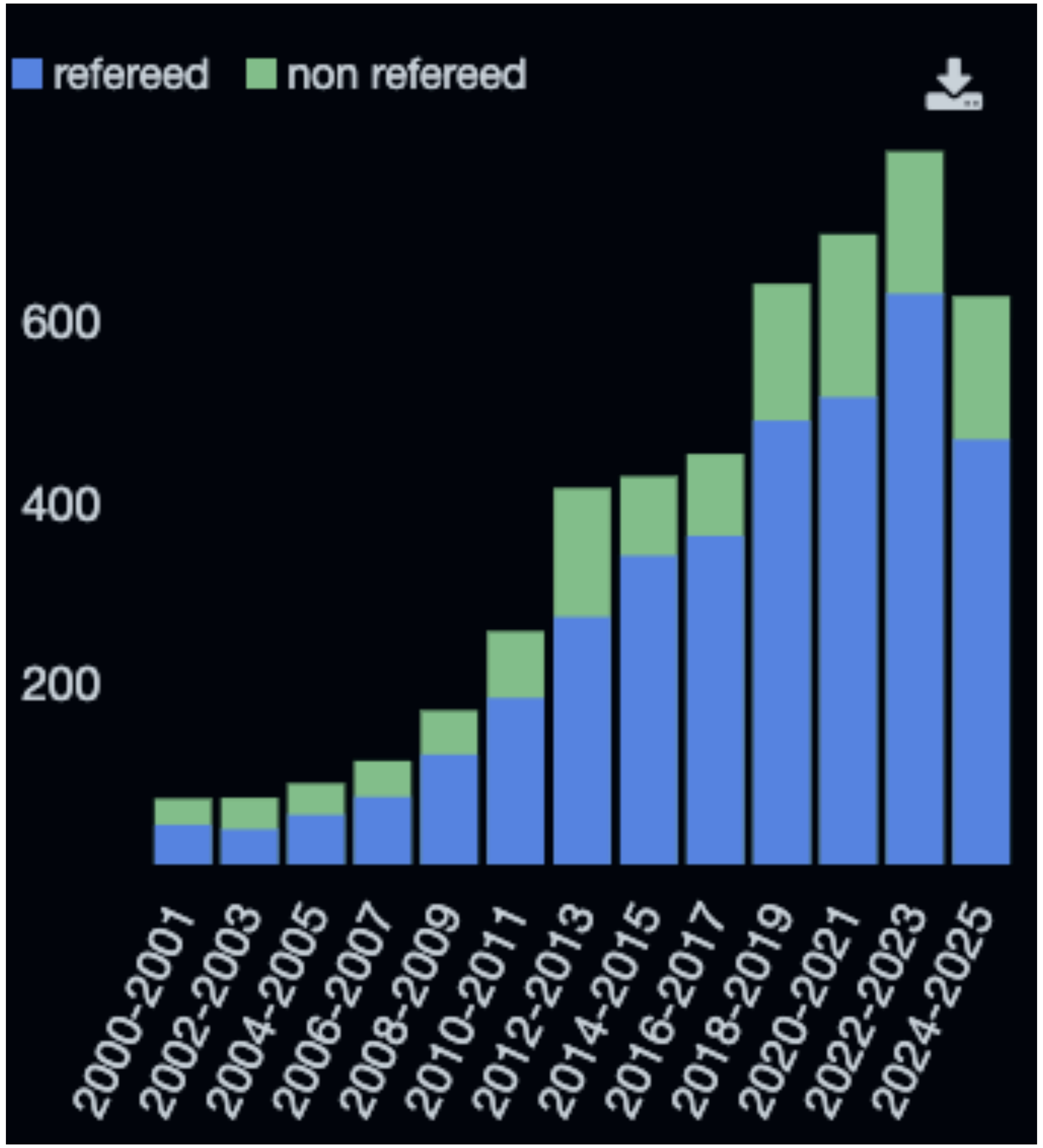
Statistics (machine learning, entropy)

Does “Bayesian” matter? Yes

On Nasa’s Astrophysics Data System (ADS):

- “keyword: statistical” and “abs: Bayesian”
- on 15/07/2025

4,855 articles, with exponential growth



Why does it matter?

Modeling

Include multiple sources or errors (measurement errors such as thermal noise or calibration errors, model mis-specification, etc.)

Account for prior information

Inference

Very natural & statistically principled way of formalising an inference problem

Manage missing or censored data

Results interpretation

Natively describe uncertainty on single parameters or on multiple parameters (variance, covariance, credibility intervals)

Marginalise over “nuisance parameters”

Evaluate probabilities from uncertainty description (including **model assessment,** also called **posterior predictive checking**)

Do I need a degree in statistics to use Bayesian methods?

No, but you need to **understand fundamental notions** of what you are working with.

The goal of this class is to provide

- an overview of what can be done with Bayesian inference
- an understanding of the core concepts and main algorithms
- some tools that you can use on your own data

Plan and goals of the class

Plan

1. Key notions of Bayesian statistics
2. Inference in the *ideal* case: Conjugate priors
3. Inference in the *non-ideal* case: Sampling methods
4. Detailing some applications

Goals: At the end of this class, you should

- Know what the prior, likelihood and posterior are
- Be able to formalise a Bayesian inference task, by identifying the main elements
- Implement the Metropolis-Hastings algorithm on a simple case, and analyse the inference results
- Know some tools to go further and solve more complex problems

Part 1: Bayesian inference & uncertainty quantification

What kind of problems are we looking at?

Notation

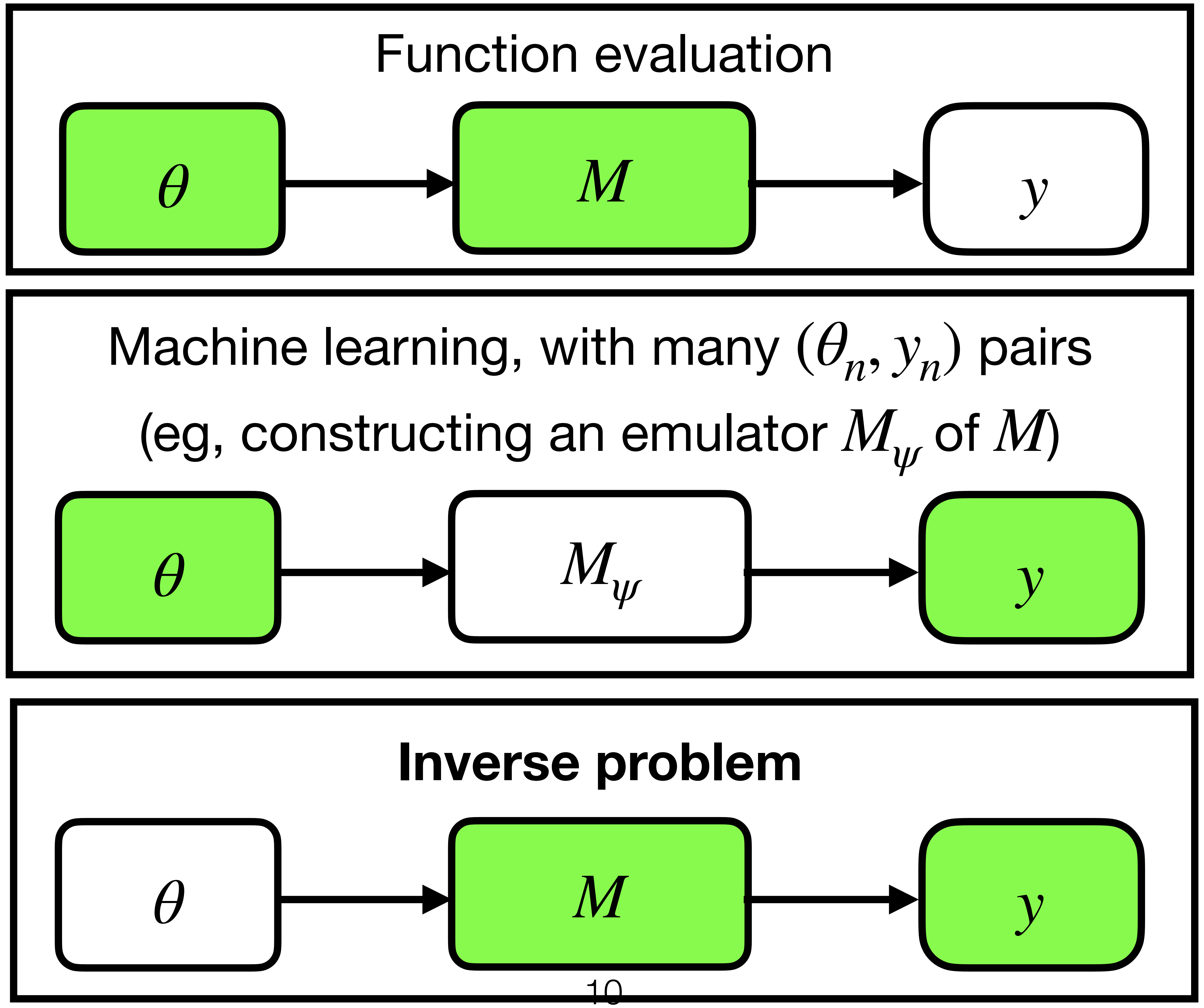
$\theta \in \Theta$: physical parameters

$y \in \mathscr{Y}$: observations

$M : \Theta \mapsto \mathscr{Y}$: a map from the parameter space to observation space (Eg, an astrophysical simulator such as CLOUDY, RADEX, Meudon PDR

M is often called "forward model"

What kind of problem are we interested in?



What kind of problems are we looking at?

Notation

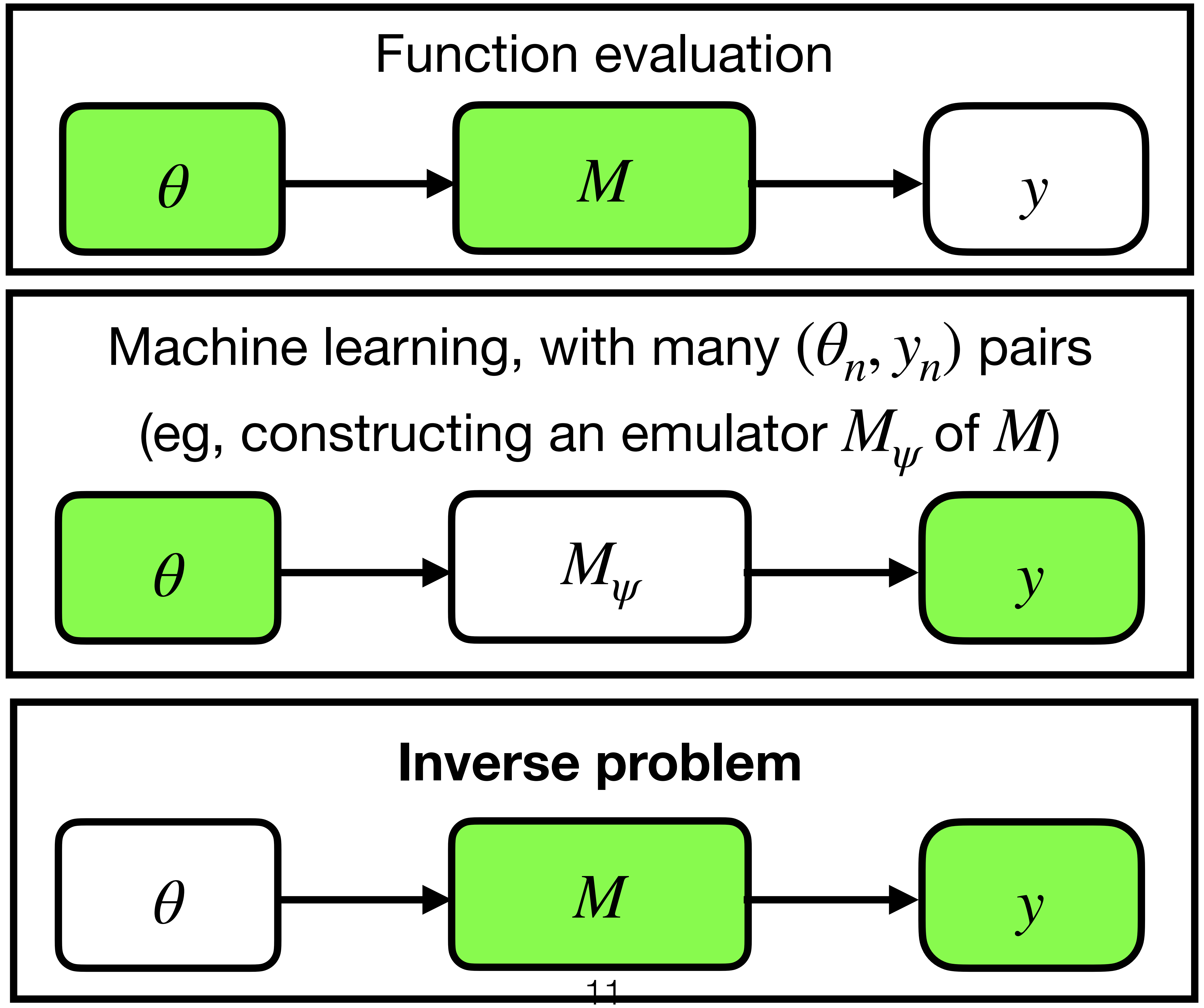
$\theta \in \Theta$: physical parameters

$y \in \mathscr{Y}$: observations

$M : \Theta \mapsto \mathscr{Y}$: a map from the parameter space to observation space (Eg, an astrophysical simulator such as CLOUDY, RADEX, Meudon PDR

M is often called "forward model"

What kind of problem are we interested in?



What about uncertainty?

Assume uncertainty on y and θ. How to describe those?

If I have uncertainty on θ, How does it propagate to y?

If I have a small dataset, how confident should I be with my emulator?

If y is affected by noise and M Is not invertible, what can I say about θ ?

Conceptually, what does the "probability of a random event" quantify?

Random event = an occurence that cannot be predicted with certainty

that is, where we don't know everything with absolute precision. This includes deterministic processes with limited knowledge of the physics or of the initial conditions

Exemples of random events:

Detecting at least 1 gravitational wave signal next week

($\simeq$ Coin toss)

The visual extinction in the Orion Bar nebula is ≥ 10 mag

Conceptually, what does the "probability of a random event" quantify?

Random event = an occurence that cannot be predicted with certainty

that is, where we don't know everything with absolute precision. This includes deterministic processes with limited knowledge of the physics or of the initial conditions

Two different visions of **Probability**

	Limit relative frequency of occurence	Degree of belief in occurence
	Frequentist paradigm	Bayesian paradigm
Exemples of random events:		
Detecting at least 1 gravitational wave signal next week ($\simeq$ Coin toss)	$\simeq$ ✓	✓
The visual extinction in the Orion Bar nebula is ≥ 10 mag	✗	✓

Describing uncertainty:
Random variables and probability distributions

Random variable (rv) = numerical representation of the outcomes of a random event

that is, a quantity with uncertain value due e.g., to lack of observation or to measurement errors

In this presentation: 2 main types of rv

The observations

y

The unknown parameter that we want to estimate

θ

Describing uncertainty
Random variables and probability distributions

Random variable (rv) = numerical representation of the outcomes of a random event

that is, a quantity with uncertain value due e.g., to lack of observation or to measurement errors

In this presentation: 2 main types of rv

The observations

y

The unknown parameter that we want to estimate

θ

Probability distribution = describes the uncertainty on a random variable

For a discrete rv

$\pi(\cdot)$: probability mass, verifies

$$\forall k \in \mathbb{N},\ \pi(k) \geq 0$$

And $\displaystyle\sum_{k=0}^{\infty} \pi(k) = 1$

Examples:

→ Bernoulli distribution

→ Poisson distribution

For a continuous rv

$\pi(\cdot)$: probability density, verifies

$$\forall x,\ \pi(x) \geq 0$$

And $\displaystyle\int \pi(x)\ dx = 1$

Examples:

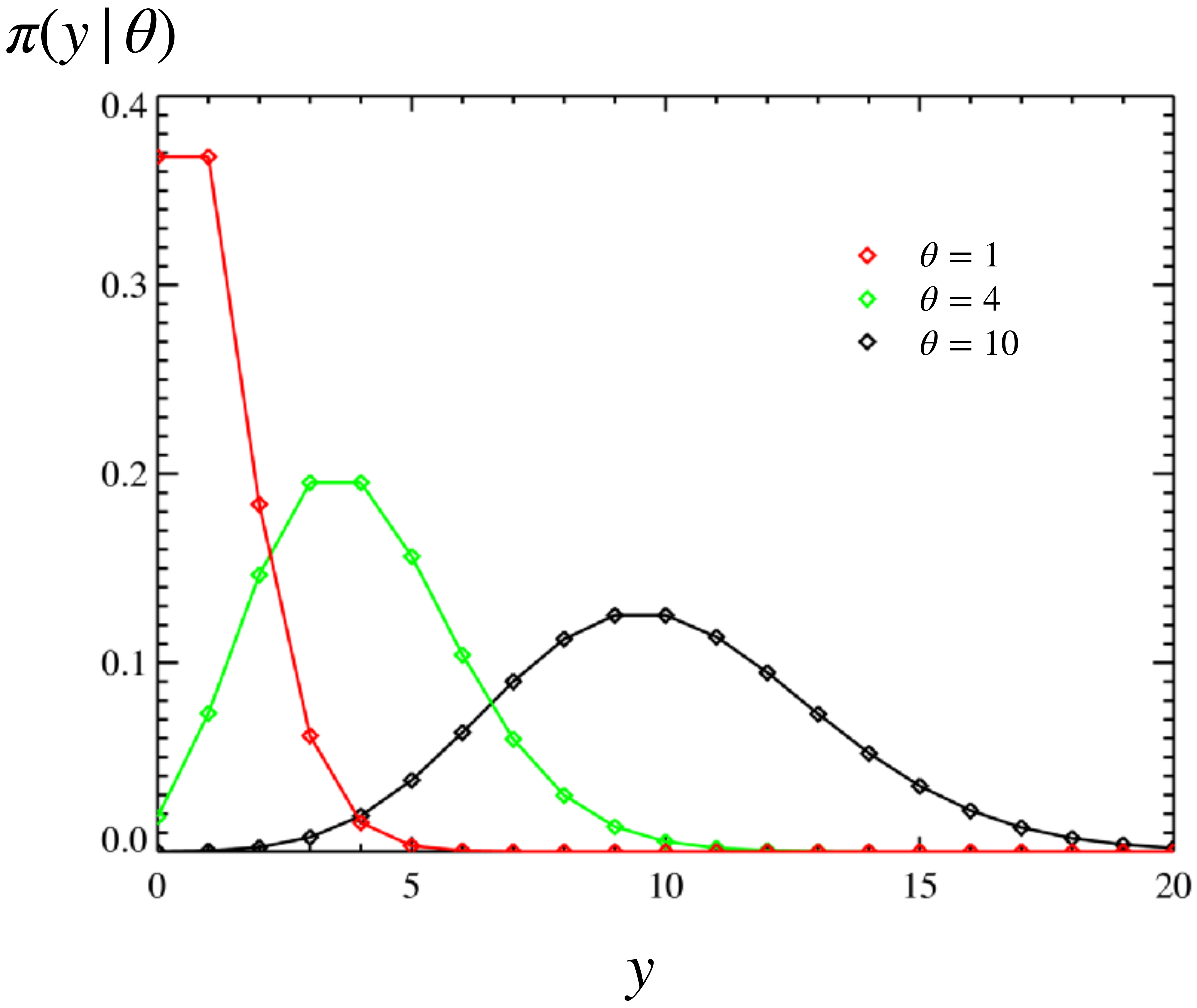
→ Gaussian distribution

→ Beta & Gamma distributions

Examples of probability distributions

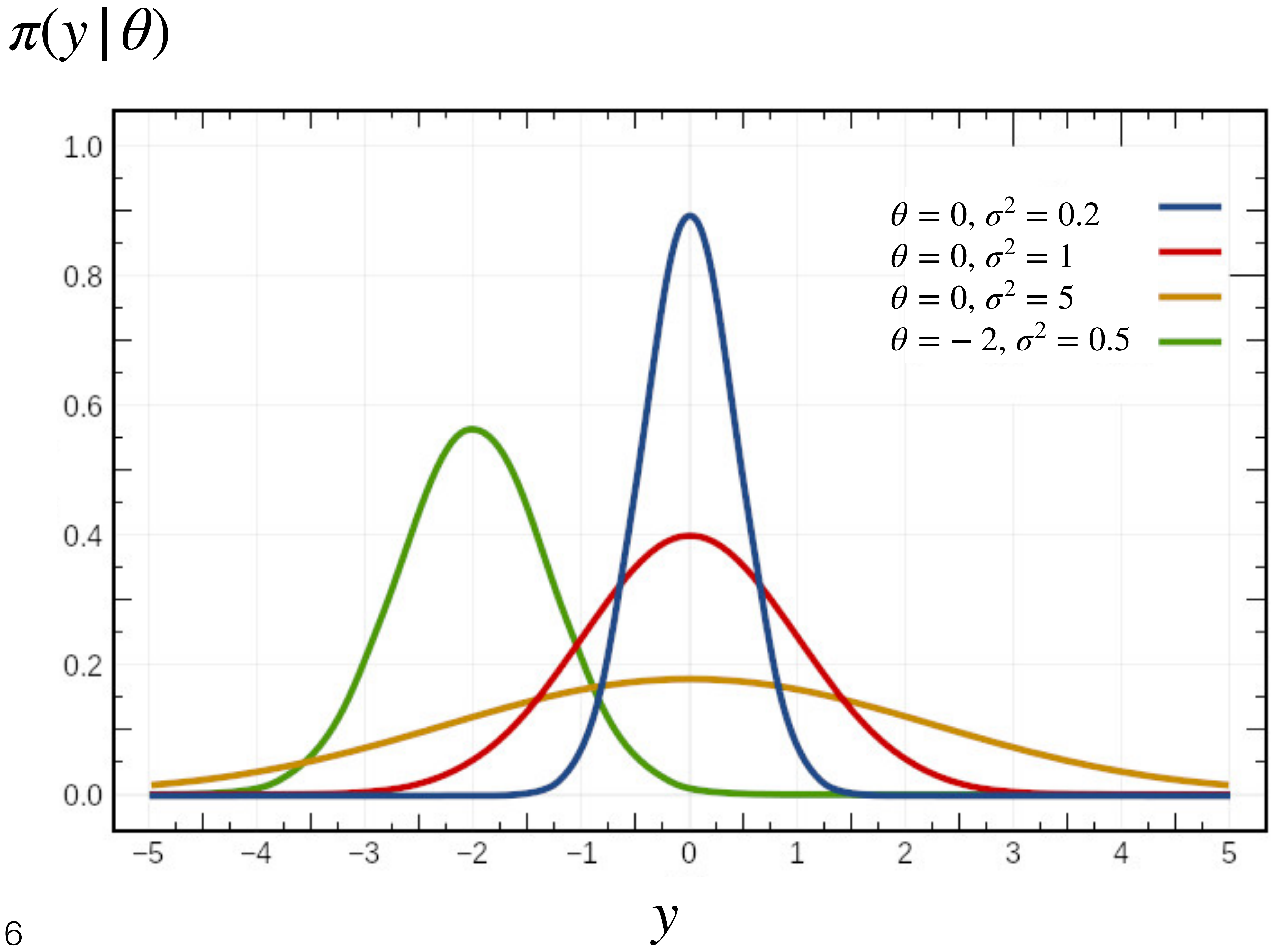
discrete rv: Poisson distribution

$$y \sim \text{Poisson}(\theta),\ \theta > 0$$



continuous rv: The normal distribution

$$y \sim \mathcal{N}(\theta, \sigma^2),\ \theta \in \mathbb{R}$$



The Bayes theorem

Joint probability of θ and y

Proportional to (when y is known)

$$\pi(\theta \mid y) = \frac{\pi(\theta, y)}{\pi(y)} = \frac{\pi(y \mid \theta)\ \pi(\theta)}{\pi(y)} \propto \pi(y \mid \theta)\ \pi(\theta)$$

Probability of θ when y is known

The Bayes theorem

Posterior distribution

the target: my updated knowledge after including my observation

Likelihood function

How surprising my observation y is for this value of θ

Prior distribution

What I already know

$$\pi(\theta \mid y) = \frac{\pi(\theta, y)}{\pi(y)} = \frac{\pi(y \mid \theta)\ \pi(\theta)}{\pi(y)} \propto \pi(y \mid \theta)\ \pi(\theta)$$

Probability of θ when y is known

Bayesian evidence

Normalisation constant

The Bayes theorem

Posterior distribution

the target: my updated knowledge after including my observation

Likelihood function

How surprising my observation y is for this value of θ

Prior distribution

What I already know

$$\pi(\theta | y) = \frac{\pi(\theta, y)}{\pi(y)} = \frac{\pi(y | \theta)\ \pi(\theta)}{\pi(y)} \propto \pi(y | \theta)\ \pi(\theta)$$

Probability of θ when y is known

Bayesian evidence

Normalisation constant

Bayesian inference / estimation / reconstruction

=

Define the posterior and then extract information from it

The 3 steps in Bayesian inference

1/ Describe the observation model → defines the **likelihood function** $\pi(y \mid \theta)$

$\Longrightarrow$ typically in astro: $y \mid \theta \sim \mathcal{N}\left(M(\theta), \sigma^2\right)$, with M an astrophysical simulator (RADEX, Meudon PDR, etc.)

2/ Choose a **prior distribution** $\pi(\theta)$

$\Longrightarrow$ typically in astro: uniform on validity intervals, spatial regularisation for images, etc.

3/ Extract estimators from the **posterior distribution** $\pi(\theta \mid y)$

$\Longrightarrow$ typically the mean, variance, credibility intervals, or the probability of a random event

2 cases for Step 3: is the prior **conjugate** to the likelihood function?

Yes (simple case)

(There is a list of them)

The posterior distribution is from the same distribution family as the prior, everything comes in closed-form expressions

No (almost every time)

(As soon as there is a non-linear model involved)

need to evaluate estimators numerically (e.g., with MCMC algo.)

Summary of part 1

Random event = where we don't know everything with absolute precision

Probability of a random event (in Bayesian paradigm) = degree of belief of occurence

Random variable = a quantity with uncertain value

Probability distribution: describes the uncertainty in a random variable

Bayesian inference: update the *uncertainty description* on a *rv* after an observation, from a prior one to a posterior one, thanks to *Bayes theorem:*

$$\pi(\theta \mid y) \propto \pi(y \mid \theta)\ \pi(\theta)$$

Prior: Initial uncertainty description on θ

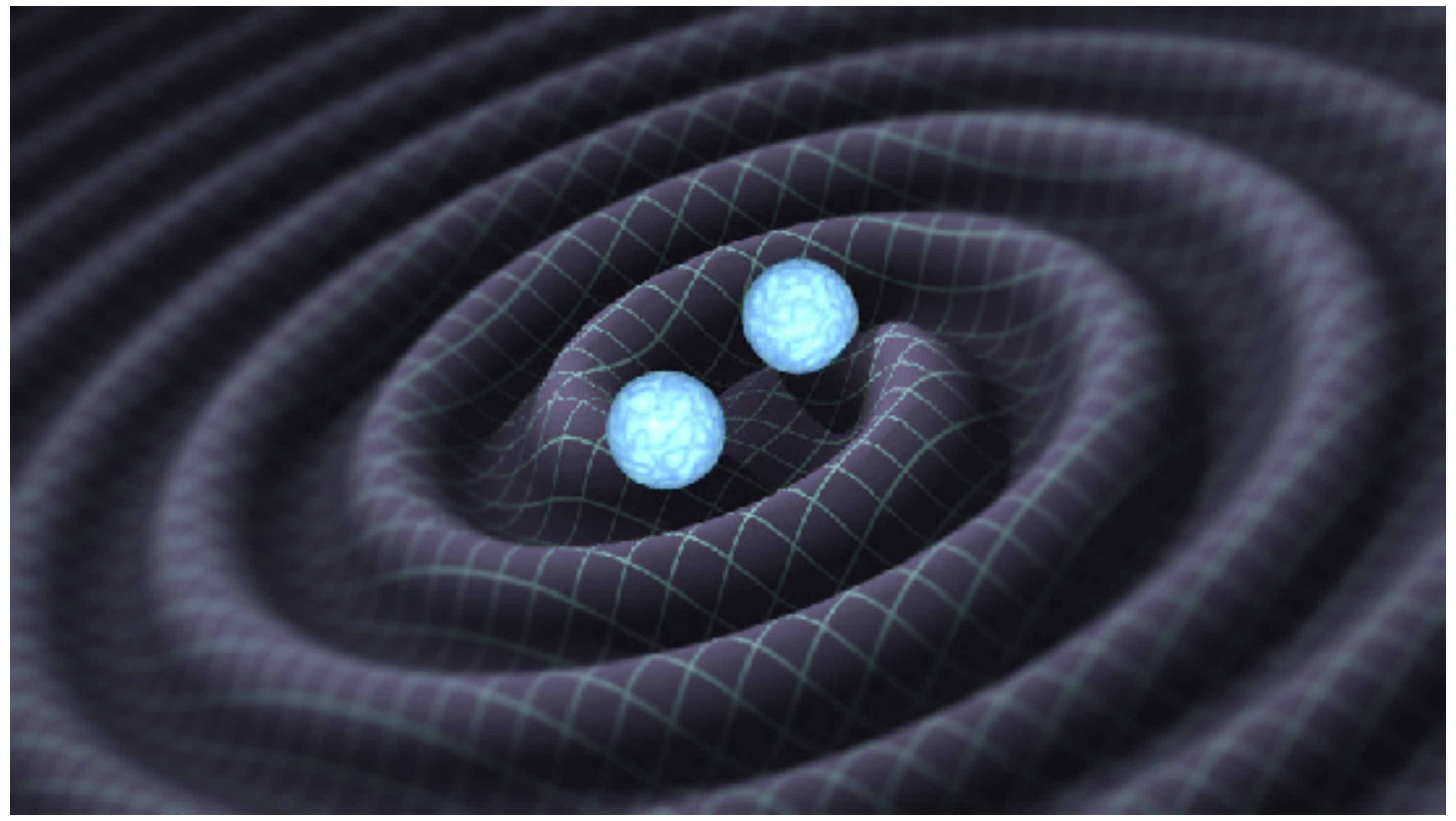
Likelihood: How surprising observing y is for a given value of θ

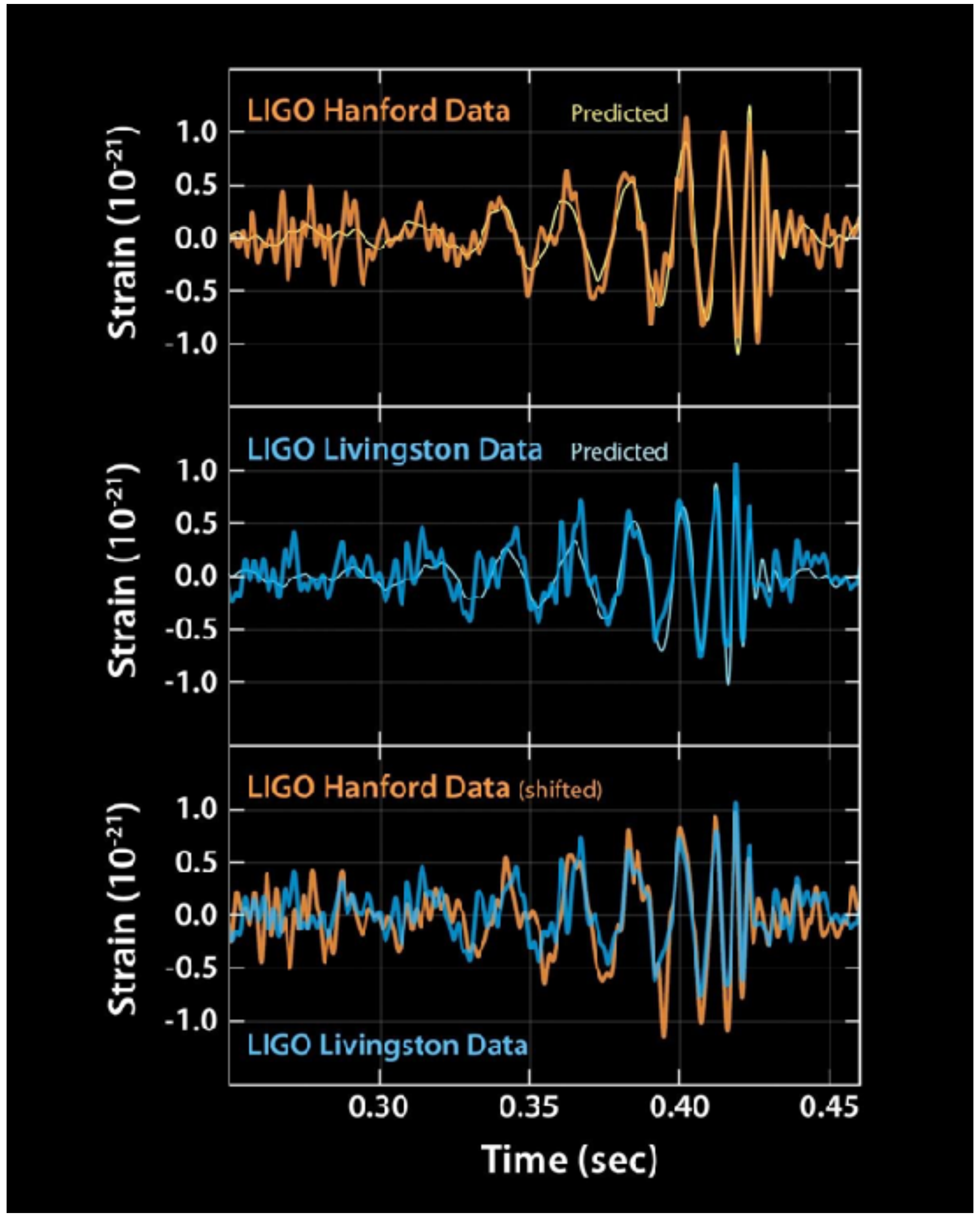
Posterior: Updated uncertainty description on θ

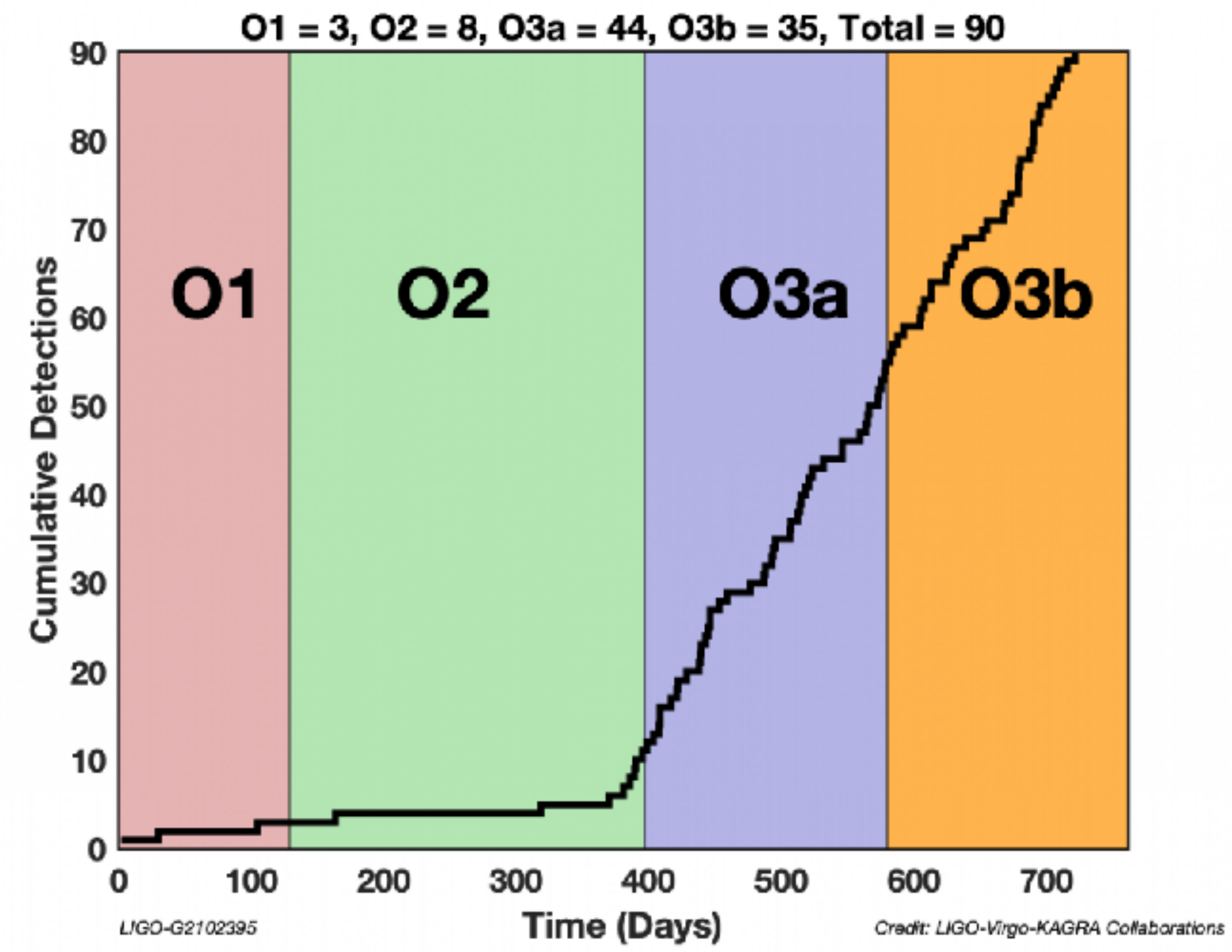
Estimators from the posterior can have closed-form if the prior is conjugate to the likelihood, otherwise need to evaluate them numerically

Part 2: Bayesian inference with conjugate priors: the ideal case

Example 1: probability of detecting at least one gravitational wave signal within the coming week







Example 1: probability of detecting at least one gravitational wave signal within the coming week

1/ Describe the observation model

Random event: “did I observe a GW signal on week n?”

Associated random variable: $y \in \{0,1\}$

Data: N binary observations $y_n \in \{0,1\}$

Hypothesis: assume the observations of the N weeks to be *independent* and *identically distributed*

Model: each y_n had the same probability $\theta \in [0,1]$ to detect at least one GW signal → **we want to estimate** θ

→ *identically distributed*: Bernoulli distribution $y_n | \theta \sim \text{Ber}(\theta)$

→ *independent*: data distribution is $\pi\left(\{y_n\}_{n=1}^{N} | \theta\right) = \prod_{n=1}^{N} \pi\left(y_n | \theta\right)$

Example 1: probability of detecting at least one gravitational wave signal within the coming week

1/ Describe the observation model

Random event: "did I observe a GW signal on week n?"

Associated random variable: $y \in \{0,1\}$

Data: N binary observations $y_n \in \{0,1\}$

Hypothesis: assume the observations of the N weeks to be *independent* and *identically distributed*

Model: each y_n had the same probability $\theta \in [0,1]$ to detect at least one GW signal $\rightarrow$ **we want to estimate** θ

$\rightarrow$ *identically distributed*: Bernoulli distribution $y_n | \theta \sim \text{Ber}(\theta)$

$\rightarrow$ *independent*: data distribution is $\pi\left(\{y_n\}_{n=1}^{N} | \theta\right) = \prod_{n=1}^{N} \pi\left(y_n | \theta\right)$

2/ Choose the prior: the conjugate prior to a Bernoulli likelihood is the **Beta distribution** $\theta \sim \text{Beta}(\alpha, \beta)$

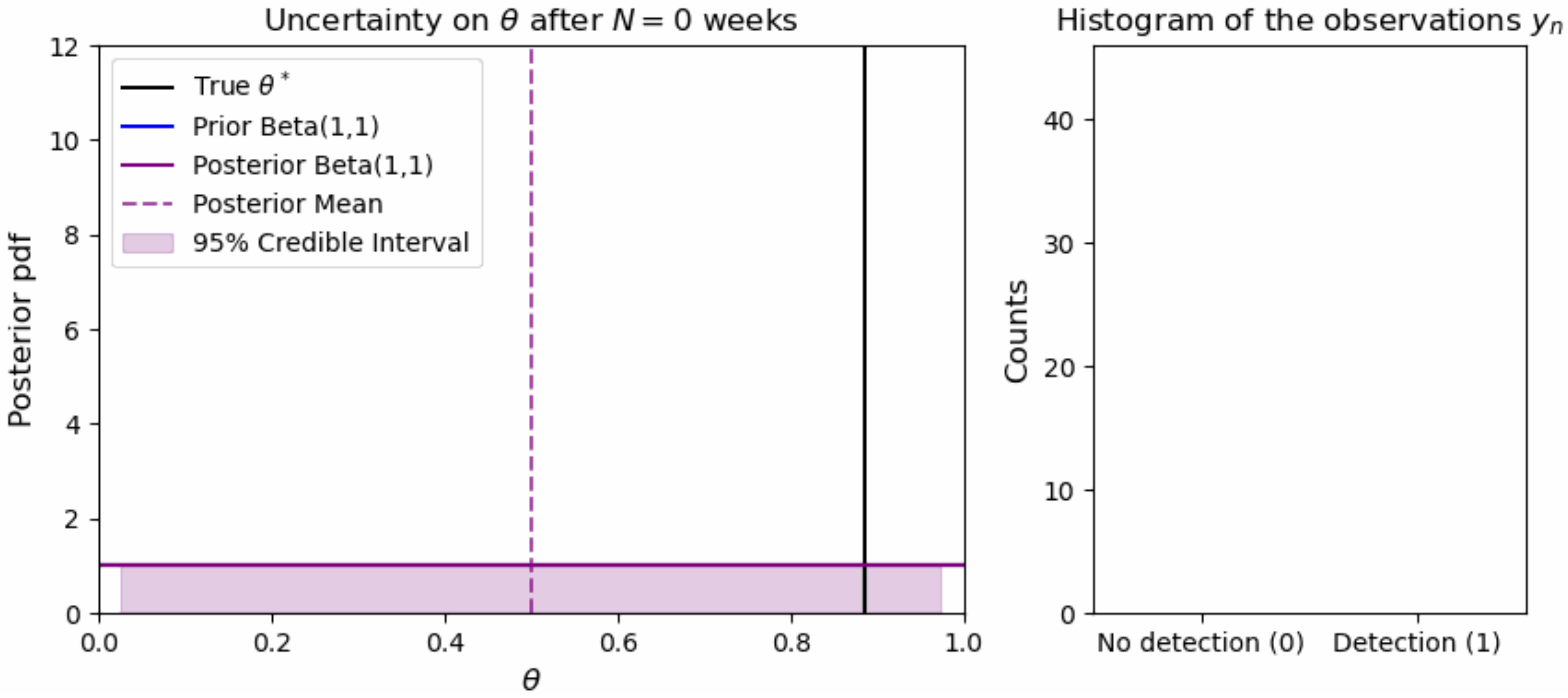
In this case, the posterior is:

$$\theta | \{y_n\}_{n=1}^{N} \sim \text{Beta}\left(\alpha + \sum_{n=1}^{N} y_n, \beta + N - \sum_{n=1}^{N} y_n\right)$$

Example 1: probability of detecting at least one gravitational wave signal within the coming week

$$\text{posterior } \theta \mid \{y_n\}_{n=1}^{N} \sim \text{Beta}\left(\alpha + \sum_{n=1}^{N} y_n,\ \beta + N - \sum_{n=1}^{N} y_n\right)$$

A priori: uniform distribution on $[0,1]$ with $\alpha, \beta = 1$



Each new frame = "new week", $N \to N+1$

$y_n = 0 \implies$ small move to the left

$y_n = 1 \implies$ small move to the right

Remarks:

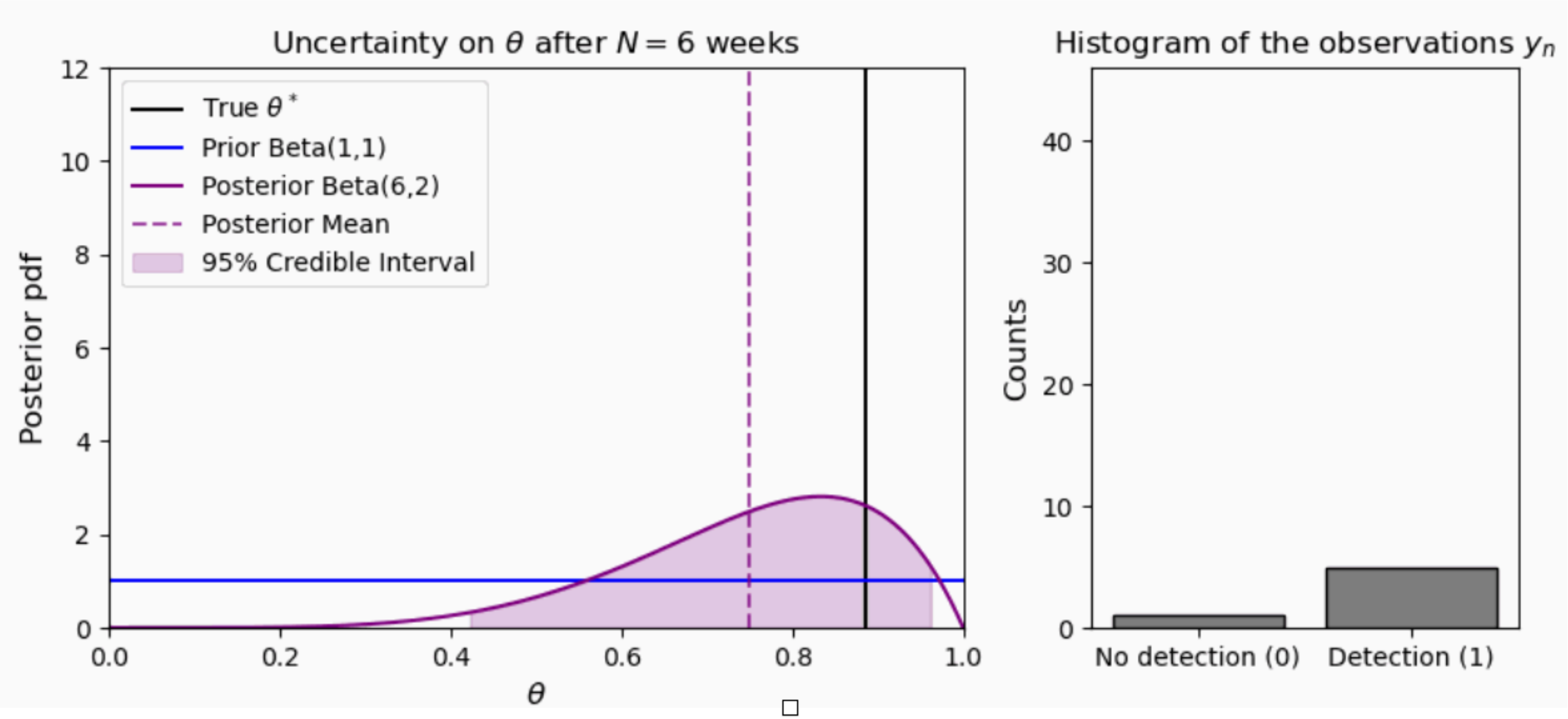
1/ For every value of N, the posterior can describe the uncertainty on θ

2/ As N increases, the variance decreases and the posterior mean converges to the true value

Example 1: probability of detecting at least one gravitational wave signal within the coming week

$$\text{posterior } \theta \,|\, \{y_n\}_{n=1}^N \sim \text{Beta}\left(\alpha + \sum_{n=1}^N y_n, \beta + N - \sum_{n=1}^N y_n\right)$$

A priori: uniform distribution on $[0,1]$ with $\alpha, \beta = 1$



Each new frame = "new week", $N \rightarrow N+1$

$y_n = 0 \implies$ small move to the left

$y_n = 1 \implies$ small move to the right

Remarks:

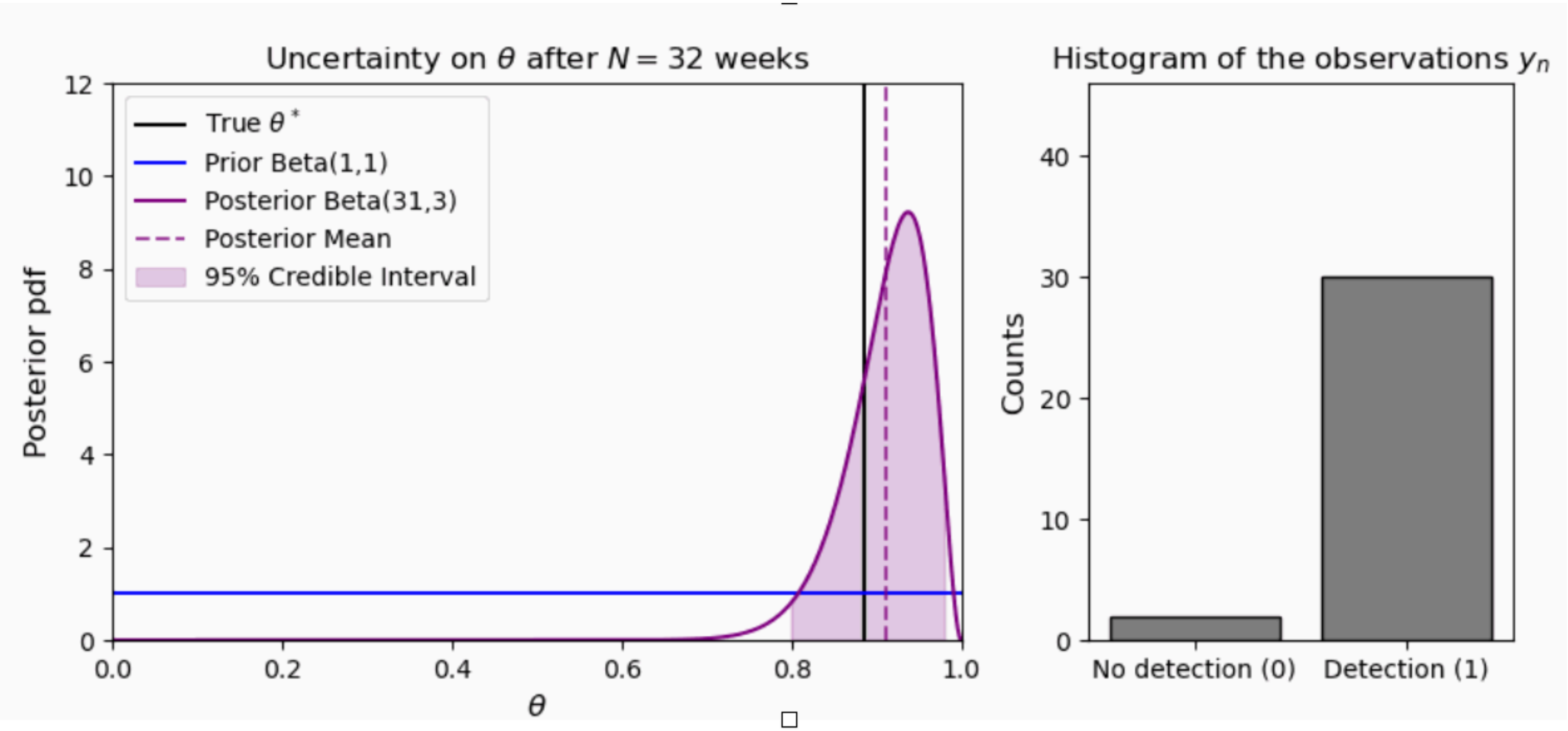
1/ For every value of N, the posterior can describe the uncertainty on θ

2/ As N increases, the variance decreases and the posterior mean converges to the true value

Example 1: probability of detecting at least one gravitational wave signal within the coming week

$$\text{posterior } \theta \,|\, \{y_n\}_{n=1}^{N} \sim \text{Beta}\left(\alpha + \sum_{n=1}^{N} y_n,\ \beta + N - \sum_{n=1}^{N} y_n\right)$$

A priori: uniform distribution on $[0,1]$ with $\alpha, \beta = 1$



Each new frame = "new week", $N \rightarrow N+1$

$y_n = 0 \Longrightarrow$ small move to the left

$y_n = 1 \Longrightarrow$ small move to the right

Remarks:

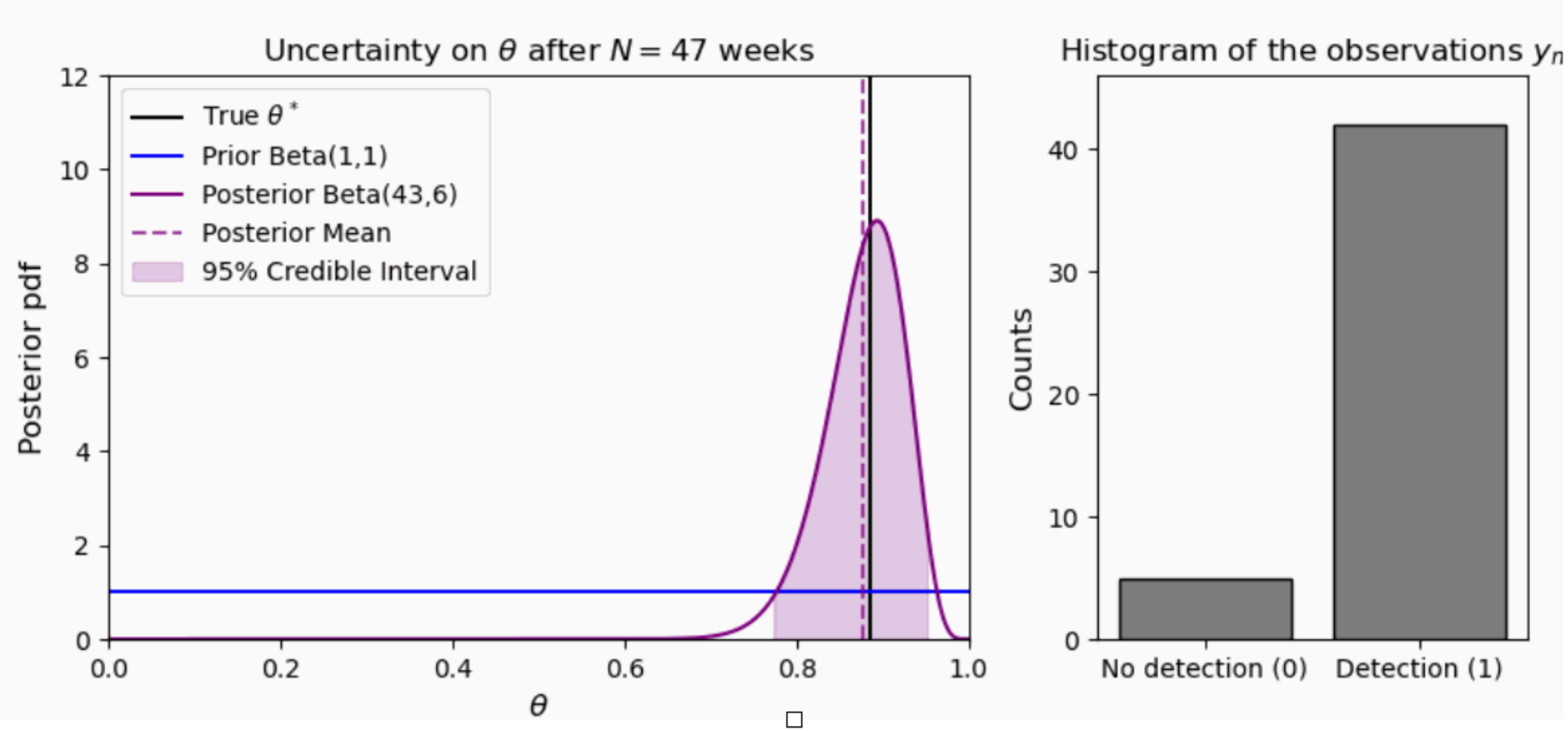
1/ For every value of N, the posterior can describe the uncertainty on θ

2/ As N increases, the variance decreases and the posterior mean converges to the true value

Example 1: probability of detecting at least one gravitational wave signal within the coming week

posterior $\theta \mid \{y_n\}_{n=1}^{N} \sim \text{Beta}\left(\alpha + \sum_{n=1}^{N} y_n, \beta + N - \sum_{n=1}^{N} y_n\right)$

A priori: uniform distribution on $[0,1]$ with $\alpha, \beta = 1$



Each new frame = "new week", $N \to N + 1$

$y_n = 0 \Longrightarrow$ small move to the left

$y_n = 1 \Longrightarrow$ small move to the right

Remarks:

1/ For every value of N, the posterior can describe the uncertainty on θ

2/ As N increases, the variance decreases and the posterior mean converges to the true value

Example 1: probability of detecting at least one gravitational wave signal within the coming week

Bayesian approach: for any $N \geq 0$

$$\text{posterior } \theta \mid \{y_n\}_{n=1}^{N} \sim \text{Beta}\left(\alpha + \sum_{n=1}^{N} y_n, \beta + N - \sum_{n=1}^{N} y_n\right)$$

$$\mathbb{E}[\theta \mid \{y_n\}] = \frac{\alpha + \sum_{n=1}^{N} y_n}{\alpha + \beta + N} = \frac{1}{N}\sum_{n=1}^{N} y_n \text{ if } \alpha = \beta = 0$$

$$\text{Var}[\theta \mid \{y_n\}] = \frac{\left(\alpha + \sum_{n=1}^{N} y_n\right)\left(\beta + N - \sum_{n=1}^{N} y_n\right)}{(\alpha + \beta + N)^2(\alpha + \beta + N + 1)}$$

Frequentist approach

Estimator (minimum variance unbiased estim., MLE)

$$\hat{\theta}_N = \frac{1}{N}\sum_{n=1}^{N} y_n$$

Asymptotic convergence:

$$\hat{\theta}_N \to \theta^* \quad (N \to +\infty)$$

Asymptotic variance of estimator: with Central Limit theorem

$$\hat{\theta}_N - \theta^* \sim \mathcal{N}\left(0, \frac{\sigma^2}{N}\right) \quad (N \to +\infty) \quad (\text{for some } \sigma > 0)$$

Example 2: detection rate of gravitational wave signals within a week

1/ Describe the observation model

Random event: “how many GW signals did we observe on week n?”

Associated random variable: integer $y \in \mathbb{N}$

Data: N observations, $y_n \in \mathbb{N}$

Hypothesis: assume the observations of the N weeks to be *independent* and *identically distributed*

Model: each y_n had the same detection rate $\theta > 0$ $\rightarrow$ **we want to estimate** θ

$\rightarrow$ *identically distributed*: Poisson distribution $y_n \,|\, \theta \sim \text{Poisson}(\theta)$

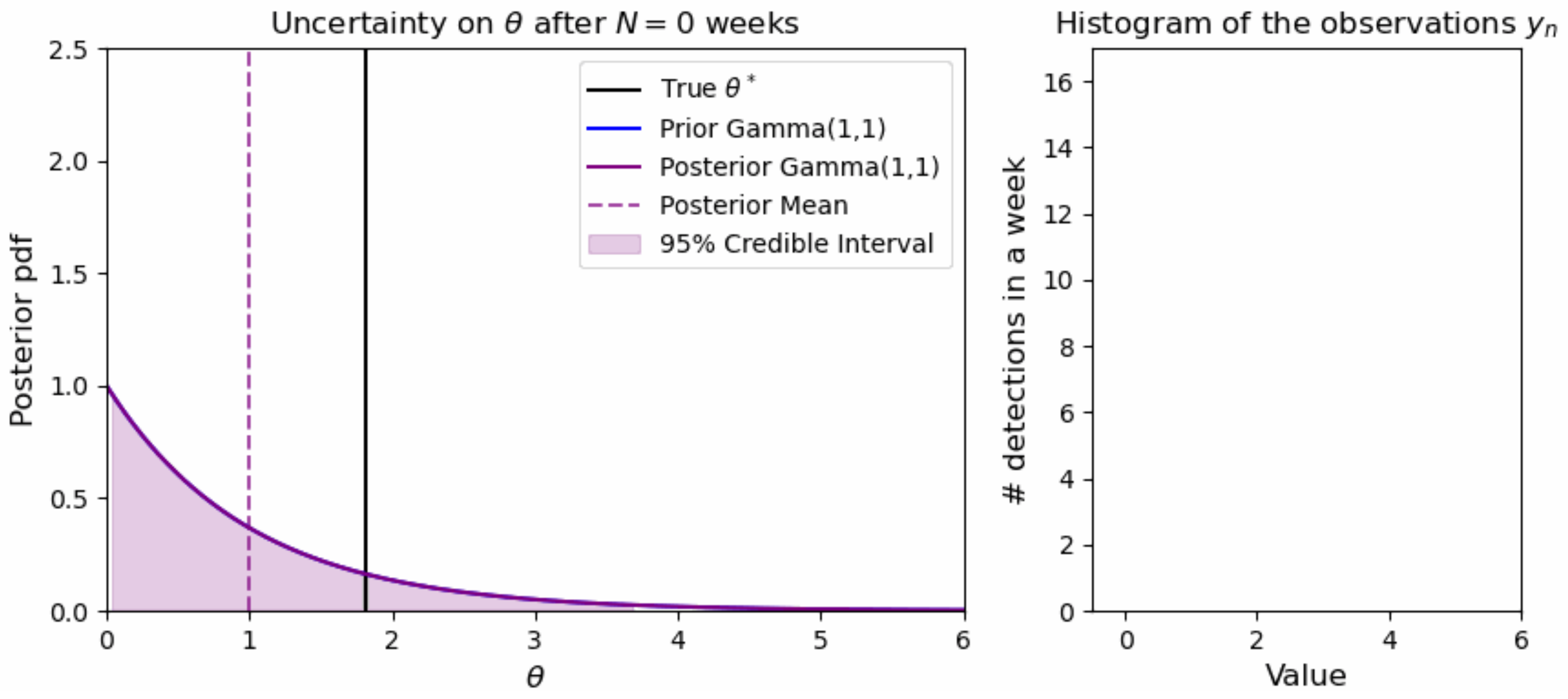
$\rightarrow$ *independent*: data distribution is $\pi\left(\{y_n\}_{n=1}^{N} \,|\, \theta\right) = \prod_{n=1}^{N} \pi\left(y_n \,|\, \theta\right)$

2/ Choose the prior: the conjugate prior to a Poisson likelihood is the **Gamma distribution** $\theta \sim \text{Gamma}(\alpha, \lambda)$

In this case, the posterior is: $\theta \,|\, \{y_n\}_{n=1}^{N} \sim \text{Gamma}\left(\alpha + \sum_{n=1}^{N} y_n, \beta + N\right)$

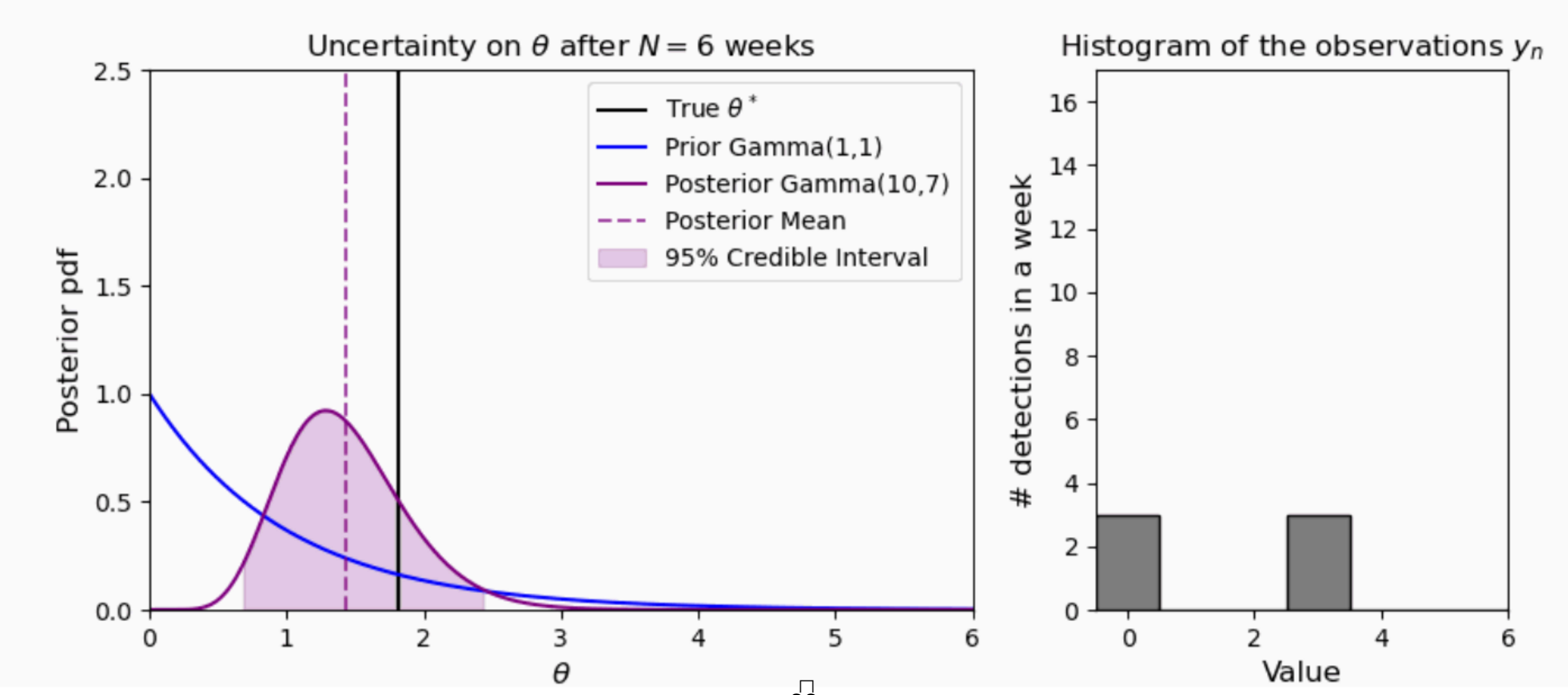
Example 2: detection rate of gravitational wave signals within a week

posterior $\theta \mid \{y_n\}_{n=1}^{N} \sim \text{Gamma}\left(\alpha + \sum_{n=1}^{N} y_n, \beta + N\right)$



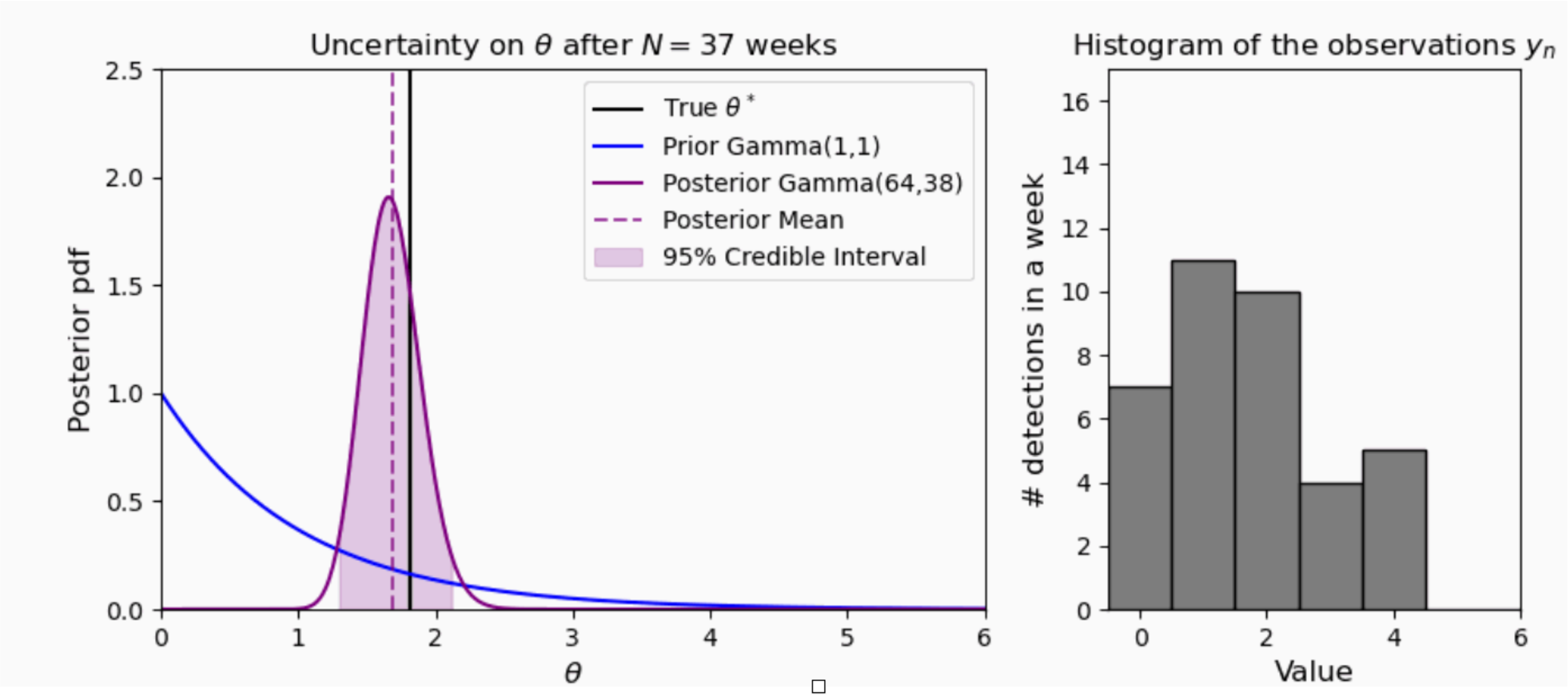
Example 2: detection rate of gravitational wave signals within a week

$$\text{posterior } \theta \mid \{y_n\}_{n=1}^{N} \sim \text{Gamma}\left(\alpha + \sum_{n=1}^{N} y_n, \beta + N\right)$$



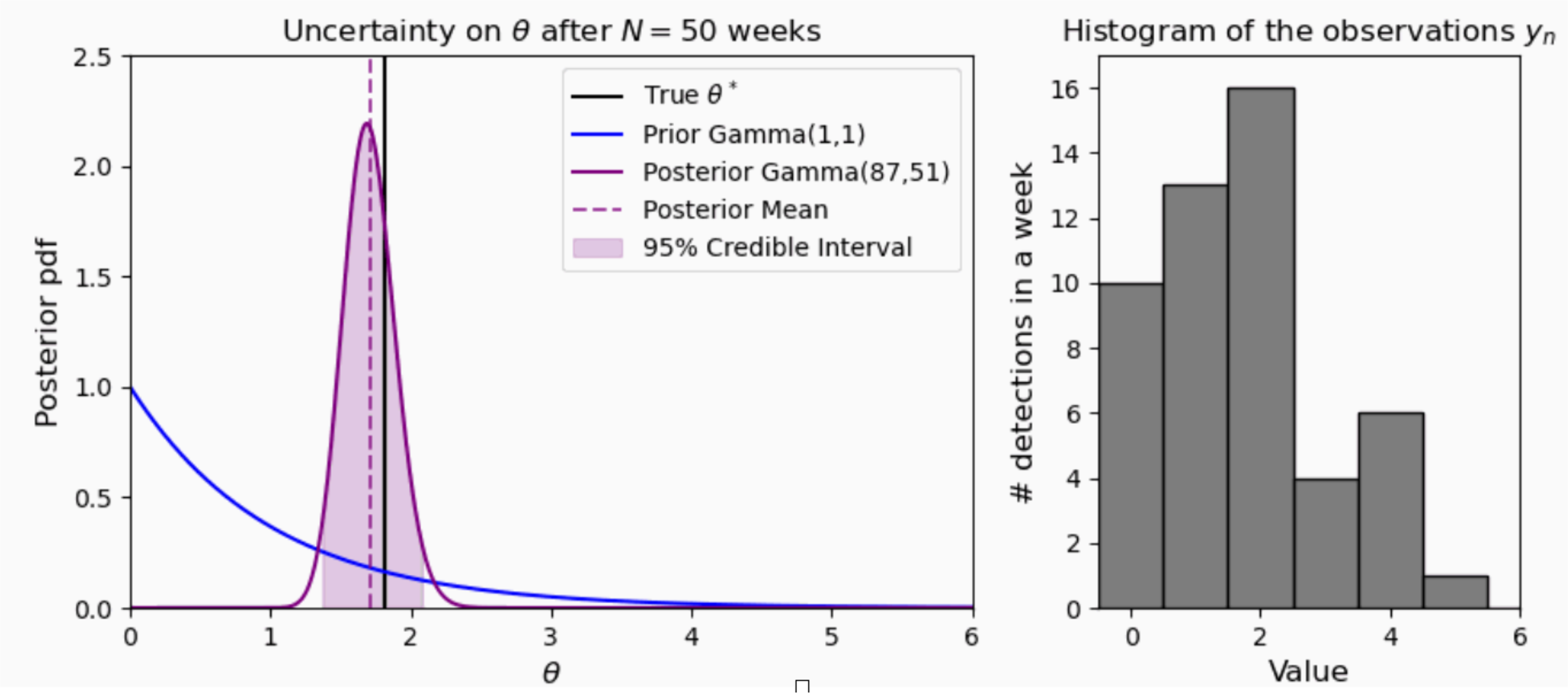
Example 2: detection rate of gravitational wave signals within a week

$$\text{posterior } \theta \mid \{y_n\}_{n=1}^{N} \sim \text{Gamma}\left(\alpha + \sum_{n=1}^{N} y_n, \beta + N\right)$$



Example 2: detection rate of gravitational wave signals within a week

$$\text{posterior } \theta \mid \{y_n\}_{n=1}^{N} \sim \text{Gamma}\left(\alpha + \sum_{n=1}^{N} y_n, \beta + N\right)$$



Summary parts 2 + Questions

Likelihood: when multiple observations of the same random variable, common to assume

- independent: the value of y_n does not influence the value of y_{n+1}
- identically distributed: all taken from the same distribution (eg, Poisson(θ))

Prior:

- to know wether there exist a conjugate priors in your case, check wikipedia's "conjugate prior" page
- As soon as you have non-linearity (eg, an astrophysical simulation M), there is no conjugate prior

$\Longrightarrow$ Need to evaluate estimators numerically.

Bayesian approach: you can derive estimators and describe uncertainties **no matter the amount of data**!

Even in case of degenerecies, even if less observations than unknowns, even with zero observations (from prior)

Part 3: Bayesian inference without conjugate priors: the non-ideal case

numerical evaluation of estimators with sampling algorithms

When there is no conjugate prior

Back to Bayes theorem: $\pi(\theta \,|\, y) \propto \pi(y \,|\, \theta)\ \pi(\theta)$ with no parametric description of the posterior. To derive estimators:

Option 1: abandon uncertainty quantification

Estimate the mode of the posterior, called maximum a posteriori (MAP), defined as

$$\hat{\theta}_{\text{MAP}}(y) = \arg\max_{\theta \in \Theta} \pi(\theta \,|\, y) = \arg\min_{\theta \in \Theta} \left[-\log \pi(\theta \,|\, y)\right]$$
$$= \arg\min_{\theta \in \Theta} \left[-\log \pi(y \,|\, \theta)\ -\log \pi(\theta)\right]$$

For instance, say:

$y \,|\, \theta \sim \mathcal{N}\left(M(\theta), \sigma^2 I_L\right)$ (Gaussian additive likelihood)

and $\pi(\theta) \propto \mathbf{1}_{\mathscr{C}}(\theta)$ (uniform prior on some set $\mathscr{C} \subset \Theta$)

$$\hat{\theta}_{\text{MAP}}(y) = \arg\min_{\theta \in \mathscr{C}} \frac{1}{2\sigma^2} \sum_{\ell=1}^{L} \left(y_\ell - M_\ell(\theta)\right)^2$$

Scales very well (standard approach in many areas)

but no uncertainty description

When there is no conjugate prior

Back to Bayes theorem: $\pi(\theta \mid y) \propto \pi(y \mid \theta)\ \pi(\theta)$ with no parametric description of the posterior. To derive estimators:

Option 1: abandon uncertainty quantification

Estimate the mode of the posterior, called maximum a posteriori (MAP), defined as

$$\hat{\theta}_{\text{MAP}}(y) = \arg\max_{\theta \in \Theta} \pi(\theta \mid y) = \arg\min_{\theta \in \Theta} \left[-\log \pi(\theta \mid y)\right]$$

$$= \arg\min_{\theta \in \Theta} \left[-\log \pi(y \mid \theta)\ -\log \pi(\theta)\right]$$

For instance, say:

$y \mid \theta \sim \mathcal{N}\left(M(\theta), \sigma^2 I_L\right)$ (Gaussian additive likelihood)

and $\pi(\theta) \propto \mathbf{1}_{\mathscr{C}}(\theta)$ (uniform prior on some set $\mathscr{C} \subset \Theta$)

$$\hat{\theta}_{\text{MAP}}(y) = \arg\min_{\theta \in \mathscr{C}} \frac{1}{2\sigma^2} \sum_{\ell=1}^{L} \left(y_\ell - M_\ell(\theta)\right)^2$$

Scales very well (standard approach in many areas) but no uncertainty description

Option 2: use an approximation

Approximate posterior with simple distribution (eg Gaussian) for which extracting estimators is simple

- ✔ Extracting parameters is easy
- ✘ The fit may be complex, especially if $\theta \in \mathbb{R}^D, D \gg 1$
- ✘ Assessing the validity of the uncertainty description may be challenging

Active area of research!

To know more on this type of approach:

see **variational methods**

When there is no conjugate prior

Back to Bayes theorem: $\pi(\theta \mid y) \propto \pi(y \mid \theta)\ \pi(\theta)$ with no parametric description of the posterior. To derive estimators:

Option 3: Compute integrals directly

$$\mathbb{E}\left[\theta \mid (y_n)_{n=1}^N\right] = \mu = \int \theta\ \pi\left(\theta \mid (y_n)_{n=1}^N\right) d\theta$$

$$\text{Var}\left[\theta \mid (y_n)_{n=1}^N\right] = \int (\theta - \mu)^2\ \pi\left(\theta \mid (y_n)_{n=1}^N\right) d\theta$$

Or $\mathbb{E}\left[f(\theta) \mid (y_n)_{n=1}^N\right] = \int f(\theta)\ \pi\left(\theta \mid (y_n)_{n=1}^N\right) d\theta$

❌ Requires to evaluate integrals for each quantity

❌ Unrealistic when $\theta \in \mathbb{R}^D, D \gg 1$

When there is no conjugate prior

Back to Bayes theorem: $\pi(\theta | y) \propto \pi(y | \theta)\ \pi(\theta)$ with no parametric description of the posterior. To derive estimators:

Option 3: Compute integrals directly

$$\mathbb{E}\left[\theta | (y_n)_{n=1}^N\right] = \mu = \int \theta\ \pi\left(\theta | (y_n)_{n=1}^N\right) d\theta$$

$$\text{Var}\left[\theta | (y_n)_{n=1}^N\right] = \int (\theta - \mu)^2\ \pi\left(\theta | (y_n)_{n=1}^N\right) d\theta$$

Or $\mathbb{E}\left[f(\theta) | (y_n)_{n=1}^N\right] = \int f(\theta)\ \pi\left(\theta | (y_n)_{n=1}^N\right) d\theta$

- ✗ Requires to evaluate integrals for each quantity
- ✗ Unrealistic when $\theta \in \mathbb{R}^D, D \gg 1$

Option 4: Monte Carlo estimators (cf risk in ML)

Generate T samples $\theta^{(t)} \sim \pi\left(\theta | (y_n)_{n=1}^N\right)$ and use Monte Carlo estimators:

$$\mathbb{E}\left[\theta | (y_n)_{n=1}^N\right] \simeq \mu_T = \frac{1}{T}\sum_{t=1}^{T} \theta^{(t)}$$

$$\text{Var}\left[\theta | (y_n)_{n=1}^N\right] \simeq \frac{1}{T-1}\sum_{t=1}^{N} \left(\theta^{(t)} - \mu_T\right)^2$$

Or $\mathbb{E}\left[f(\theta) | (y_n)_{n=1}^N\right] \simeq \frac{1}{T} f\left(\theta^{(t)}\right)$

- ✓ Guarantee to converge as $T \to \infty$
- ✓ Verifies the Central Limit theorem
- ✗ Requires numerous evaluations of likelihood

How to sample from the posterior distribution?

There are algorithms

How to sample from the posterior distribution?

Nested sampling

Rejection sampling

EMCEE and
parallel tempering-EMCEE

(Preconditioned) Metropolis-Adjusted Langevin Algorithm

There are many algorithms!

Gibbs sampling

Multiple-try Metropolis (MTM)

Normalising flows (NF)

Random Walk Metropolis-Hastings

Hamiltonian Monte Carlo (HMC)
And No-U-Turn sampler (NUTS)

Sequential Monte Carlo (SMC)

How to sample from the posterior distribution?

Nested sampling

Rejection sampling

EMCEE and
parallel tempering-EMCEE

(Preconditioned) Metropolis-Adjusted Langevin Algorithm

But one category is most fundamental

Markov chain Monte Carlo (MCMC)

Gibbs sampling

Multiple-try Metropolis (MTM)

Normalising flows (NF)

Random Walk Metropolis-Hastings

Hamiltonian Monte Carlo (HMC)
And No-U-Turn sampler (NUTS)

Sequential Monte Carlo (SMC)

A fundamental MCMC algorithm: Metropolis-Hastings

The idea

Generating independent samples from the posterior is generally not feasible.

Instead, we resort to an **iterative algorithm** that yields a **sequence of correlated samples**

Remark on the name “MCMC”:
Markov chain = the sequence
Monte Carlo = how the sequence is used

A fundamental MCMC algorithm: Metropolis-Hastings

The idea

Generating independent samples from the posterior is generally not feasible.

Instead, we resort to an **iterative algorithm** that yields a **sequence of correlated samples**

Remark on the name “MCMC”:
Markov chain = the sequence
Monte Carlo = how the sequence is used

The algorithm

$\pi(\theta \mid y) \propto \pi(y \mid \theta)\, \pi(\theta)$

At iteration t+1:

1) Generate a candidate from a distribution q (generally a Gaussian):

$$\theta^{(c)} \sim q(\theta \mid \theta^{(t)})$$

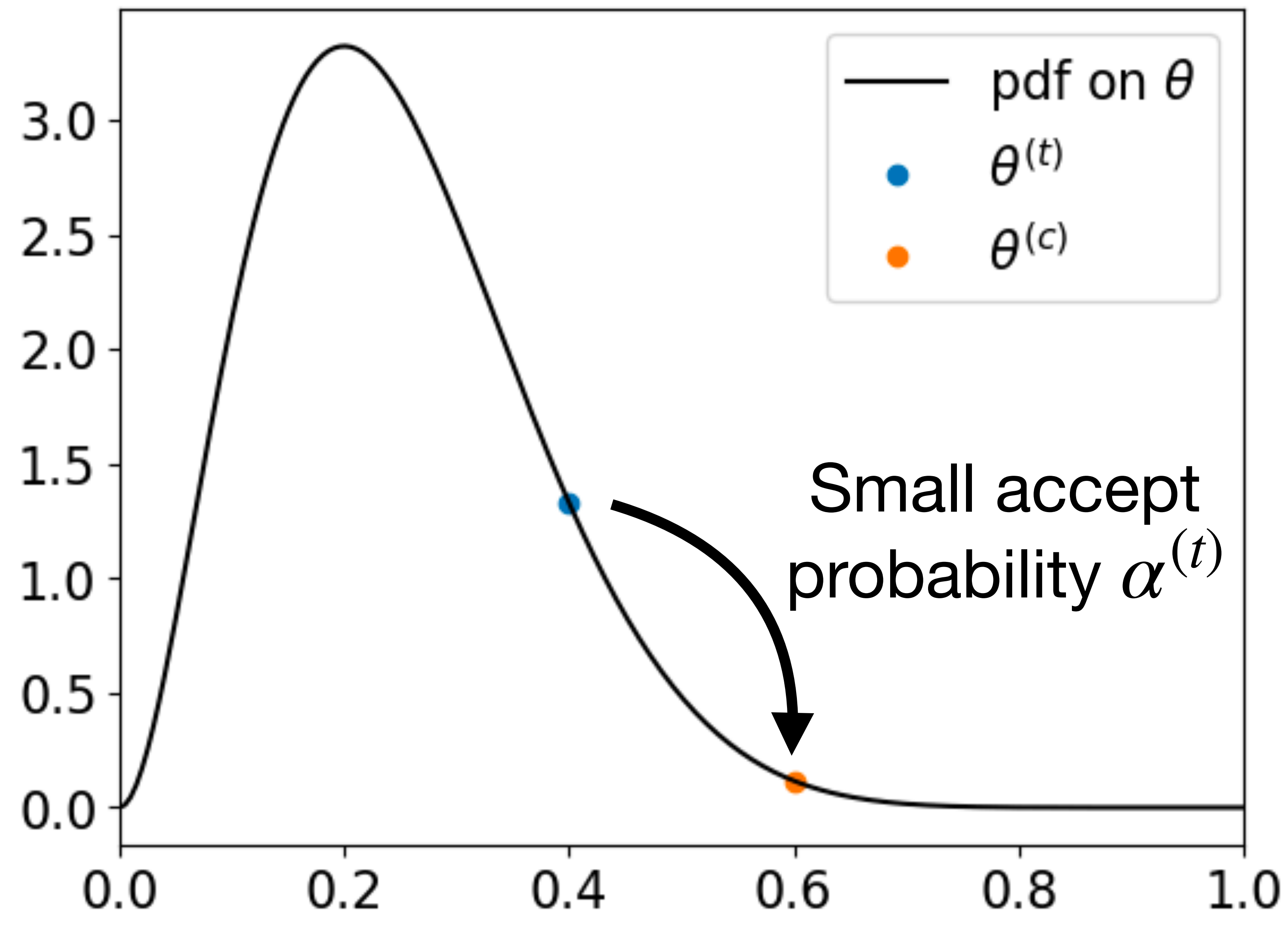
2) Compute an accept probability $\alpha^{(t)}$

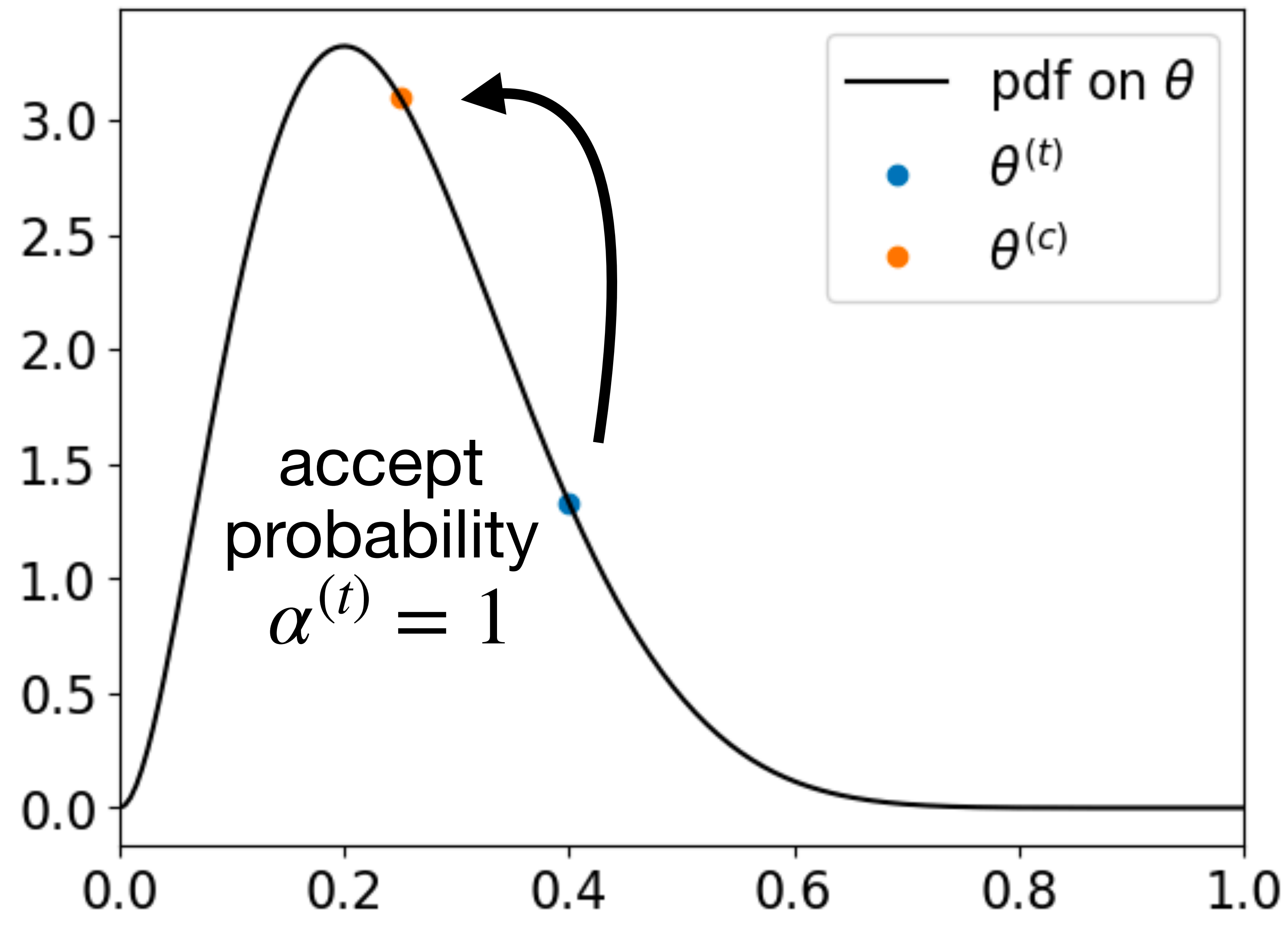
$$\alpha^{(t)} = \min \left\{ 1, \frac{\pi\left(\theta^{(c)} \mid (y_n)_{n=1}^N\right)}{\pi\left(\theta^{(t)} \mid (y_n)_{n=1}^N\right)} \frac{q(\theta^{(t)} \mid \theta^{(c)})}{q(\theta^{(c)} \mid \theta^{(t)})} \right\}$$

Note: q often set so that $q(\theta^{(c)} \mid \theta^{(t)}) = q(\theta^{(t)} \mid \theta^{(c)})$

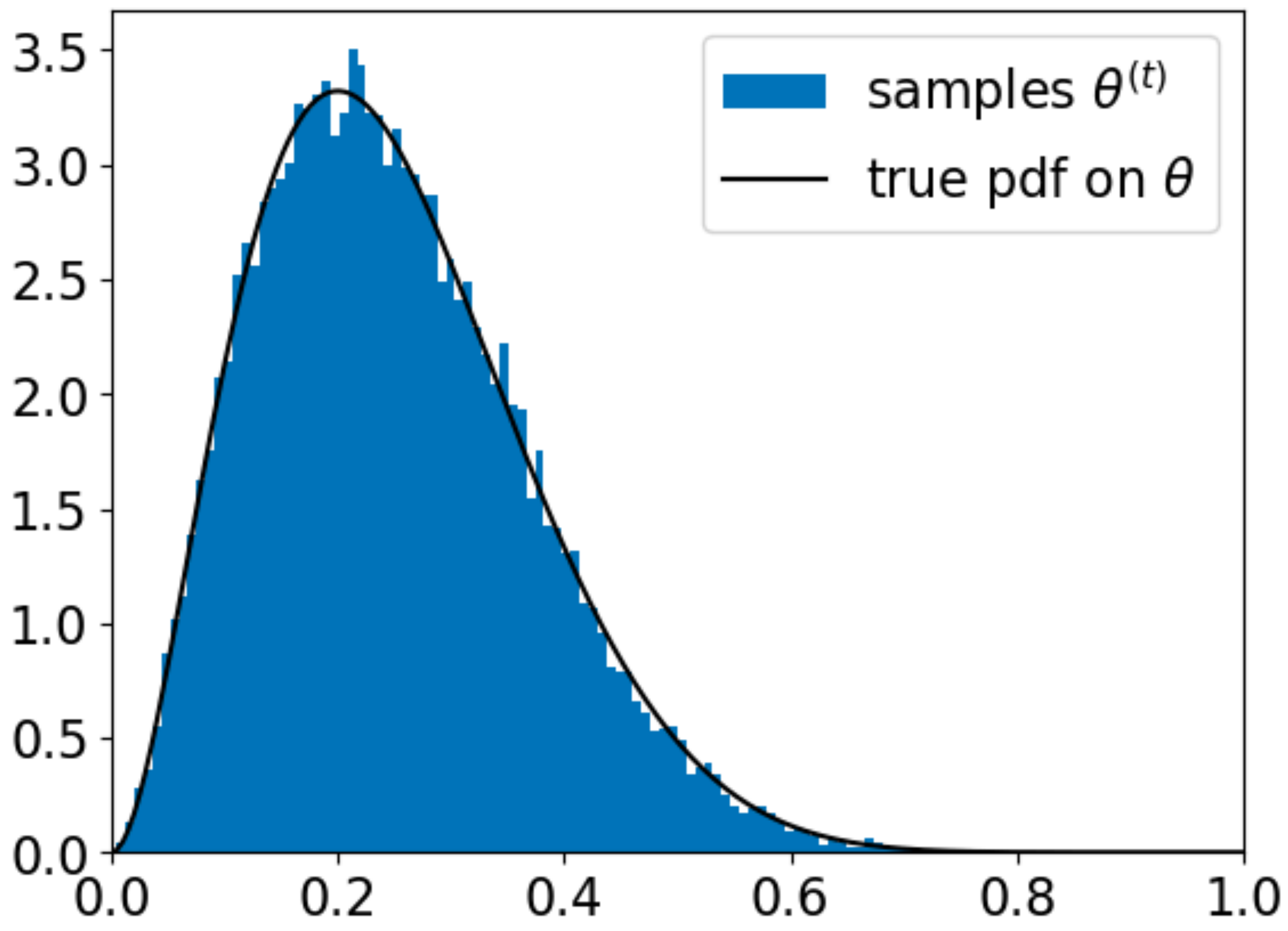
3) $\theta^{(t+1)} = \theta^{(c)}$ with proba $\alpha^{(t)}$ and $\theta^{(t)}$ with proba $1 - \alpha^{(t)}$

Illustration: Metropolis-Hastings on the Beta(3,9) distribution





Histogram of $T = 20{,}000$ samples $\theta^{(t)}$



Monte Carlo estimators are evaluated from these samples

Eg, $\mathbb{E}\left[\theta \,|\, (y_n)_{n=1}^{N}\right] = \frac{1}{T}\sum_{t=1}^{T} \theta^{(t)}$

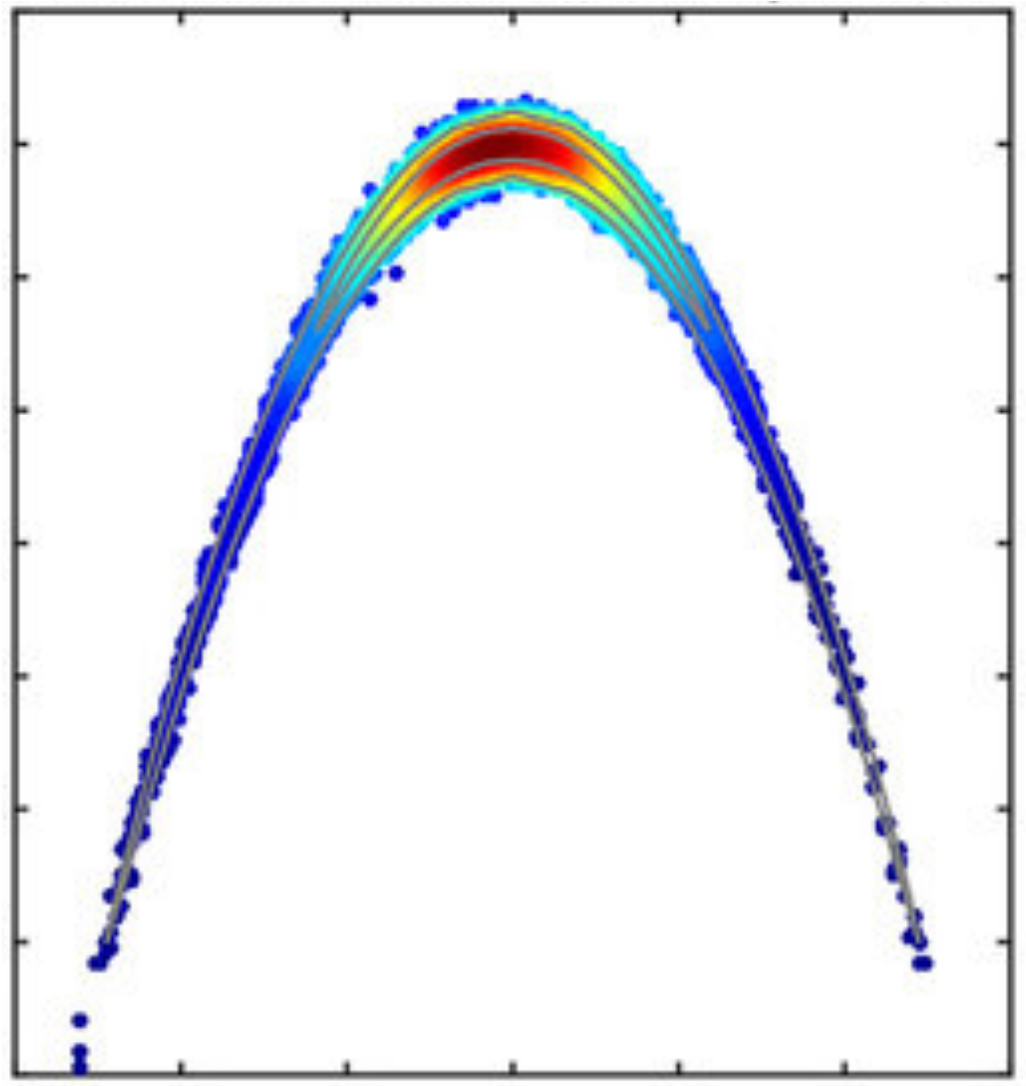
Similarly for variance, credibility intervals, specific probabilities, etc.

When to use Metropolis-Hastings (and not to)

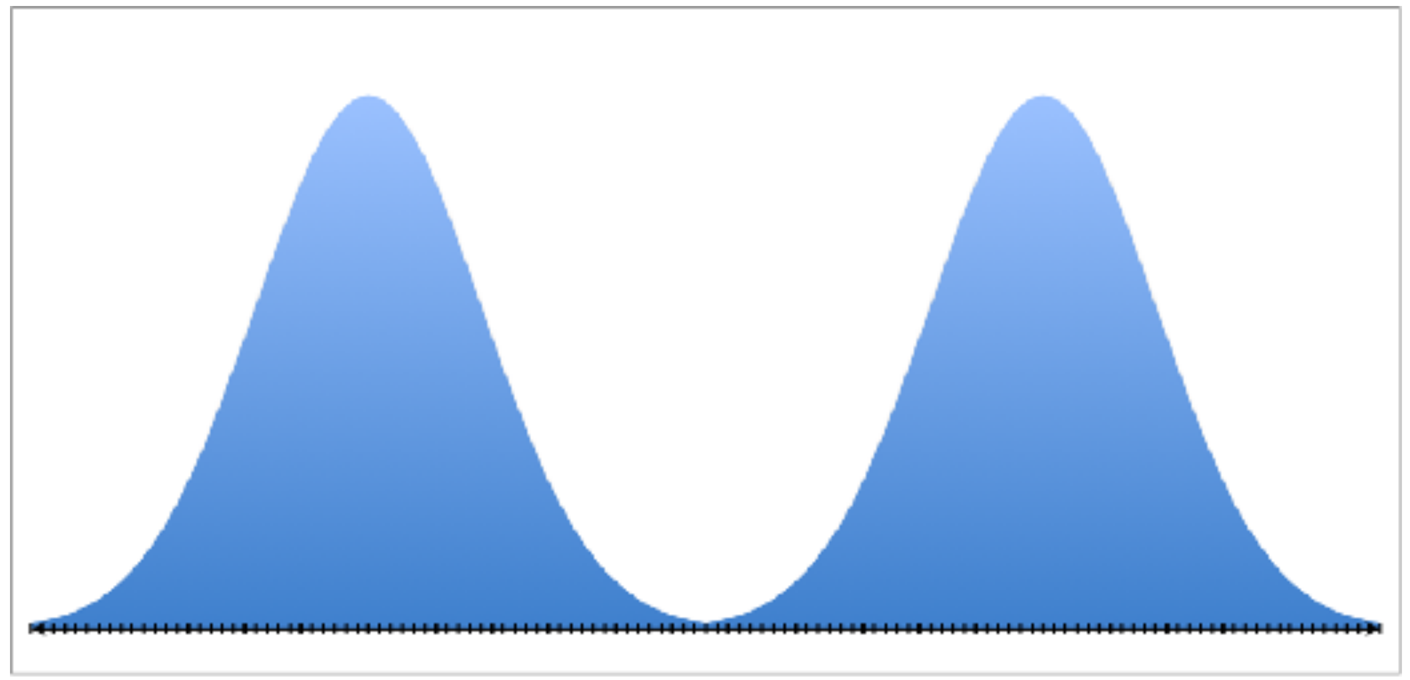
✗ If degeneracy between elements of θ in posterior

✗ If θ in high dimensions (≥ 10)

✓ If θ in low dimensions, posterior unimodal, no large degeneracy



✗ If local modes



Like standard gradient descent is an entry point to optimization methods, MH is an entry point to sampling methods.

Summary of part 3

When no conjugate prior : 3 main types of approaches to evaluate estimators from the posterior

1) Approximate the posterior with simple distribution

2) Evaluate integrals directly

3) Sample from the posterior and use Monte-Carlo estimators

MCMC algo. such as Metropolis-Hastings: generate a candidate, then accept or reject it with a certain probability

- Random walk Metropolis-Hastings: $\theta^{(c)} \sim \mathcal{N}(\theta^{(t)}, \Sigma)$
- Other MCMC algorithms use different candidate distributions (some with gradient information)

For simples cases (unimodal, low dimension & no large degeneracy in posterior), Metropolis-Hastings should work

For more complex cases, check other MCMC algorithms (EMCEE, HMC, Gibbs) and software (HerBIE, Beetroots)

Sampling algorithms require many evaluations of the likelihood (which often involves an astrophysical simulator)

→ an astrophysical simulator needs to be very fast, or one can resort to an emulator

Part 4: connecting the dots

real applications of Bayesian inference in ISM studies

Already many applications

	Topic	$\theta \in \mathbb{R}^D$	y	M	$\pi(y\|\theta)$ Noise	$\pi(\theta)$	Approach	Algorithm
Panter et al 2003	Star formation	25	galaxy spectra from SDSS	?	Gaussian	Uniform	MCMC	RWMH
Acquaviva et al 2011 (GalMC)	Star formation	5	Galaxy SEDs	GALAXEV	Gaussian	Uniform	MCMC	RWMH
Bailer-Jones et al 2011	Star properties	2	Photometry	ILIUM	Gaussian	Hertzsprung–Russell Diagram prior	Integration	
Perez-Montero 2014 (Hii-Chi-Mistry)	Hii regions	3	Emission lines	Popstar+Cloudy	Gaussian	Uniform	Integration	
Blanc 2015 (IZI)		2	Emission lines		Gaussian + mult.	Uniform	Integration	
Chevallard 2016 (BEAGLE)	Hii regions	7	Galaxy SEDs	“Simple model”	Gaussian	Uniform	Nested sampling	MultiNest
Johannesson 2016	Cosmic rays	30		GALPROP	Gaussian + mult.	Uniform	Nested sampling	MultiNest
…								

There’s plenty more: CIGALEMC, BOND, **HerBIE**, NebulaBayes, **MULTIGRIS**, **UCLCHEMCMC**

With applications to star formation history, Hii regions, molecular clouds, PDRs, galactic & extragalactic

Involving a variety of astrophysical simulators (often emulated) such as Cloudy, RADEX, Meudon PDR code

For a review, see my PhD manuscript, chapter 3

Case 1: Estimating distances from parallaxes, from Bailer-Jones (2015)

Observation y: angle in mas (eg, from Hipparcos or Gaia missions)
Physical parameter $\theta \in \mathbb{R}$: distance to star (in pc)

Likelihood function:

Astrophysical simulator: $M(\theta) = 1/\theta$

additive Gaussian noise with known std: $$y = M(\theta) + \varepsilon = \frac{1}{\theta} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Prior distribution: how does the density of stars evolve with distance to the sun?

1/ Uniform on validity interval $\pi(\theta) \propto \begin{cases} 1 \text{ if } \theta \in [0, \theta_{\max}] \\ 0 \text{ otherwise} \end{cases}$

2/ Constant star volume density $\pi(\theta) = \begin{cases} \frac{3}{\theta_{\max}^3}\theta^2 \text{ if } 0 \leq \theta \leq \theta_{\max} \\ 0 \text{ otherwise} \end{cases}$

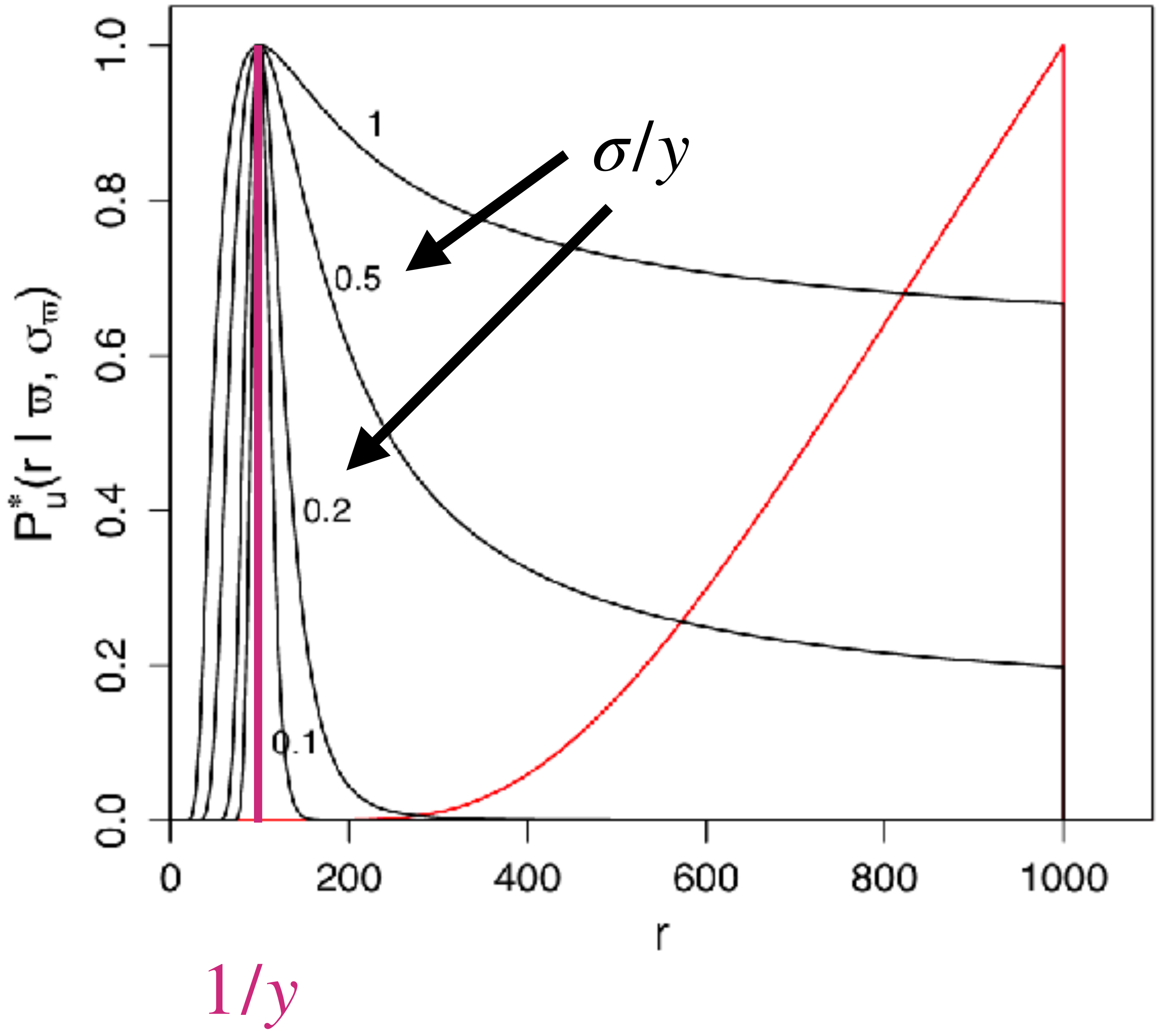
3/ Exponentially decreasing star volume density $\pi(\theta) = \begin{cases} \frac{3}{\theta_{\max}^3}\theta^2 e^{-r/L} \text{ if } 0 \leq \theta \leq \theta_{\max} \\ 0 \text{ otherwise} \end{cases} \quad (L \geq 0)$

Case 1: Estimating distances from parallaxes, from Bailer-Jones (2015)

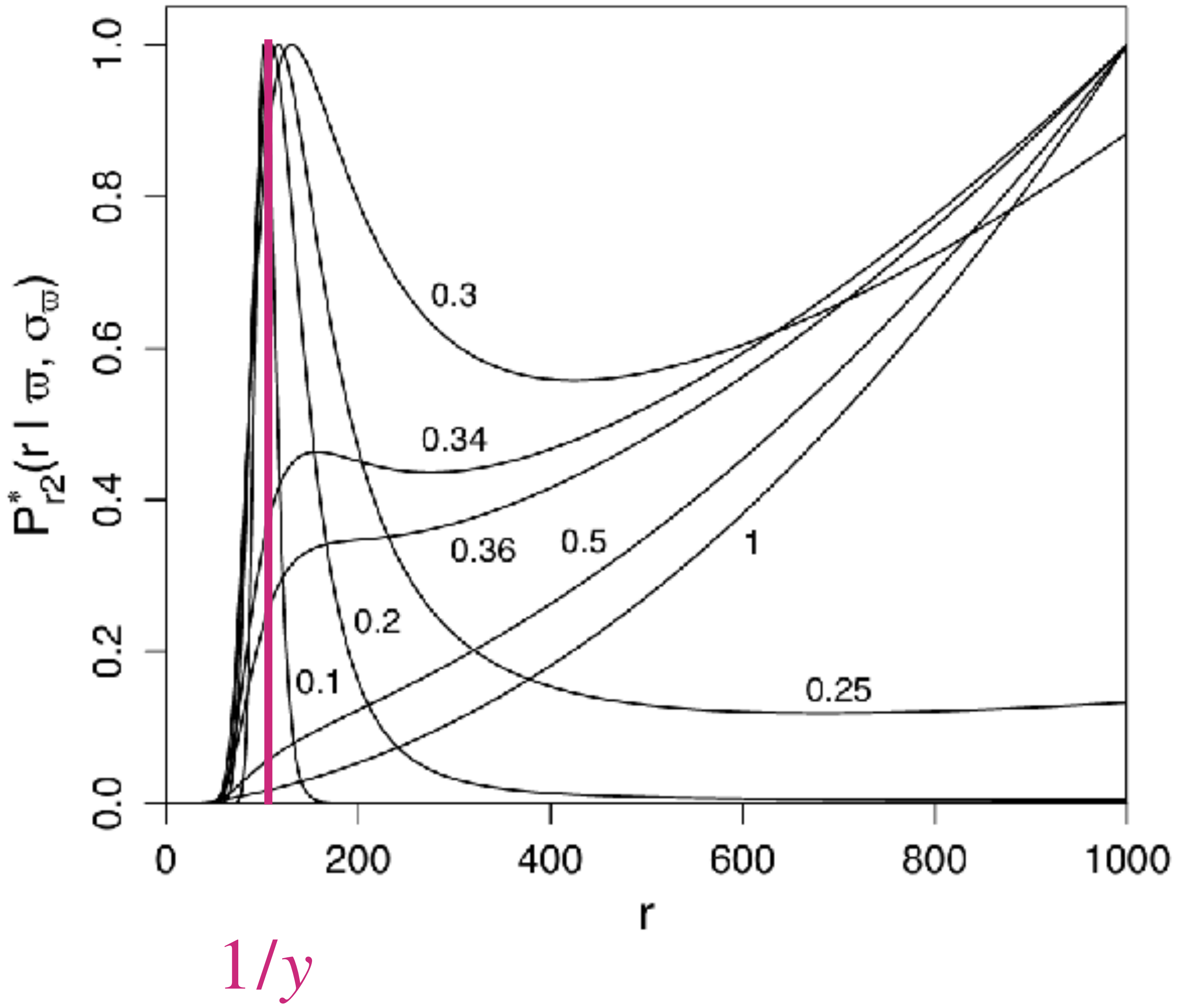
One-dimensional simple inference problem: everything can be computed with integrals

For $y = 1/100$ mas

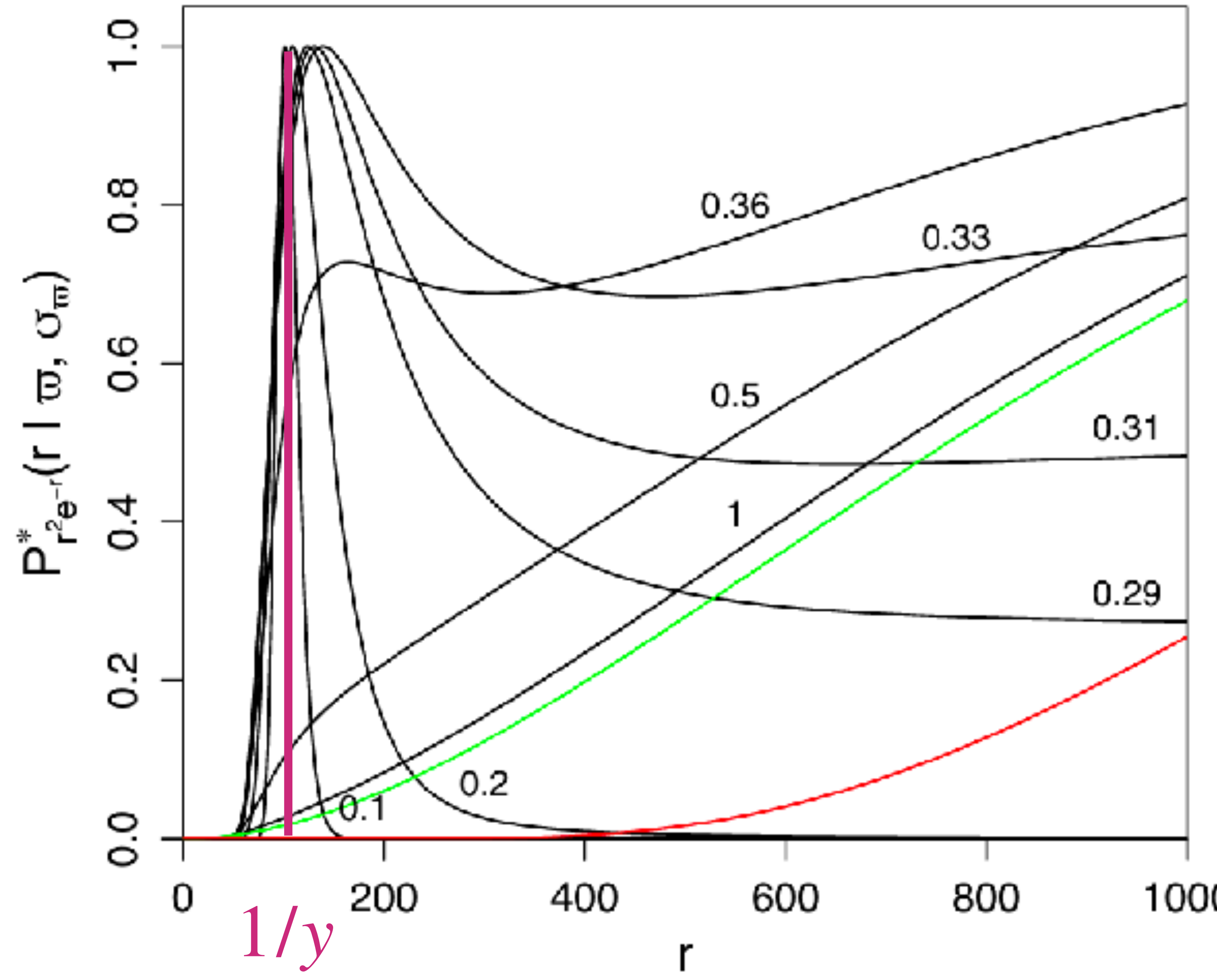
Prior: uniform on validity interval



Constant star volume density



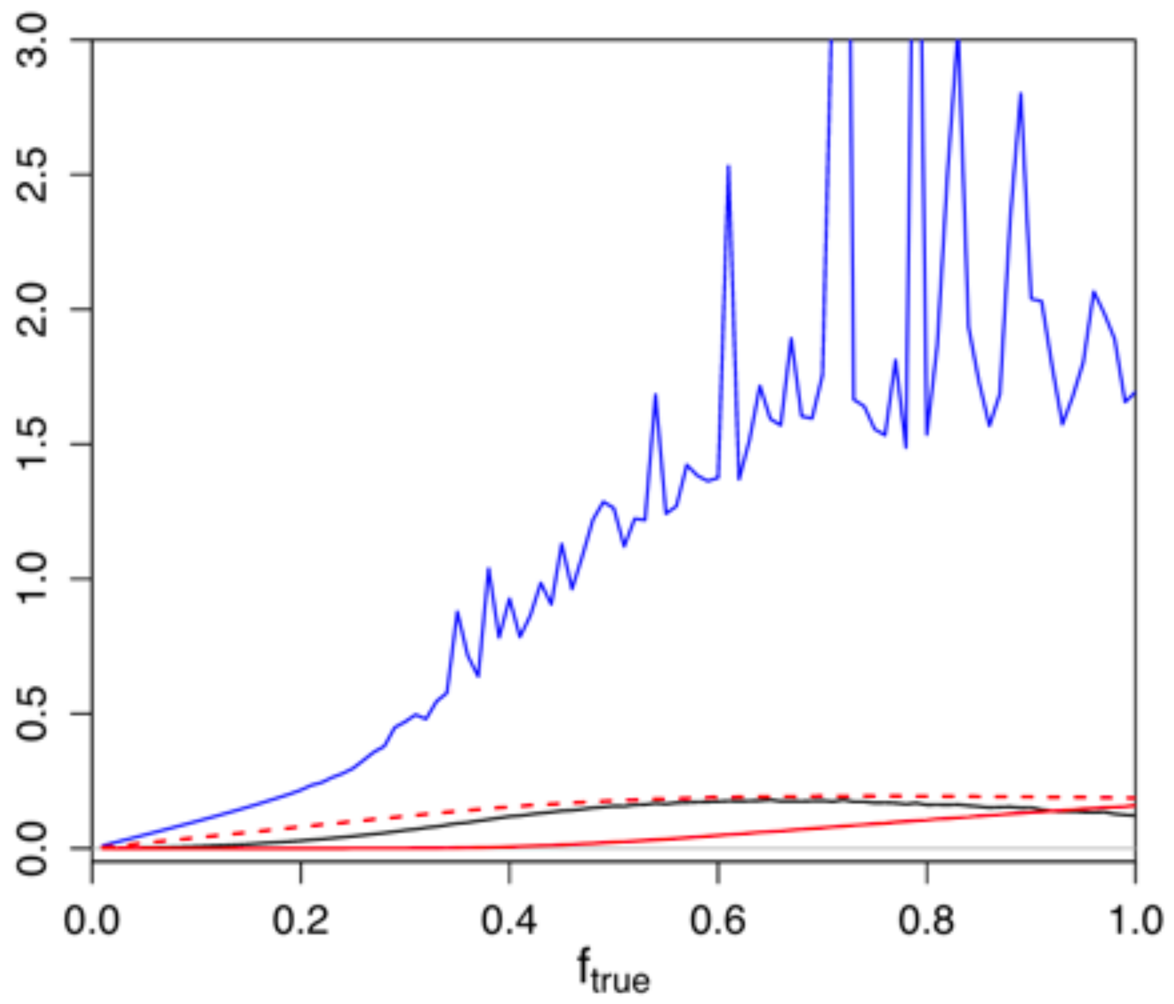
Exponentially decreasing star volume density ($L = 1000$)



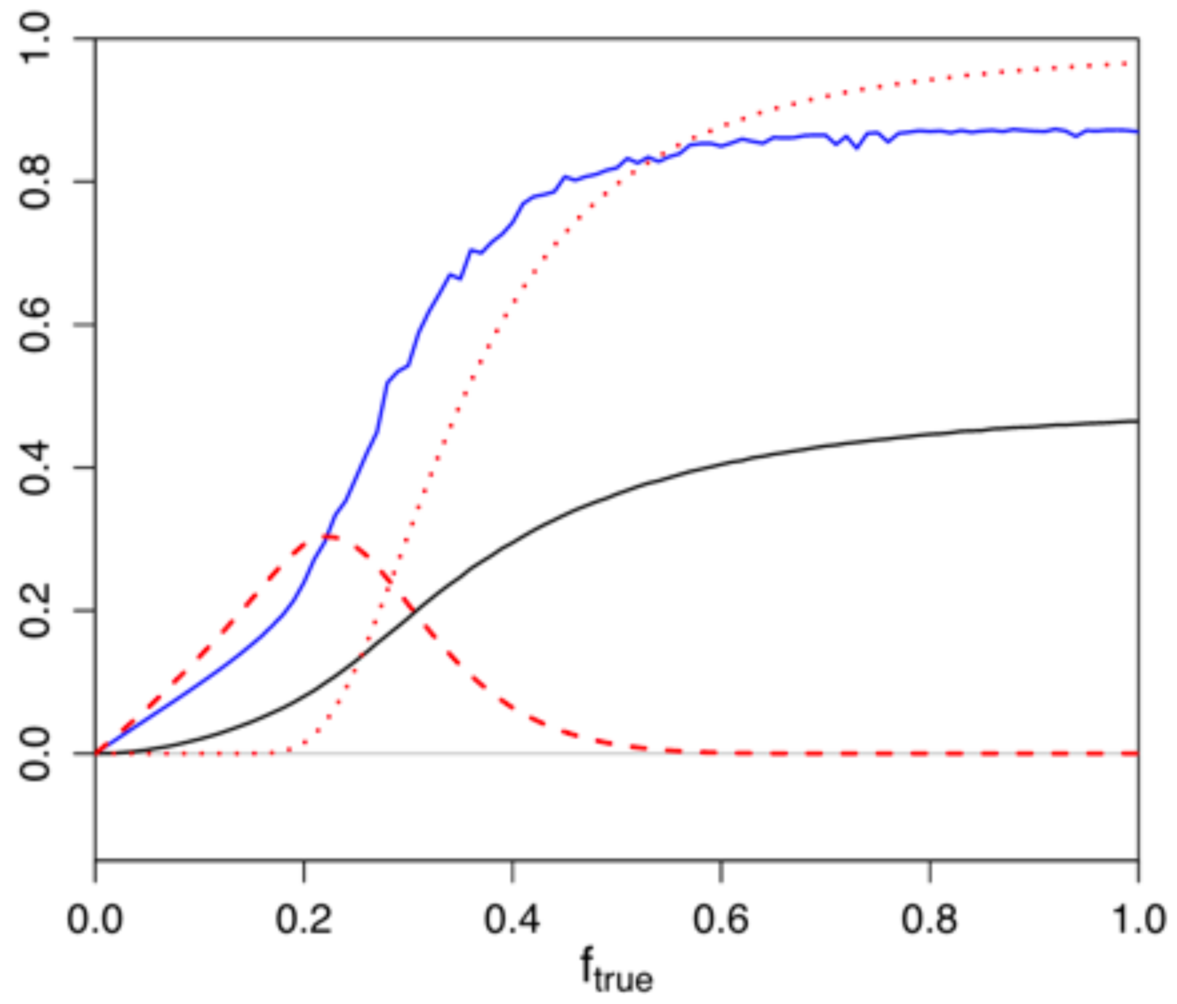
Green = prior, Red: $y = -1/100$

Case 1: Estimating distances from parallaxes, from Bailer-Jones (2015)

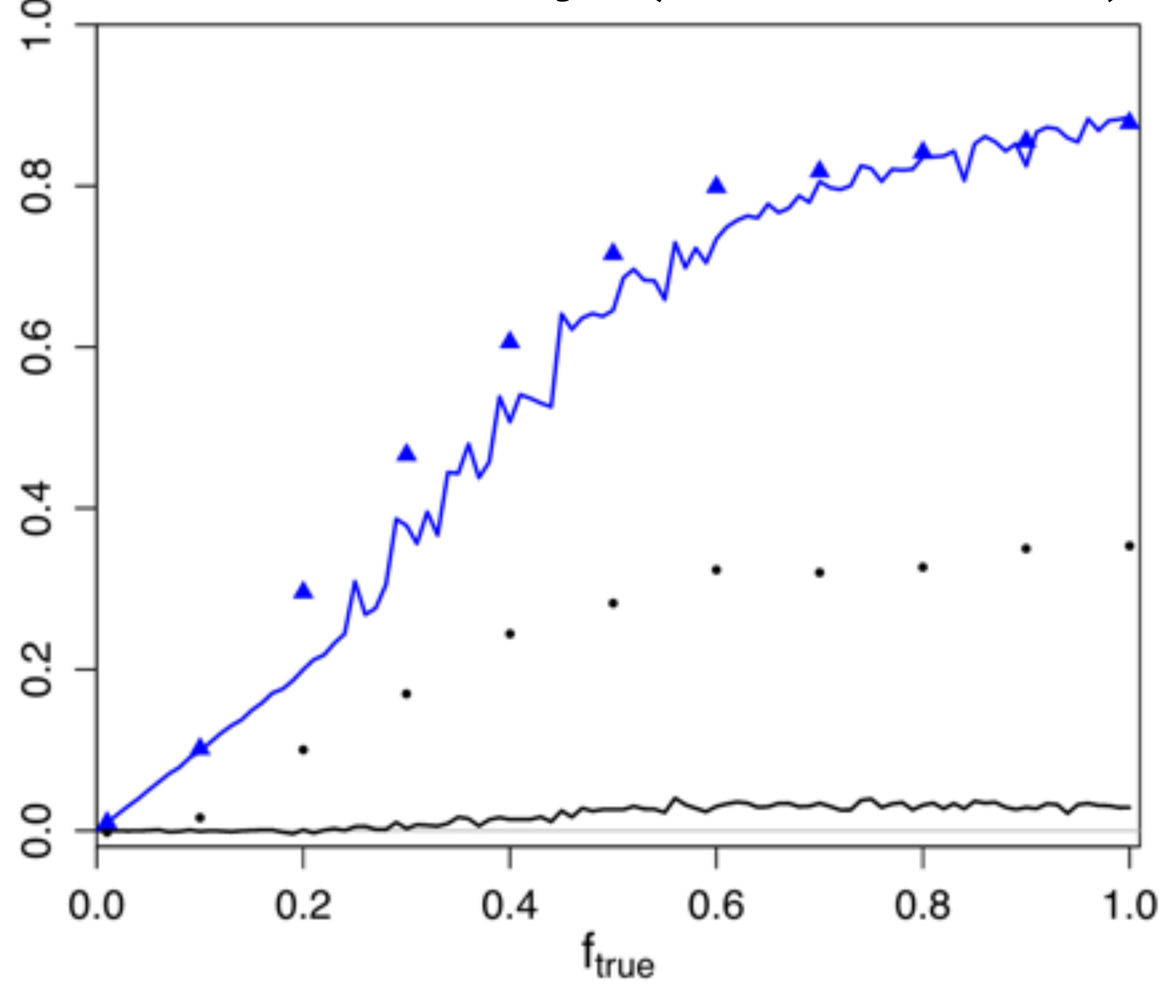
uniform on validity interval



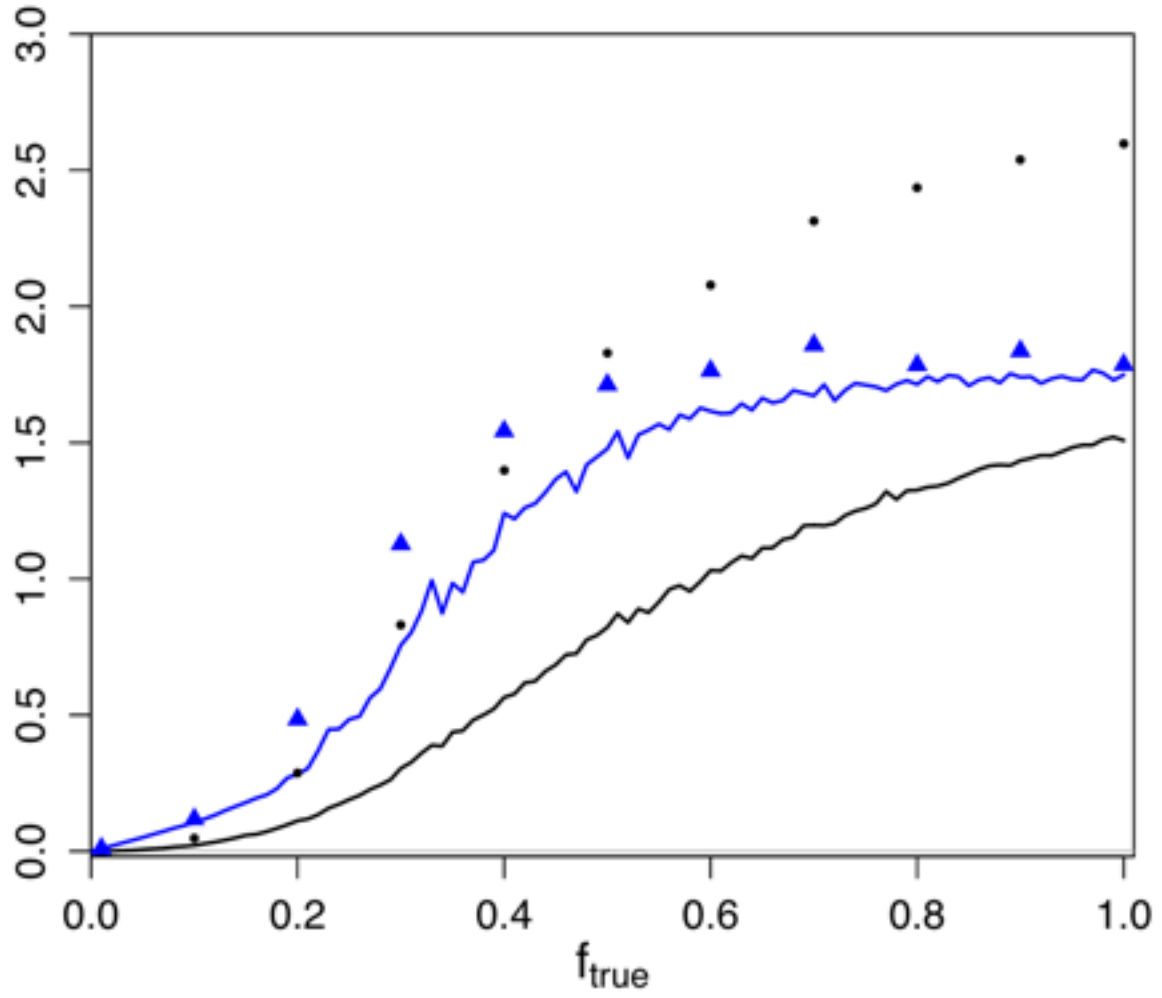
Constant star volume density



Exponentially decreasing star volume density $(L = 1000)$



Data drawn from same prior



Data drawn from constant star volume density prior

For mode estimator on multiple cases:

Black = bias $\mathbb{E}_{\theta^*,y}\left[\hat{\theta}_{\text{MAP}}(y) - \theta^*\right]$

Blue = variance $\mathbb{E}_{\theta^*,y}\left[\left(\hat{\theta}_{\text{MAP}}(y) - \theta^*\right)^2\right]$

Case 2: Analysis of prestellar core L1455, from Keil et al (2022)

θ = 4 parameters: volume density, Temperature, CR ionisation rate, radius of the assumed spherical cloud R_{out}

Observations $y \in \mathbb{R}^L : L = 12$ molecular emission lines (single pixel)

Likelihood:

Astrophysical simulator $M(\theta)$ = RADEX + UCLCHEM. → fast enough to be used directly in inference process

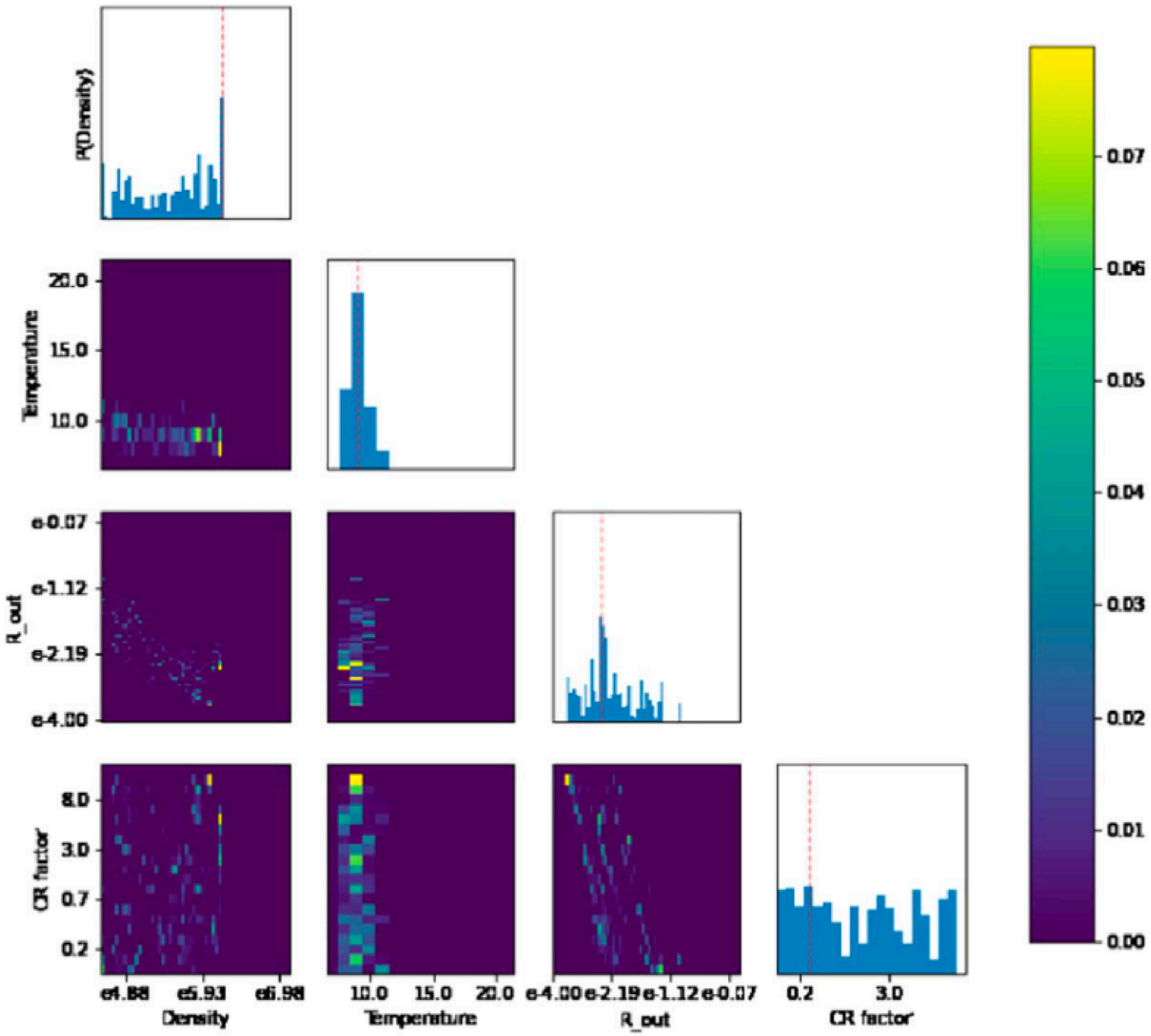
Observation model: $\forall \ell, \quad y_\ell = M_\ell(\theta) + \varepsilon_\ell, \quad \varepsilon_\ell \sim \mathcal{N}(0, \sigma_\ell^2)$

Prior: log-uniform on validity intervals

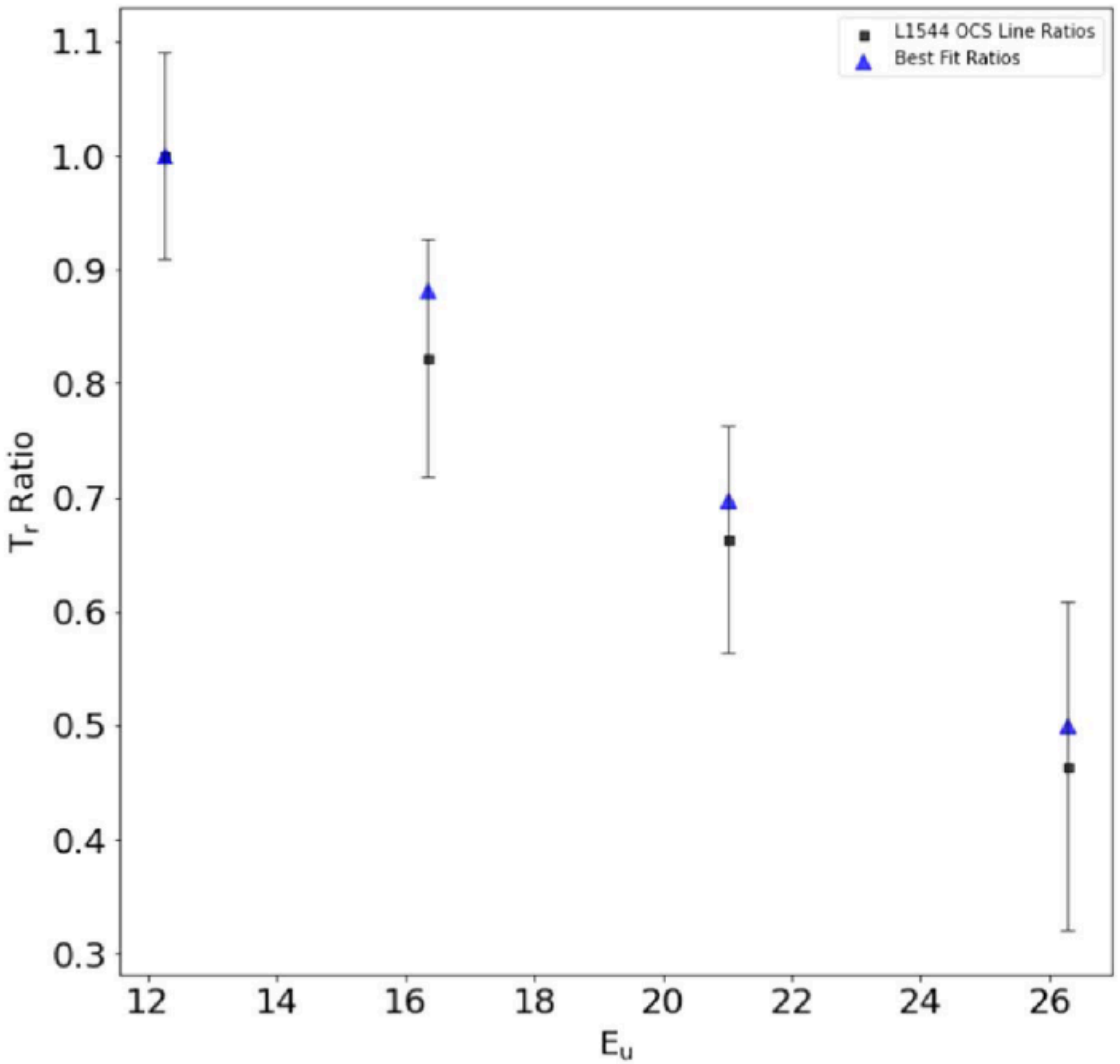
Sampling algorithm: EMCEE

Case 2: Analysis of prestellar core L1455, from Keil et al (2022)

Posterior samples



Model check: do the reconstruction reproduce the observations?

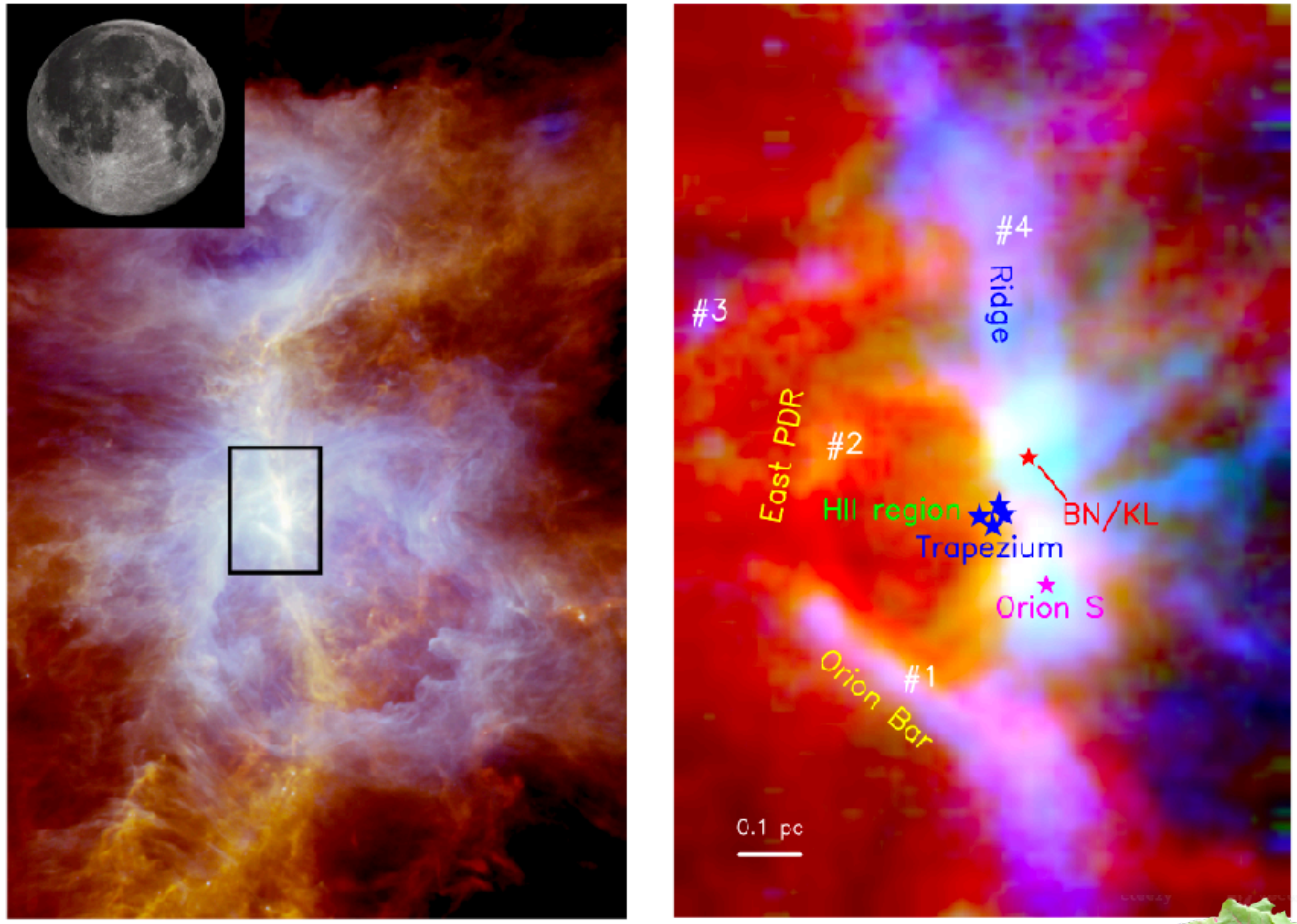


Case 3: Analysing maps of physical parameters with Beetroots: application to OMC-1, from Palud et al (2025)

Observation $y \in \mathbb{R}^{N \times L}$: multispectral image
($N \simeq 2400$, $L = 5$)

Physical parameter maps $\theta \in \mathbb{R}^{N \times D}$:

- Scaling factor κ (includes beam dilution factor)
- Thermal pressure P_{th}
- Intensity of UV radiative field G_0
- Visual extinction A_V^{tot}



Likelihood:
Astrophysical model: $M(\theta)$ = emulator of the Meudon PDR Code (built with a neural network)

Observation model: $\forall n, \ell, \quad y_{n\ell} = \varepsilon_{n\ell}^{(m)} M_\ell(\theta_n) + \varepsilon_{n\ell}^{(a)}$

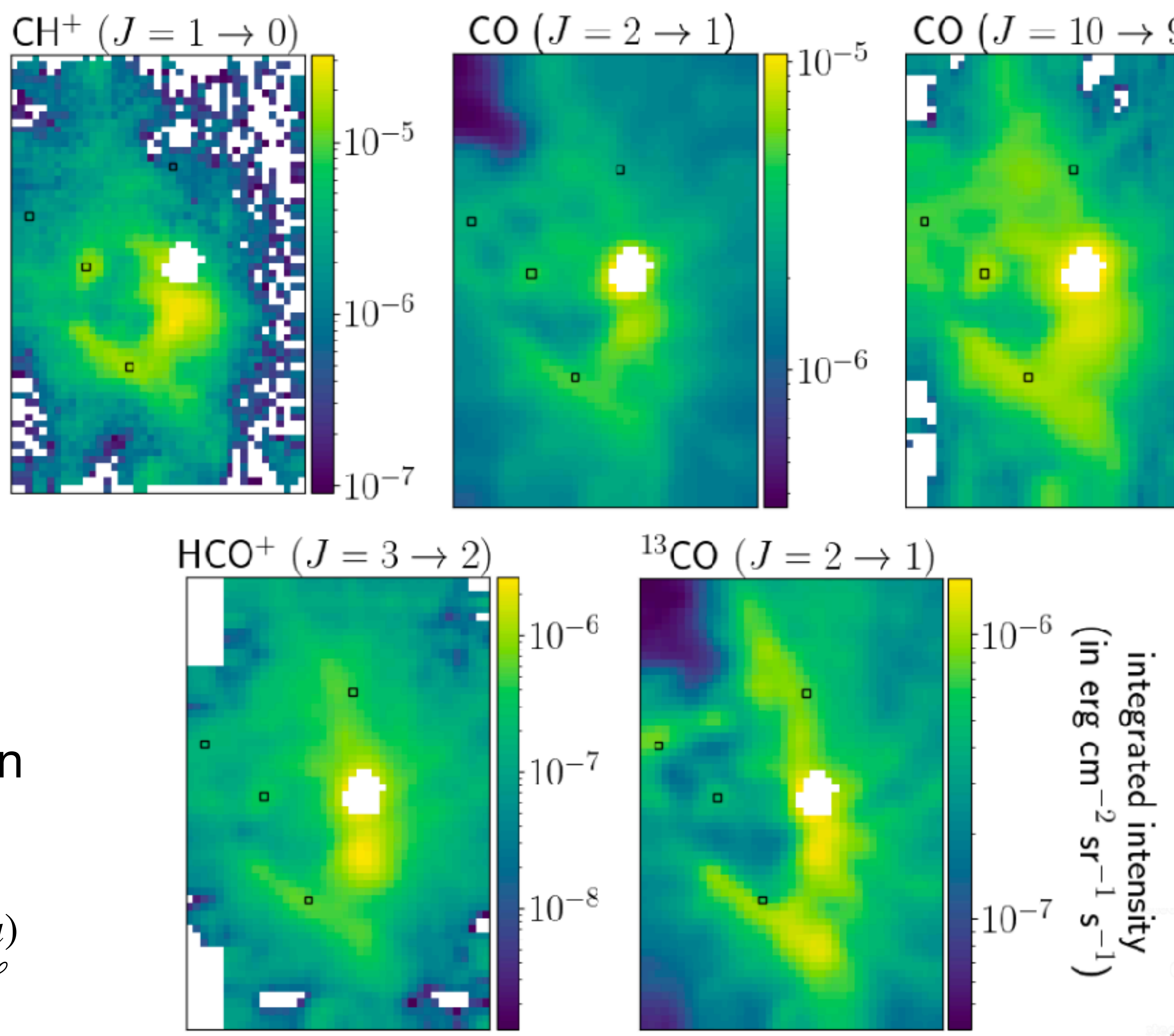
Prior:
validity intervals + spatial regularization

Case 3: Analysing maps of physical parameters with Beetroots: application to OMC-1, from Palud et al (2025)

Observation $y \in \mathbb{R}^{N \times L}$: multispectral image ($N \simeq 2400$, $L = 5$)

Physical parameter maps $\theta \in \mathbb{R}^{N \times D}$:

- Scaling factor κ (includes beam dilution factor)
- Thermal pressure P_{th}
- Intensity of UV radiative field G_0
- Visual extinction A_V^{tot}



Likelihood:

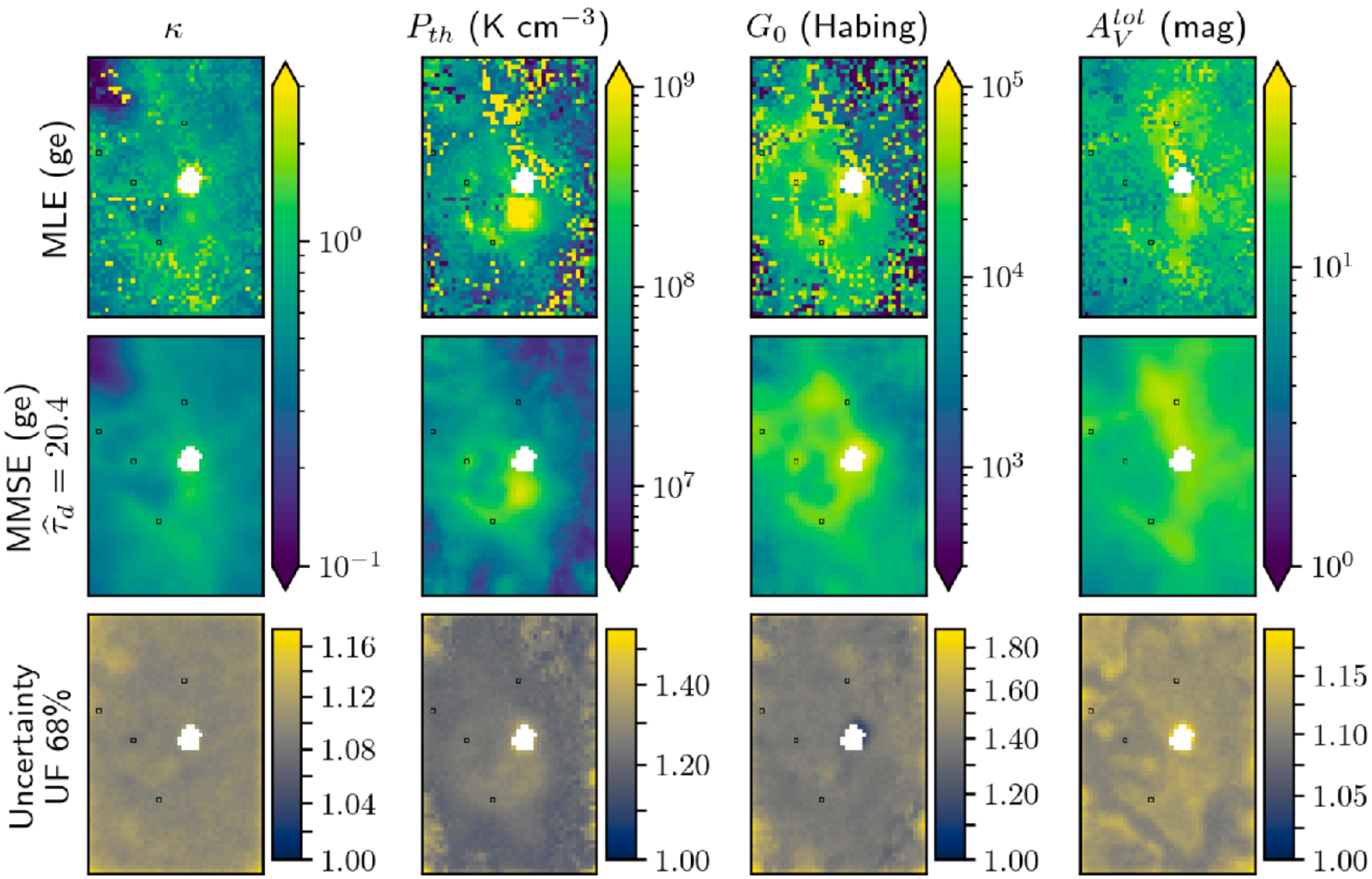
Astrophysical model: $M(\theta) =$ emulator of the Meudon PDR Code (built with a neural network)

Observation model: $\forall n, \ell, \quad y_{n\ell} = \varepsilon_{n\ell}^{(m)} M_\ell(\theta_n) + \varepsilon_{n\ell}^{(a)}$

Prior:

validity intervals + spatial regularization

Case 3: Analysing maps of physical parameters with Beetroots: application to OMC-1, from Palud et al (2025)

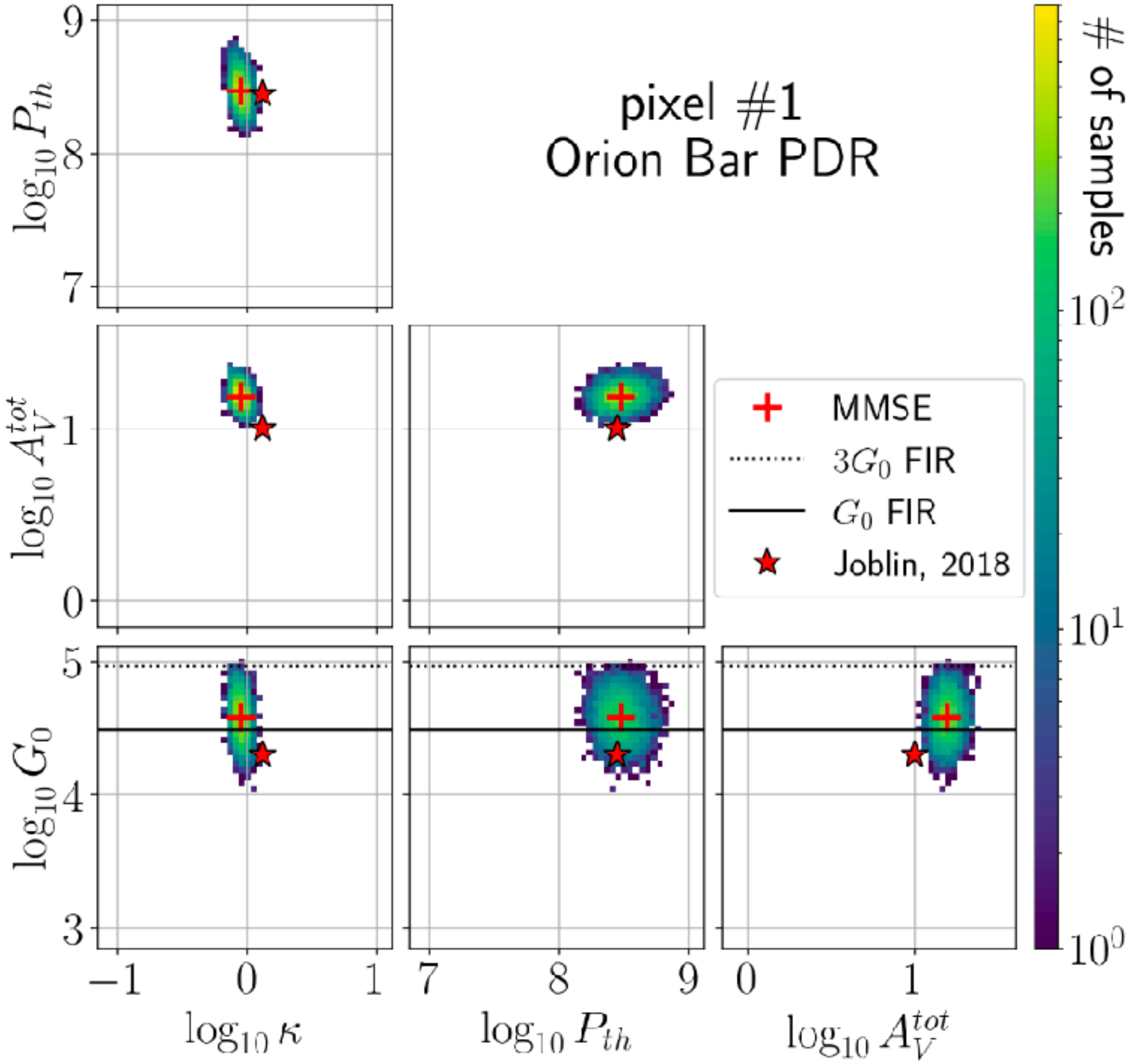


Good estimator

With uncertainty description

Case 3: Analysing maps of physical parameters with Beetroots: application to OMC-1, from Palud et al (2025)

Exploring results for a single pixel



Random event: The visual extinction A_V in the Orion Bar nebula is ≥ 10 mag

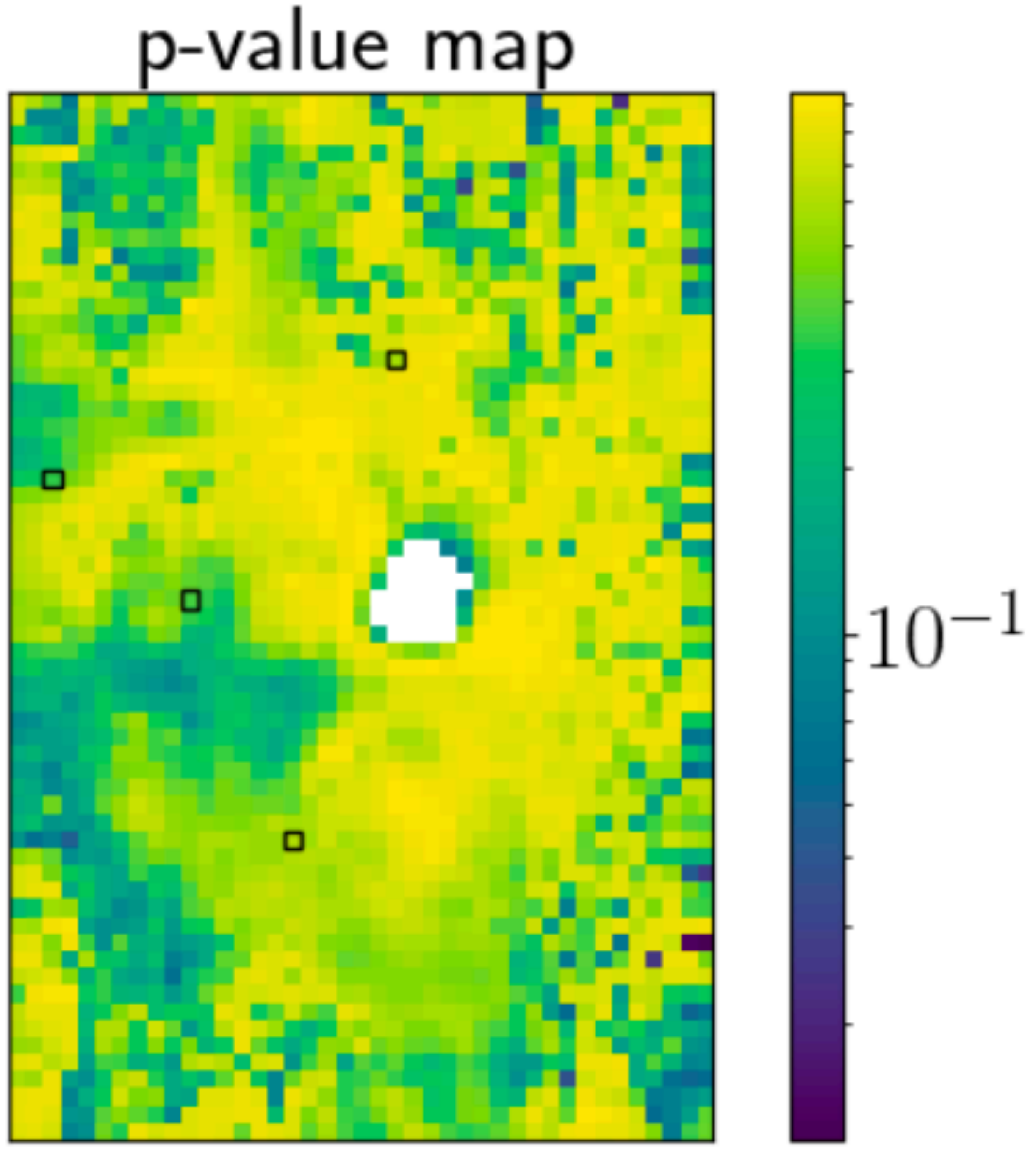
After sampling, **the probability of this random event** (given the many assumptions on observation model, choice of astrophysical simulator, prior distribution) is

$$\simeq \frac{\text{\# samples such that } A_V \geq 10}{\text{total number of samples}}$$

(Here, seems very close to 1)

Case 3: Analysing maps of physical parameters with Beetroots: application to OMC-1, from Palud et al (2025)

Posterior predictive checks: Can I reproduce my observations from the Meudon PDR code, my spatial regularisation prior and observation model ?

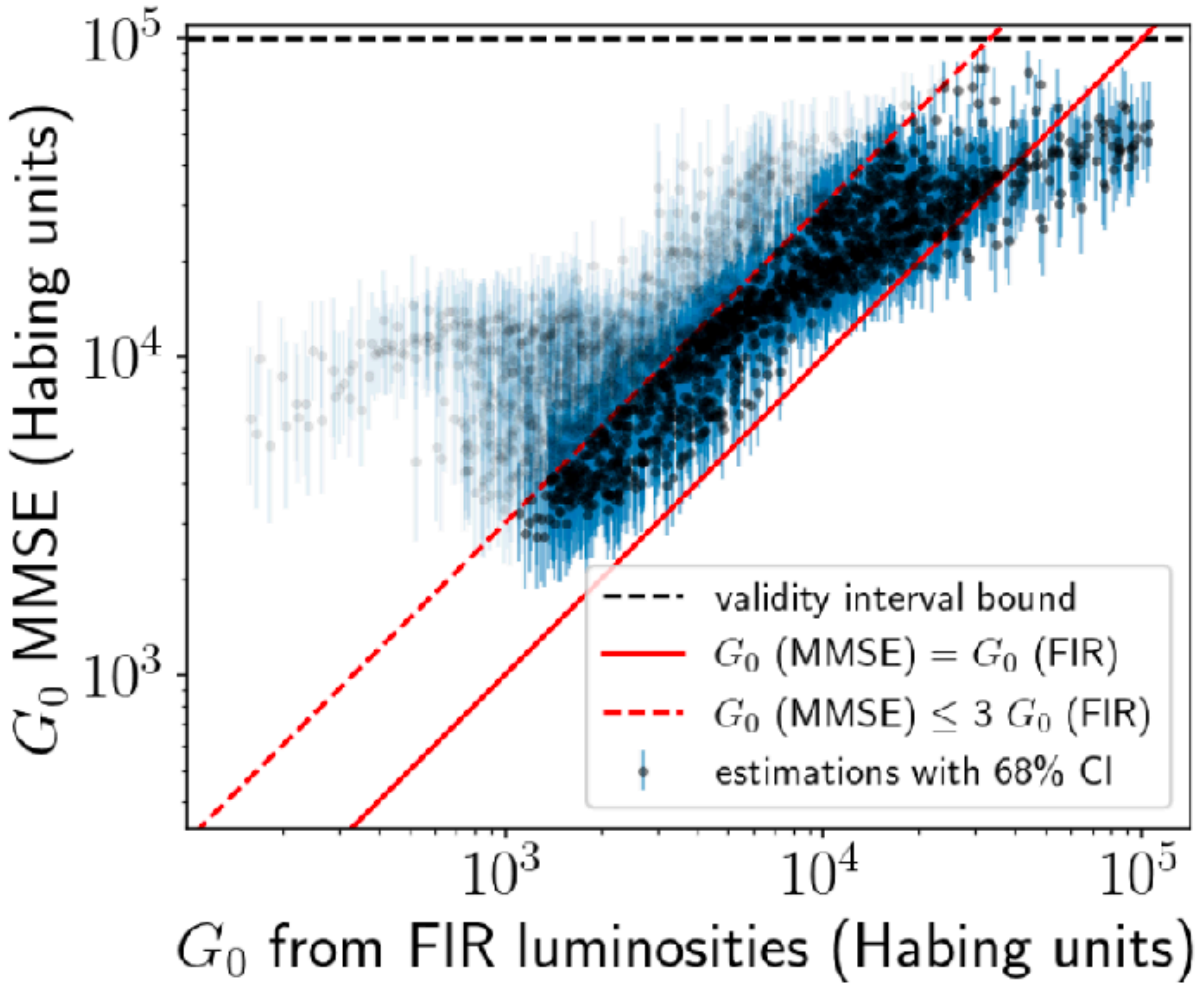


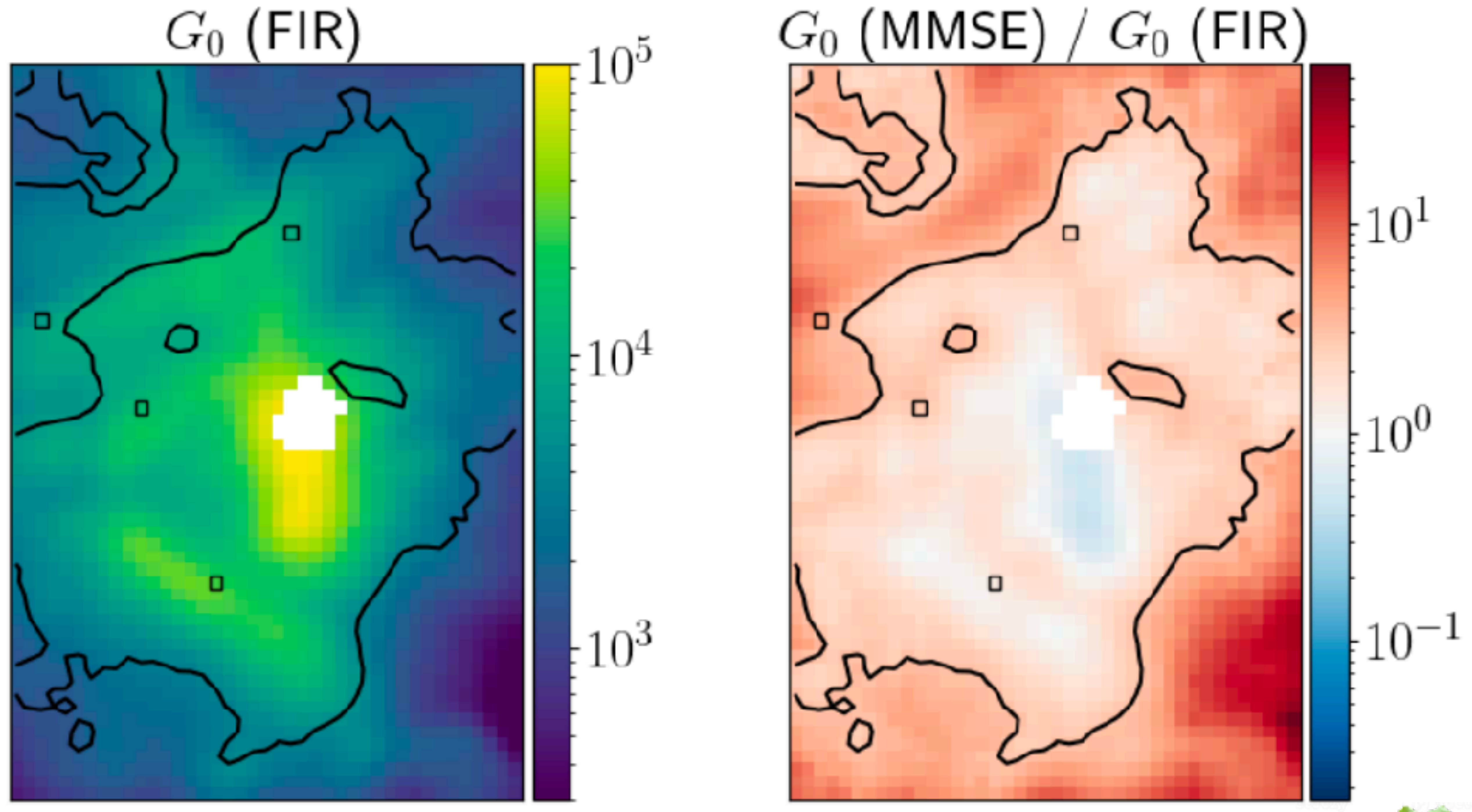
For more information on this topic:

- For a review in ISM: see chapter 3, section 3.3
- For this specific method: see chapter 5, section 5.3

Case 3: Analysing maps of physical parameters with Beetroots: application to OMC-1, from Palud et al (2025)

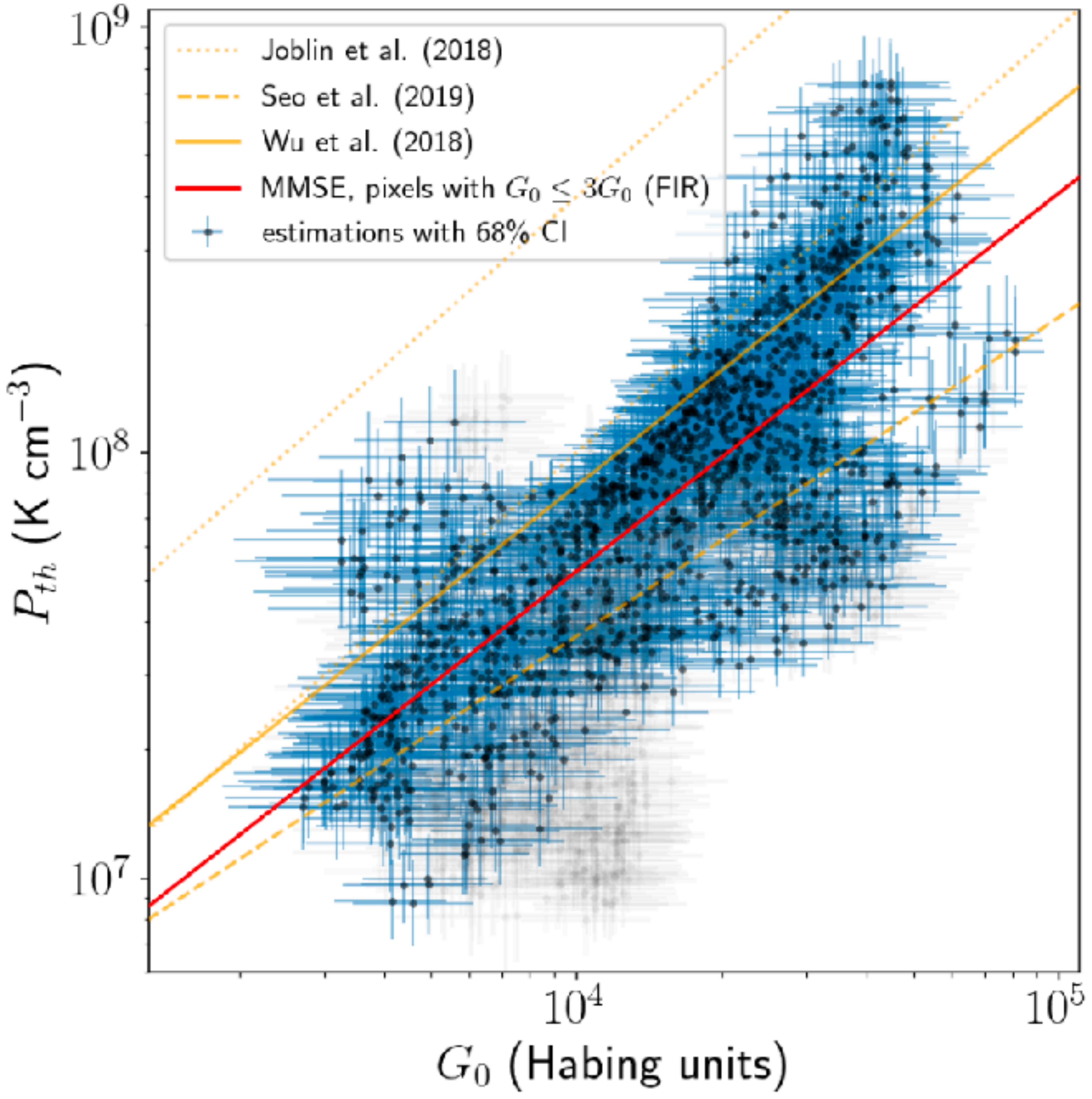
Compatibility with independent estimations, from other tracers

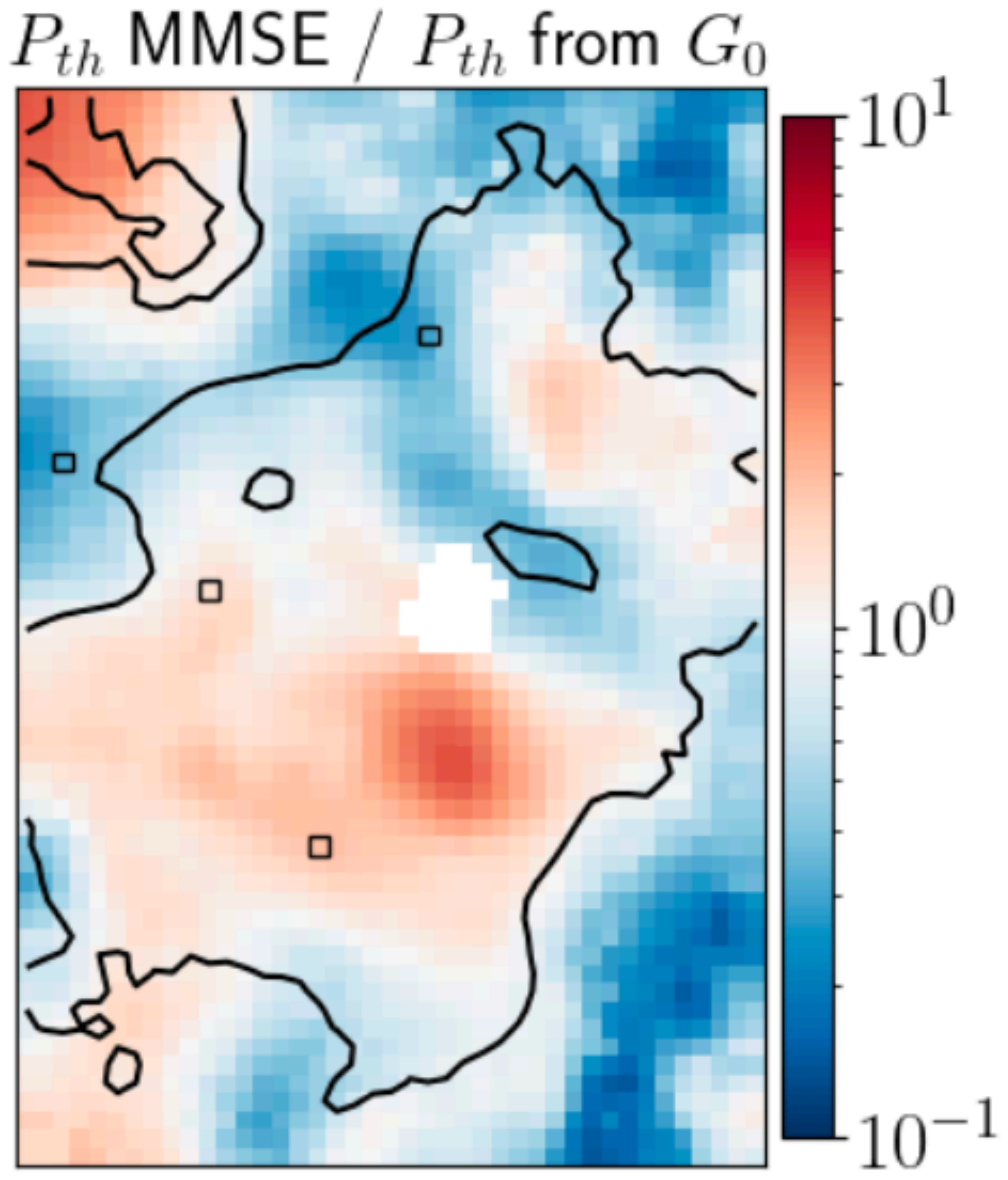




Case 3: Analysing maps of physical parameters with Beetroots: application to OMC-1, from Palud et al (2025)

Checking astrophysical relationships between variables





Summary of Part 4 + Conclusion

Goals: At the end of this class, you should

- Know what the **prior**, **likelihood** and **posterior** are
- Be able to formalise a Bayesian inference task, by identifying the main elements
- Implement the Metropolis-Hastings algorithm on a simple case, and analyse the inference results
- Know some tools to go further and solve more complex problems

In the Hands-on session:

→ You will transform the description of a use case to a likelihood function and prior distribution

→ You will implement MH, visualise the results and evaluate Monte-Carlo estimators

→ Check out this tutorial on Beetroots https://github.com/pierrePalud/beetroots-tuto/tree/main

Thanks!

Check out my webpage : https://pierrepalud.github.io/