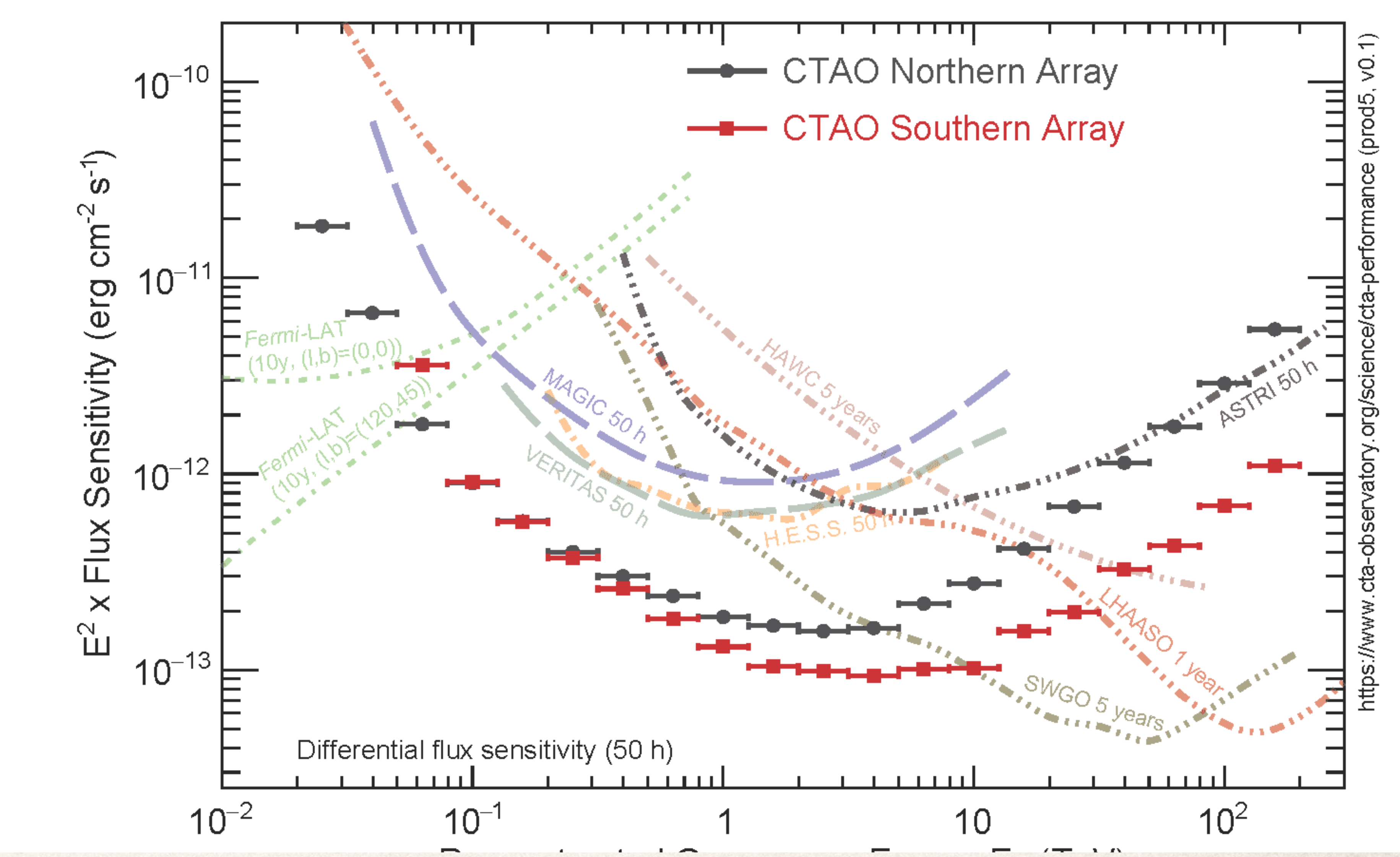
Statistics IV

Atreyee Sinha

Sensitivity and detection potential

- Sensitivity:
  - "What is the faintest flux an instrument can reliably detect (or exclude) after a given observing time?"
  - OR - 5-sigma sensitivity = 5-sigma upper limits for a background only source
  - Upper limits (or values) below the sensitivity are likely to be incorrect
- Detection potential:
  - "If the signal truly exists at some strength, what is the probability the experiment achieves discovery significance?"



Significance and Power

- Significance level = $P(reject\ H_0 | H_0)$ = P(Type 1 error)
  - P-value: probability of type 1 error
- Power = Correctly reject H_0 = 1- P(type II error)
  - Higher power —> better test

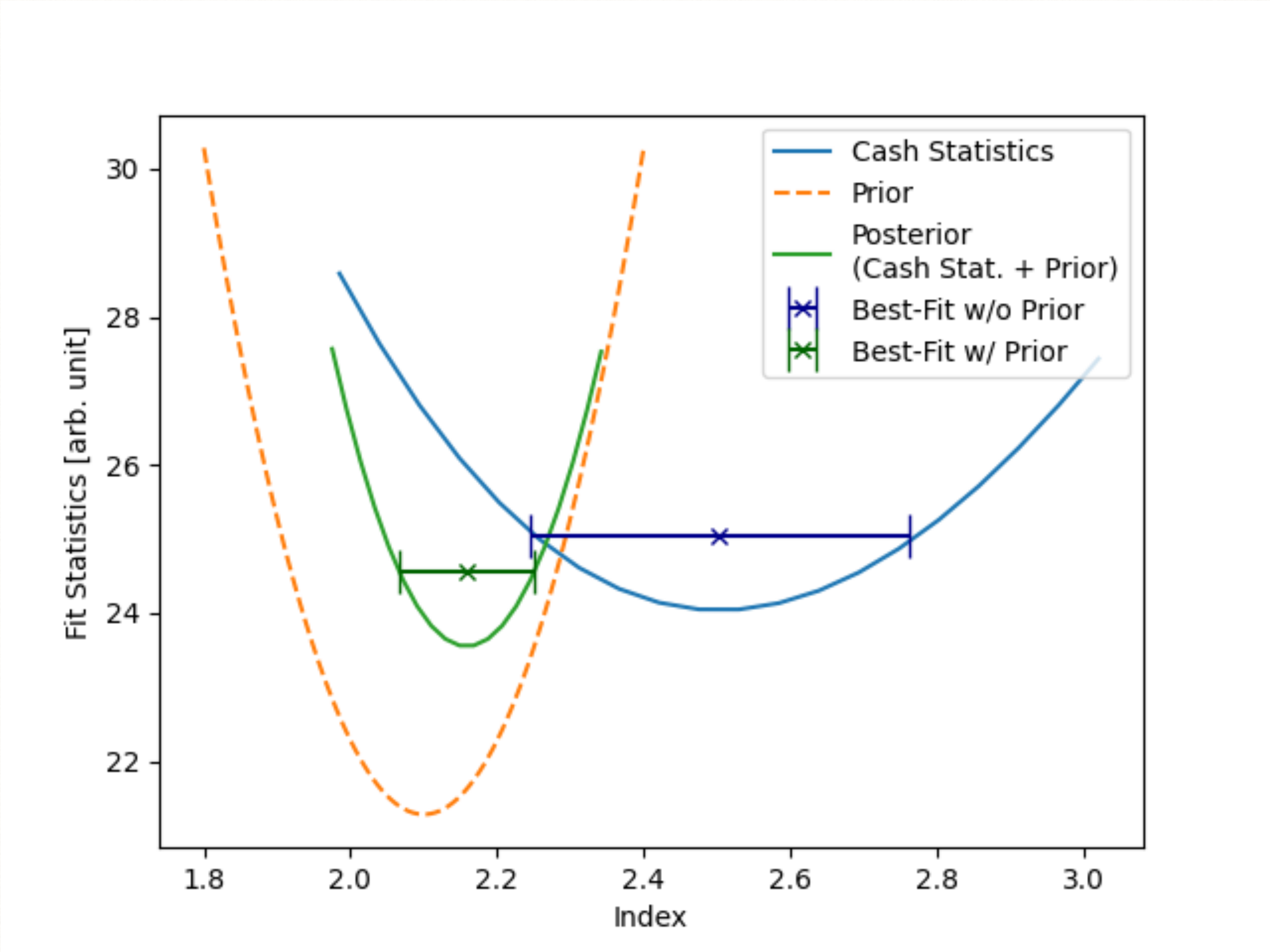
		True state of nature	
		H_0	H_A
Our decision	Reject H_0	Type I error	correct decision
	'Don't reject' H_0	correct decision	Type II error

Type1: false rejection of $H0$

Type II: false non-rejection ('acceptance') of $H0$

MAP: Maximum a priori

- MLE: Estimate parameters with no knowledge on their allowed values
- Reality:
  - Some constrains on the parameters
    - Eg: Flux cannot be negative
    - Bad way to implement: Put bounds on parameters in MLE
    - Why bad: Minuit works between $-\infty$ *to* ∞. If bounds are put, variable transformation is done —> changes well behaved linearisation into non linear problems —> derivative 0 at the edges —> miminizer gets stuck
  - Good way MAP:
    - Add a prior on the parameter —> eg: step function, gaussian, etc etc
    - Penalise the likelihood outside the range
    - Posterior (= Likelihood * Prior) is minimised
      - (add the negative log of the prior probability density to the minimization function
    - Allows a frequentist interpretation while allowing for prior information



A simple example of Bayesian updating

- Lets say I have 5 coins:
  - 2 of Type A: p(H) = 0.5
  - 2 of Type B: p(H) = 0.6
  - 2 of Type C: p(H) = 0.9
- I take one coin at random. What's the probabilit
  - Prior knowledge
- I toss it (my experiment). Output is H.
  - Now what's the probability of it being of eac
    - Posterior

$$P(\mathcal{H} \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \mathcal{H}) P(\mathcal{H})}{P(\mathcal{D})}$$

hypothesis	prior	likelihood	Bayes numerator	posterior (numerator/$P(\mathcal{D})$)
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\|\mathcal{D})$
A	0.4	0.5	0.2	0.3226
B	0.4	0.6	0.24	0.3871
C	0.2	0.9	0.18	0.2903
total	1	NO SUM	$P(\mathcal{D}) = 0.62$	1

$$\boxed{\text{posterior} \propto \text{likelihood} \times \text{prior}}$$

Update…

- Now, I toss it again and get H
  - Previous posterior —> current prior
  - Repeat!

hypothesis	prior	likelihood 1	Bayes numerator 1	likelihood 2	Bayes numerator 2	posterior 2
θ	$p(\theta)$	$p(x_1 = 1 \mid \theta)$	$p(x_1 = 1 \mid \theta)p(\theta)$	$p(x_2 = 1 \mid \theta)$	$p(x_2 = 1 \mid \theta)p(x_1 = 1 \mid \theta)p(\theta)$	$p(\theta \mid x_1 = 1, x_2 = 1)$
0.5	0.4	0.5	0.2	0.5	0.1	0.2463
0.6	0.4	0.6	0.24	0.6	0.144	0.3547
0.9	0.2	0.9	0.18	0.9	0.162	0.3990
total	1	NO SUM		NO SUM	0.406	1

Predictive probabilities

- Probabilistic prediction simply means assigning a probability to each possible outcomes of an experiment.
  - Predictive prior: (probabilistic) prediction of what will happen if the coin is tossed
  - Predictive posterior: (probabilistic) prediction of what will happen if the coin is tossed *again*

Bayesian updating for continuous priors

- We have a bent coin with unknown probability θ of heads. The value of θ is random with prior pdf $f(\theta) = 2\theta$. Suppose we flip the coin three times and get the sequence HTT . Compute the posterior pdf for θ.

hypothesis	range	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}_\theta$	θ range	$f(\theta)\,d\theta$	$p(x = 1,1,0\|\theta)$	$p(x = 1,1,0\|\theta)f(\theta)\,d\theta$	$f(\theta\|x = 1,1,0)\,d\theta$
$\mathcal{H}_\theta$	$[0,1]$	$2\theta\,d\theta$	$\theta^2(1-\theta)$	$2\theta^3(1-\theta)\,d\theta$	$20\theta^3(1-\theta)\,d\theta$
total	$[0,1]$	$\int_a^b f(\theta)\,d\theta = 1$	no sum	$p(x = 1,1,0) = \int_0^1 2\theta^3(1-\theta)\,d\theta = 1/10$	1

Therefore the posterior pdf (after seeing HHT) is $\boxed{f(\theta|x) = 20\theta^3(1-\theta)}$.

Constructing a prior: subjective!

- The Experiment: Measuring g using a simple pendulum.
- The New Data: A single (hasty) measurement yields a value with a large uncertainty: ($g = 12.0 \pm 2.0$ m/s^2) .
- What should you take as prior?
  - Person A: Agnostic —> Uniform prior: $P_A(g) = \dfrac{1}{30} \quad \text{for } 0 \le g \le 30 \text{ m/s}^2$
  - Person B: Physics student —> Gaussian Prior $P_B(g) = \dfrac{1}{0.1\sqrt{2\pi}} \exp\left(-\dfrac{(g-9.8)^2}{2(0.1)^2}\right)$
    - Note: choice of $\sigma = 0.1$ is also subjective
- Effect of prior:
  - Person A: $g \sim 12.0\ m/s^2$
  - Person B: $g \sim 9.8 m/s^2$

Conjugate functions

- For most functional forms, analytical computation of the evidence is often impossible
  - Choose a prior that matches the algebraic form of the likelihood.
  - The posterior belongs to the same probability family as the prior.
  - Bypasses the denominator integral entirely —> normalisation is known

$$\underbrace{P(\theta|D)}_{\text{Posterior}} = \frac{\overbrace{P(D|\theta)}^{\text{Likelihood}}\overbrace{P(\theta)}^{\text{Prior}}}{\underbrace{\int P(D|\theta)P(\theta)d\theta}_{\text{Evidence}}}$$

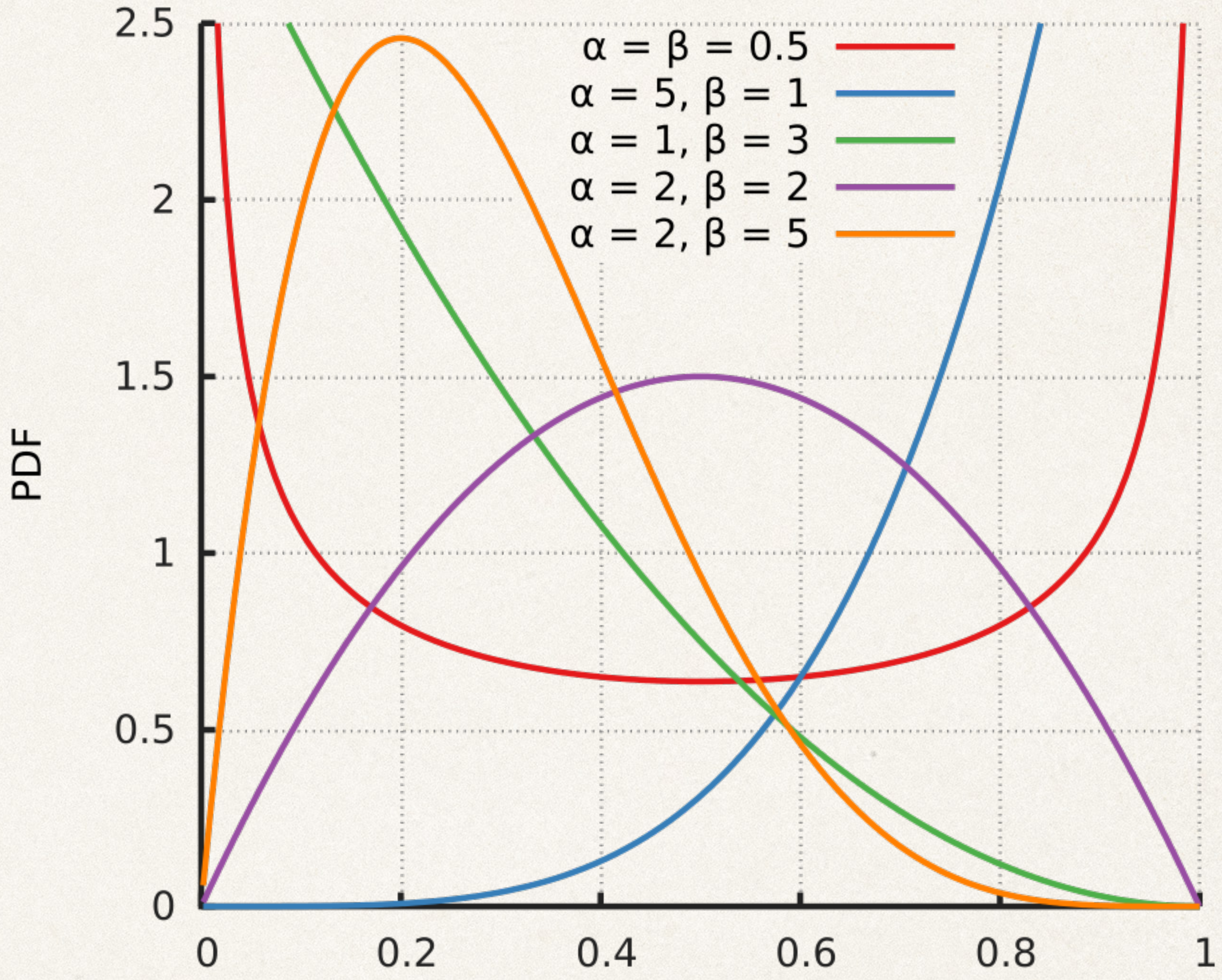
Common conjugate functions

Likelihood (Data Model)	**Prior Distribution**	**Posterior Distribution**	**Physical Application Example**
Poisson (Counts)	**Gamma**	**Gamma**	UHE Cosmic Ray / Photon arrivals
Gaussian (Known Variance)	**Gaussian**	**Gaussian**	CMB Power Spectrum / Calibration
Binomial (Success/Failure)	**Beta**	**Beta**	Detector trigger / Selection efficiencies
Exponential (Time intervals)	**Gamma**	**Gamma**	Particle lifetimes / Decays

Beta distribution

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

- Definition: A continuous probability distribution defined on the bounded interval [0, 1].
- Physics Context: Ideal for modeling parameters that represent fractions, probabilities, or bounded efficiencies (e.g., detector trigger efficiencies).
- Conjugacy: It is the conjugate prior for Binomial data models (Success / Failure outcomes).



$$\text{Mean: } E[x] = \frac{\alpha}{\alpha+\beta}$$

$$\text{Mode: } M_o = \frac{\alpha-1}{\alpha+\beta-2} \quad (\text{for } \alpha,\beta > 1)$$

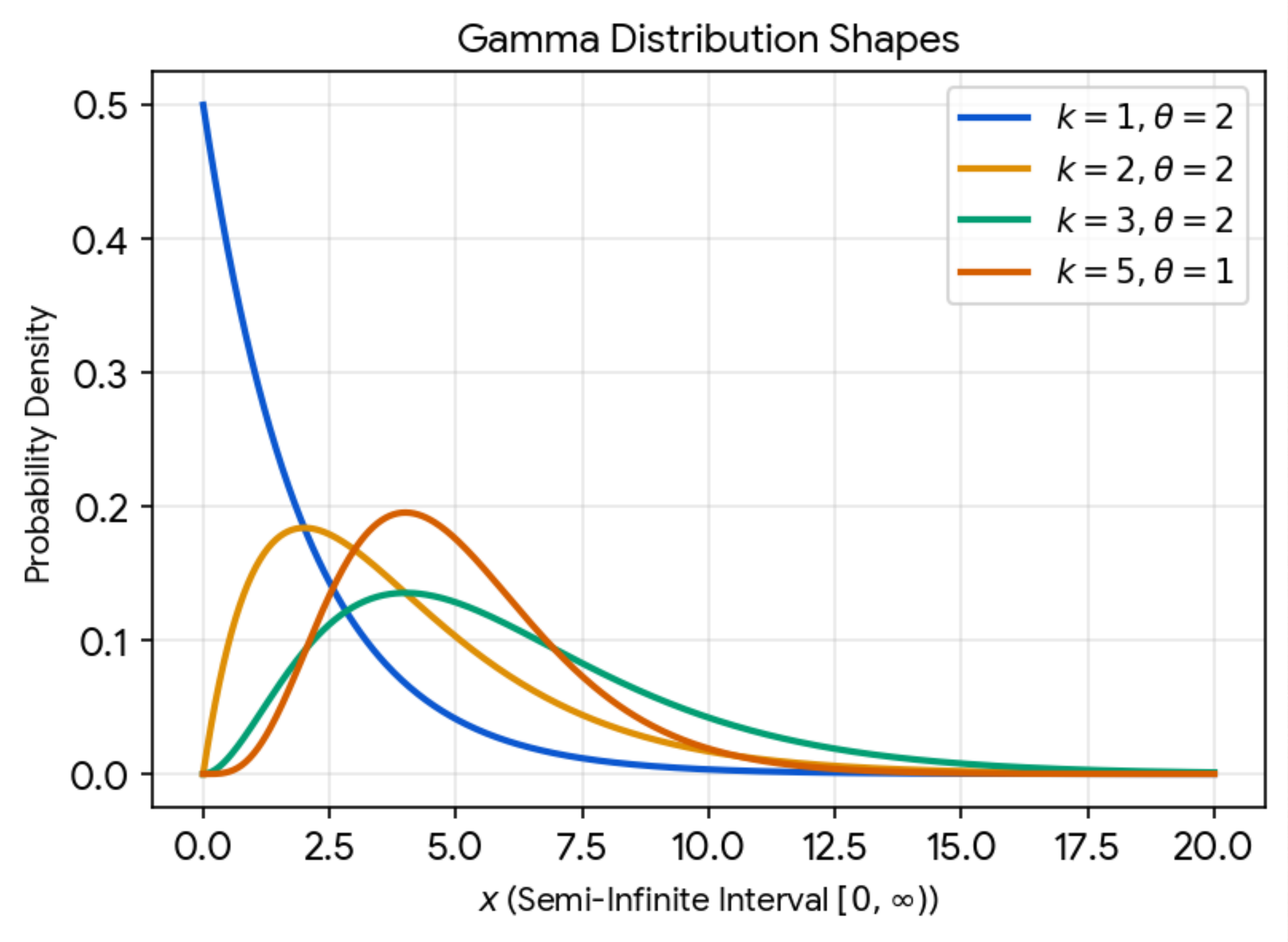
Gamma distribution

$$f(x; k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}$$

- Definition: A continuous probability distribution defined on the semi-infinite interval [0,∞].

- Physics Context: Ideal for modeling physically positive parameters like particle decay rates, cosmic ray flux scales, or expected arrival rates.

- Conjugacy: It is the conjugate prior for both Poisson (event counts) and Exponential (decay lifetimes) data models.



Mean: $E[x] = k\theta$

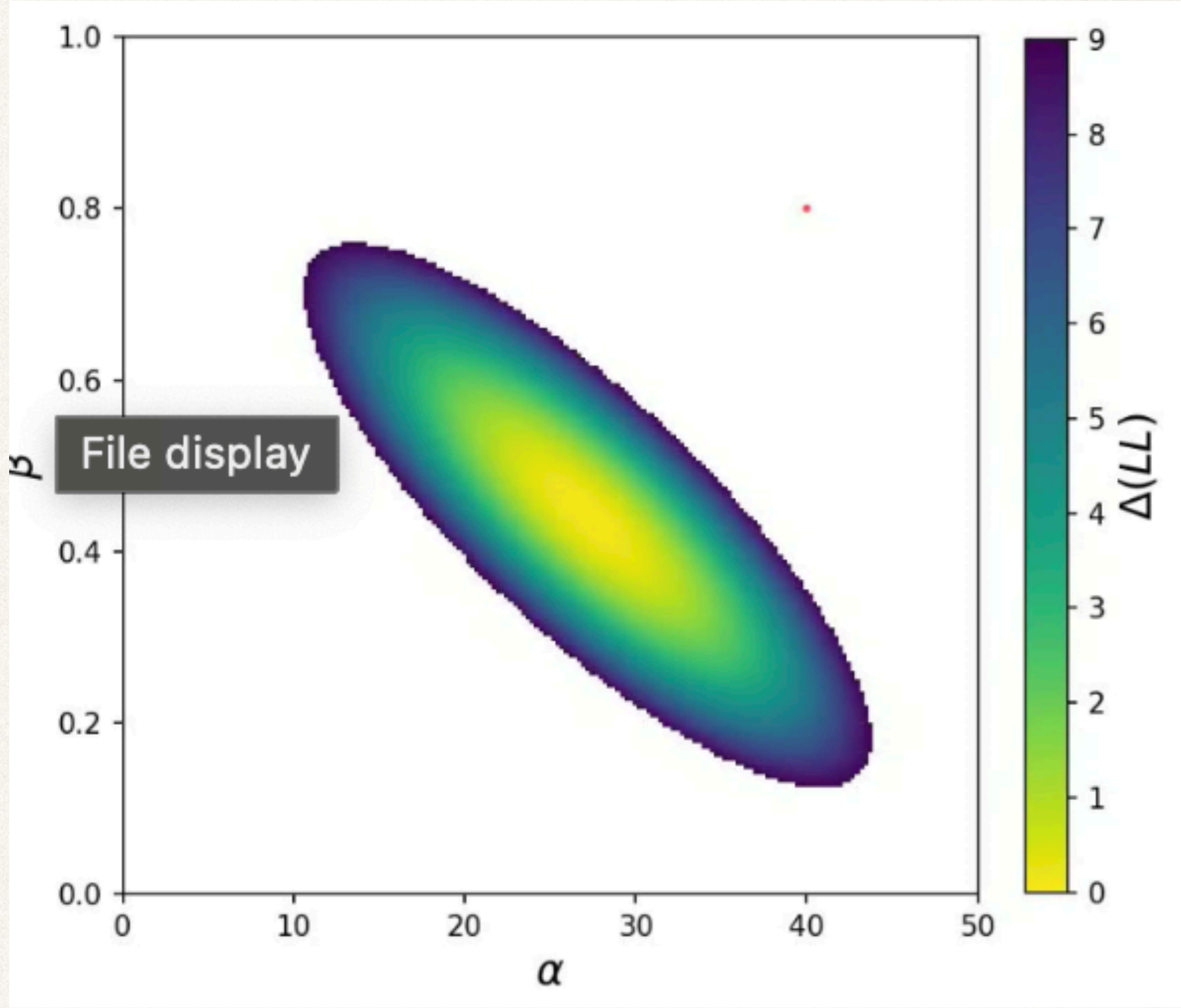
Mode: $M_o = (k-1)\theta \quad (\text{for } k \geq 1)$

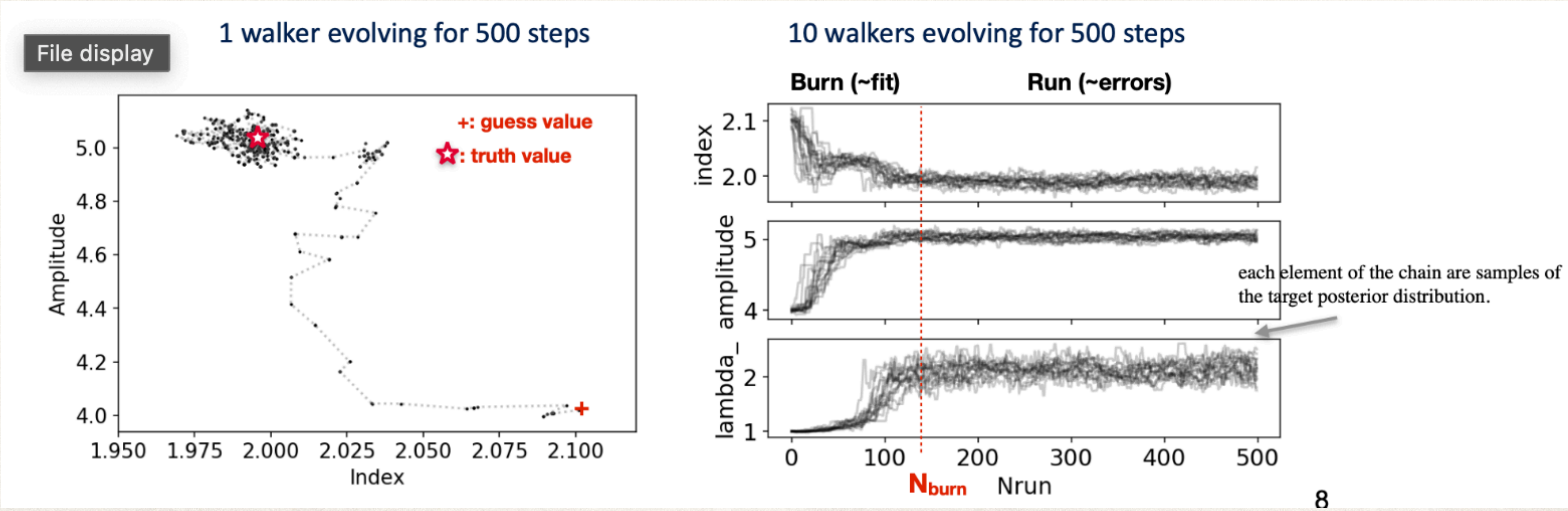
Updating the conjugate prior

- Multiplying a conjugate prior by its matched likelihood yields a posterior where you simply add the data straight to the prior's parameters.
- No Integration Needed: The complex denominator integral cancels out because the functional form is preserved.
- Binomial Data × Beta Prior →Beta Posterior
  - Prior: $\beta_{\text{new}} = \beta_{\text{old}} + (n - k)$
- Poisson Data × Gamma Prior → Gamma Posterior
  - Prior: $\theta_{\text{new}} = \dfrac{\theta_{\text{old}}}{T \cdot \theta_{\text{old}} + 1}$

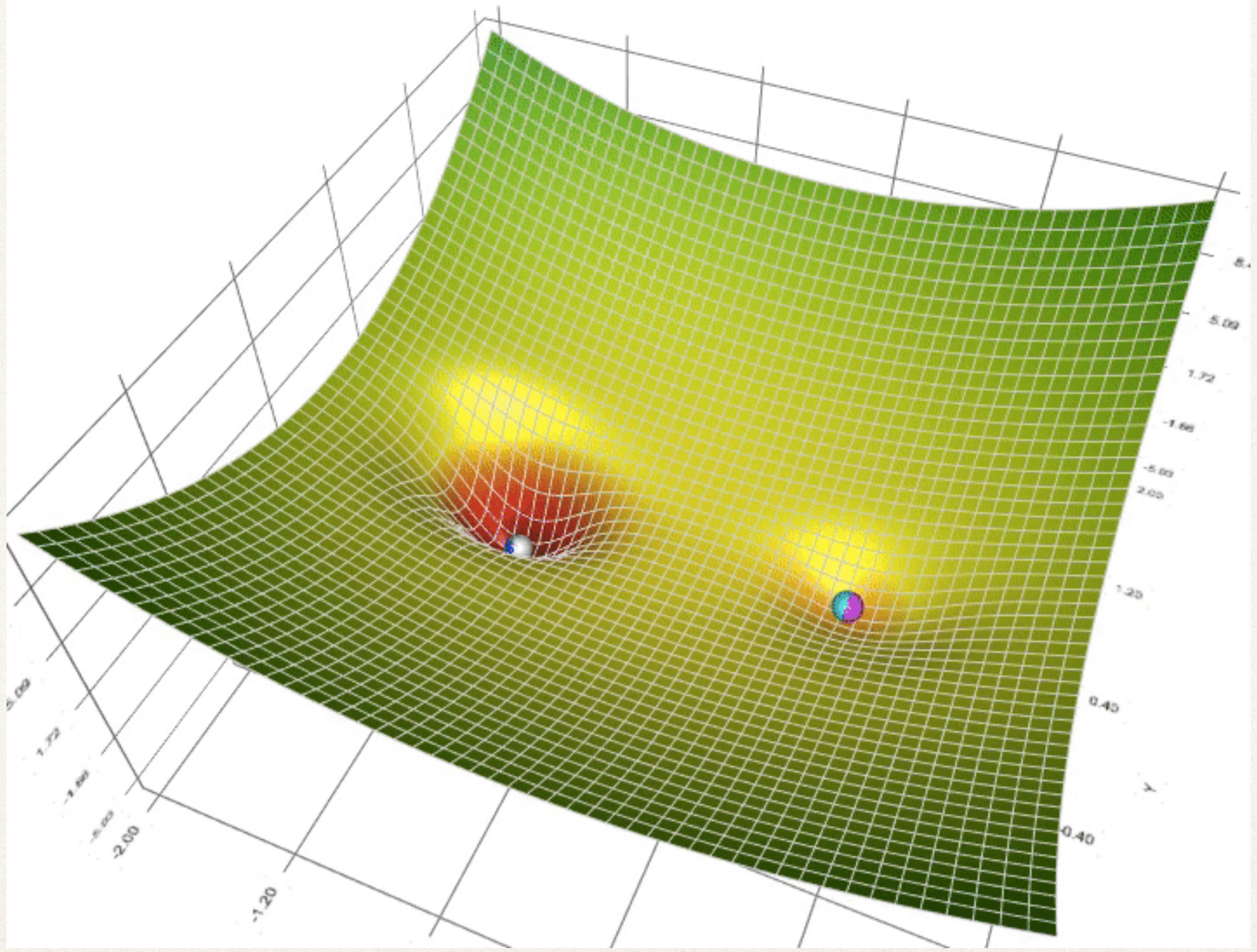
Computational Sampling: MCMC

- Often conjugates unavailable:
  - Numerical sampling is required
- MCMC
  - Explores the parameter space via a random walk proportional to the posterior height.
  - Maps the peak shapes and calculates parameter uncertainties.
  - Need to define an initial point
    - Might stay trapped —> miss multimodal posteriors
  - Evolve walkers for long to sample target posteriors



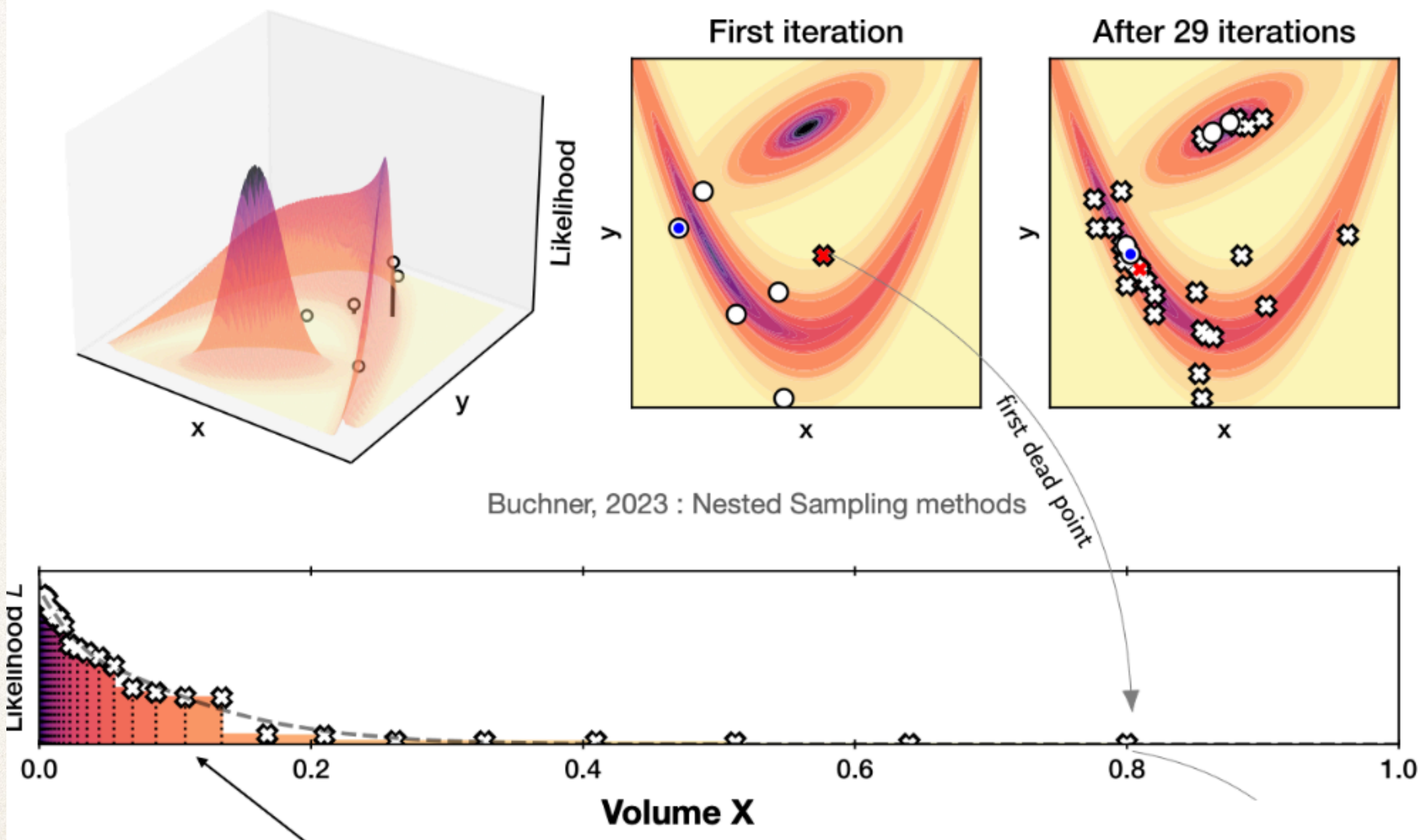


Ending up in local minima



Nested Sampling

Basic steps of Nested Sampling



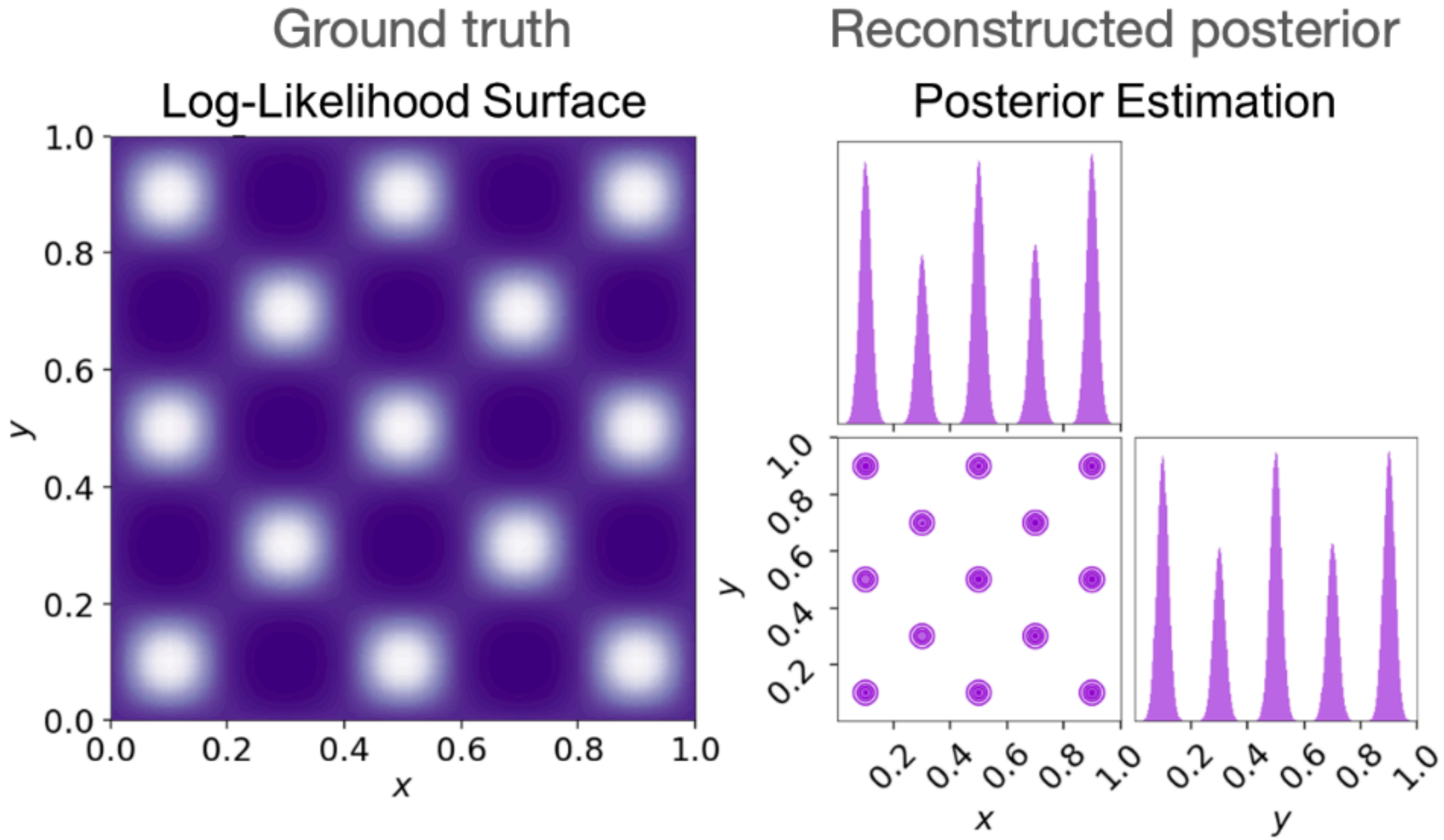
Buchner, 2023 : Nested Sampling methods

$$Z = \sum L \Delta X$$ **Integration via trapezoid rule**

Z = Bayesian evidence = ∫ Like(Data|θ) * Prior(θ) dθ

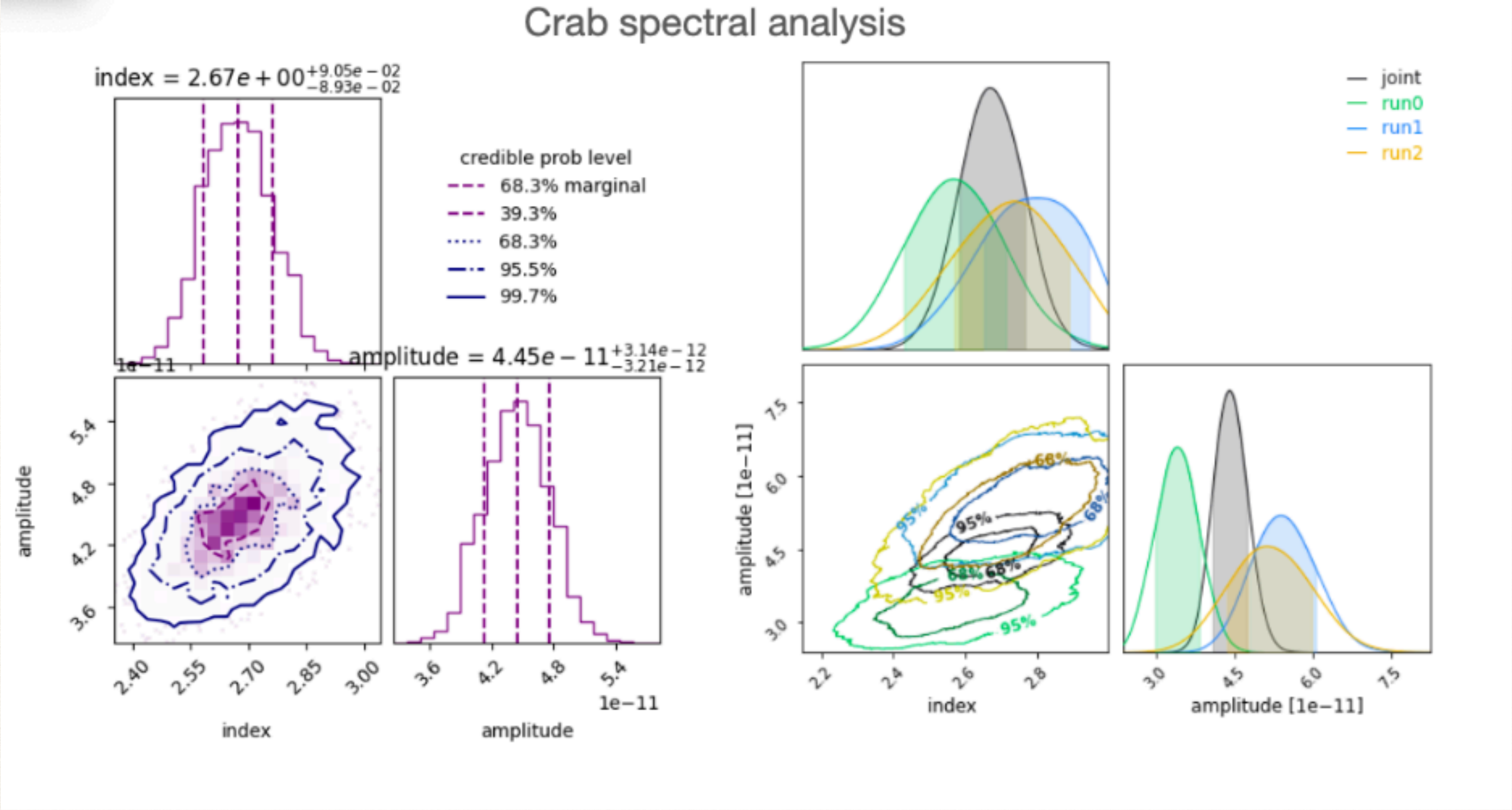
- **Define priors on params**
- **Transform param space to volume unity cube X**
- **Draw random uniform N_{live} (~400-1000)**
- **Iterate until stop criteria:**
  - **Remove worst LogLike point**
  - **Draw new point with a better LL**
    - **Likelihood-restricted prior sampling**
    - **(The tricky part)**
  - **Increment Z= Z + L*ΔX**
  - **Stop criteria : ΔZ/Z<tolerance**

- **Works perfectly well in landscapes that would drive crazy any MCMC or gradient descent method**



Speagle, 2020

Generally shown in terms of corner plots



Frequentist vs Bayesian

Frequentist

- Objective: Everyone will agree on the p-value
- Demands careful description of experiment
- Additional data needs full analysis
- Ad-hoc and often misses implementation of known information

Bayesian

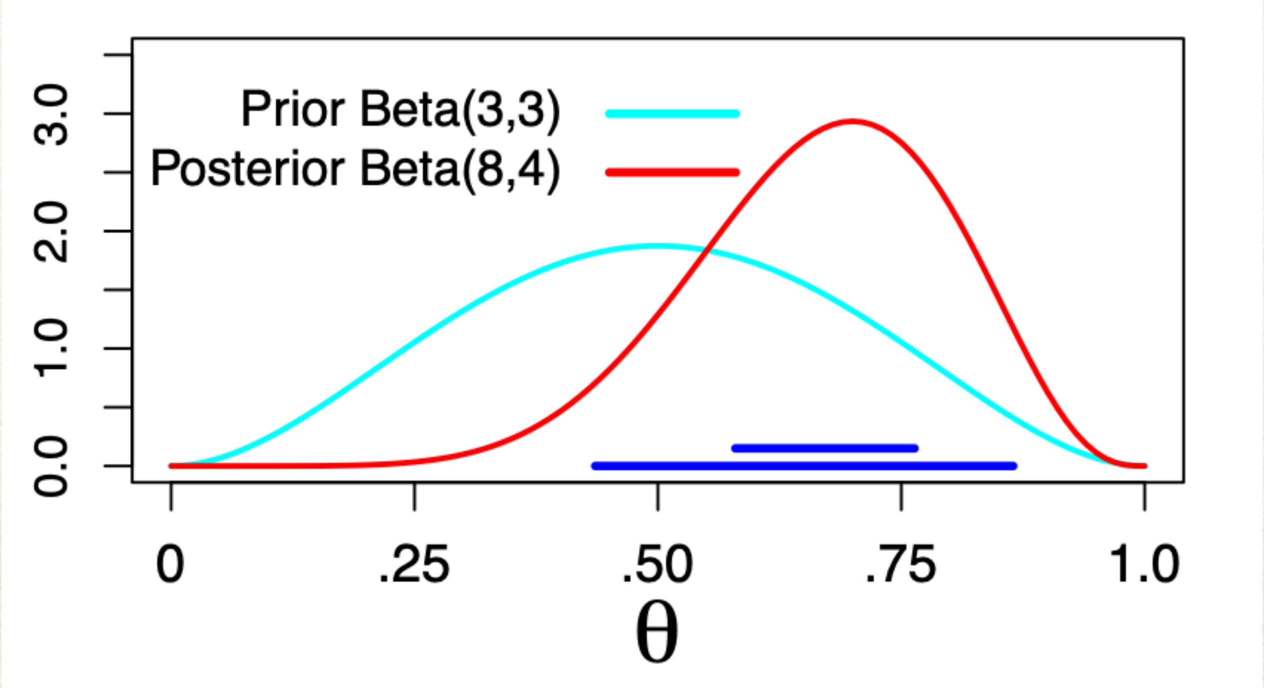
- Prior: Subjective choice
  - Different posteriors and conclusions
- Logically rigorous
- Data can be updated as available
- Evidence derived less dependent on experimental setup

Stopping rules

Result of my coin toss is HHHHHT. Is it a biased coin?

- Frequentist:
  - $H_0 : \hat{\theta} = 0.5$ & $H_1 : \hat{\theta} > 0.5$
  - What was the experiment?
    1. Toss coin 6 times and find now many H.
       - $H_0 : 6 * p^5 * (1 - p) \sim 0.1$
    2. Toss till you get T:
       - $H_0 : 6 * p^5 = 0.03$
  - Conclusion:
    - At 95% confidence ($p = 0.05$); reject Case (2) but not Case (1)

- Bayesian:
  - Choose a prior:: $\beta(3,3)$
  - What was the experiment?
    - Doesn't matter! Keep doing Bayesian updating
  - Conclusion:
    - The posterior probability that the coin is biased toward H = 0.89.



Discussion on Frequentist vs Bayesian

- Lets say you are asked to predict daily ridership for a new metro line.
- Frequentist:
  - p is fixed, N total number of people in region.
  - $X \sim Binomial(N, p)$; $E[X] = Np$
  - p estimated historical data/surveys/etc…
- Bayesian:
  - p is not fixed, its a random variable
  - $p \sim \beta(a, b)$; a : People who did take the metro, b : People who did not take metro
  - Choice of (a, b) decides strength of the prior.
    - $a/(a + b)$: match expected ridership
    - $(a + b)$: How confident you are of the expectation
  - After first ride, update the expectation
    - Lets say k out of n people took metro

$$p(\theta) \sim \beta(a, b)$$

$$p(\theta) \sim \beta(a + l, b + n - k)$$

$$E[X] = N * (a + k)/a + b + n$$