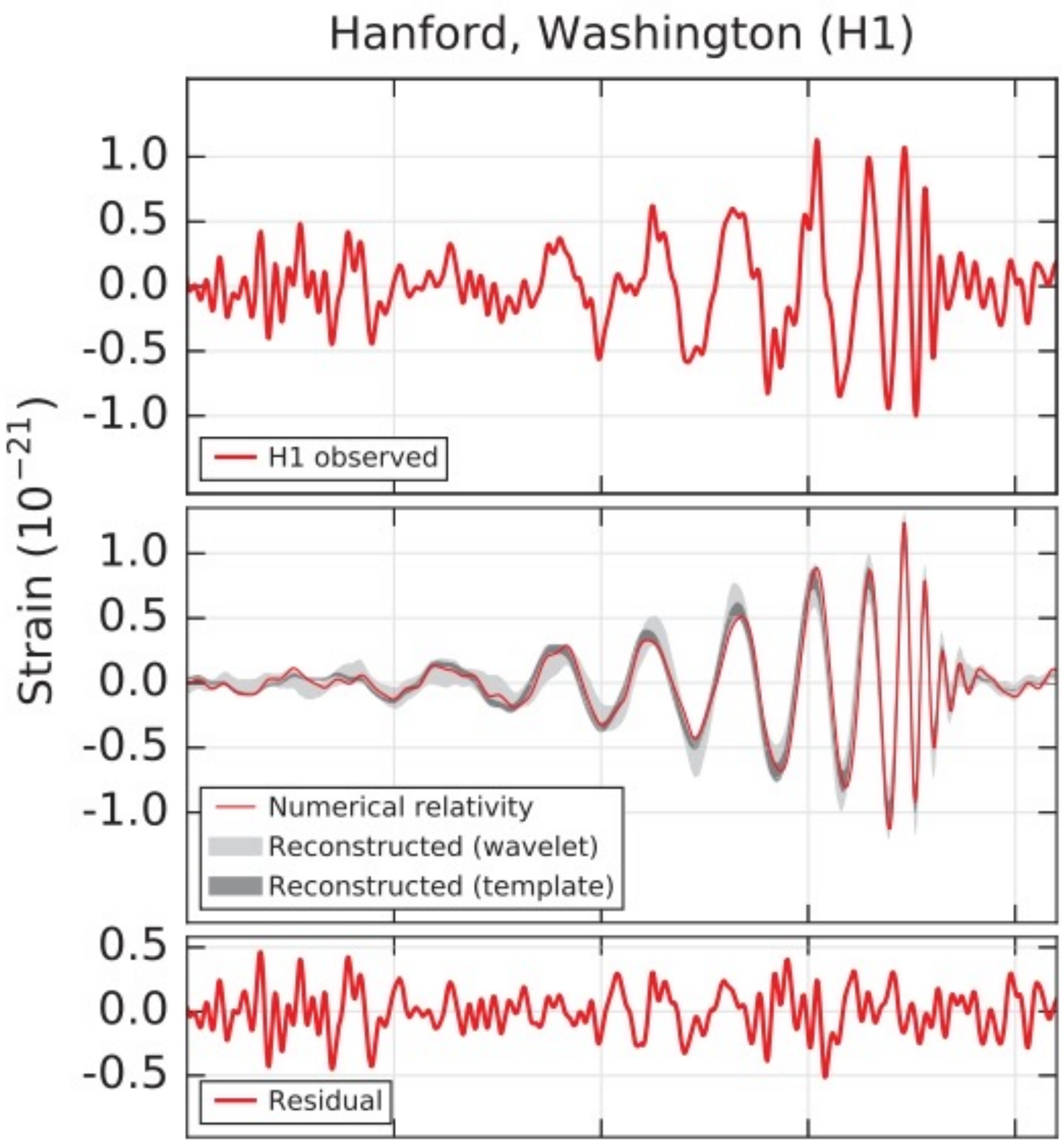
Estimation of Gravitational Wave parameters

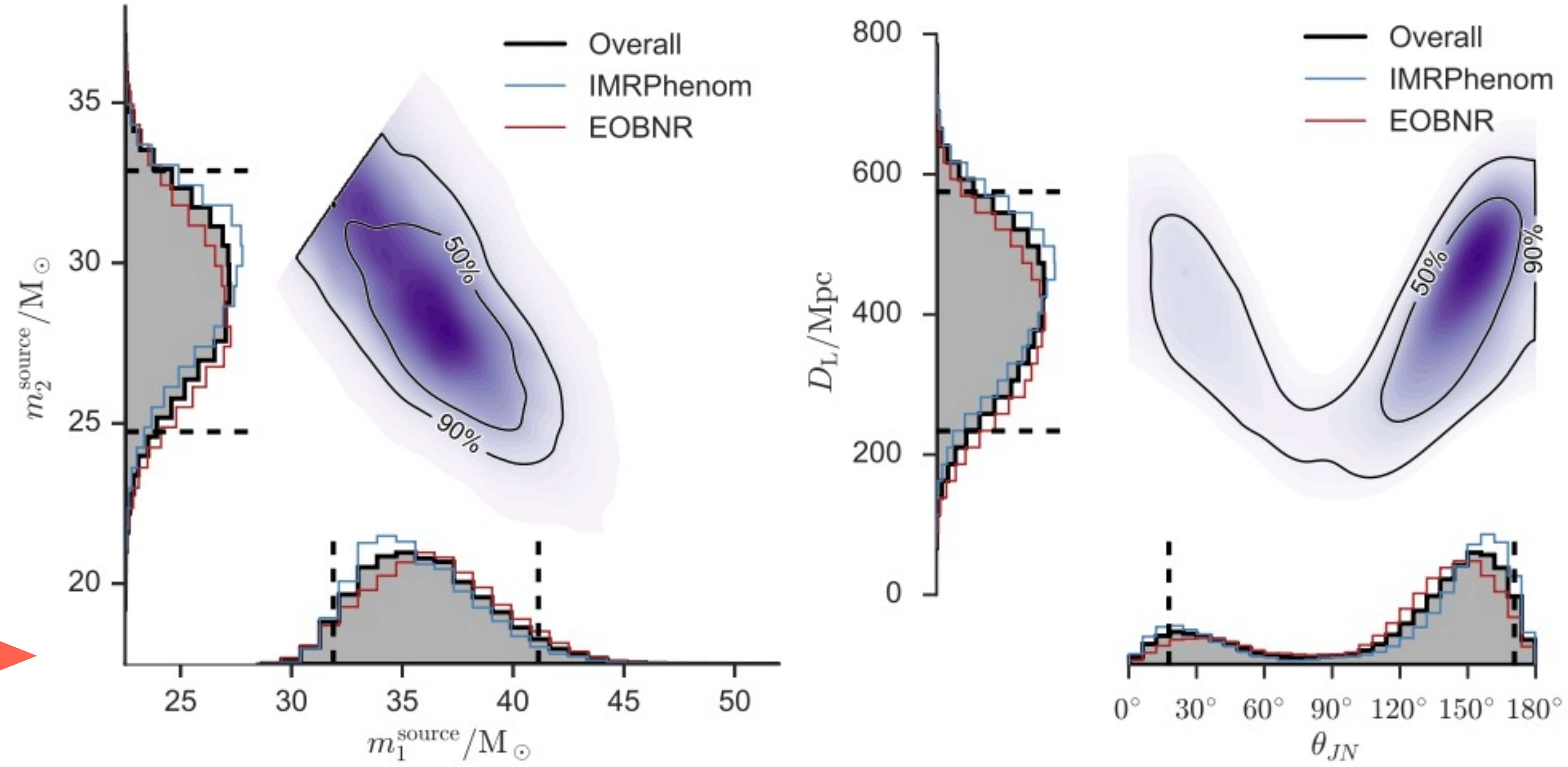
GW open data workshop
14 May 2025

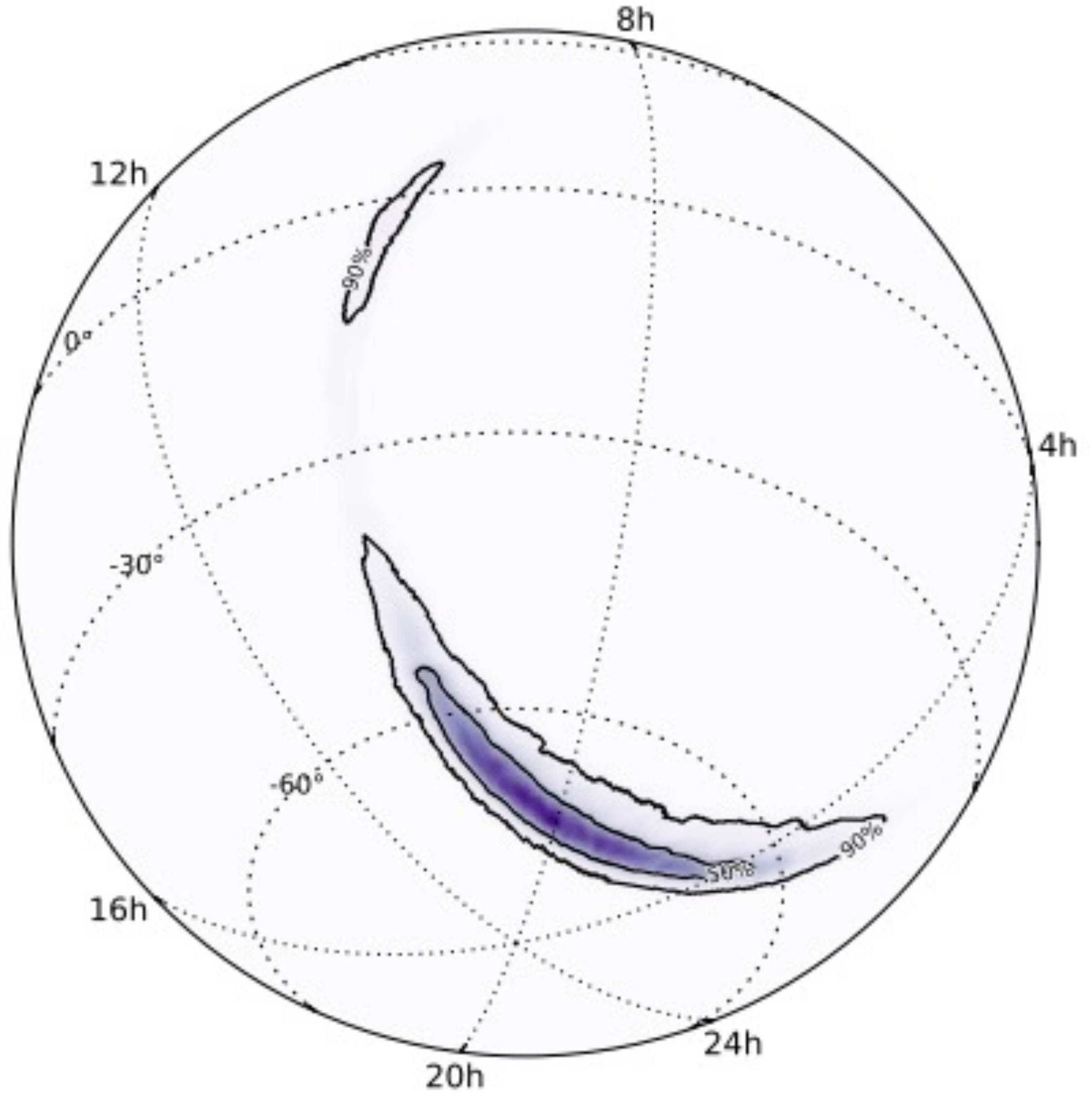
Neha Sharma
Neha.sharma@icts.res.in

INTERNATIONAL CENTRE *for* THEORETICAL SCIENCES
TATA INSTITUTE OF FUNDAMENTAL RESEARCH

From signal to source : Extracting parameters







arXiv:1602.03837

arXiv:1602.03840

Why estimate the parameters ?

- **Understanding the source** :

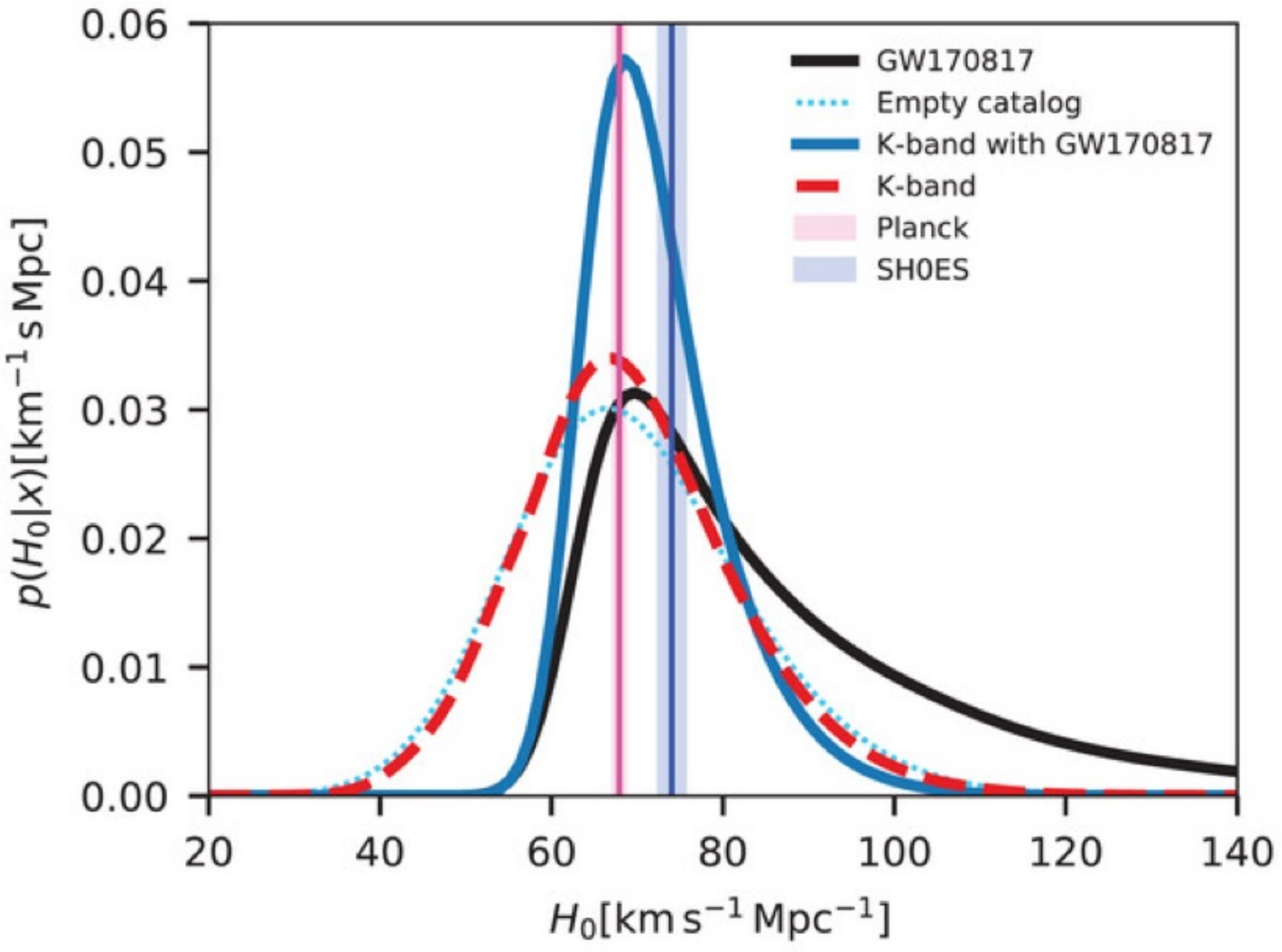
  1. The masses, spins of compact object, distance and inclination angle etc.

  2. The equation of state of neutron stars

- **Tests of General Relativity** : Searching for derivation for GR

- **Cosmology**

And much more!



Abbott et al. (2023b)

GW: Data, Signal and Bayes theorem

Data Model

$$d(t) = h(t;\vec{\theta}) + n(t)$$

Data $d(t)$

Signal $h(t;\vec{\theta})$

Noise $n(t)$

Signal : 15+ Parameters

- **Intrinsic parameters**:
    1. Masses (m_1, m_2)
    2. Spins $(\vec{\chi_1}, \vec{\chi_2})$
    3. Tidal deformability (Λ_1, Λ_2) - Only for NS

- **Extrinsic parameters**:
    1. Orbital inclination angle (θ_{JN}), Luminosity distance (d_L)
    2. Sky location (α, δ), Polarisation (ψ)
    3. Coalescence time (t_c) and phase (ϕ_c)

$$h(t;\vec{\theta}) = \sum_{i=+,\times} F_i(\alpha,\delta,\psi)h_i(t;\vec{\theta})$$

Signal

Antenna Pattern

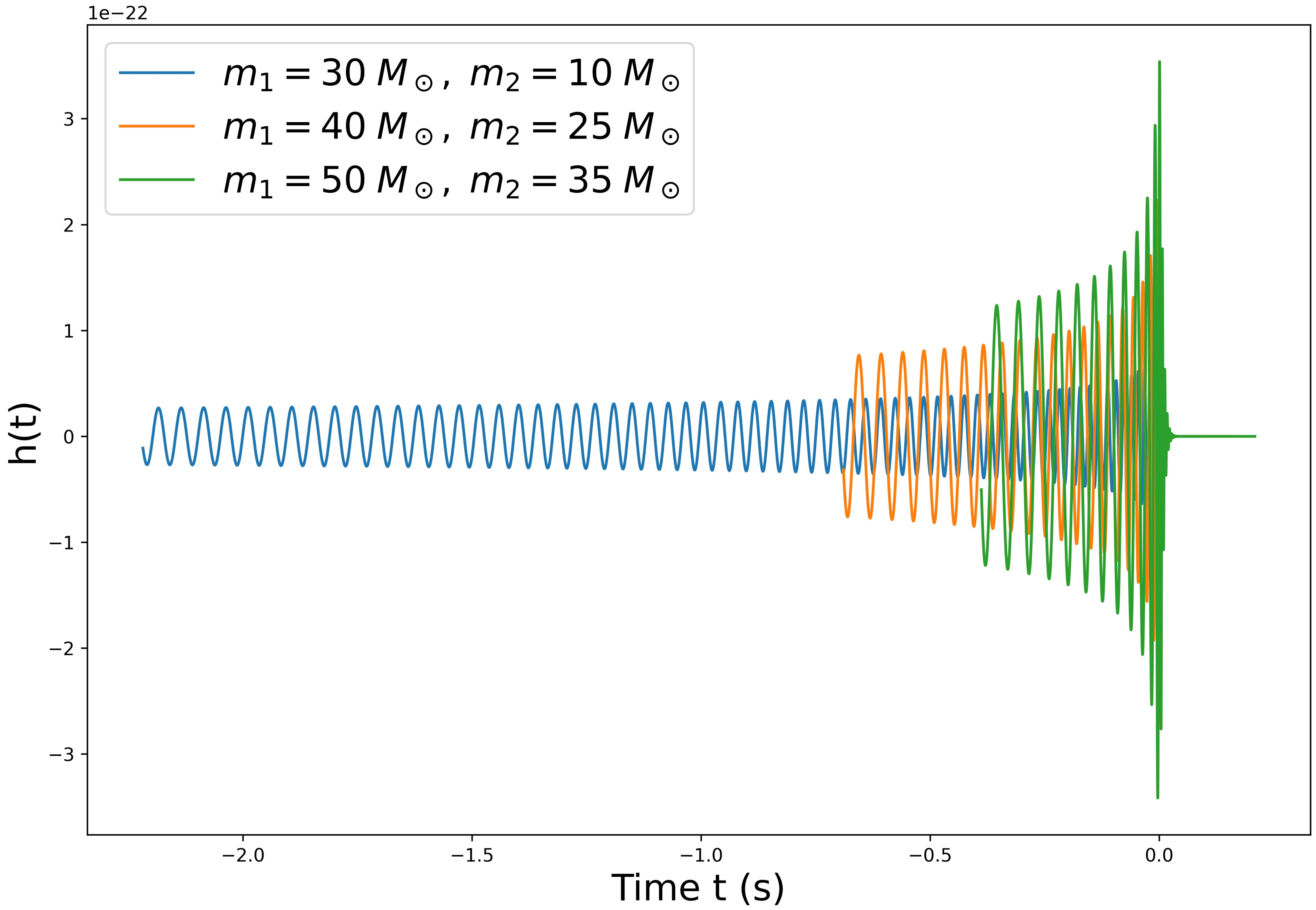
Polarizations

Impact of intrinsic parameters

Masses : m_1, m_2

Higher mass $\Rightarrow$ Higher amplitude

Higher mass $\Rightarrow$ Shorter signal



Impact of intrinsic parameters

Credit : Vijay Varma et al,
Binary Black Hole Explorer

Spins $\overrightarrow{\chi_1}, \overrightarrow{\chi_2}$:

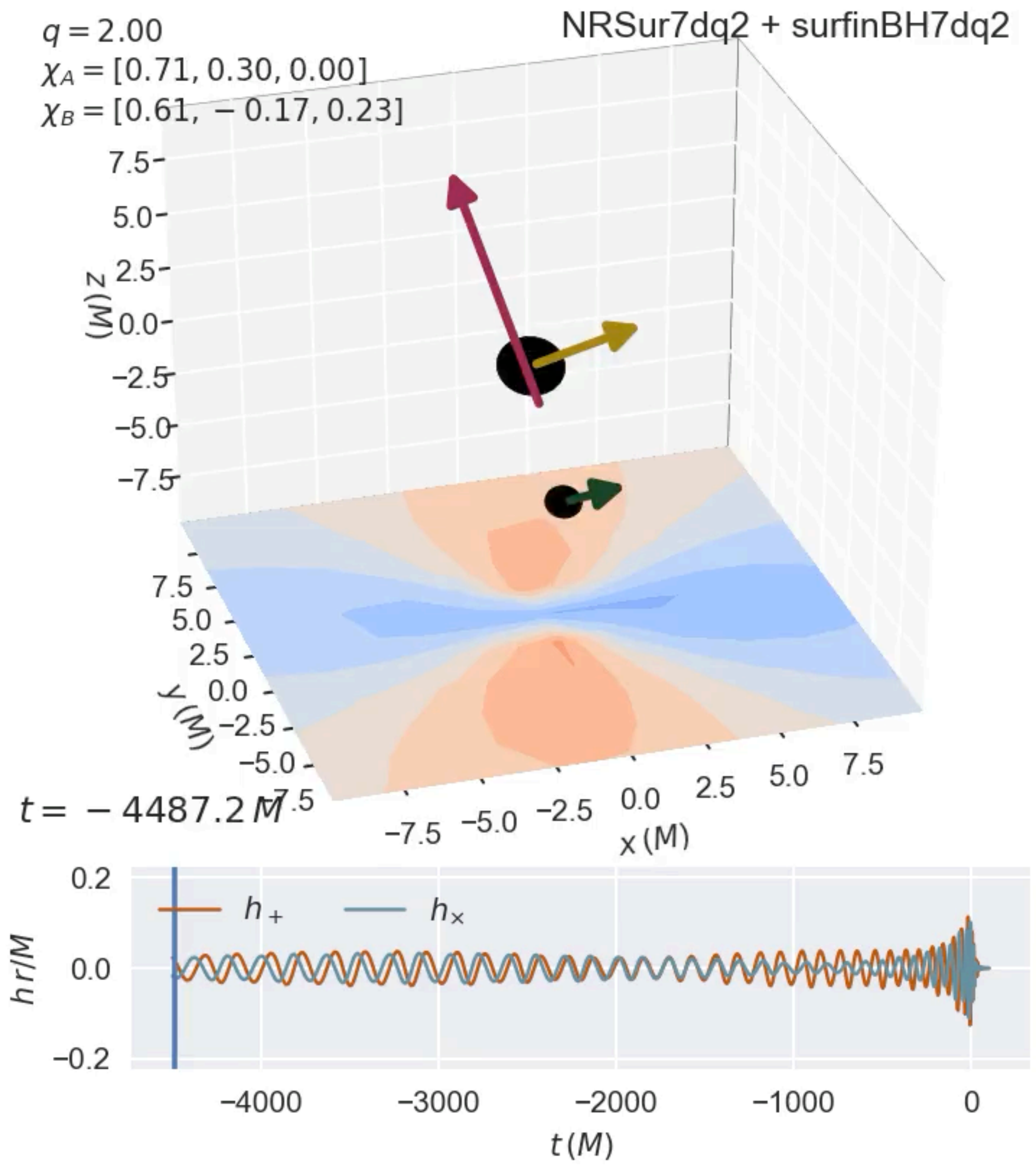
Spins aligned with orbital angular momentum ⇒ Longer Signal

Impact of intrinsic parameters

Spins $\overrightarrow{\chi_1}, \overrightarrow{\chi_2}$:

Precession $\Rightarrow$ Amplitude modulation

Credit : Vijay Varma et al,
Binary Black Hole Explorer



Bayesian Inference : Likelihood, evidence and posteriors

Bayes' Theorem

Likelihood

Prior

Posterior distribution

$$p(\vec{\theta}\,|\,d, H) = \frac{\mathscr{L}(d\,|\,\vec{\theta}, H)\pi(\vec{\theta}\,|\,H)}{p(d\,|\,H)}$$

Evidence ($\mathscr{Z}_H$)

Parameters

Data

$$p(d\,|\,H) = \mathscr{Z}_H = \int \mathscr{L}(d\,|\,\vec{\theta}, H)\pi(\vec{\theta}\,|\,H)d\theta$$

The Likelihood

Likelihood is the probability of obtaining data d give the parameters $\vec{\theta}$

Residual

$$\mathscr{L}\left(d(f_i)\,|\,\vec{\theta}\right) = \frac{2}{T\pi S_n(f_i)} \exp\left(-\frac{2\,|\tilde{d}(f_i) - \tilde{h}(f_i\,;\vec{\theta})|^2}{TS_n(f_i)}\right)$$

$$\mathscr{L}\left(d\,|\,\vec{\theta}\right) = \prod_i \mathscr{L}\left(d(f_i)\,|\,\vec{\theta}\right)$$

The residual after subtracting the best-fit signal from the data should look like noise. It should follow Gaussian distribution.

The Priors

A distribution that encodes the initial beliefs or prior knowledge about the parameters.

Examples:

1. **Masses** :
    1. Uniform in components masses (m_1, m_2)
    2. Uniform in chirp mass $(\mathscr{M})$ and mass ratio (q).
2. **Spins** :
    1. Uniform in spin magnitude $\Rightarrow \chi \in [0,0.99]$
    2. Isotropic prior on spin orientation.

The Priors

A distribution that encodes the initial beliefs or prior knowledge about the parameters.

Examples:

3. **Sky Location** :
    1. Uniform in sky area.
4. **Inclination Angle (θ_{jn})** :
    1. Uniform in $\cos(\theta_{jn})$ over [-1,1]
5. **Luminosity Distance** :
    1. Uniform in volume : $p(d_L) \propto d_L^2$

And so on !

The Evidence ($\mathcal{Z}_H$)

The Evidence is defined as

$$\mathcal{Z}_H = \int \mathcal{L}(d\,|\,\vec{\theta}, H)\ \pi(\vec{\theta}\,|\,H)\ d\theta$$

- The evidence is a single number.
- It is Marginalised likelihood function.
- Useful when comparing one evidence with another evidence.
- Serves as a normalization constant for posterior $p(\vec{\theta}\,|\,d, H)$.
- Mainly used for Model Selection. Which model is statistically preferred by data and by how much?

Bayes Factor

The ratio of the evidence for two different models is called Bayes factor. Assume a model M_A and model M_B. The A/B Bayes factor is

$$BF_B^A = \frac{\mathcal{Z}_A}{\mathcal{Z}_B}$$

Where

$$\mathcal{Z}_A = \int d\theta_1 \, \mathscr{L}(d \,|\, \theta_1, M_A) \, \pi(\theta_1)$$

$$\mathcal{Z}_B = \int d\theta_2 \, \mathscr{L}(d \,|\, \theta_2, M_B) \, \pi(\theta_2)$$

Example - *Signal Model* vs *Noise Model*

There is a BBH/BNS signal present in the data.

There is no signal in data, just noise.

$$\text{Signal Evidence} : \mathcal{Z}_S = \int d\theta \, \mathcal{L}(d\,|\,\theta) \, \pi(\theta)$$

$$\text{Noise Evidence} : \mathcal{Z}_N = \mathcal{L}(d\,|\,0) \quad \text{“Null Likelihood”}$$

$$\text{Bayes Factor} : \log BF^A_B = \log(\mathcal{Z}_S) - \log(\mathcal{Z}_N)$$

If $\log BF^A_B$ is large, the Signal model is preferred over the Noise model.

A threshold of $|\log BF| = 8$ is used as the level of “Strong Evidence”.

Marginalizing the Posterior

- 15+ parameters. What if we want to study a subset of parameters?
- **Marginalization** : Integrating out the unwanted parameters ("*nuisance parameters*").
- The marginalized posterior

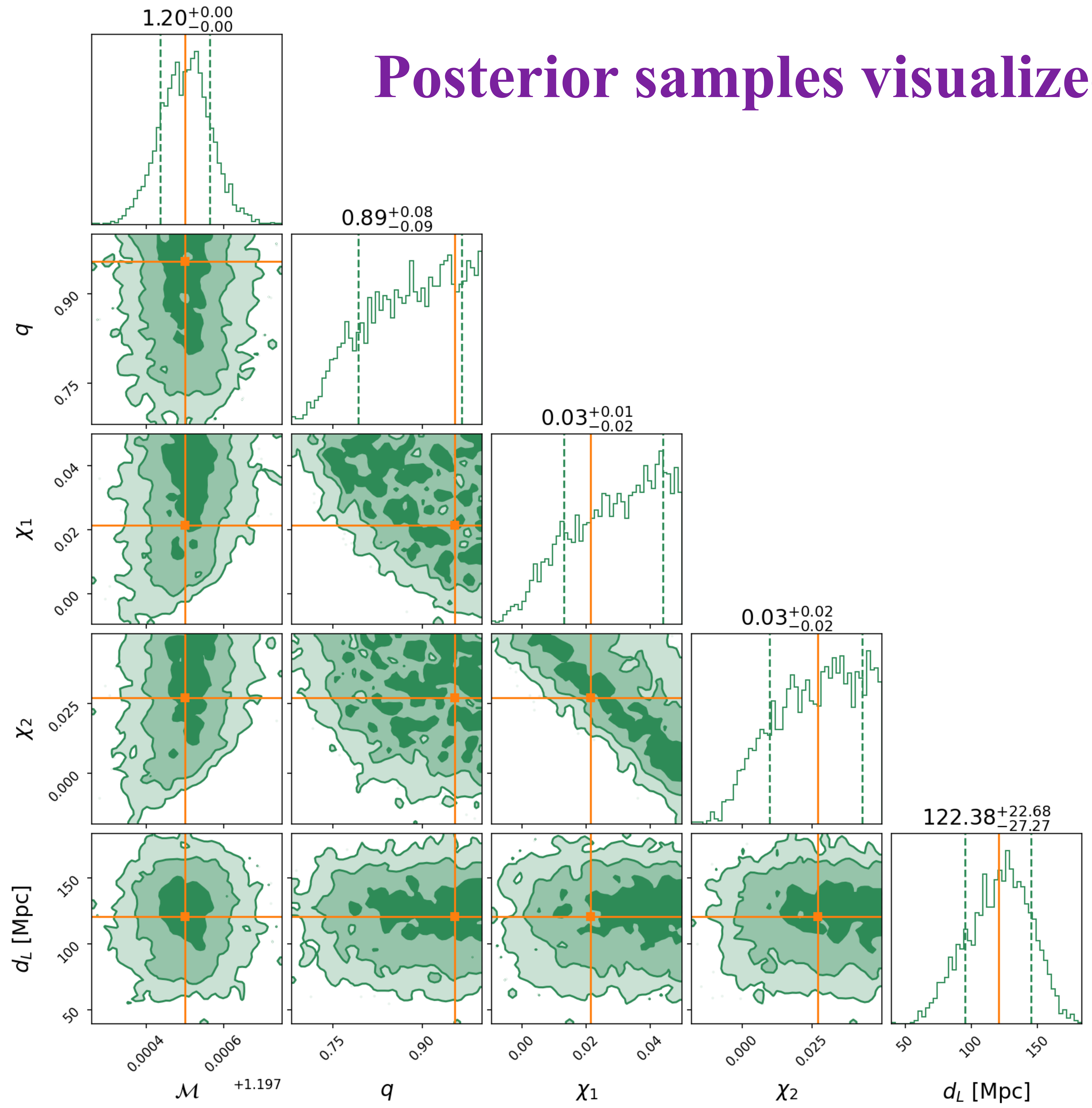
$$p(\theta_i | d) = \int \left(\prod_{k \neq i} d\theta_k \right) p(\theta | d) = \frac{\mathscr{L}(d | \theta_i)\pi(\theta_i)}{\mathscr{Z}}$$

- The quantity $\mathscr{L}(d | \theta_i)$ is called "marginalized likelihood".

$$\mathscr{L}(d | \theta_i) = \int \left(\prod_{k \neq i} d\theta_k \right) \pi(\theta_k)\mathscr{L}(d | \theta)$$

- Marginalized distributions appear as 1D histograms in a corner plot.

Posterior samples visualized using a Corner plot.



- Diagonal : Marginalized 1D histograms of individual parameters.
- Off-diagonal : 2D joint-posterior distributions with contours corresponding to the 68% and 95% credible region.

Example : Is this a fair coin?

Coin Tossing : Inferring Coin Fairness

A small experiment gives me 4 heads in 11 independent coin tosses.

1. Is this a fair coin? How sure are we?
2. If not fair :
    1. How biased is the coin?
    2. How confident are we in the estimate?

Coin Tossing : Inferring Coin Fairness

Let H denote the probability of getting head such that

$$H = 0 \Rightarrow \text{A coin that produces only TAIL.}$$

$$H = 1 \Rightarrow \text{A coin that produces only HEAD.}$$

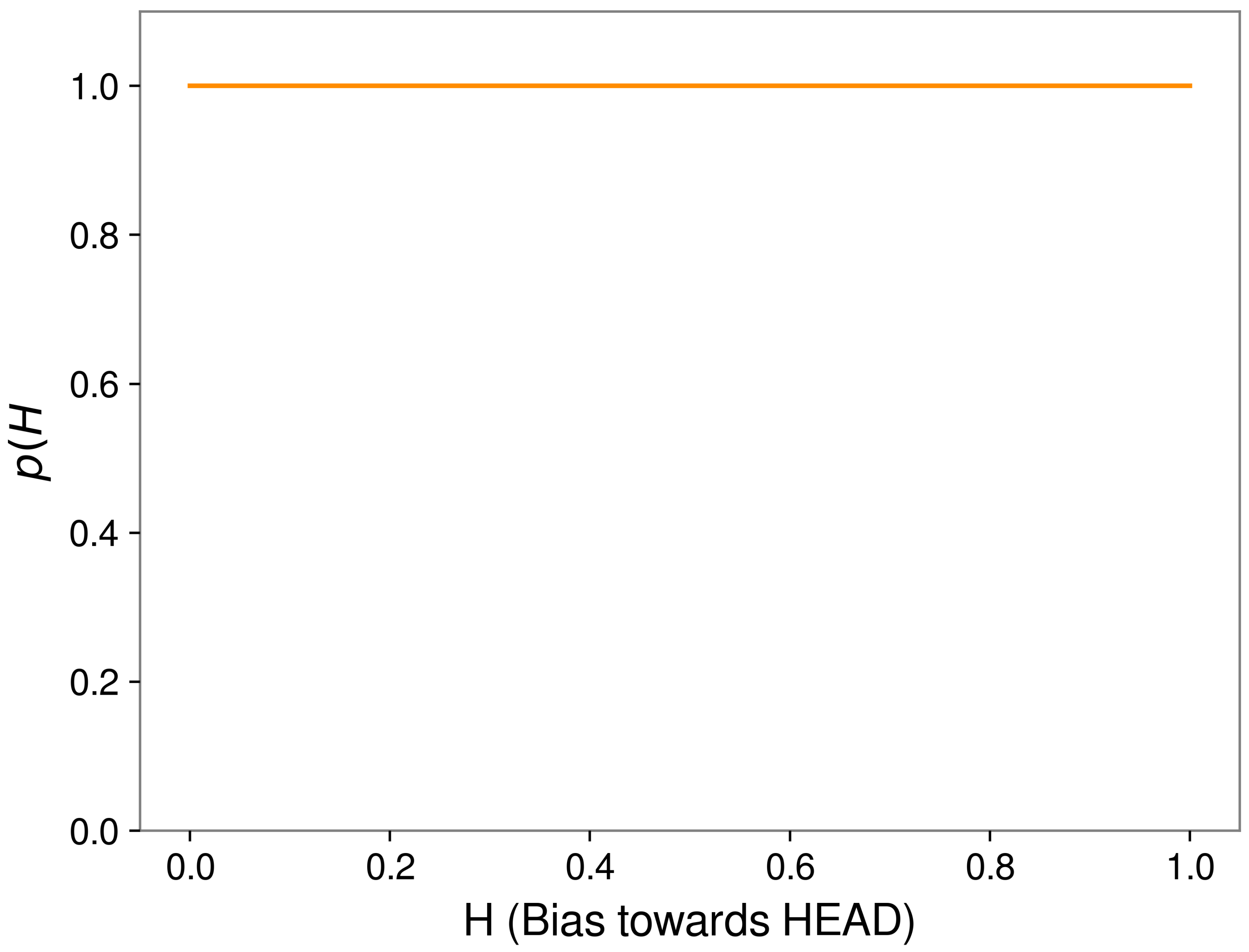
$$H = \frac{1}{2} \Rightarrow \text{A FAIR coin}$$

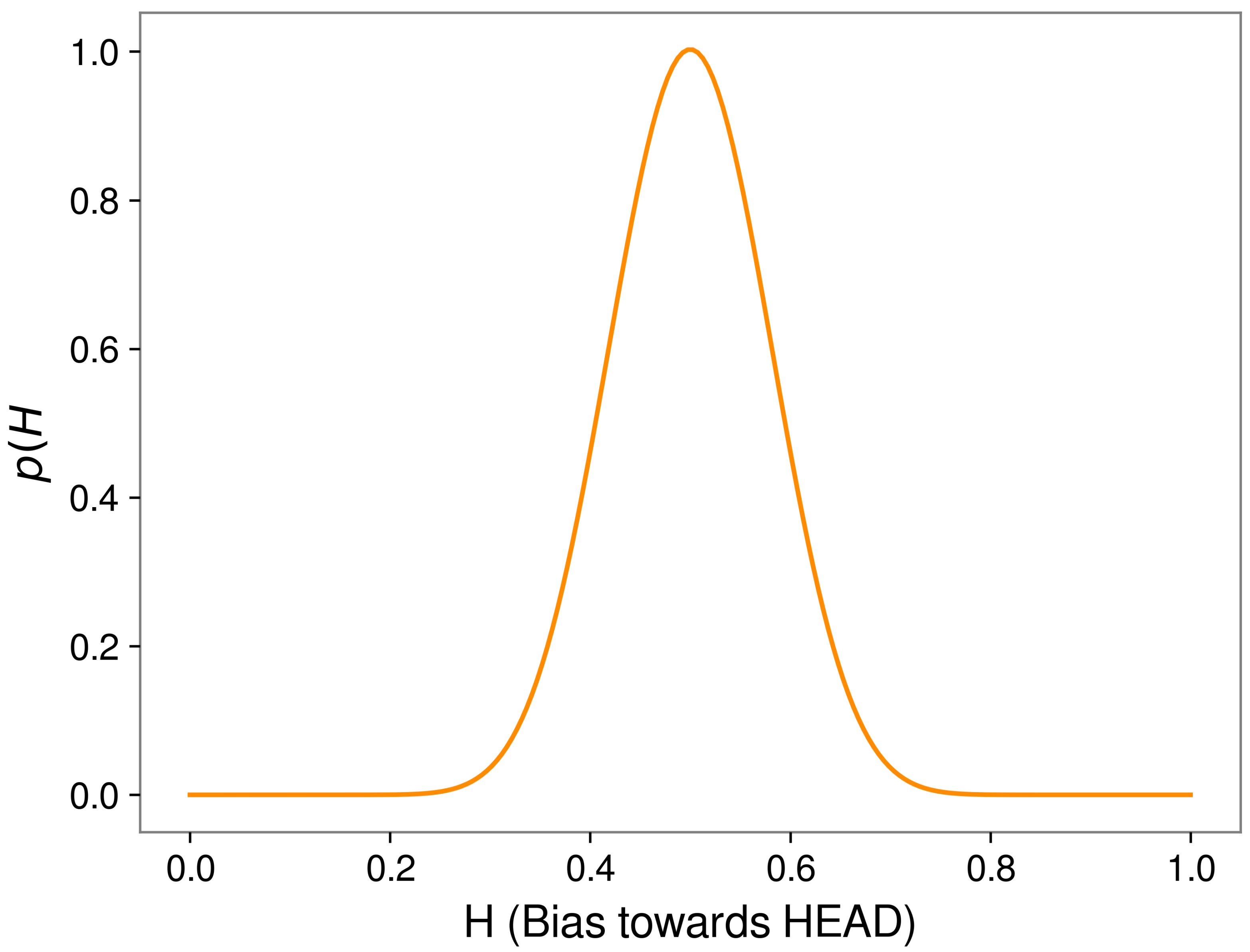
Our goal : To compute the conditional pdf - $p(H|d)$

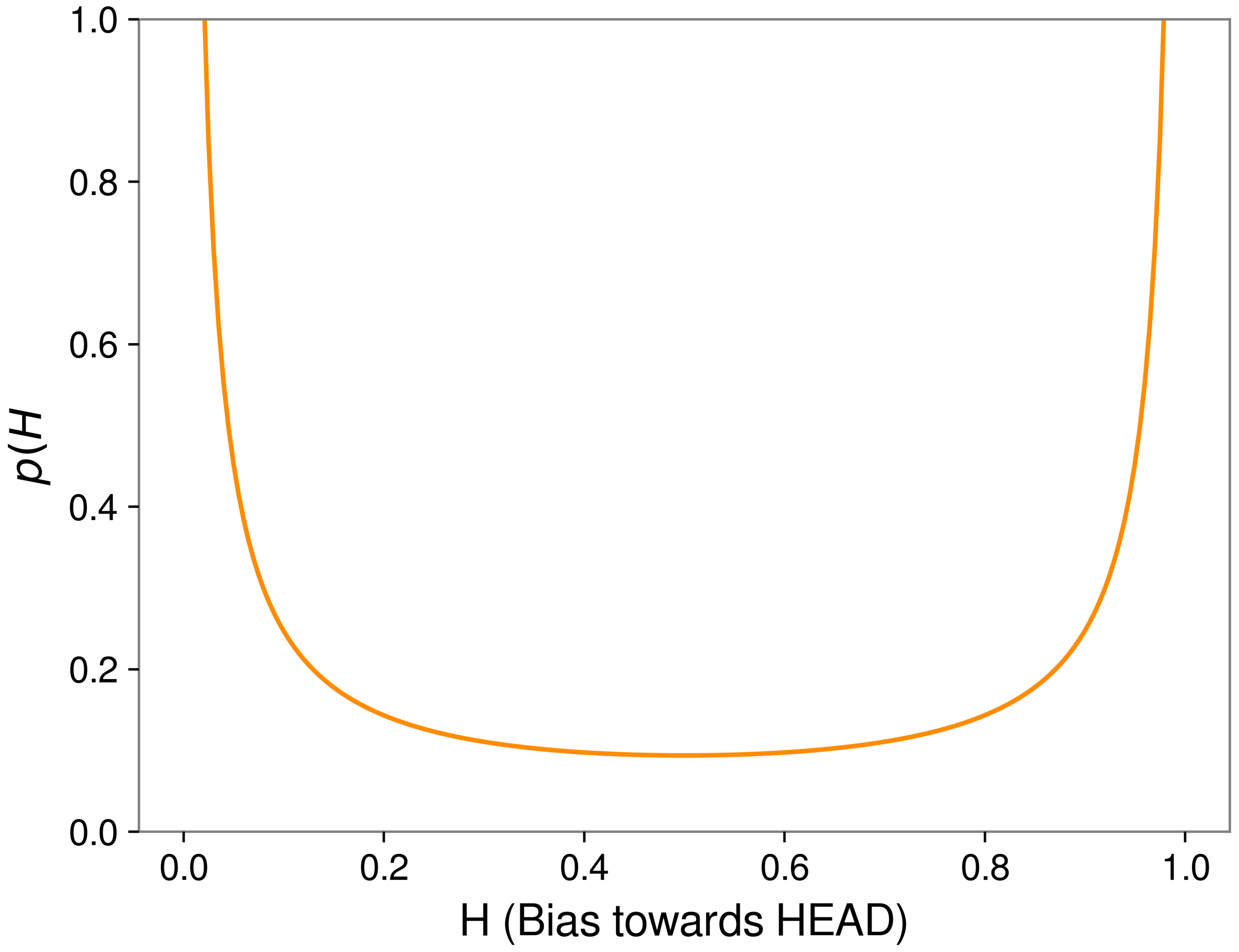
Using Bayes' theorem : $p(H|d) \propto p(d|H)\, p(H)$

Coin Tossing : Inferring Coin Fairness

Priors :

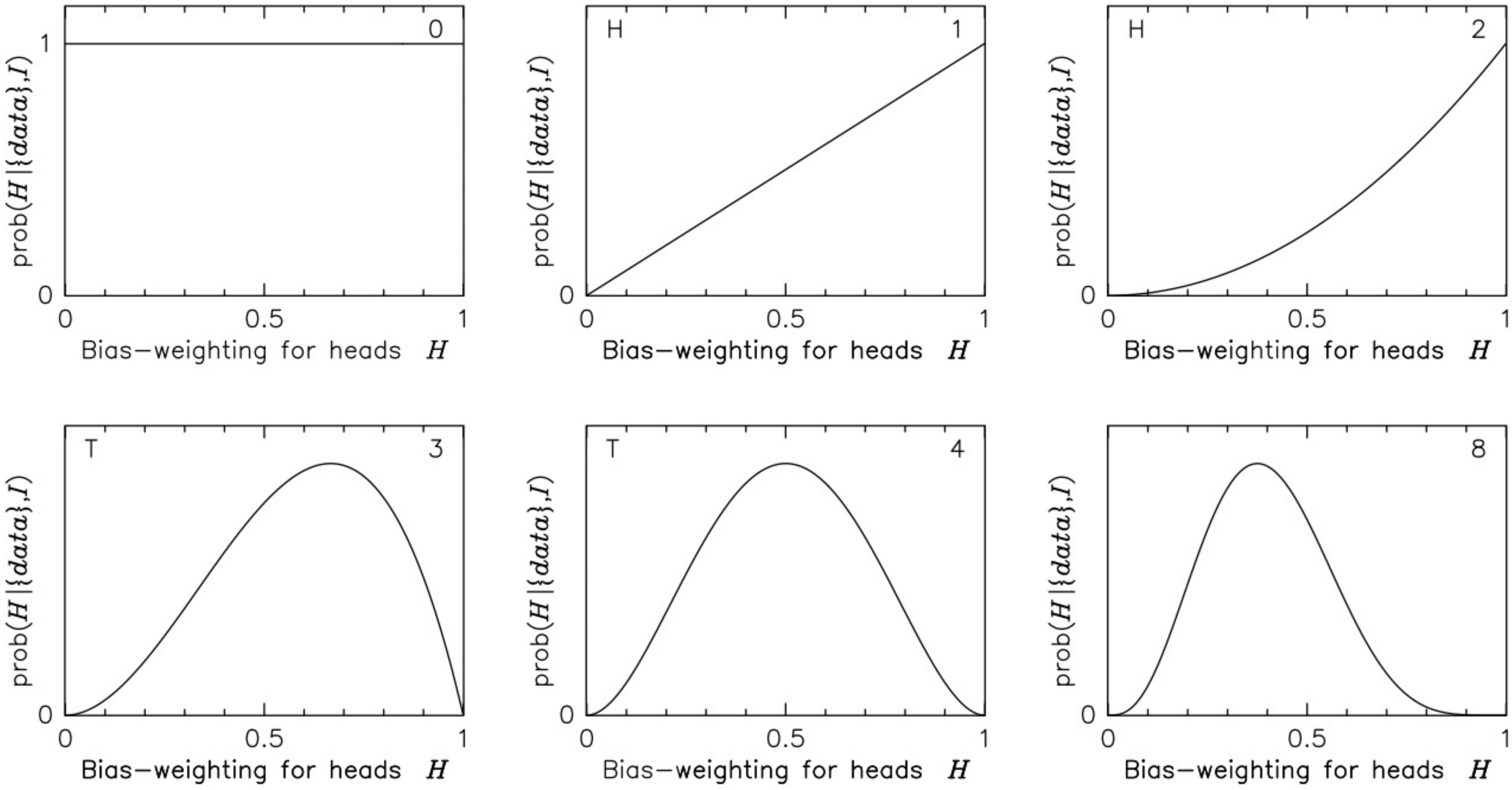






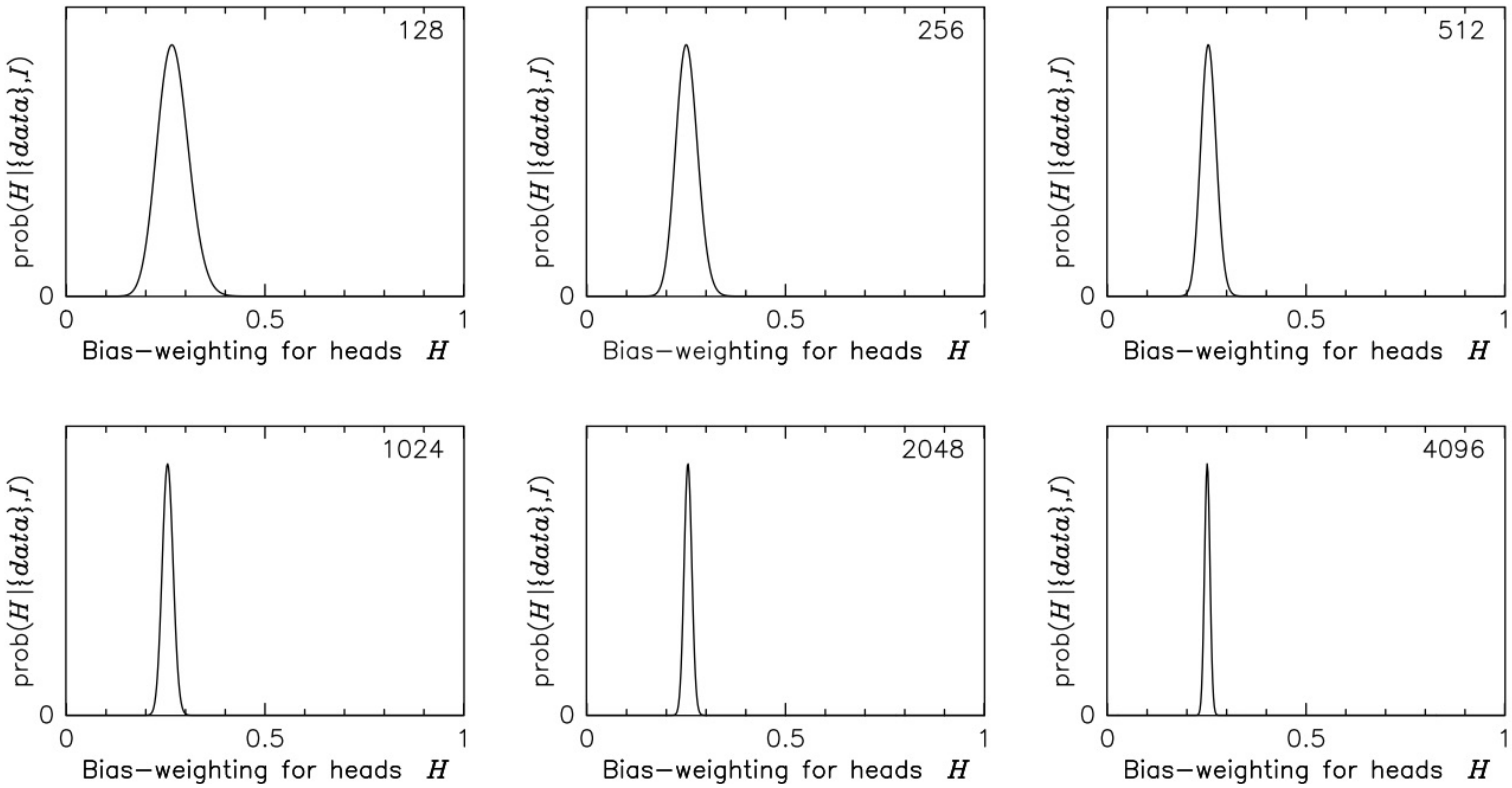
Likelihood : $p(d|H) \propto H^{N_H}(1-H)^{N-N_H}$

Coin Tossing : Inferring Coin Fairness



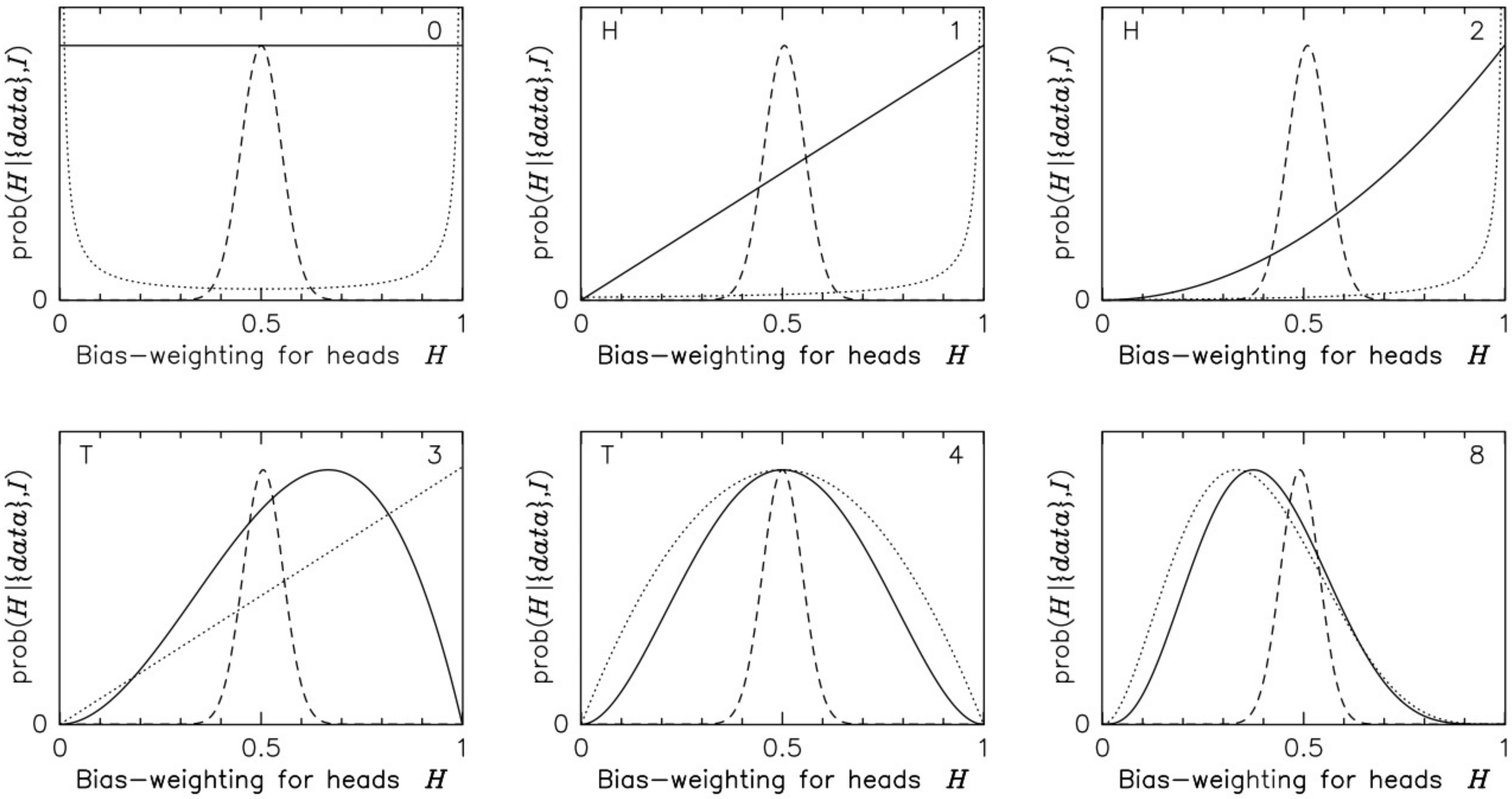
Source : Data Analysis : A Bayesian tutorial, D.S. Sivia with J. Skilling, 2nd edition

Coin Tossing : Inferring Coin Fairness



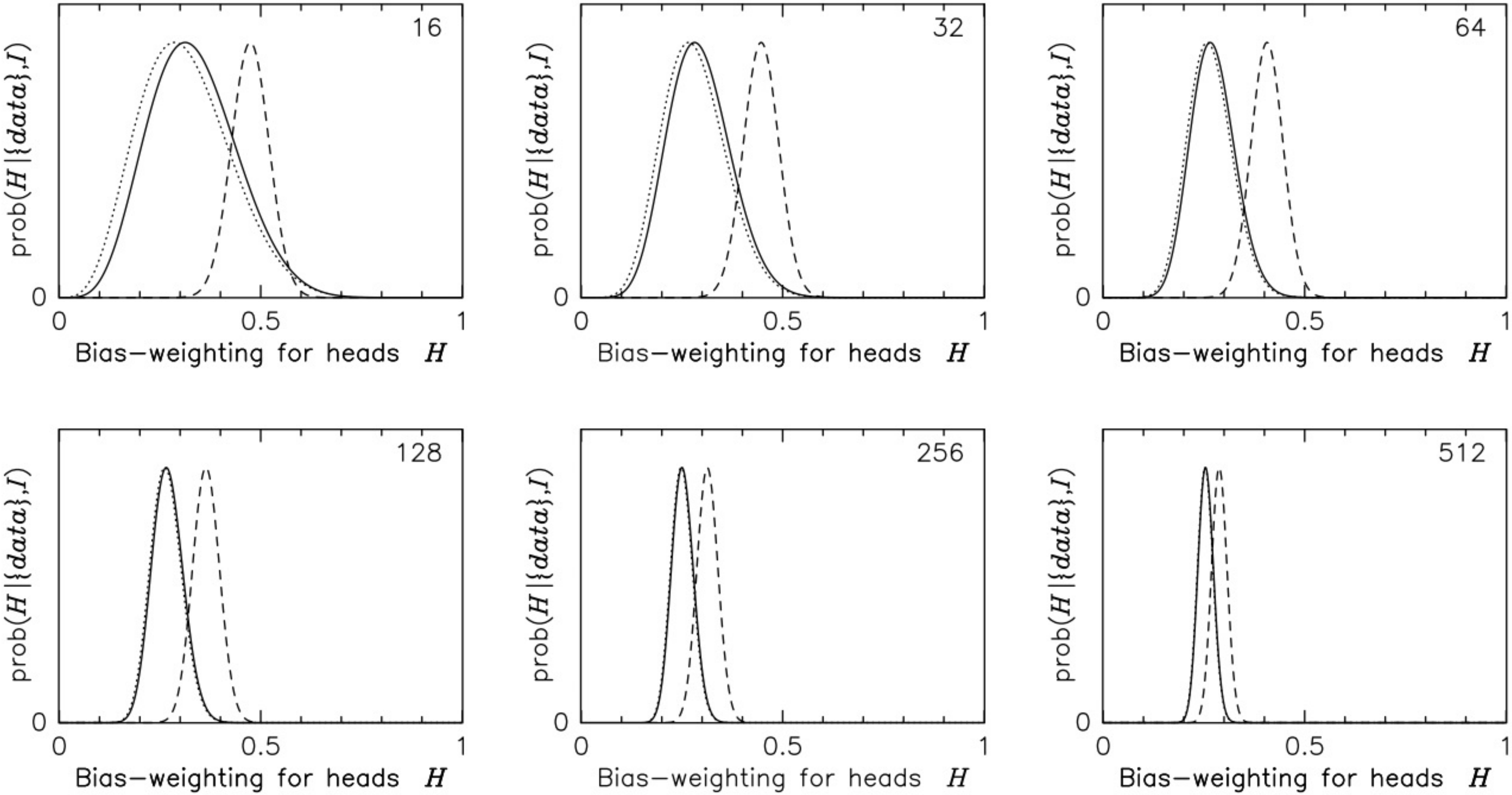
Source : Data Analysis : A Bayesian tutorial, D.S. Sivia with J. Skilling, 2nd edition

Coin Tossing : Inferring Coin Fairness



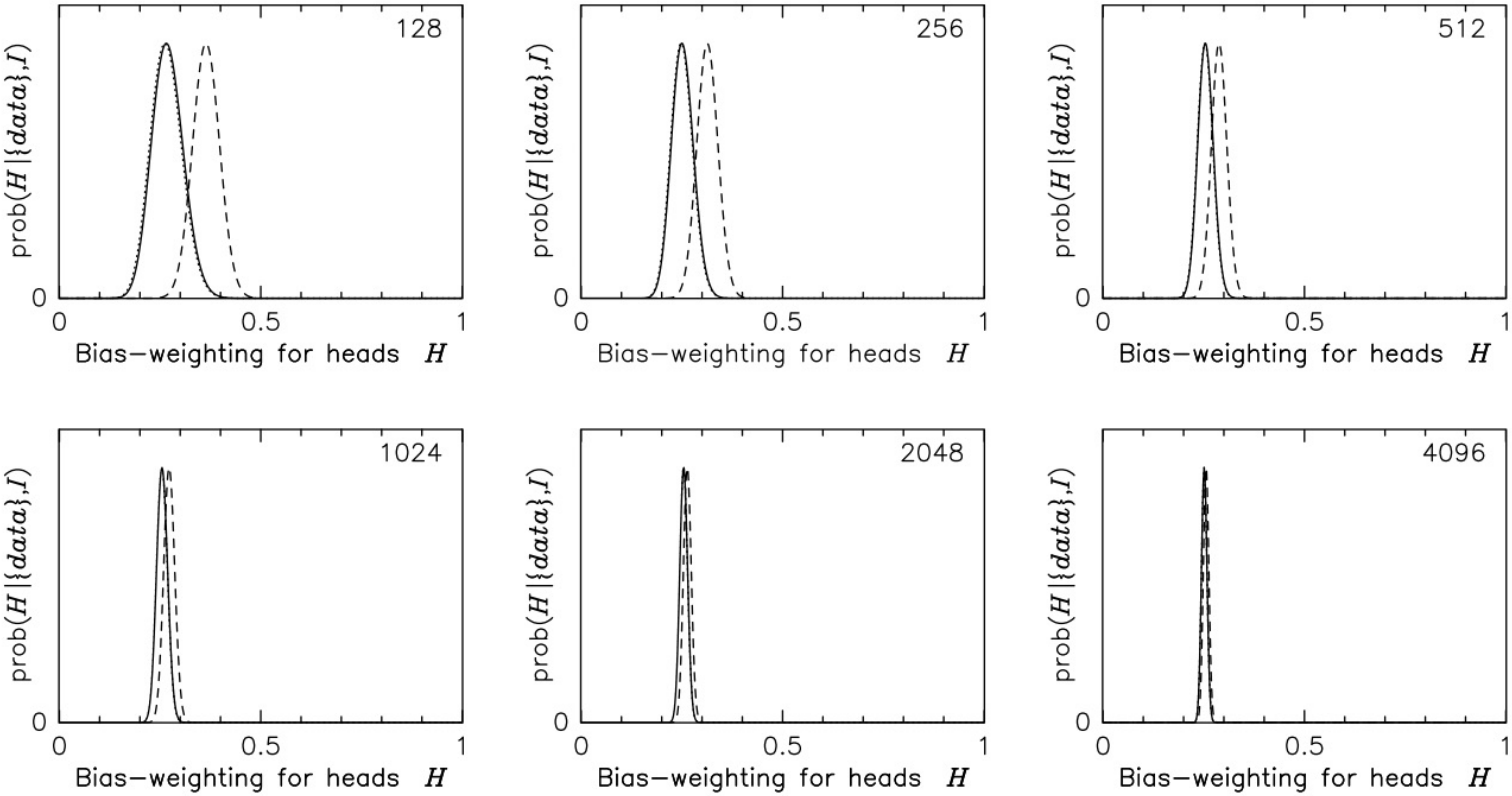
Source : Data Analysis : A Bayesian tutorial, D.S. Sivia with J. Skilling, 2nd edition

Coin Tossing : Inferring Coin Fairness



Source : Data Analysis : A Bayesian tutorial, D.S. Sivia with J. Skilling, 2nd edition

Coin Tossing : Inferring Coin Fairness

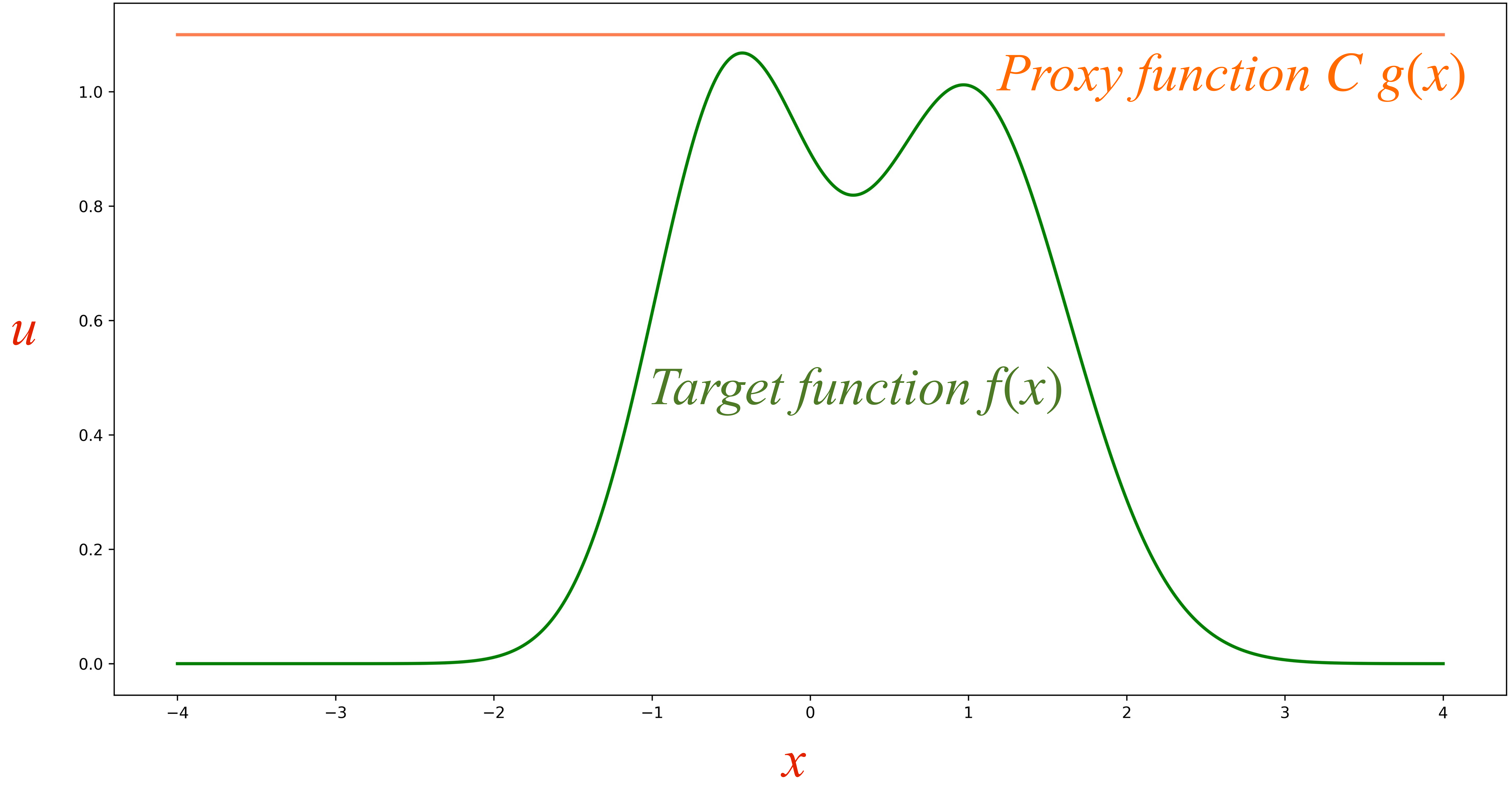


Source : Data Analysis : A Bayesian tutorial, D.S. Sivia with J. Skilling, 2nd edition

Sampling

Rejection Sampling

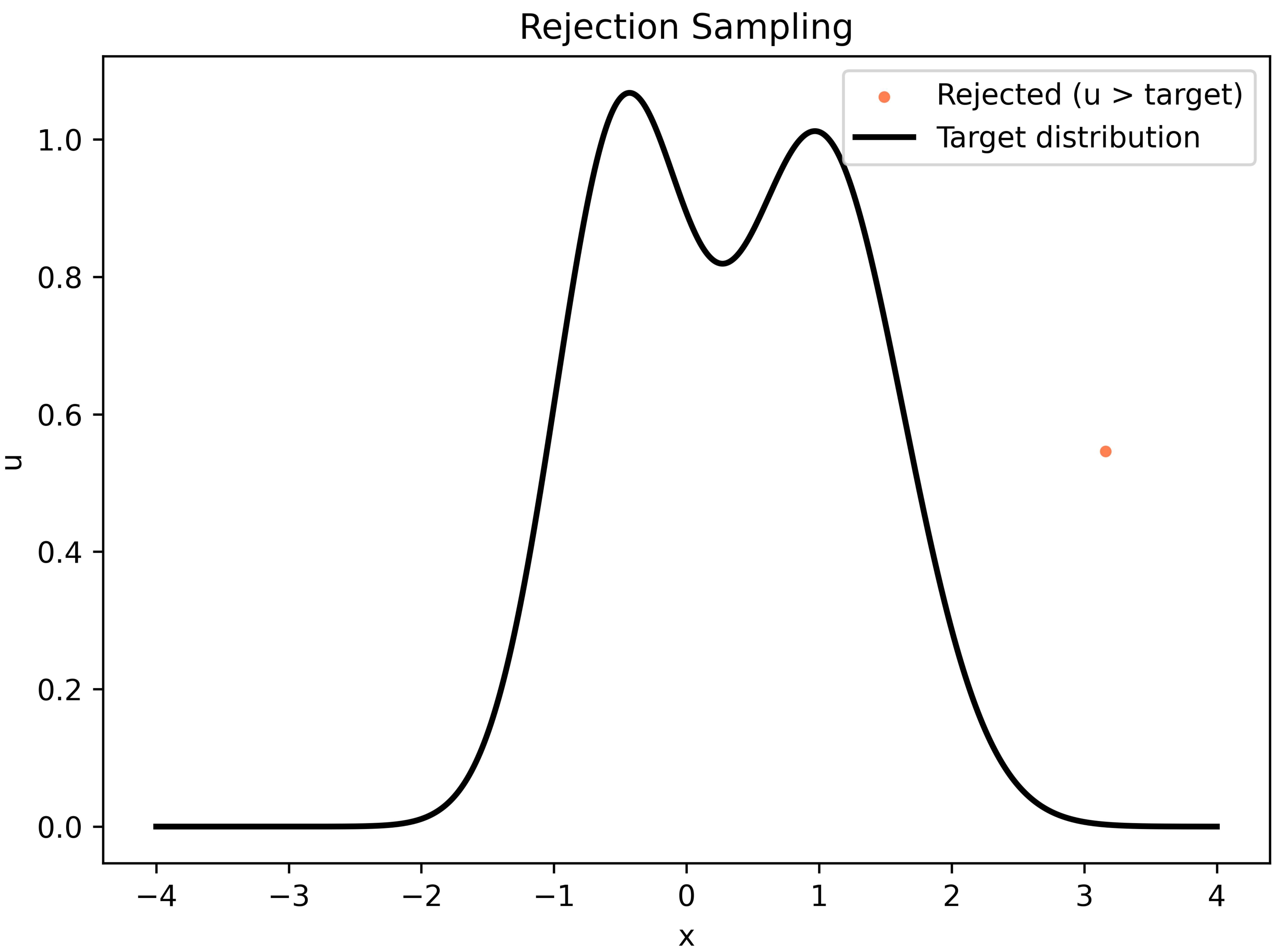
Sample a target distribution that is difficult to sample directly, with the help of a proxy distribution.



Rejection Sampling : Algorithm

One sample - Rejected

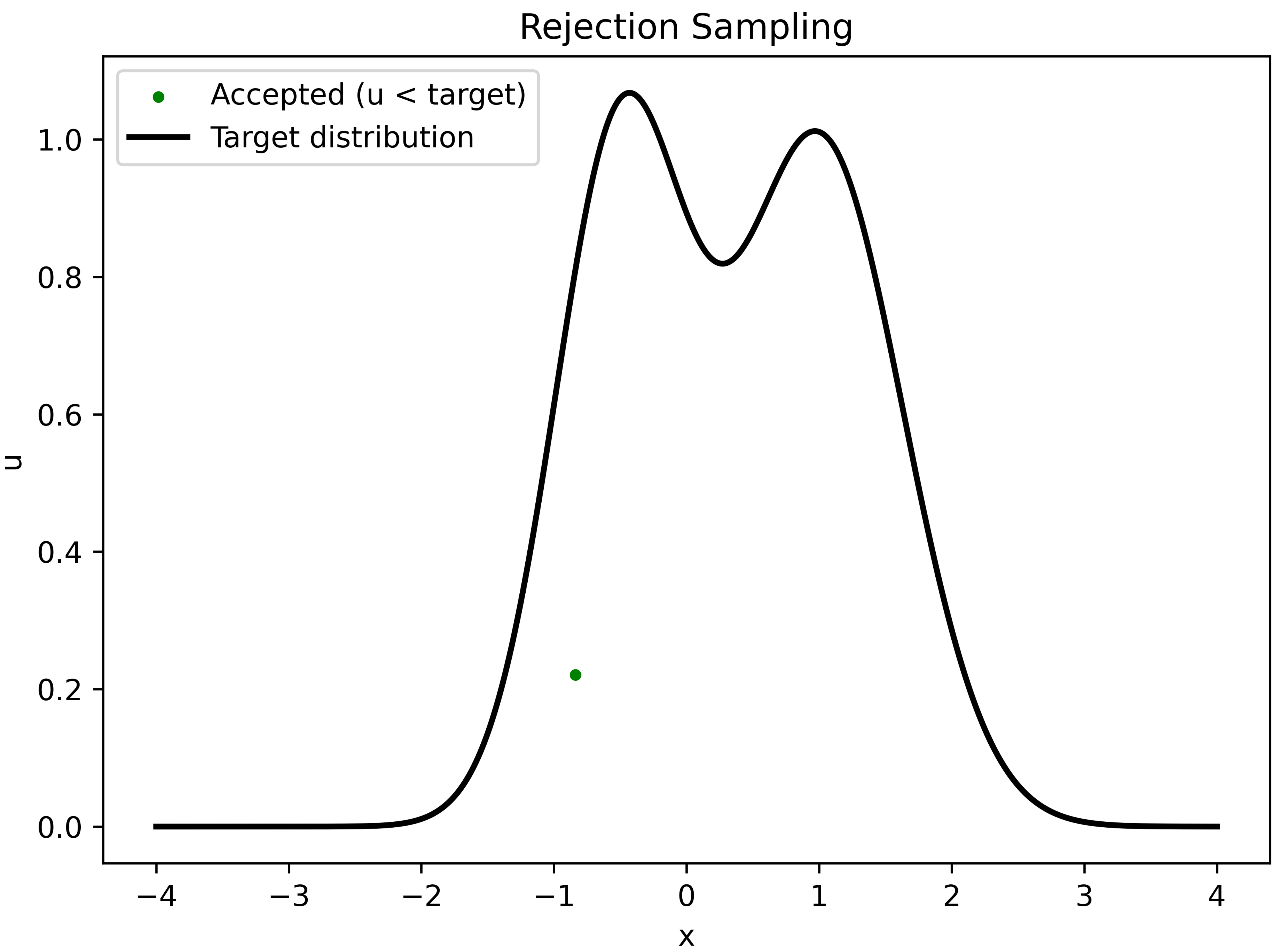
1. Sample x
2. Sample u from $U(0,1)$
3. Check if $u \leq \dfrac{f(x)}{C\,g(x)}$
4. And repeat.



Rejection Sampling : Algorithm

One sample - Accepted

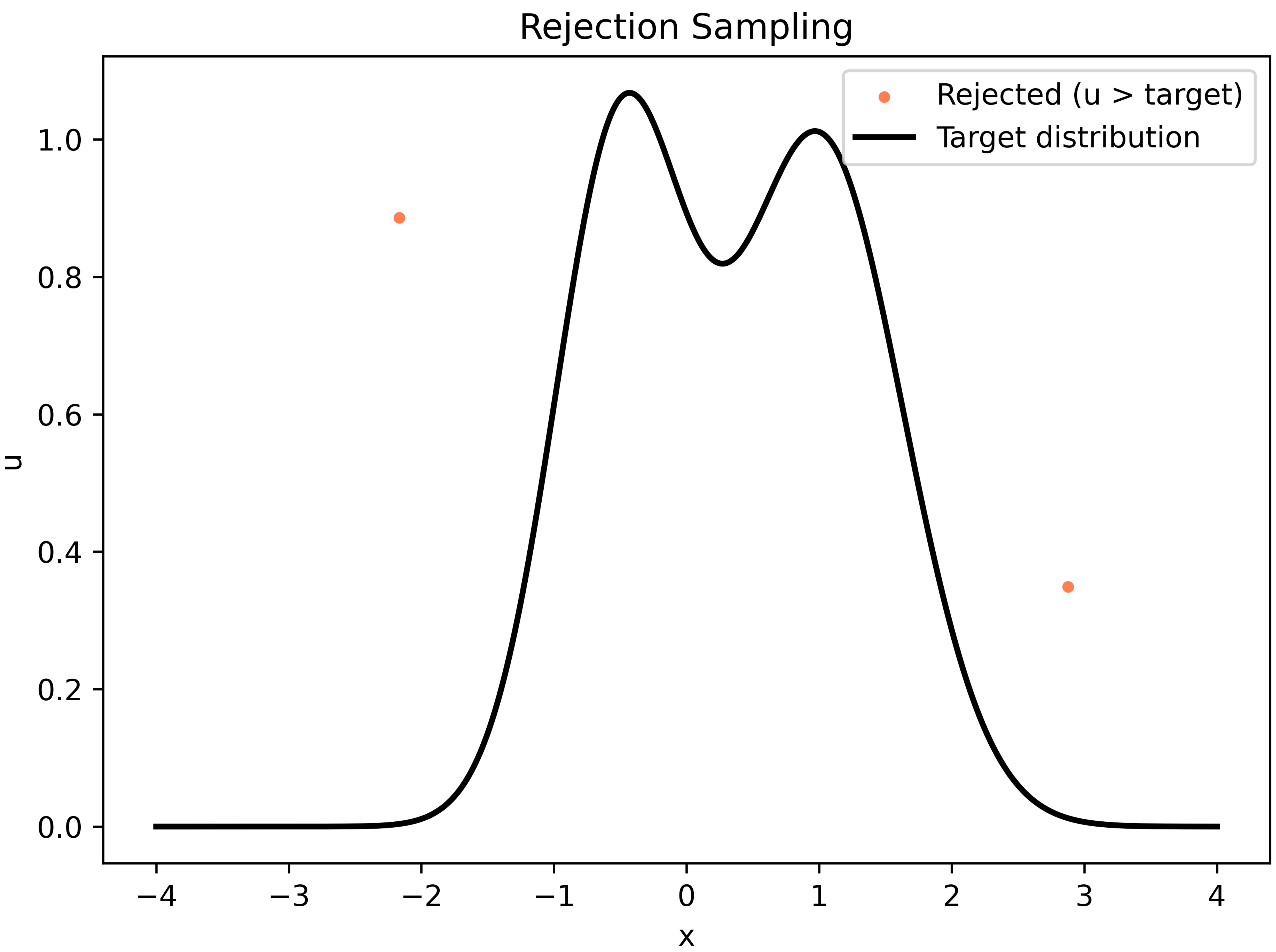
1. Sample x
2. Sample u from $U(0,1)$
3. Check if $u \leq \dfrac{f(x)}{C\,g(x)}$
4. And repeat.



Rejection Sampling : Algorithm

2 samples - Rejected

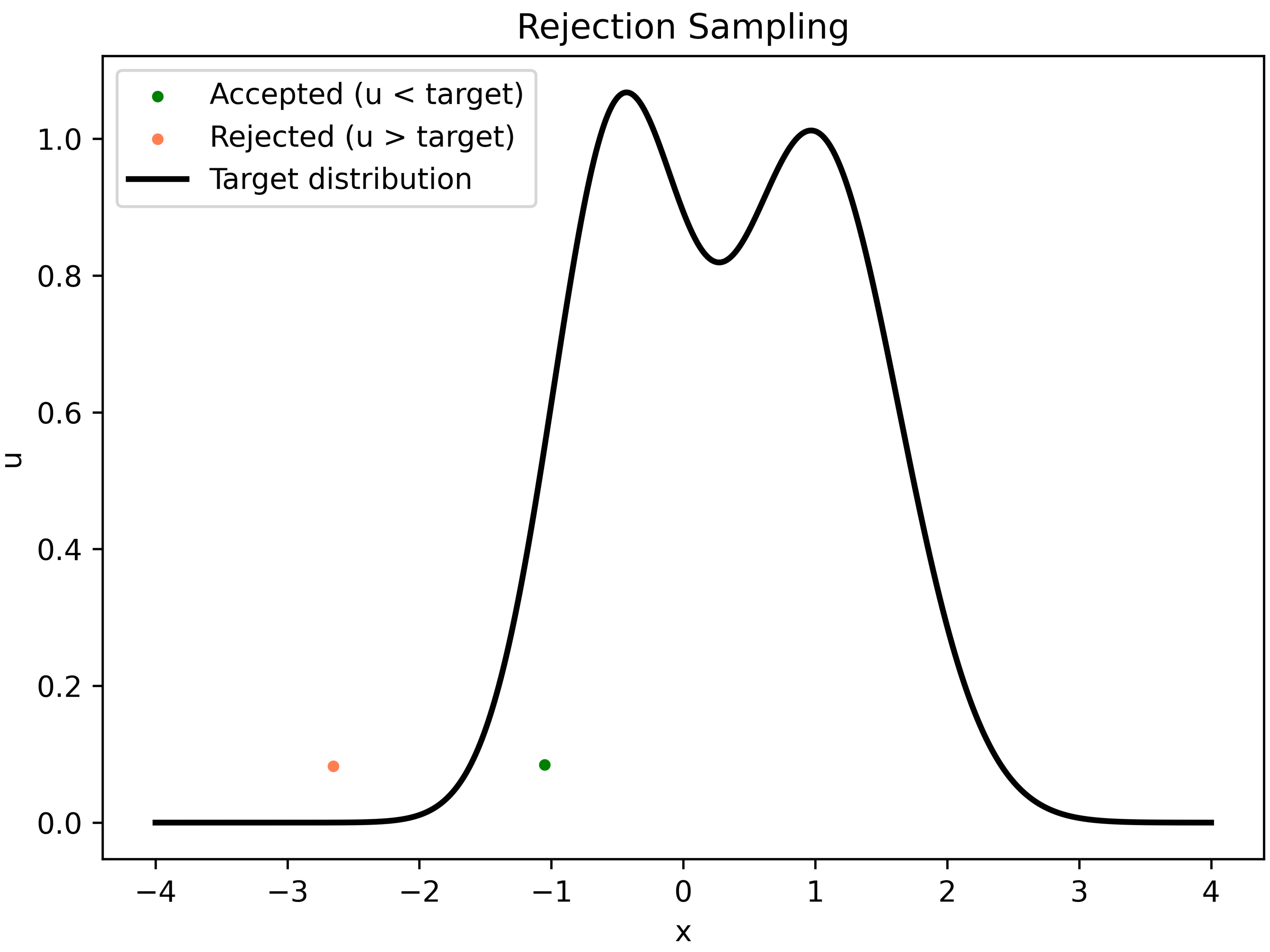
1. Sample x
2. Sample u from $U(0,1)$
3. Check if $u \leq \dfrac{f(x)}{C\, g(x)}$
4. And repeat.



Rejection Sampling : Algorithm

2 samples - 1 Rejected, 1 Accepted

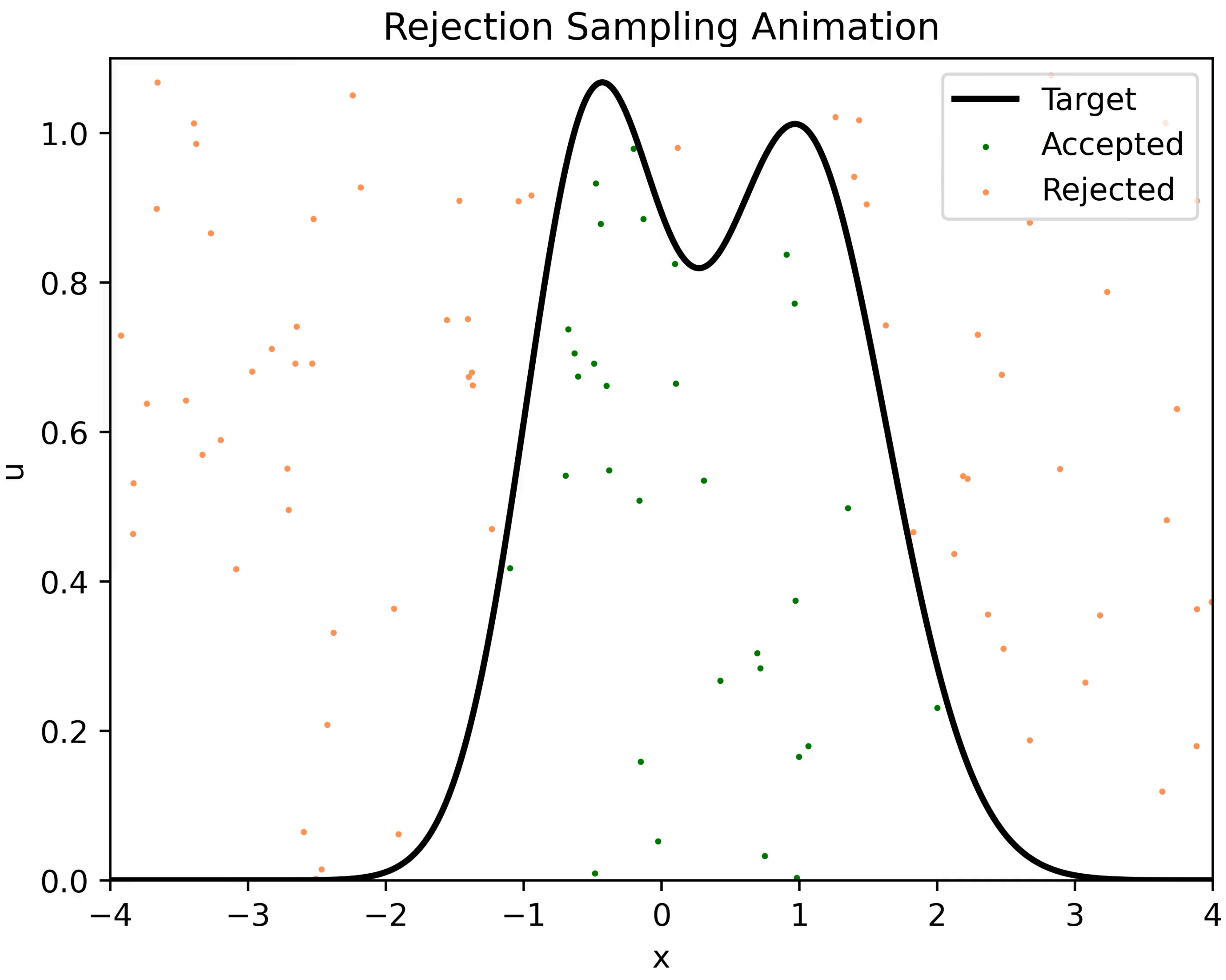
1. Sample x
2. Sample u from $U(0,1)$
3. Check if $u \leq \dfrac{f(x)}{C\, g(x)}$
4. And repeat.



Rejection Sampling : Algorithm

Proxy distribution - Uniform

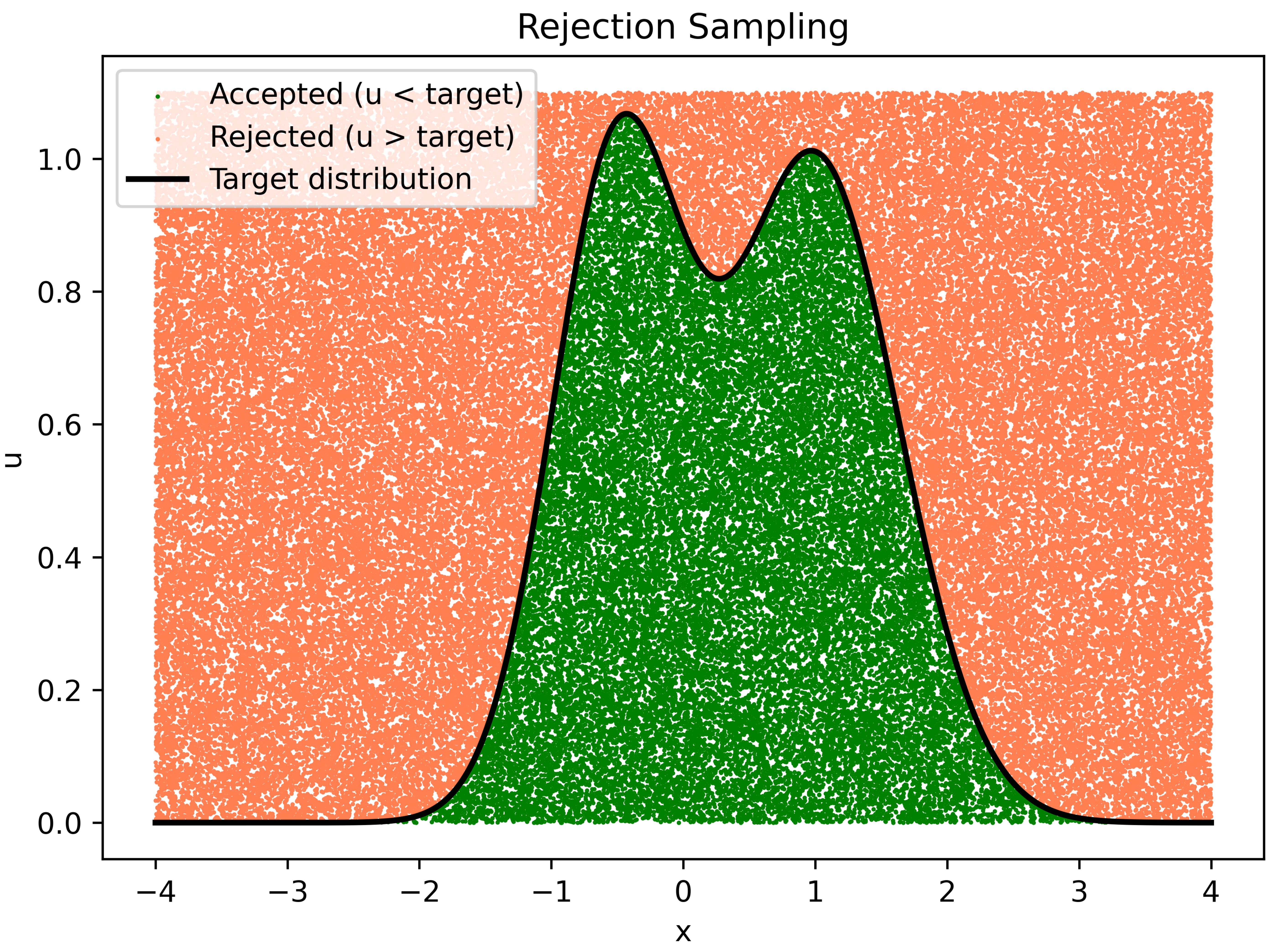
1. Sample x
2. Sample u from $U(0,1)$
3. Check if $u \leq \dfrac{f(x)}{C\,g(x)}$
4. And repeat.



Rejection Sampling : Algorithm

100000 samples

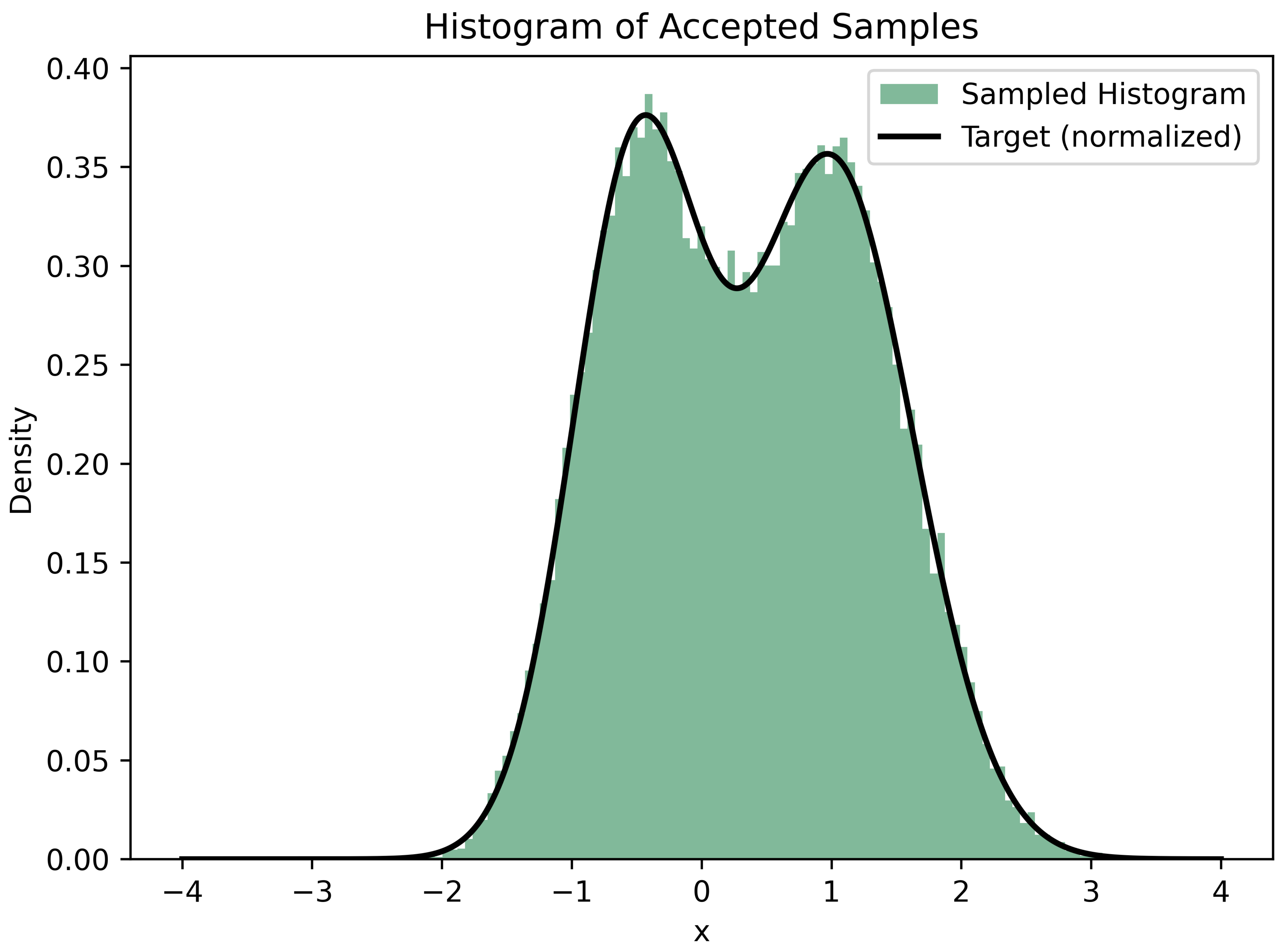
1. Sample x
2. Sample u from $U(0,1)$
3. Check if $u \leq \dfrac{f(x)}{C\, g(x)}$
4. And repeat.



Rejection Sampling : Algorithm

100000 samples

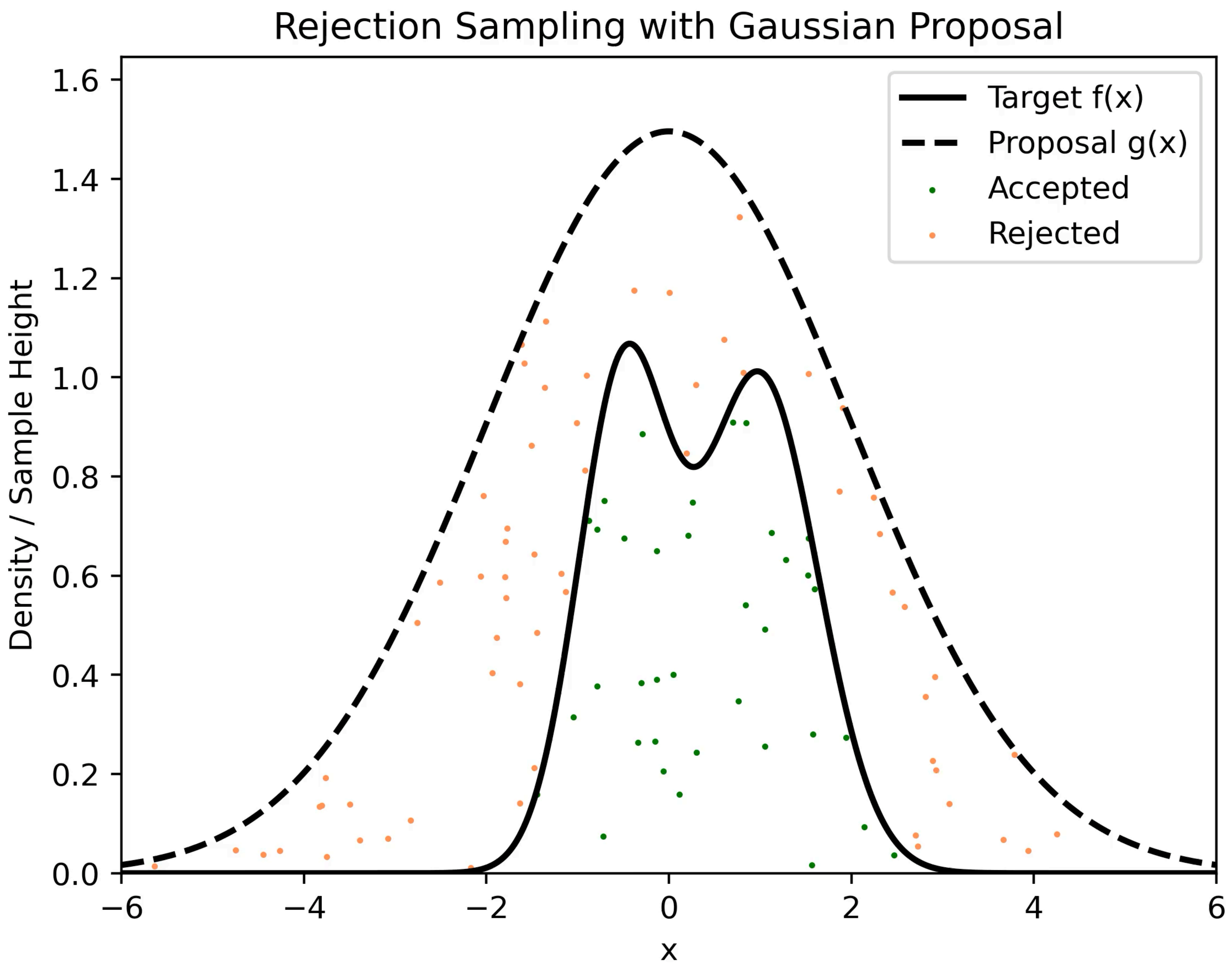
1. Sample x
2. Sample u from $U(0,1)$
3. Check if $u \leq \dfrac{f(x)}{C\,g(x)}$
4. And repeat.



Rejection Sampling : Algorithm

Proxy distribution - Gaussian

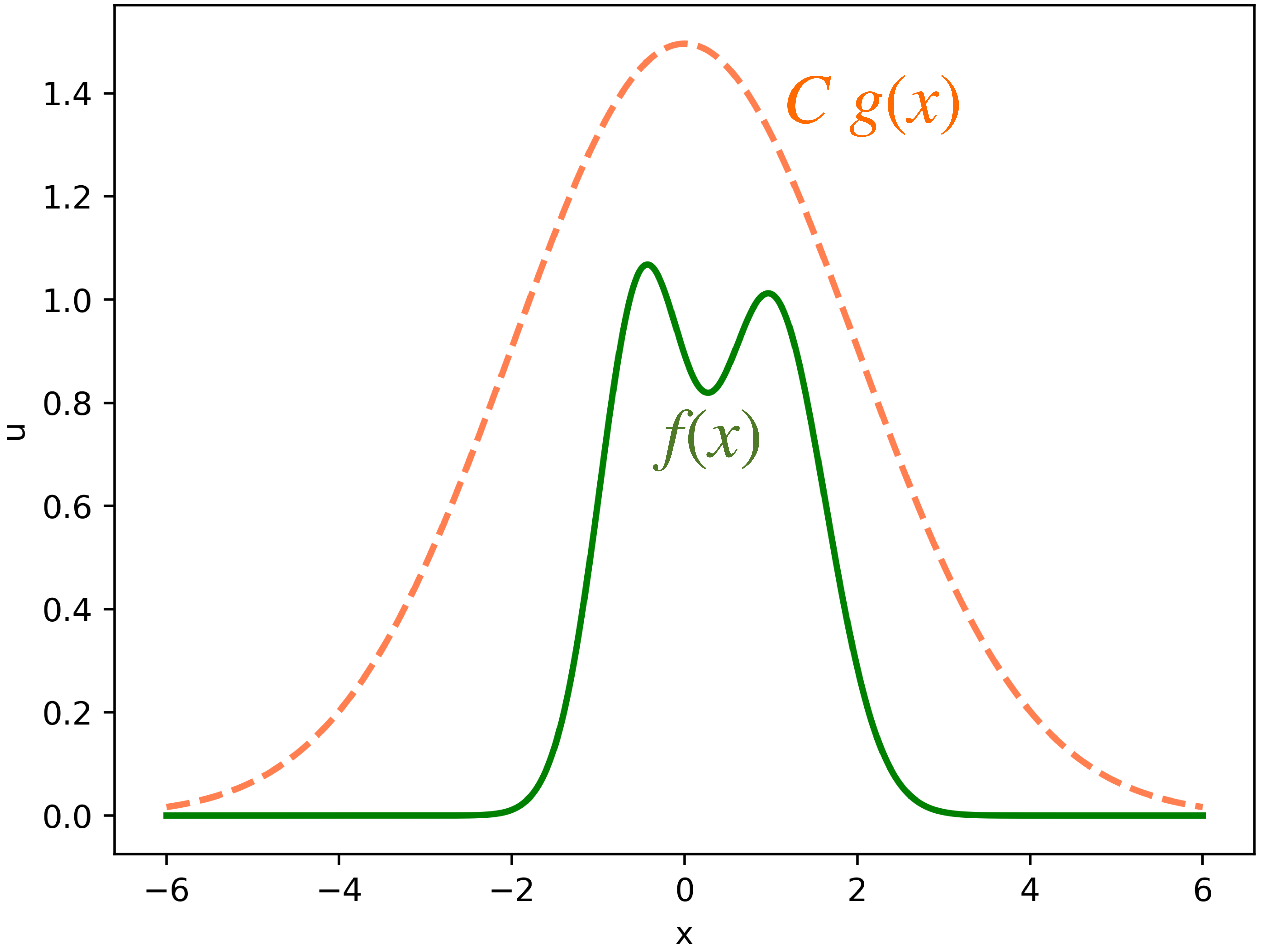
1. Sample x
2. Sample u from $U(0,1)$
3. Check if $u \leq \dfrac{f(x)}{C\,g(x)}$
4. And repeat.



Rejection Sampling : Limitations

1. High Rejection Rates, especially in higher dimensions!

2. Large C (for N dimension) $\Rightarrow$ Very inefficient. Prob of accepting a sample is low for large C.

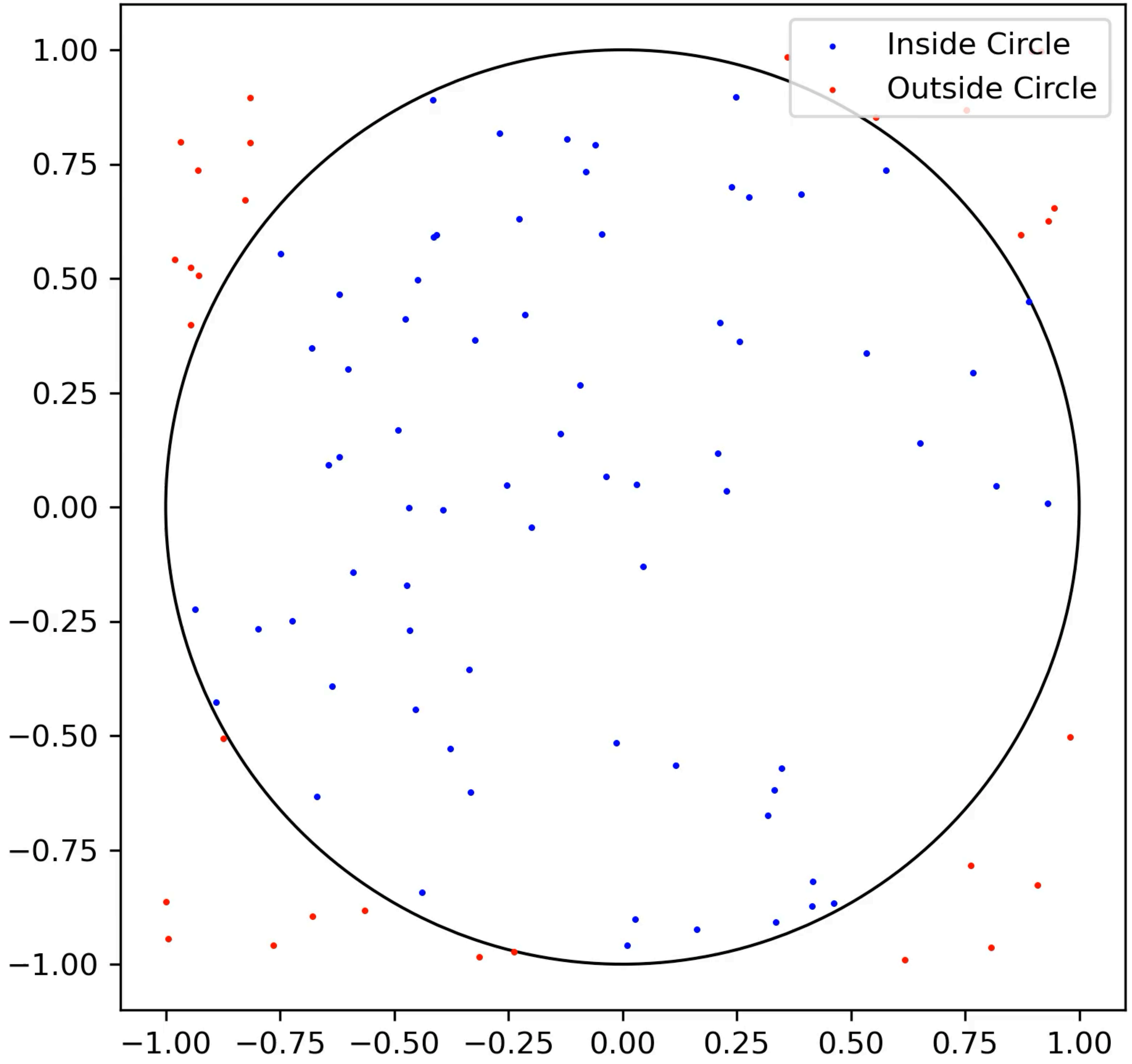


MCMC : Markov Chain Monte Carlo

A family of algorithms that uses Markov chains to perform Monte-Carlo estimate.

Monte Carlo Sampling

Uses randomness to draw samples.



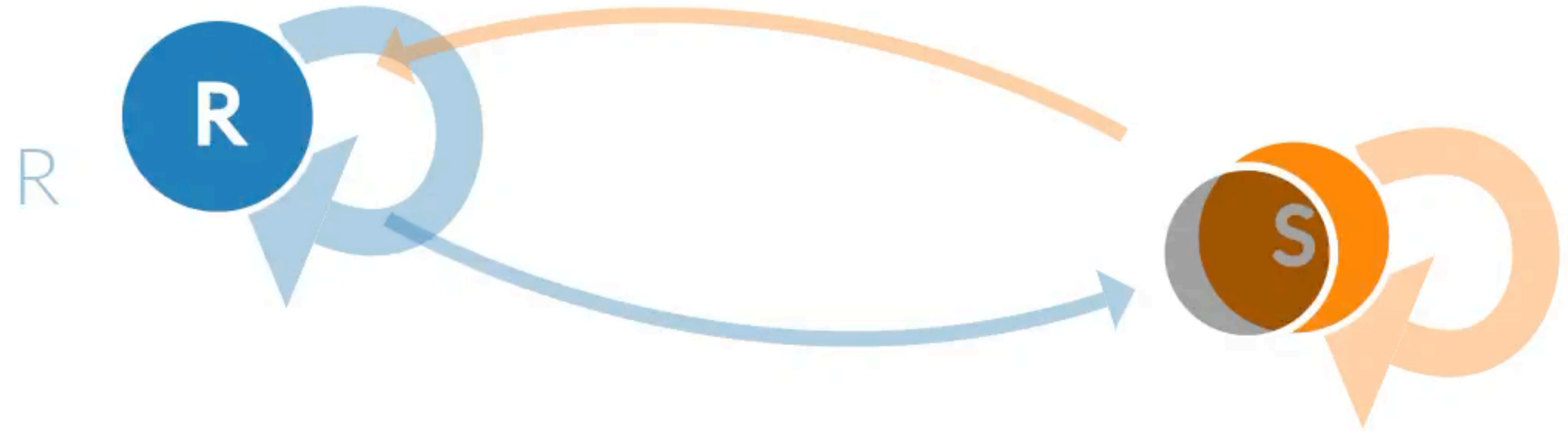
Markov Chain

Any process that exhibits **Markov property** is **Markov chain.**

Markov Property : The probability of jumping from one state to other depends on current state, not on the previous states.

Transition Matrix

	R	S
R	P(R|R) = 0.9	P(S|R) = 0.1
S	P(R|S) = 0.1	P(S|S) = 0.9



Credit : Victor Powell, Lewis Lehe

Source:https://setosa.io/ev/markov-chains/

Why use Markov Chain?

$\Rightarrow$ because of **Stationary Distribution of Markov Chains**

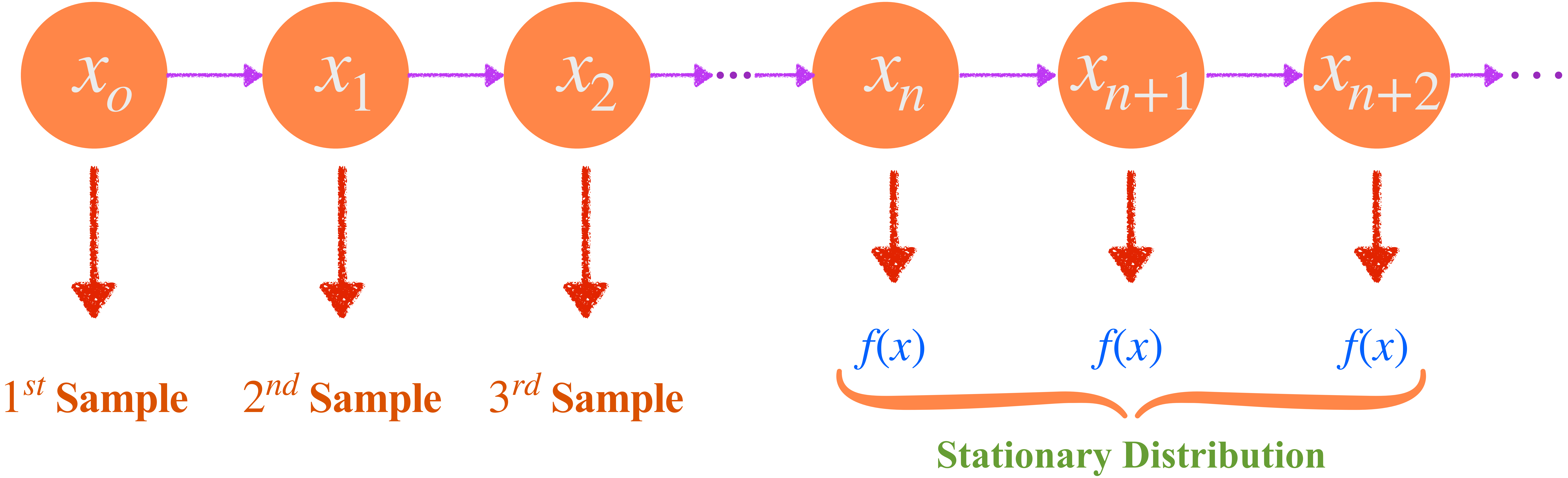
A stationary distribution is a probability distribution that remains unchanged in the Markov chain as time progresses.

Condition :

$$p\,T = p$$

Where p is probability distribution and T is transition matrix.

MCMC : Markov Chain Monte Carlo



Markov Chain is designed such that the Stationary distribution is same as the Target distribution. $\Rightarrow p(x_o)T(x_1 \mid x_o) = p(x_1)T(x_o \mid x_1)$

MCMC : Metropolis - Hastings algorithm

STEP 1 : Trial sample $x_t = x_n + \delta$.

MCMC : Metropolis - Hastings algorithm

STEP 1 : Trial sample $x_t = x_n + \delta$.

STEP 2 : Compute acceptance Probability $P_a = \min\left(\frac{p(x_t)}{p(x_n)}, 1\right)$

MCMC : Metropolis - Hastings algorithm

STEP 1 : Trial sample $x_t = x_n + \delta$.

STEP 2 : Compute acceptance Probability $P_a = \min\left(\dfrac{p(x_t)}{p(x_n)}, 1\right)$

STEP 3 : Draw a number of r from $U(0,1)$. Accept if $r < P_a$, else reject.

MCMC : Metropolis - Hastings algorithm

STEP 1 : Trial sample $x_t = x_n + \delta$.

STEP 2 : Compute acceptance Probability $P_a = \min\left(\frac{p(x_t)}{p(x_n)}, 1\right)$

STEP 3 : Draw a number of r from $U(0,1)$. Accept if $r < P_a$, else reject.

STEP 4 : Next sample, $x_{n+1} = \begin{cases} x_t & \text{if } r < P_a \\ x_n & \text{if } r > P_a \end{cases}$

Nested Sampling

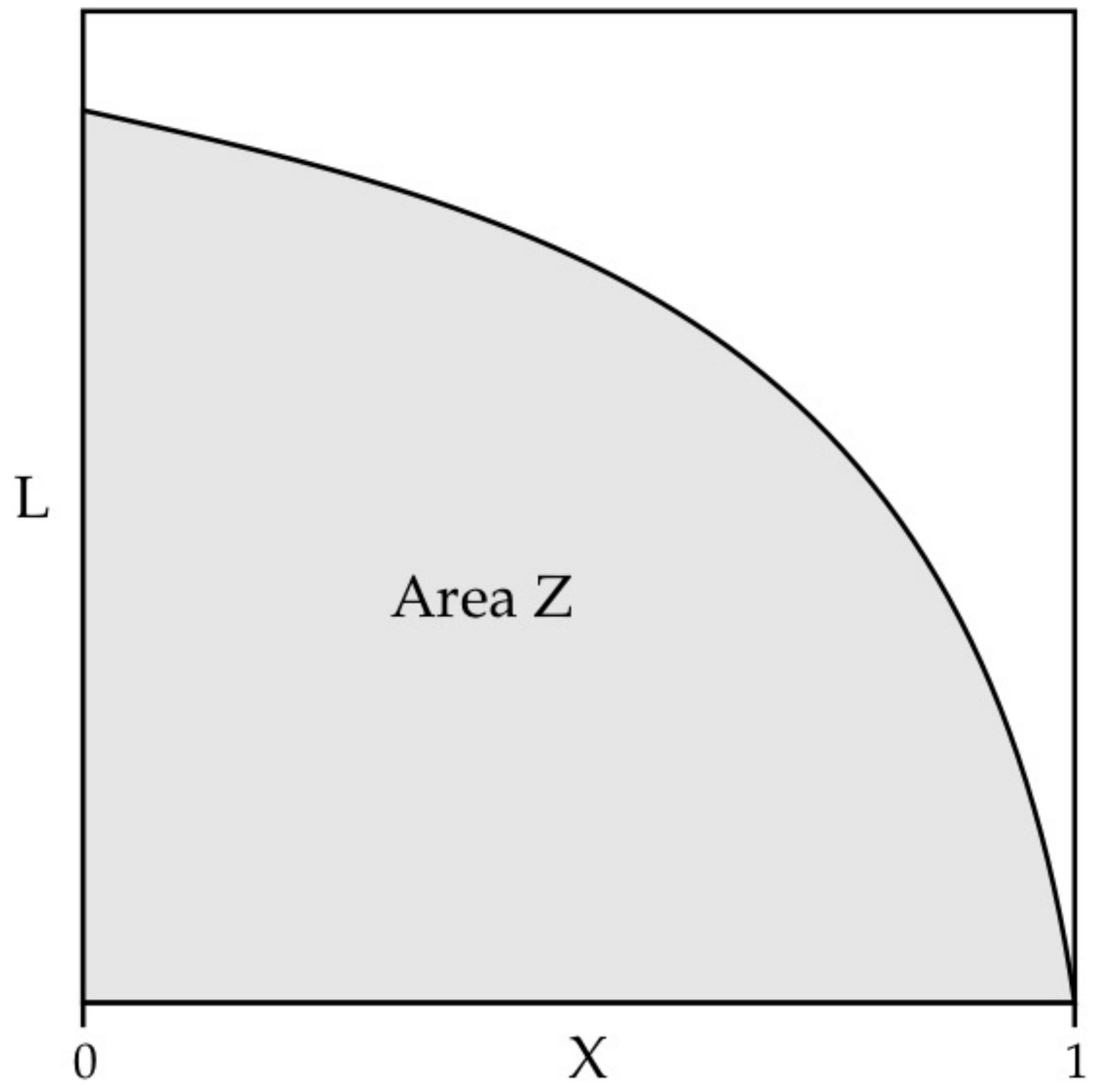
$$\mathcal{Z} = \int \mathcal{L}(d\,|\,\theta)\pi(\theta)d\theta \longrightarrow \mathcal{Z} = \int_0^1 \mathcal{L}(X)\, dX$$

Where

$$X(\lambda) = \int_{\mathcal{L}(\theta)>\lambda} \pi(\theta)\, d\theta$$

1. N-dimensional integral to 1D integral.
2. $X(\lambda)$ is known as prior mass.
3. As λ increases, the X decreases from 1 to 0.

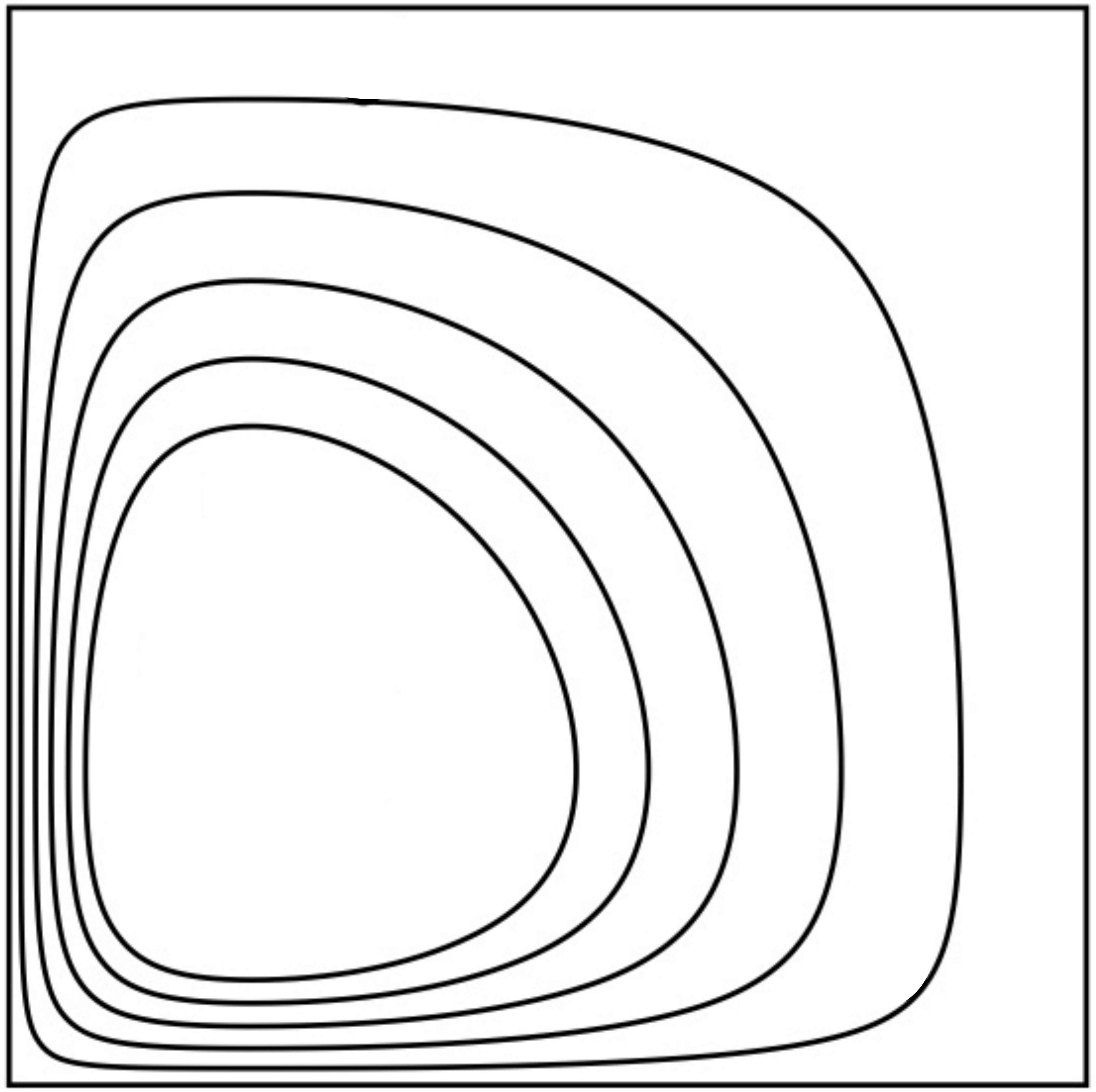
Source : Nested sampling for general Bayesian computation, John Skilling

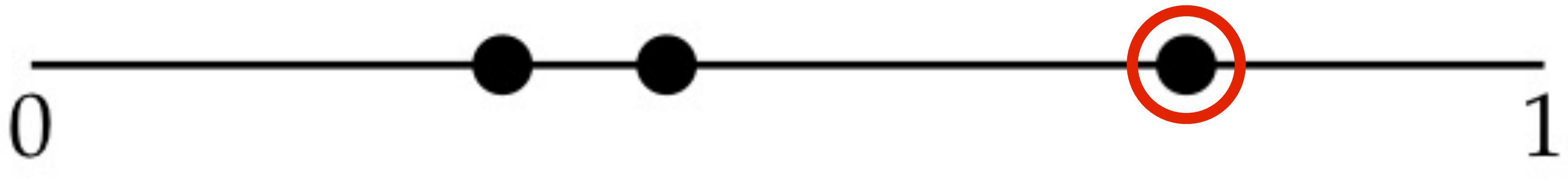


Nested Sampling procedure

Figure illustrating nested sampling method with 3 points.

Source : Nested sampling for general Bayesian computation, John Skilling

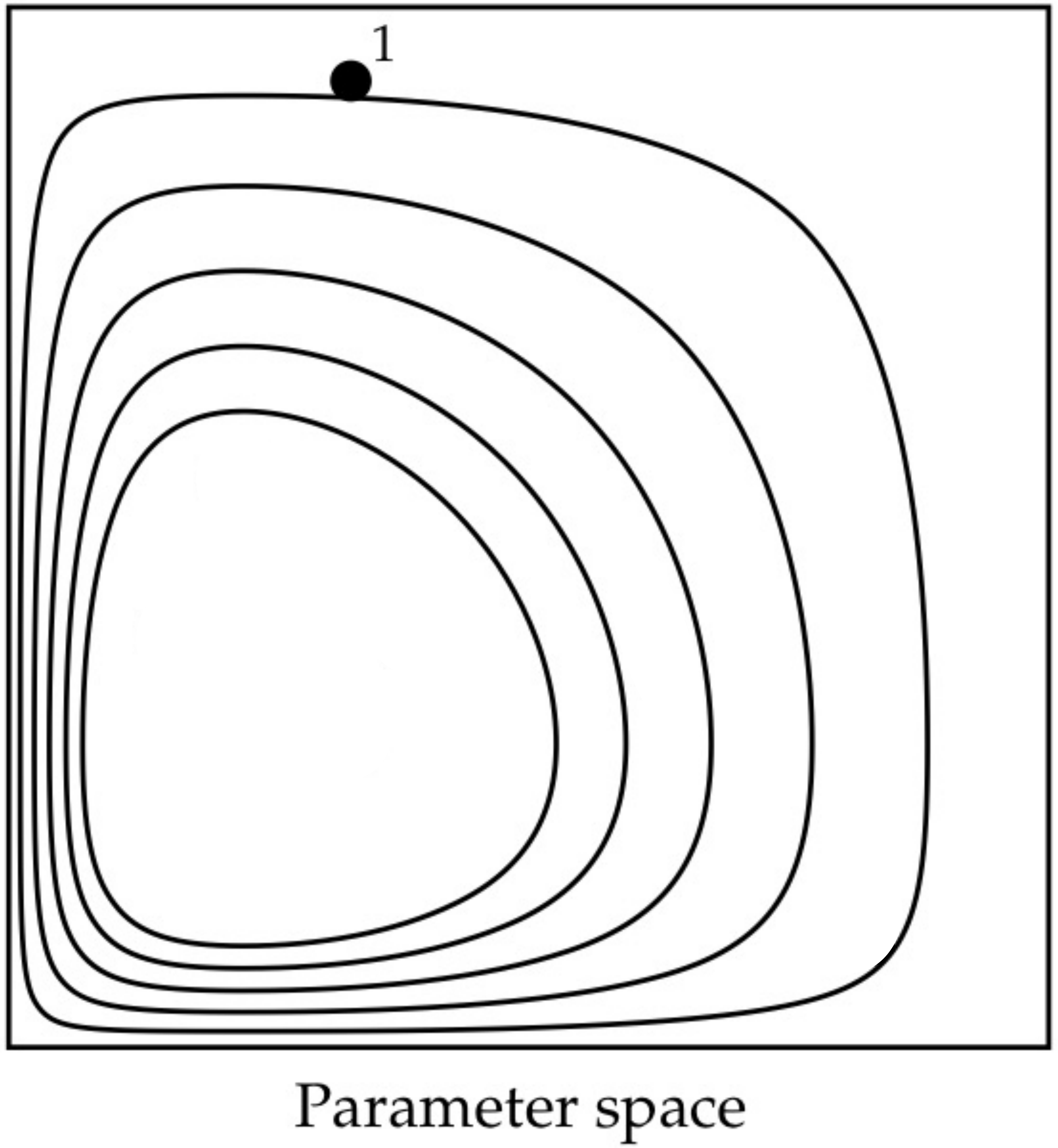


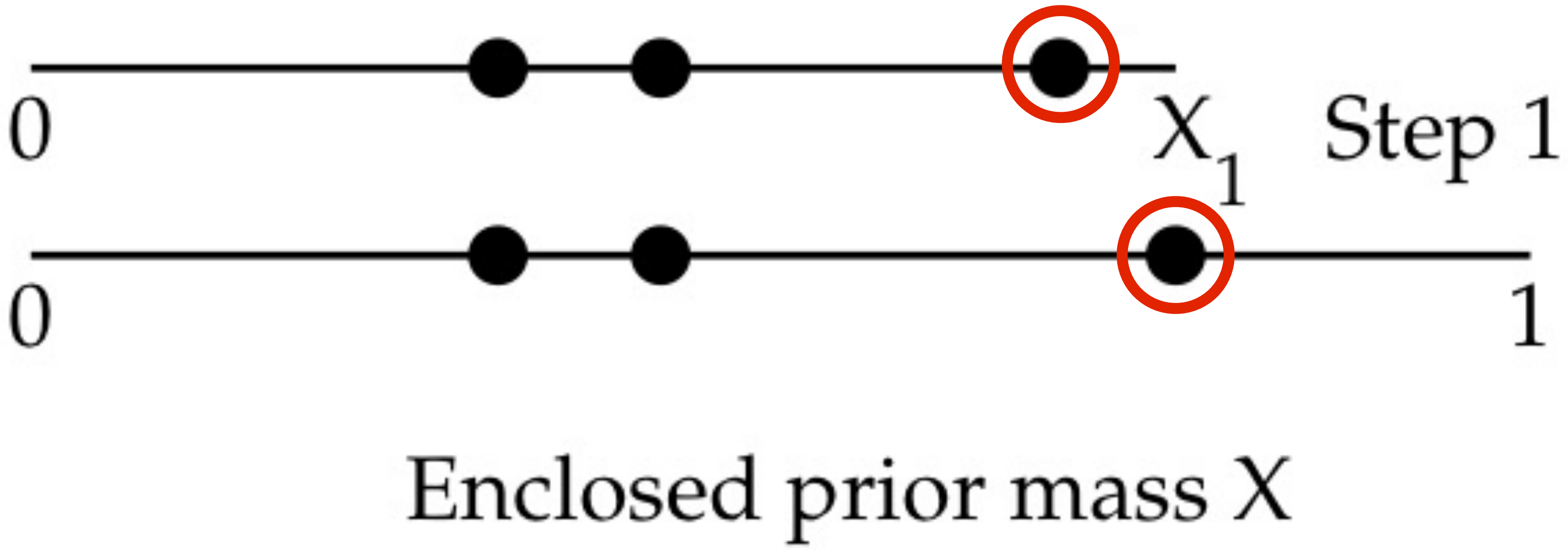


Nested Sampling procedure

Figure illustrating nested sampling method with 3 points.

Source : Nested sampling for general Bayesian computation, John Skilling

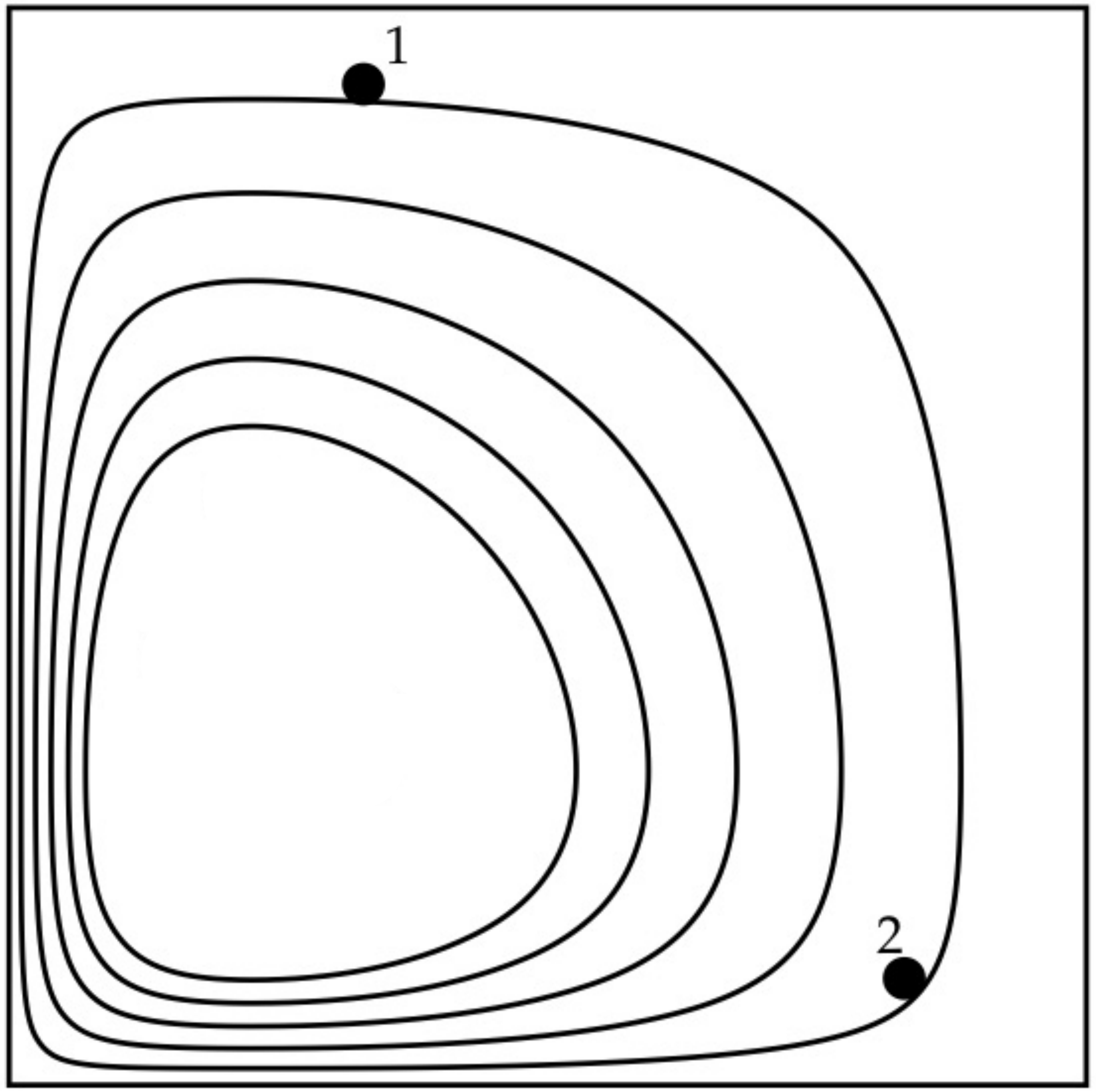




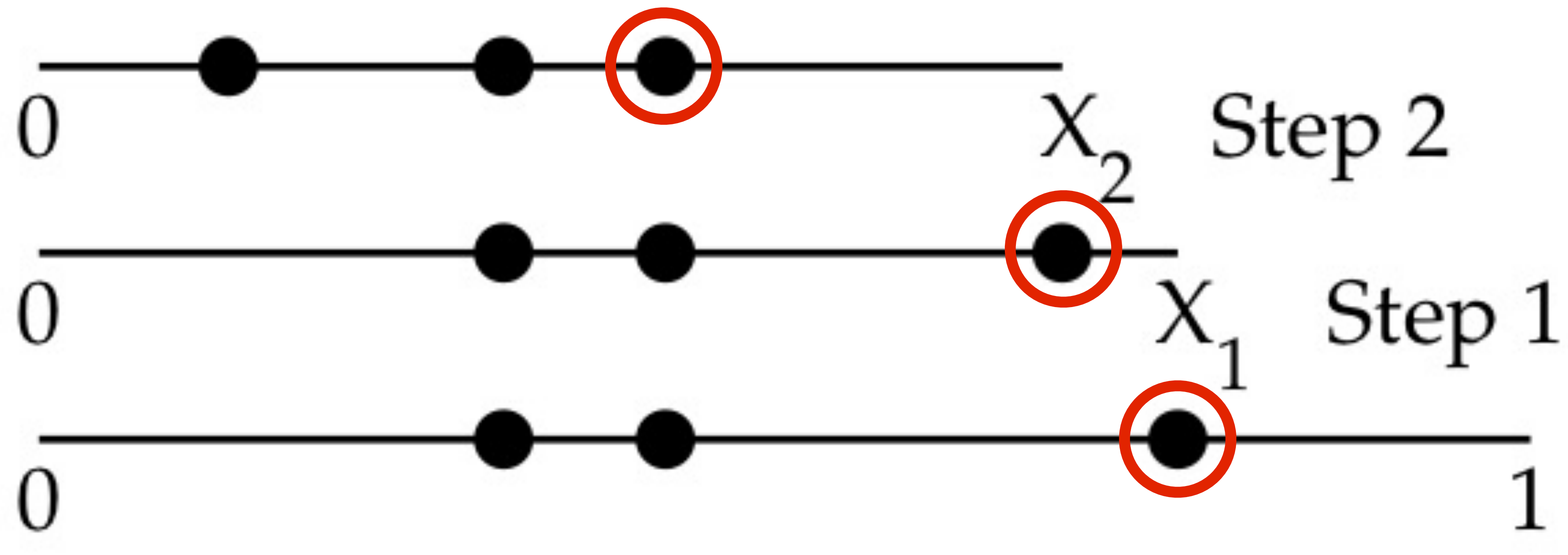
Nested Sampling procedure

Figure illustrating nested sampling method with 3 points.

Source : Nested sampling for general Bayesian computation, John Skilling



Parameter space

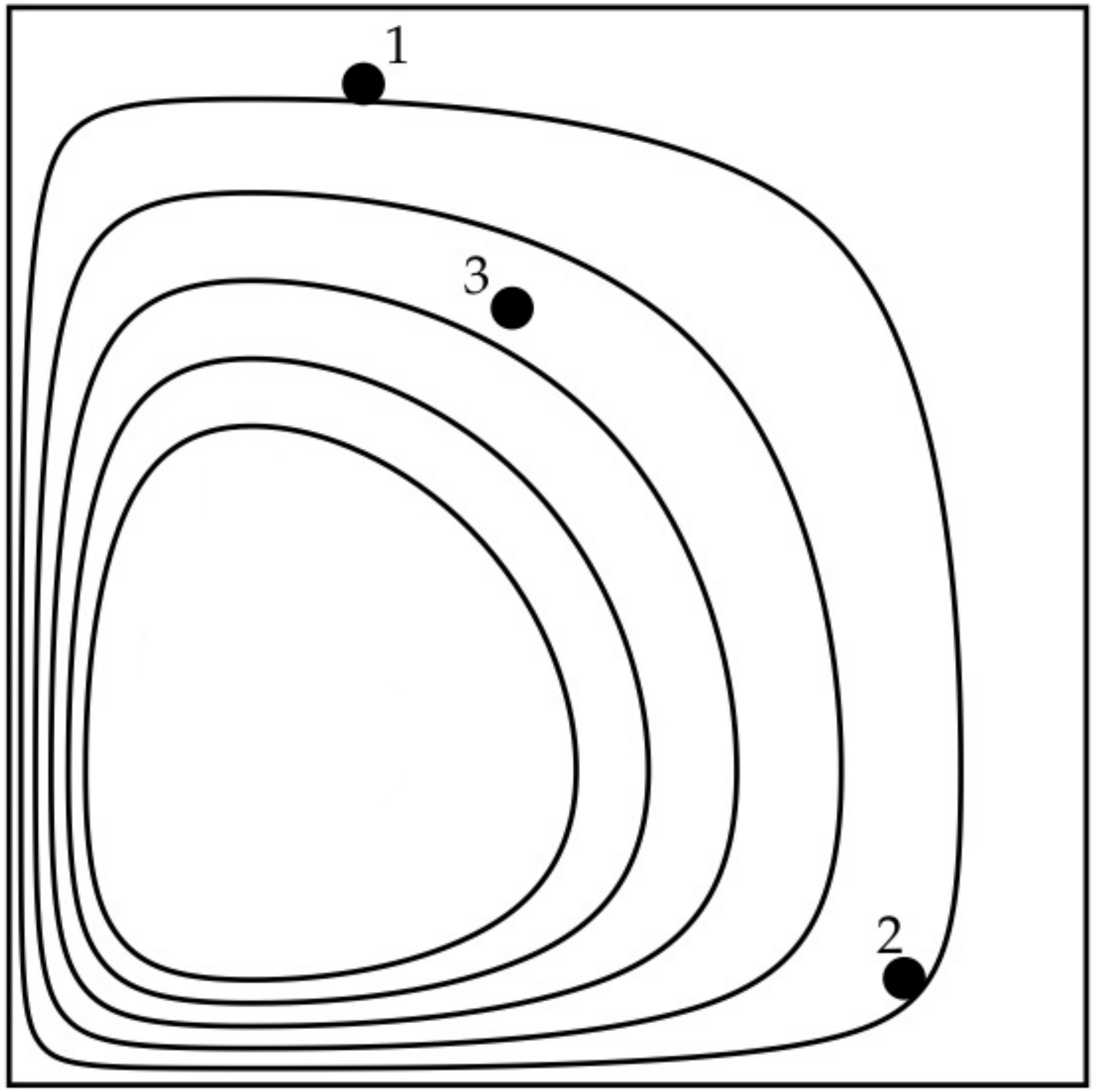


Enclosed prior mass X

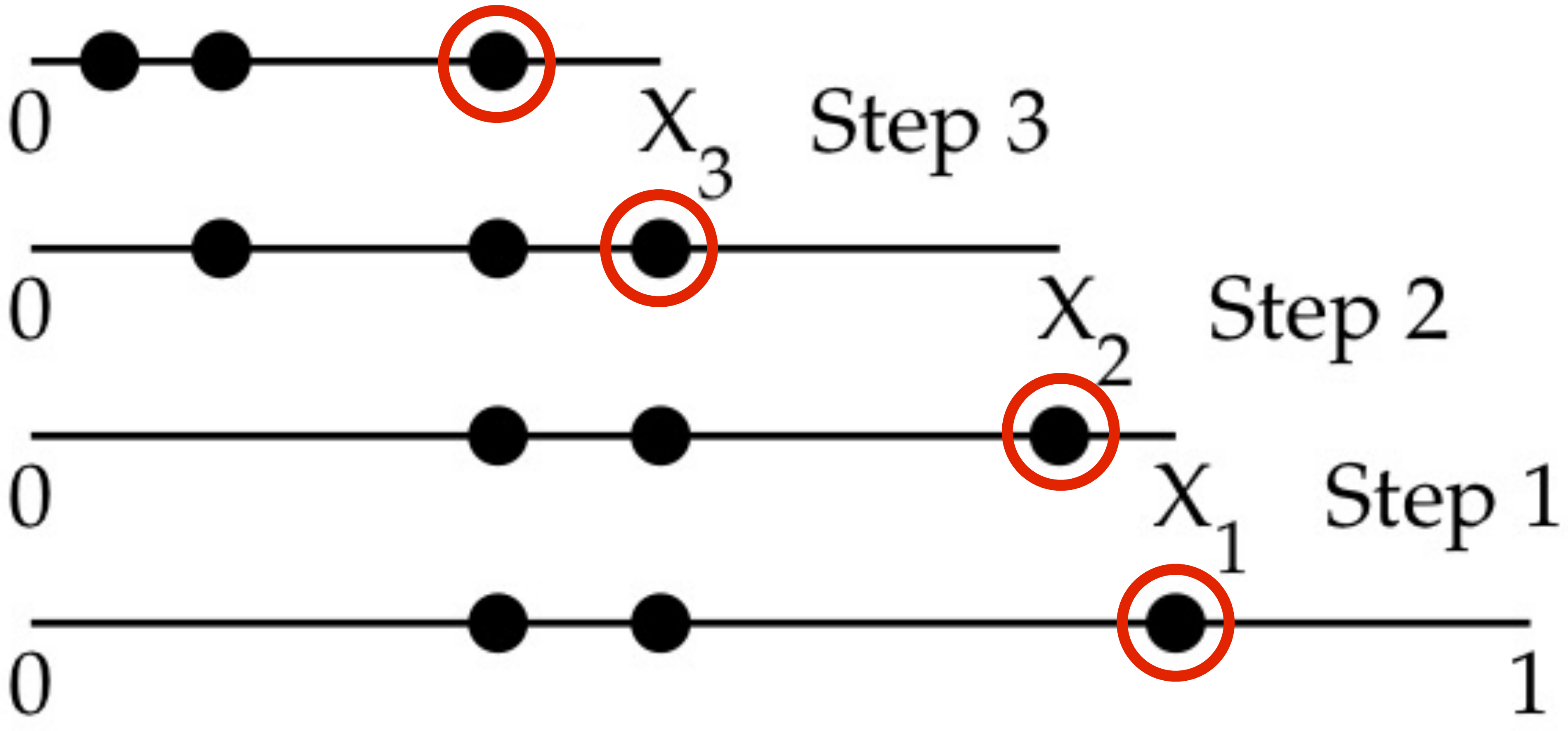
Nested Sampling procedure

Figure illustrating nested sampling method with 3 points.

Source : Nested sampling for general Bayesian computation, John Skilling



Parameter space

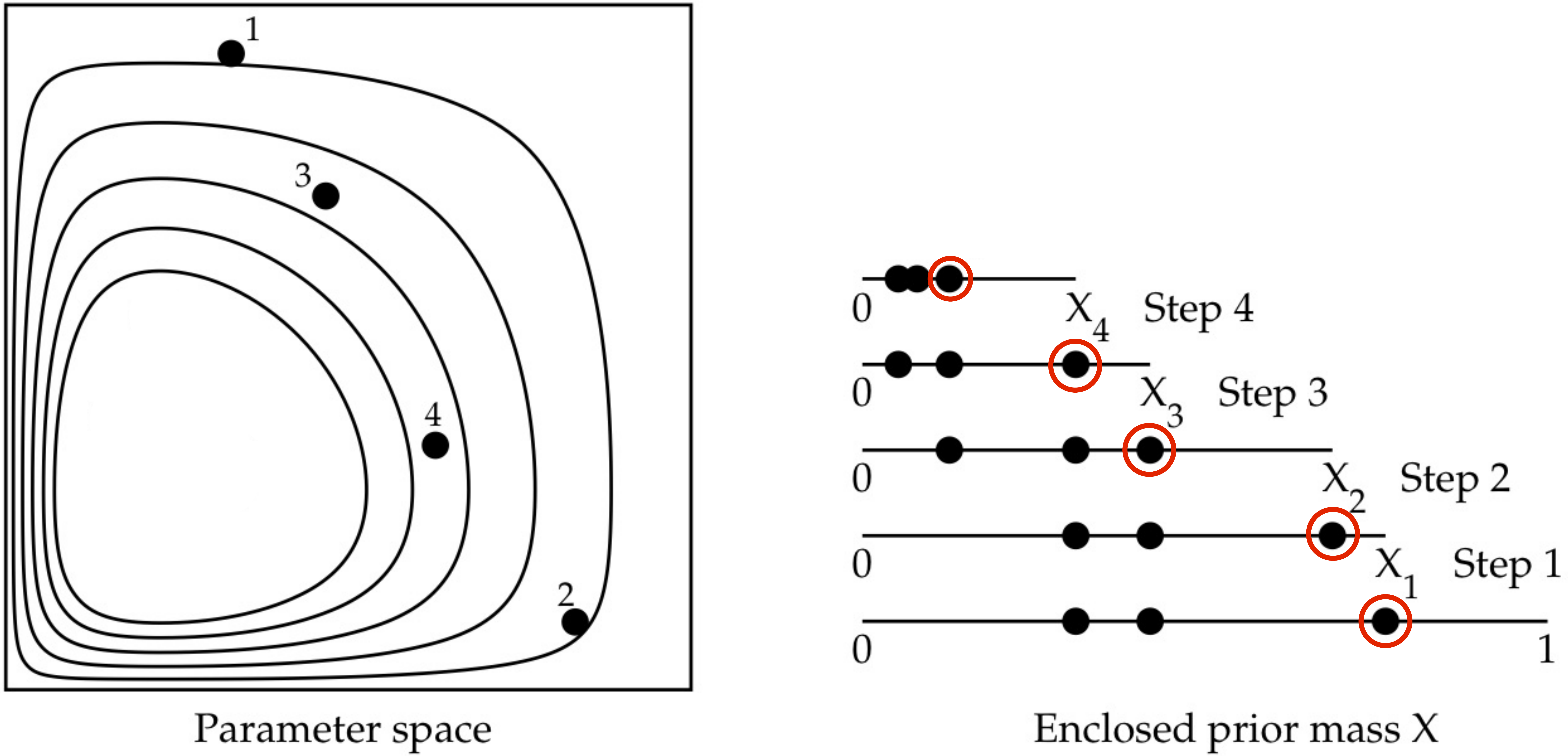


Enclosed prior mass X

Nested Sampling procedure

Figure illustrating nested sampling method with 3 points.

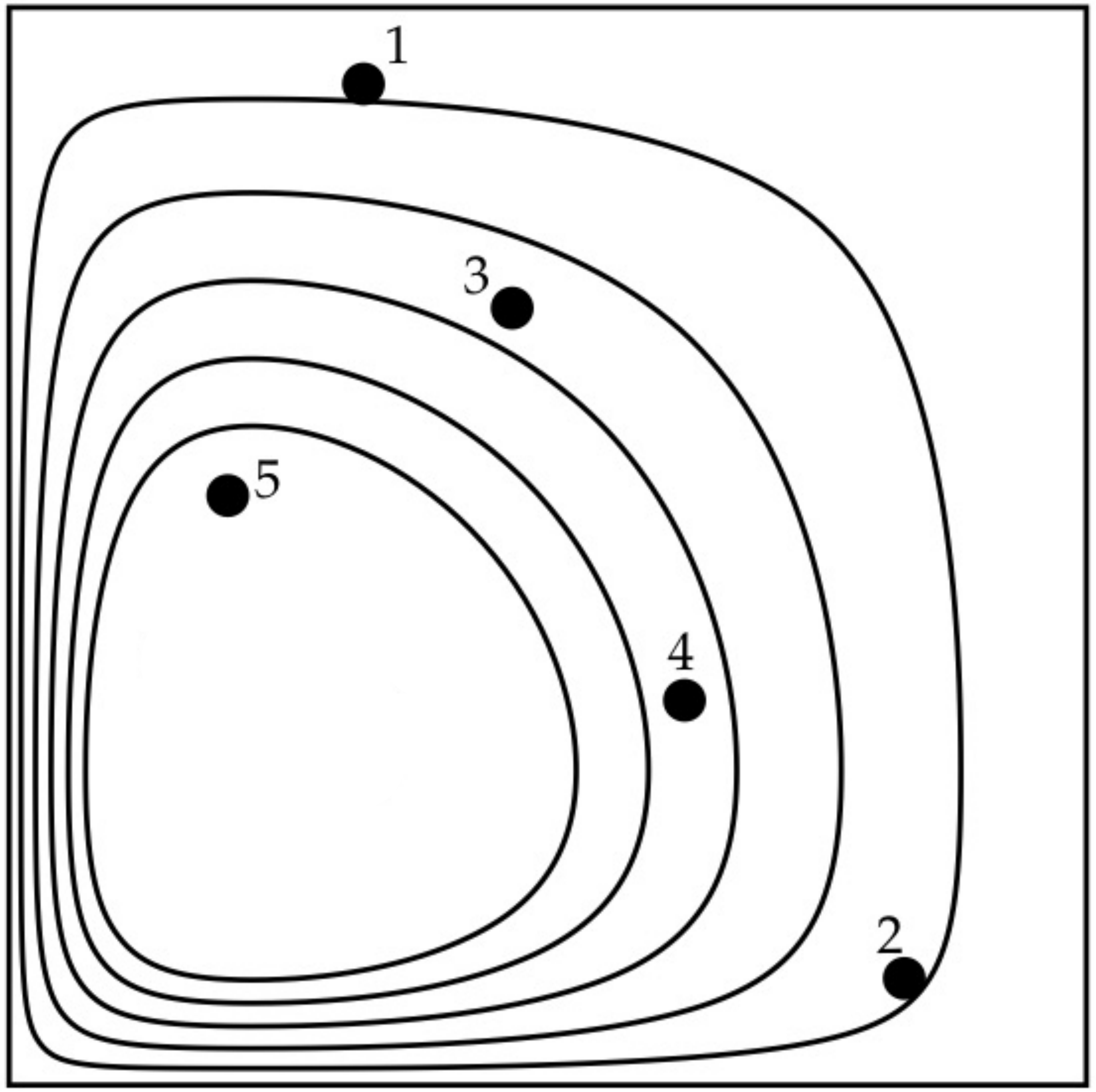
Source : Nested sampling for general Bayesian computation, John Skilling



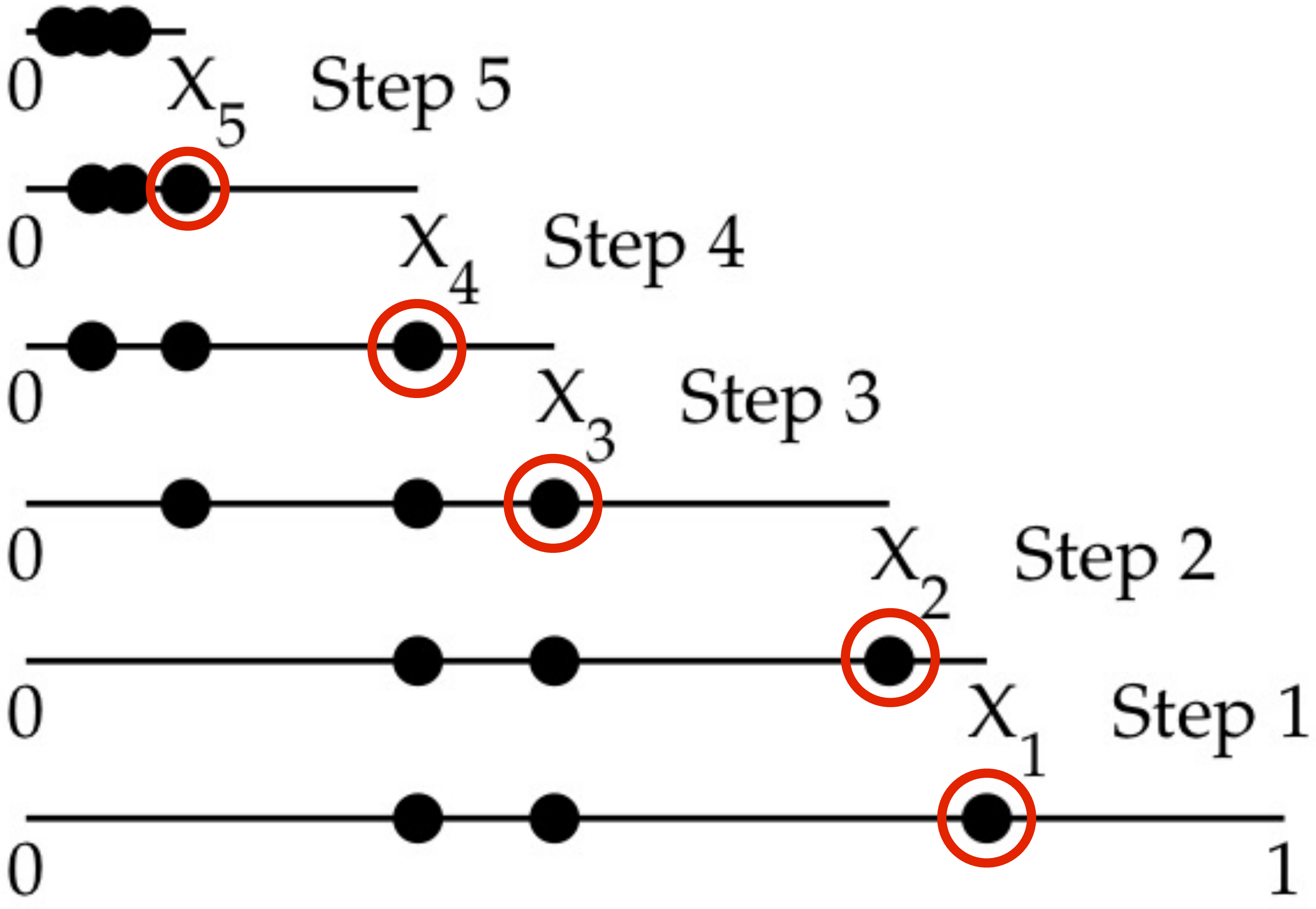
Nested Sampling procedure

Figure illustrating nested sampling method with 3 points.

Source : Nested sampling for general Bayesian computation, John Skilling



Parameter space

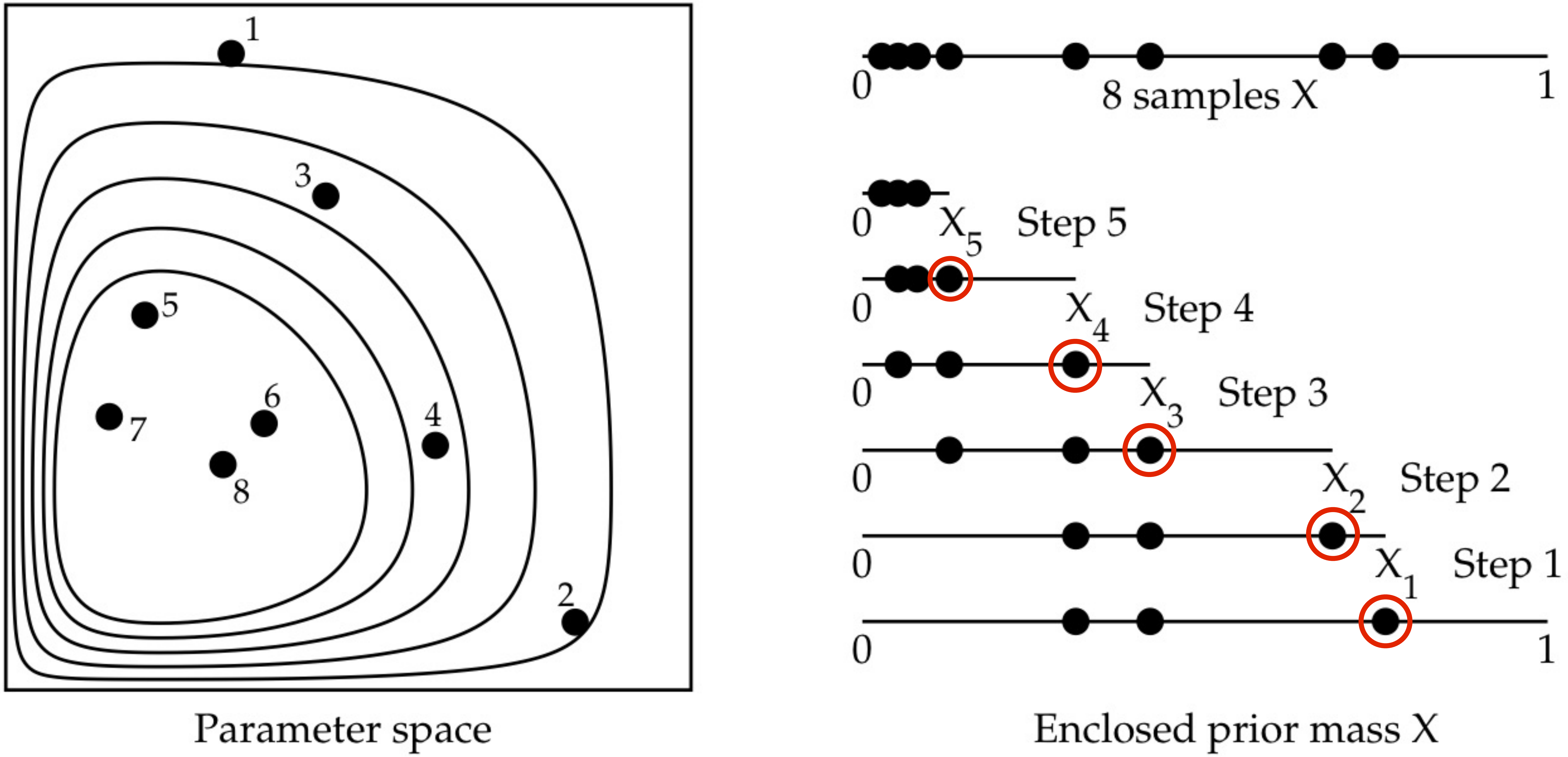


Enclosed prior mass X

Nested Sampling procedure

Figure illustrating nested sampling method with 3 points.

Source : Nested sampling for general Bayesian computation, John Skilling



Nested Sampling Procedure

STEP 1: Start with random N points $\theta_1, \ldots \theta_N$. Set $\mathscr{Z} = 0$, $X_o = 1$

Nested Sampling Procedure

STEP 1: Start with random N points $\theta_1, \ldots \theta_N$. Set $\mathscr{Z} = 0$, $X_o = 1$

STEP 2: Find the lowest value of $\mathscr{L}$ from N points. Set $X_1 = \exp(-1/N)$ or sample it.

Nested Sampling Procedure

STEP 1: Start with random N points $\theta_1, \ldots \theta_N$. Set $\mathscr{Z} = 0$, $X_o = 1$

STEP 2: Find the lowest value of $\mathscr{L}$ from N points. Set $X_1 = \exp(-1/N)$ or sample it.

STEP 3: $\mathscr{Z} += \mathscr{L}(X_o - X_1)$

Nested Sampling Procedure

STEP 1: Start with random N points $\theta_1, \ldots \theta_N$. Set $\mathscr{Z} = 0$, $X_o = 1$

STEP 2: Find the lowest value of $\mathscr{L}$ from N points. Set $X_1 = \exp(-1/N)$ or sample it.

STEP 3: $\mathscr{Z} + = \mathscr{L}(X_o - X_1)$

STEP 4: Repeat steps 2 and 3, $i = 1,2,3,...,j$ times.
$X_i = \exp(-i/N)$, $\mathscr{Z} + = \mathscr{L}_i(X_{i-1} - X_i)$

Nested Sampling Procedure

STEP 1: Start with random N points $\theta_1, \ldots \theta_N$. Set $\mathscr{Z} = 0$, $X_o = 1$

STEP 2: Find the lowest value of $\mathscr{L}$ from N points. Set $X_1 = \exp(-1/N)$ or sample it.

STEP 3: $\mathscr{Z} += \mathscr{L}(X_o - X_1)$

STEP 4: Repeat steps 2 and 3, $i = 1,2,3,\ldots,j$ times.
$X_i = \exp(-i/N)$, $\mathscr{Z} += \mathscr{L}_i(X_{i-1} - X_i)$

STEP 5: After j steps, $\mathscr{Z} += N^{-1}(\mathscr{L}(\theta_1) + \ldots + \mathscr{L}(\theta_N))$

Thank you for listening!