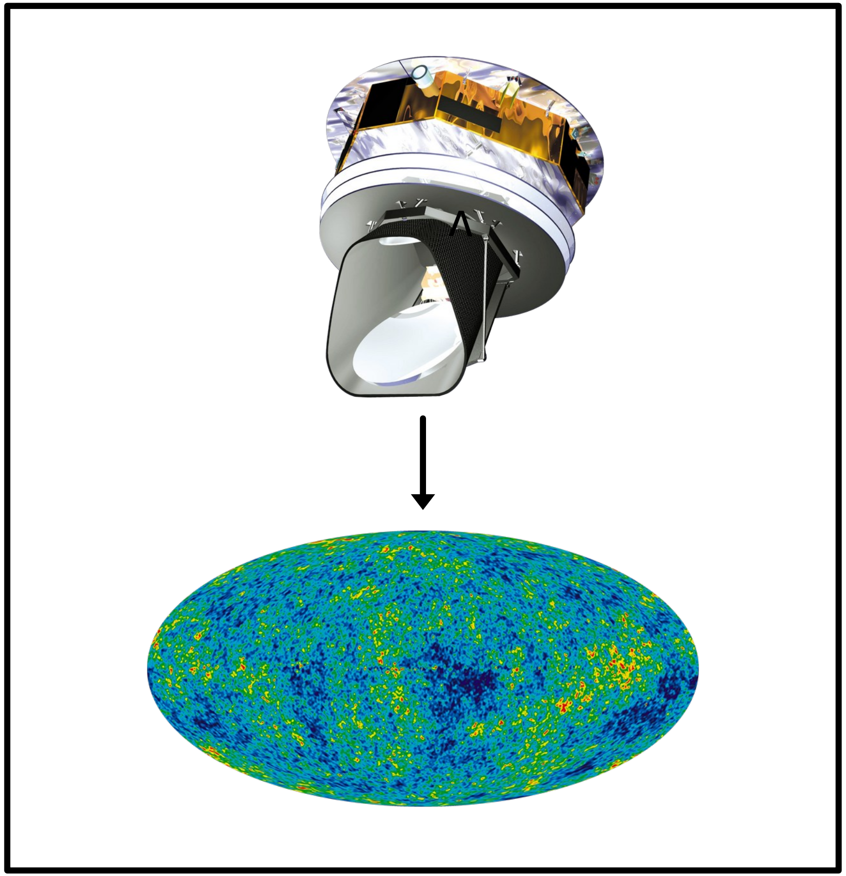
Data Modeling

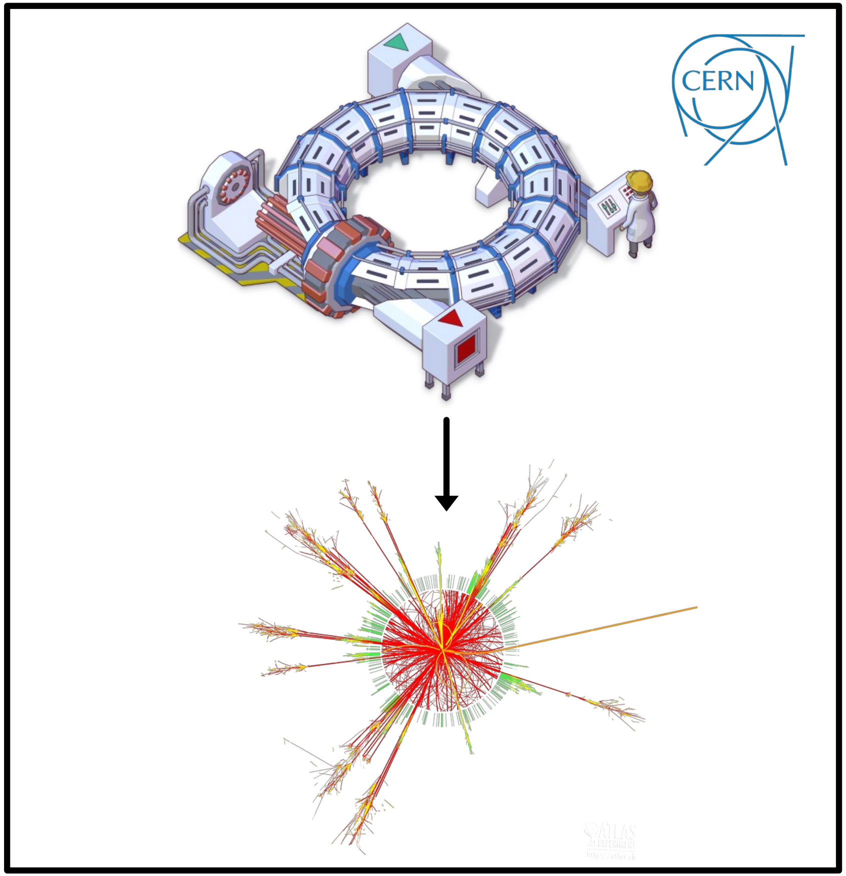
Parameter Estimation

Optimization

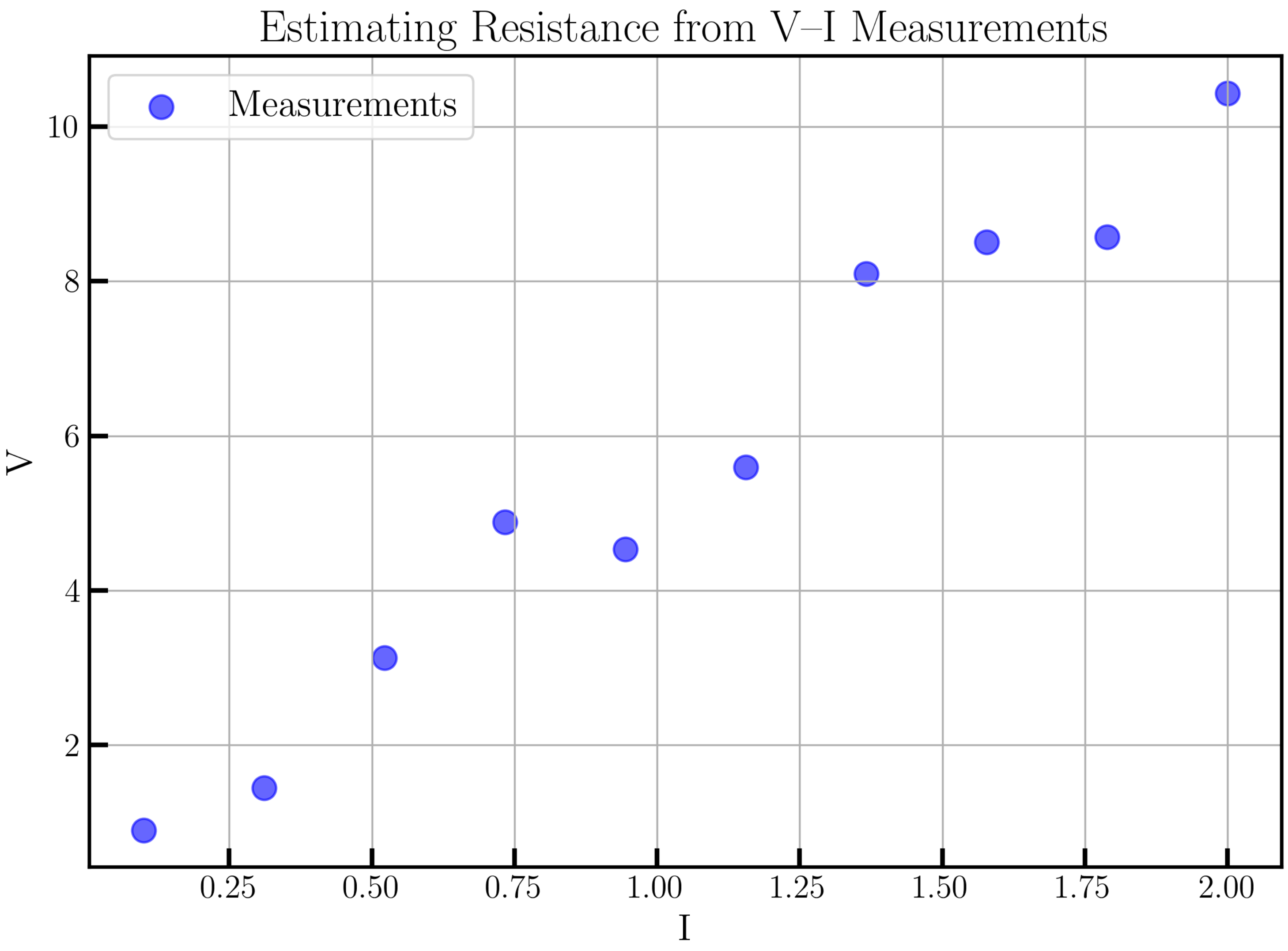
By: M.H. Jalali

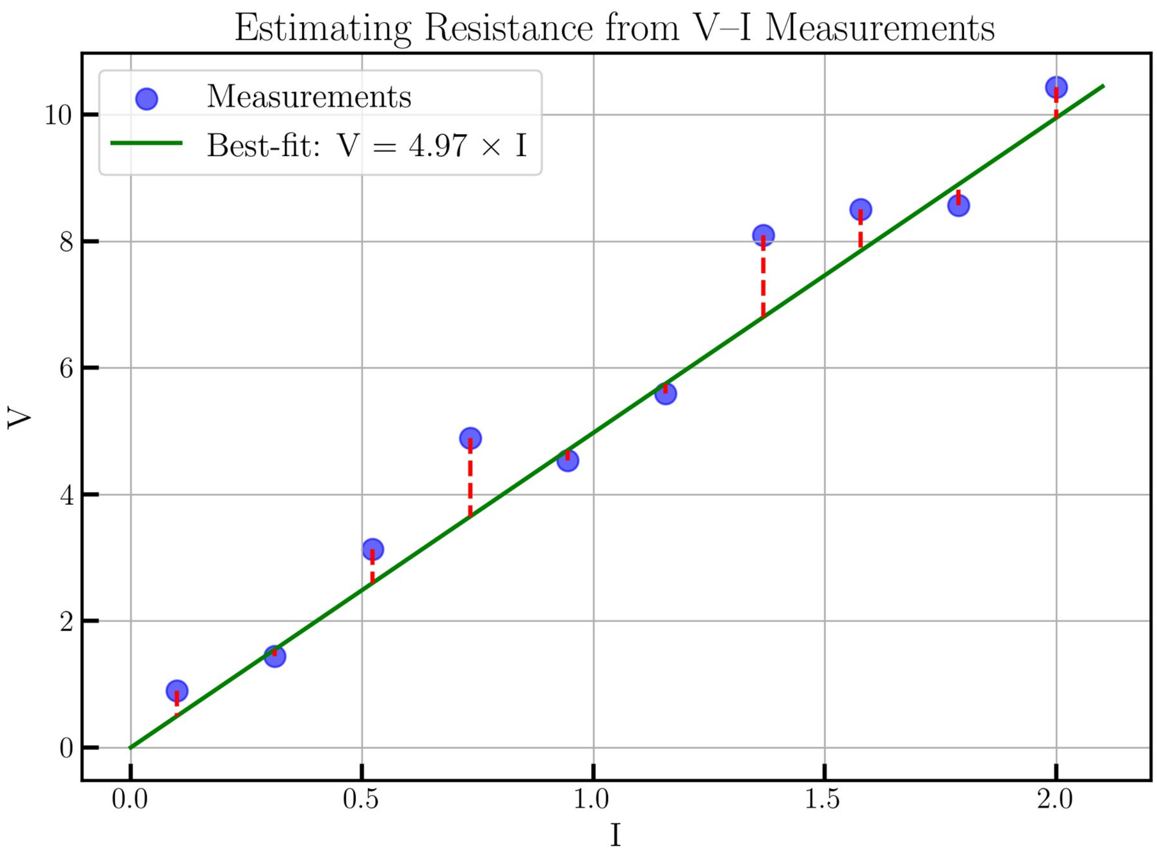
Parameter Estimation





How can this observation be described using a given physical model M with free parameters θ?


Estimating Resistance from V–I Measurements
Measurements
V
I
10
8
6
4
2
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

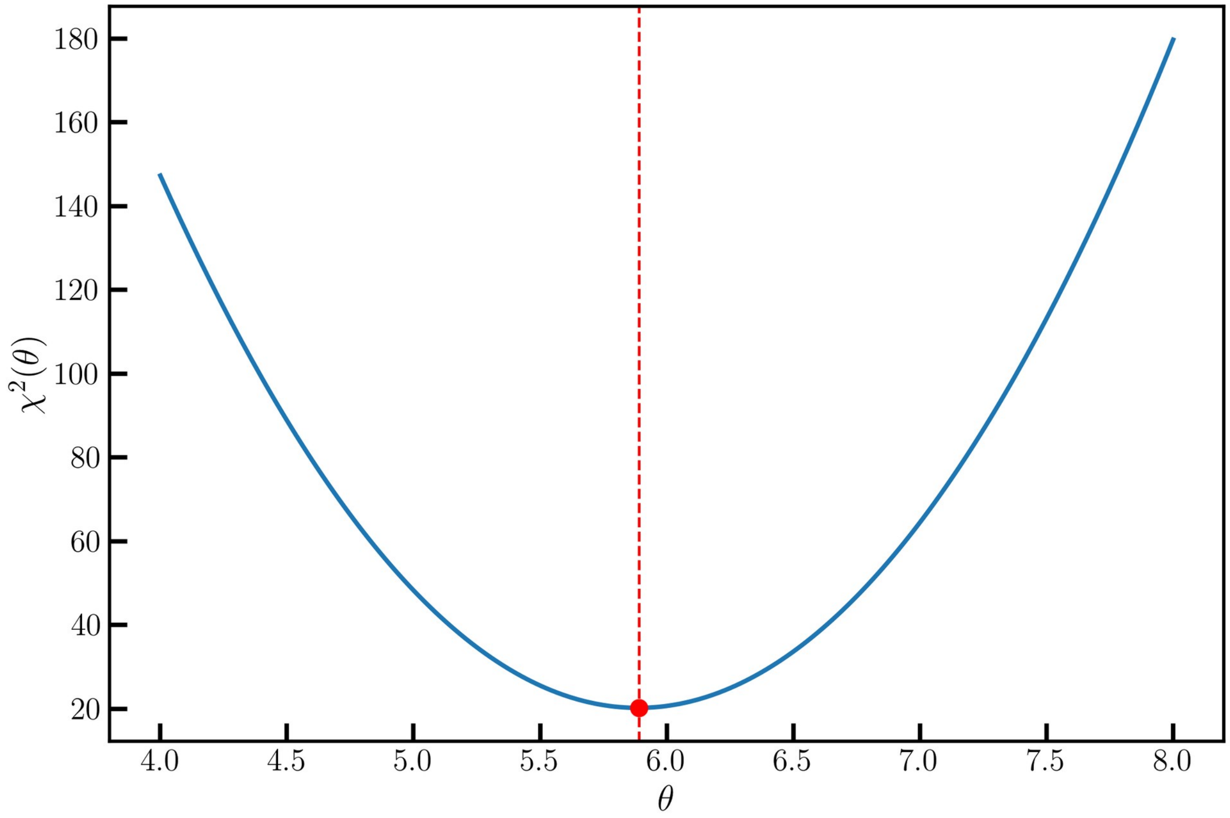


A criterion for assessing the similarity between data and model

$$\chi^2(R) = \sum_{i=1}^{N} \frac{(V_i - R\,I_i)^2}{\sigma_{V,i}^2}$$

How well the model predicts the observed data?

$$\chi^2(\theta) = \sum_{i=1}^{N} \frac{\left[y_i^{(\mathrm{obs})} - y_i^{(\mathrm{th})}(\theta) \right]^2}{\sigma_i^2}$$



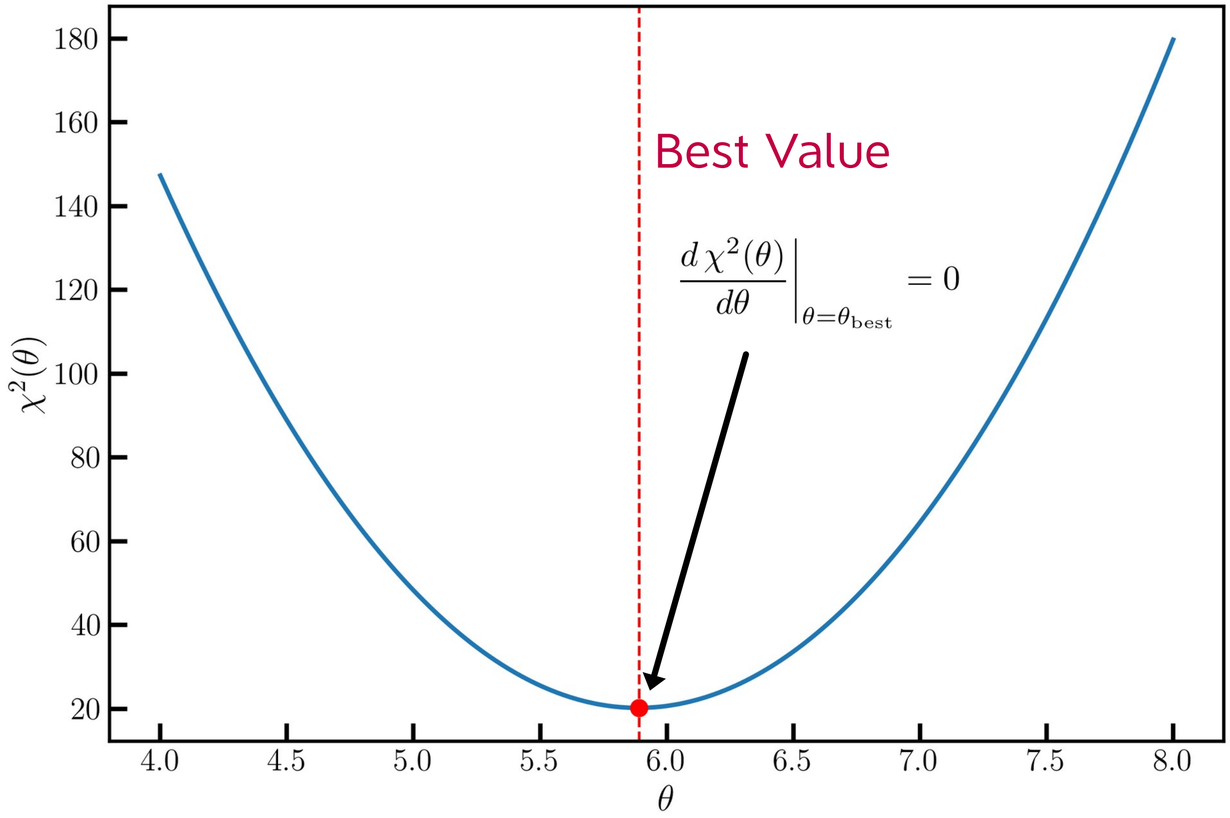
$$\chi^2(\theta) = \sum_{i=1}^{N} \frac{\left[y_i^{(\mathrm{obs})} - y_i^{(\mathrm{th})}(\theta) \right]^2}{\sigma_i^2}$$

$$\downarrow$$

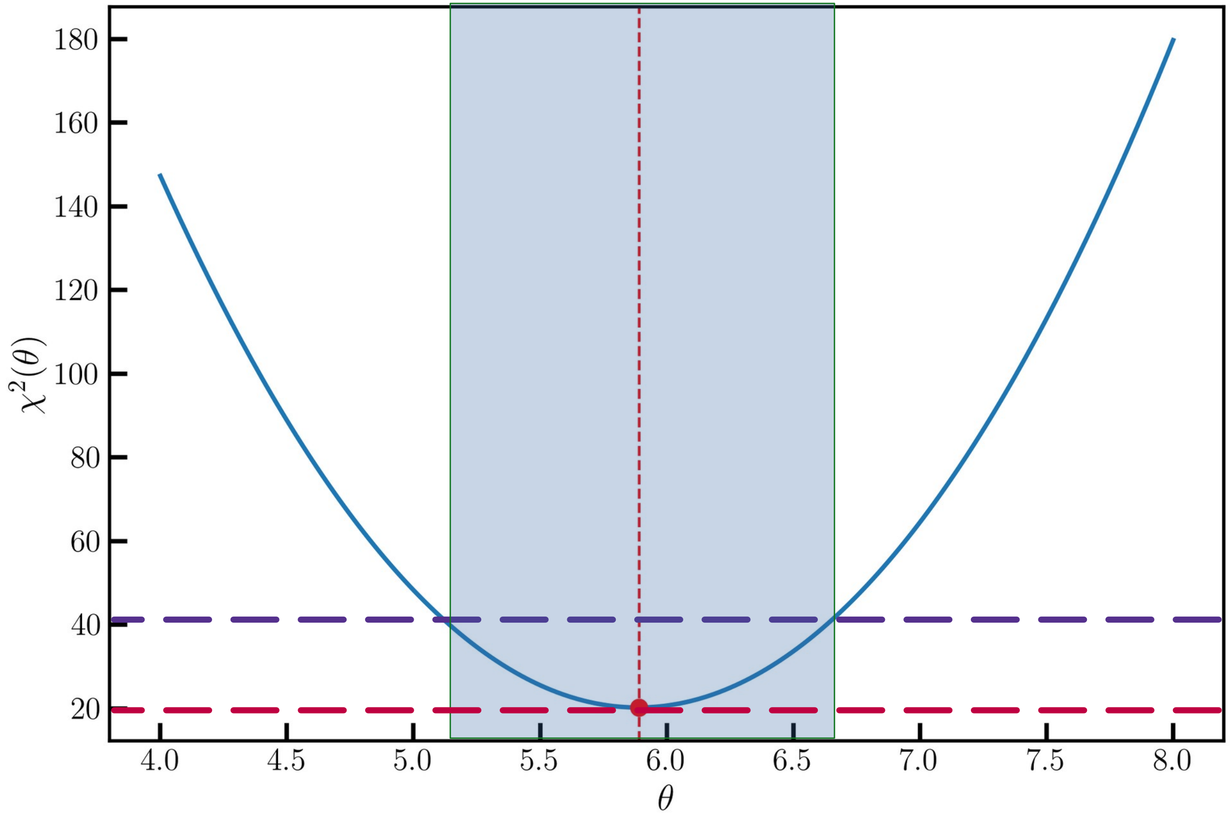
$$\left. \frac{d\,\chi^2(\theta)}{d\theta} \right|_{\theta=\theta_{\mathrm{best}}} = 0$$

$$\downarrow$$

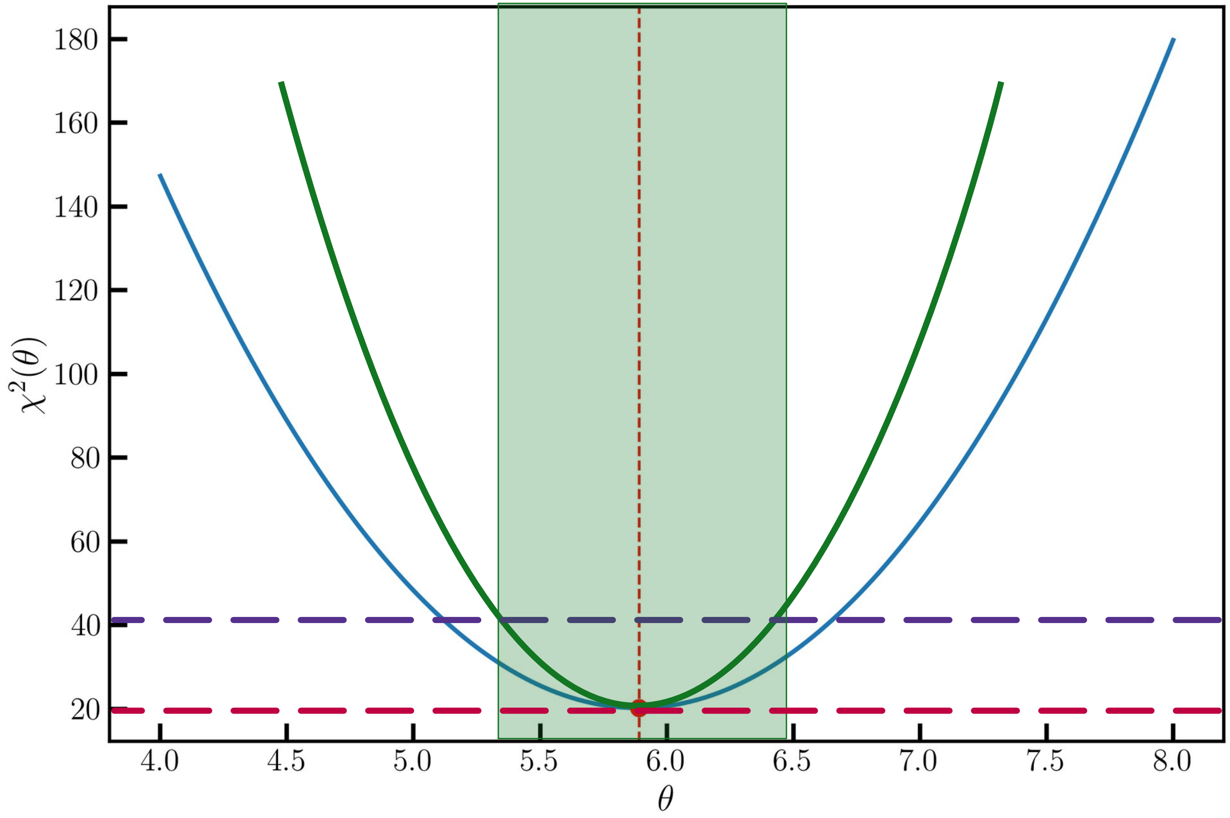
$$\boxed{\theta_{\mathrm{best}} = \frac{\displaystyle\sum_{i=1}^{N} \frac{\theta_i \, y_i^{\mathrm{obs}}}{\sigma_i^2}}{\displaystyle\sum_{i=1}^{N} \frac{\theta_i^2}{\sigma_i^2}}}$$



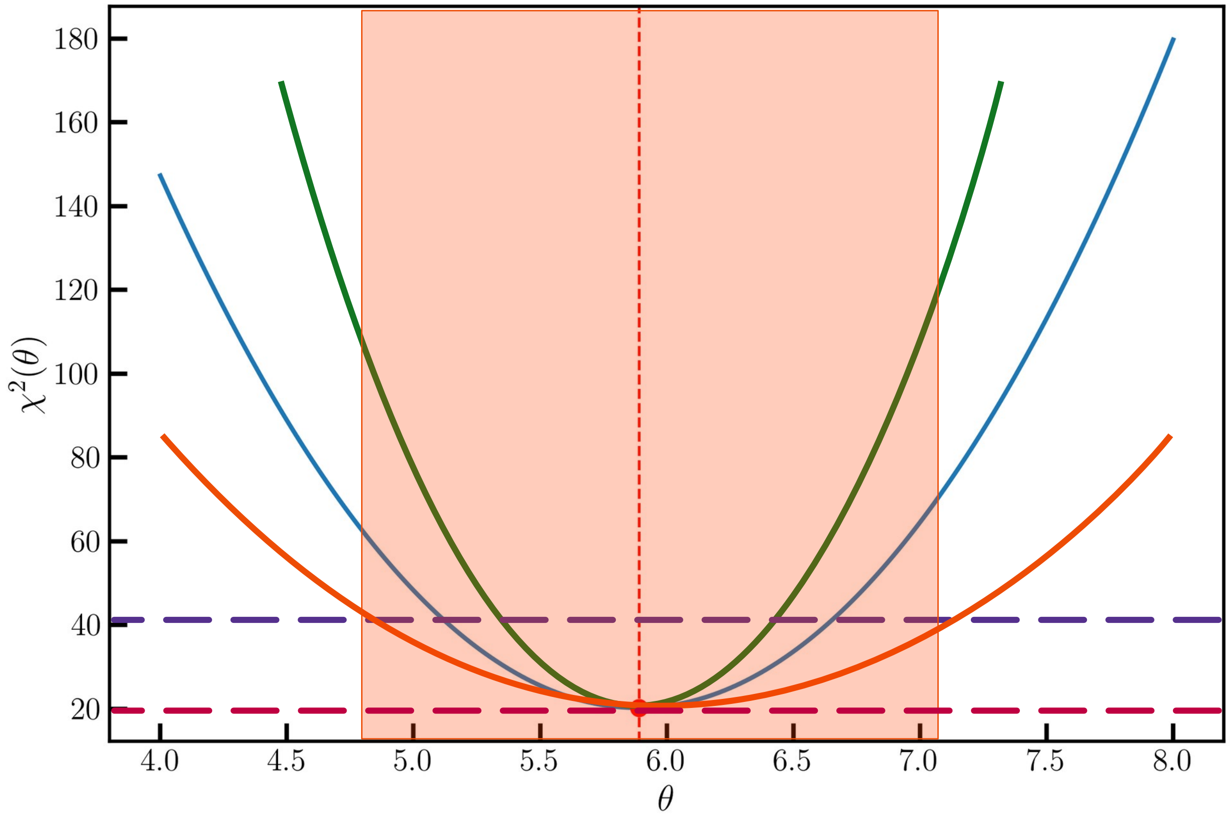
Error Estimation:



Error Estimation:



Error Estimation:



Error Estimation:

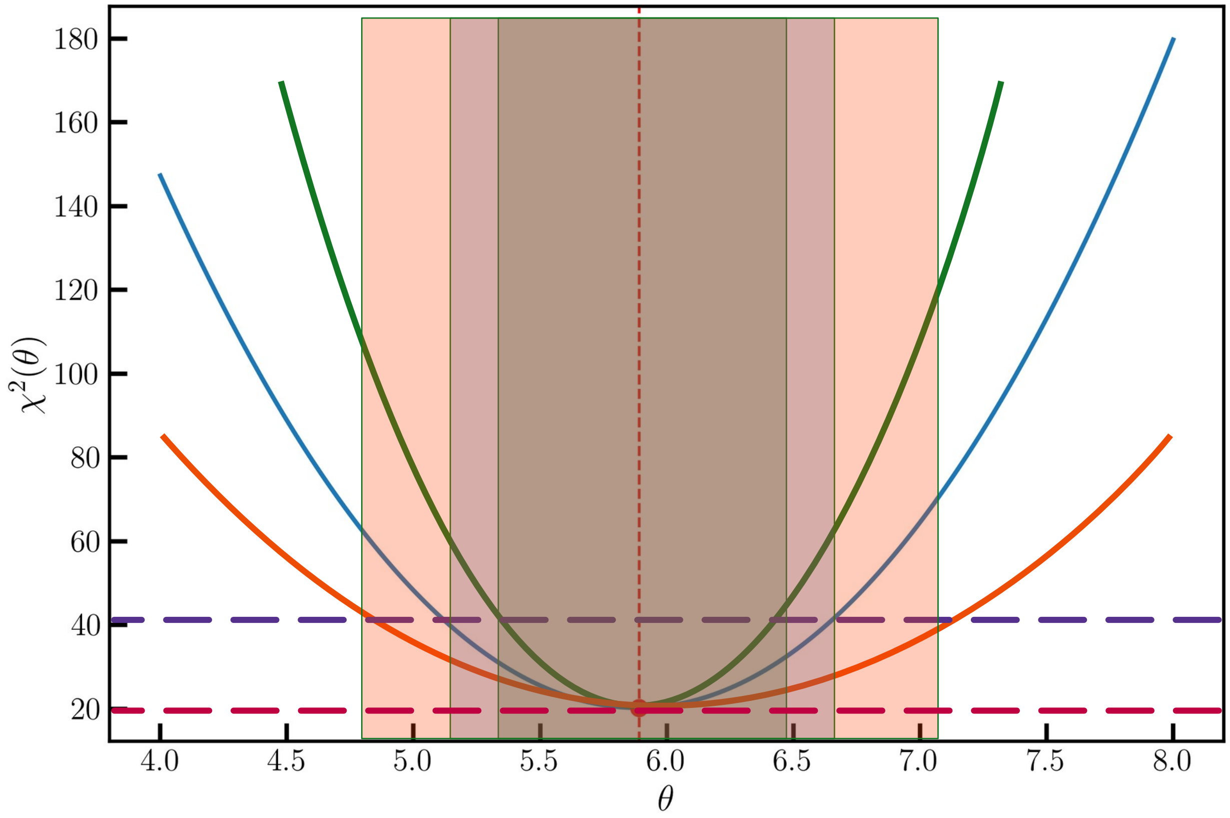
$$\left.\frac{d\,\chi^2(\theta)}{d\theta}\right|_{\theta=\theta_{\text{best}}} = 0$$

$$\chi^2(\theta) \approx \chi^2_{\min} + \frac{1}{2}\left.\frac{d^2\chi^2}{d\theta^2}\right|_{\theta_{\text{best}}}(\theta-\theta_{\text{best}})^2$$

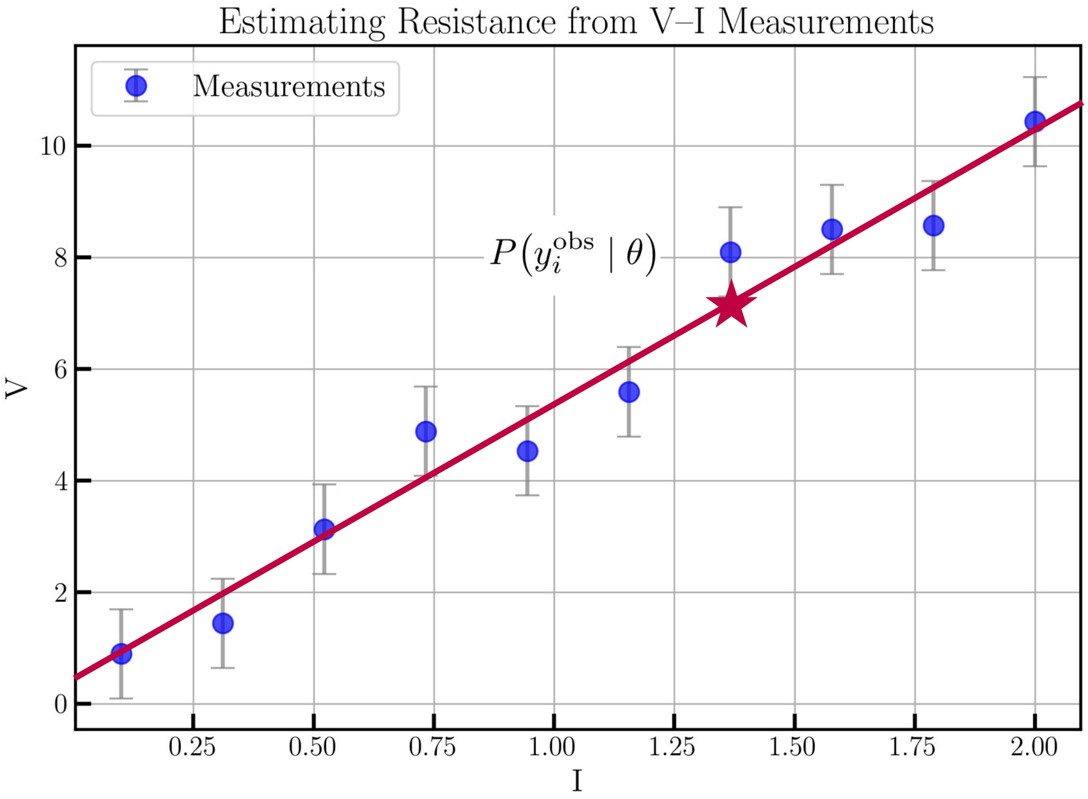
$$\chi^2(\theta_{\text{best}} \pm \sigma_\theta) = \chi^2_{\min} + 1$$

$$\frac{1}{2}\left.\frac{d^2\chi^2}{d\theta^2}\right|_{\theta_{\text{best}}}(\sigma_\theta)^2 = 1$$

$$\boxed{\sigma_\theta = \sqrt{\frac{2}{\left.\frac{d^2\chi^2}{d\theta^2}\right|_{\theta_{\text{best}}}}}}$$



Probabilistic Approach :

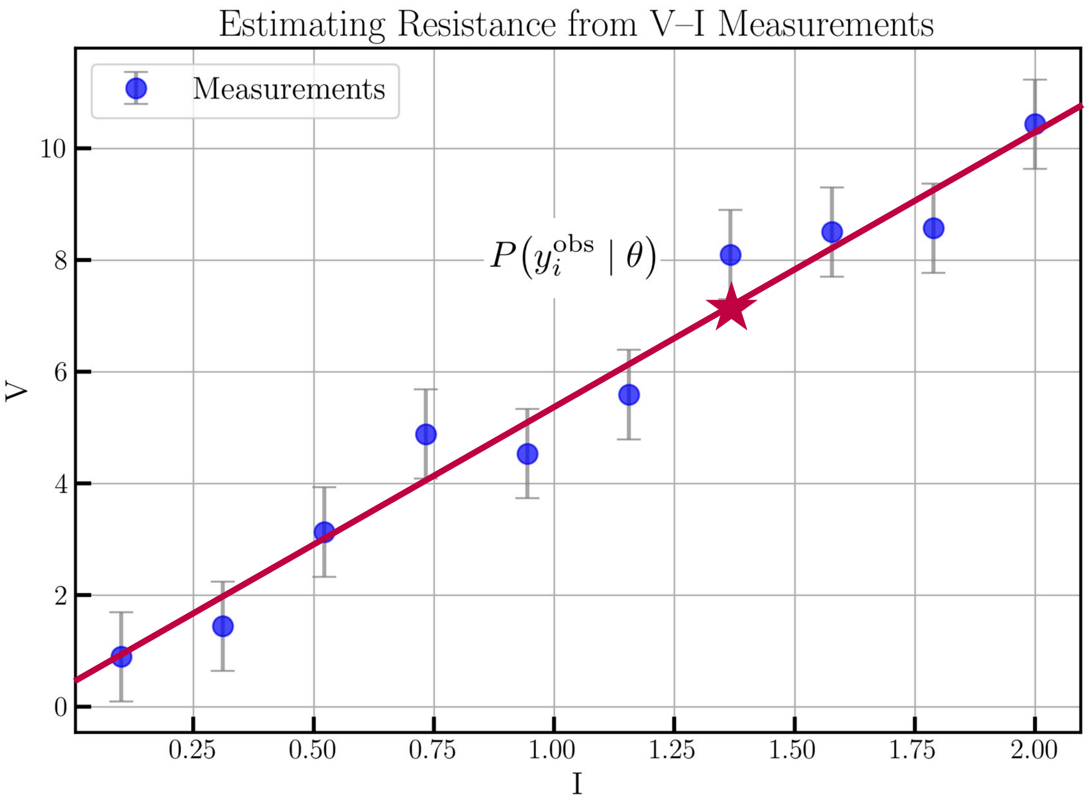


Probabilistic Approach :

Likelihood

$$L(D \mid \theta) = \prod_{i=1}^{N} P\left(y_i^{\text{obs}} \mid \theta\right)$$

Probability of observing the data given a specific parameter value



Probabilistic Approach :

Likelihood

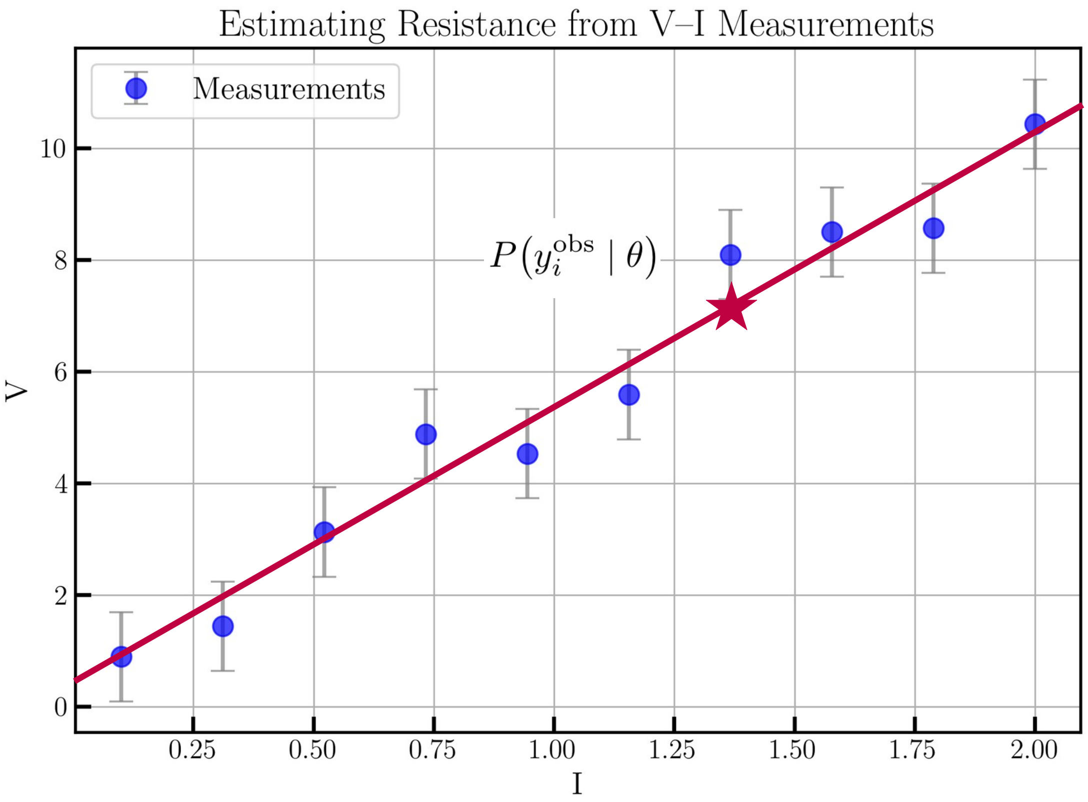
$$L(D \mid \theta) = \prod_{i=1}^{N} P\left(y_i^{\mathrm{obs}} \mid \theta\right)$$

Probability of observing the data given a specific parameter value

Central Limit Theorem

$$P\left(y_i^{\mathrm{obs}} \mid \theta\right) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{\left(y_i^{\mathrm{obs}} - y_i^{\mathrm{theory}}(\theta)\right)^2}{2\,\sigma_i^2}\right]$$

$$L(\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{\left(y_i^{\mathrm{obs}} - y_i^{\mathrm{theory}}(\theta)\right)^2}{2\,\sigma_i^2}\right]$$



Probabilistic Approach :

Likelihood

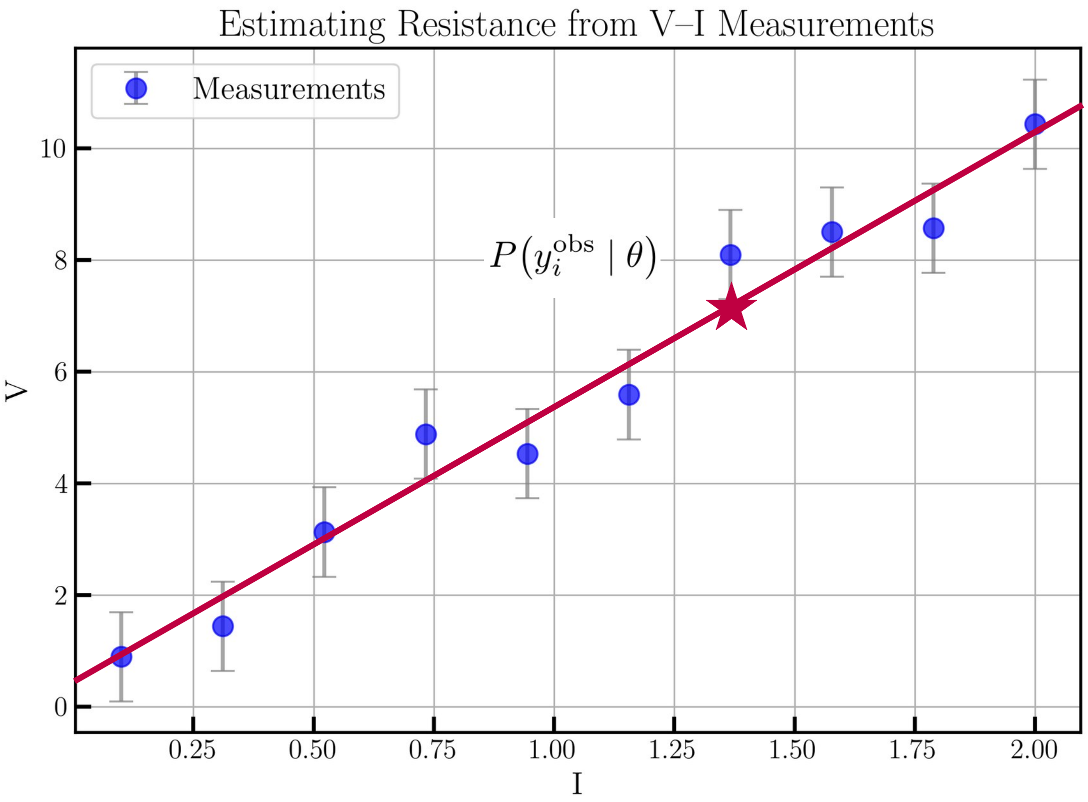
$$L(D \mid \theta) = \prod_{i=1}^{N} P\left(y_i^{\mathrm{obs}} \mid \theta\right)$$

Probability of observing the data given a specific parameter value

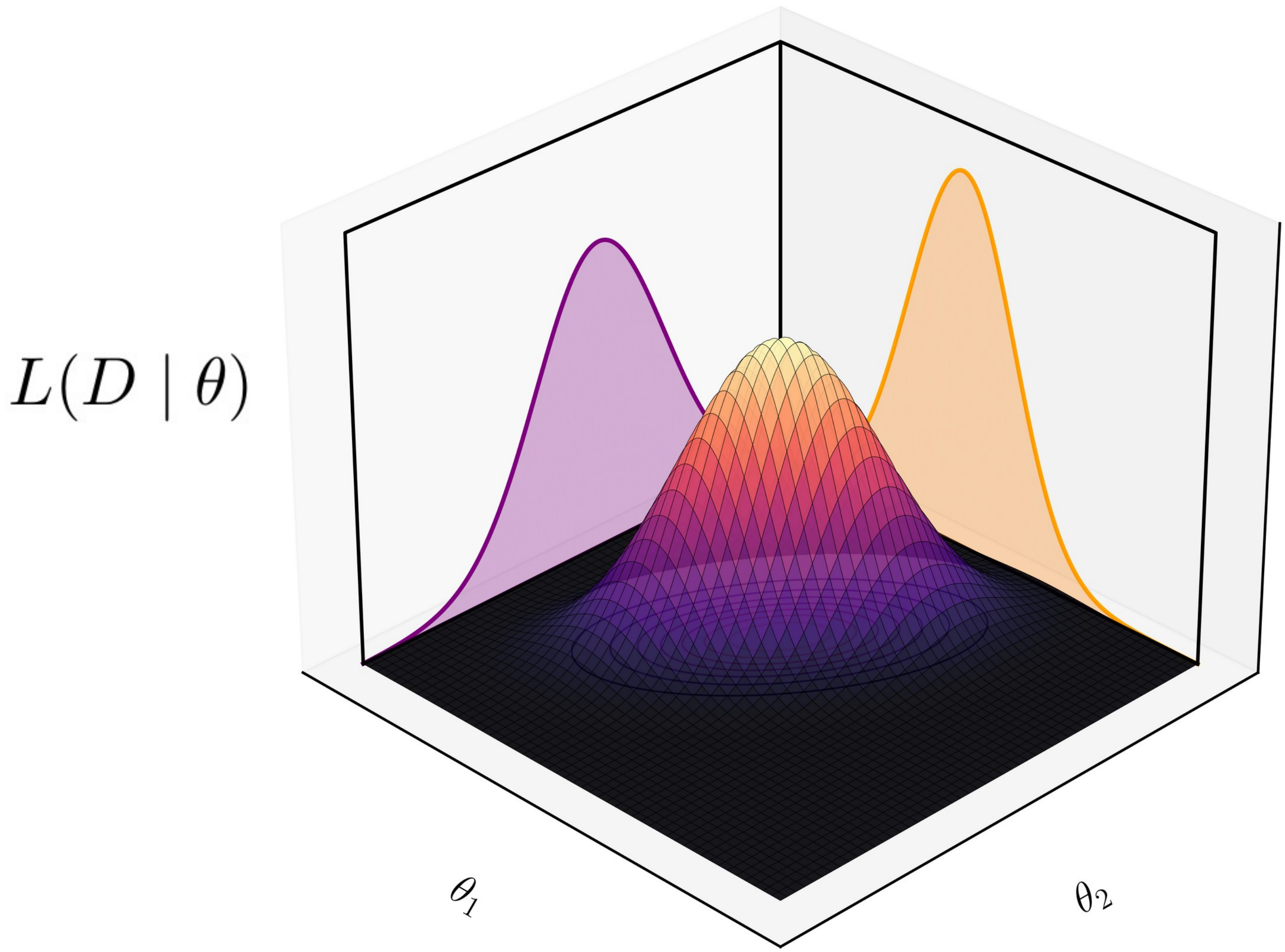
Central Limit Theorem

$$P\left(y_i^{\mathrm{obs}} \mid \theta\right) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{\left(y_i^{\mathrm{obs}} - y_i^{\mathrm{theory}}(\theta)\right)^2}{2\,\sigma_i^2}\right]$$

$$L(\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{\left(y_i^{\mathrm{obs}} - y_i^{\mathrm{theory}}(\theta)\right)^2}{2\,\sigma_i^2}\right]$$

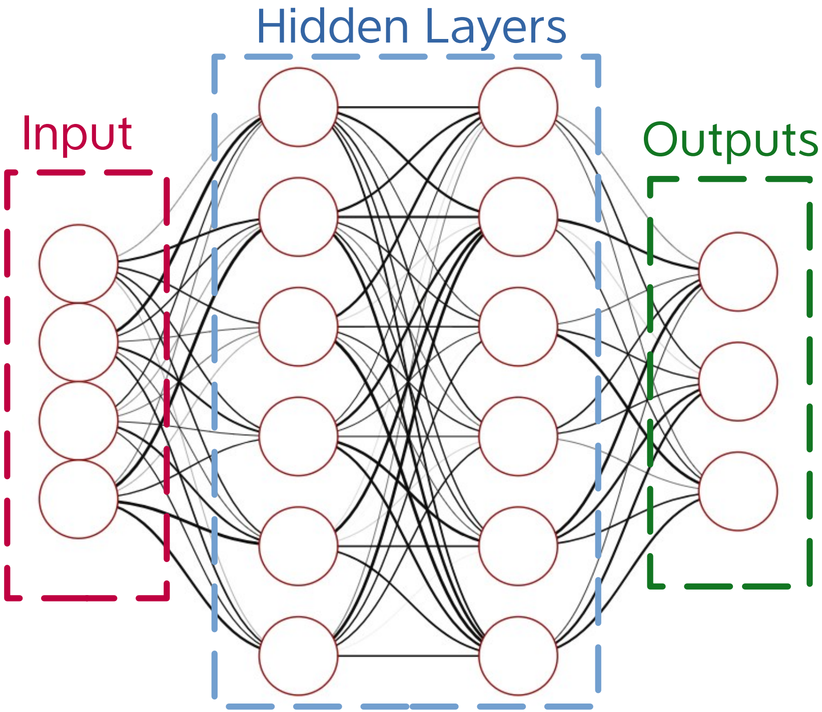


$$L(\theta) \propto \exp\left(-\frac{1}{2}\chi^2(\theta)\right)$$

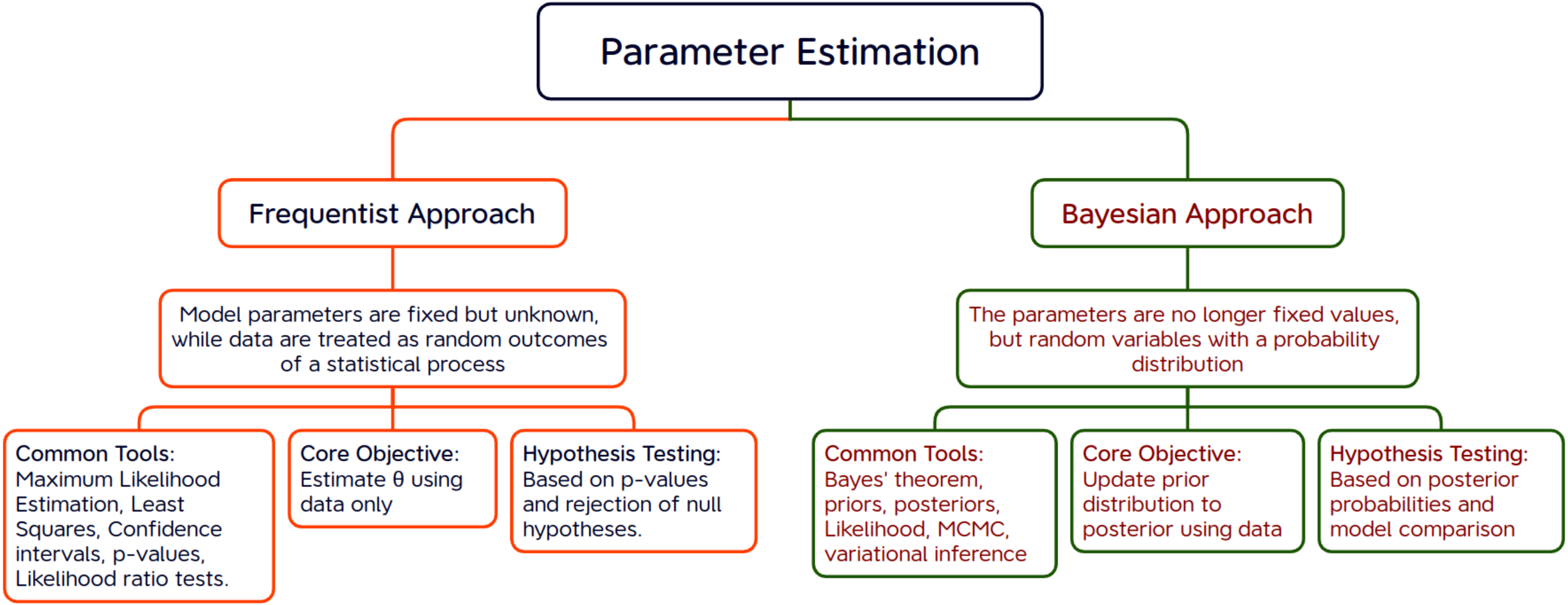

$L(D \mid \theta)$
θ_1
θ_2

Machine Learning Is Nothing But Optimization

Training Data
+
Labels



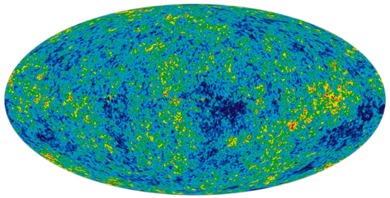
Labels → Loss Function ← Outputs


Parameter Estimation
Frequentist Approach
Model parameters are fixed but unknown, while data are treated as random outcomes of a statistical process
Common Tools: Maximum Likelihood Estimation, Least Squares, Confidence intervals, p-values, Likelihood ratio tests.
Core Objective: Estimate θ using data only
Hypothesis Testing: Based on p-values and rejection of null hypotheses.
Bayesian Approach
The parameters are no longer fixed values, but random variables with a probability distribution
Common Tools: Bayes' theorem, priors, posteriors, Likelihood, MCMC, variational inference
Core Objective: Update prior distribution to posterior using data
Hypothesis Testing: Based on posterior probabilities and model comparison

Aspect	Frequentist Approach	Bayesian Approach
Philosophy	Parameters are fixed but unknown constants; randomness comes from data.	Parameters are random variables with probability distributions.
Probability Interpretation	Long-run frequency of outcomes over repeated experiments.	Degree of belief or uncertainty about parameters.
Parameters (θ)	Treated as fixed, unknown quantities.	Treated as random variables with prior distributions.
Data	Random samples from a fixed distribution.	Used to update prior beliefs via Bayes' theorem.
Core Objective	Estimate θ using data only.	Update prior distribution to posterior using data.
Prior Information	Not used; inference is based solely on data.	Prior information is explicitly included in the analysis.
Hypothesis Testing	Based on p-values and rejection of null hypotheses.	Based on posterior probabilities and model comparison (e.g., Bayes factor).
Common Tools	MLE, confidence intervals, p-values, likelihood ratio tests.	Bayes' theorem, priors, posteriors, MCMC, variational inference.
Computation	Often simpler and faster; analytic solutions for many problems.	Can be computationally intensive, especially in high dimensions.
Interpretability	Some frequentist measures (e.g., p-values) are often misinterpreted.	Probabilistic interpretation aligns better with intuitive understanding.
Typical Use Cases	Classical experiments, large-sample statistics, regulated fields (e.g., pharma).	Complex models, small datasets, simulation-based inference, hierarchical models.

Traditional Bayesian Inference

Observation



Adopting Model

Estimating Parameters

Bayesian Approach

$$P(\theta|D)$$

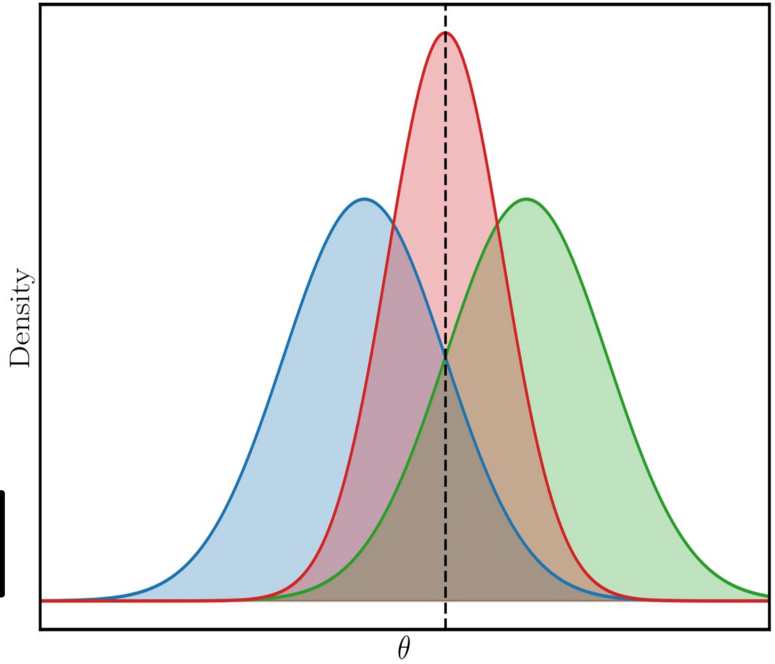
Posterior

Bayesian Theorem

$$P(\theta|D) \sim P(D|\theta) \times P(\theta)$$

Posterior Likelihood Prior

Likelihood Selection



Prior

Your initial belief about a parameter before seeing any data

Likelihood

Probability of observing the data given a specific parameter value

Posterior

Your updated belief about the parameter after seeing the data

$$\ln \mathcal{L}(D|\theta) = \frac{1}{2}\left[\left(D - D^{(th)}(\theta)\right)^T . C^{-1} . \left(D - D^{(th)}(\theta)\right)\right]$$

Gaussian Likelihood

Classical Bayesian methods require an **explicit analytical** form of how the data depends on the parameters in order to compute the likelihood

The Gaussian likelihood is often justified by the **Central Limit Theorem**

Analytical Solution

A closed-form approach to compute the posterior distribution exactly--applicable only when the prior and likelihood are mathematically tractable

Grid-based Methods

Discretize the parameter space into a grid and evaluate the posterior at each point -- Simple and intuitive -- Useful for low-dimensional problems -- Computationally expensive in high dimensions

Variational Inference

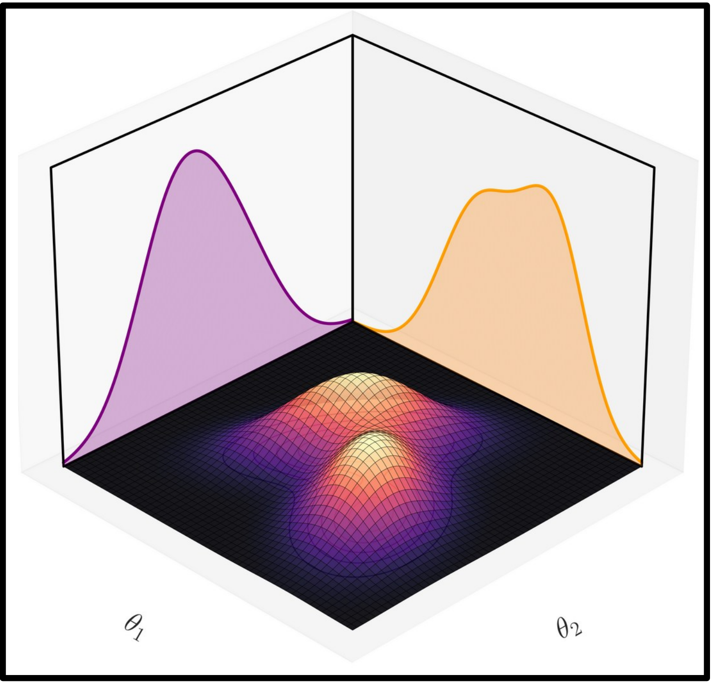
An approximate Bayesian inference method that replaces a complex posterior distribution with a simpler, parameterized distribution and then optimizes that approximation

Sampling-Based Methods

Sampling-based methods aim to approximate the posterior distribution by generating a large number of representative samples from it.

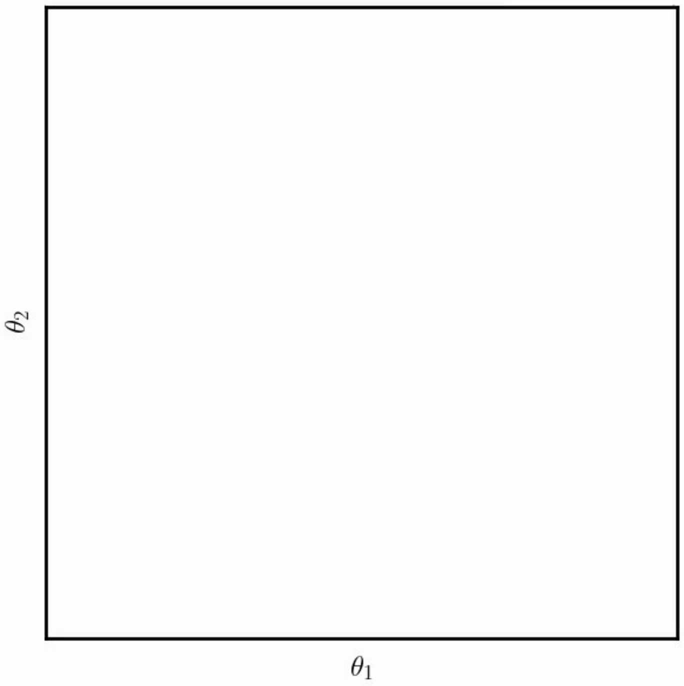
- **Rejection Sampling**
- **Importance Sampling**
- **MCMC Methods**
  - **Metropolis-Hastings**
  - **Gibbs Sampling**
  - **Hamiltonian Monte Carlo (HMC)**
  - **NUTS (No-U-Turn Sampler)**

Estimated Posterior



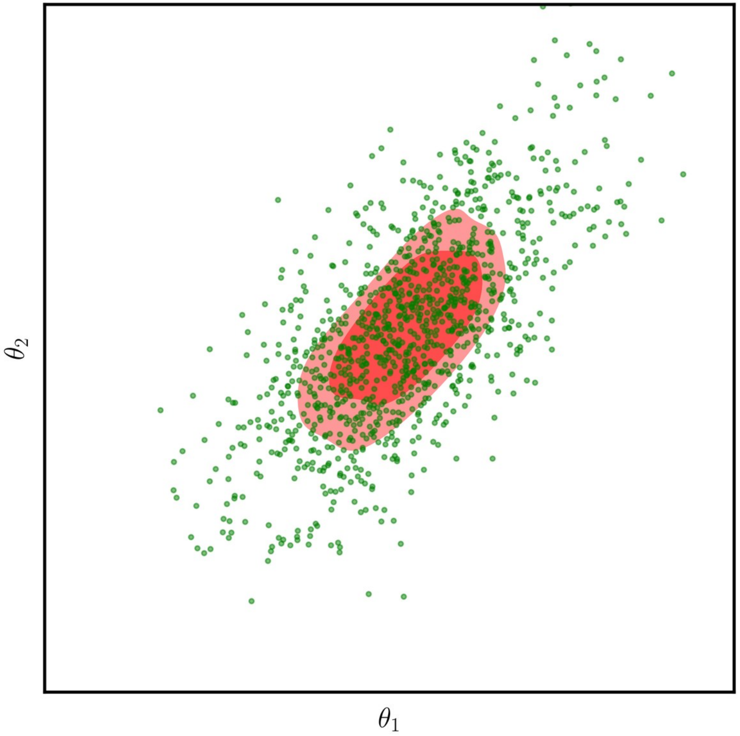
Metropolis-Hastings Markov Chain Monte Carlo (MCMC)

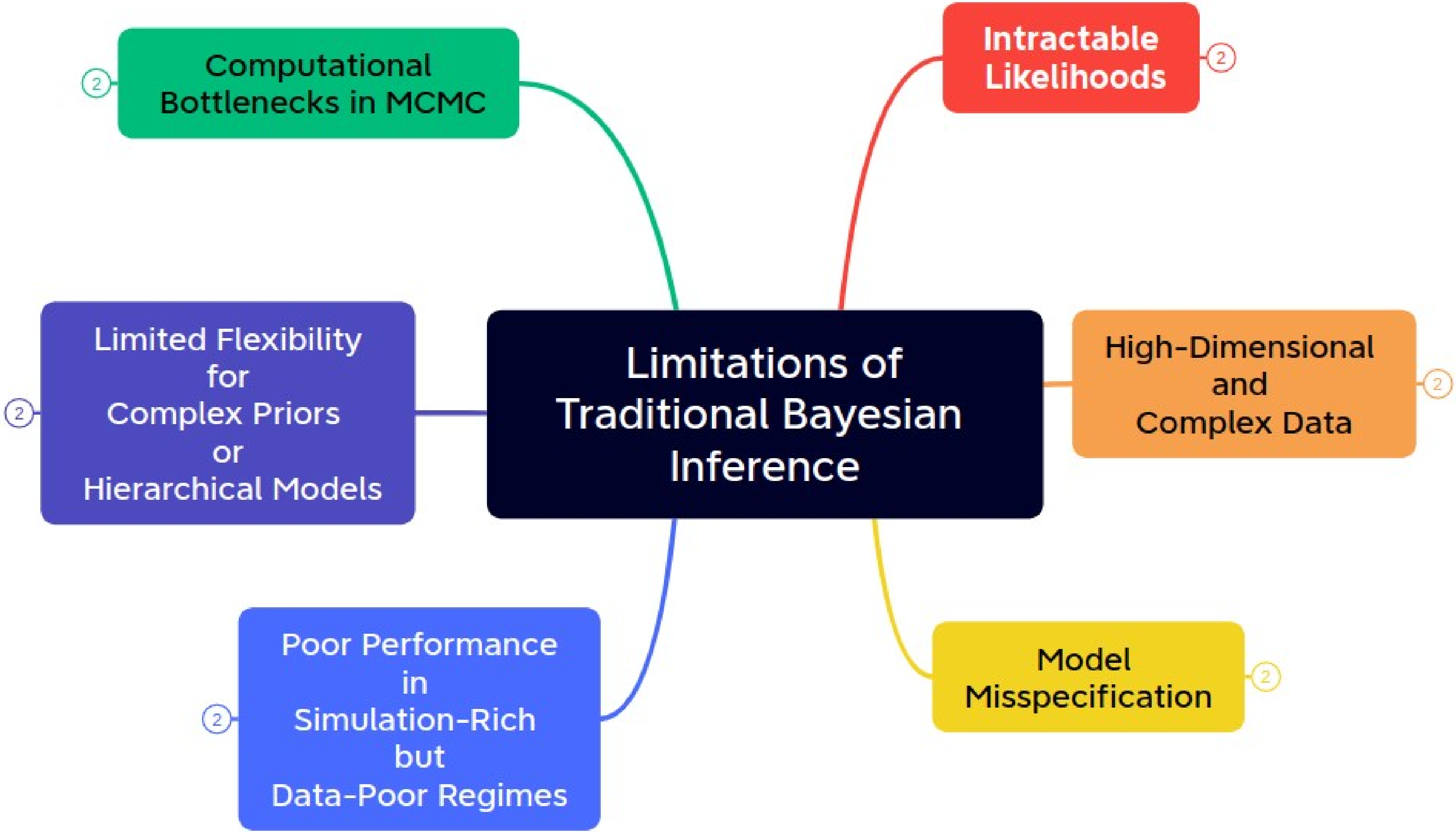
Import Data
Select initial parameters $\{\theta_{\text{old}}\}$
Compute $\chi^2_{\text{old}} = \chi^2(\{\theta_{\text{old}}\})$
Compute $L_{\text{old}} = e^{-\chi^2_{\text{old}}/2}$
while not converged **do**
    Select $\{\theta_{\text{new}}\}$ using a proposal based on $\{\theta_{\text{old}}\}$
    Compute $\chi^2_{\text{new}} = \chi^2(\{\theta_{\text{new}}\})$
    $\Delta\chi^2 \leftarrow \chi^2_{\text{new}} - \chi^2_{\text{old}}$
    Acceptance rate: $R = \min\left\{1,\ e^{-\Delta\chi^2/2}\right\}$
    Draw $r \sim \text{Uniform}(0, 1)$
    if $s \leq R$ **then**
        $\{\theta_{\text{old}}\} = \{\theta_{\text{new}}\}$
        $\chi^2_{\text{old}} = \chi^2_{\text{new}}$
    end if
    Write $\{\theta_{\text{old}}\},\ \chi^2_{\text{old}}$ to chain
end while

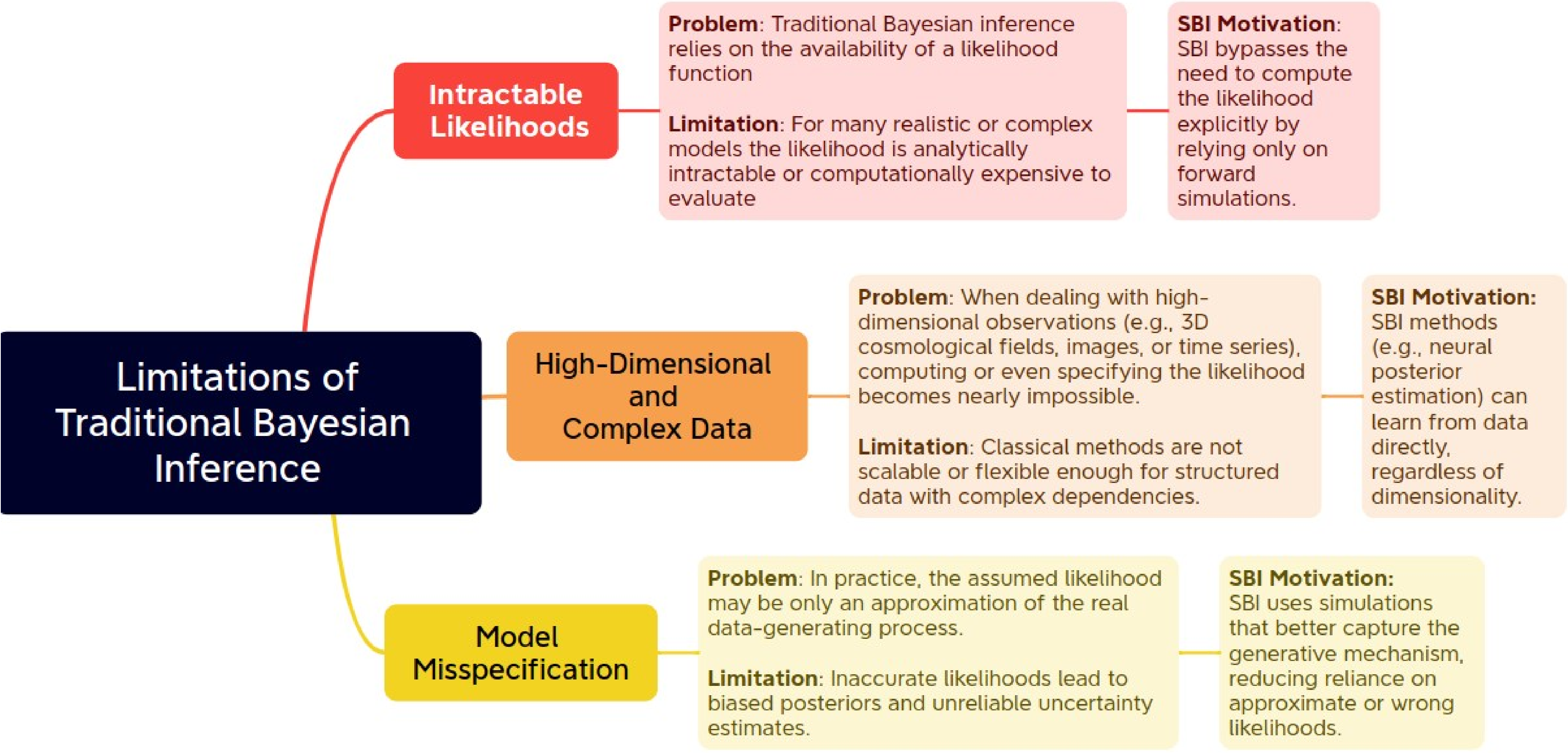


Metropolis-Hastings Markov Chain Monte Carlo (MCMC)

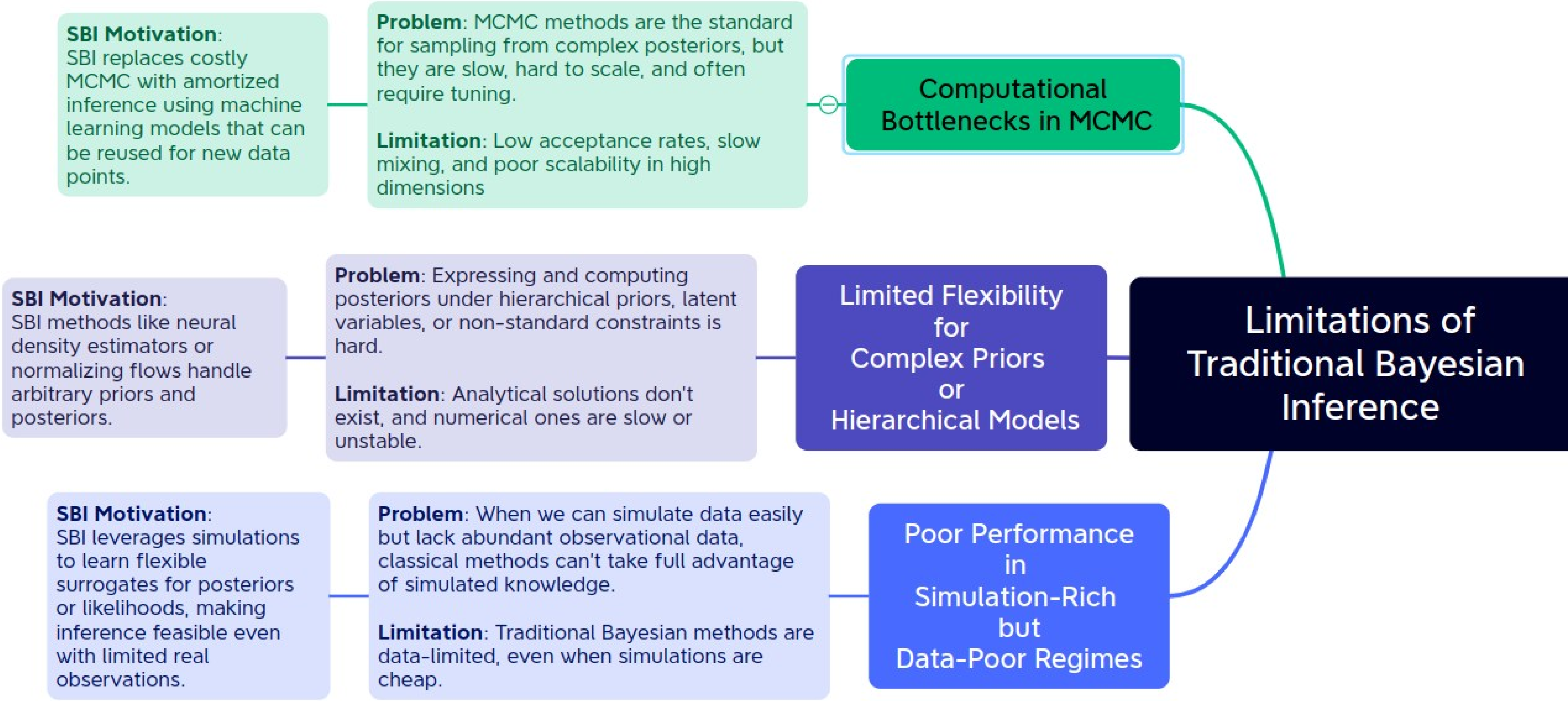
Import Data
Select initial parameters $\{\theta_{\rm old}\}$
Compute $\chi^2_{\rm old} = \chi^2(\{\theta_{\rm old}\})$
Compute $L_{\rm old} = e^{-\chi^2_{\rm old}/2}$
while not converged **do**
    Select $\{\theta_{\rm new}\}$ using a proposal based on $\{\theta_{\rm old}\}$
    Compute $\chi^2_{\rm new} = \chi^2(\{\theta_{\rm new}\})$
    $\Delta\chi^2 \leftarrow \chi^2_{\rm new} - \chi^2_{\rm old}$
    Acceptance rate: $R = \min\left\{1,\ e^{-\Delta\chi^2/2}\right\}$
    Draw $r \sim \mathrm{Uniform}(0,1)$
    if $s \leq R$ **then**
        $\{\theta_{\rm old}\} = \{\theta_{\rm new}\}$
        $\chi^2_{\rm old} = \chi^2_{\rm new}$
    end if
    Write $\{\theta_{\rm old}\},\ \chi^2_{\rm old}$ to chain
end while



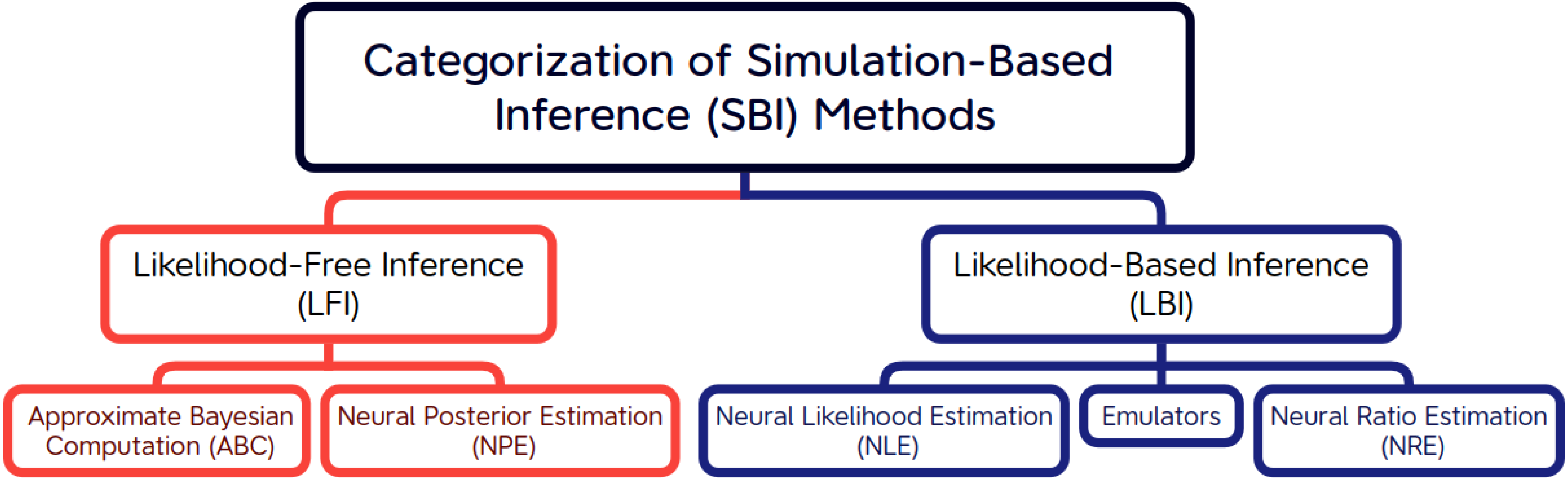

Limitations of Traditional Bayesian Inference
Computational Bottlenecks in MCMC
Intractable Likelihoods
Limited Flexibility for Complex Priors or Hierarchical Models
High-Dimensional and Complex Data
Poor Performance in Simulation-Rich but Data-Poor Regimes
Model Misspecification
2
2
2
2
2
2



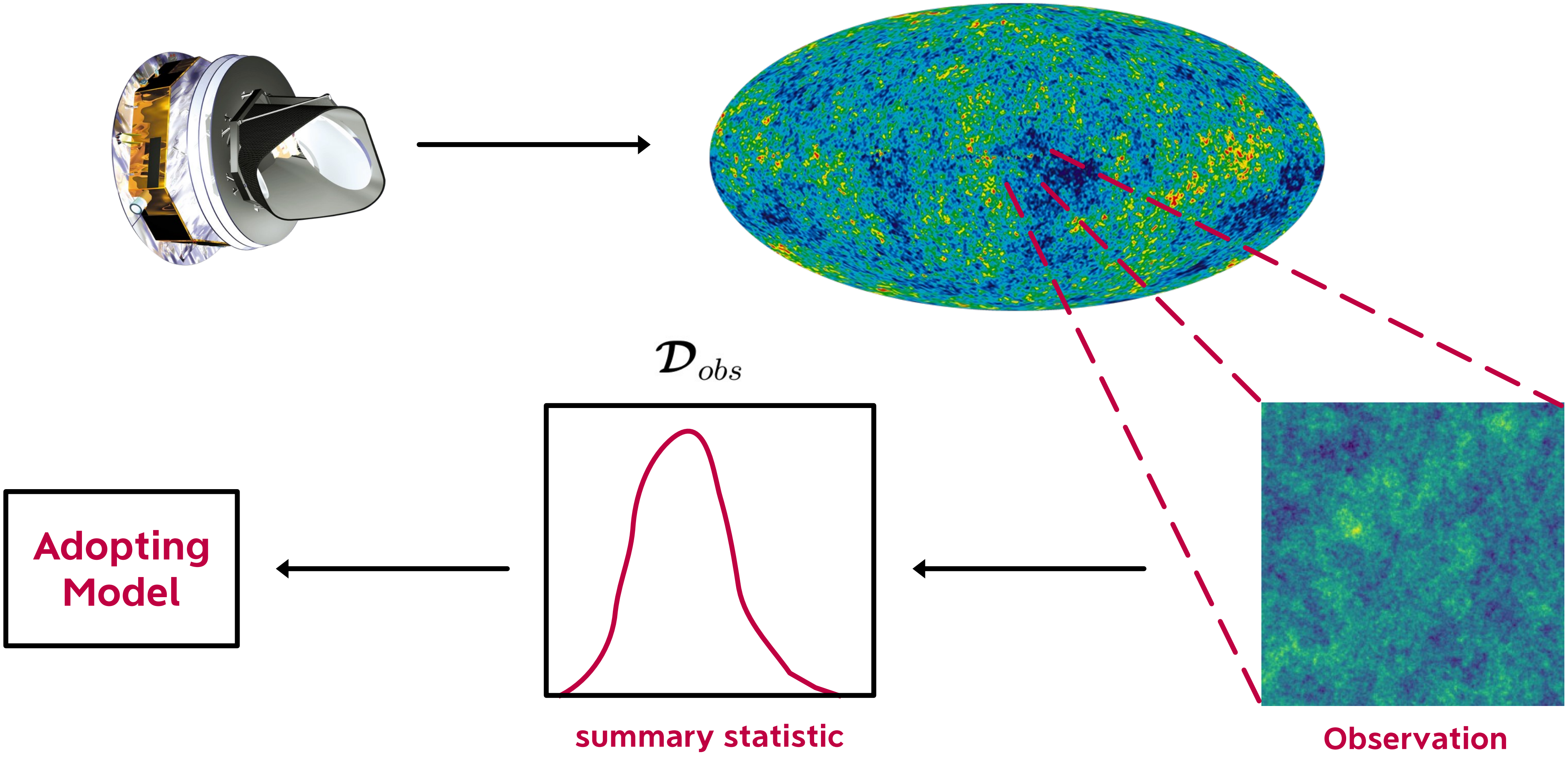
Limitations of Traditional Bayesian Inference
Intractable Likelihoods
Problem: Traditional Bayesian inference relies on the availability of a likelihood function
Limitation: For many realistic or complex models the likelihood is analytically intractable or computationally expensive to evaluate
SBI Motivation: SBI bypasses the need to compute the likelihood explicitly by relying only on forward simulations.
High-Dimensional and Complex Data
Problem: When dealing with high-dimensional observations (e.g., 3D cosmological fields, images, or time series), computing or even specifying the likelihood becomes nearly impossible.
Limitation: Classical methods are not scalable or flexible enough for structured data with complex dependencies.
SBI Motivation: SBI methods (e.g., neural posterior estimation) can learn from data directly, regardless of dimensionality.
Model Misspecification
Problem: In practice, the assumed likelihood may be only an approximation of the real data-generating process.
Limitation: Inaccurate likelihoods lead to biased posteriors and unreliable uncertainty estimates.
SBI Motivation: SBI uses simulations that better capture the generative mechanism, reducing reliance on approximate or wrong likelihoods.


Limitations of Traditional Bayesian Inference
Computational Bottlenecks in MCMC
Problem: MCMC methods are the standard for sampling from complex posteriors, but they are slow, hard to scale, and often require tuning.
Limitation: Low acceptance rates, slow mixing, and poor scalability in high dimensions
SBI Motivation: SBI replaces costly MCMC with amortized inference using machine learning models that can be reused for new data points.
Limited Flexibility for Complex Priors or Hierarchical Models
Problem: Expressing and computing posteriors under hierarchical priors, latent variables, or non-standard constraints is hard.
Limitation: Analytical solutions don't exist, and numerical ones are slow or unstable.
SBI Motivation: SBI methods like neural density estimators or normalizing flows handle arbitrary priors and posteriors.
Poor Performance in Simulation-Rich but Data-Poor Regimes
Problem: When we can simulate data easily but lack abundant observational data, classical methods can't take full advantage of simulated knowledge.
Limitation: Traditional Bayesian methods are data-limited, even when simulations are cheap.
SBI Motivation: SBI leverages simulations to learn flexible surrogates for posteriors or likelihoods, making inference feasible even with limited real observations.

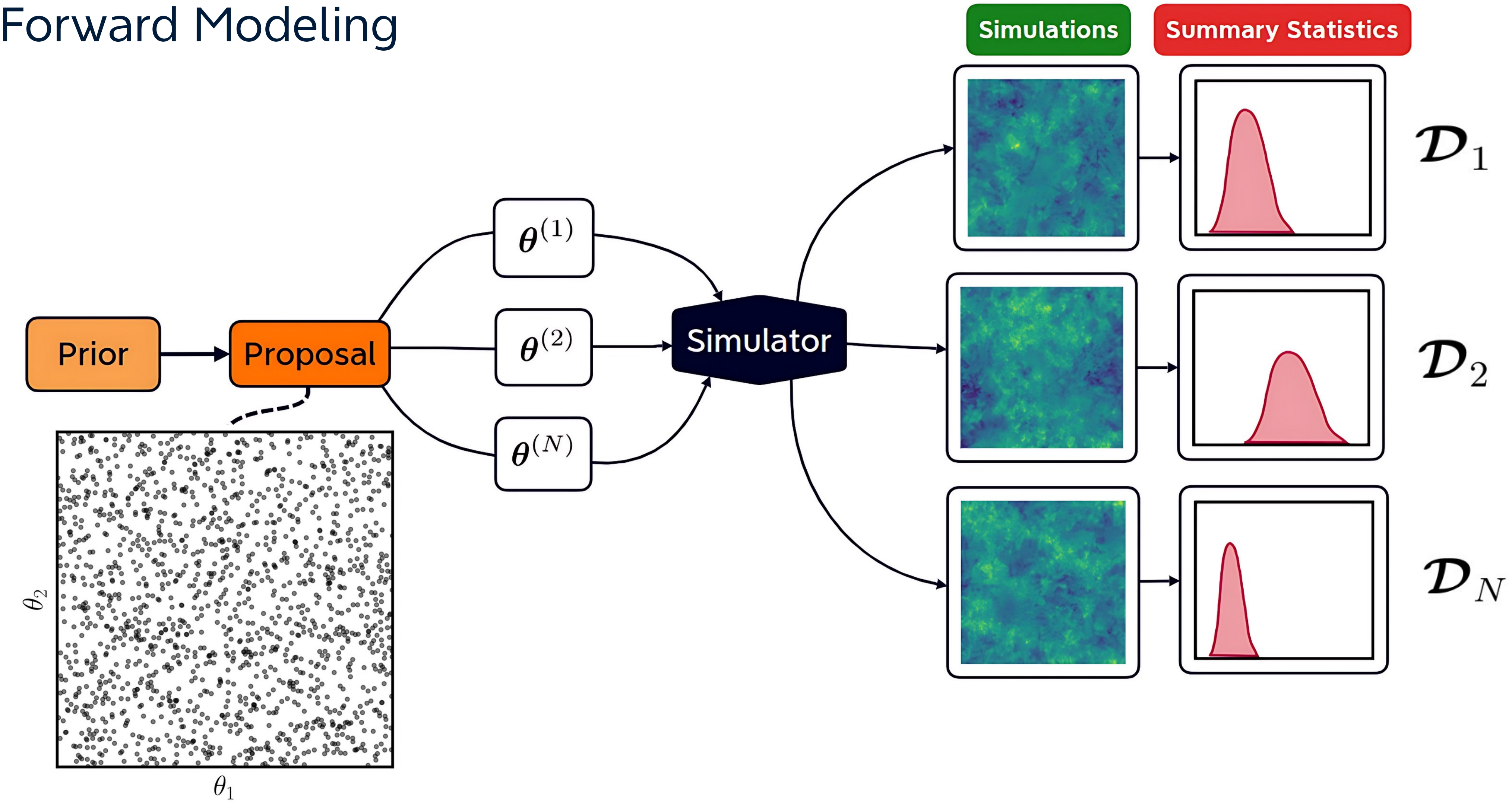
Diving into the SBI Framework !


Categorization of Simulation-Based Inference (SBI) Methods
Likelihood-Free Inference (LFI)
Likelihood-Based Inference (LBI)
Approximate Bayesian Computation (ABC)
Neural Posterior Estimation (NPE)
Neural Likelihood Estimation (NLE)
Emulators
Neural Ratio Estimation (NRE)

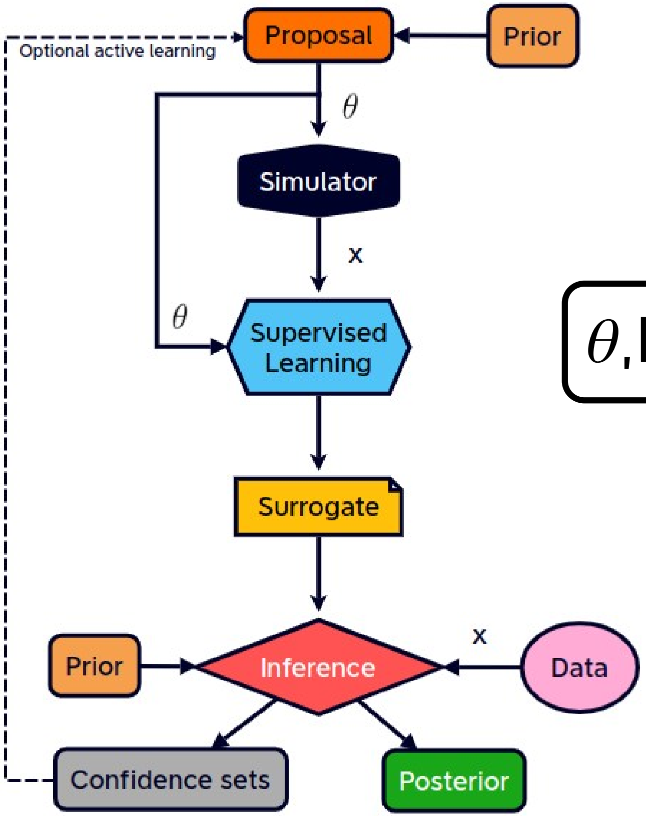
Starting Point to SBI

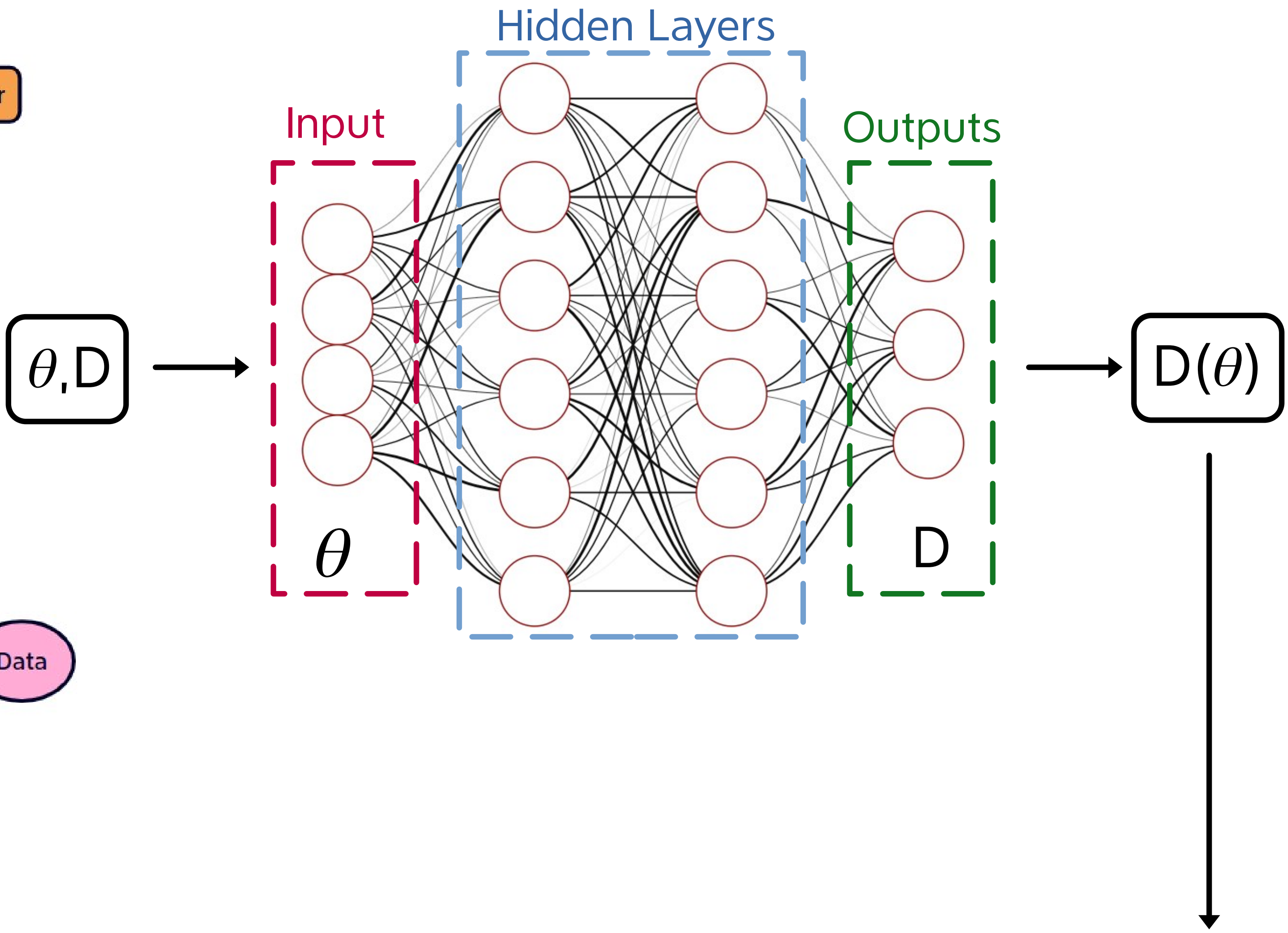


Forward Modeling

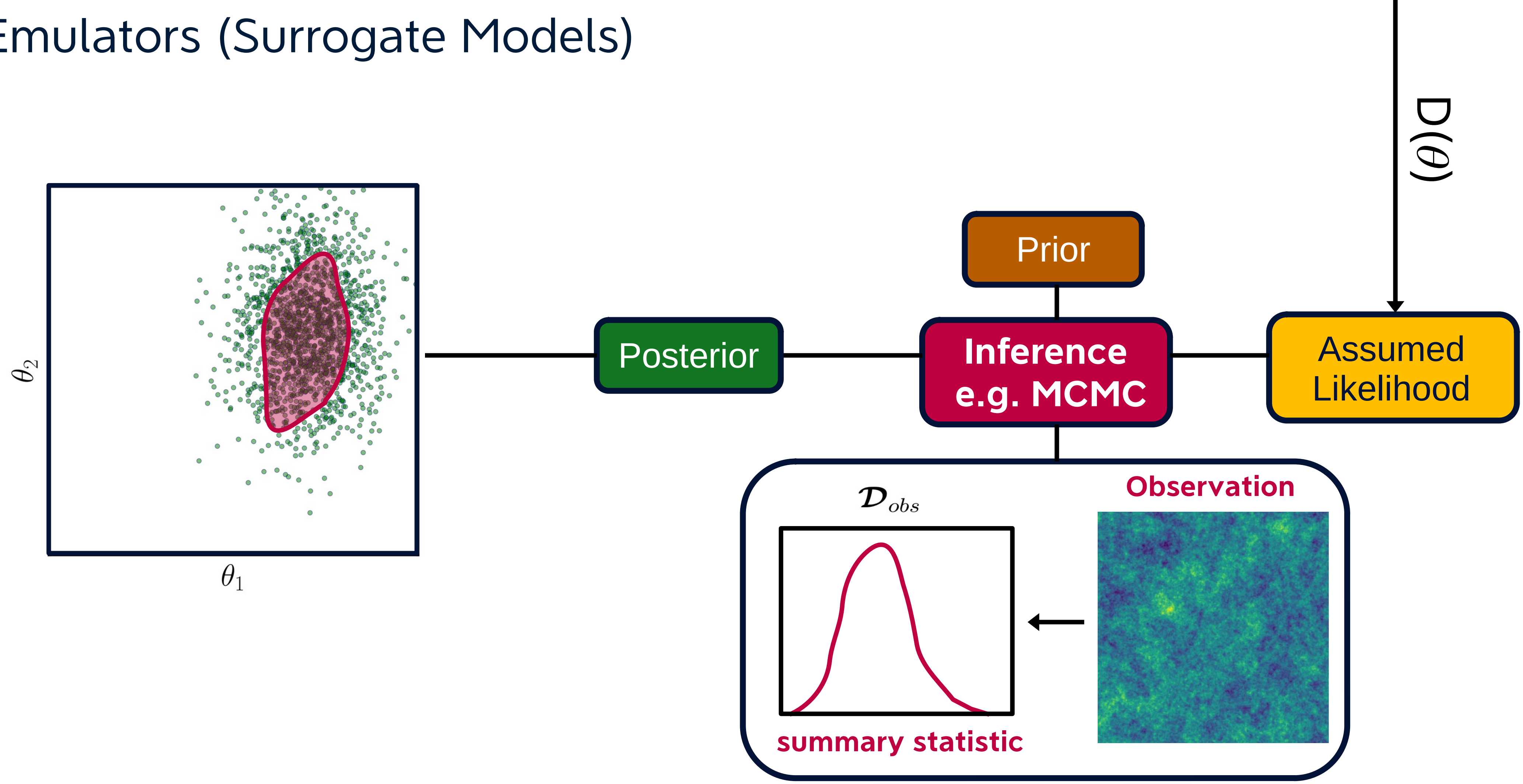


Emulators (Surrogate Models)





Emulators (Surrogate Models)



سپاس از توجه شما