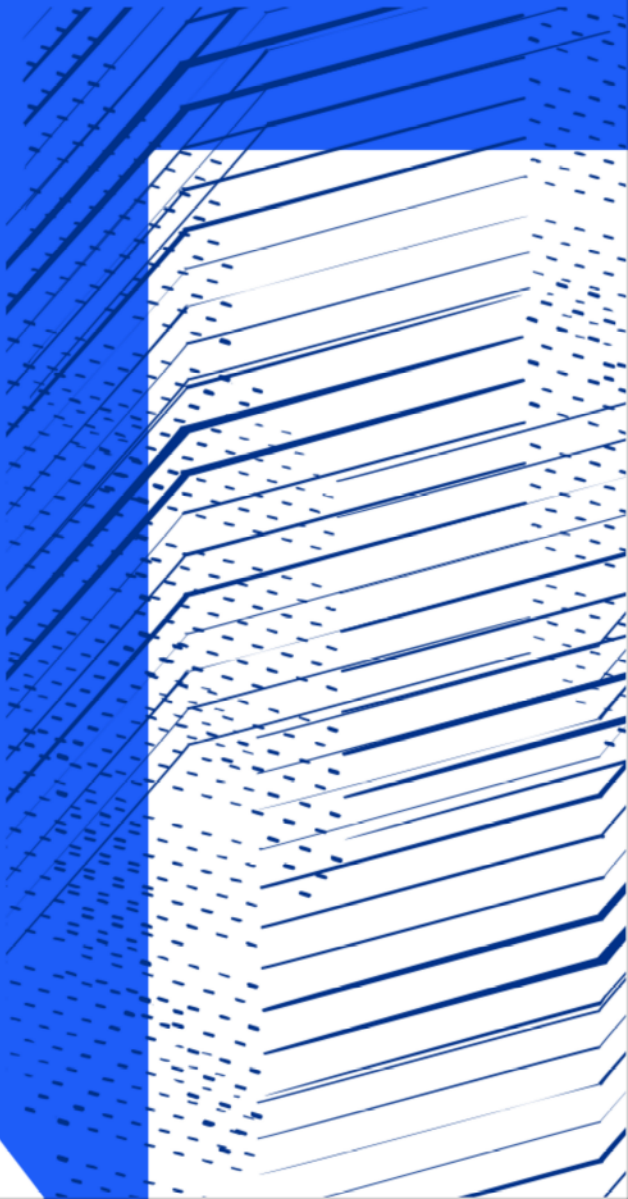
Science and Technology Facilities Council

# Stochastic Model Fitting

Taking methods back to physical basics



Science and Technology Facilities Council

# Agenda

## 1 Goals

The what and why.

## 2 Methods

## 3 Procedures

Ensuring and documenting progress


Science and Technology Facilities Council

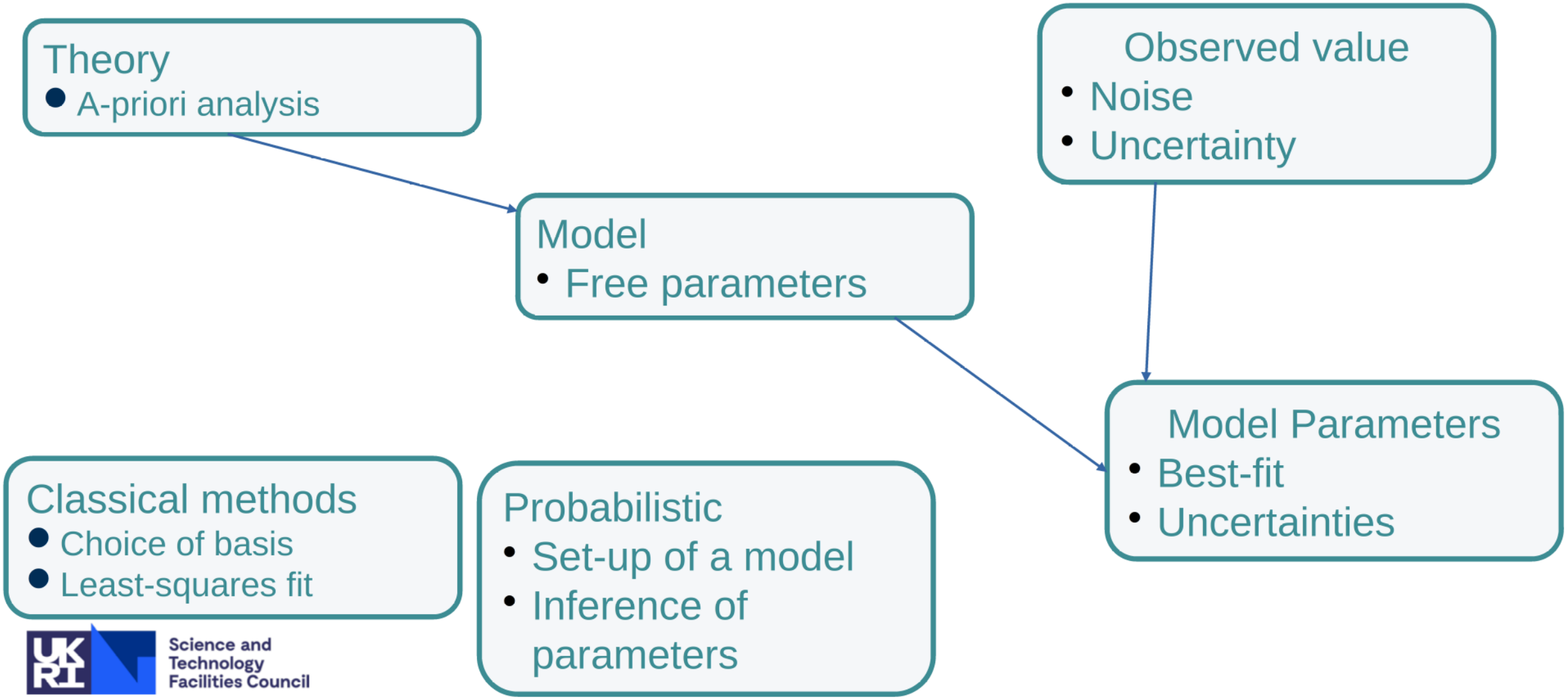
Science and Technology Facilities Council

# The Overarching Goal

What use case do we consider?

Science and Technology Facilities Council

# Fitting probabilistic models to data




Science and Technology Facilities Council

# Inference from measurements – Bayes' rule

Probability of B occuring if A has occured – given a model

$$P\left[B_i \middle| A\right] = \frac{P\left[A \middle| B_i\right] P\left[B_i\right]}{\sum_{j=1}^{n} P\left[A \middle| B_j\right] P\left[B_j\right]}$$

The same, but with densities

$$p_{X|Y}\left(x \middle| y\right) = \frac{p_X\left(x\right) p_{Y|X}\left(y \middle| x\right)}{p_Y\left(y\right)}$$

Not solvable by hand for a n arrow's flight

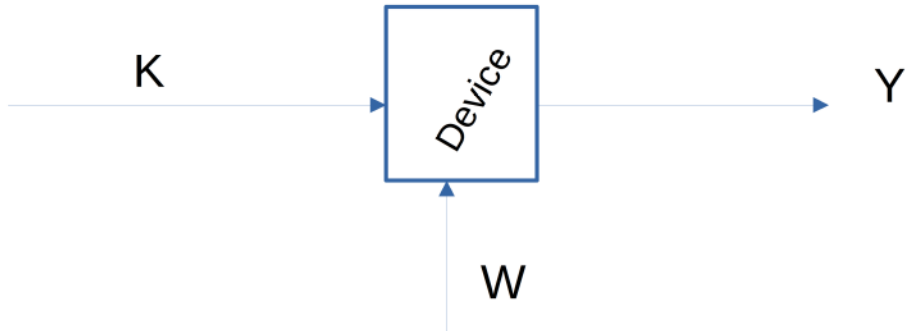
$$Z_i = Z_0 + V_0\left(t - t_0\right) + A_0\left(t - t_0\right)^2 + W_i$$

$$p_{(Z,V,A)|(z_1,\ldots,z_n)}\left((z,v,a)\middle|(t_1,\ldots,t_n)\right) = \frac{\prod_{i=1}^{3} p_{(Z,V,A)_i}\left((z,v,a)_i\right) \prod_{j=1}^{N} p_{Z_i|(Z,V,A)}\left(z_i \middle| (z,v,a)\right)}{p_{(Z,V,A)}\left((z,v,a)\right)}$$

Science and Technology Facilities Council

# Bayes' rule example: Noisy digital output

## Discrete signal with noise




- Input: signal value K equally likely to be +1 or -1
- Additive, normally distributed noise W
- Measure Y = K + W
- **Given measurement y what is the p'bility of the signal K being 1?**

$$p_K(k) = \frac{1}{2}$$

$$p_{Y|K}(y|k) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(y-k)^2}$$

$$p_Y(y) = \frac{1}{2\sqrt{2\pi}} e^{\frac{-1}{2}(y+1)^2} + \frac{1}{2\sqrt{2\pi}} e^{\frac{-1}{2}(y-1)^2}$$

$$p_{K|Y}(1|y) = \frac{1}{1+e^{-2y}}$$

Science and Technology Facilities Council

Science and Technology Facilities Council

# Methods

What will we be developing?

Science and Technology Facilities Council

# Methods to fit a model to existing data

**Deterministic**

- Polynomial fitting
- Polynomial interpolation
- Trigonometric Interpolation

**Statistics**

- Machine Learning
- Bayesian Inference
- Markov Chain methods

Determining the minimum/ground state of an unknown potential

Science and Technology Facilities Council

# Problems of the common methods

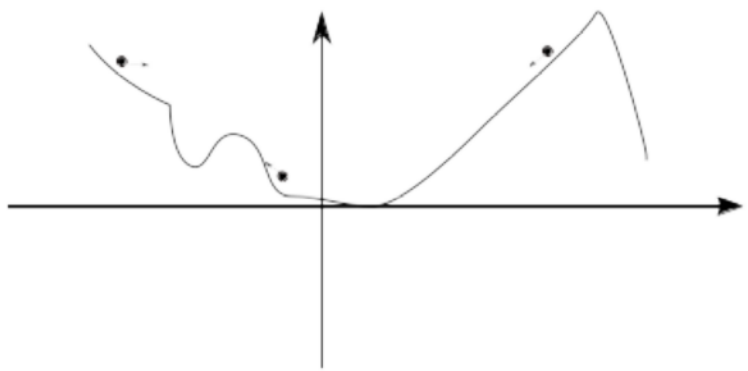
**Deterministic**

- Overfitting
- Numerical stability
- Computational cost

**Statistics**

- Large datasets
- Slow convergence
- Computational cost

**Solution:** Consider the problem as an energy minimization problem
Finding the minimum of an arbitrary potential.



Science and Technology Facilities Council

# Solving parameter estimation by finding the ground state of a Hamiltonian

$$H(q,p) = T(p) + V(q) = \sum p_i^2/(2m_i) + V(q_i)$$

## Random Walk

- Direction and step size chosen at random
- Step (almost) always accepted

## Markov Chain MC

- Fixed step size
- Direction chosen at random
- Metropolis-Hastings for acceptance

## Hamiltonian MC

- Sampling trajectories in phase space
- Extensive propagation
- Metropolis-Hastings / Gibbs

Science and Technology Facilities Council

# Hamiltonian Monte Carlo – the method of choice

$$H(q,p) = T(p) + V(q) = \sum p_i^2/(2m_i) + V(q_i)$$

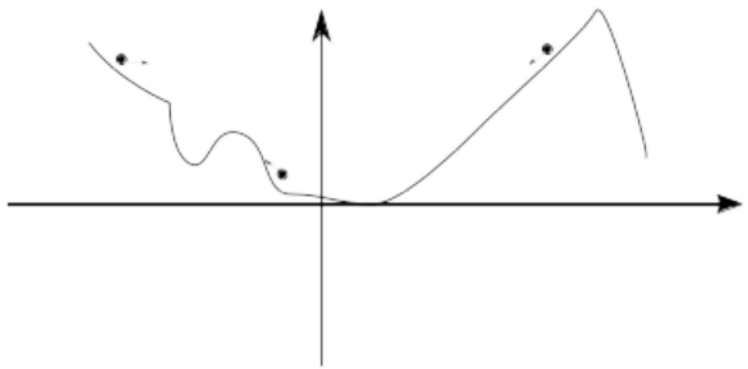
**Algorithm 1** Hamiltonian Monte Carlo

Given $\theta^0$, $\epsilon$, $L$, $\mathcal{L}$, $M$:
**for** $m = 1$ to $M$ **do**
  Sample $r^0 \sim \mathcal{N}(0, I)$.
  Set $\theta^m \leftarrow \theta^{m-1}, \tilde{\theta} \leftarrow \theta^{m-1}, \tilde{r} \leftarrow r^0$.
  **for** $i = 1$ to $L$ **do**
    Set $\tilde{\theta}, \tilde{r} \leftarrow \text{Leapfrog}(\tilde{\theta}, \tilde{r}, \epsilon)$.
  **end for**
  With probability $\alpha = \min\left\{1, \frac{\exp\{\mathcal{L}(\tilde{\theta}) - \frac{1}{2}\tilde{r}\cdot\tilde{r}\}}{\exp\{\mathcal{L}(\theta^{m-1}) - \frac{1}{2}r^0\cdot r^0\}}\right\}$, set $\theta^m \leftarrow \tilde{\theta}$, $r^m \leftarrow -\tilde{r}$.
**end for**

**function** Leapfrog$(\theta, r, \epsilon)$
Set $\tilde{r} \leftarrow r + (\epsilon/2)\nabla_\theta \mathcal{L}(\theta)$.
Set $\tilde{\theta} \leftarrow \theta + \epsilon\tilde{r}$.
Set $\tilde{r} \leftarrow \tilde{r} + (\epsilon/2)\nabla_\theta \mathcal{L}(\tilde{\theta})$.
**return** $\tilde{\theta}, \tilde{r}$.

$$P_C[(q,p)] = \frac{1}{Z} e^{-\beta H(q,p)}$$

$$Z = \sum e^{-\beta H(q_i, p_i)}$$



Science and Technology Facilities Council

# Hamiltonian Monte Carlo – problems

$$H(q,p) = T(p) + V(q) = \sum p_i^2/(2m_i) + V(q_i)$$

**Algorithm 1** Hamiltonian Monte Carlo

Given $\theta^0$, $\epsilon$, $L$, $\mathcal{L}$, $M$:
**for** $m = 1$ to $M$ **do**
  Sample $r^0 \sim \mathcal{N}(0, I)$.
  Set $\theta^m \leftarrow \theta^{m-1}, \tilde{\theta} \leftarrow \theta^{m-1}, \tilde{r} \leftarrow r^0$.
  **for** $i = 1$ to $L$ **do**
    Set $\tilde{\theta}, \tilde{r} \leftarrow \text{Leapfrog}(\tilde{\theta}, \tilde{r}, \epsilon)$.
  **end for**
  With probability $\alpha = \min\left\{1, \frac{\exp\{\mathcal{L}(\tilde{\theta}) - \frac{1}{2}\tilde{r}\cdot\tilde{r}\}}{\exp\{\mathcal{L}(\theta^{m-1}) - \frac{1}{2}r^0\cdot r^0\}}\right\}$, set $\theta^m \leftarrow \tilde{\theta}$, $r^m \leftarrow -\tilde{r}$.
**end for**

**function** Leapfrog$(\theta, r, \epsilon)$
Set $\tilde{r} \leftarrow r + (\epsilon/2)\nabla_\theta \mathcal{L}(\theta)$.
Set $\tilde{\theta} \leftarrow \theta + \epsilon\tilde{r}$.
Set $\tilde{r} \leftarrow \tilde{r} + (\epsilon/2)\nabla_\theta \mathcal{L}(\tilde{\theta})$.
**return** $\tilde{\theta}, \tilde{r}$.

$$P_C[(q,p)] = \frac{1}{Z} e^{-\beta H(q,p)}$$

$$Z = \sum e^{-\beta H(q_i, p_i)}$$

**Methodological**

- Step size
- Number of steps
- Initial conditions ( position, momentum)

**Pedagogical**

- Mathematical mutilation

**Computational**

- Stability dictated by integrator
- Expensive evaluation of potentials
- Parameter ranging
- Mediocre scalablity
  - Synchronisation for global averages

Science and Technology Facilities Council

# Ensemble-based extension – and other improvements

$$H(q,p) = T(p) + V(q) = \sum p_i^2/(2m_i) + V(q_i)$$

## Ensemble-based HMC

- μ-canonical ensemble
- Ensemble-averages
- Divide-and-conquer

## Extensions

- GPU offloading
- Variational Quantum Encoding
- Simlated/Quantum Annealing

## Physical extensions

- Canonical ensemble
- Variable step-size
- Optimal step-size choice
- Initialization
- Max-entropy principle

Science and Technology Facilities Council

Science and Technology Facilities Council

# Procedures

The technical part.

Science and Technology Facilities Council

# Meetings

- **Times:**
  - **Monday** 0900BST / 1000 CEST – for 1 h
  - **Thursday** 1400 BST / 1500 CEST – for 30 min – 1h
  - Upon request.
- **Mode**
  - Video conference via Zoom.

Science and Technology Facilities Council

# Basic tools – GIT, Python

Programming:
- Python (possibly C++)
- NumPy/SciPy
- **NumPyro**
  (https://num.pyro.ai/en/stable/index.html)

GIT rules:
- **No binaries**
- **Pull-Request only**
- **No acceptance without a test**
- Questions as issues.
- At least one approval required before merge.
- Method summaries as TeX files.

Collaboration / communication
- GIT repository
- Email + VC

Science and Technology Facilities Council

# Theoretical basics – literature.

**Physics**:

- Statistical physics and thermodynamics
  - Lecture notes by M. Sigrist (ETH Zürich Fall term 2012)
  - Elementary Statistical Physics (C. Kittel, Dover publications)
- Classical mechanics
  - Lecture notes on Classical Dynamics (D. Tong, U Cambridge)
  - Theoretical Physics I (L.D. Landau)
  - Fundamentals of Mechanics (J. Marsden)

**Mathematics**:

- Mathematical Foundations of Statistical Mechanics (A. Khinchin, Dover publications)
- „Efficient Sequential Monte-Carlo
- Samplers for Bayesian Inference“ (Thi Nguyen et al., IEEE transactions on signal processing 64, 2016)
- „The No-U-Turn sampler“ (Hoffman et al, Journal of Machine Learning Research 15, 2014)

Methods:

- Monte Carlo Simulation in Statistical Physics (Binder, Heerman, Springer GTP)

UKRI Science and Technology Facilities Council

Science and Technology Facilities Council

# Questions?

Science and Technology Facilities Council