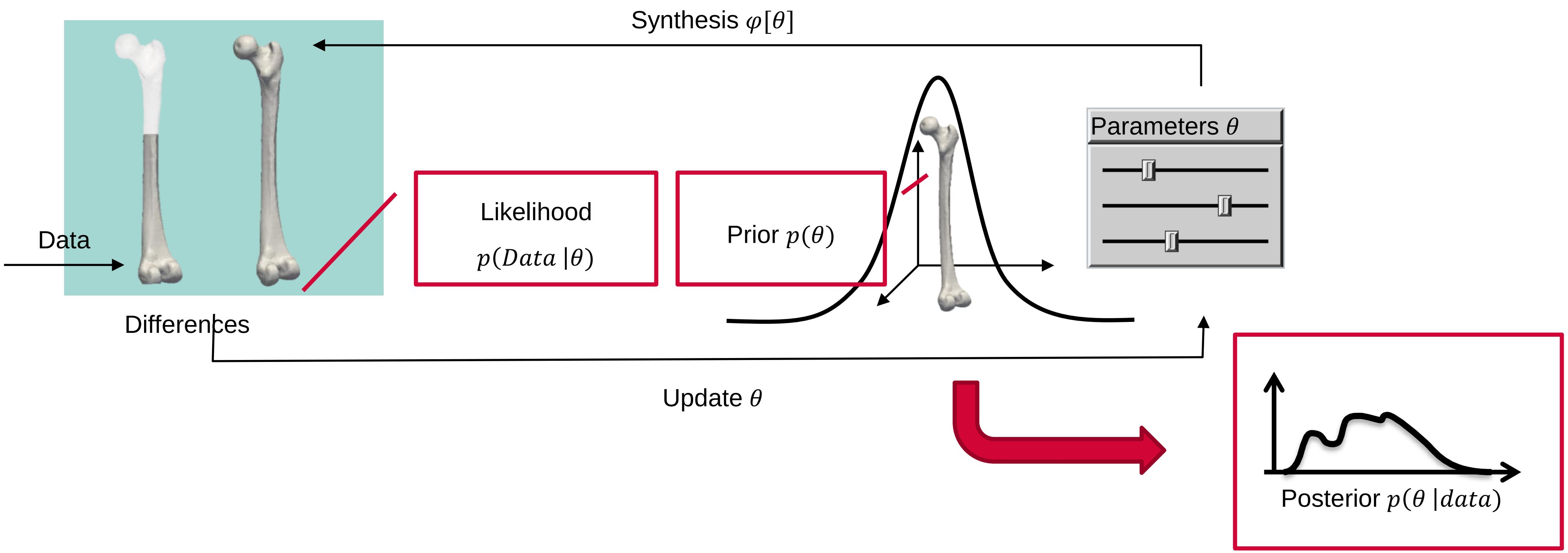
University
of Basel

The Metropolis-Hastings algorithm

Marcel Lüthi, Departement of Mathematics and Computer Science, University of Basel

Analysis by Synthesis and Bayesian Inference



Computational problem

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta}$$

Analysis by synthesis – Computational problem

- Non-linear
- Impossible to evaluate directly
- Can only be approximated

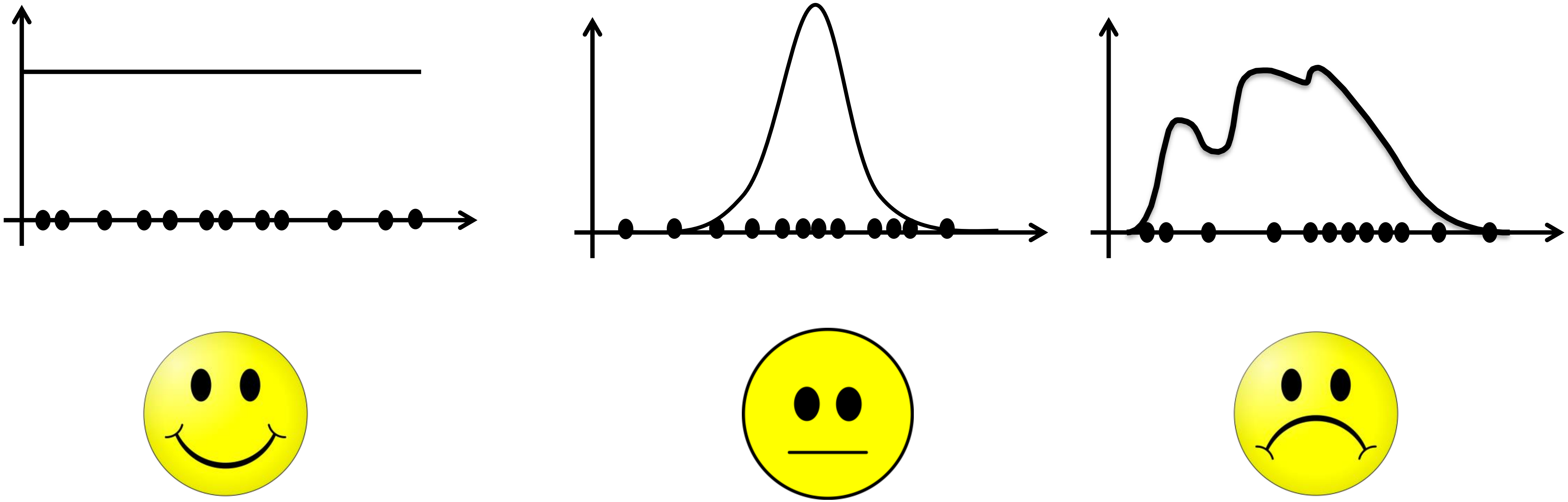
$$p(\theta|\mathrm{D}) = \frac{p(\theta)p(\mathrm{D}|\theta)}{\int p(\theta)p(D|\theta)d\theta}$$

- Usually non-linear
- Expensive to evaluate
- Can only be evaluated pointwise

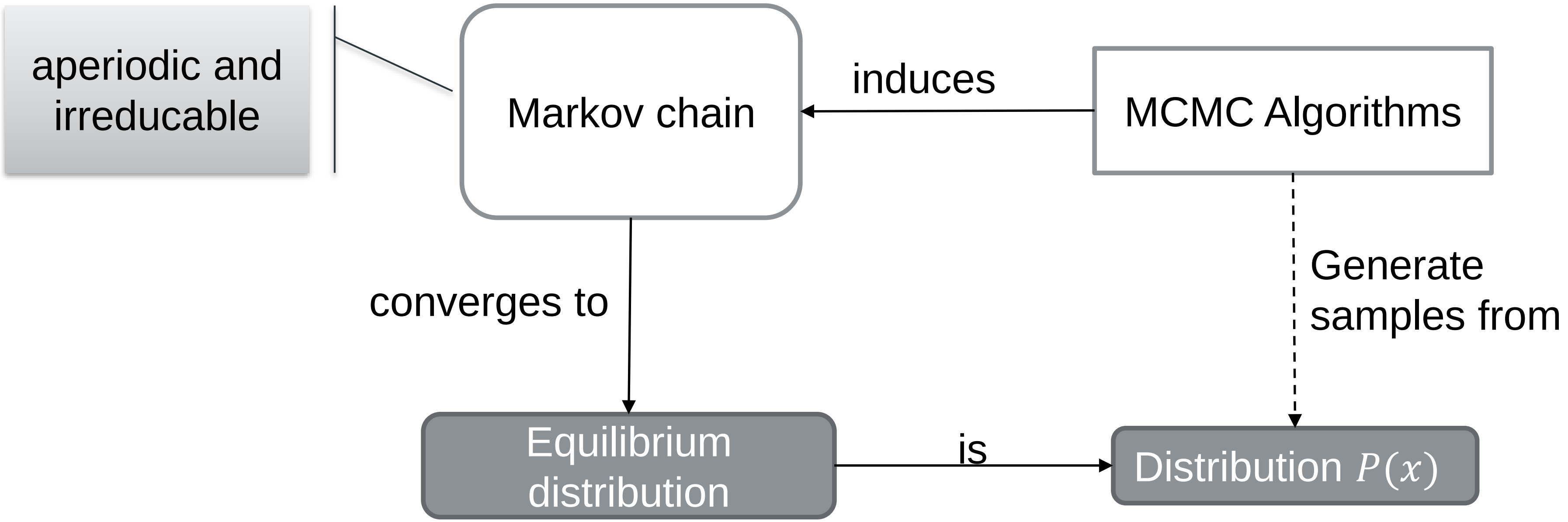
Image Source: Wikipedia
https://commons.wikimedia.org/wiki/File:Medical_X-Ray_imaging_IYN05_nevit.jpg

High dimensional

Drawing samples from a distribution



Markov Chain Monte Carlo Methods



Concept of Markov Chains: *“Use an already existing sample to produce the next one”*

The Metropolis Algorithm

Requirements:

- Proposal distribution $Q(\boldsymbol{x}'|\boldsymbol{x})$ – *must generate samples, symmetric*
- Target distribution $P(\boldsymbol{x})$ – *with point-wise evaluation*

Result:

- Stream of samples approximately from $P(\boldsymbol{x})$

Initialize with sample $\boldsymbol{x}$

Generate next sample, with current sample $\boldsymbol{x}$

1. Draw a sample $\boldsymbol{x}'$ from $Q(\boldsymbol{x}'|\boldsymbol{x})$ (“proposal”)
2. With *probability* $\alpha = \min\left\{\frac{P(\boldsymbol{x}')}{P(\boldsymbol{x})}, 1\right\}$ accept $\boldsymbol{x}'$ as new state $\boldsymbol{x}$
3. Emit current state $\boldsymbol{x}$ as sample

Example: 2D Gaussian

Target: $P(\boldsymbol{x}) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}$

Proposal: $Q(\boldsymbol{x}'|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}'|\boldsymbol{x}, \sigma^2 I_2)$ — Random walk

Target

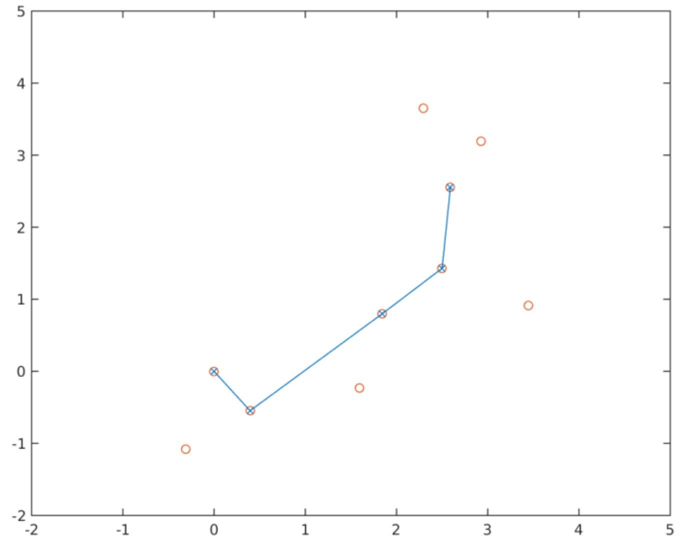
$\mu = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix}$

$\Sigma = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$

Sampled Estimate

$\hat{\mu} = \begin{bmatrix} 1.56 \\ 1.68 \end{bmatrix}$

$\hat{\Sigma} = \begin{bmatrix} 1.09 & 0.63 \\ 0.63 & 1.07 \end{bmatrix}$



The Metropolis-Hastings Algorithm

Generalization of Metropolis algorithm to asymmetric proposal distribution

$$Q(\boldsymbol{x}'|\boldsymbol{x}) \neq Q(\boldsymbol{x}|\boldsymbol{x}')$$

$$Q(\boldsymbol{x}'|\boldsymbol{x}) > 0 \Leftrightarrow Q(\boldsymbol{x}|\boldsymbol{x}') > 0$$

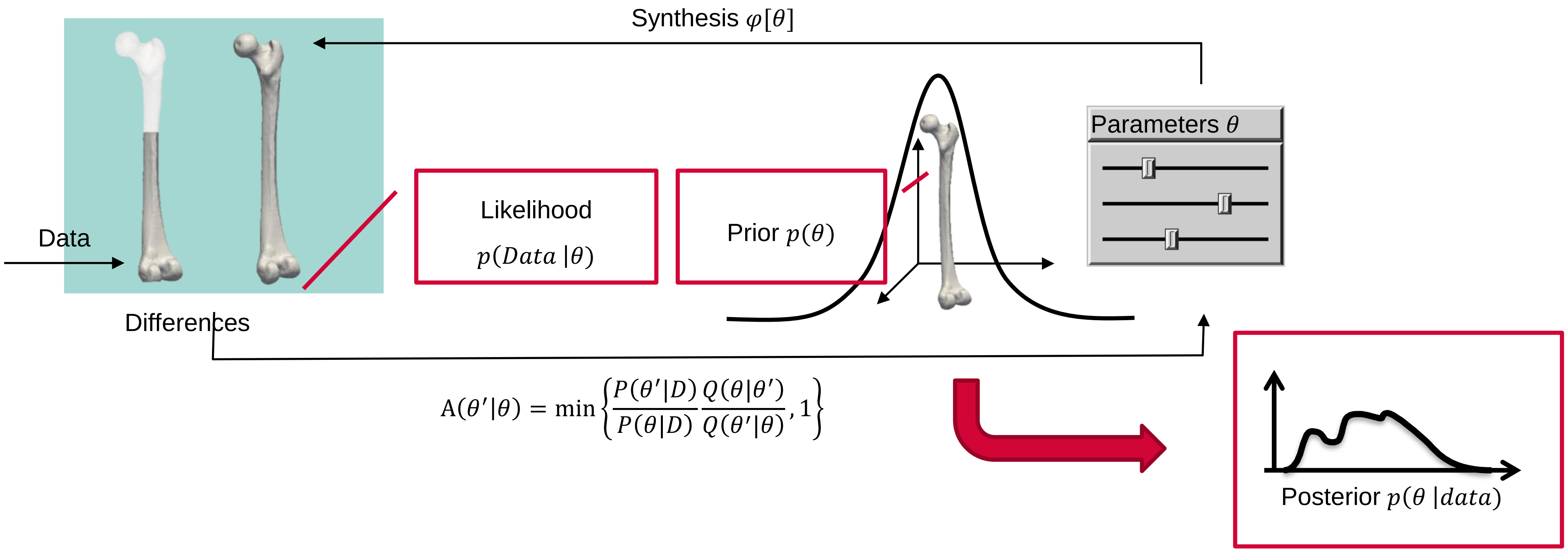
Initialize with sample $\boldsymbol{x}$
Generate next sample, with current sample $\boldsymbol{x}$

1. Draw a sample $\boldsymbol{x}'$ from $Q(\boldsymbol{x}'|\boldsymbol{x})$ (“proposal”)
2. With *probability* $\alpha = \min\left\{\frac{P(x')}{P(x)}\frac{Q(x|x')}{Q(x'|x)}, 1\right\}$ accept $\boldsymbol{x}'$ as new state $\boldsymbol{x}$
3. Emit current state $\boldsymbol{x}$ as sample

Metropolis-Hastings for analysis by synthesis

Computational problem

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta}$$



$$A(\theta'|\theta) = \min\left\{\frac{P(\theta'|D)}{P(\theta|D)}\frac{Q(\theta|\theta')}{Q(\theta'|\theta)}, 1\right\}$$