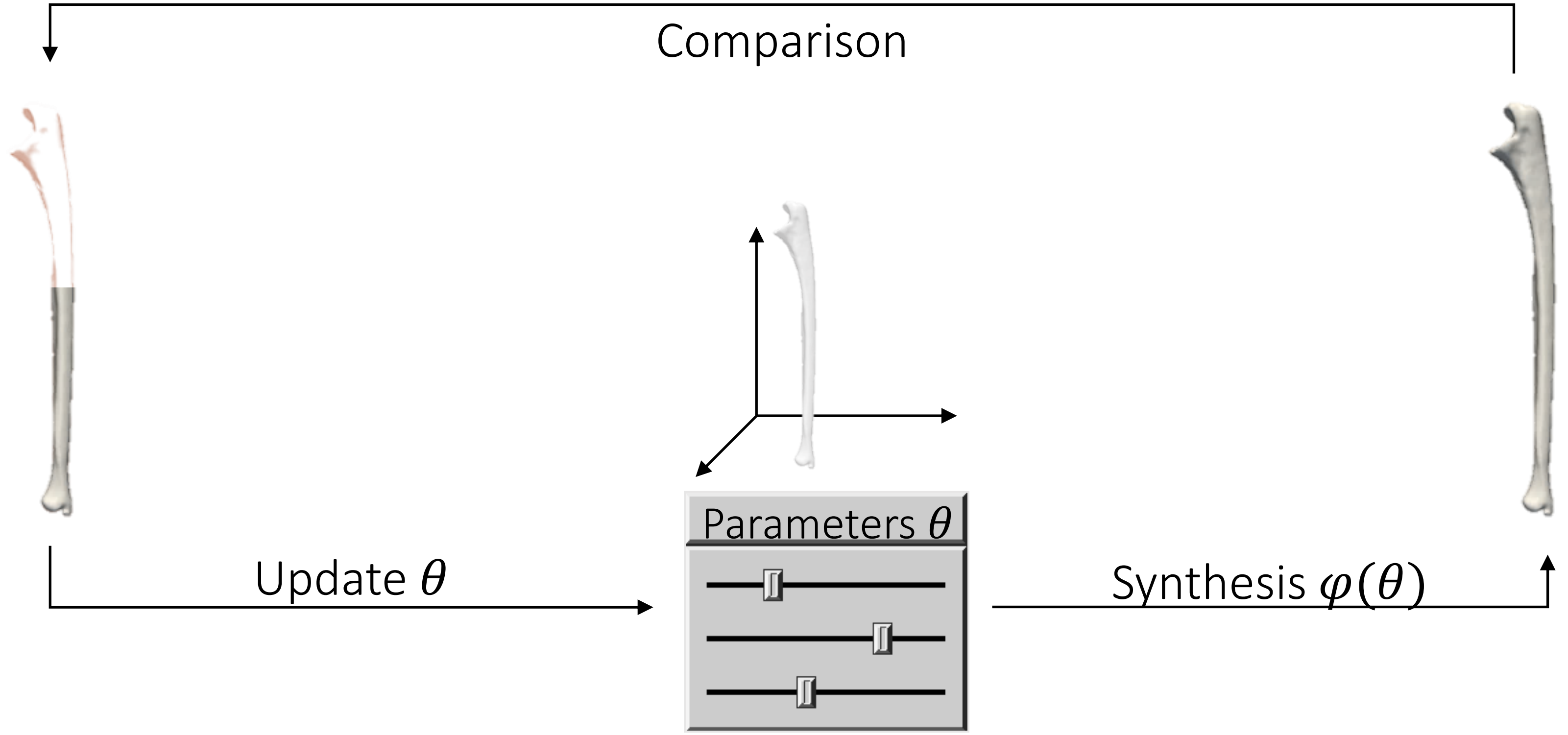
Probabilistic Shape Modelling

Summary

Marcel Lüthi

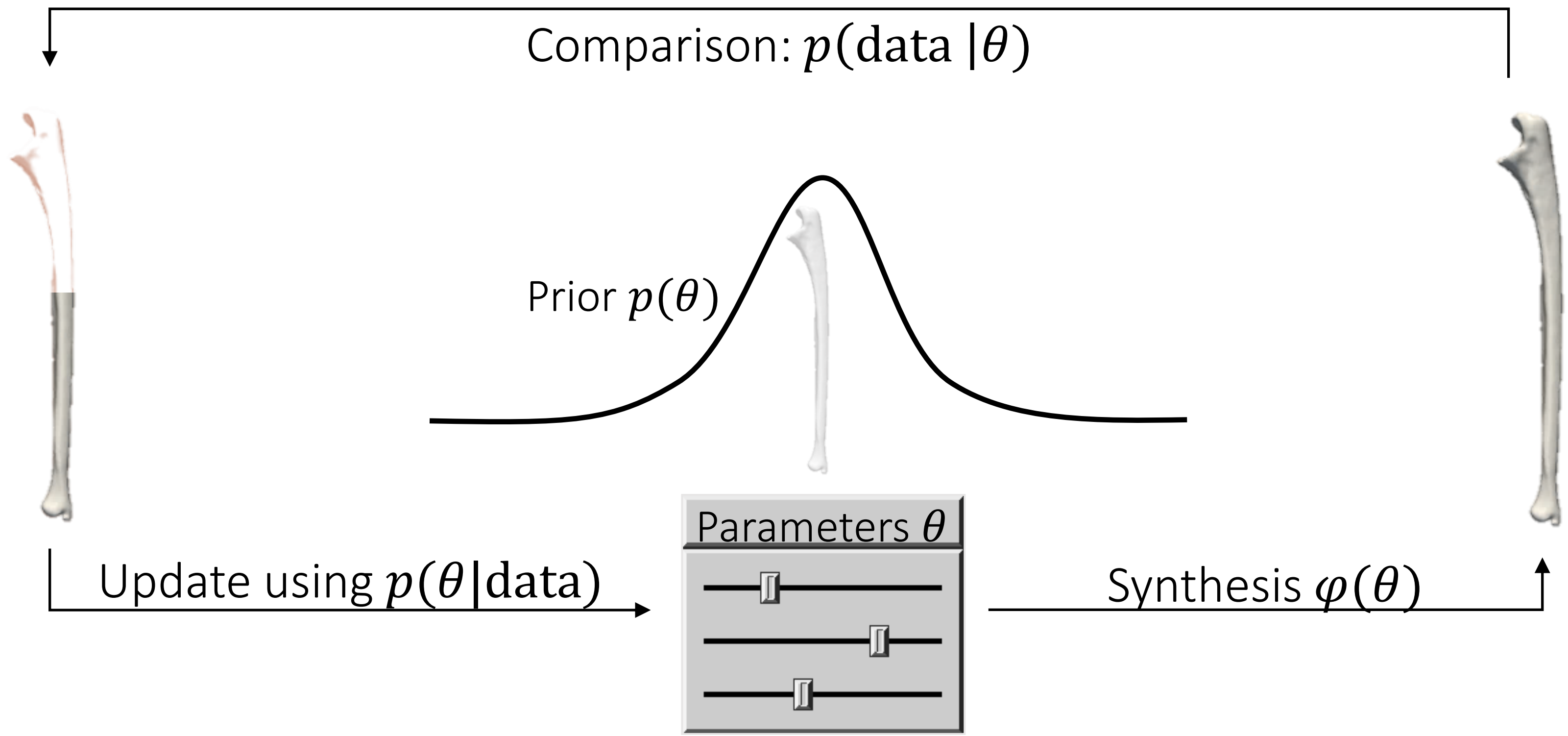
Conceptual Basis: Analysis by synthesis



Being able to synthesize data means we can understand how it was formed.

– Allows reasoning about unseen parts.

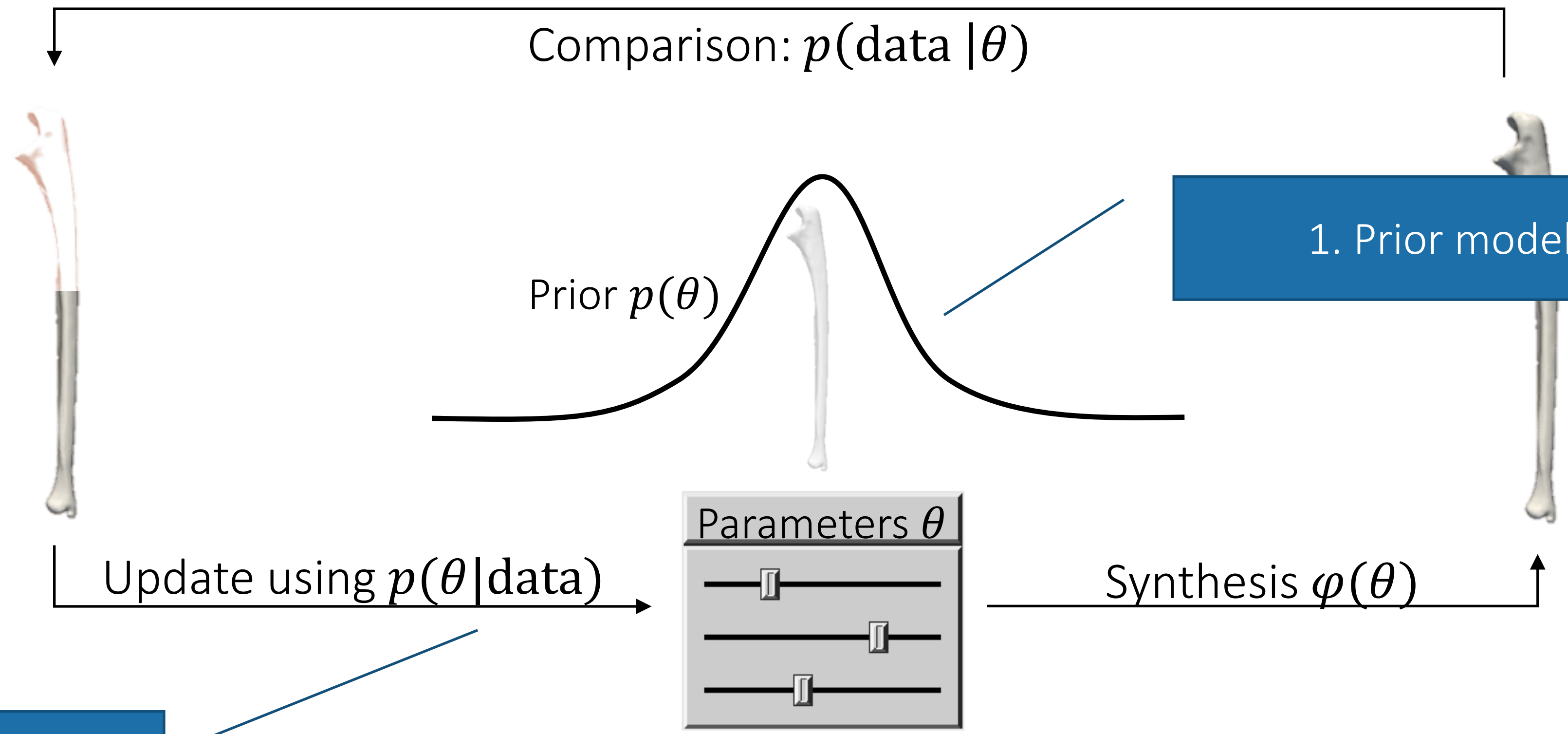
Analysis by Synthesis – Bayesian modelling



- Principled way of dealing with uncertainty.

Agenda

2. Likelihood function



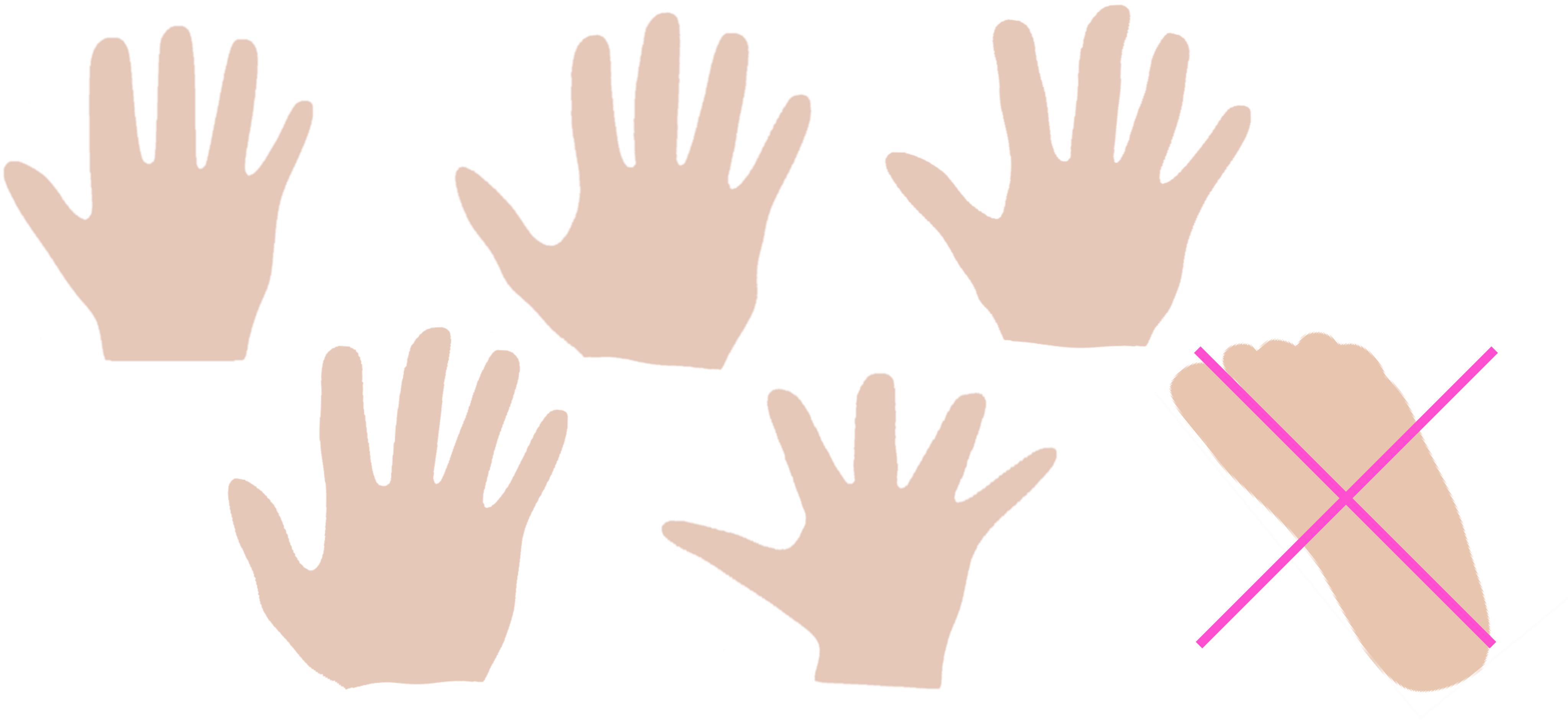
1. Prior modelling

3. Inference

1. Shape modelling using GPs

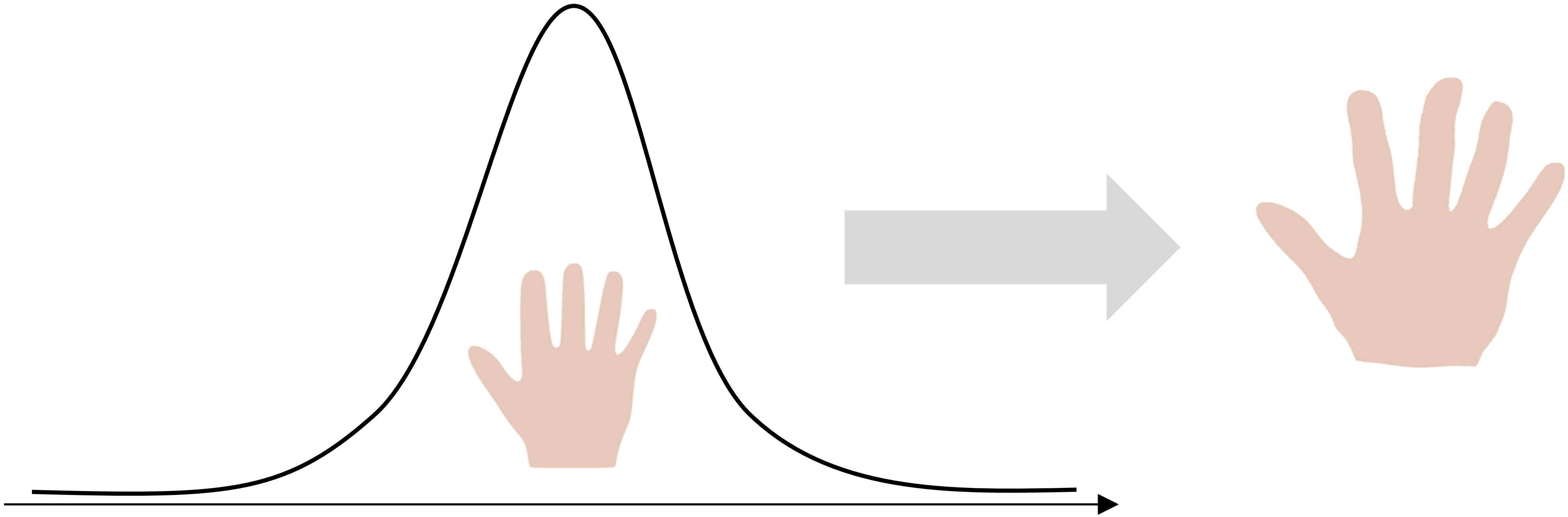
Starting point: Characterizing shape families

Starting point: Characterizing shape families



Probabilistic models of shapes

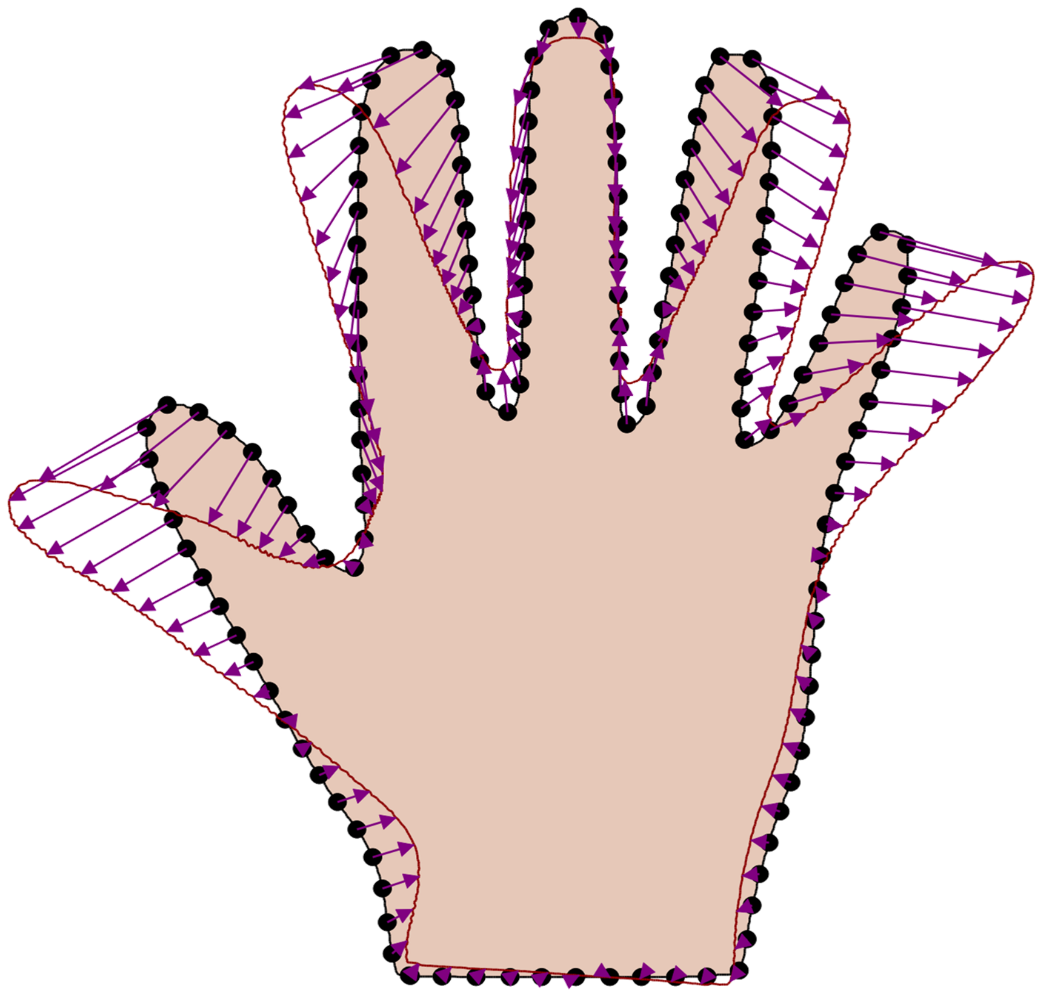
- Define how likely it is that a shape is part of the family
- Can generate new shapes



Defining the shape model

1. Generating a shape
   - Start with reference shape: $\Gamma_R = \{x \mid x \in \mathbb{R}^2\}$
   - Describe shape difference as vector field $u : \Gamma_R \to \mathbb{R}^2$

2. Defining shape model
   - Induce probability distribution on $u \sim GP(\mu, k)$



Gaussian process: Formal definition

A Gaussian process $p(u) = GP(\mu, k)$

is a probability distribution over functions

$$u : \mathcal{X} \to \mathbb{R}^d$$

such that every finite restriction to function values

$$u_X = (u(x_1), \dots, u(x_n))$$

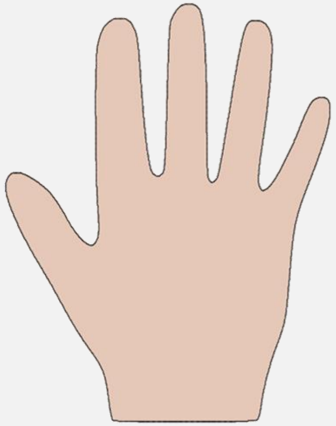
is a multivariate normal distribution

$$p(u_X) = N(\mu_X, k_{XX}).$$

From continuous to discrete

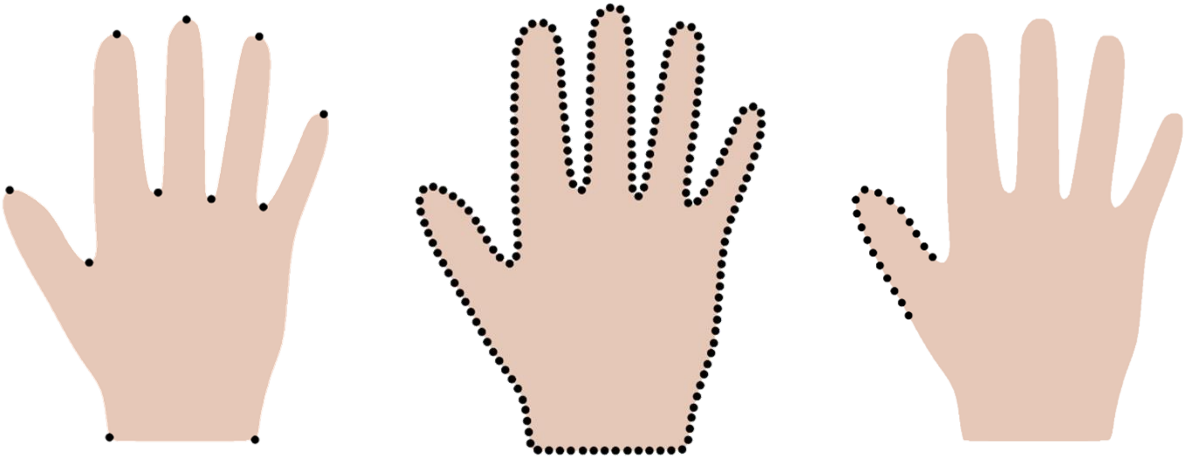
Conceptual formulation:

Continuous: $GP(\mu, k)$



Practical implementation:

Discrete: $N(\mu, K)$



Defining a Gaussian process

A Gaussian process $GP(\mu, k)$ is completely specified by a mean function μ and covariance function (or kernel) k.

- $\mu: \mathcal{X} \to \mathbb{R}^d$ defines how the average deformation looks like
- $k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}^{d \times d}$ defines how it can deviate from the mean
  - Must be positive semi-definite

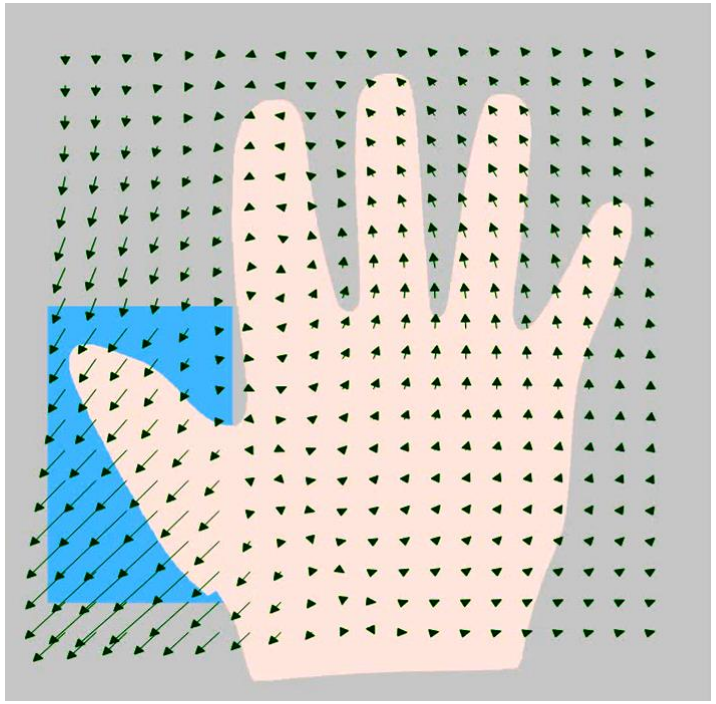
Rules for combining covariance functions

Simple kernels are not powerful enough for modelling realistic deformations.

Rules for constructing kernels:

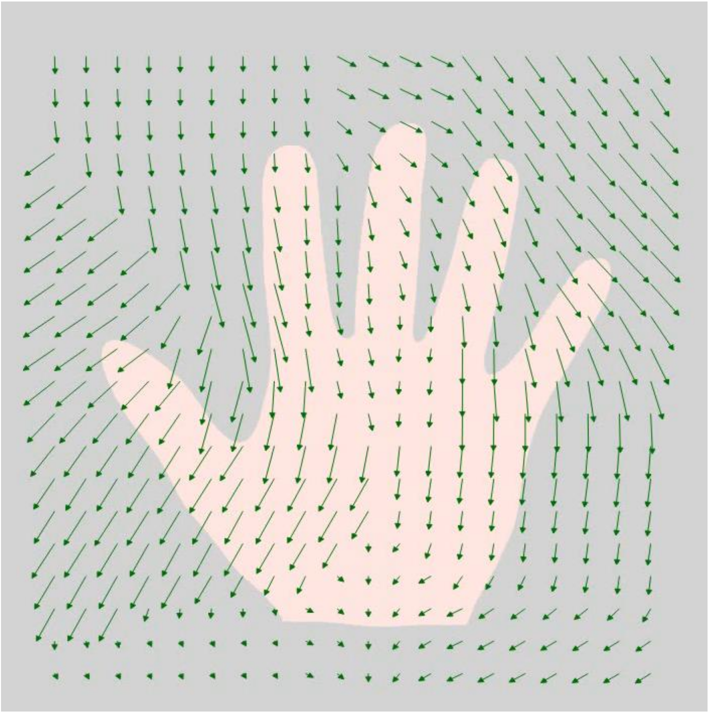
1. $k(x, x') = k_1(x, x') + k_2(x, x')$
2. $k(x, x') = \alpha k_1(x, x'), \alpha \in \mathbb{R}_+$
3. $k(x, x') = k_1(x, x') \circ k_2(x, x')$
4. $k(x, x') = f(x)\, f(x')^T, f: X \rightarrow \mathbb{R}^d$
5. $\mathrm{k}(x, x') = B^T k(x, x') B, \quad \mathrm{B} \in \mathbb{R}^{r \times d}$

Combining kernels for shape modelling



- Spatially varying smooth deformations

$$k(x,x') = \chi(x)\chi(x')k_1(x,x') + (1-\chi(x))(1-\chi(x'))\,k_2(x,x')$$



- Covariance function learned from examples

$$k_{SM}(x,x') = \frac{1}{n-1}\sum_{i}^{n}(u^i(x)-\overline{u}(x))(u^i(x')-\overline{u}(x'))^T$$

The Karhunen-Loève expansion

We can write $\mathrm{u} \sim GP(\mu, k)$

as $u \sim \mu + \sum_{i=1}^{\infty} \alpha_i \sqrt{\lambda_i}\, \phi_i,\ \alpha_i \sim N(0,1)$

- ϕ_i is the eigenfunction with associated eigenvalue λ_i of the linear operator

$$[T_k u](x) = \int k(x,s) u(s) ds$$

Low-rank approximation

$$\boxed{u = \mu + \sum_{i=1}^{r} \alpha_i \sqrt{\lambda_i}\, \phi_i, \qquad \alpha_i \sim N(0,1)}$$

Main idea: Represent process using only the first r components

- We have a finite, parametric representation of the process.
- Any deformation u is determined by the coefficients $\alpha = (\alpha_1, \dots, \alpha_r)$

$$p(u) = p(\alpha) = \prod_{i=1}^{r} \frac{1}{\sqrt{2\pi}} \exp(-\alpha_i^2/2)$$

Summary – Gaussian processes

- Gaussian processes are an extremely rich toolbox for modelling functions / deformation fields
  - Possible to build complex models out of simple building blocks
  - Defining good prior assumptions is on us => Difficult part

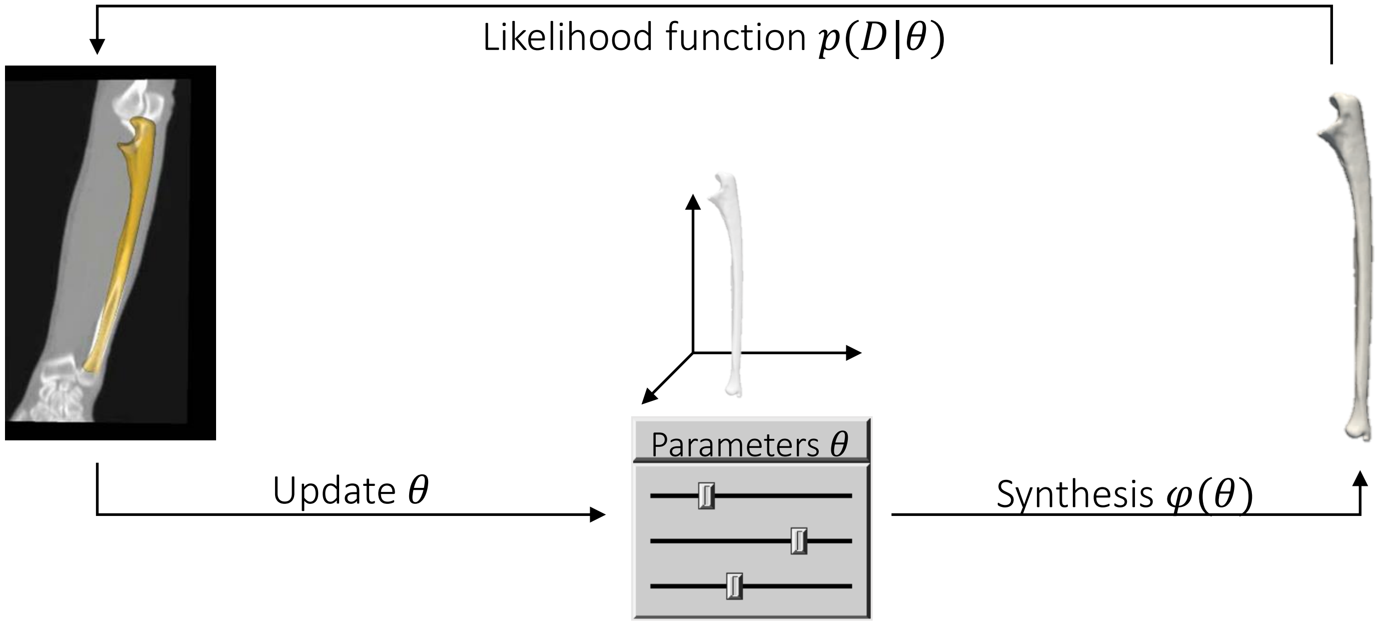
- Marginalization property and low-rank approximation allow for practical and efficient implementations

2. Likelihood functions

Likelihood functions

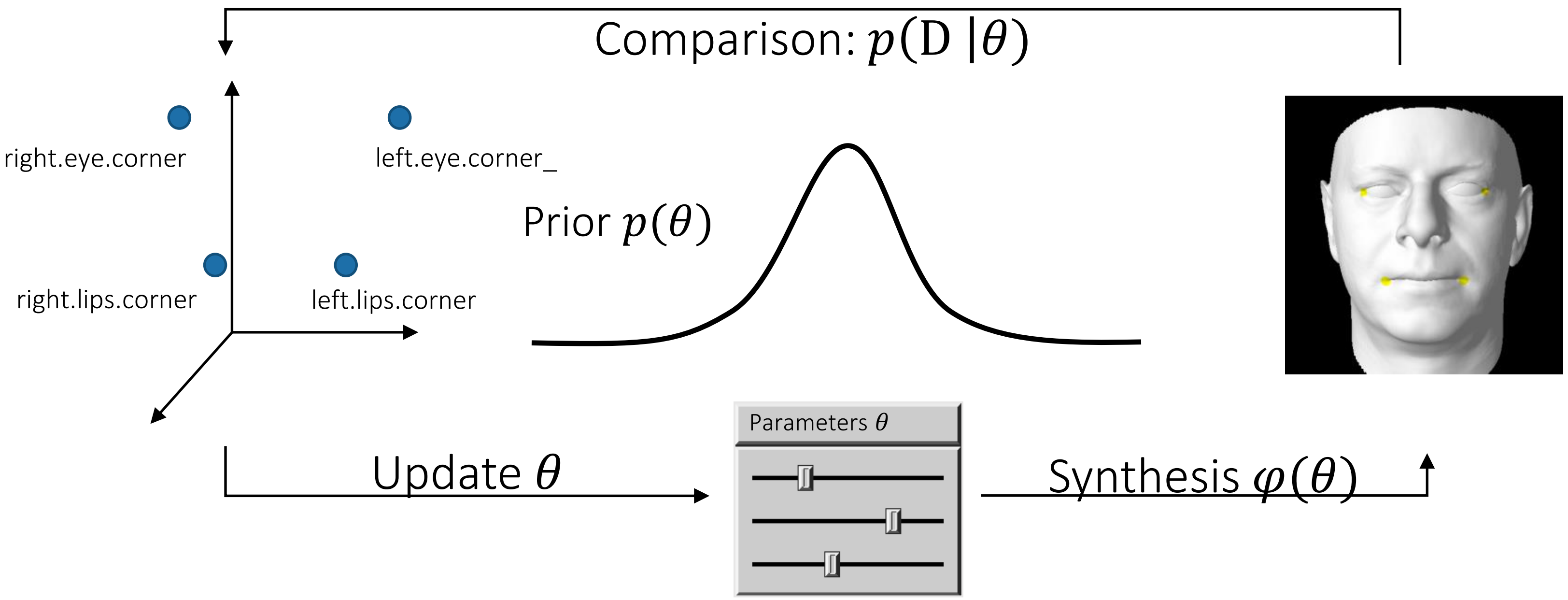
- The likelihood function $p(D|\theta)$ captures how we think the observation D arises from a given model instance defined by θ.
  - Split into synthesis function (deterministic) and probabilistic model

$$p(D|\theta) = p(D|\varphi[\theta])$$



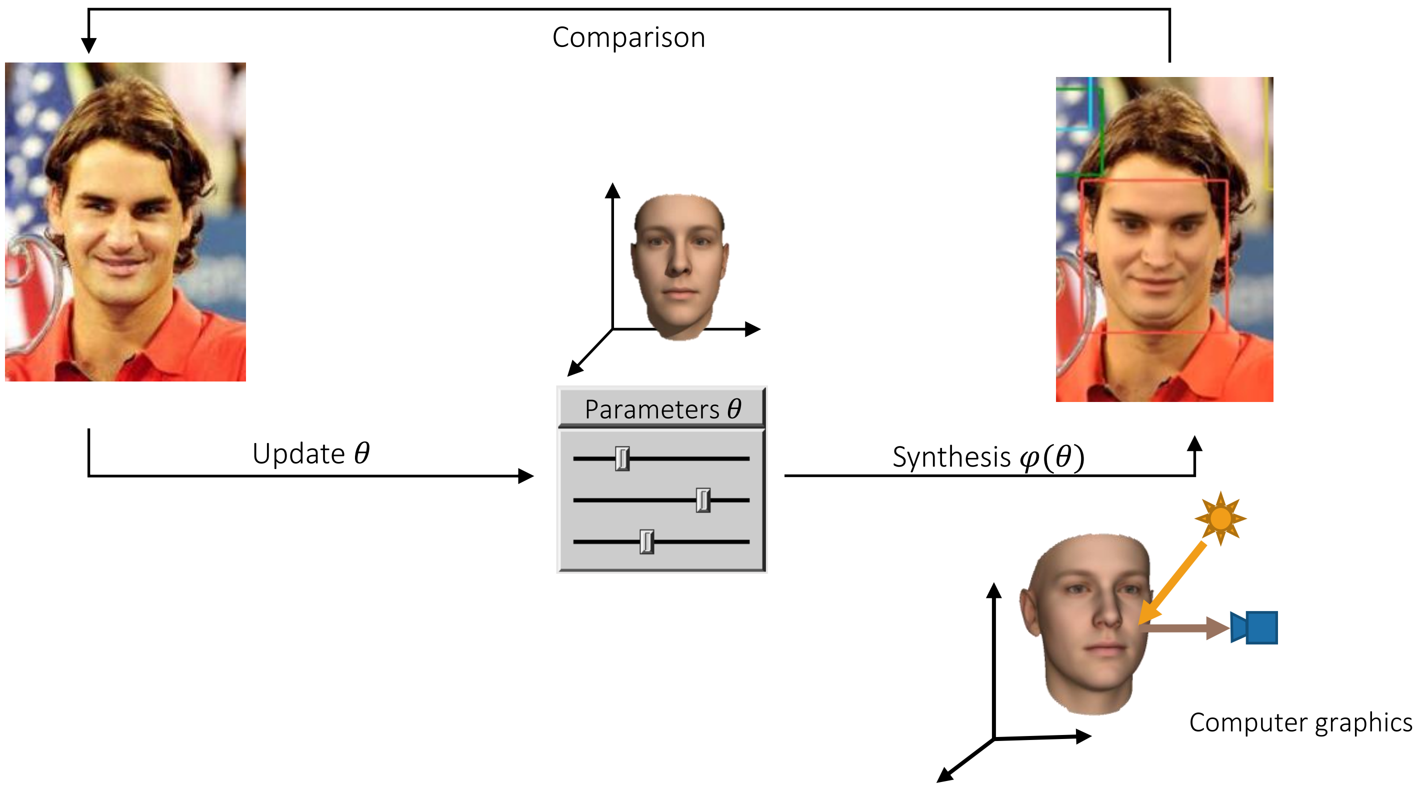
Likelihood functions

- Synthesis function can be very simple
  - Example: 3D Landmarks in correspondence



Likelihood functions

- Synthesis function can be very complex
  - Example: Complete computer graphics rendering pipeline



Typical approach to define likelihood

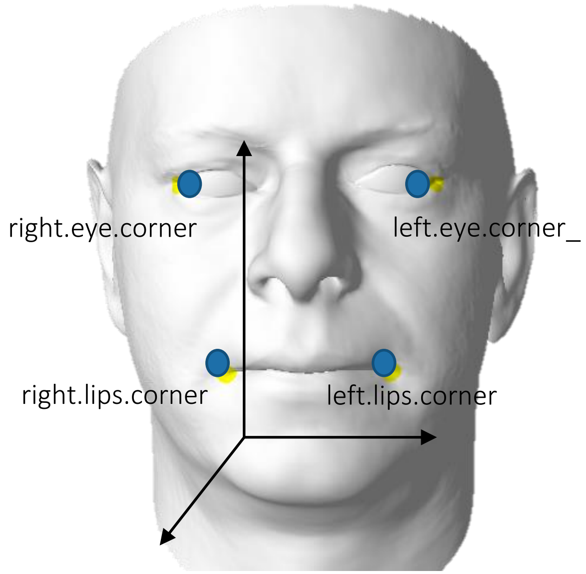
Quantify uncertainty after synthesizing individual landmark $\varphi[\theta](l_i^R)$

$$p(l_i^T | \theta, l_i^R) = N(\varphi[\theta](l_i^R), I_{3x3}\sigma^2)$$

Assume independence

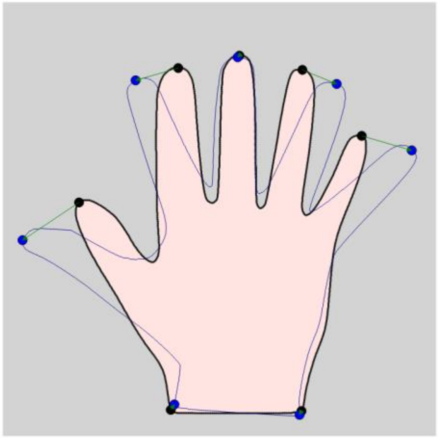
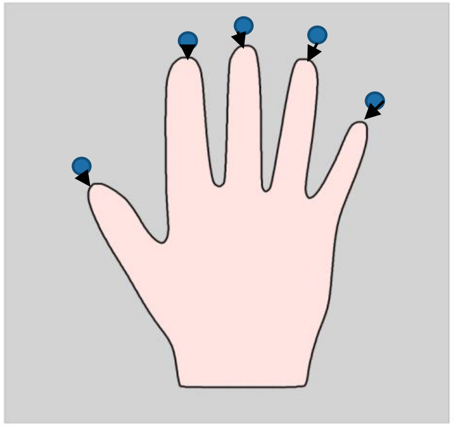
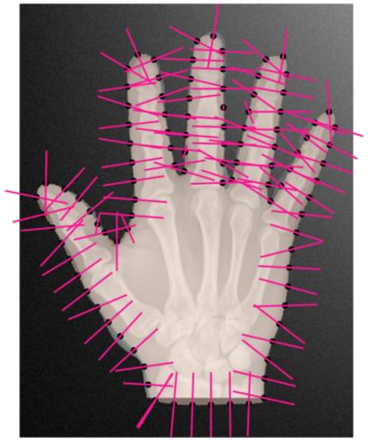
$$(l_1^T, \dots, l_n^T | \theta, l_1^R, \dots, l_n^R) = \prod_i N(\varphi[\theta](l_i^R), I_{2x2}\sigma^2)$$

Landmarks match target position up to zero-mean Gaussian noise.



Reminder: Likelihood functions

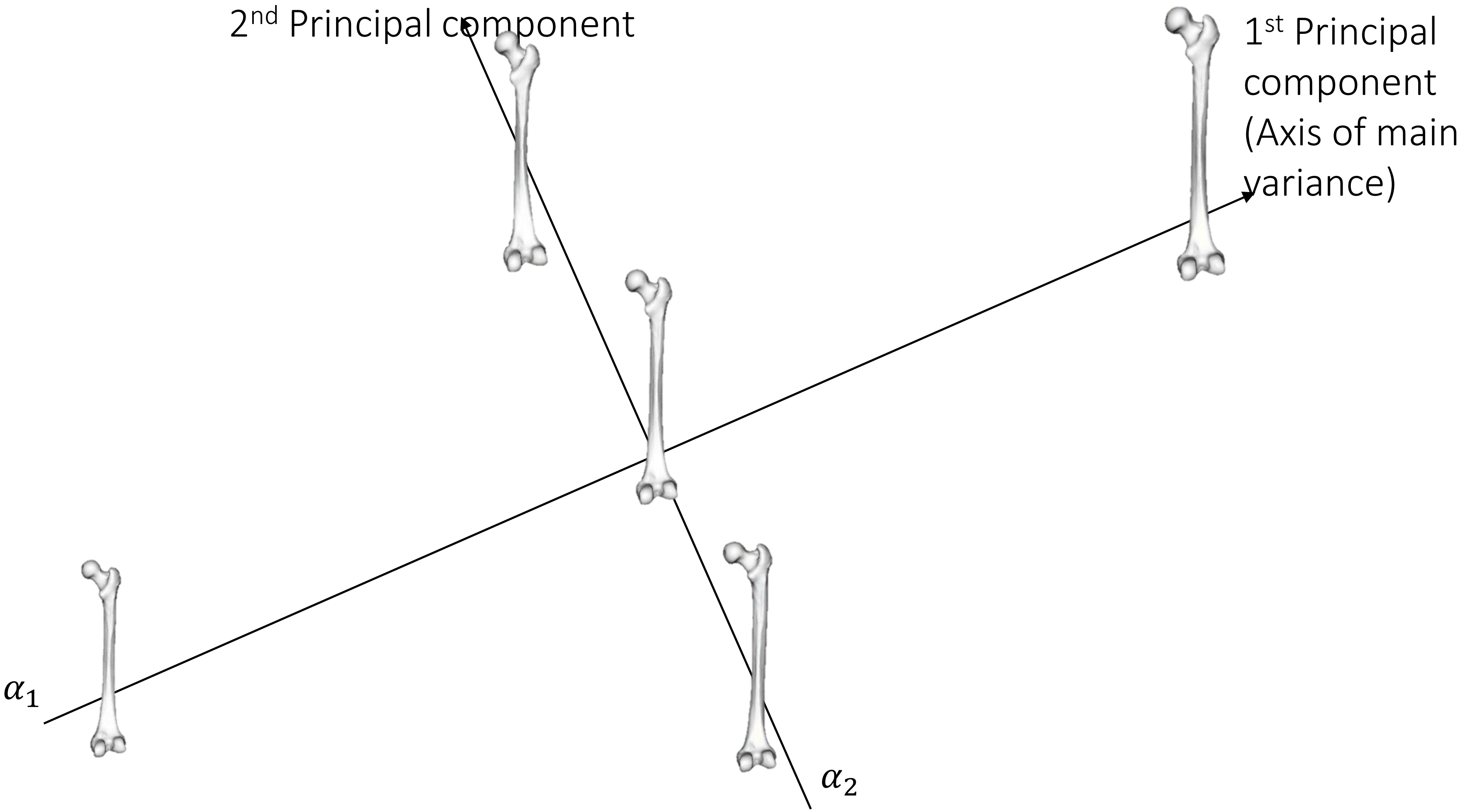
Likelihood function: $p(I_T|\theta, I_R)$

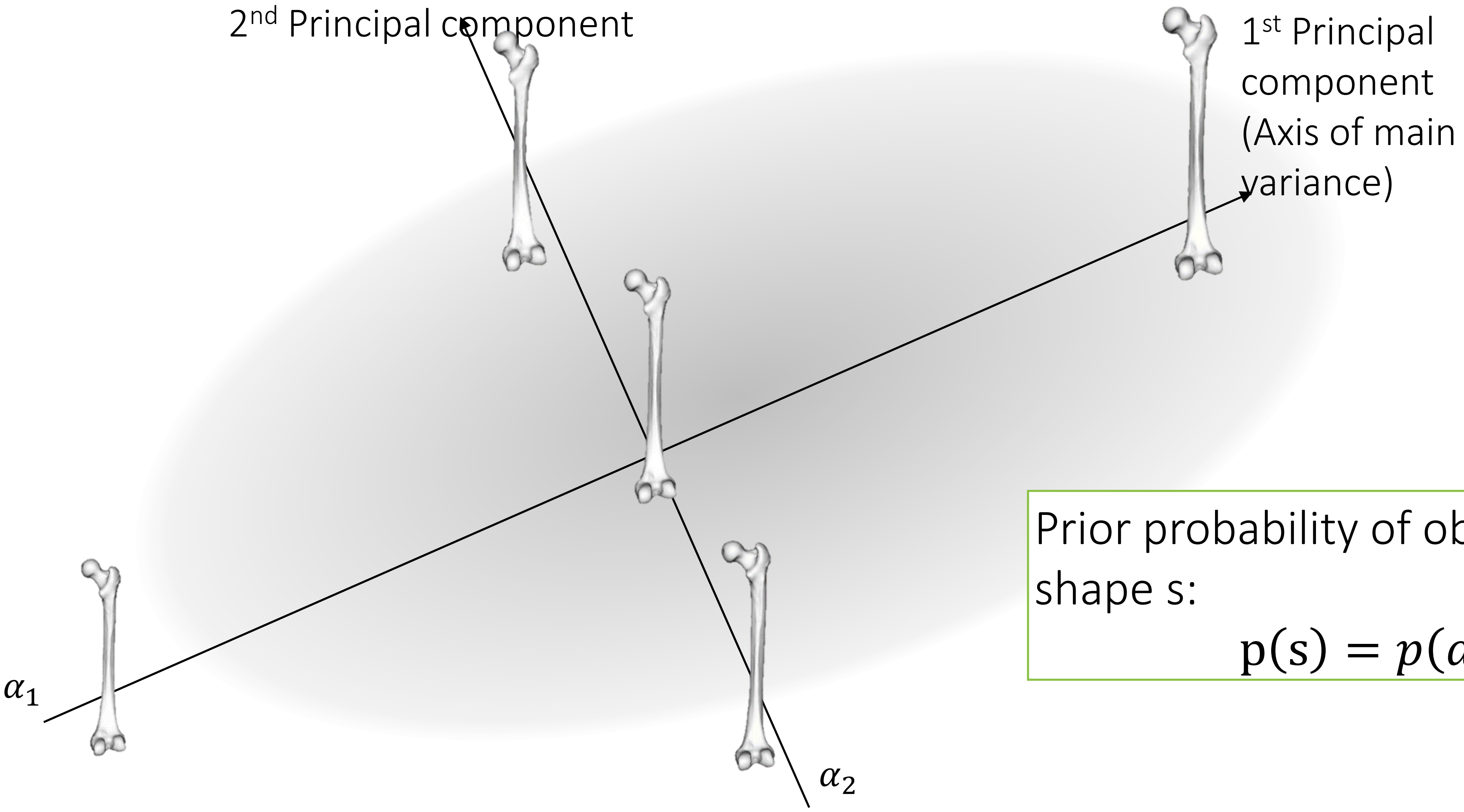
Comparison	Landmarks / landmark	Points / contour	Contour to image
Stochastic component	Noise on landmark points	Noise on point position	Deviation of image intensity from learned profiles

3. Inference

Statistical shape model



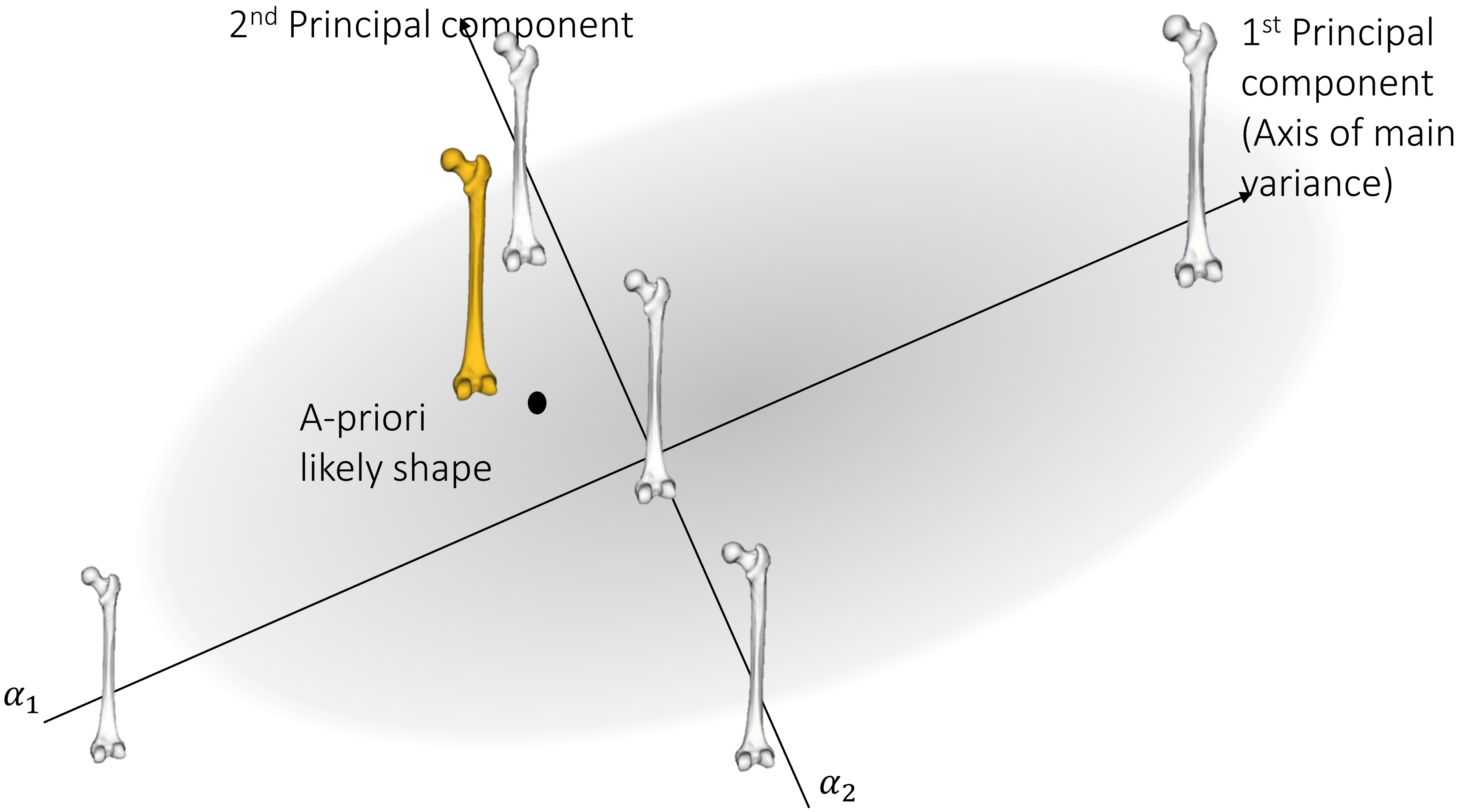
Probability before seeing data



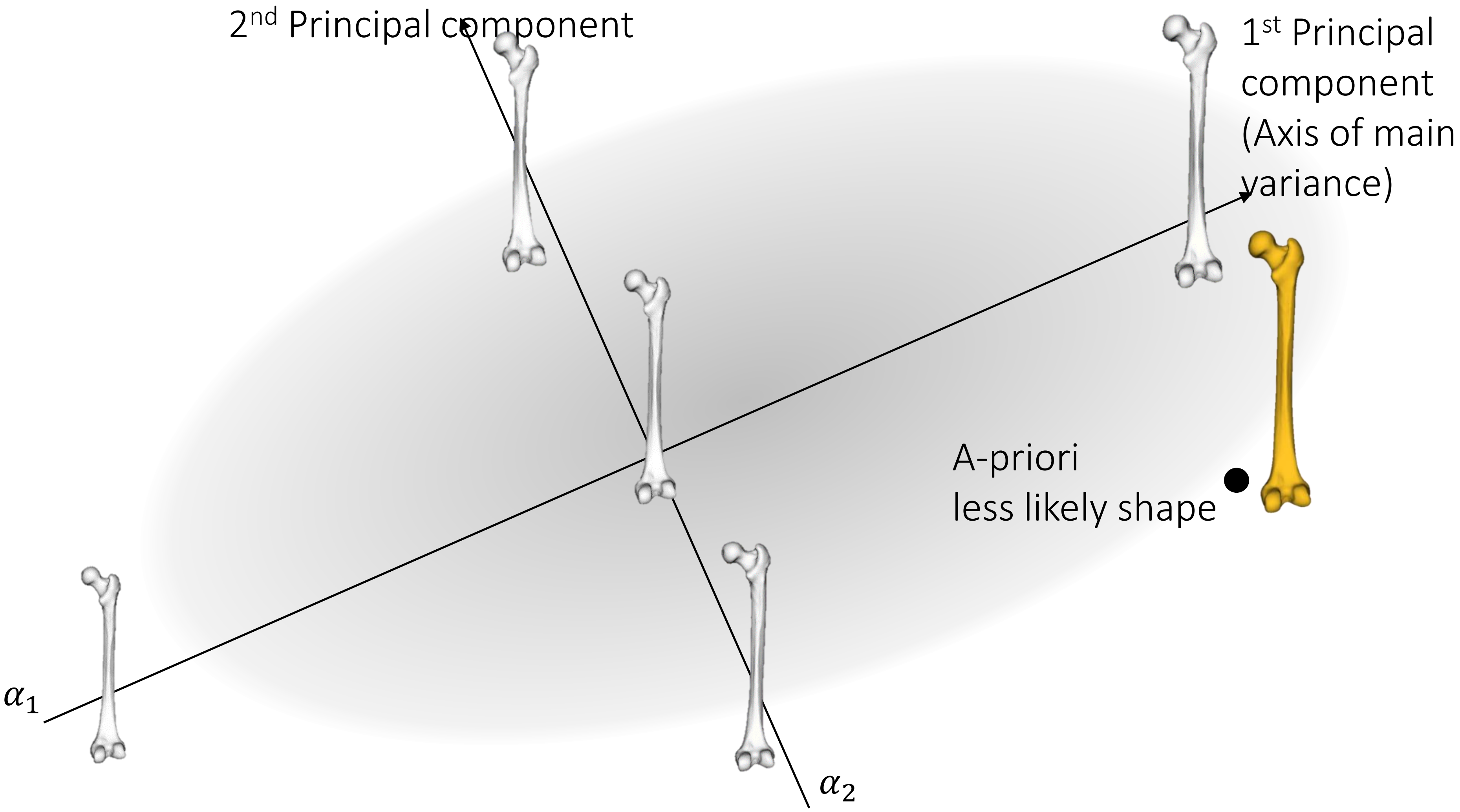
Prior probability of observing shape s:

$$\mathrm{p}(s) = p(\alpha)$$

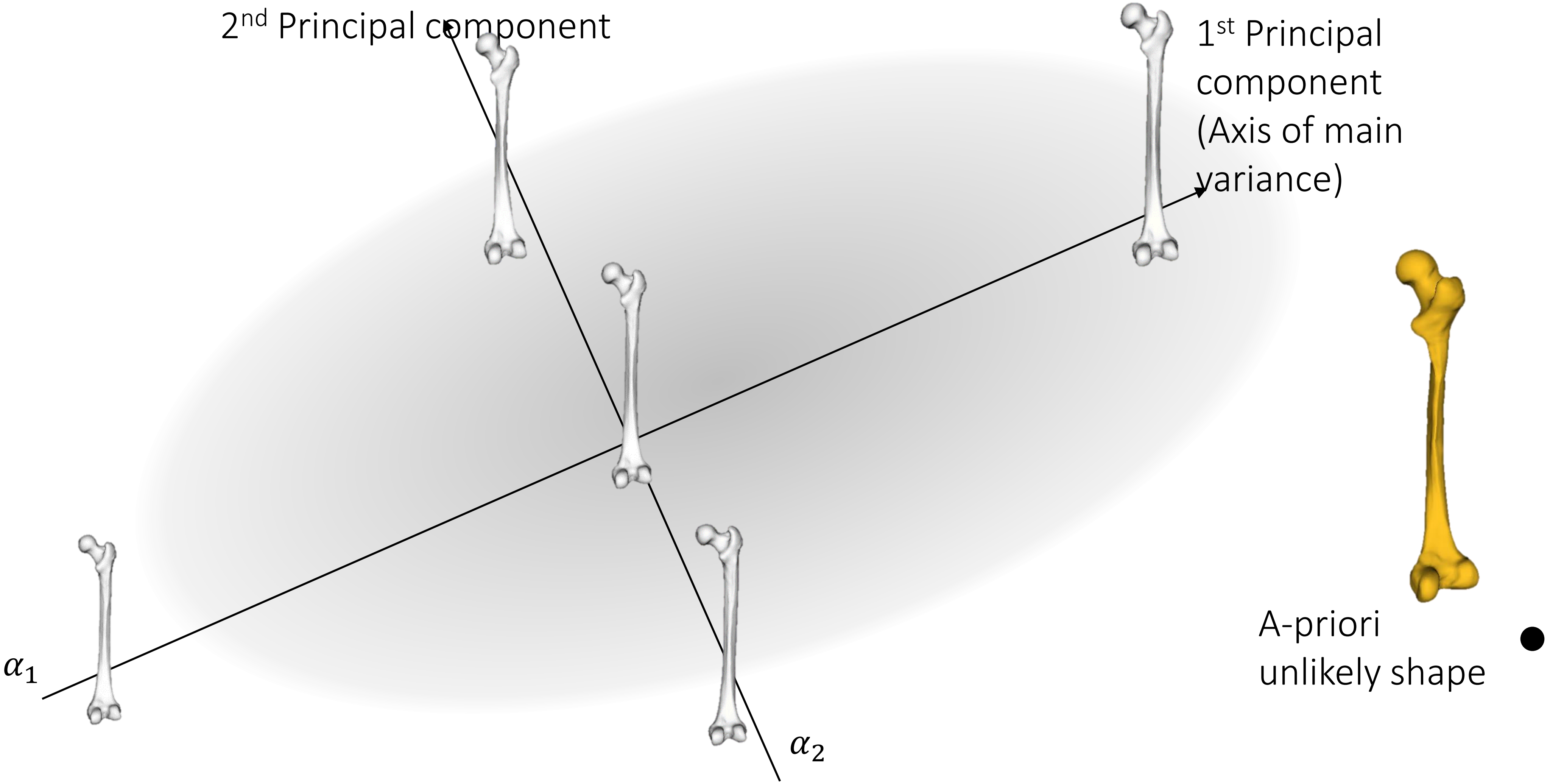
Probability before seeing data



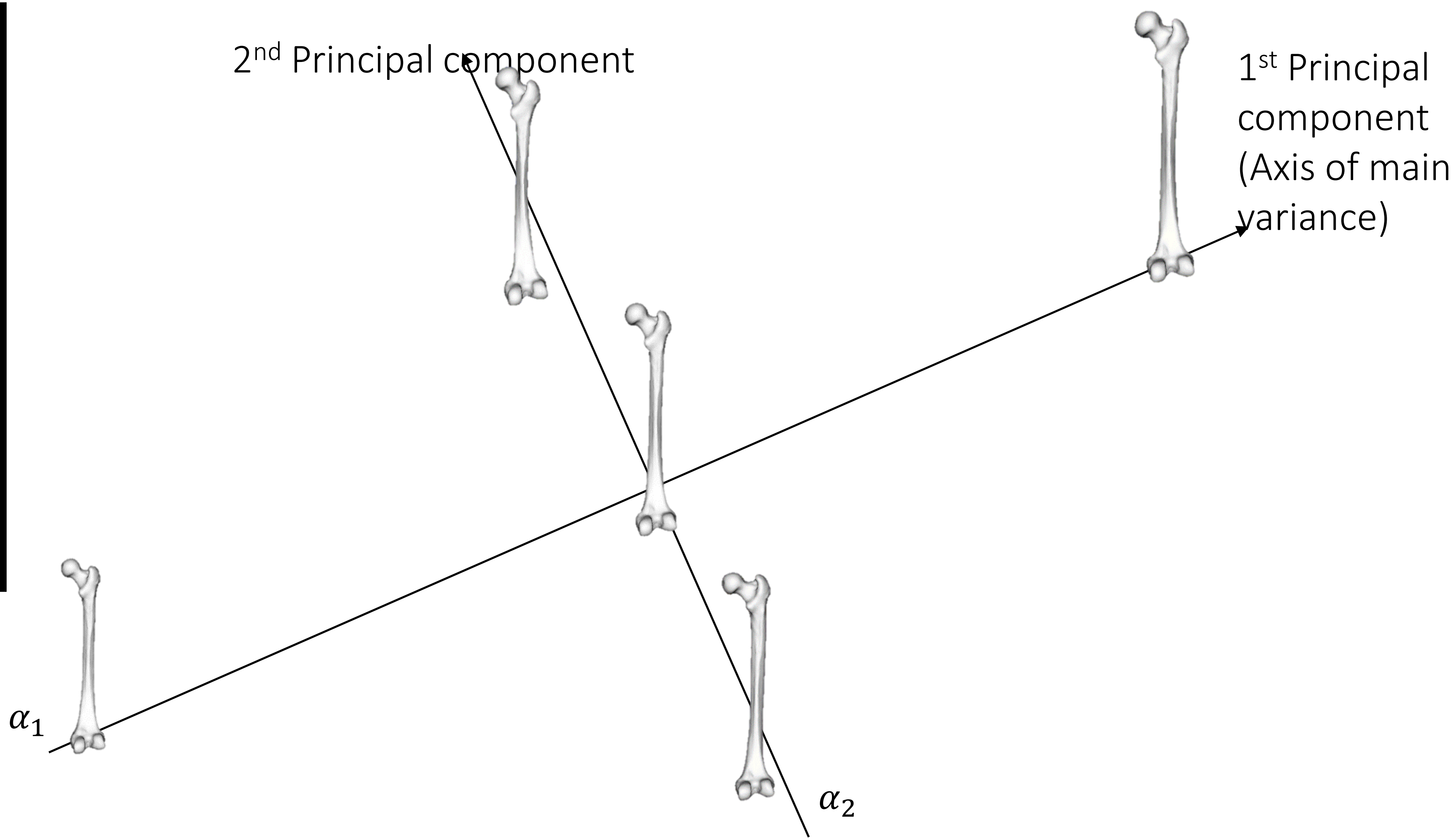
Probability before seeing data



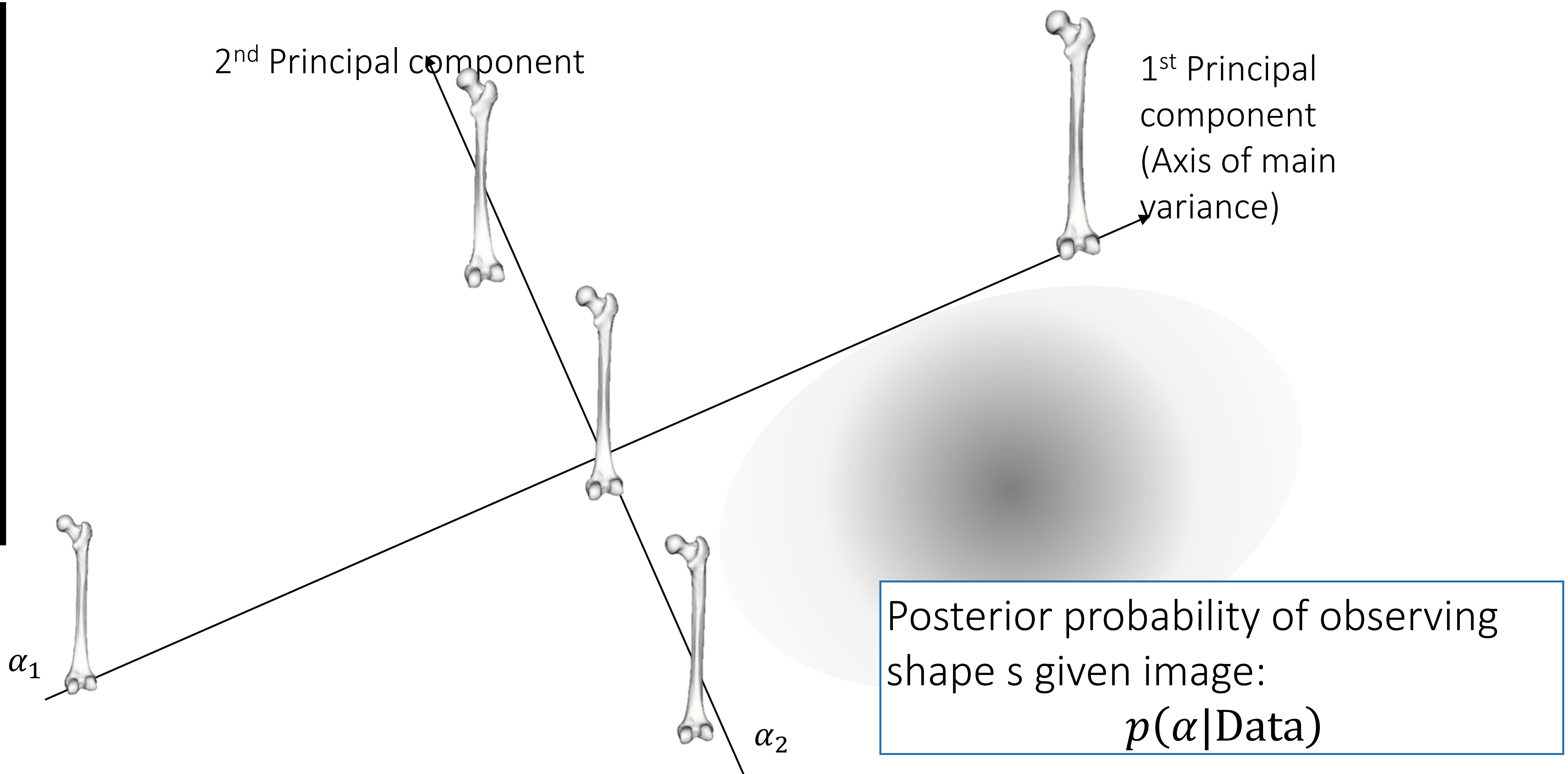
Probability before seeing data



Observing Data



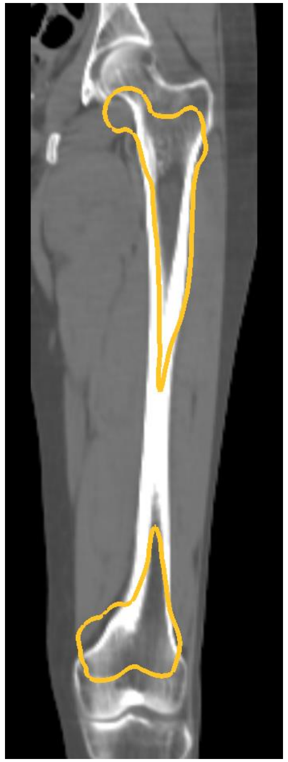
Probability after observing data

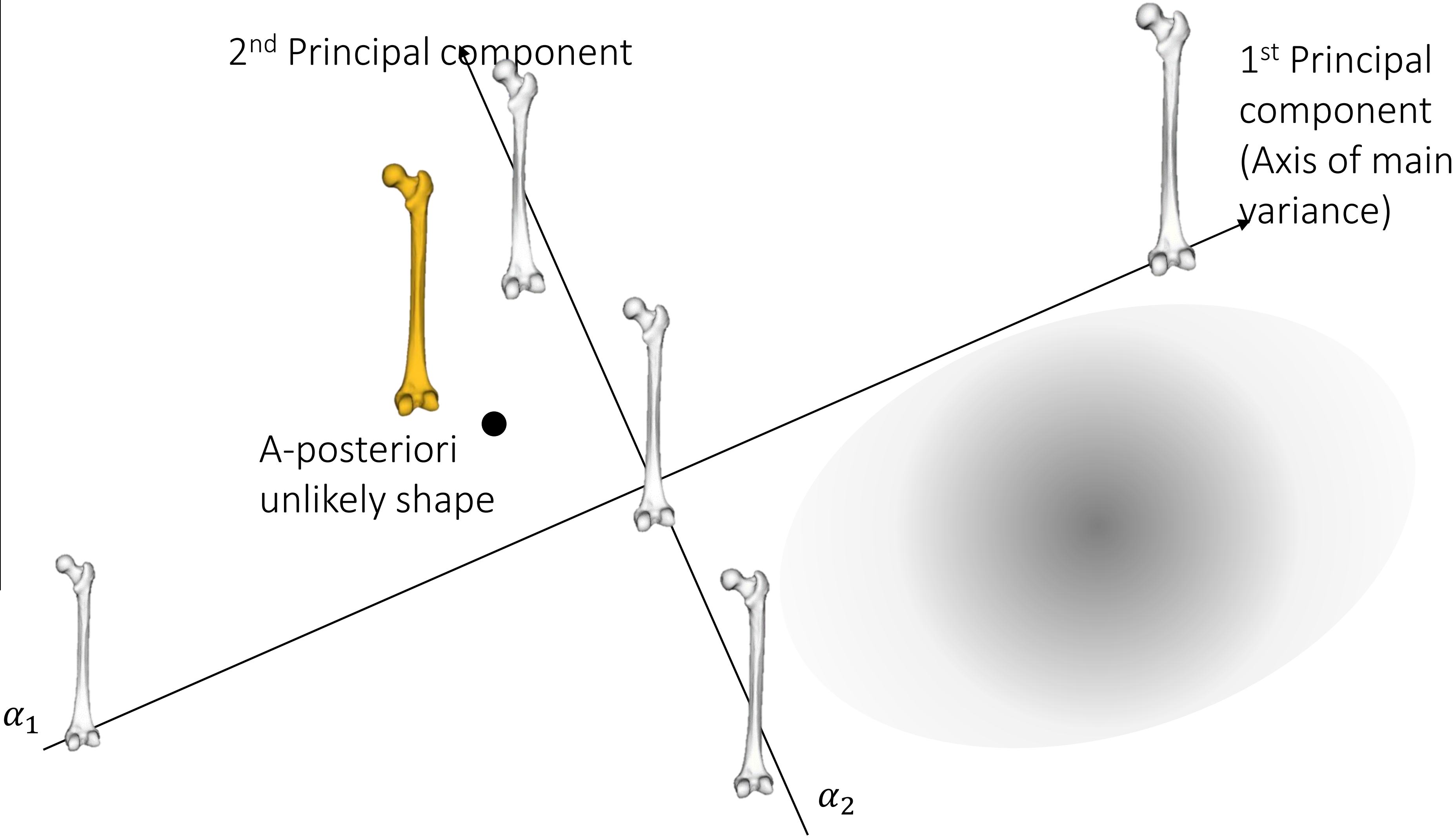


Posterior probability of observing shape s given image:

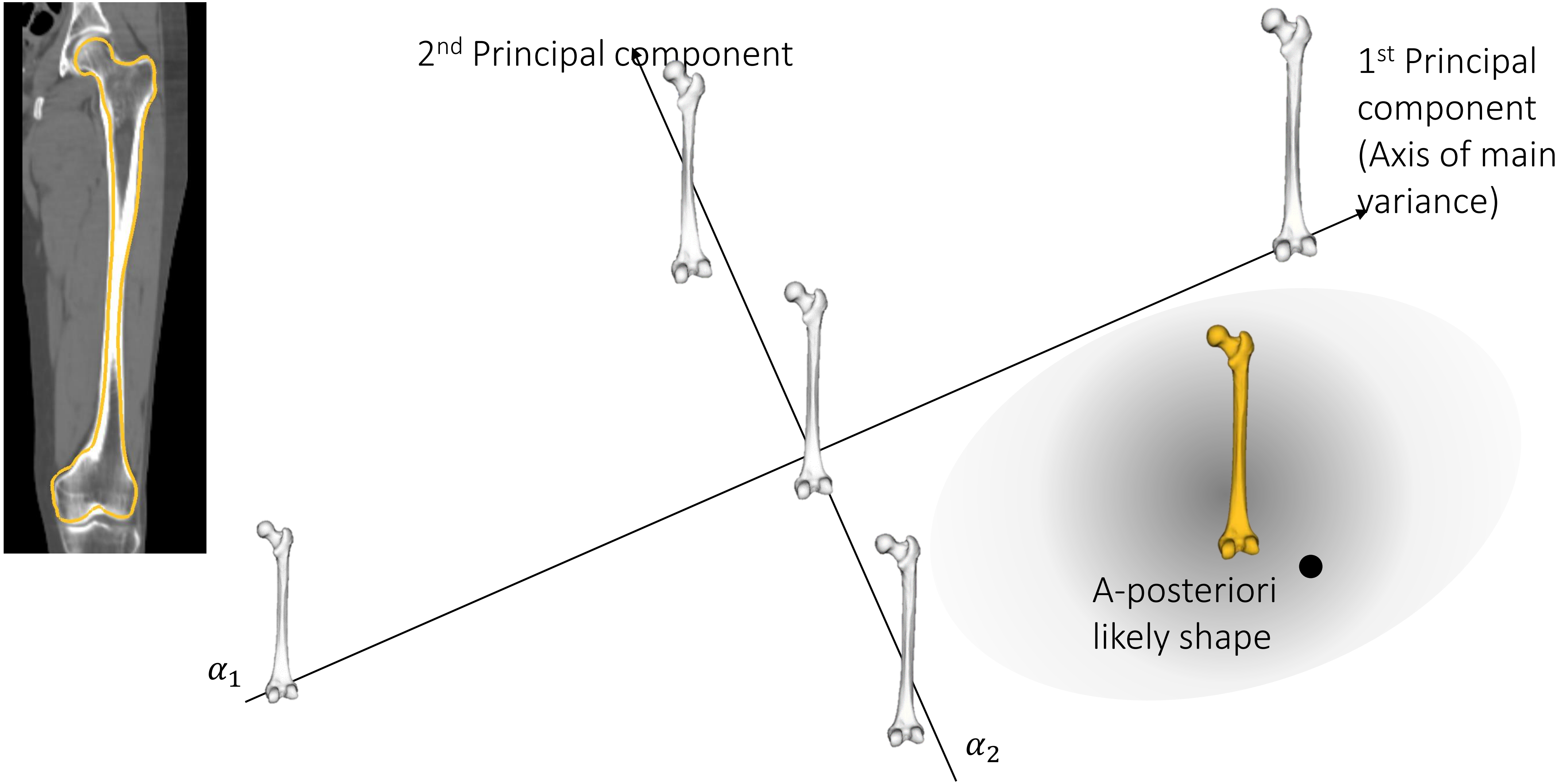
$$p(\alpha|\text{Data})$$

Probability after observing data





Probability after observing data



Model-based data analysis – a Bayesian approach

How well did we explain the data with the model (parameters)

Prior belief: Statistical shape model

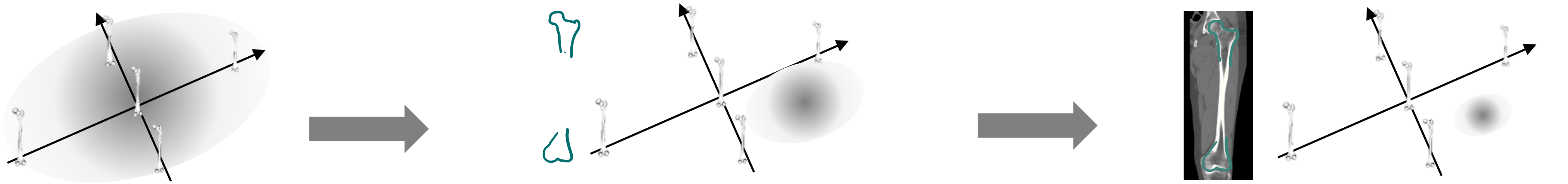
$$P(\alpha|\text{Data}) = \frac{P(\text{Data}|\alpha)P(\alpha)}{P(\text{Data})}$$

Posterior belief

Normalization term (marginal likelihood)

Model-based data– a Bayesian approach

Can introduce new data one by one.

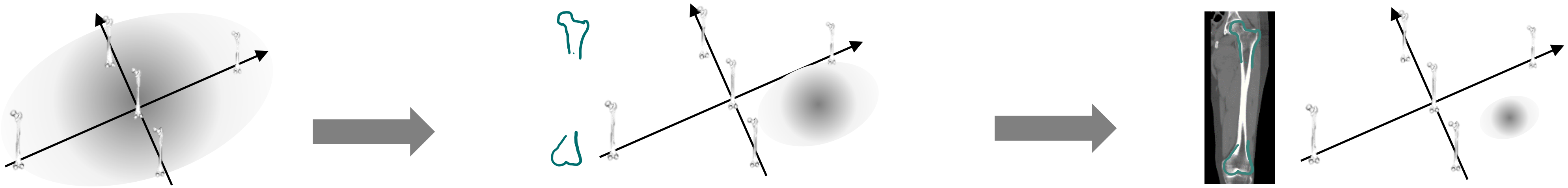


$P(\alpha)$ $\quad$ $P(\alpha|\text{Annotations})$ $\quad$ $P(\alpha|\text{Annotation},\text{Image})$

- Uncertainty is reduced in every step.
- Bayesian inference gives mathematically sound way of updating our knowledge.

Model-based data– a Bayesian approach

Can introduce new data one by one.



$P(\alpha)$ $\quad$ $P(\alpha|\text{Annotations})$ $\quad$ $P(\alpha|\text{Annotation},\text{Image})$

- Challenges
- How do we model shape variations?
- How do we update probabilities?
- How do we make this applicable and useful in practice?

Computational problem:

$$P(\alpha|\text{Data}) = \frac{P(\text{Data}|\alpha)P(\alpha)}{\int \dots \int P(\text{Data}|\alpha_1, \dots, \alpha_n)P(\alpha_1, \dots, \alpha_n)d\alpha_1, \dots, d\alpha_n}$$

Metropolis-Hastings algorithm

- *Formalizes propose-and-verify*

Draw a sample x' from $Q(x'\|x)$	Propose
With *probability* $\alpha = \min\left\{\frac{P(x')}{P(x)}\frac{Q(x\|x')}{Q(x'\|x)}, 1\right\}$ accept $\boldsymbol{x'}$ as new sample	Verify

- Very useful concept to integrate unreliable proposals!
  - Can deal with heuristics which are not always right
  - Can deal with unreliable data
- All assumptions about the problem in proposals ⇒ Extremely important to design them well

Conclusion

Analysis by synthesis is a generic approach to shape and image analysis

- Based on Bayesian framework
  1. Model prior (which shape do we expect to see)
  2. Model likelihood (how do we expect it to appear in an image)
  3. Compute posterior (what are the likely shapes given the image)
- Extremely flexible main components
  - Gaussian process (prior)
  - Metropolis Hastings (fitting / sampling from the posterior)
- Rigorous theoretical framework, which helps us to navigate in the space
  But: Finding good solutions for practical image analysis is hard work!

There is nothing more practical than a good theory

V. Vapnik