THE UNIVERSITY of TENNESSEE KNOXVILLE

Lecture 25: Coin Toss and Crystal Structure - the Bayesian Way

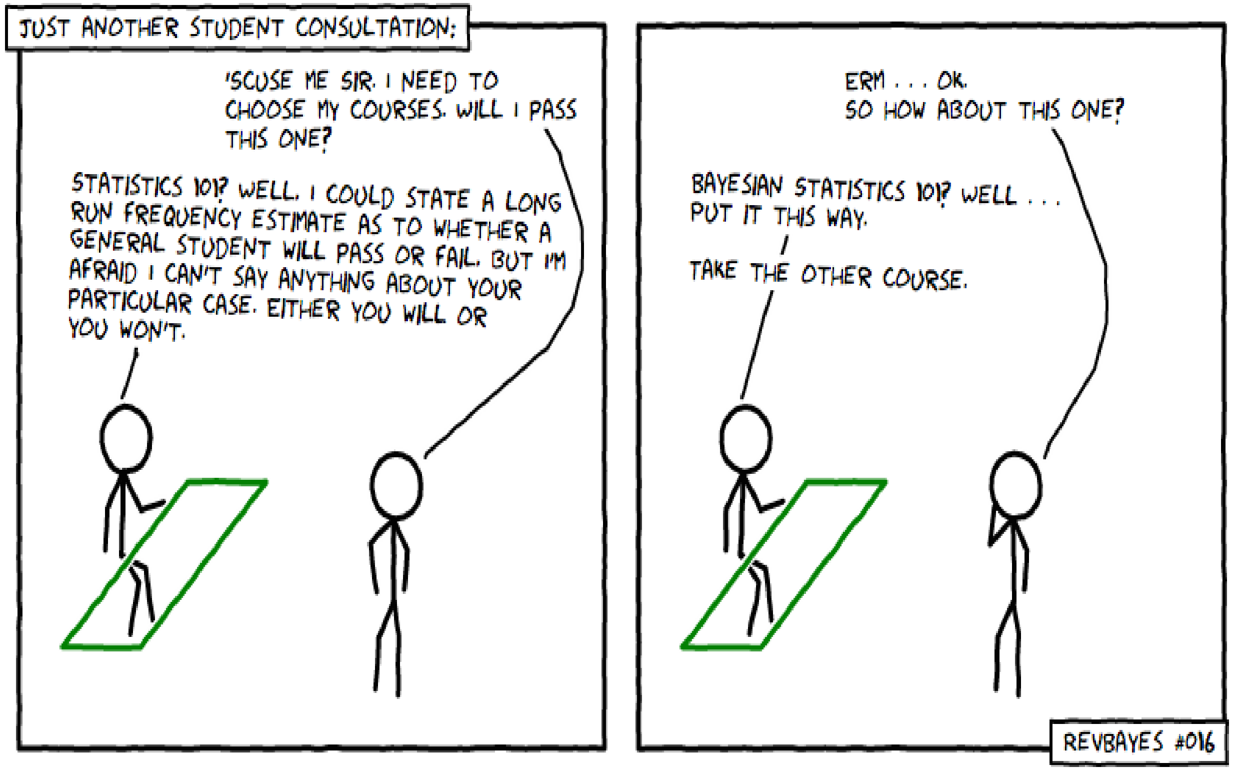
Instructor: Sergei V. Kalinin

Special thanks to Mani Valleti for the crystal structure example

Frequentist vs. Bayesian

Frequentist:

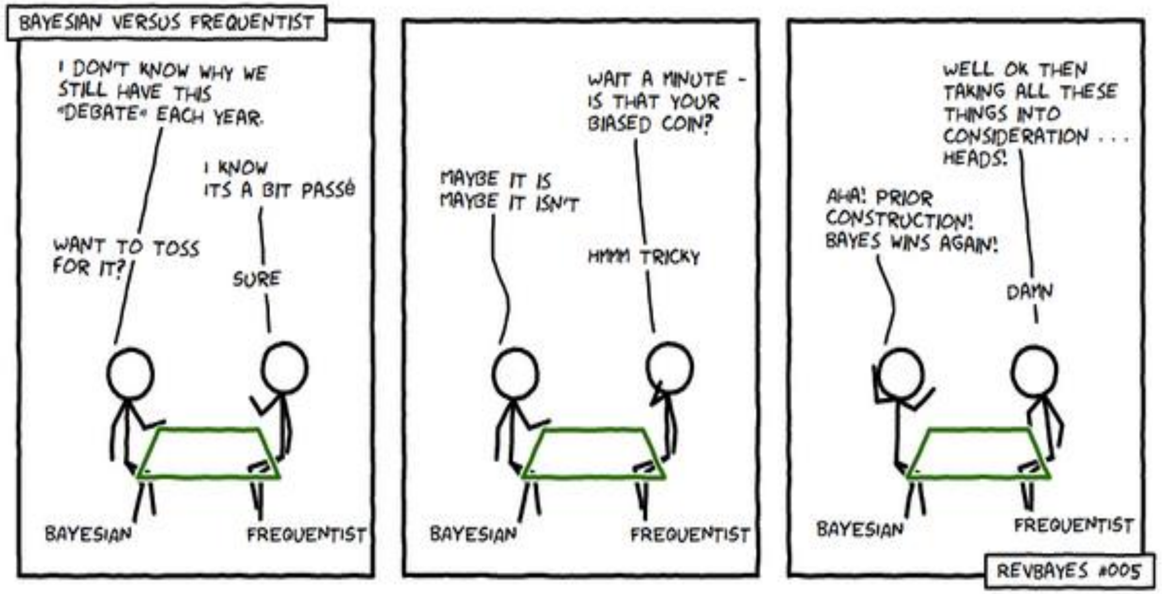
- Probabilities are defined as long term frequencies
- Model parameters are treated to be fixed while data itself is treated as the variable
- "The probability distribution of a parameter" has no meaning in this context
- 95% confidence interval means very little when we cannot repeat the experiments
- Can answer the questions regarding the coin flipping and dice rolling but has no answers when it comes to probabilities of events of one-time occurrence.



Frequentist vs. Bayesian

Bayesian

- Probabilities are treated as the degree of beliefs
- The prior knowledge of the experiments when translated into probability distribution functions have great implications when working with a limited dataset.
- 95% credible interval actually means what it sounds like
- **At infinitely large data sets, they both are in accord.**



Bayes theorem

Conditional probabilities: $P(a,b) = P(a|b)\,P(b) = P(a)P(b|a)$

Thus, $P(a|b) = P(a,b)/P(b) = P(a)P(b|a)/P(b)$

But what is P(b)? $P(b) = \sum_a P(a,b) = \sum_a P(a)P(b|a)$

$$\boxed{P(a|b) = \frac{P(a)P(b|a)}{P(b)} = \frac{P(a)P(b|a)}{\sum_{a^*} P(a^*)P(b|a^*)}}$$

- **Prior**
- **Posterior**
- **Likelihood**
- **Evidence**

Bayesian paradigm in science

- Bayes' theorem can be usefully re-written for science as:

Posterior: probability of the model given the data

Likelihood: probability of the data given the model

Prior: probability of the model

$$P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model})\,P(\text{model})}{P(\text{data})}$$

Evidence [can typically be absorbed into the normalization of the posterior]

Main Elements of Bayesian Models

- **Prior Distribution** – use probability to quantify uncertainty about unknown quantities (parameters)

- **Likelihood** – relates all variables into a “full probability model”

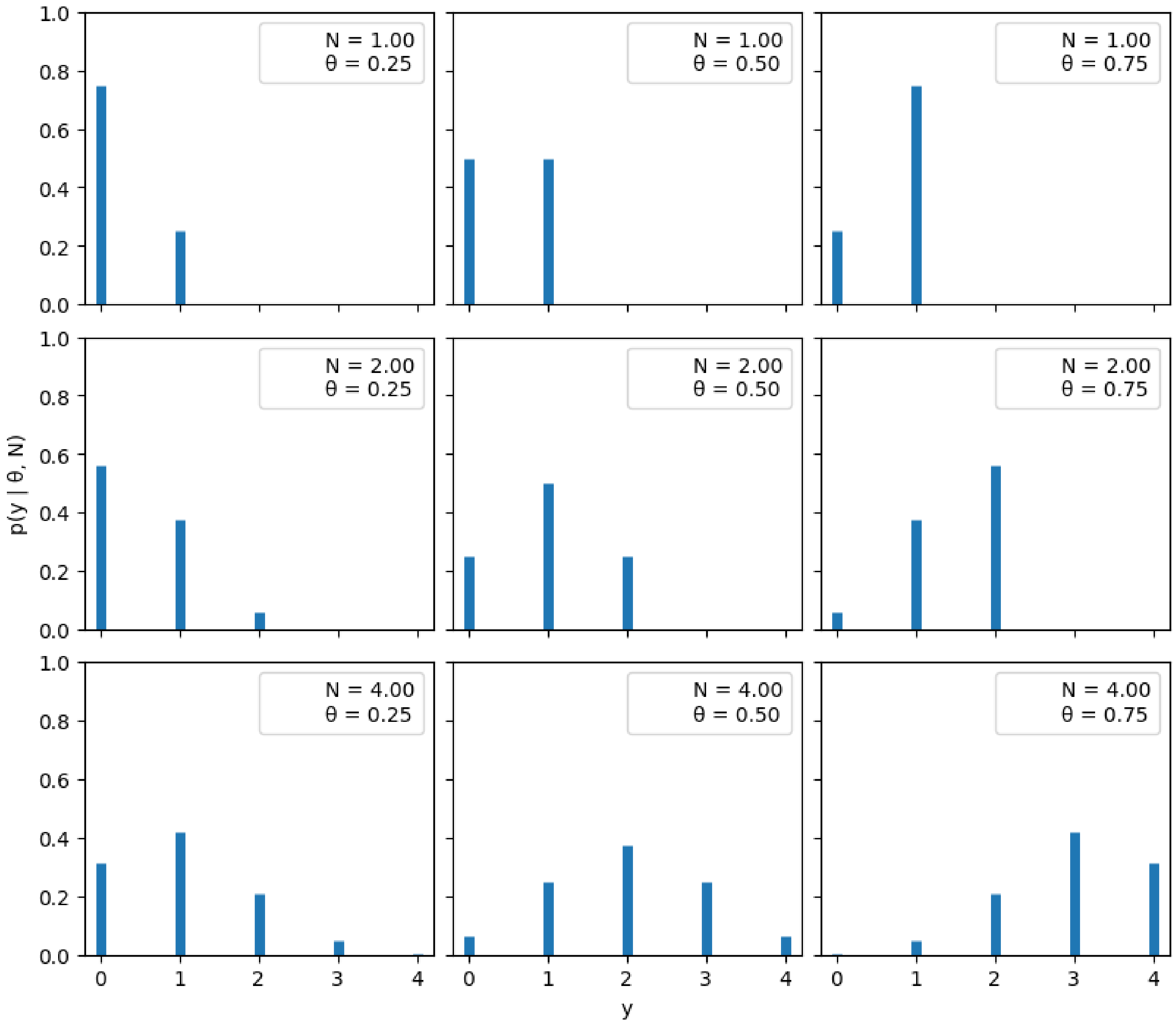
- **Posterior Distribution** – result of using data to update information about unknown quantities (parameters)

Coin Toss

- **Probability of tails:** p
- **Probability of heads:** 1-p

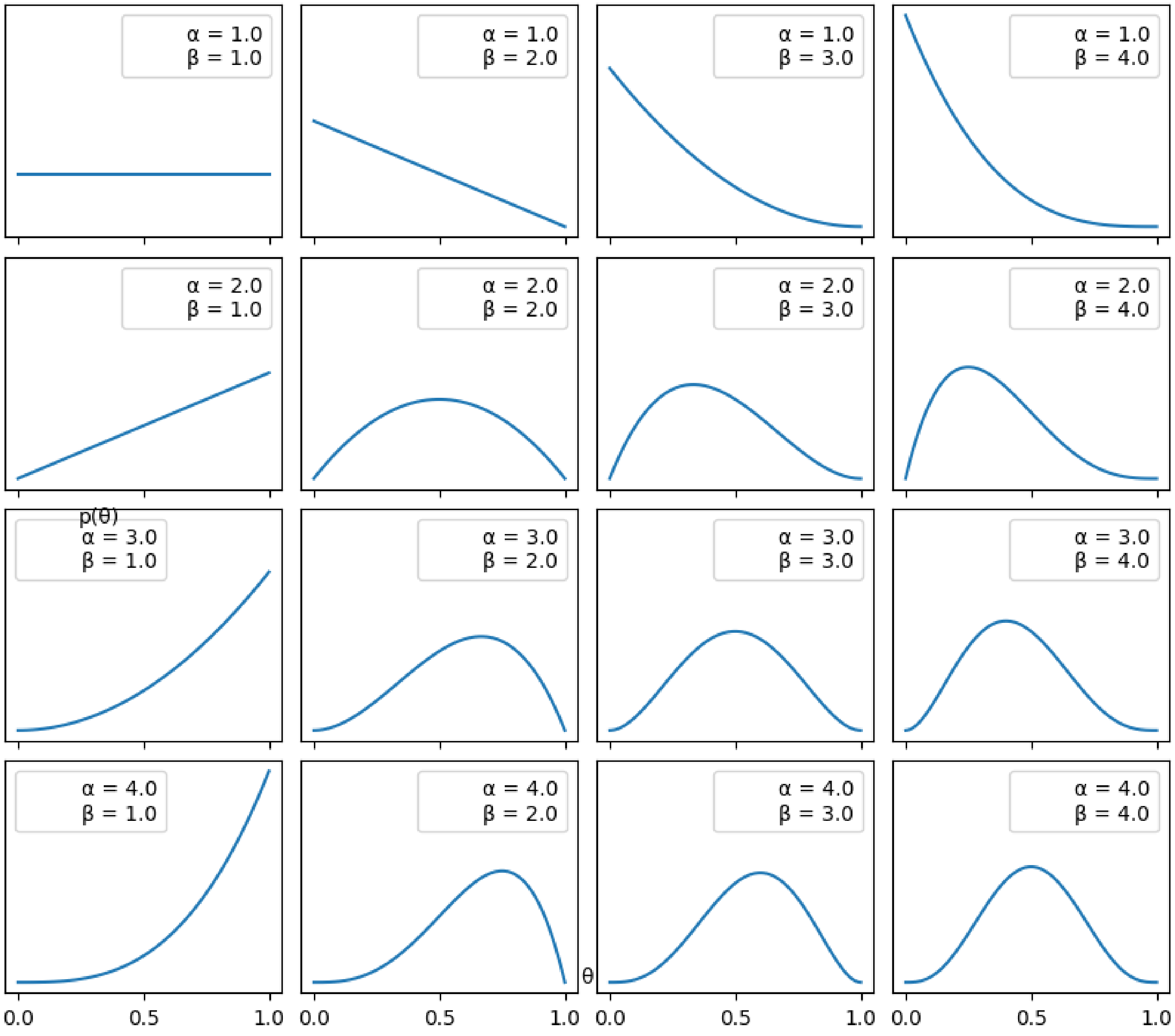
- Probability of getting n tails out of N tosses of coin:

$$C_N^n p^n (1-p)^{N-n}$$



Beta distribution: conjugate to binomial

$$
\begin{aligned}
f(x;\alpha,\beta) &= \text{constant}\cdot x^{\alpha-1}(1-x)^{\beta-1} \\
&= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha-1}(1-x)^{\beta-1} \\
&= \frac{1}{\mathrm{B}(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}
\end{aligned}
$$



https://en.wikipedia.org/wiki/Beta_distribution

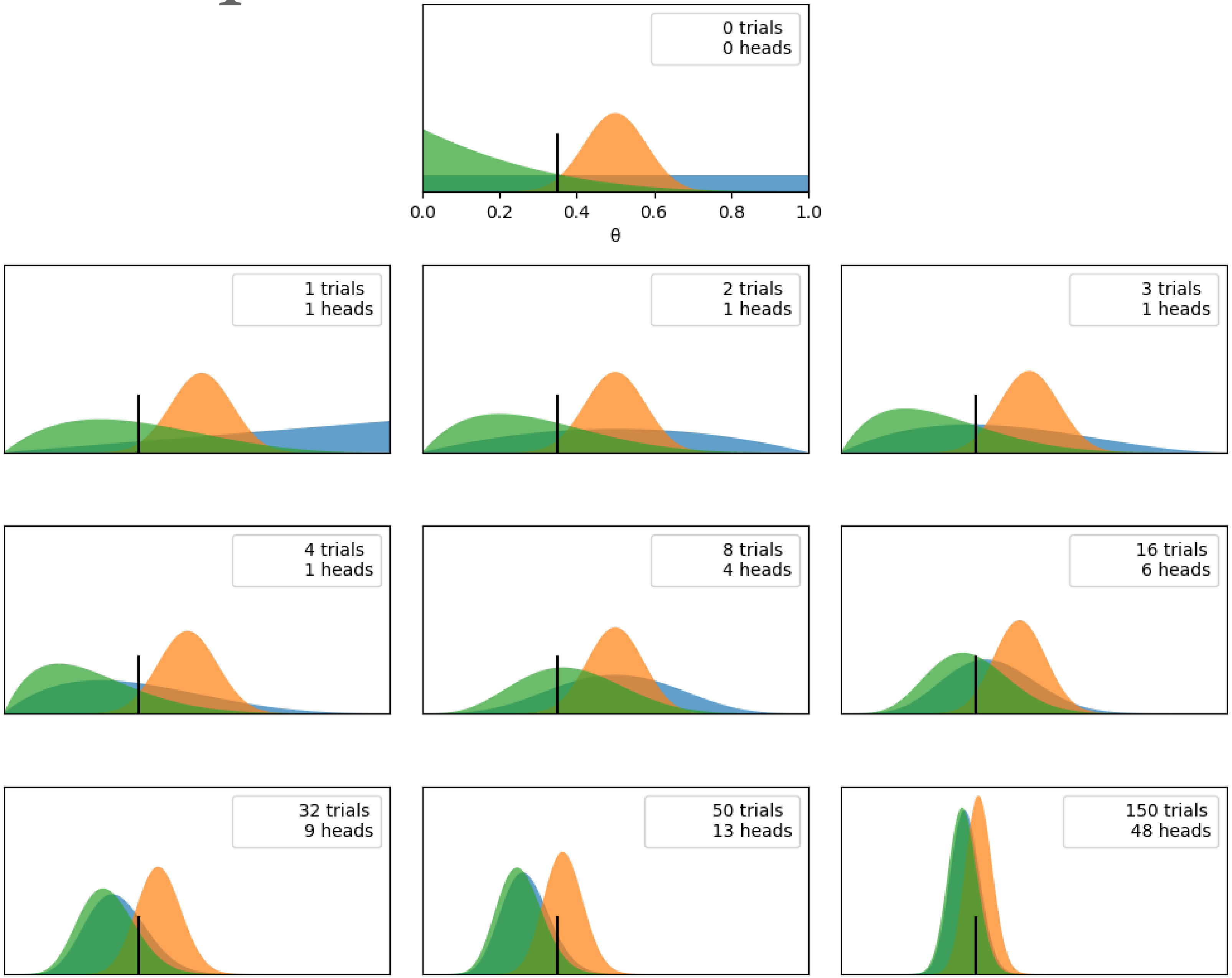
Conjugate Prior: A prior distribution is said to be conjugate to a likelihood function if the resulting posterior distribution is in the same family as the prior. In other words, if you start with a certain type of distribution as your prior, and after observing data and updating your beliefs (via Bayes' theorem), your posterior is still of that same type, then the prior is a conjugate prior for that likelihood function.

- **Computational Convenience**: Using conjugate priors can greatly simplify the mathematical computation required to find the posterior distribution. This can be especially useful in situations where you're continually updating your beliefs with new data; with conjugate priors, you can easily update your posterior without complex integrals or advanced sampling methods.
- **Analytical Solutions**: Many standard problems in Bayesian statistics can be solved analytically using conjugate priors, leading to exact posterior distributions.

Examples:

1. **Beta distribution is conjugate to the Binomial likelihood**: This means that if you have a Binomial likelihood (e.g., flipping coins) and a Beta-distributed prior on the probability of heads, the resulting posterior distribution after observing some data will also be a Beta distribution.
2. **Gamma distribution is conjugate to the Poisson likelihood**: If you're observing the number of events occurring in fixed intervals of time or space (modeled by a Poisson distribution) and have a Gamma-distributed prior on the rate parameter, the posterior will also be Gamma-distributed.
3. **Normal distribution is conjugate to itself**: If both the likelihood and the prior are normally distributed, then the posterior will also be normally distributed.

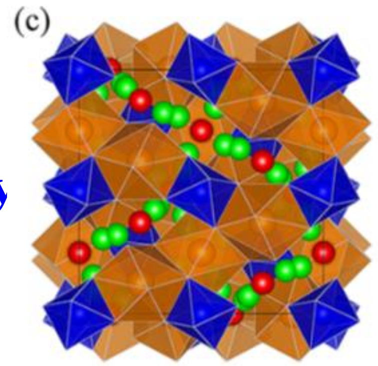
Can we learn p from several coin tosses?



Symmetry in materials science

- Crystal structure and symmetry play a crucial role in material science.
- Knowing chemical composition and crystal structure - the way atoms are arranged in space is an essential ingredient for predicting properties of a material.
  - Phase transitions
  - Order parameters
  - Physical properties
  - Vibrations and quasiparticles
- It is well known fact that the crystal structure has a direct impact on materials properties.

Cubic
High Li ion conductivity



LLZO

Tetragonal
Conductivity two orders of magnitude lower

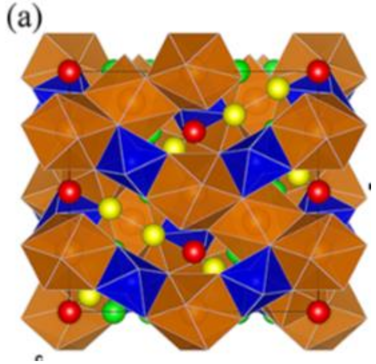


Image: J. Phys. Chem. C 2018, 122, 1963–1972

Usual way: scattering

- Diffraction techniques were predominantly used to identify crystal structure and symmetry in condensed matter physics community.

- These techniques have been successfully applied to atomic[1], magnetic[2], superconducting vortex[3] and protein[4] lattices and their structures were disinterred.

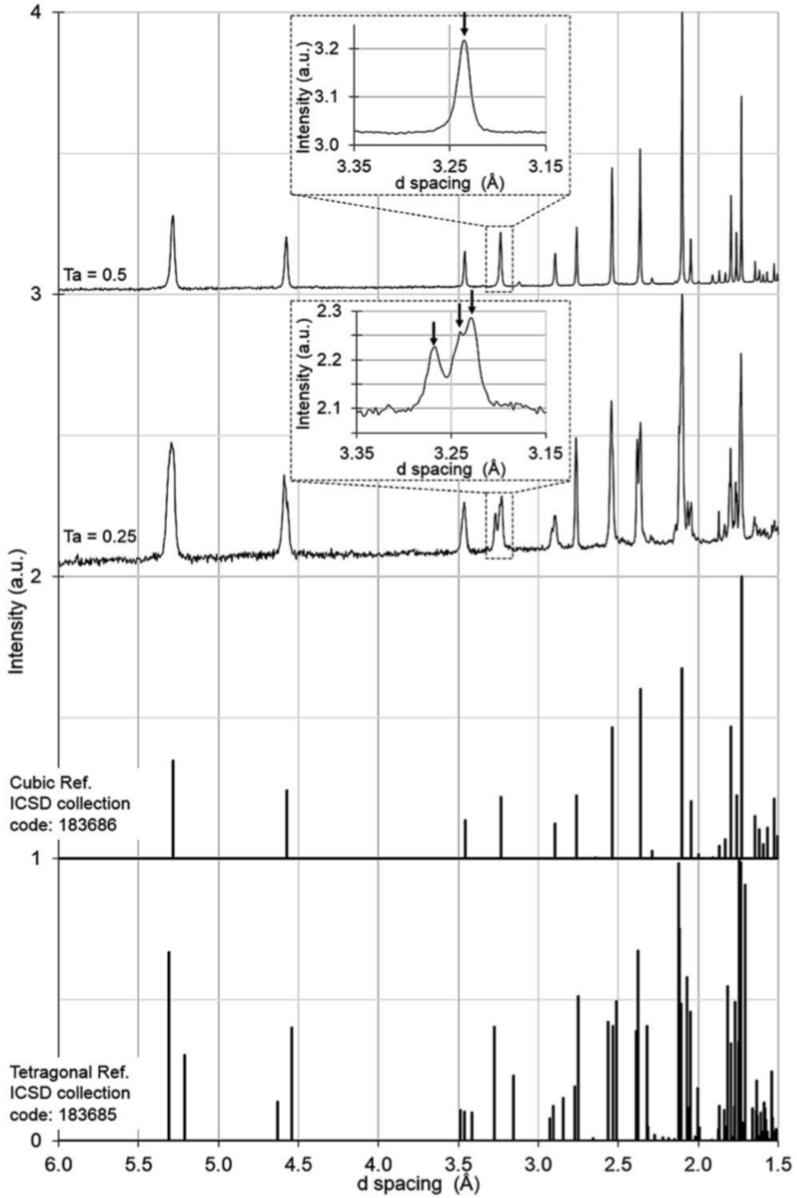
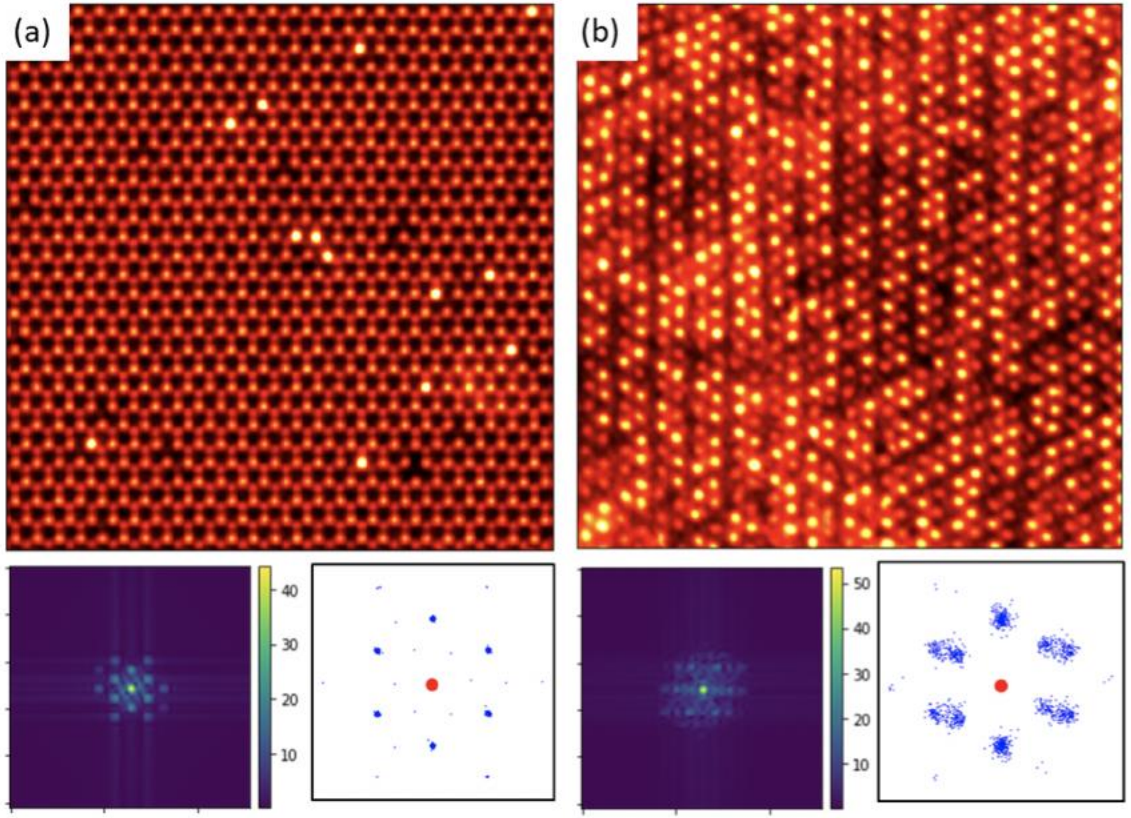


Image: J. Mater. Chem. A, 2014, 2, 13431–13436

1) Journal of Applied Crystallography, 1969. **2**(2): p. 65-71.
2) Physica B: Condensed Matter, 1993. **192**(1): p. 55-69.
3) Reports on Progress in Physics, 2011. **74**(12)
4) Nature, 1958. **181**(4610): p. 662-666.

… But what about microscopy?

- Scanning Transmission Electron Microscopy has gained significant traction in the last decades and resolution in the order of picometers was achieved.
- Real space images or Fourier transformed images can now be used to determine symmetry in crystal structures.
- Machine learning techniques (mostly CNNs) were applied on real space and k-space images to determine underlying symmetries.
- But: in all cases we essentially apply macroscopic criterion developed for systems with large number of atoms to systems with small number of atoms
- What is the limit?

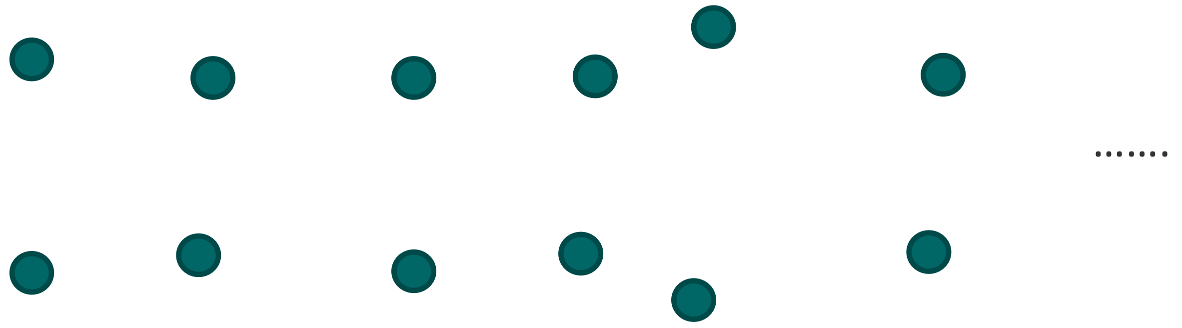


High-resolution scanning transmission electron microscopy images of $Mo_{1-x}Ru_xS_2$ with x = 0.05, 0.55 and their corresponding Fourier transforms.

… But what about microscopy?

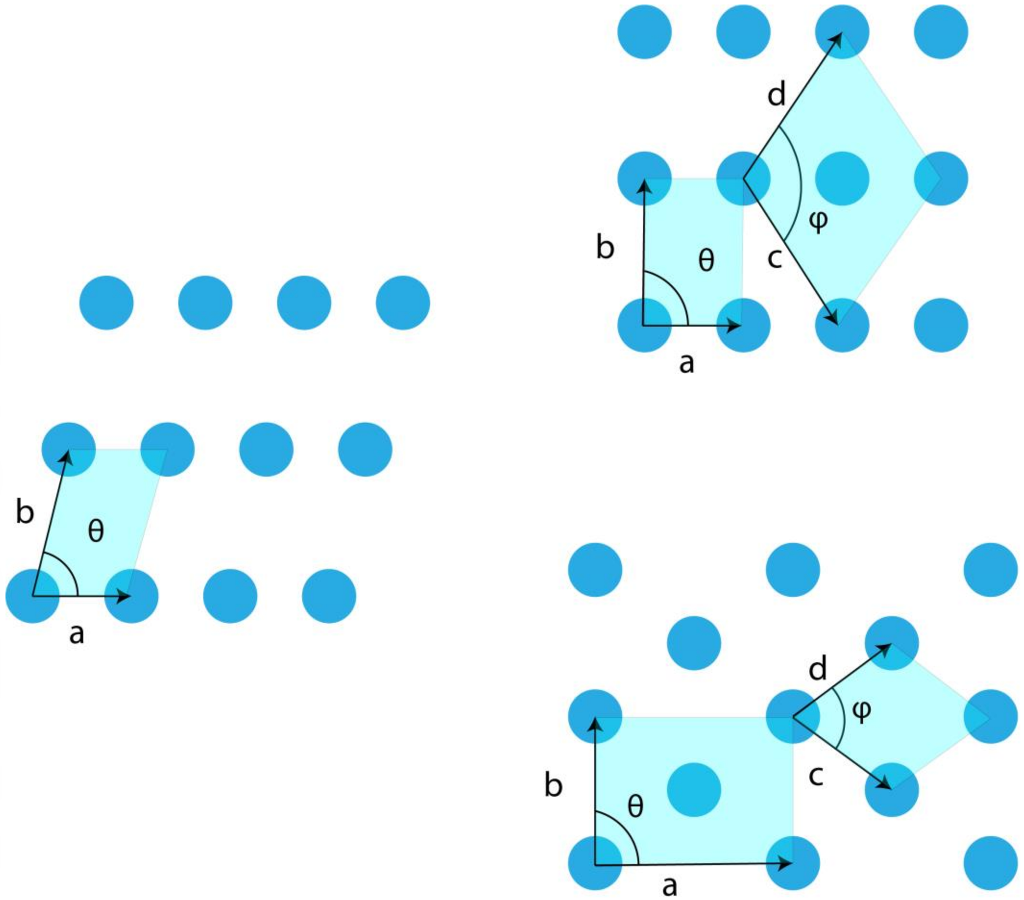
- Can we start talking symmetry as a microscopic property?

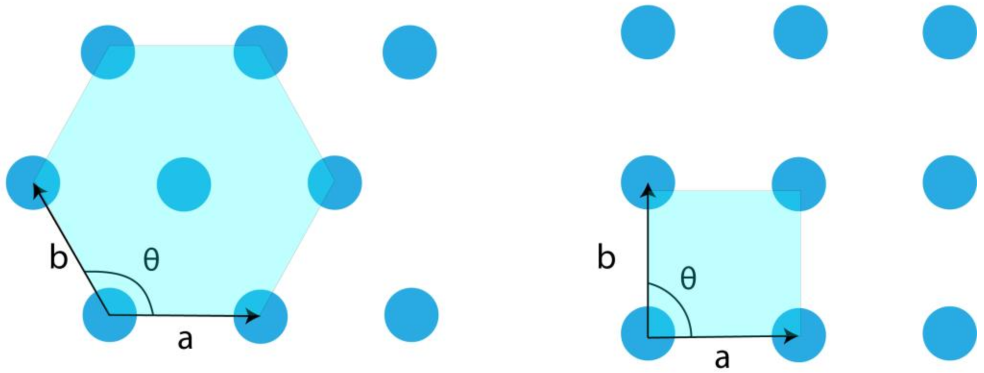
- If so, how many lattice units do we need before we define a particular symmetry in a crystal?



2D Bravais lattices

Let's start by the definition (Kittel):
Look at lattice parameters:
Square?: a = b, $\boldsymbol{\theta}$ = 90
Rectangular, a != b,
And so on…

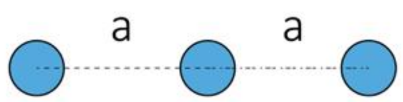




- In macroscopic systems, these equalities determine peak splitting in scattering data
- What about real space images?
- Especially when our data is limited?

2D Bravais lattices

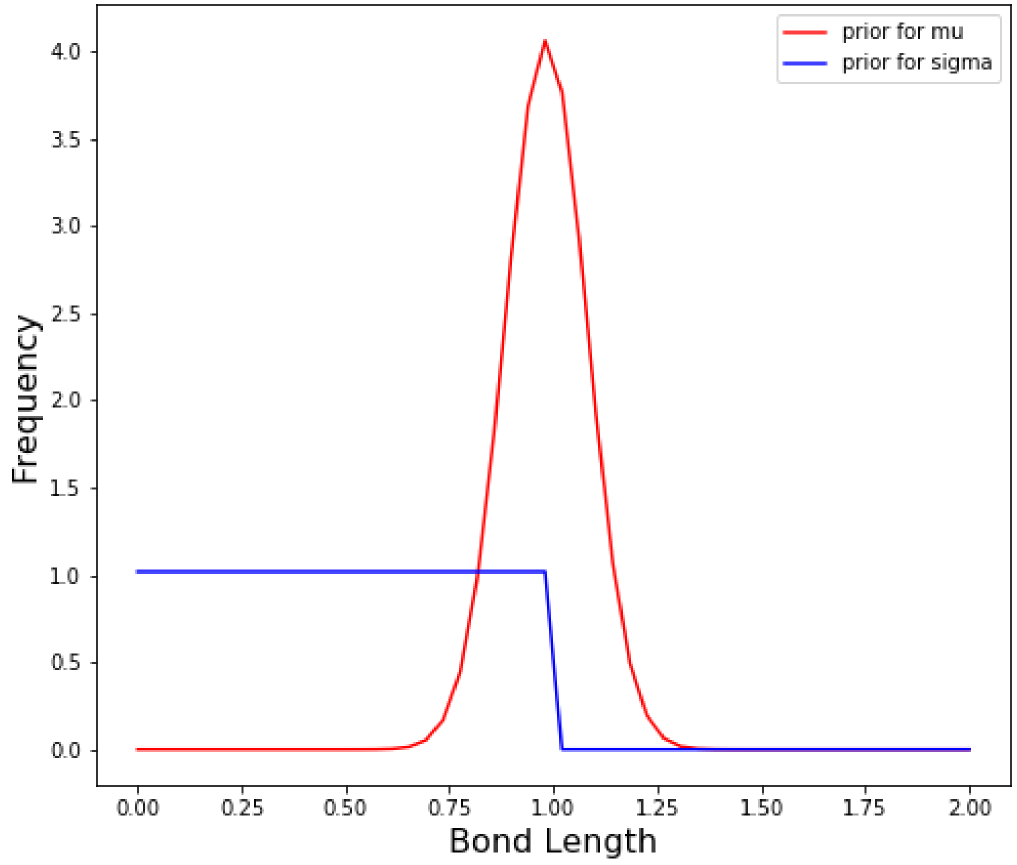
- An array of hundred bond lengths is generated using a normal distribution
- Mean of the Gaussian is $\boldsymbol{\mu}^* = 1.0$
- Whereas standard deviation is $\boldsymbol{\sigma}^* = 0.1$
- Standard deviation represents the bond disorder present in the system whereas mean represents the average bond length



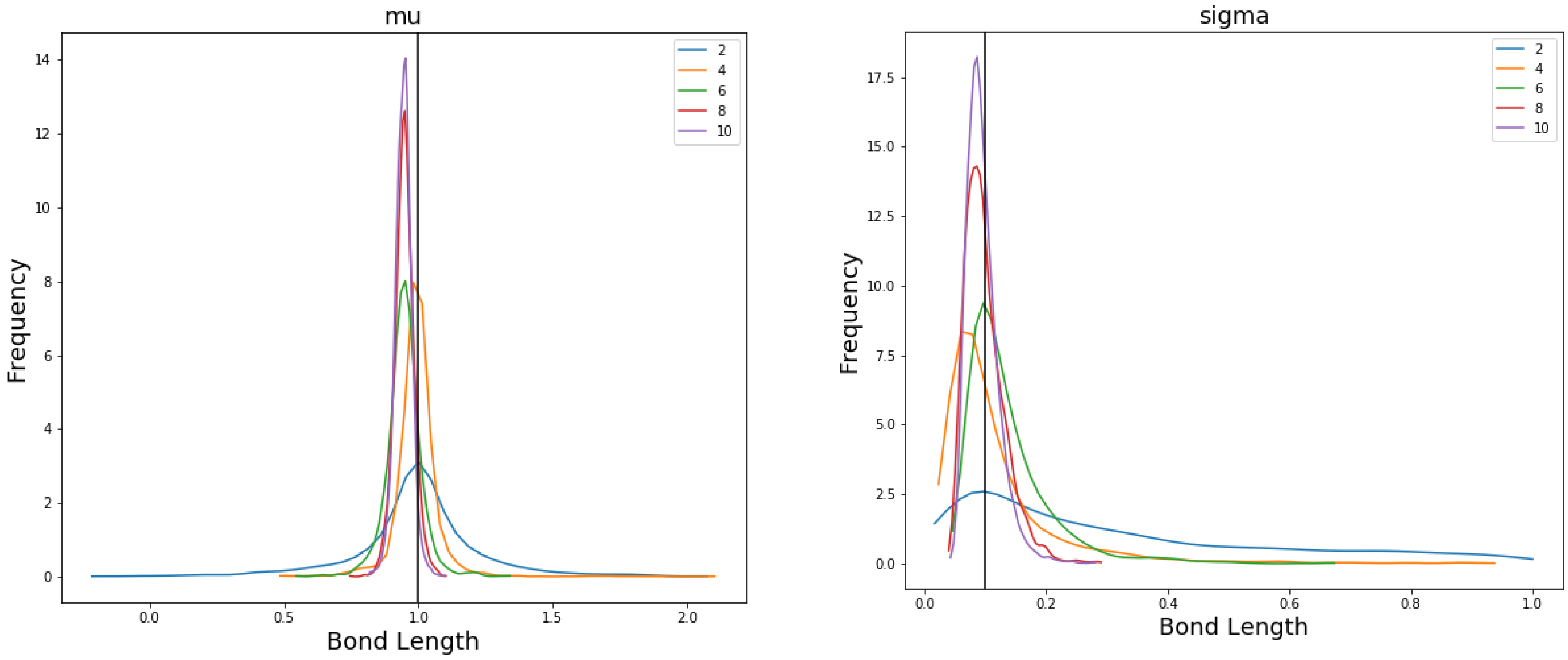
1-D Array (100 bonds)

Priors

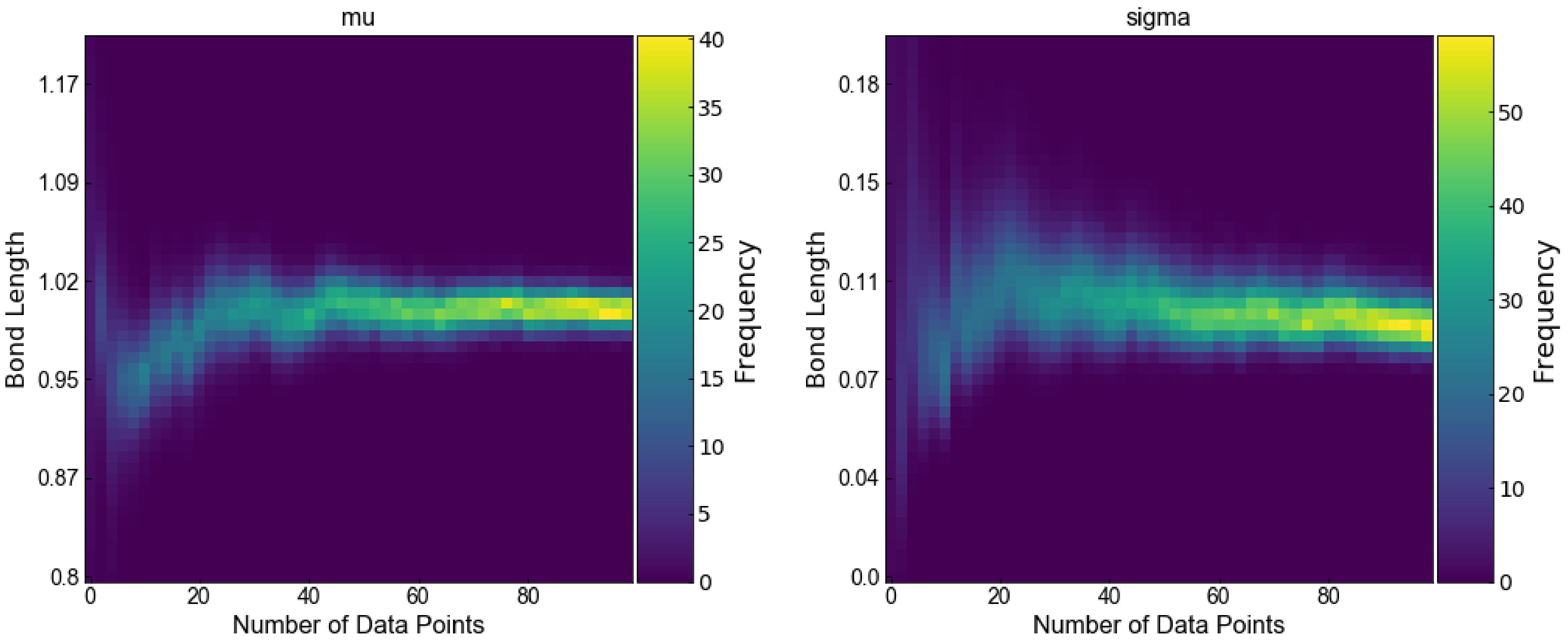
- Prior information of microscopes and of the material under consideration can be used in generating priors.
- Or if you know nothing like Jon Snow, the priors can be uniform
- For this analysis, priors are formed using the first twenty data points of the dataset generated.
- A gaussian prior for bond lengths and an uniform priors for bond disorders are considered.



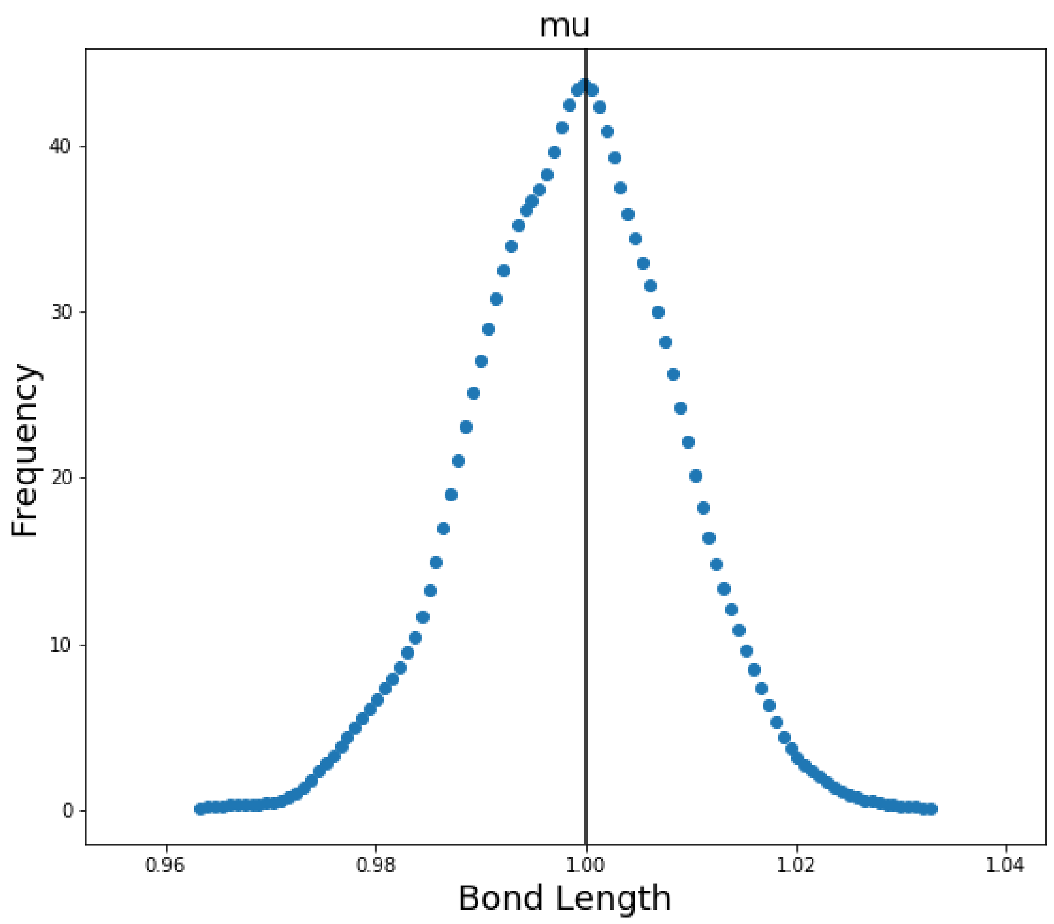
Few initial iterations

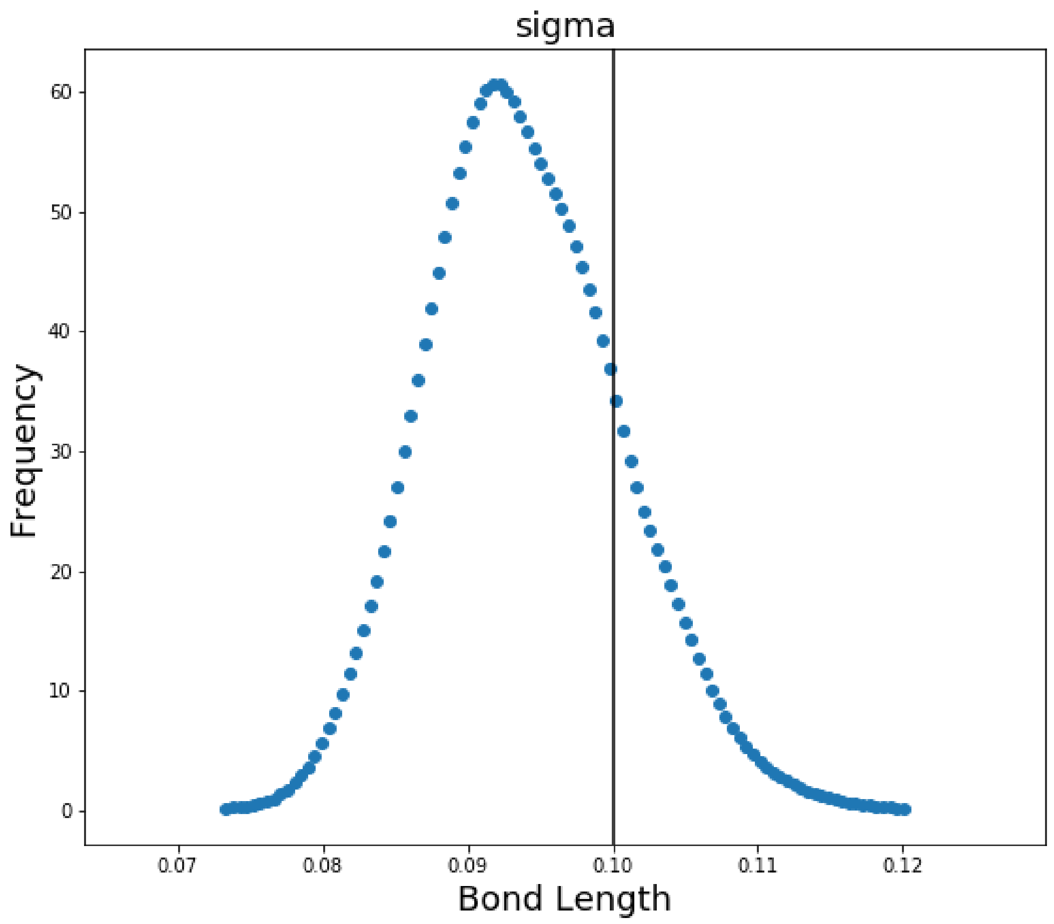


More iterations….



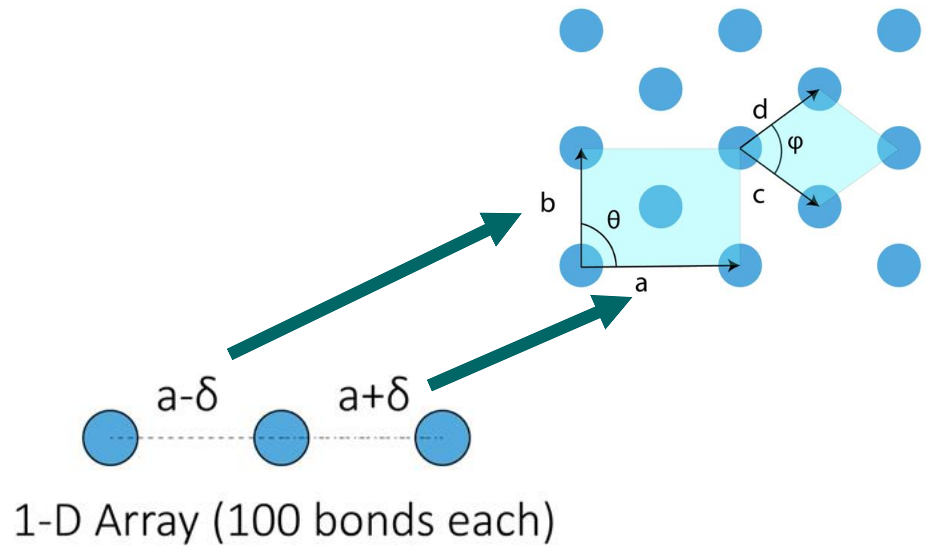
Final posterior distribution



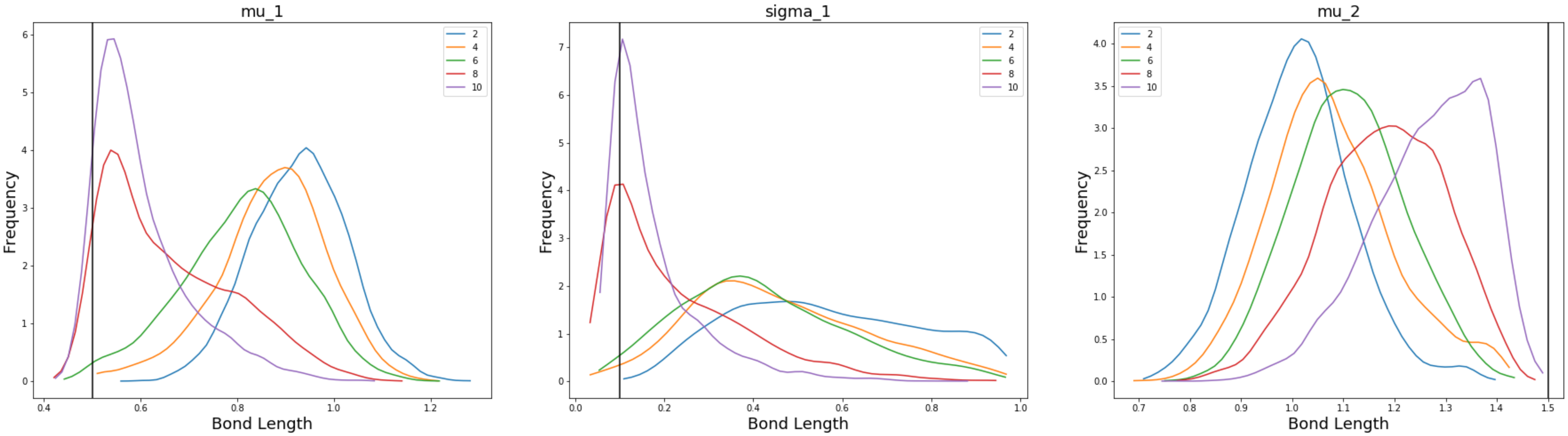


Now, let's try rectangular lattice

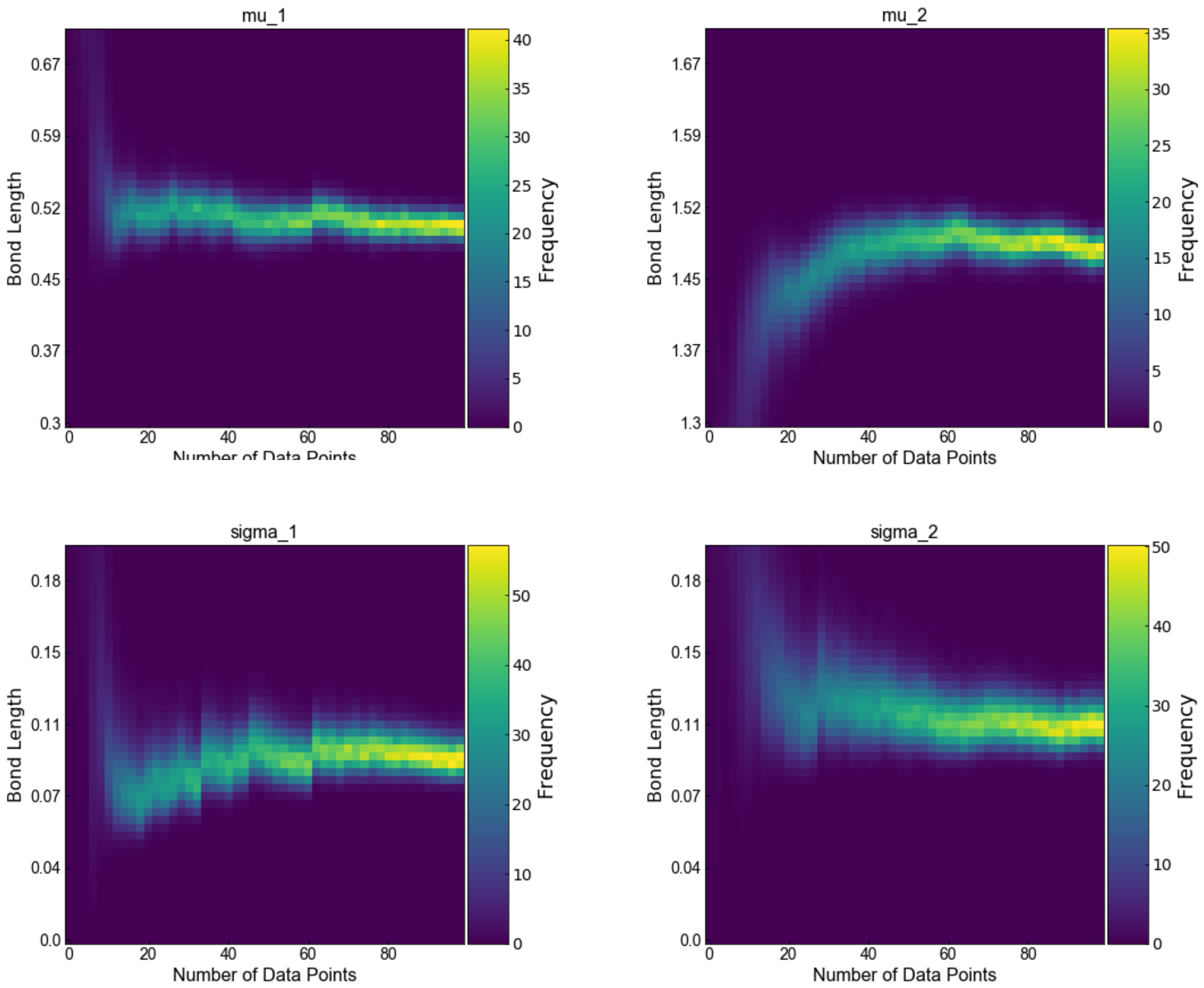
- Two arrays of bond lengths are generated with different lattice parameters
- a = 1.0 and $\delta = 0.5$
- Means of the gaussians are $\boldsymbol{\mu}1 = a+\delta$ and $\boldsymbol{\mu}2 = a-\delta$
- Whereas standard deviation is $\boldsymbol{\sigma}^* = 0.1$
- Priors are constructed based on the first 10 data points from each set
- These bond lengths are analogous to rectangular bravais lattice



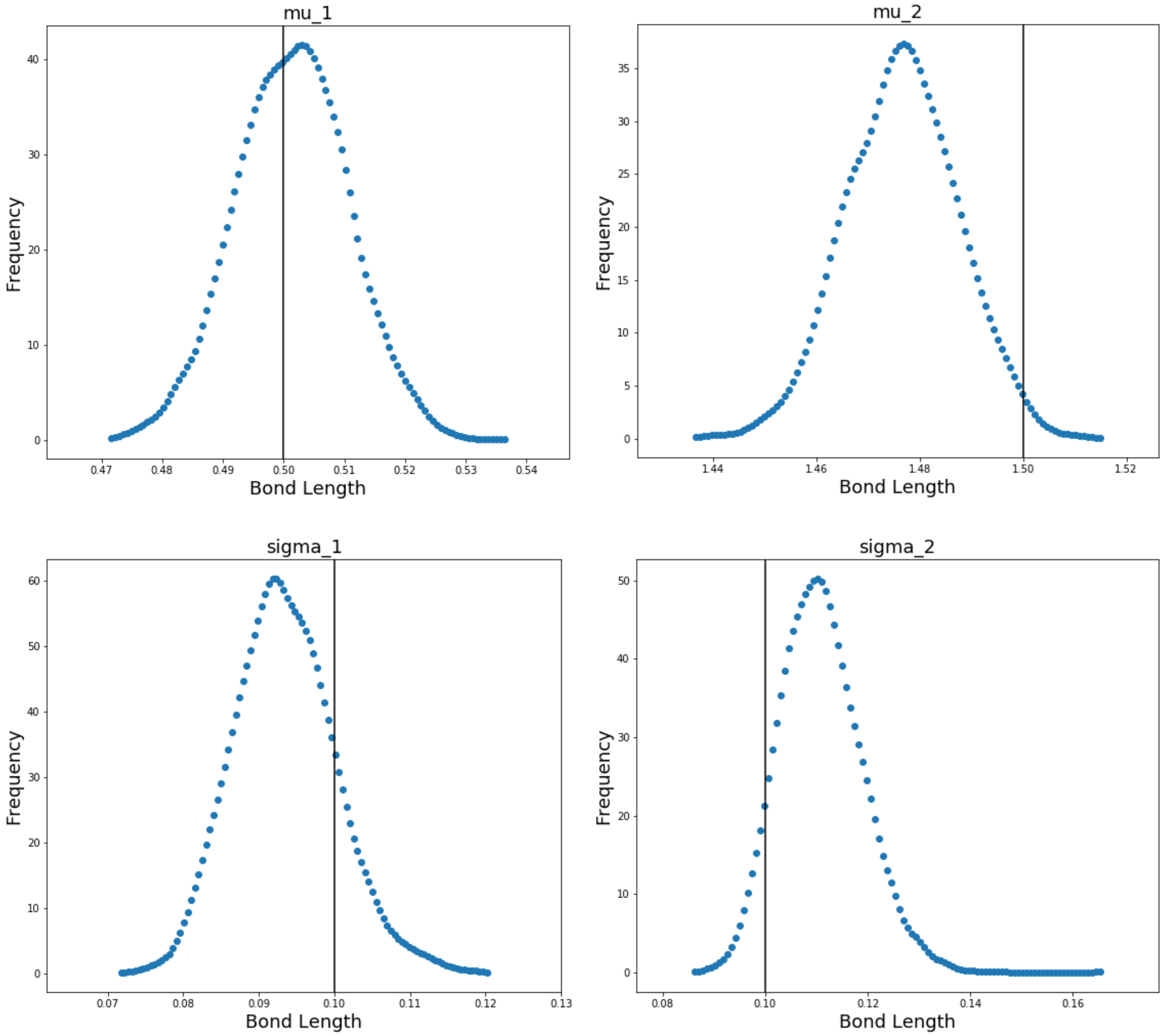
First several iterations



More iterations



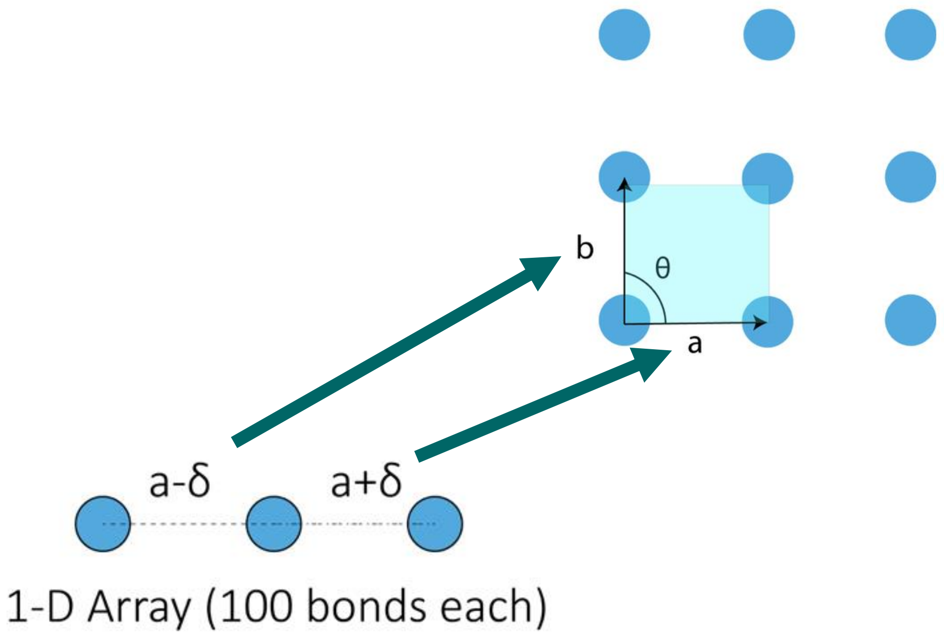
Posteriors



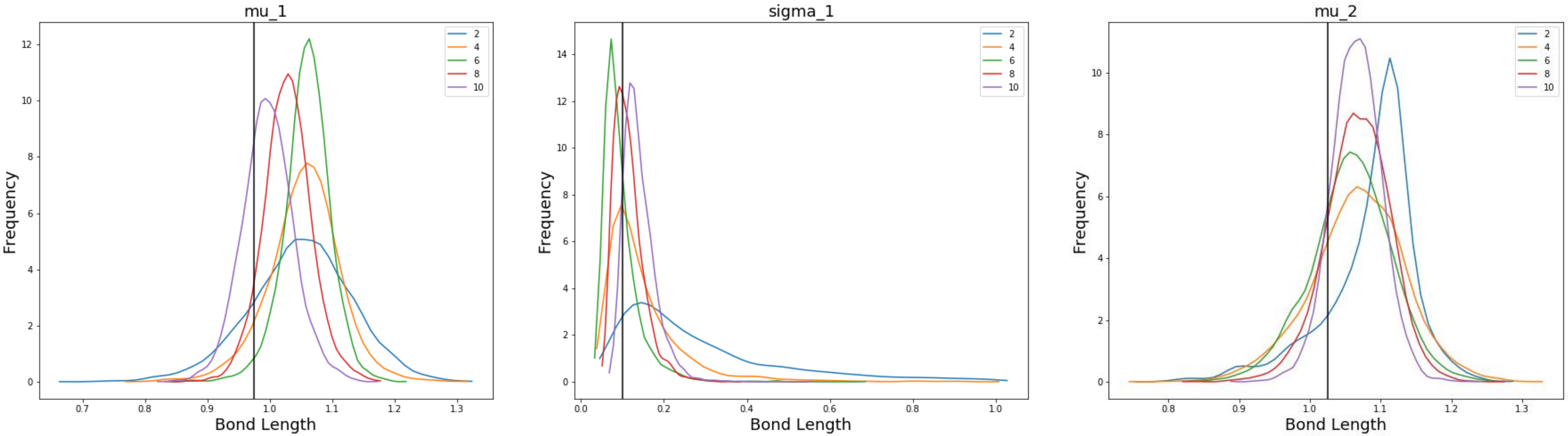
mu_1
Frequency
Bond Length
mu_2
Frequency
Bond Length
sigma_1
Frequency
Bond Length
sigma_2
Frequency
Bond Length

Pseudo-square lattice

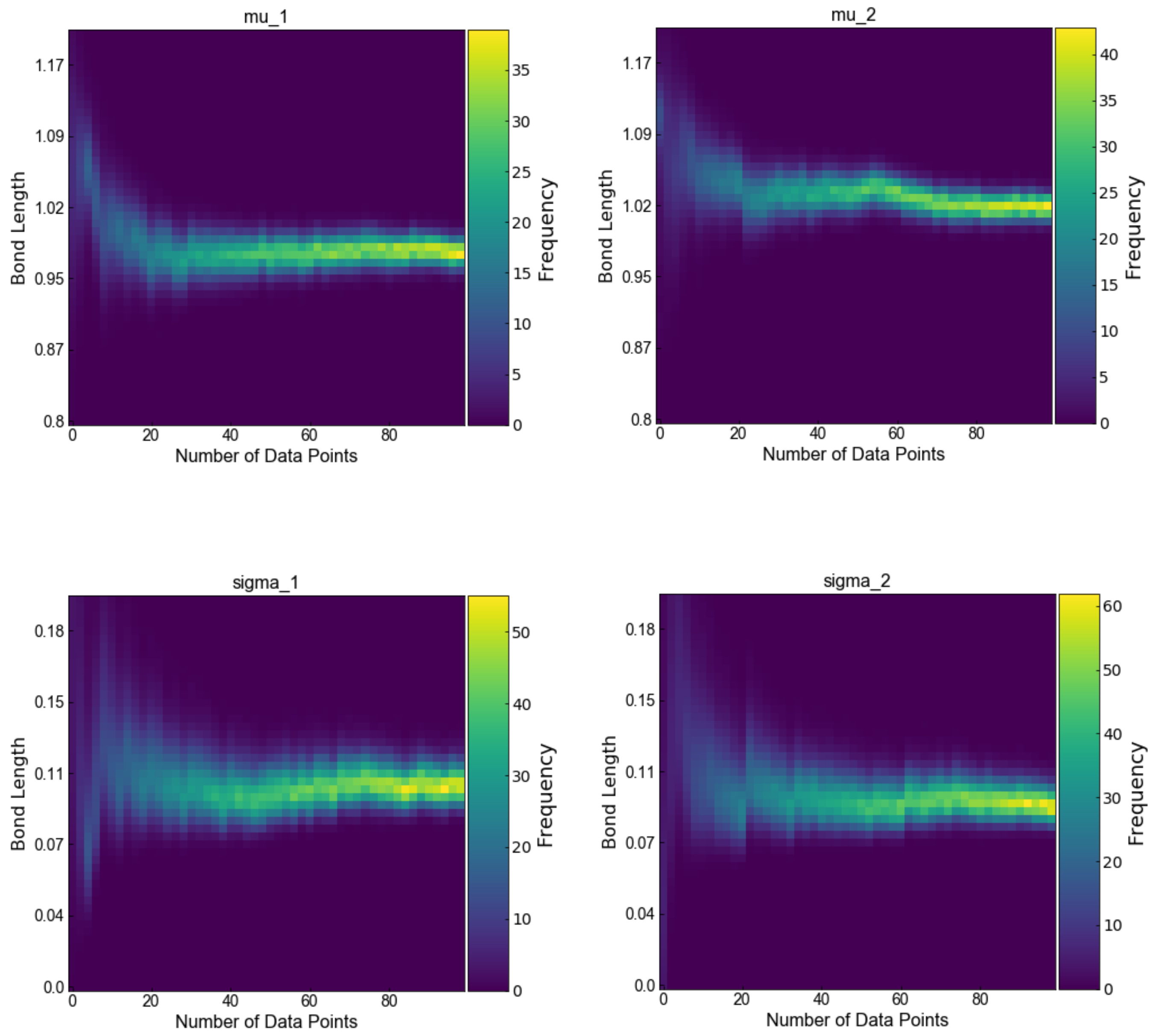
- Similar as of the previous case but the parameters are close to each other and are comparable to the bond disorder
- a = 1.0 and $\delta = 0.025$
- Standard deviation is $\boldsymbol{\sigma}^* = 0.1$
- Priors are constructed based on the first 10 data points from each set
- These bond lengths are analogous to square Bravais lattice



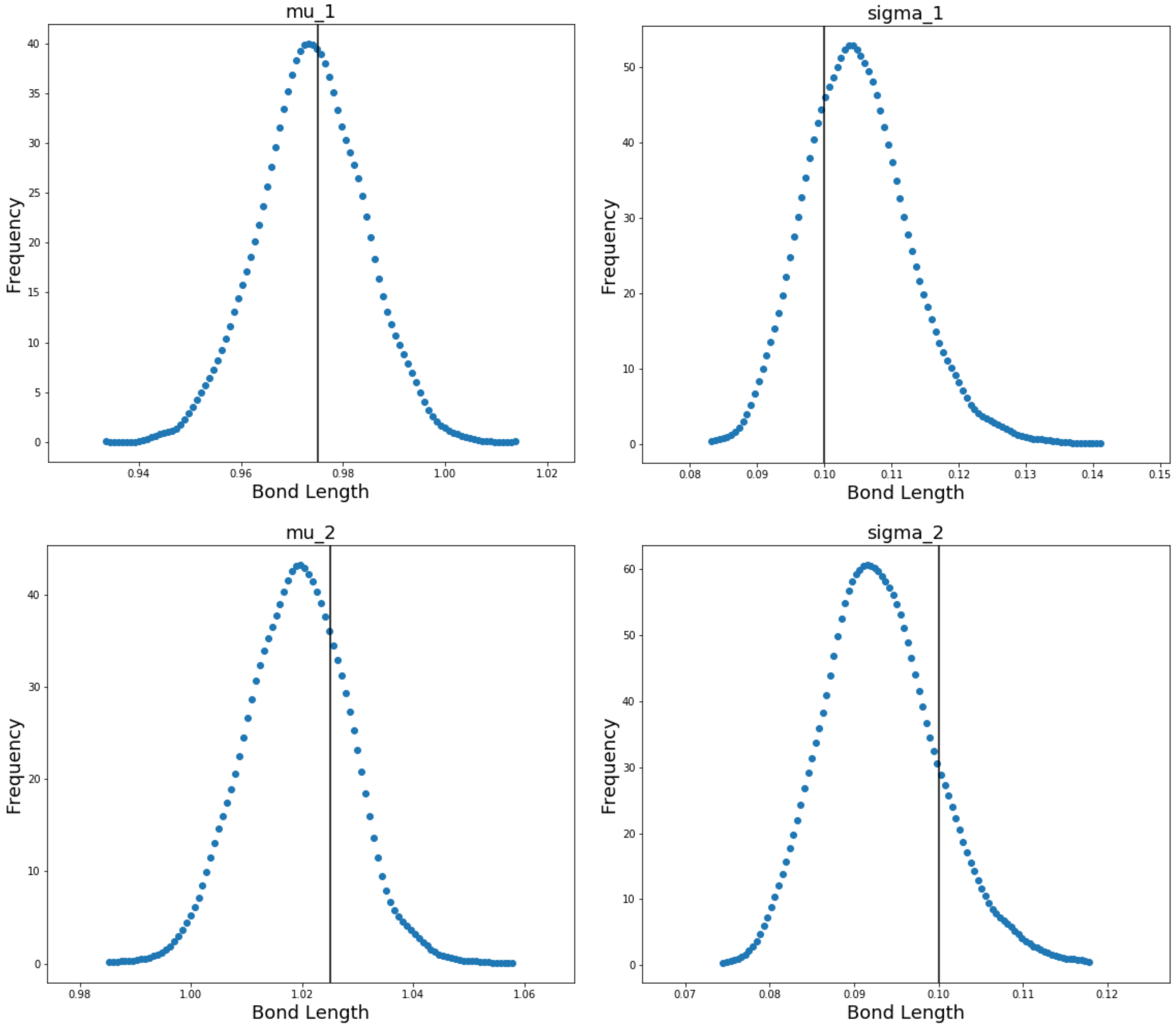
First iterations



More iterations

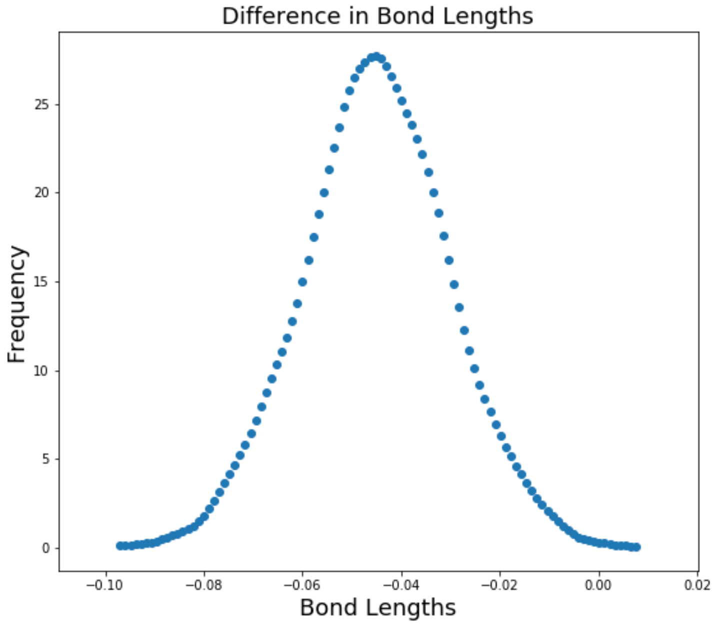


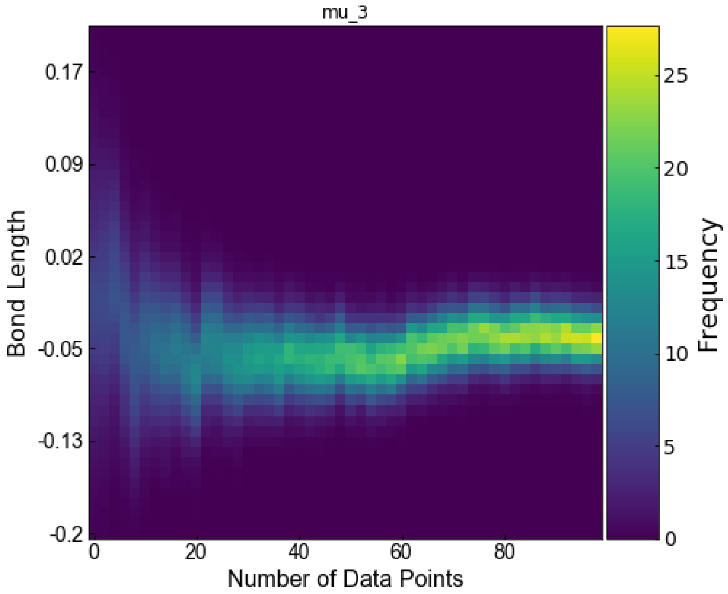
Posteriors


mu_1
sigma_1
mu_2
sigma_2
Frequency
Bond Length
0.94
0.96
0.98
1.00
1.02
0.08
0.09
0.10
0.11
0.12
0.13
0.14
0.15
0.98
1.00
1.02
1.04
1.06
0.07
0.08
0.09
0.10
0.11
0.12
0
5
10
15
20
25
30
35
40
50
60

At which point we say lattice is square?

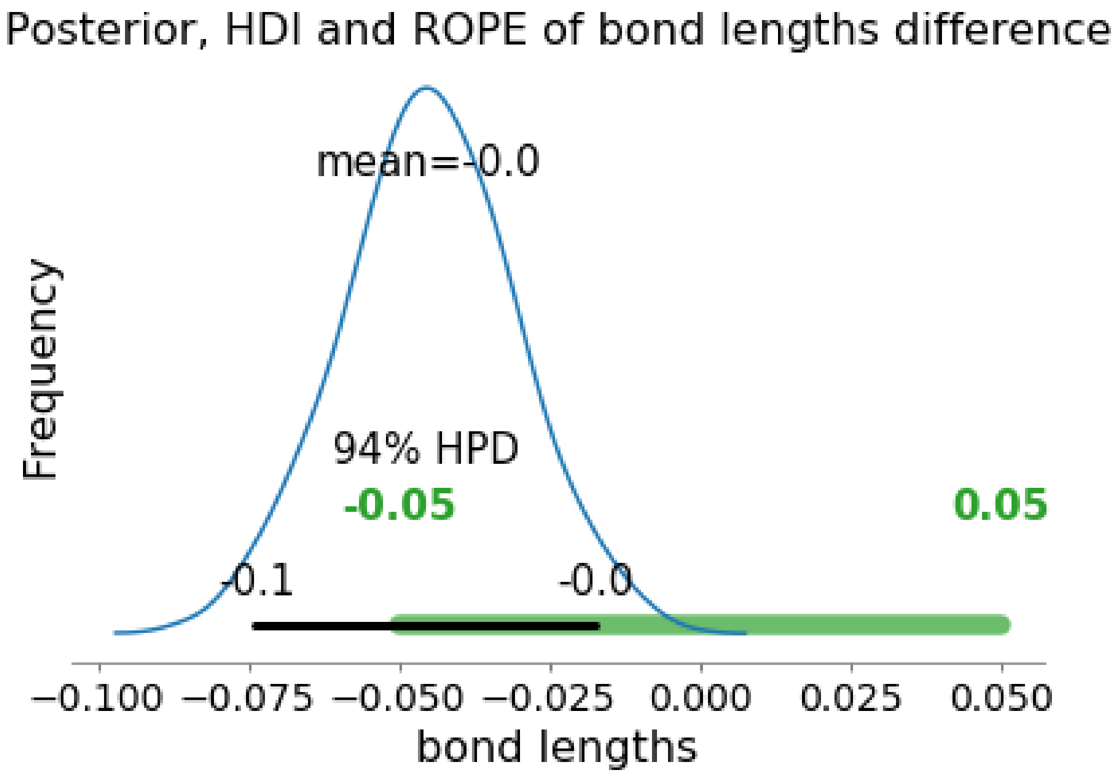
- Posterior distributions can be analyzed in many different ways
- Subjective to different phenomena like prior knowledge of the material or the person doing the analysis etc.
- Consider the posterior distribution of difference of bond lengths $\boldsymbol{\mu}_3 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$





Decision making: ROPE

- We construct an interval (ROPE) around the hypothesis and a decision can be made by comparing the intervals of HDI (94% credible interval) and ROPE.
- Decision rules for different criteria are listed in “Bayesian Analysis with Python” by Osvaldo Martin



Priors are the key!

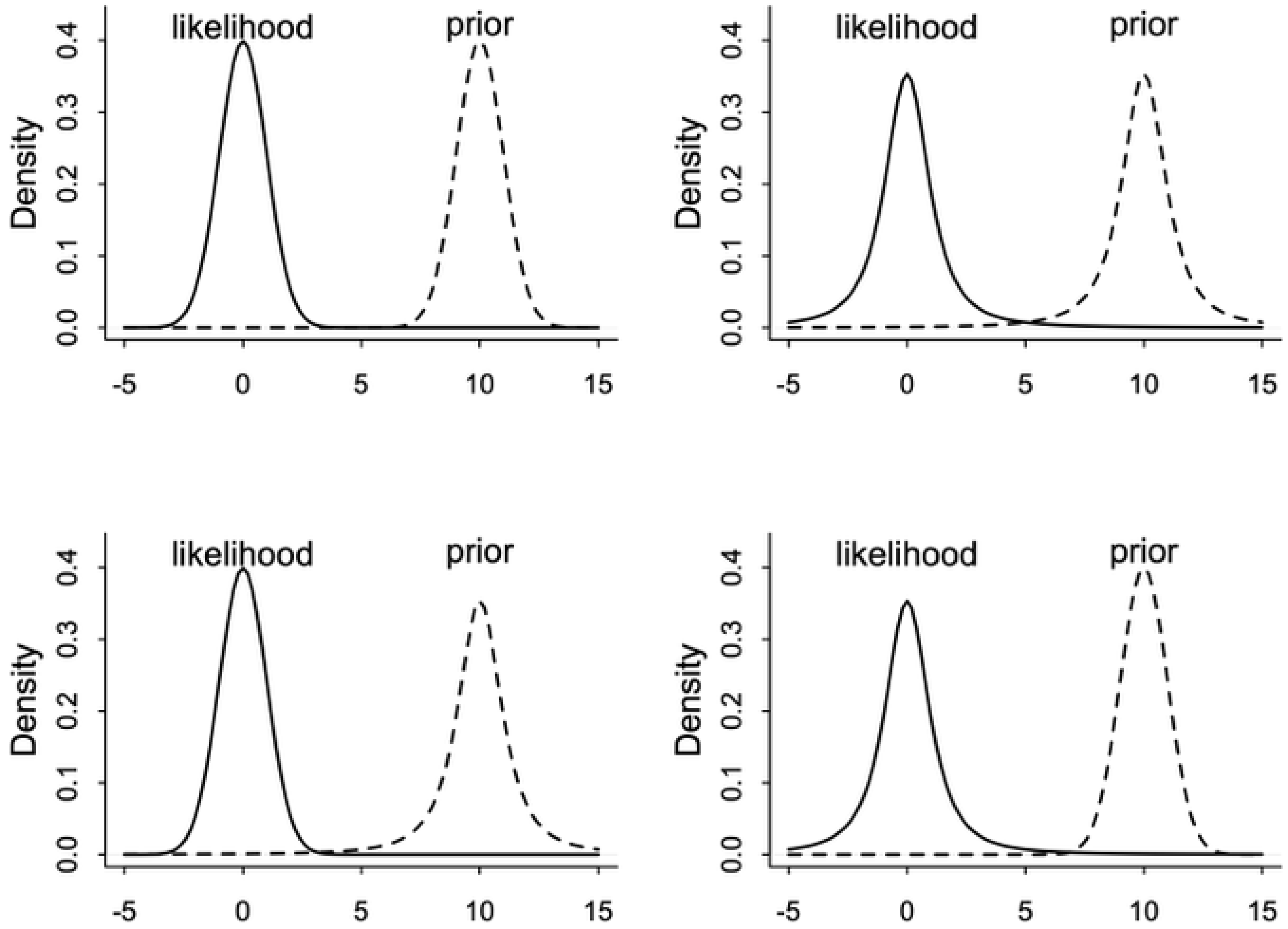
- Most Bayesian analysts assume "uninformative priors"
  - no strong assumptions about the parameter estimates other than the shape of their distributions
- When we use uninformative priors and analyze the data using both a traditional approach and a Bayesian approach, the resulting parameter estimates are the same (for all practical purposes) = *no strong rationale for Bayesian*
  - Uninformative priors means the results are strongly determined by the current experiment's data

When to use strong priors

- When you are willing to specify informative priors
- When there is no existing analysis for your design

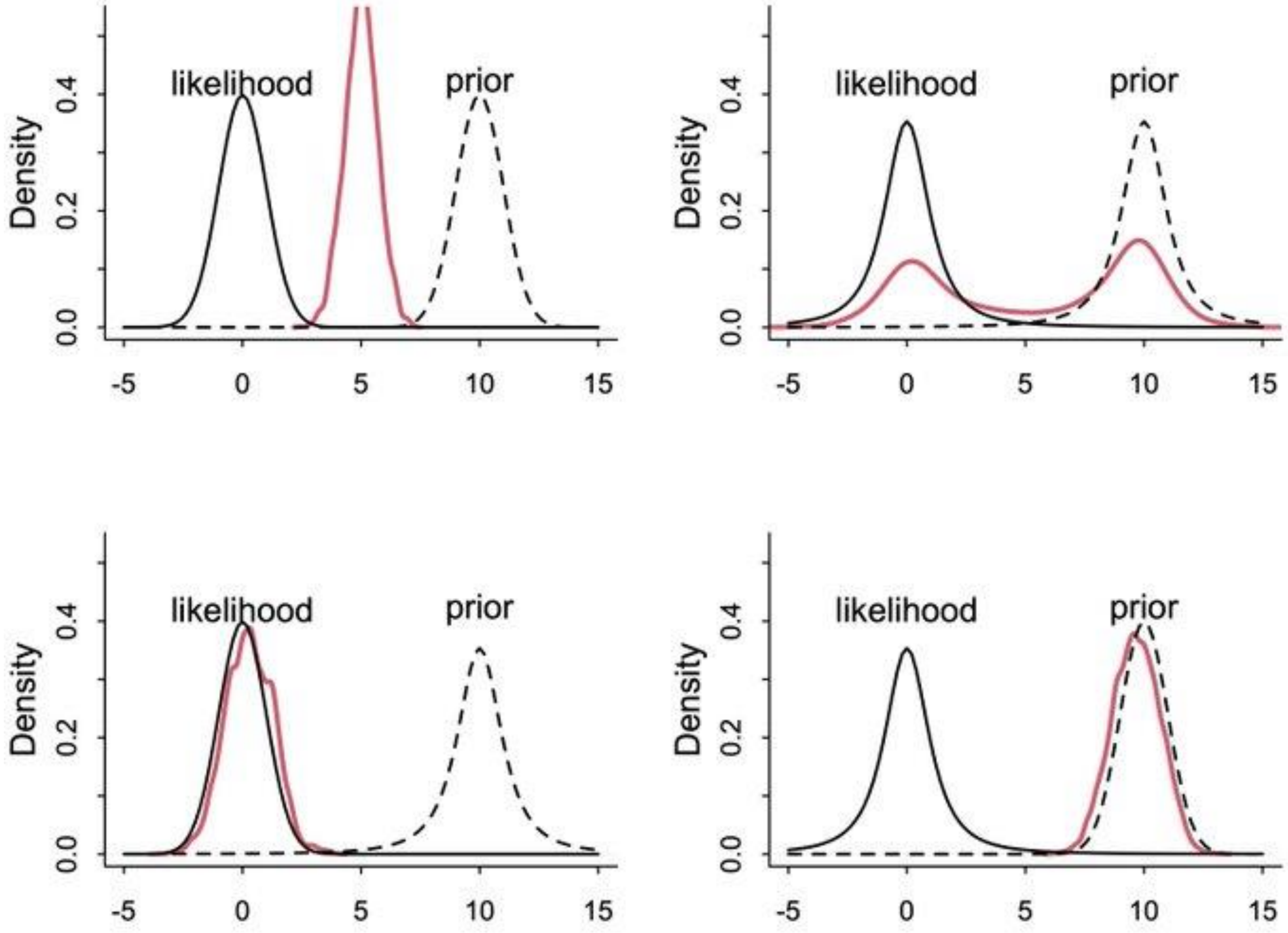
- Narrower priors will have a bigger effect on the posterior estimation
- The effect will be larger if the new data is limited or highly variable
  - This situation indicates that the new data are equivocal and offer little new

McElreath Quartet



https://twitter.com/rlmcelreath/status/1701165075493470644

McElreath Quartet



https://twitter.com/rlmcelreath/status/1701165075493470644