Bayesian Inference in systems biology

Theory

What is Bayesian inference?

Probability 101

Bayes' theorem in biology

Priors and likelihood

Interpretation of posteriors

Markov Chain Monte Carlo

Applications

Population dynamics in microscopic and macroscopic communities

Phylogeny and model selection

ODE generative models: Morphogen patterning of embryonic tissues

Bayes' Theorem

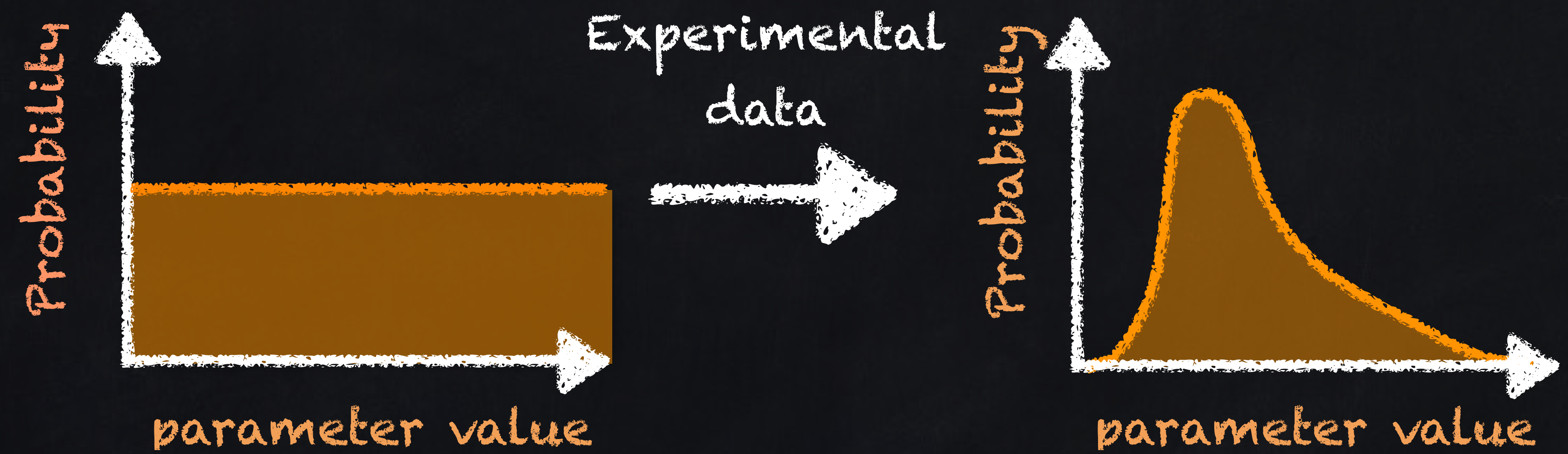
$$P(X,Y) = P(X)P(Y|X)$$
$$P(X,Y) = P(Y)P(X|Y)$$

$$\frac{\cancel{P(X,Y)}}{\cancel{P(X,Y)}} = \frac{P(X)P(Y|X)}{P(Y)P(X|Y)}$$

$$\boxed{P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)}}$$

Bayesian Inference

$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)}$$



Likelihood

Prior probability

$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

Posterior probability

Y : Experimental data

X : Parameter value (δ)

Bayesian Inference

$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)}$$

Likelihood

Prior probability

Posterior probability

$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

- What about $P(data)$? ✓
- How do I choose the prior? ✓
- What am I supposed to do with the posterior? ✓
- Okay, but how do apply this with my laptop? ✓
- Can you give examples of the likelihood? ✓

What about $P(data)$? ✓

$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

Marginal likelihood

$$P(data) = \int P(data, \delta)\mathrm{d}\delta = \int P(data|\delta)P(\delta)\mathrm{d}\delta$$

Let's compare the posterior probabilities of two values :

$$\frac{P(\delta = \delta_1|data)}{P(\delta = \delta_2|data)} = \frac{\frac{P(\delta_1)P(data|\delta_1)}{\cancel{P(data)}}}{\frac{P(\delta_2)P(data|\delta_2)}{\cancel{P(data)}}}$$

What about $P(data)$?

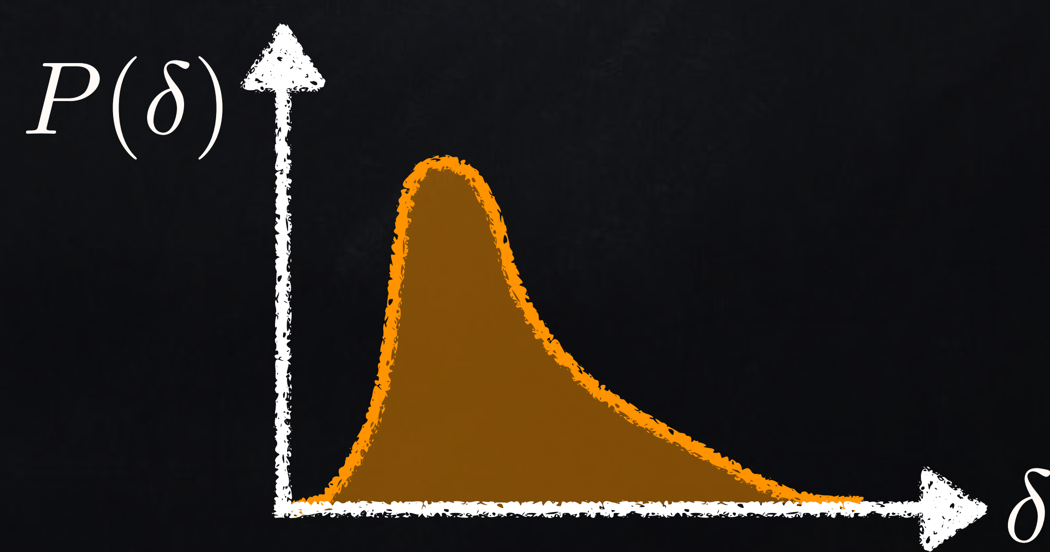
$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

In practice is not necessary

How do I choose the prior?

Reflecting your prior knowledge

- Noninformative priors
- Previous studies
- Previous Bayesian analysis



What about $P(data)$? ✔

How do I choose the prio

Refle

- Noninformative priors
- Previous studies
- Previous Bayesian analys

Question:

- Two different studies report that the mean diameter of the trunk of oaks in a particular forest

$D = 0.7 \pm 0.4$ m
$D = 1.0 \pm 0.1$ m

How would you encode this in a prior distribution?
What if the studies report:

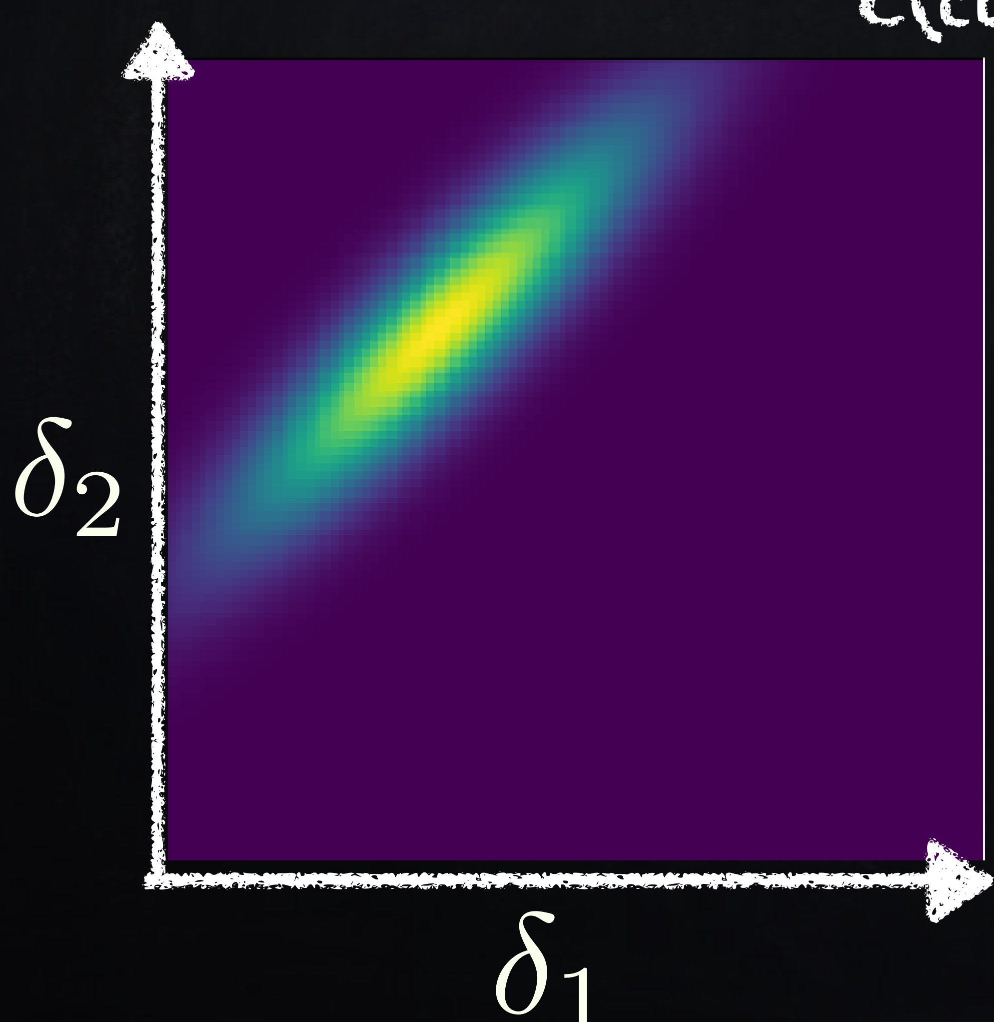
$D = 0.7 \pm 0.1$ m
$D = 1.0 \pm 0.1$ m

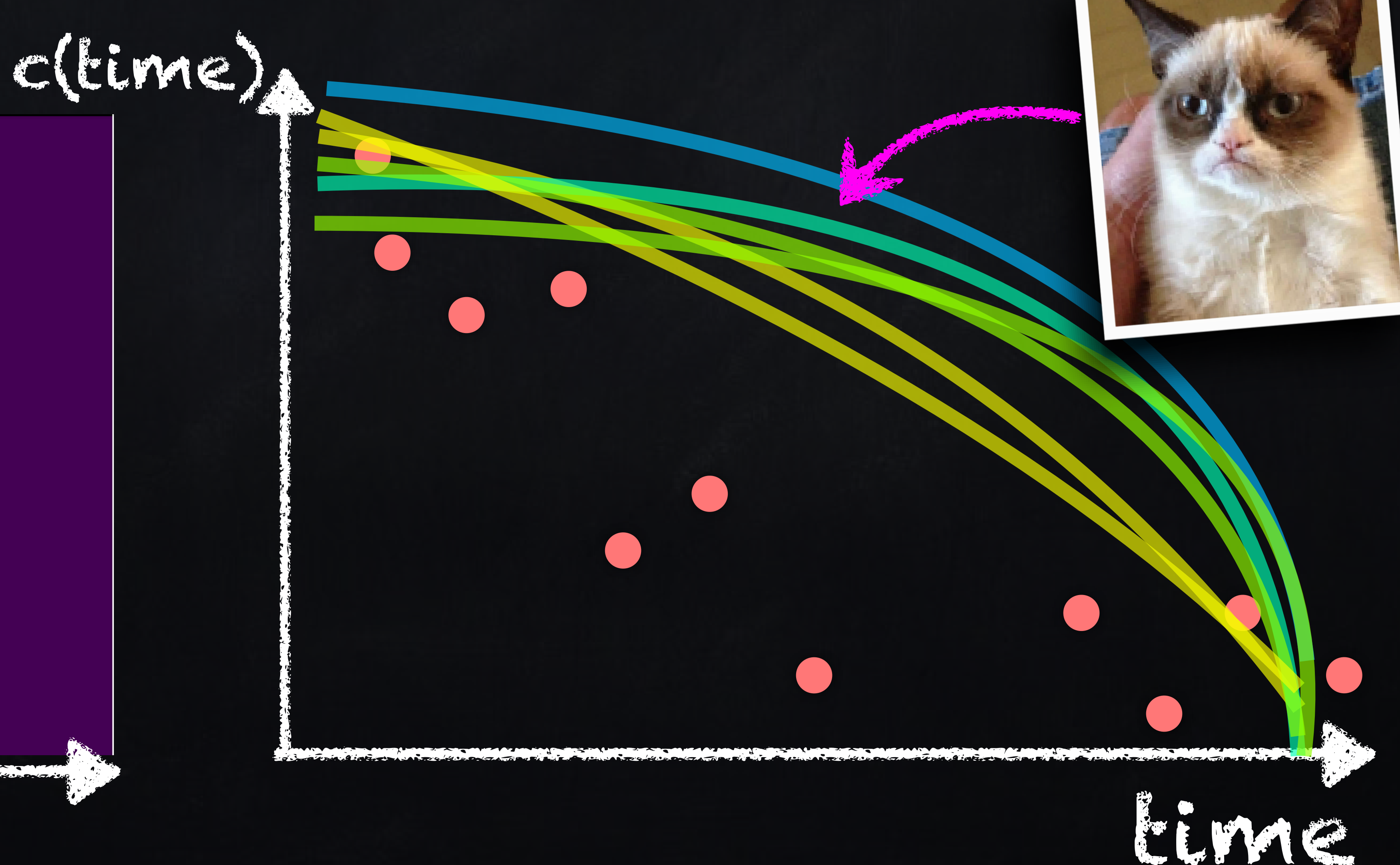
What am I supposed to do with the posterior? ✓

First, we need to check that the results make sense

Posterior predictive checks

Correct priors?
Correct likelihood/model?



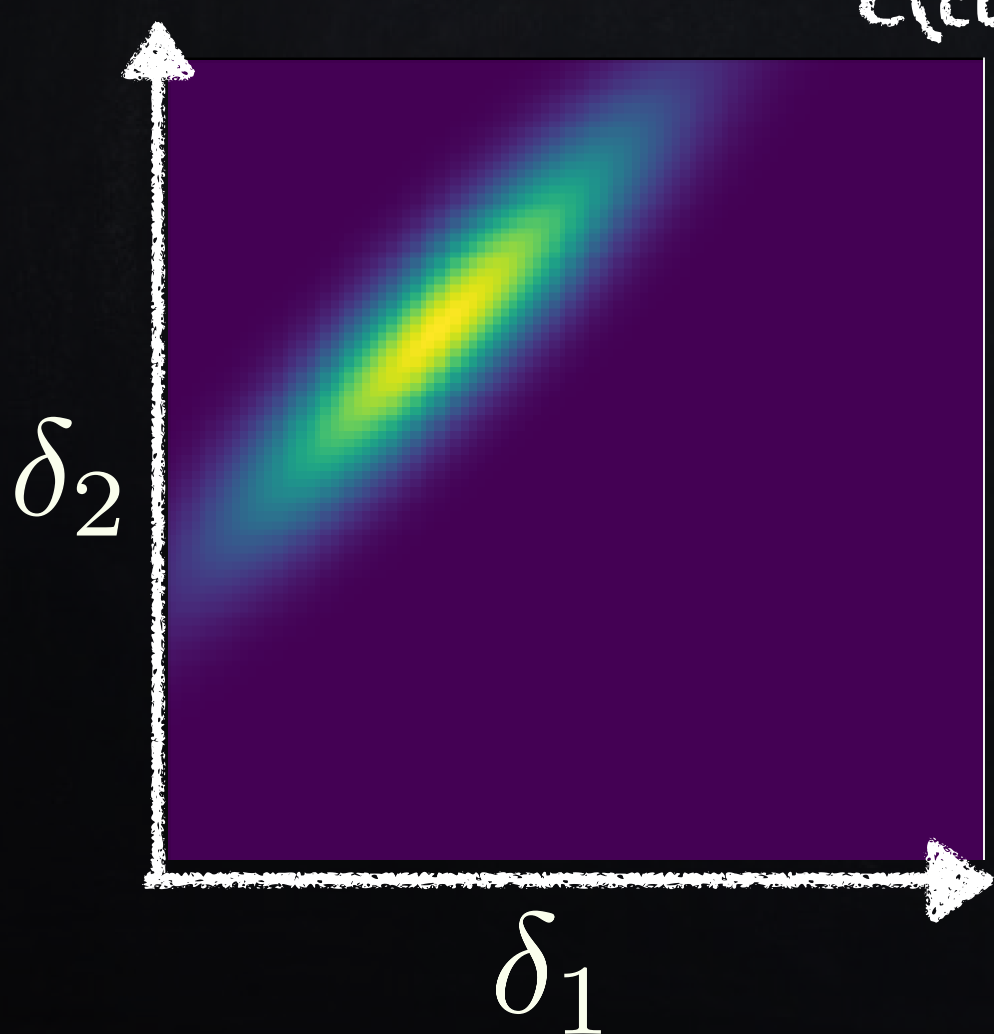


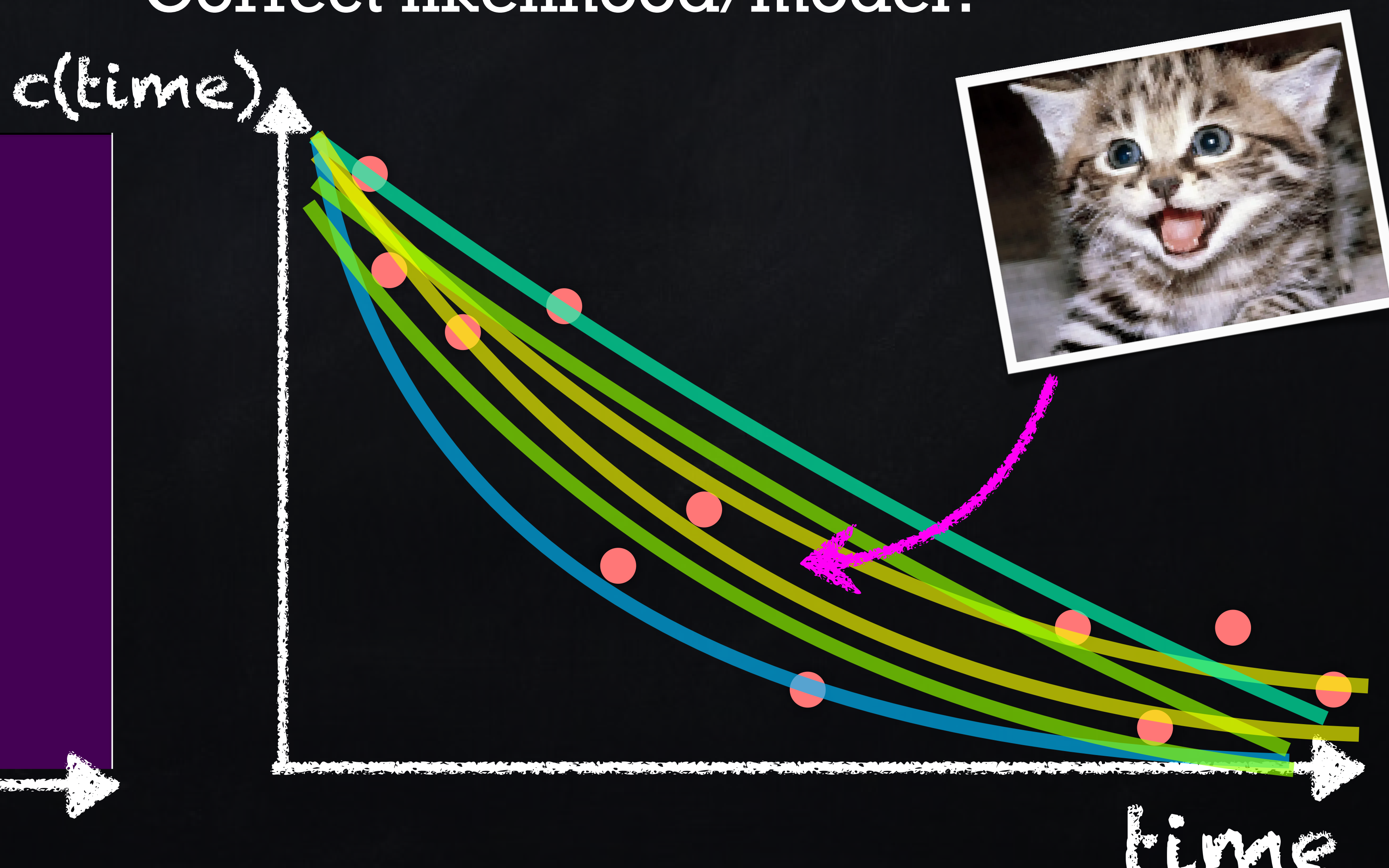
What am I supposed to do with the posterior?

First, we need to check that the results make sense

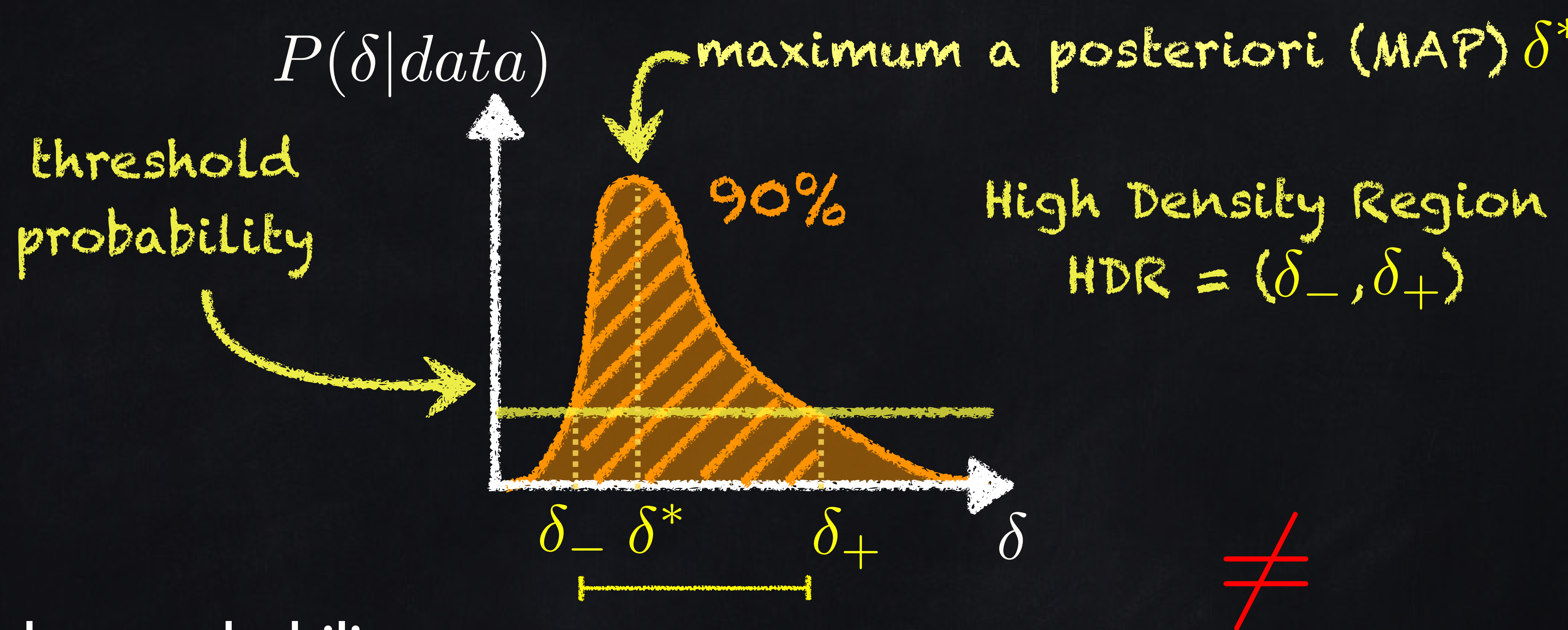
Posterior predictive checks

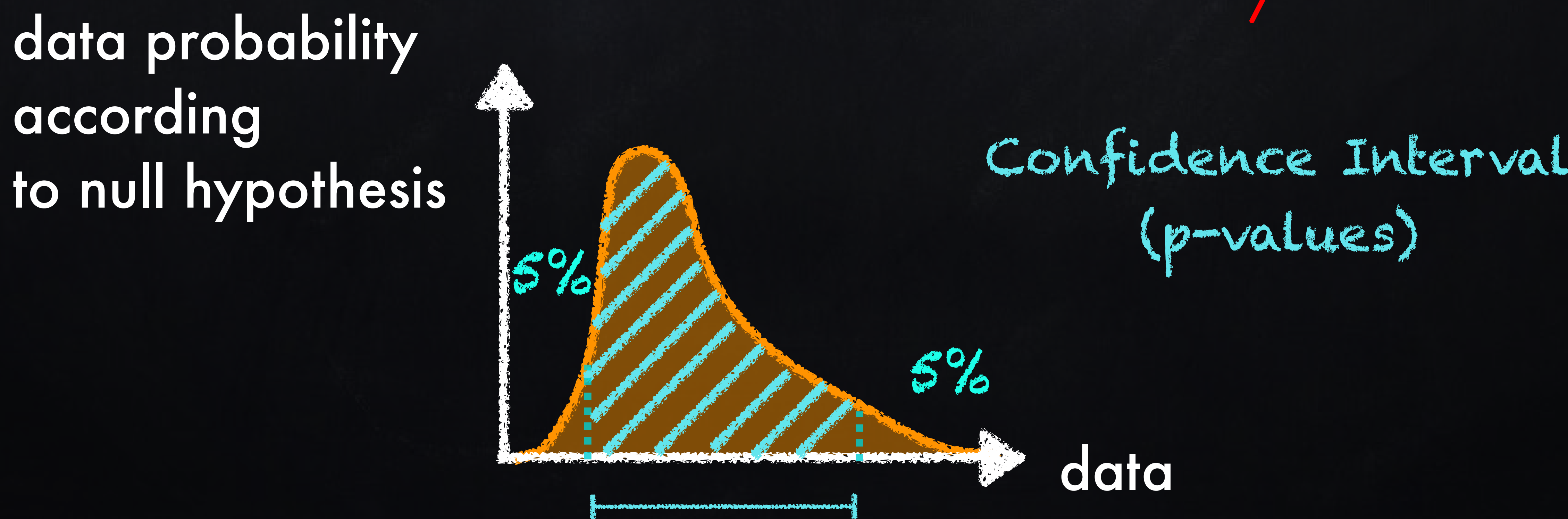
Correct priors?
Correct likelihood/model?



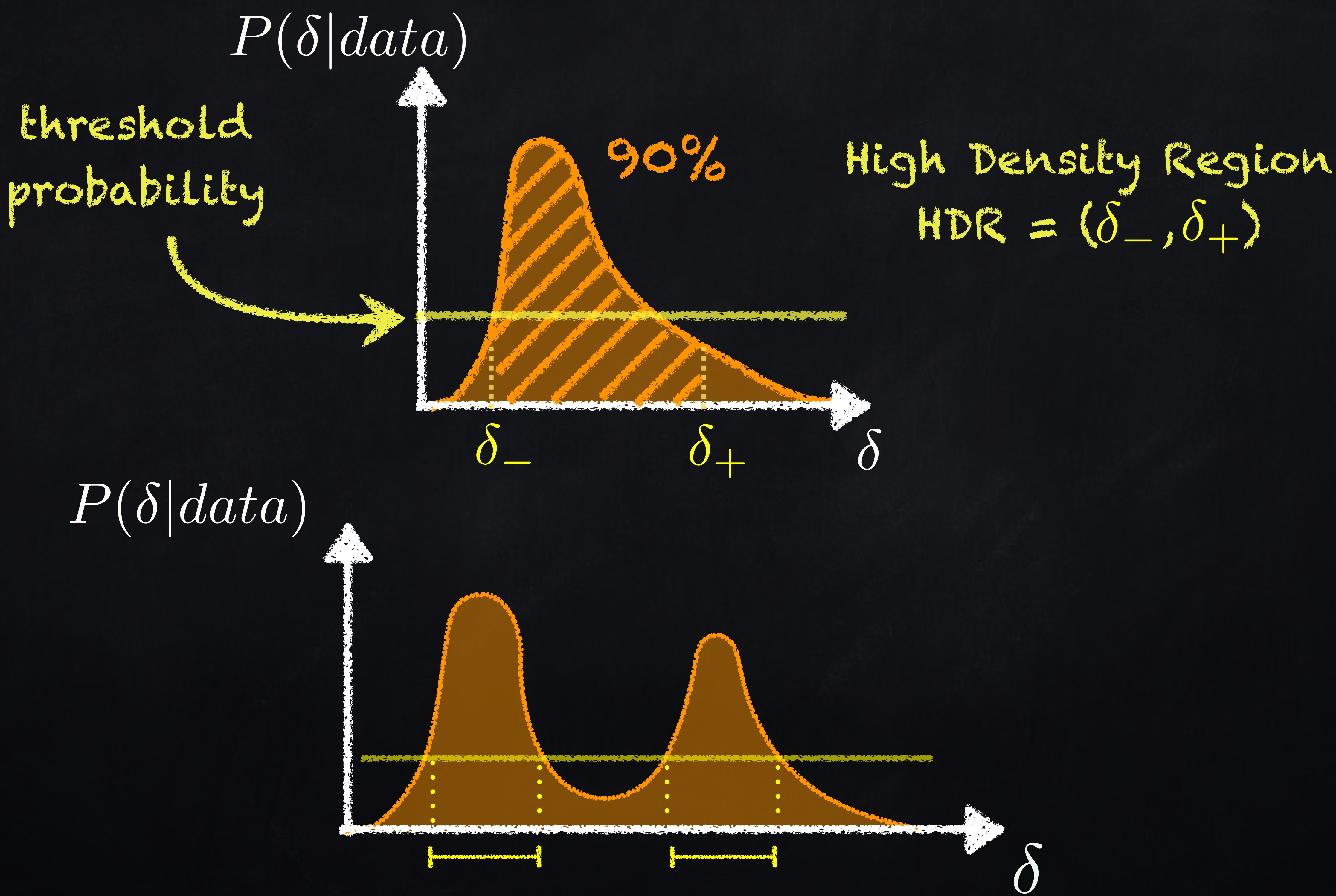


What am I supposed to do with the posterior?



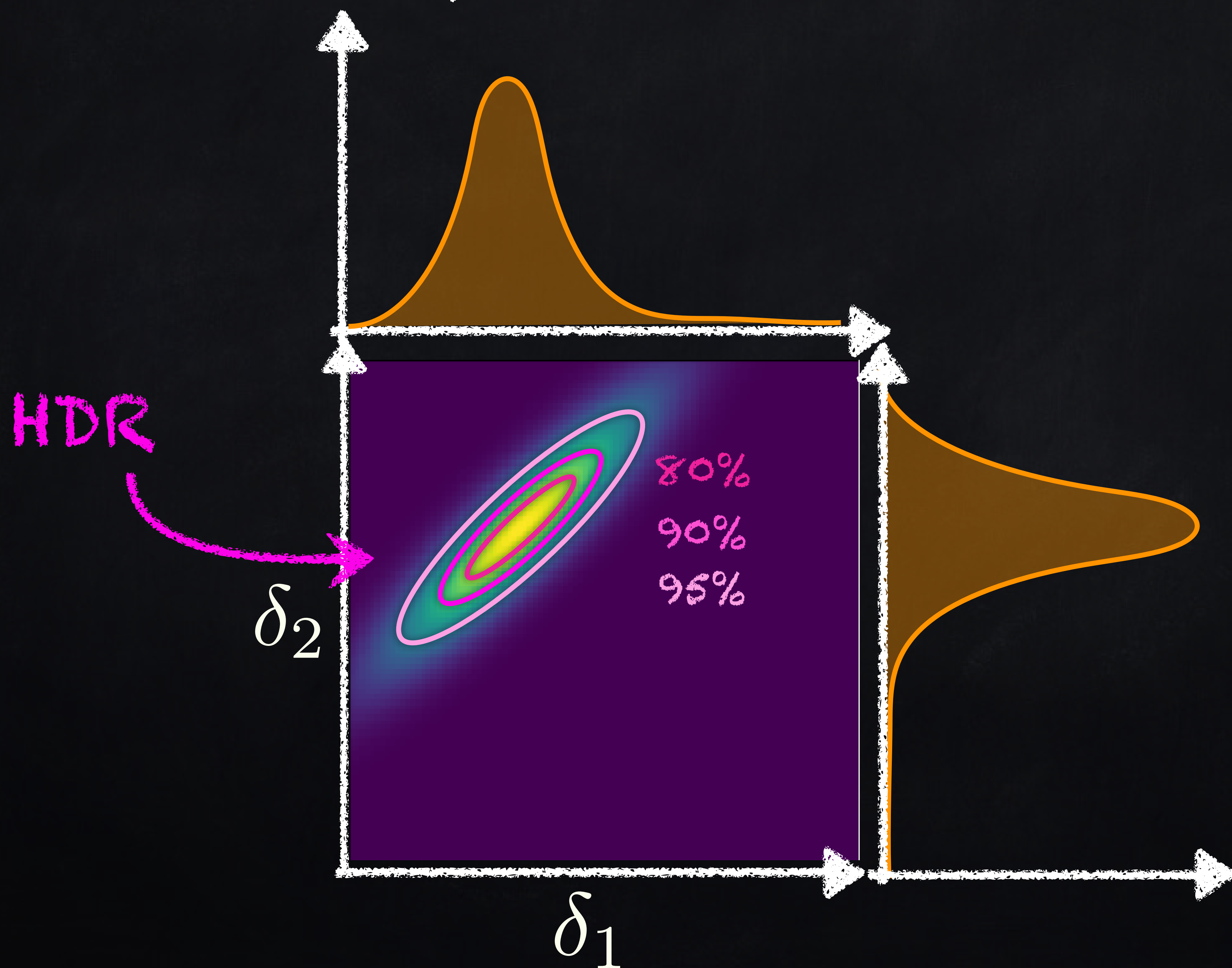


What am I supposed to do with the posterior?



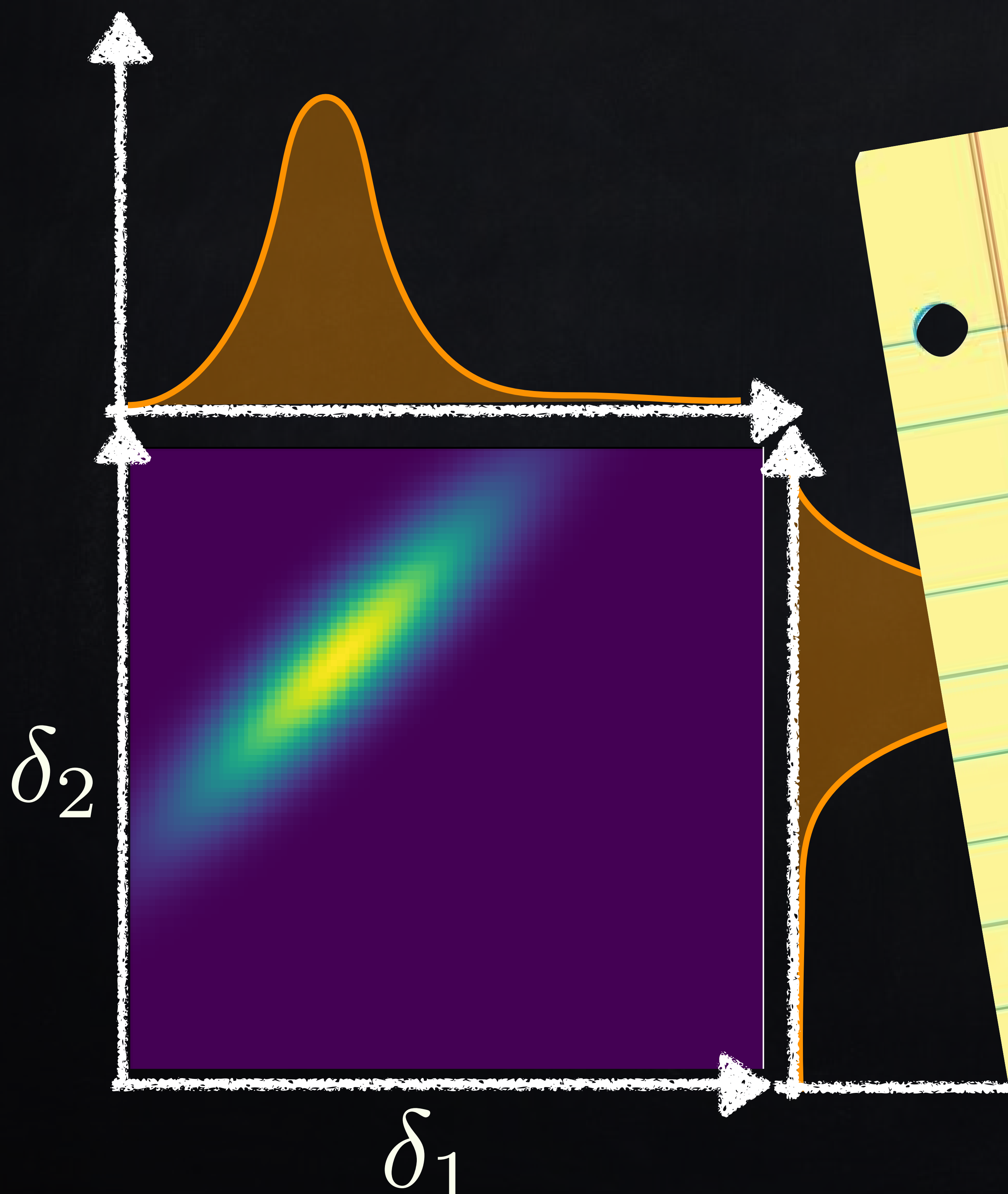
What am I supposed to do with the posterior? ✓

Inference of 2 parameters → $P(\delta_1, \delta_2 | data)$



What am I supposed to do with the posterior? ✔

Inference of 2 parameters ⟶ $P(\delta_1, \delta_2 | data)$



Question:

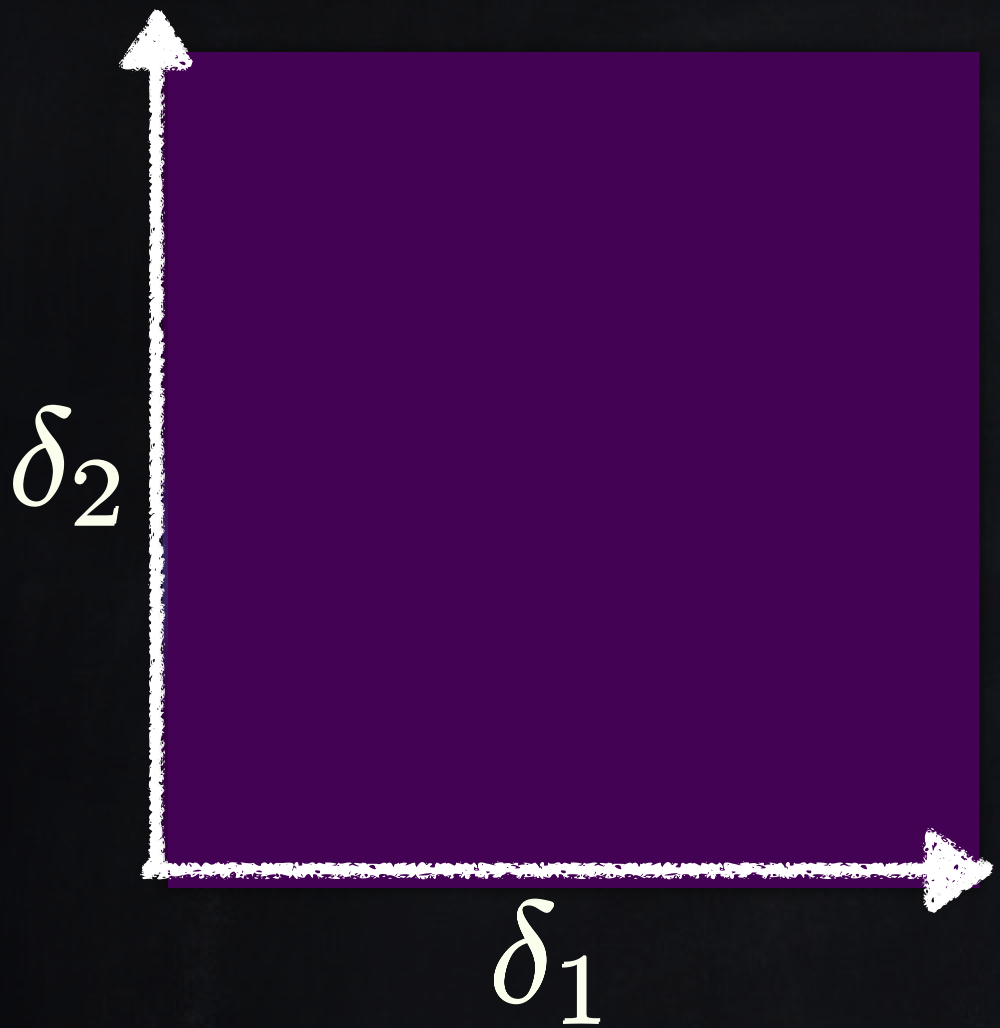
Hypothesis testing: How would you analyze the probabilities of the following hypotheses?

$\delta_2 > A, A \in \mathbb{R}$

$\delta_2 > \delta_1$

$\delta_2 = \delta_1$

How the algorithms behind them work?

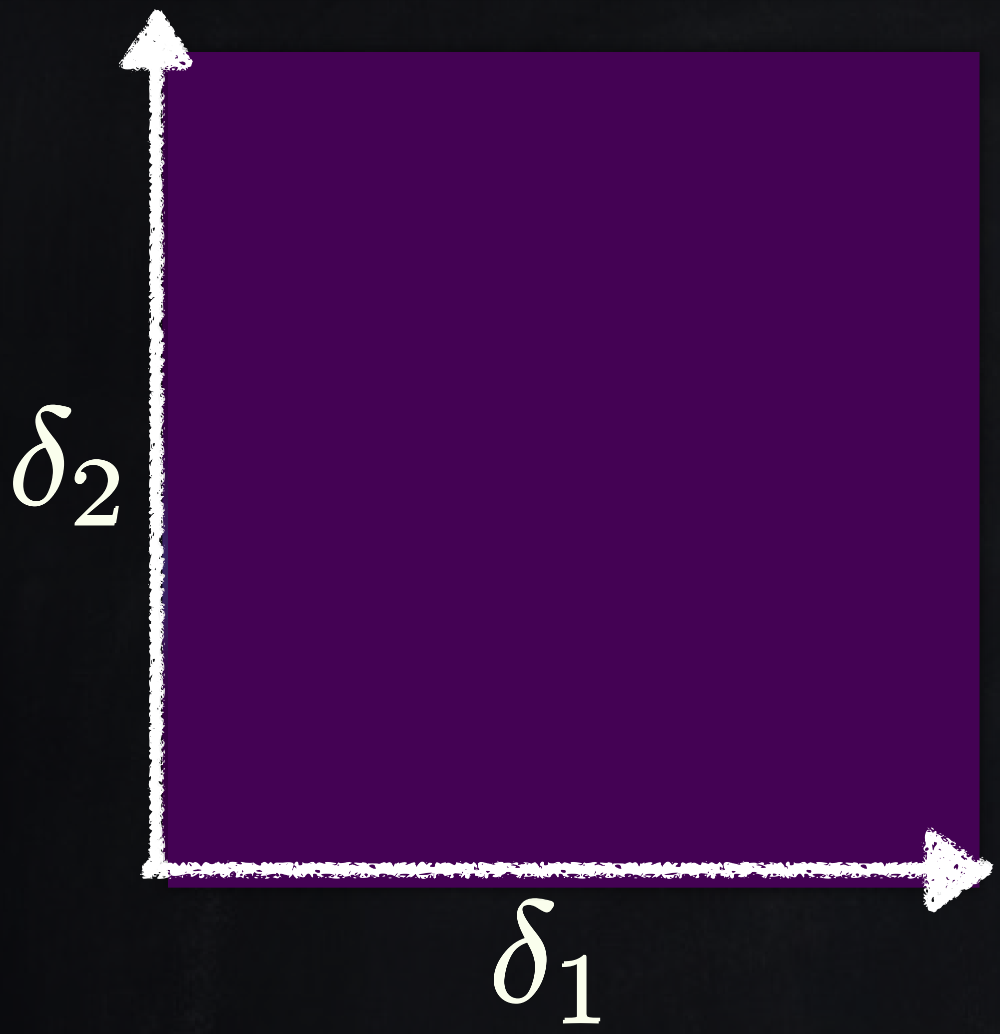


$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

Random sampling !

Markov Chain Monte Carlo (MCMC)

How the algorithms behind them work? ✓



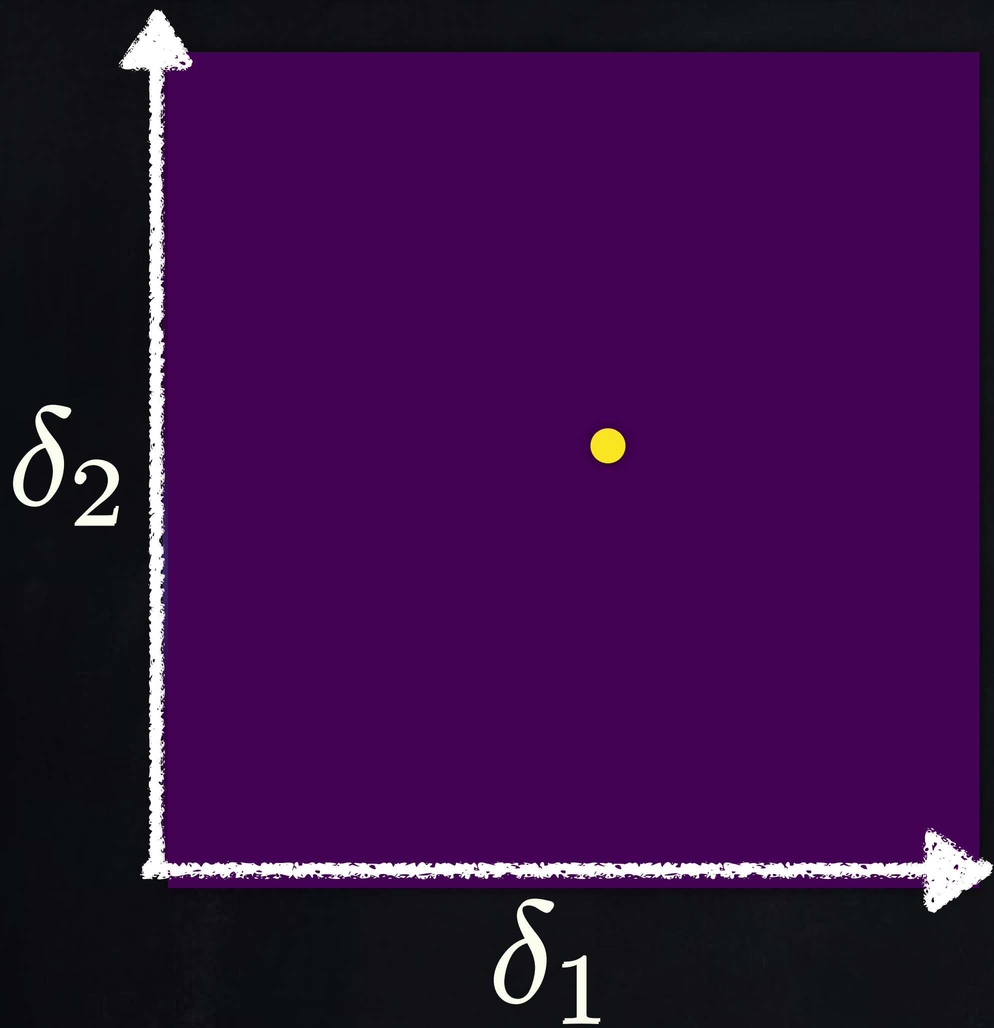
$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

Random sampling !

Markov Chain Monte Carlo (MCMC)

1. Set counter for accepted points $n = 0$

How the algorithms behind them work? ✓



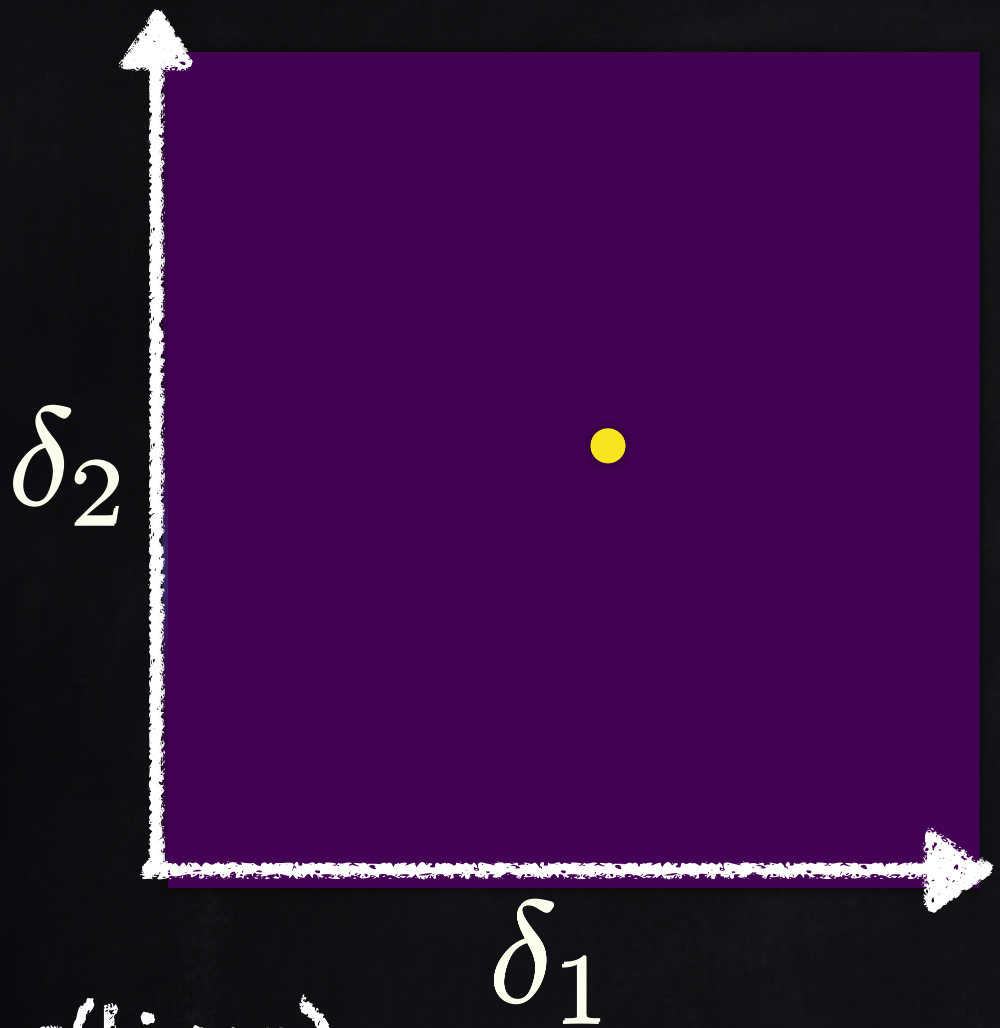
$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

Random sampling !

Markov Chain Monte Carlo (MCMC)

1. Set counter for accepted points $n = 0$
2. Select initial parameter point $\delta^{(0)} = (\delta_1^{(0)}, \delta_2^{(0)})$ from prior $P(\delta)$

How the algorithms behind them work?

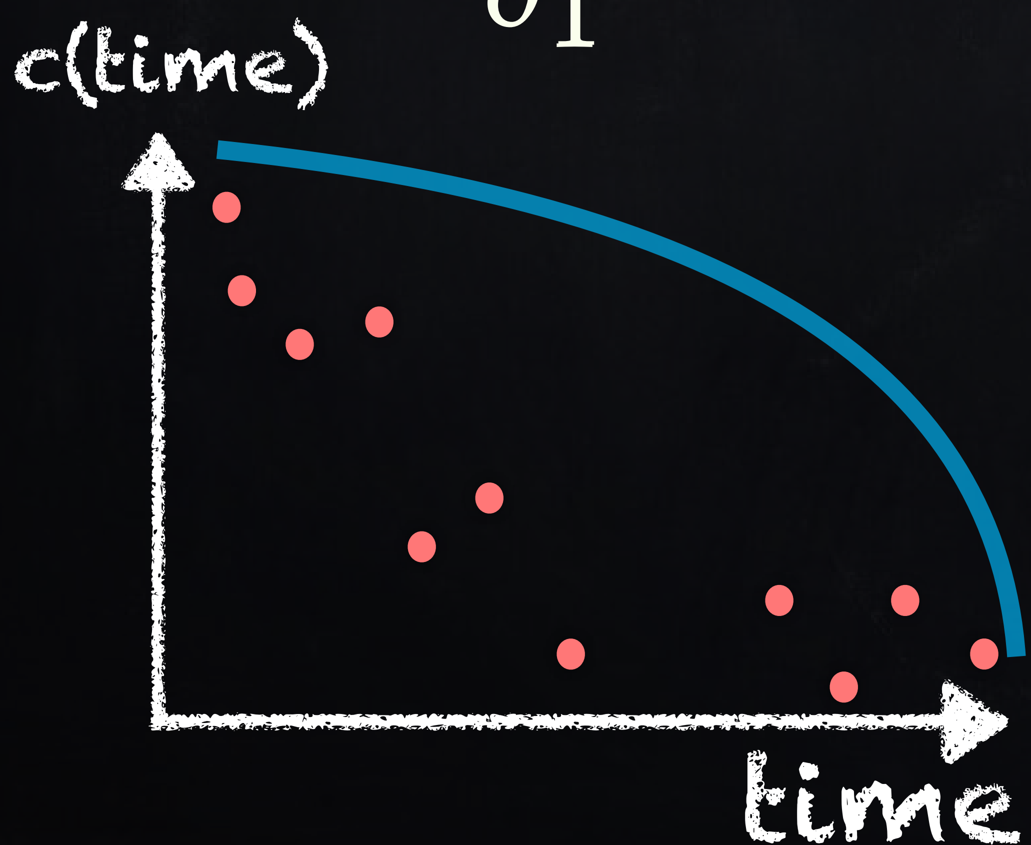


$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

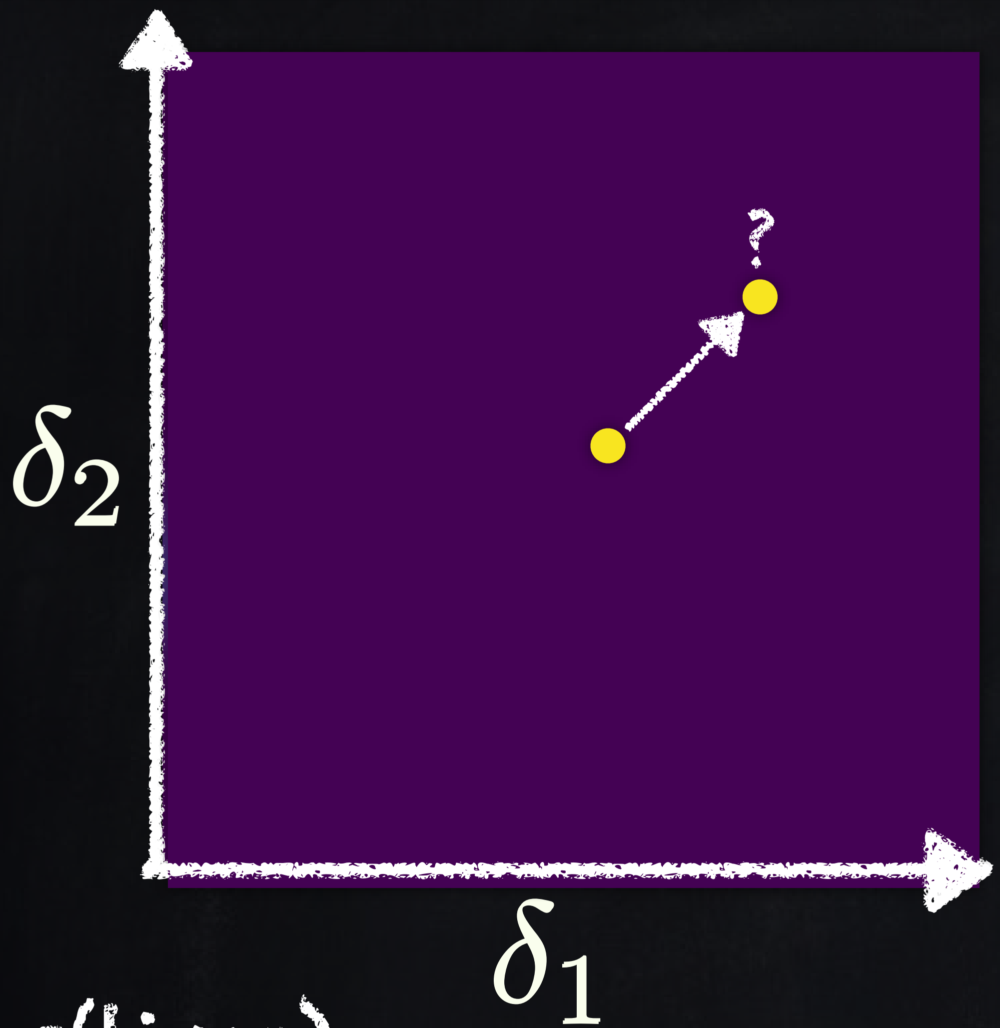
Random sampling !

Markov Chain Monte Carlo (MCMC)

1. Set counter for accepted points $n = 0$
2. Select initial parameter point $\delta^{(0)} = (\delta_1^{(0)}, \delta_2^{(0)})$ from prior $P(\delta)$



How the algorithms behind them work? ✓

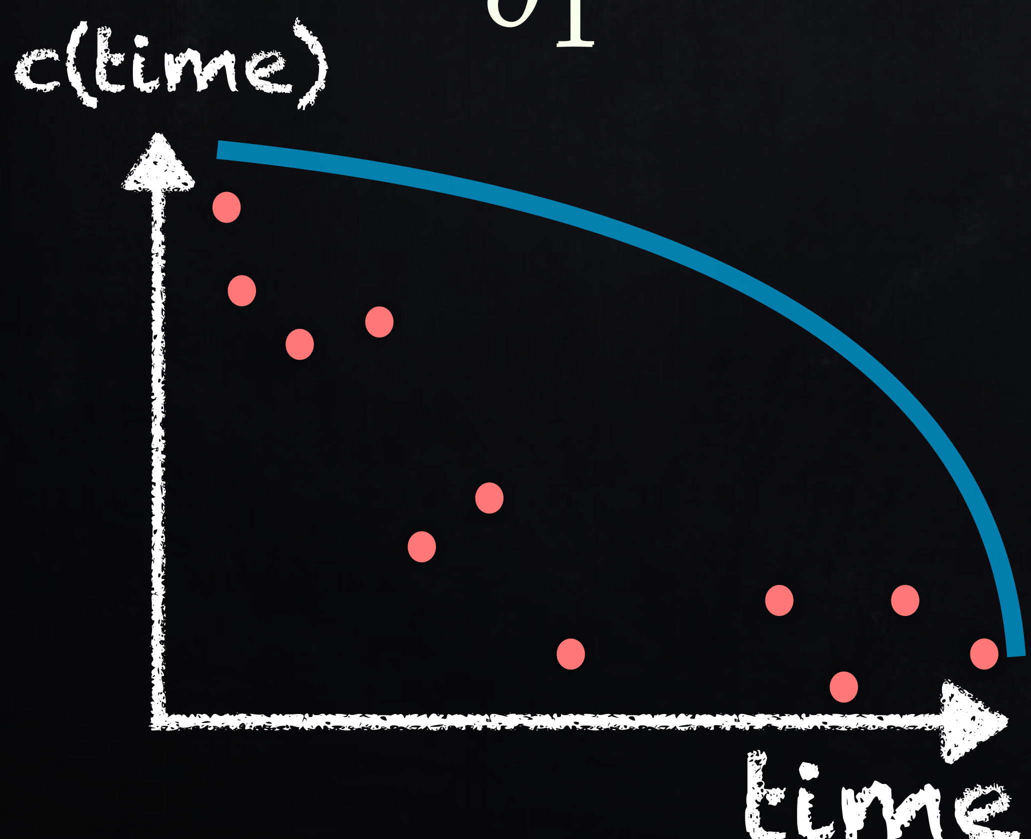


$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

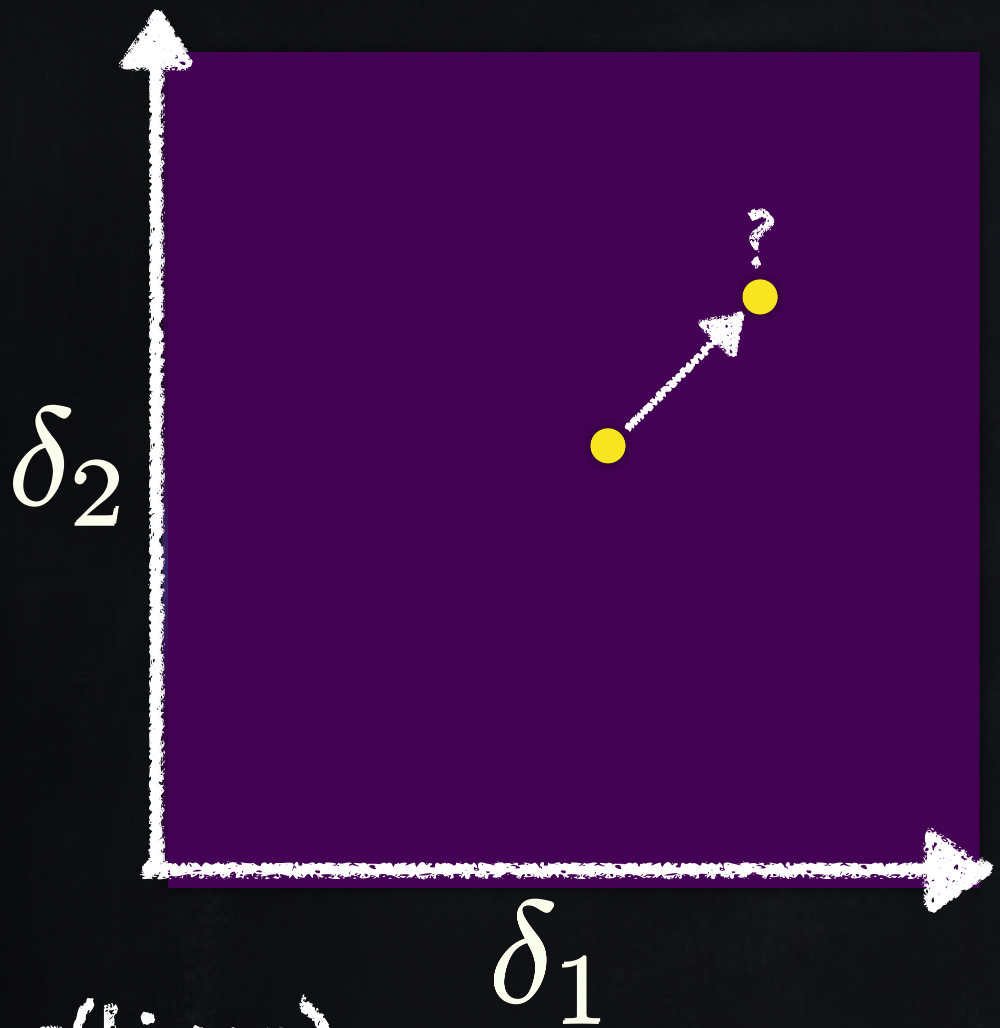
Random sampling !

Markov Chain Monte Carlo (MCMC)

1. Set counter for accepted points $n = 0$
2. Select initial parameter point $\delta^{(0)} = (\delta_1^{(0)}, \delta_2^{(0)})$ from prior $P(\delta)$
3. Propose new parameter $\delta^{(n+1)}$ using symmetric kernel $q(\delta^{(n+1)}|\delta^{(n)})$ e.g. Gaussian



How the algorithms behind them work? ✓

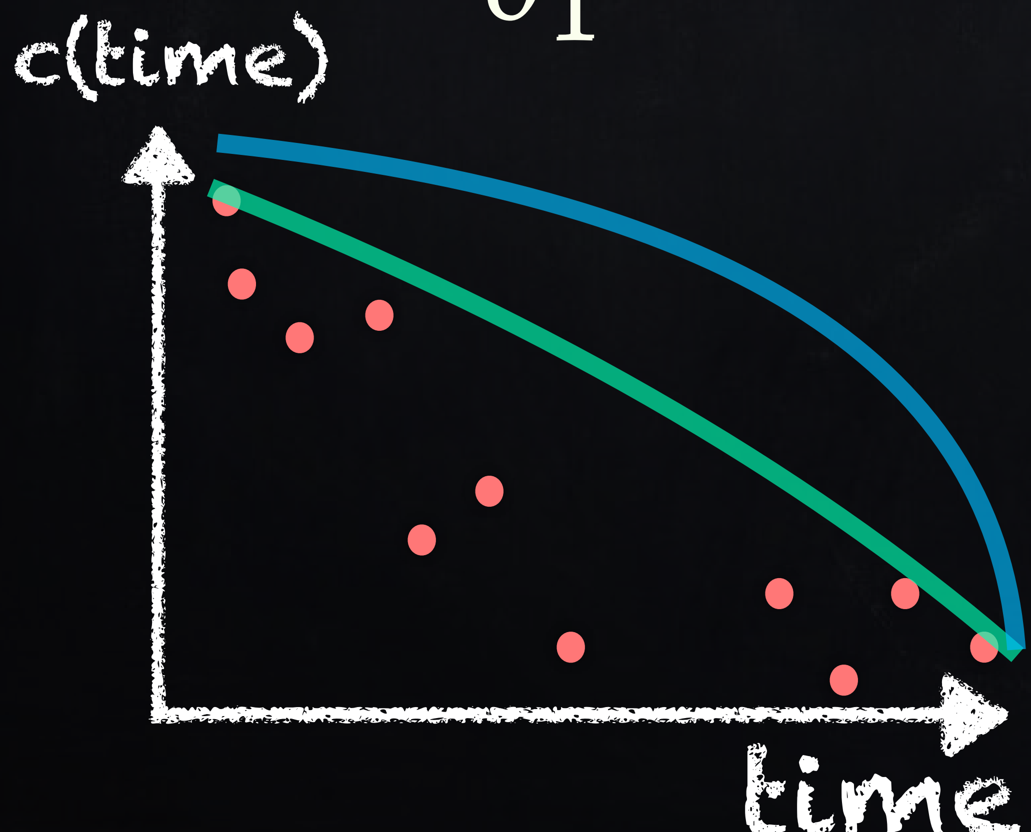


$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

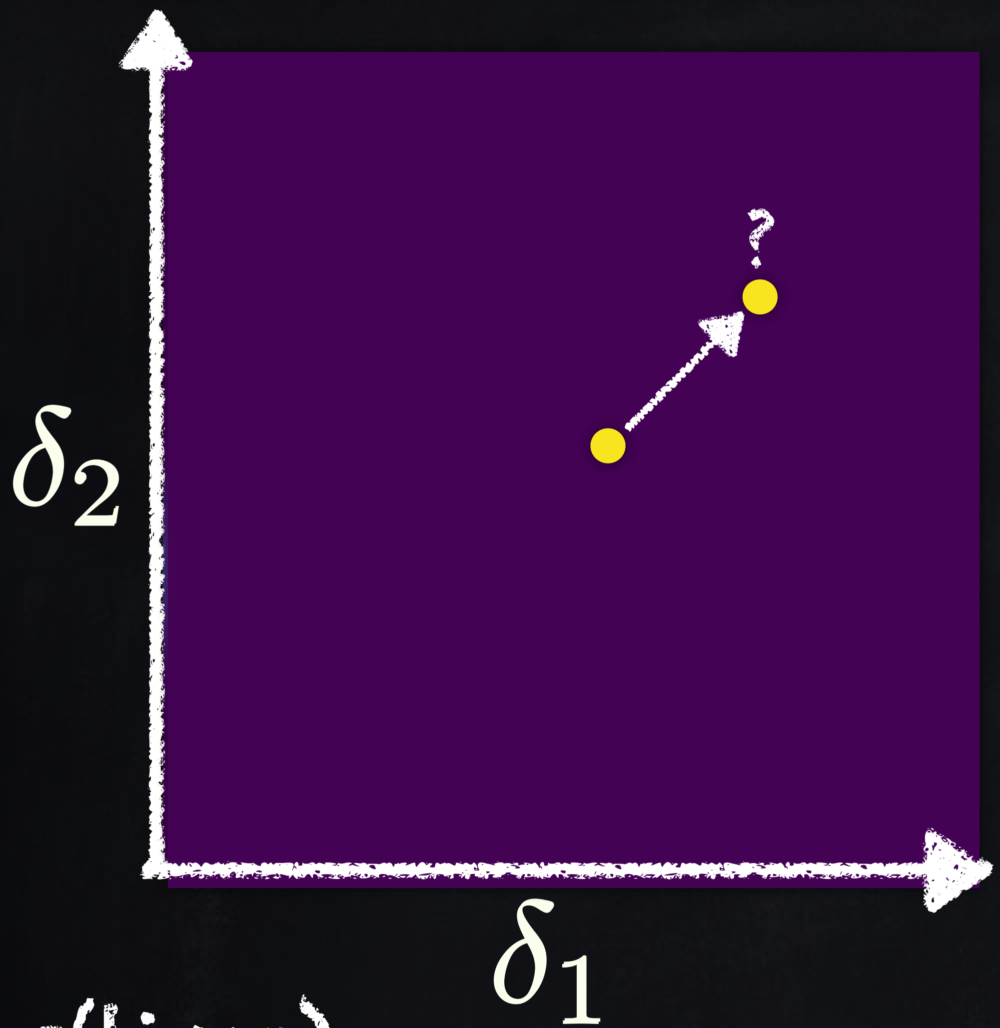
Random sampling !

Markov Chain Monte Carlo (MCMC)

1. Set counter for accepted points $n = 0$
2. Select initial parameter point $\delta^{(0)} = (\delta_1^{(0)}, \delta_2^{(0)})$ from prior $P(\delta)$
3. Propose new parameter $\delta^{(n+1)}$ using symmetric kernel $q(\delta^{(n+1)}|\delta^{(n)})$ e.g. Gaussian



How the algorithms behind them work?



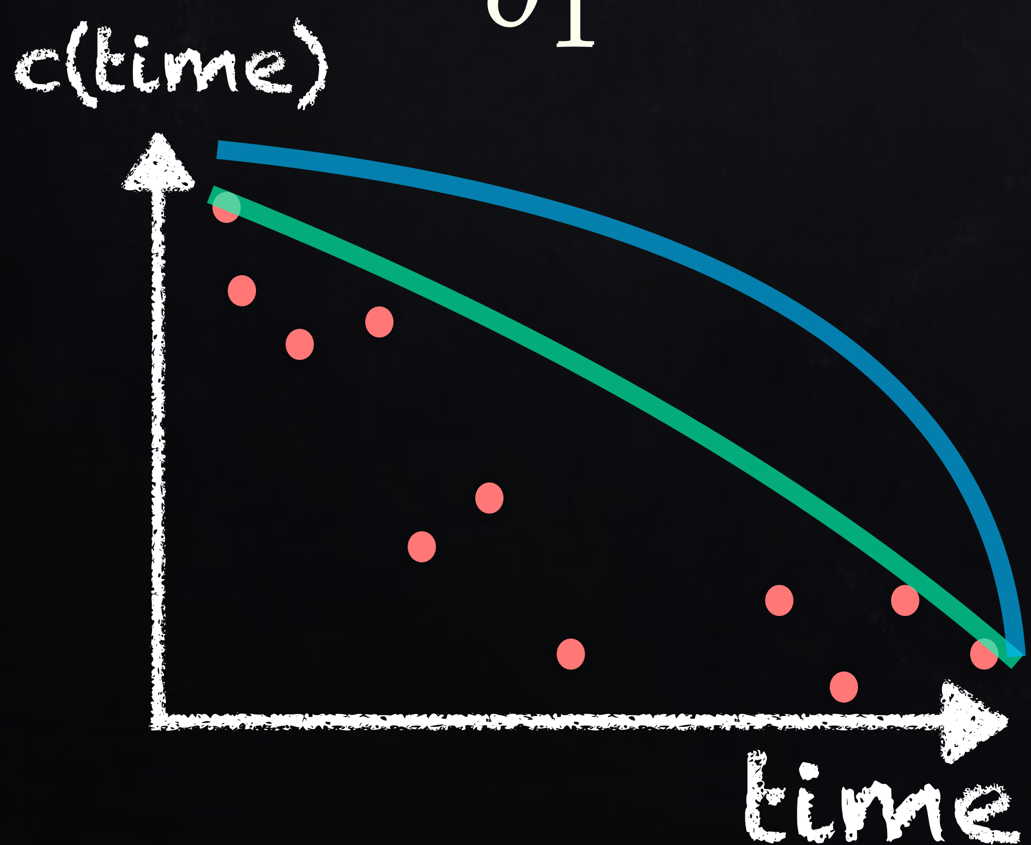
$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

Random sampling !

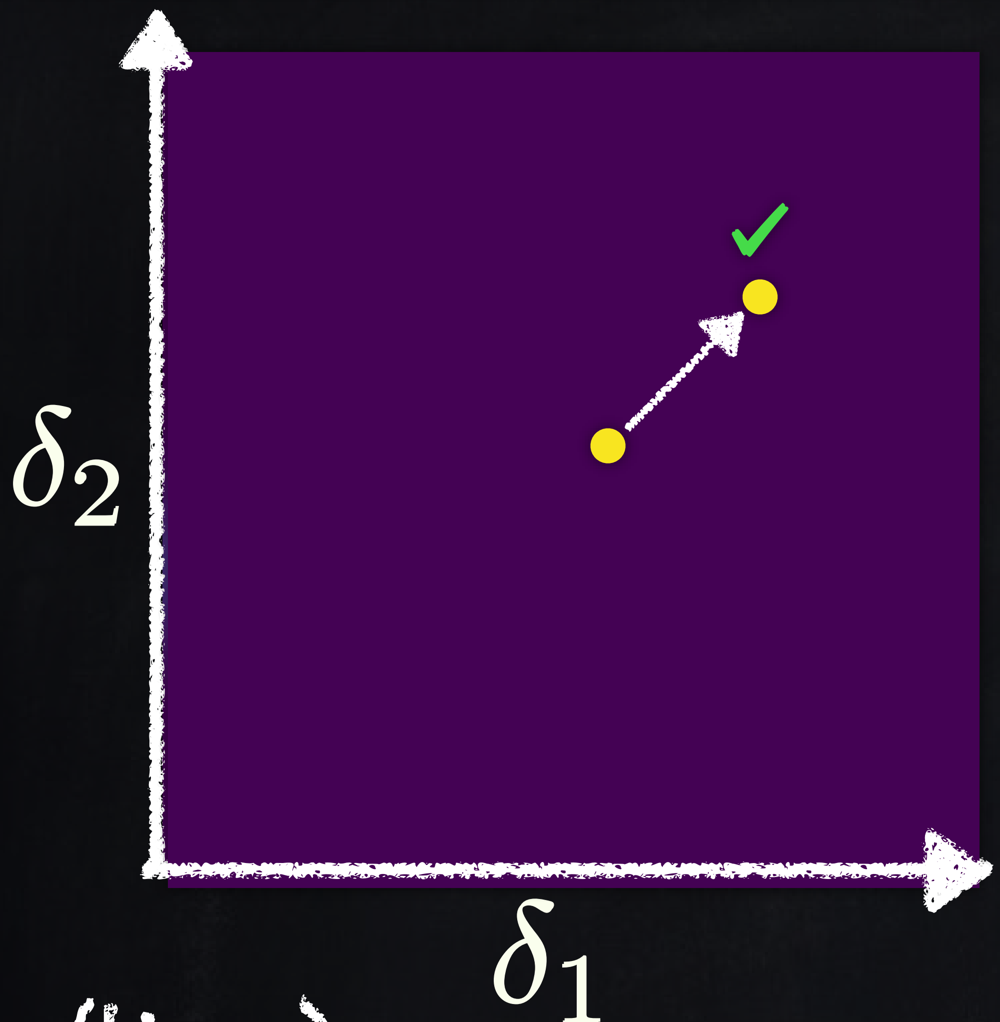
Markov Chain Monte Carlo (MCMC)

1. Set counter for accepted points $n = 0$
2. Select initial parameter point $\delta^{(0)} = (\delta_1^{(0)}, \delta_2^{(0)})$ from prior $P(\delta)$
3. Propose new parameter $\delta^{(n+1)}$ using symmetric kernel $q(\delta^{(n+1)}|\delta^{(n)})$ e.g. Gaussian
4. Evaluate posterior ratio

$$r = \frac{P(\delta^{(n+1)}|data)}{P(\delta^{(n)}|data)} = \frac{P(data|\delta^{(n+1)})P(\delta^{(n+1)})}{P(data|\delta^{(n)})P(\delta^{(n)})}$$



How the algorithms behind them work?



$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

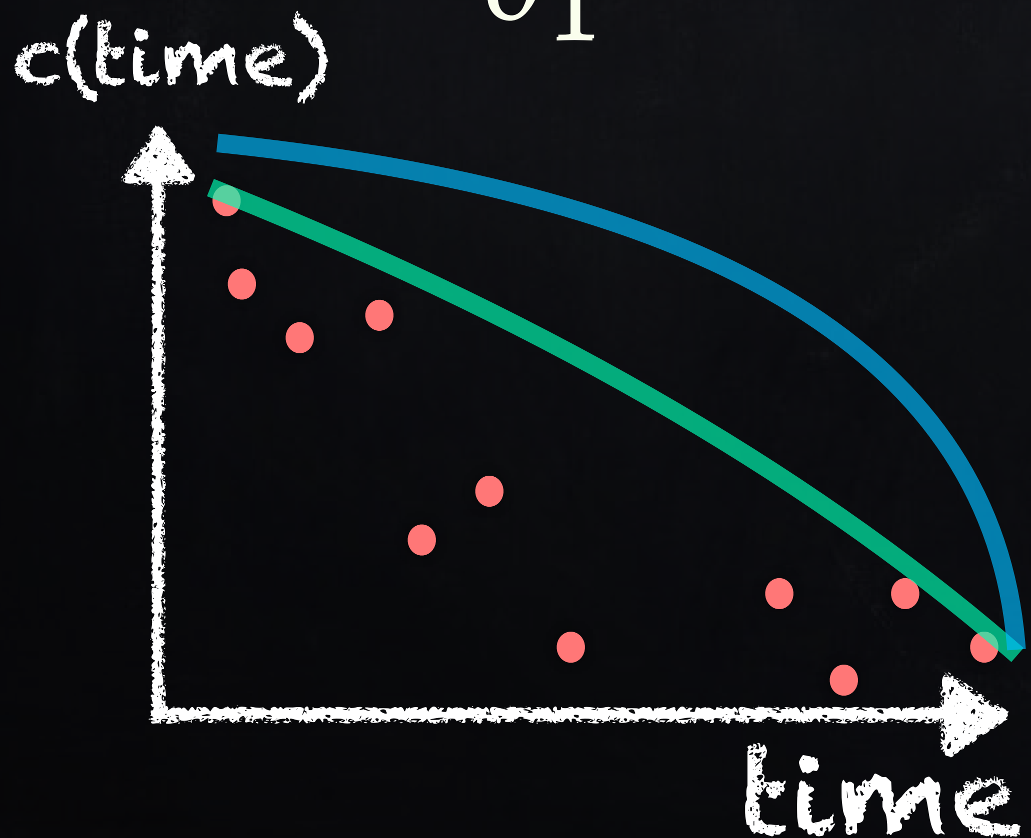
Random sampling !

Markov Chain Monte Carlo (MCMC)

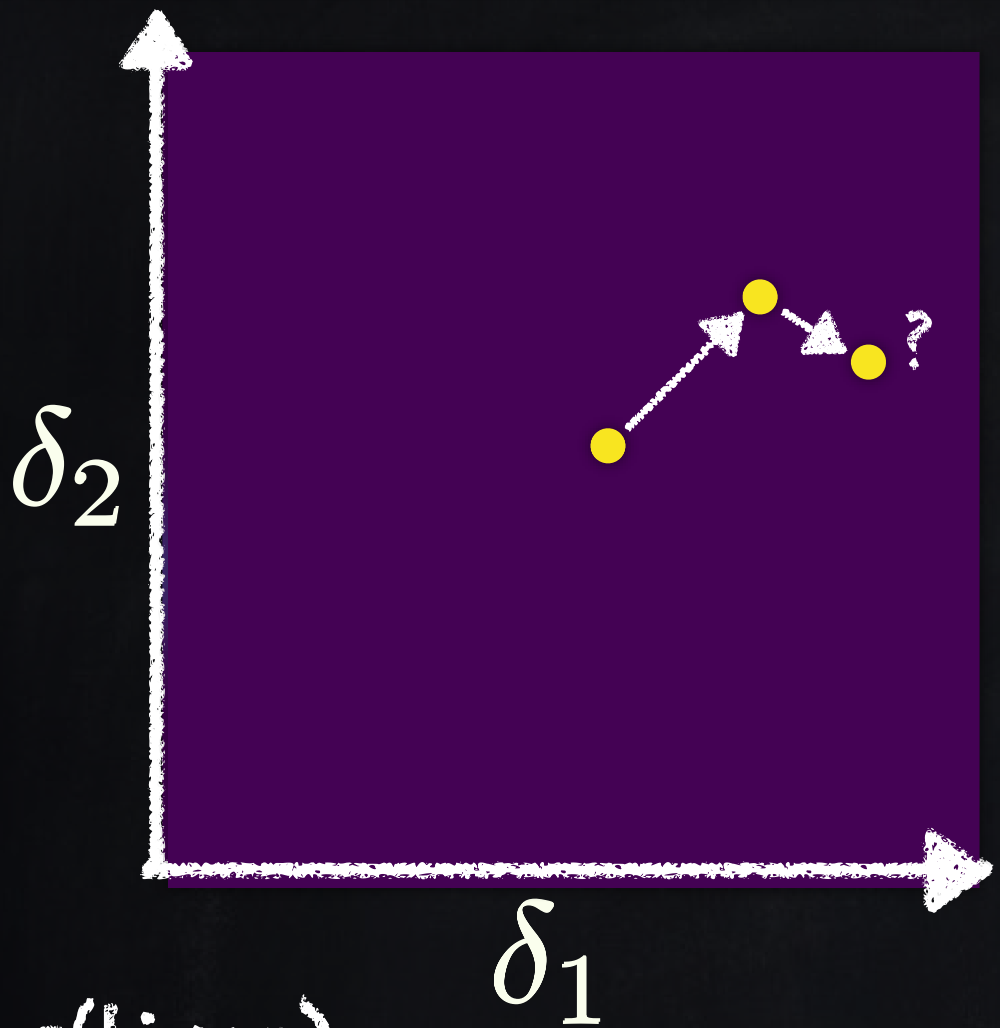
1. Set counter for accepted points $n = 0$
2. Select initial parameter point $\delta^{(0)} = (\delta_1^{(0)}, \delta_2^{(0)})$ from prior $P(\delta)$
3. Propose new parameter $\delta^{(n+1)}$ using symmetric kernel $q(\delta^{(n+1)}|\delta^{(n)})$ e.g. Gaussian
4. Evaluate posterior ratio

$$r = \frac{P(\delta^{(n+1)}|data)}{P(\delta^{(n)}|data)} = \frac{P(data|\delta^{(n+1)})P(\delta^{(n+1)})}{P(data|\delta^{(n)})P(\delta^{(n)})}$$

5. Accept point with probability min(1,r)



How the algorithms behind them work? ✓



$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

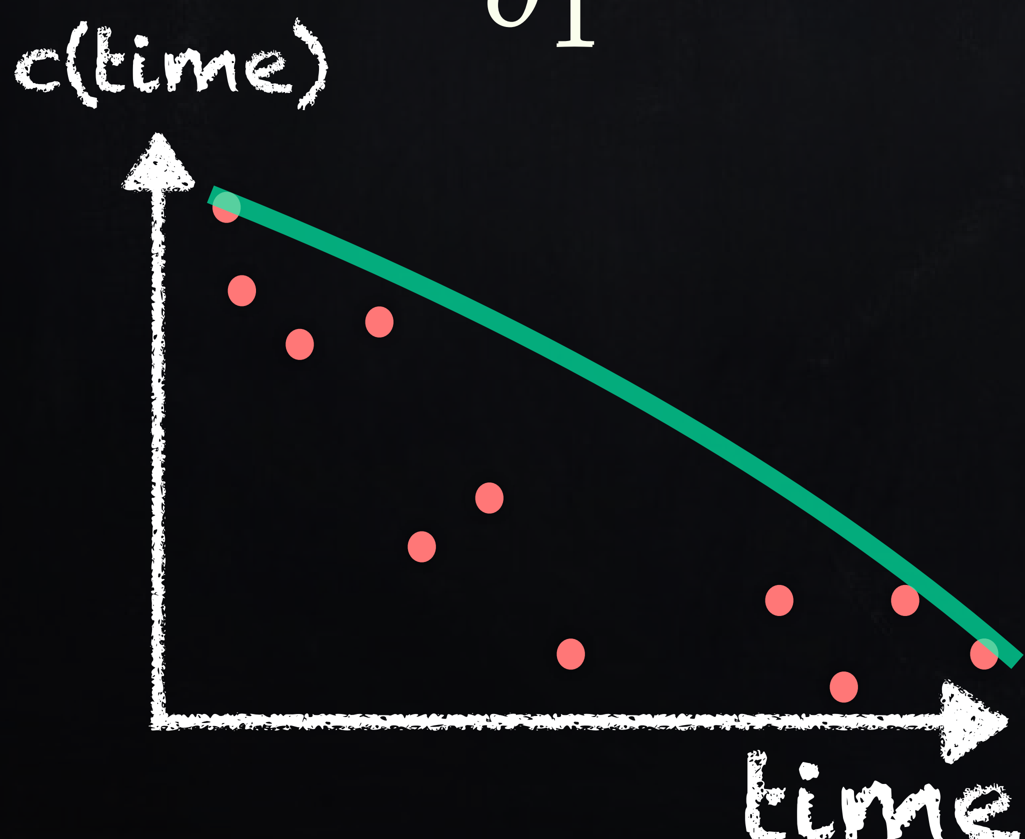
Random sampling !

Markov Chain Monte Carlo (MCMC)

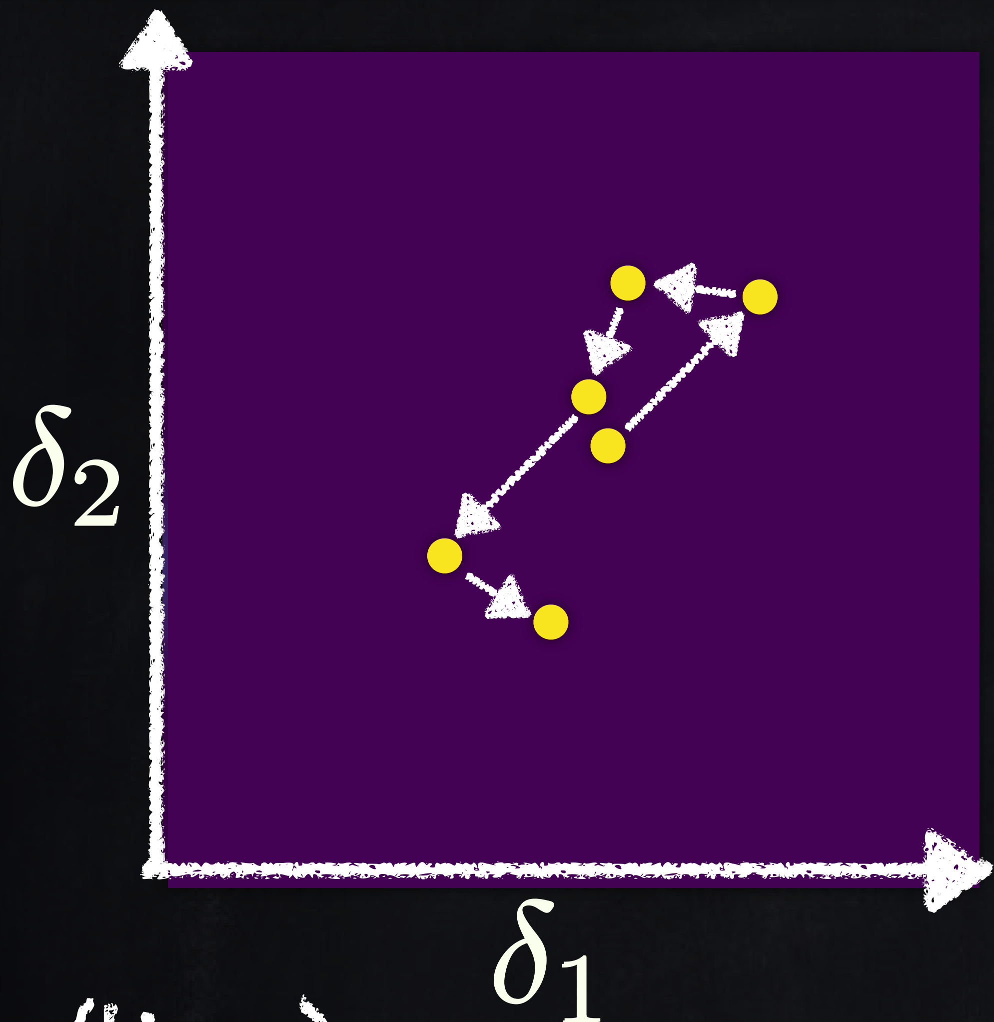
1. Set counter for accepted points $n = 0$
2. Select initial parameter point $\delta^{(0)} = (\delta_1^{(0)}, \delta_2^{(0)})$ from prior $P(\delta)$
3. Propose new parameter $\delta^{(n+1)}$ using symmetric kernel $q(\delta^{(n+1)}|\delta^{(n)})$ e.g. Gaussian
4. Evaluate posterior ratio

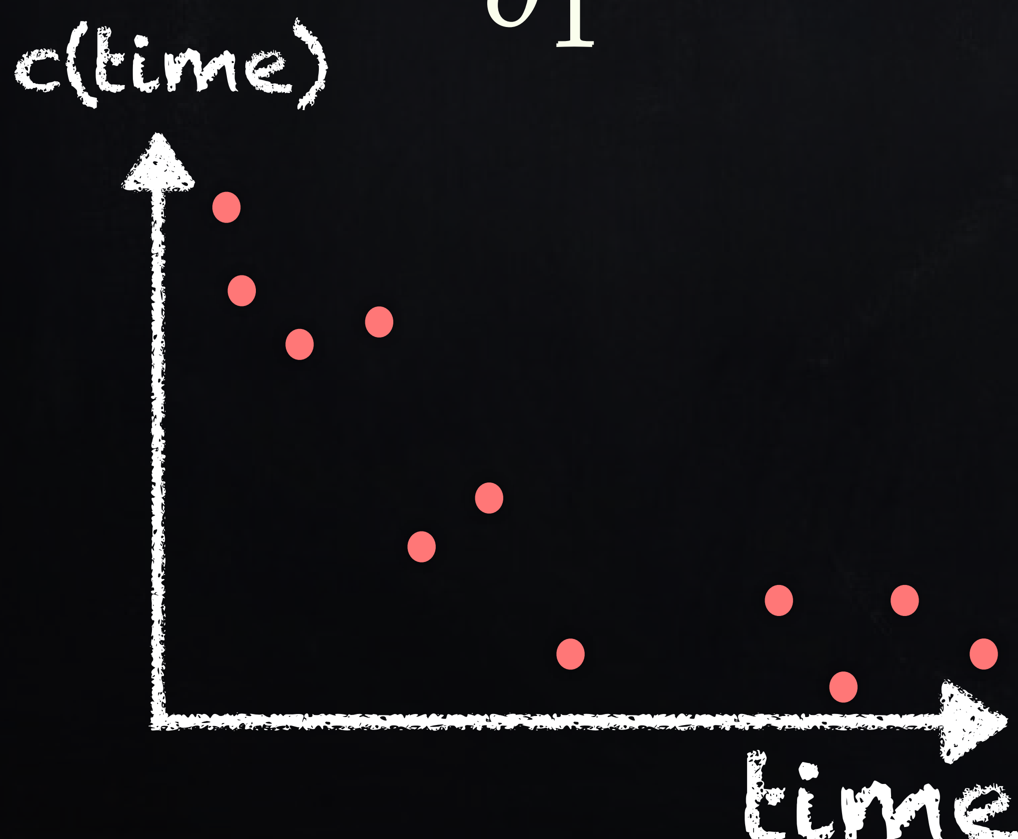
$$r = \frac{P(\delta^{(n+1)}|data)}{P(\delta^{(n)}|data)} = \frac{P(data|\delta^{(n+1)})P(\delta^{(n+1)})}{P(data|\delta^{(n)})P(\delta^{(n)})}$$

5. Accept point with probability min(1,r)
6. If successful acceptance: $n = n + 1$
7. Propose next point



How the algorithms behind them work?





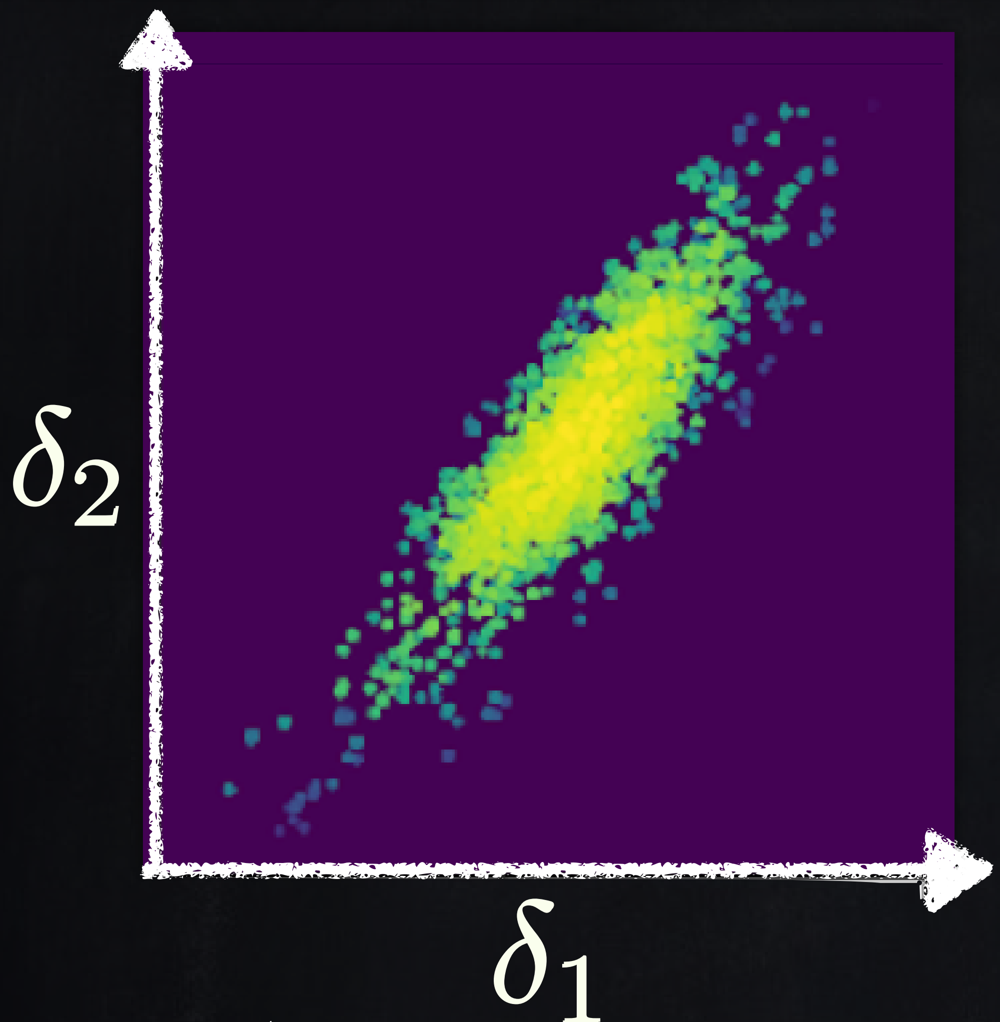
$$\frac{P(\delta)P(data|\delta)}{P(data)} = P(\delta|data)$$

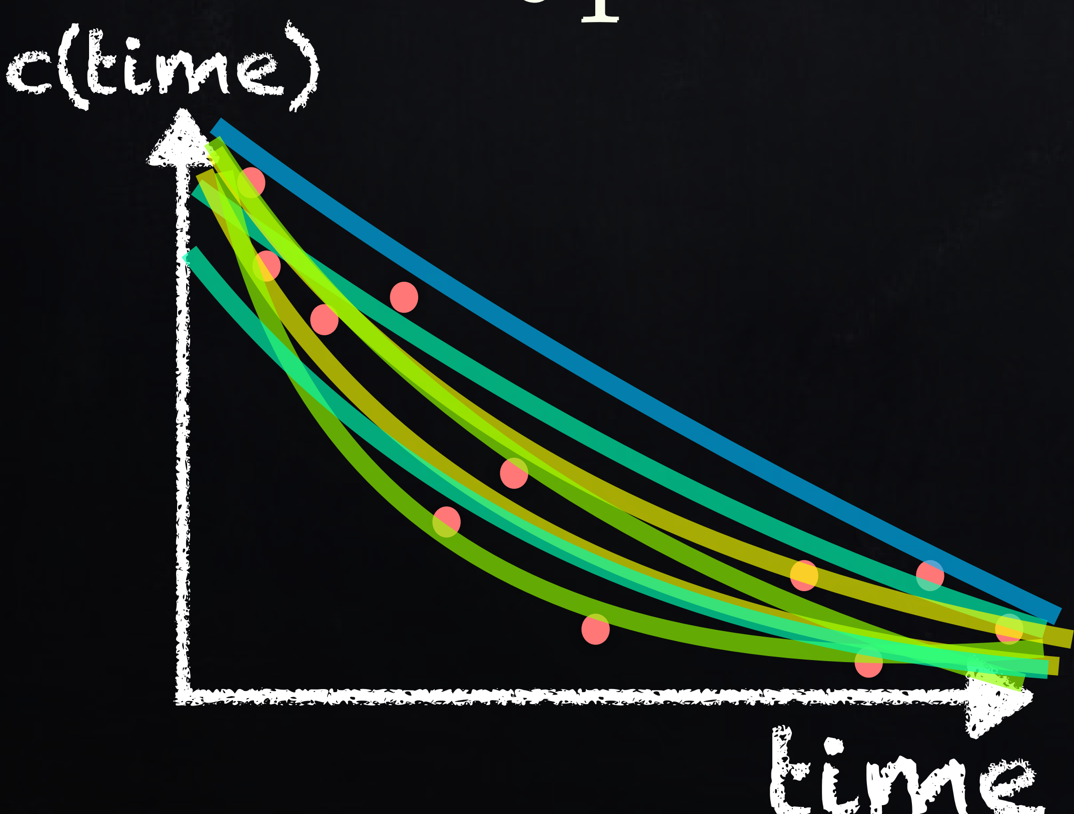
Random sampling !

Markov Chain Monte Carlo (MCMC)

1. Set counter for accepted points $n = 0$
2. Select initial parameter point $\delta^{(0)} = (\delta_1^{(0)}, \delta_2^{(0)})$ from prior $P(\delta)$
3. Propose new parameter $\delta^{(n+1)}$ using symmetric kernel $q(\delta^{(n+1)}|\delta^{(n)})$ e.g. Gaussian
4. Evaluate posterior ratio
$$r = \frac{P(\delta^{(n+1)}|data)}{P(\delta^{(n)}|data)} = \frac{P(data|\delta^{(n+1)})P(\delta^{(n+1)})}{P(data|\delta^{(n)})P(\delta^{(n)})}$$
5. Accept point with probability min(1,r)
6. If successful acceptance: $n = n + 1$
7. Propose next point

How the algorithms behind them work? ✓





Markov Chain Monte Carlo (MCMC)

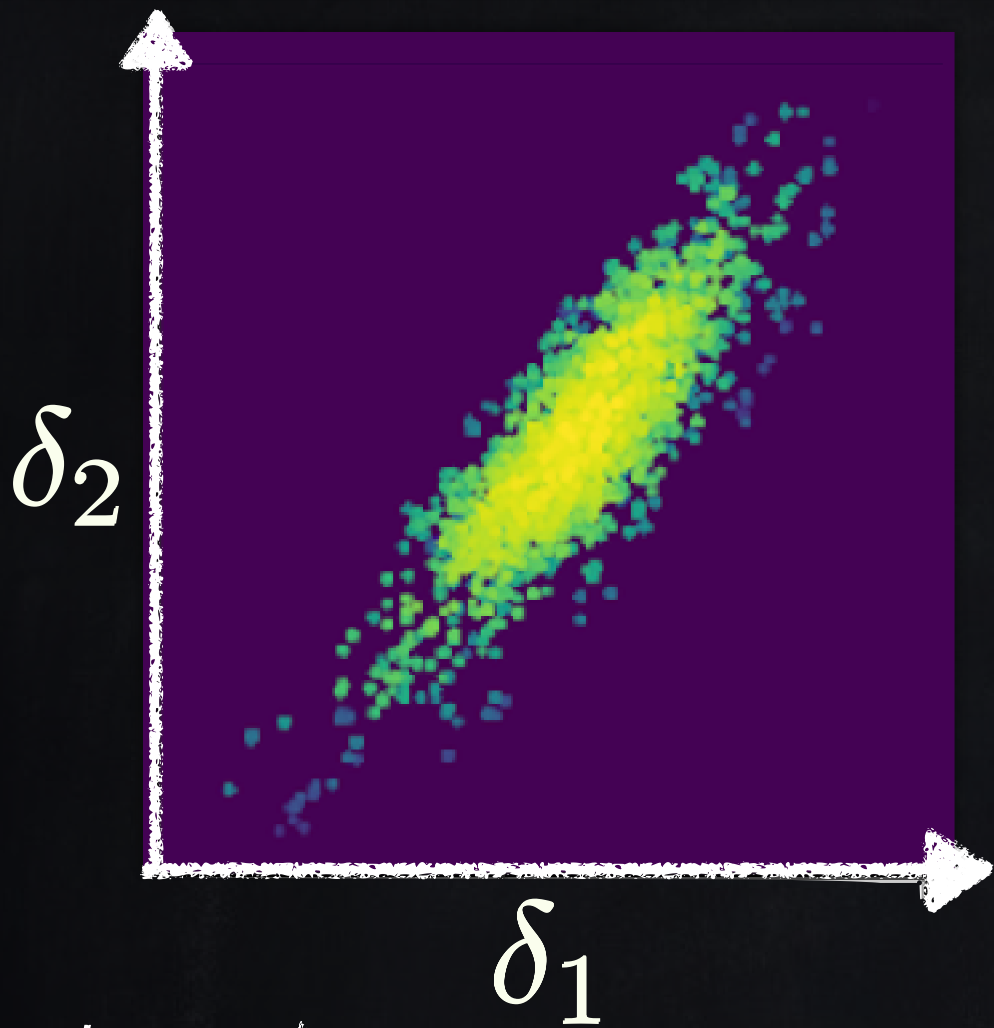
$$\delta^{(1)}, \delta^{(2)}, \delta^{(3)}, \delta^{(4)}, \ldots \sim P(\delta|data)$$

1. Set counter for accepted points $n = 0$
2. Select initial parameter point $\delta^{(0)} = (\delta_1^{(0)}, \delta_2^{(0)})$ from prior $P(\delta)$
3. Propose new parameter $\delta^{(n+1)}$ using symmetric kernel $q(\delta^{(n+1)}|\delta^{(n)})$ e.g. Gaussian
4. Evaluate posterior ratio

$$r = \frac{P(\delta^{(n+1)}|data)}{P(\delta^{(n)}|data)} = \frac{P(data|\delta^{(n+1)})P(\delta^{(n+1)})}{P(data|\delta^{(n)})P(\delta^{(n)})}$$

5. Accept point with probability min(1,r)
6. If successful acceptance: $n = n + 1$
7. Propose next point

How the algorithms behind them work? ✓

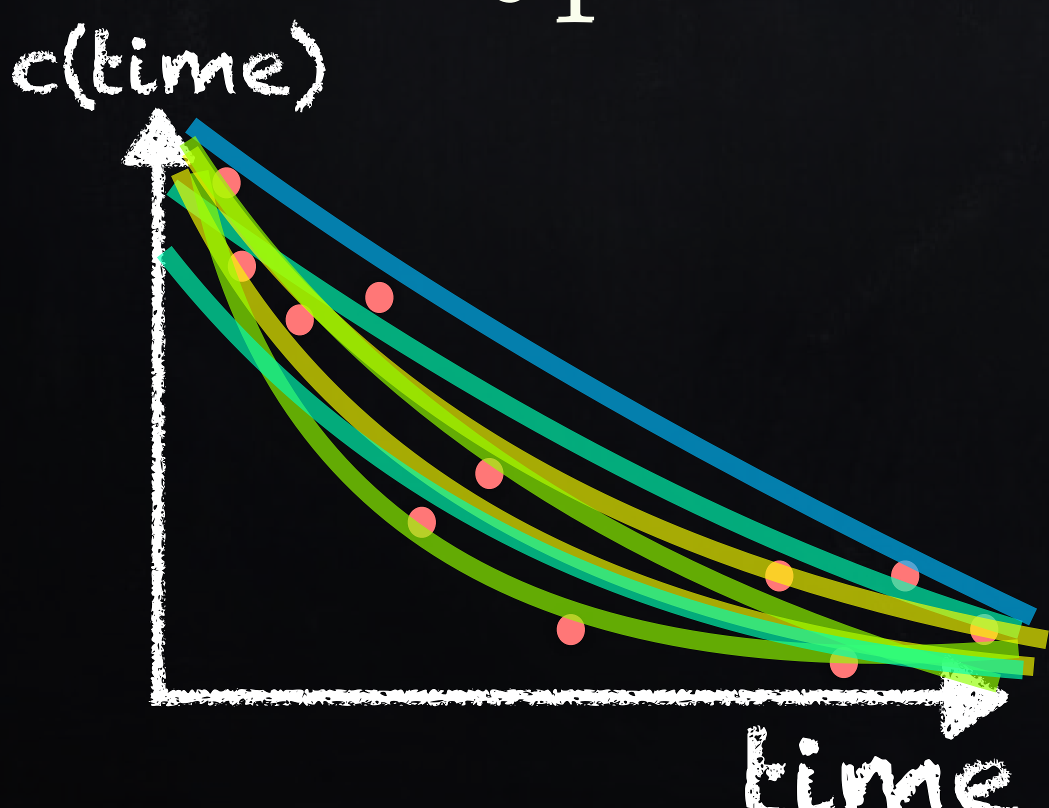


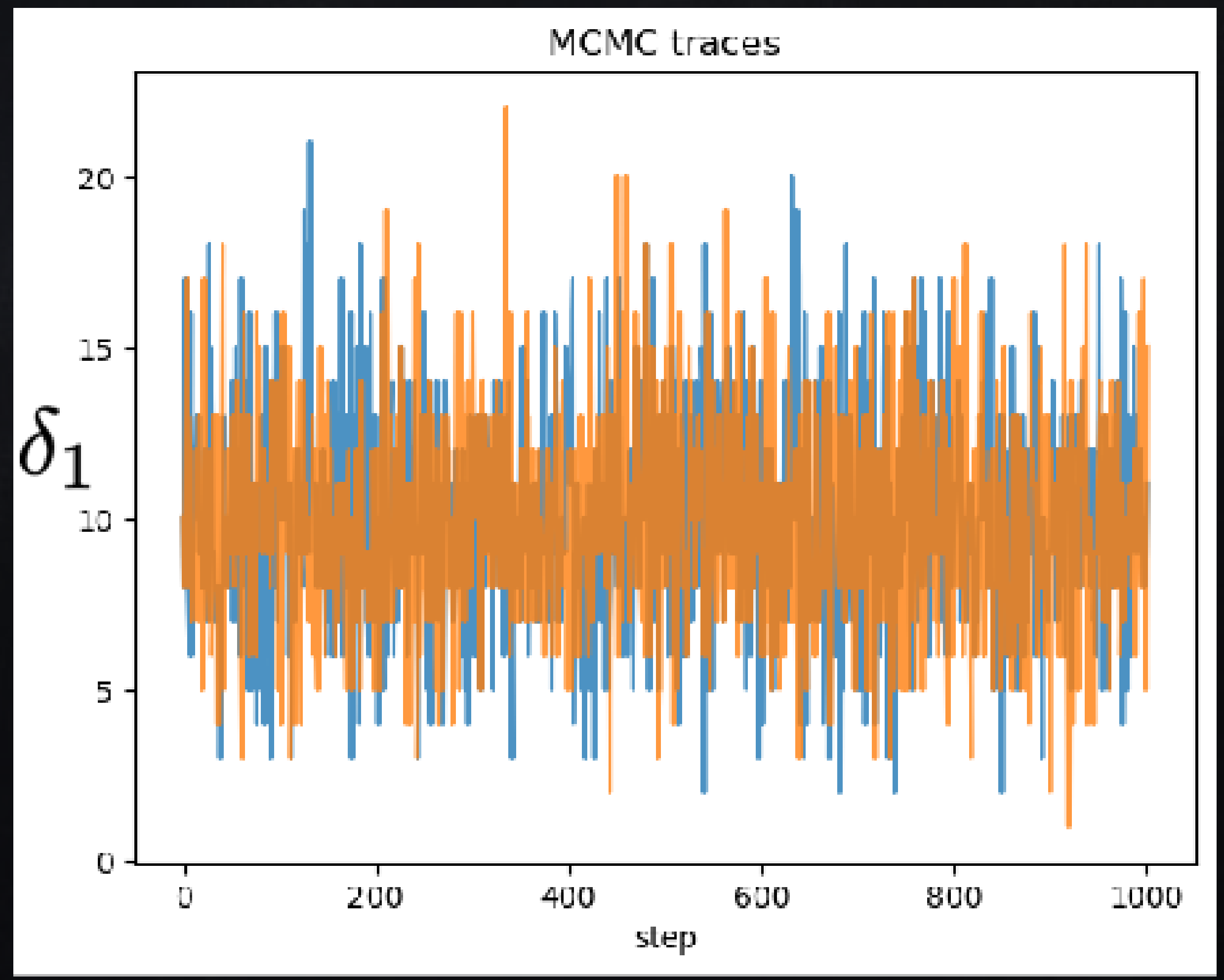
Markov Chain Monte Carlo (MCMC)

$$\delta^{(1)}, \delta^{(2)}, \delta^{(3)}, \delta^{(4)}, \ldots \sim P(\delta|data)$$

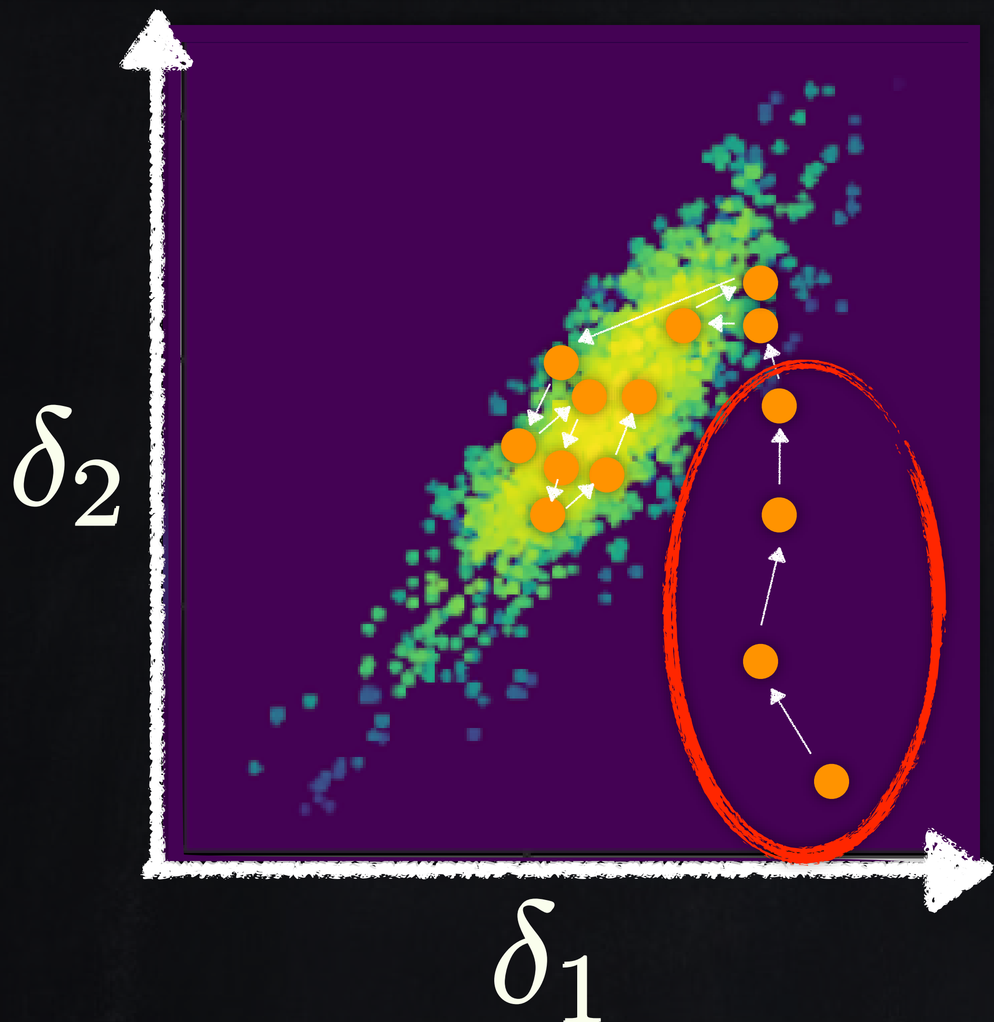
MCMC diagnosis

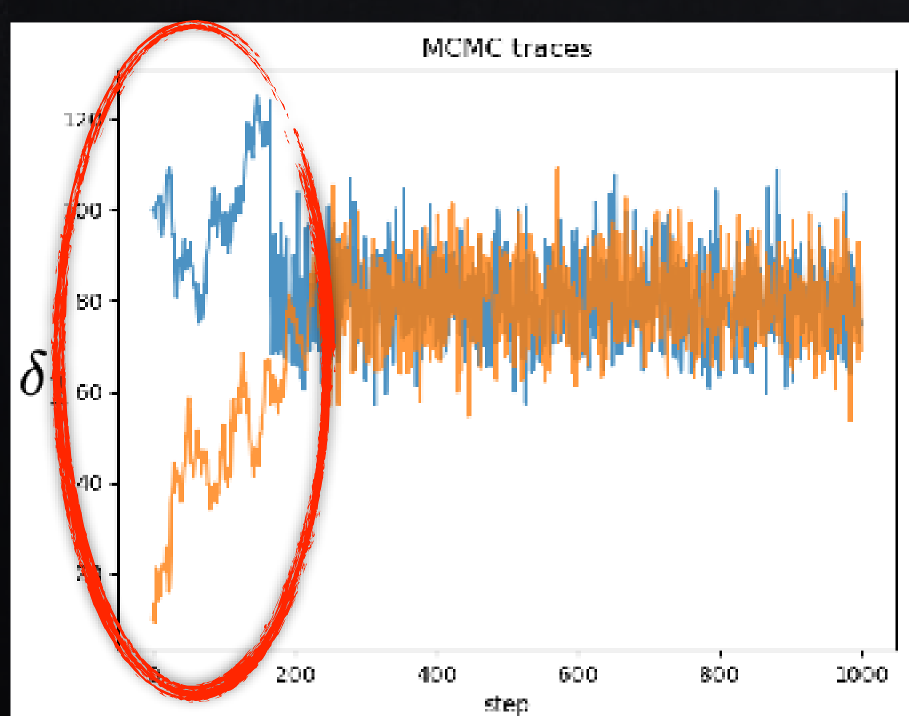
Trace plot : plot of Markov chain progression





How the algorithms behind them work?





Markov Chain Monte Carlo (MCMC)

$$\delta^{(1)}, \delta^{(2)}, \delta^{(3)}, \delta^{(4)}, \ldots \sim P(\delta|data)$$

MCMC diagnosis

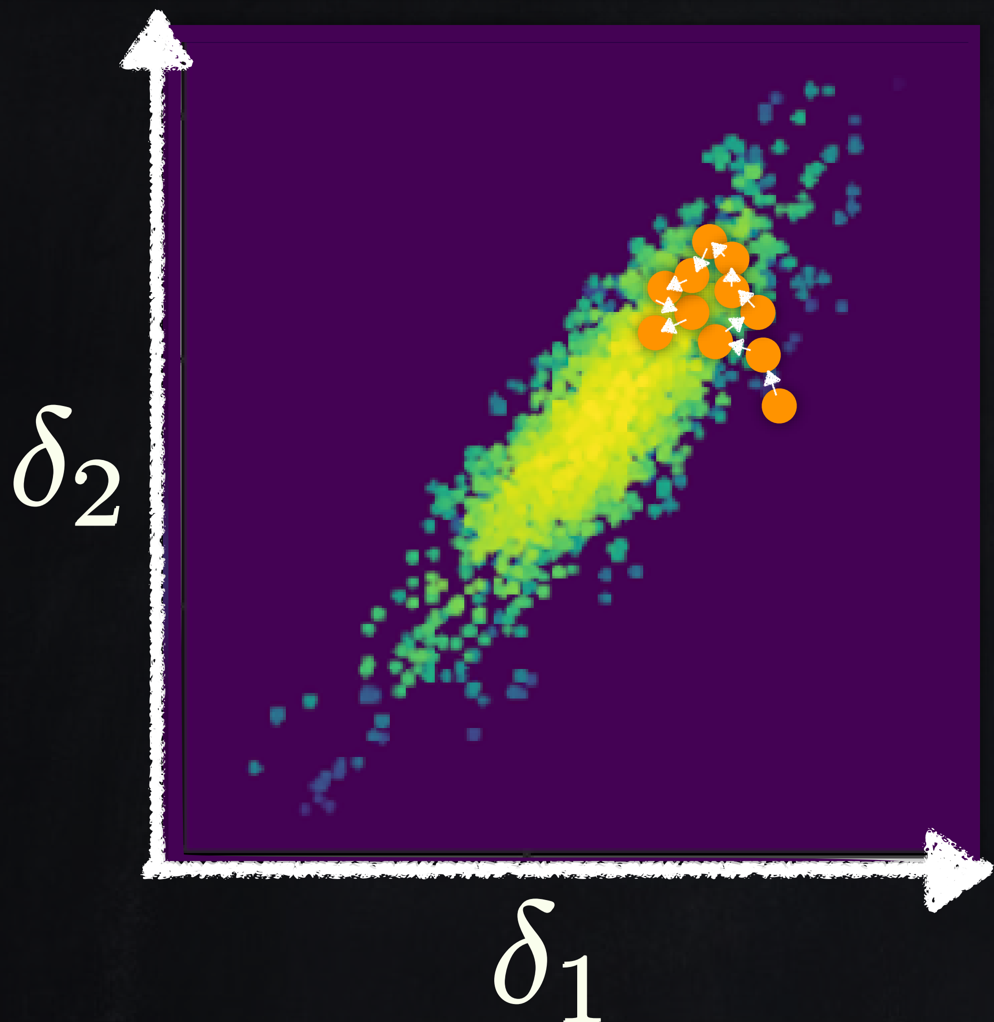
Case 1: “Burn-in”

Problem: Initial part of trace not representative of $P(\delta|data)$

Solution: Remove an initial fraction of points from the MCMC

Objective measure: Variance should be constant along the chain

How the algorithms behind them work?



Markov Chain Monte Carlo (MCMC)

$$\delta^{(1)}, \delta^{(2)}, \delta^{(3)}, \delta^{(4)}, \ldots \sim P(\delta|data)$$

MCMC diagnosis

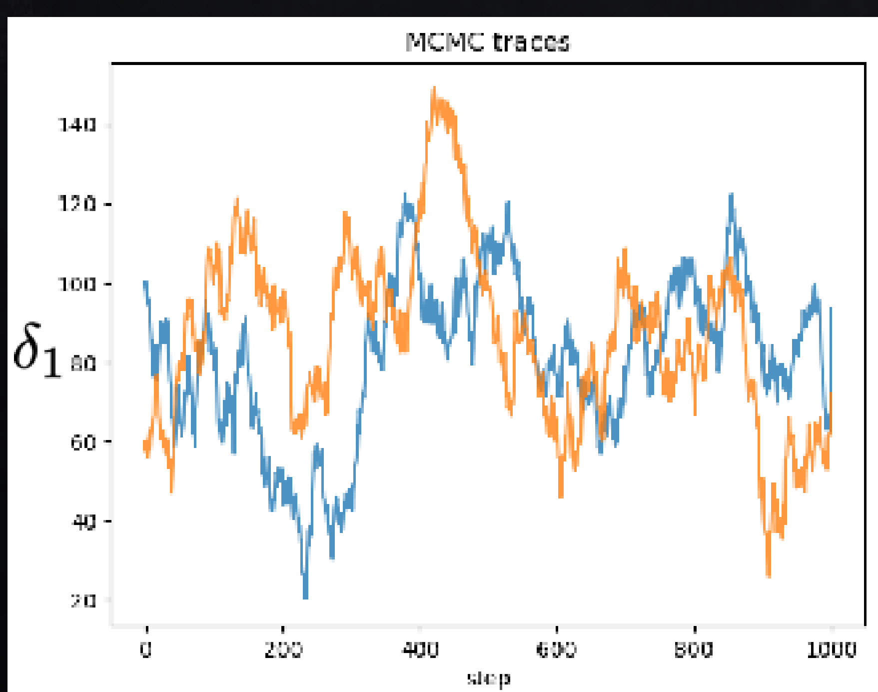
Case 2: Poor mixing

Problem: Points are correlated, limiting the information of the chain

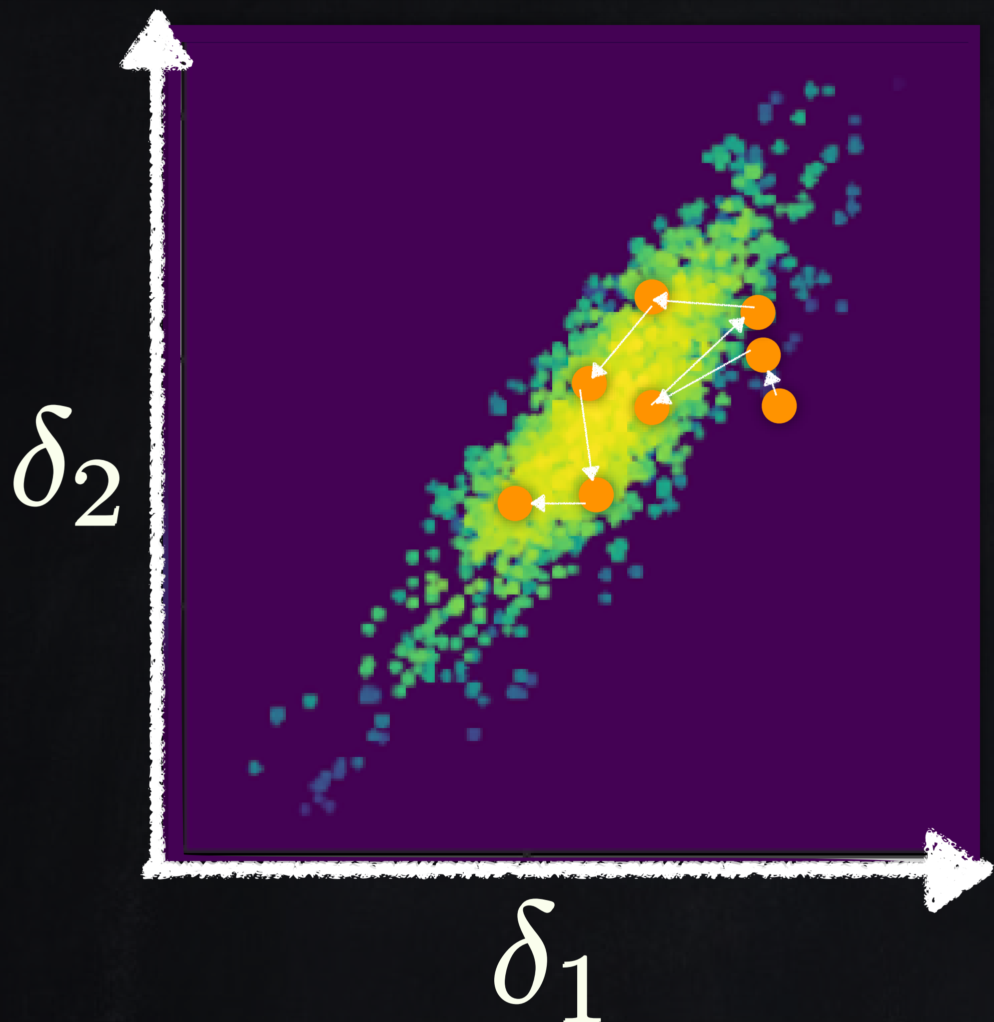
Solution: Increase size of proposal step kernel

Objective measure:
Effective sampling size

$$N_{\mathrm{eff}} = \frac{N}{1 + 2\sum_{\tau} \mathrm{corr}(\delta_i, \delta_{i+\tau})} > 1000$$



How the algorithms behind them work? ✓



Markov Chain Monte Carlo (MCMC)

$$\delta^{(1)}, \delta^{(2)}, \delta^{(3)}, \delta^{(4)}, \ldots \sim P(\delta|data)$$

MCMC diagnosis

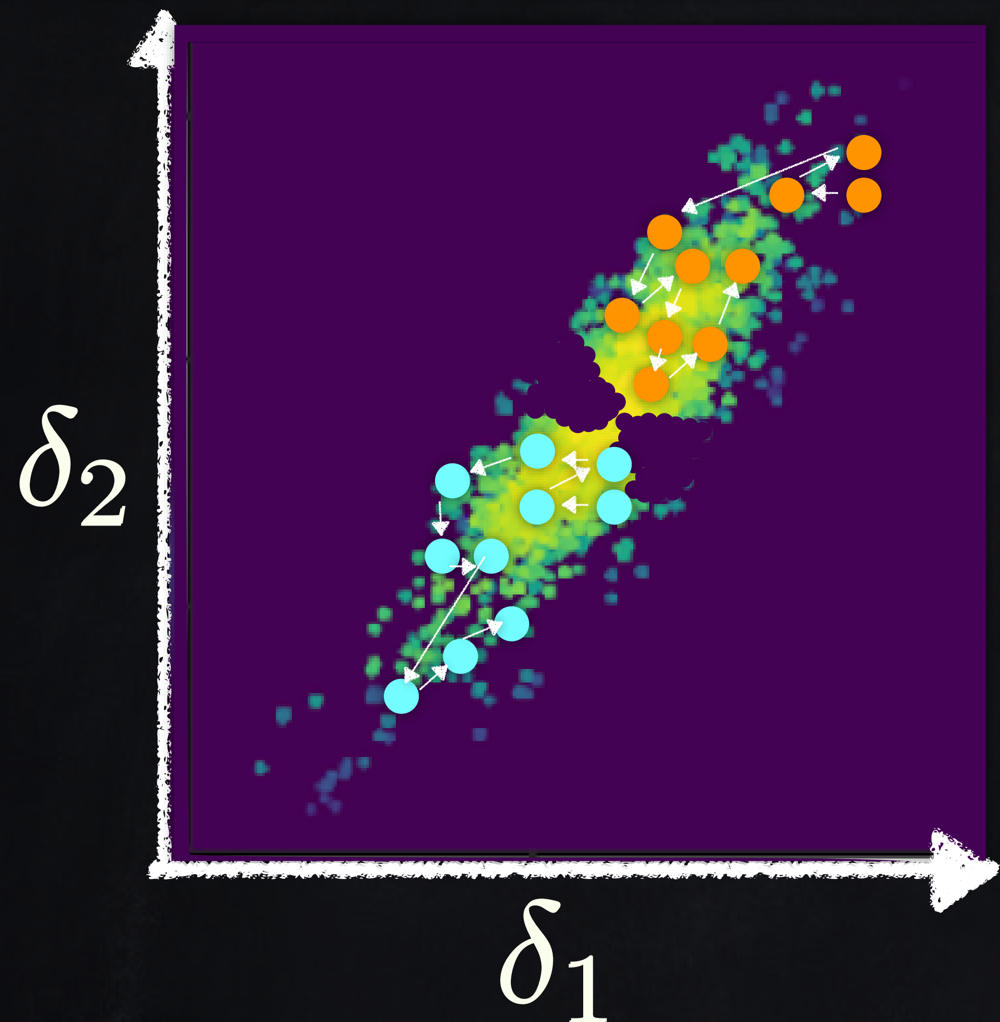
Case 3: Slow convergence

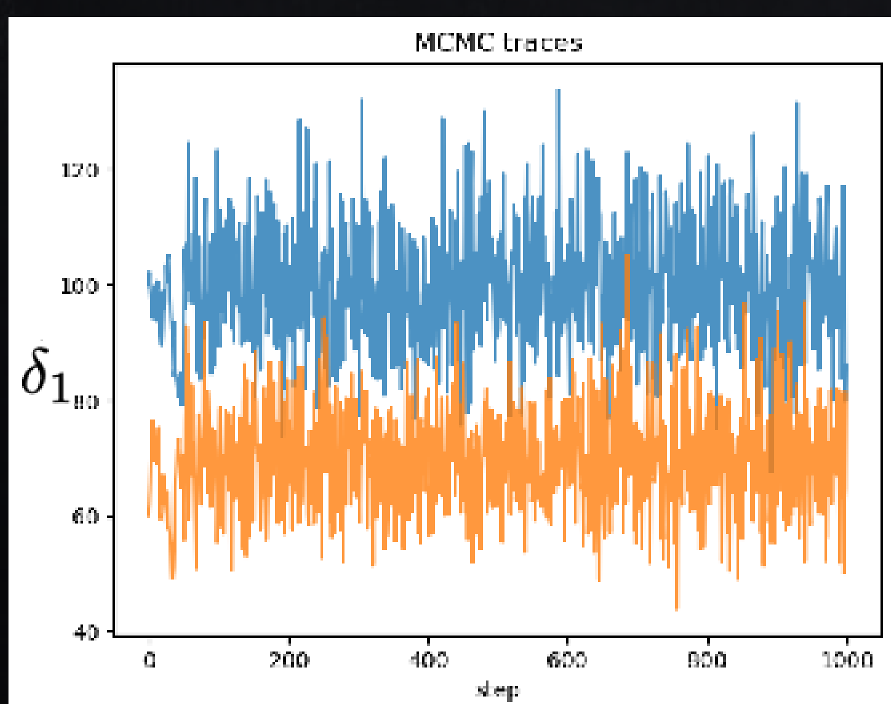
Problem: Most points proposals are rejected

Solution: Decrease size of proposal step kernel

Objective measure:
Rate of acceptance <10%

How the algorithms behind them work? ✓





Markov Chain Monte Carlo (MCMC)

$$\delta^{(1)}, \delta^{(2)}, \delta^{(3)}, \delta^{(4)}, \ldots \sim P(\delta|data)$$

MCMC diagnosis

Case 4: Multimodality / irregular posterior shape

Problem: Different convergence across chains
Not clear steady chain convergence

Solution: Increase chain number or chain size

Objective measure:

Gelman-Rubin statistic ($\hat{R}$)
Compares variance within chains with variance between chains

$$1 < \hat{R} < 1.1$$

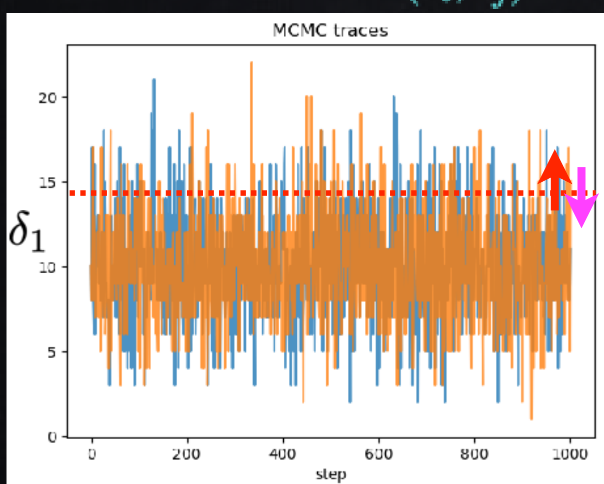
How the algorithms behind them work? ✓

OK, but WHY does MCMC work?? ✓

MCMC steady state probability   Posterior probability

$$P^{MC}(\delta) \overset{?}{=} P(\delta)$$

$$P^{MC}(\delta^{(i)} \to \delta^{(j)}) = P^{MC}(\delta^{(j)} \to \delta^{(i)}) \equiv P^{MC}(\delta_i, \delta_j)$$



$$P^{MC}(\delta^{(n)}, \delta^{(n+1)}) = P^{MC}(\delta^{(n)}) P^{MC}(\delta^{(n+1)} | \delta^{(n)})$$

MCMC algorithm

$$P^{MC}(\delta^{(n)}, \delta^{(n+1)}) = P^{MC}(\delta^{(n+1)}) P^{MC}(\delta^{(n)} | \delta^{(n+1)}) \qquad \div$$

$$\frac{P^{MC}(\delta^{(n)})}{P^{MC}(\delta^{(n+1)})} = \frac{P^{MC}(\delta^{(n)} | \delta^{(n+1)})}{P^{MC}(\delta^{(n+1)} | \delta^{(n)})} = \frac{\cancel{q(\delta^{(n)}, \delta^{(n+1)})}\min(1, P(\delta^{(n)})/P(\delta^{(n+1)}))}{\cancel{q(\delta^{(n+1)}, \delta^{(n)})}\min(1, P(\delta^{(n+1)})/P(\delta^{(n)}))}$$

$$= \frac{P(\delta^{n})}{P(\delta^{n+1})}$$ ✓ q.e.d

fun!! Can you give more examples of Likelihood? ✓

Bayesian inference methods for ecology

Theory

What is Bayesian inference?

Probability 101

Bayes' theorem in biology

Priors and likelihood

Interpretation of posteriors

Markov Chain Monte Carlo

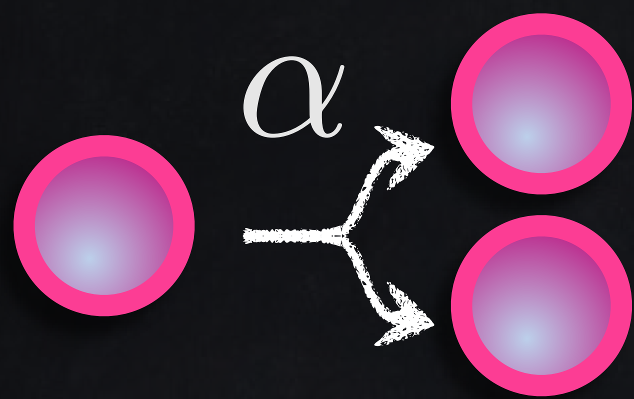
Applications

Population dynamics in microscopic and macroscopic communities

ODE generative models: Morphogen patterning of embryonic tissues

Phylogeny and model selection

Example 1.1: Determining a proportion



probability of proliferation

probability of apoptosis

$$p = \frac{\alpha}{\alpha + \beta}$$

$$q = 1 - p = \frac{\beta}{\alpha + \beta}$$

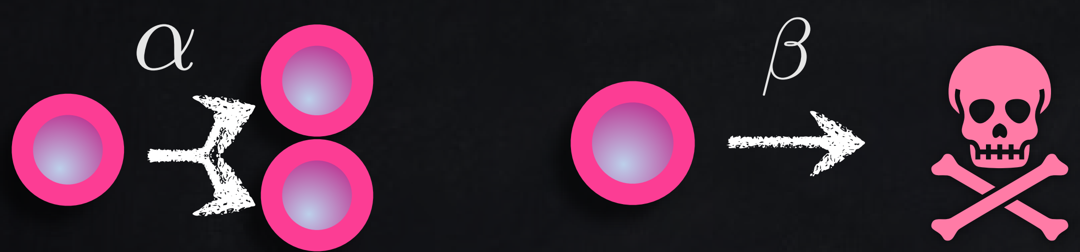
This example applies to any experiment in which you are measuring the fraction of a trait with two possible outcomes

e.g. fraction of males in a population of penguins ← practical session

fraction of trees in a forest with a specific disease

fraction of mating attempts that are successful

Example 1.1: Determining a proportion



$$p = \frac{\alpha}{\alpha + \beta}$$

N experiments

N_α proliferations

Proliferated	Died
X X X X	X X X X
X X X X	
X X X X	
X X	

$$P(p|N_\alpha) = \frac{P(N_\alpha|p)P(p)}{P(N_\alpha)}$$

Example 1.1: Determining a proportion

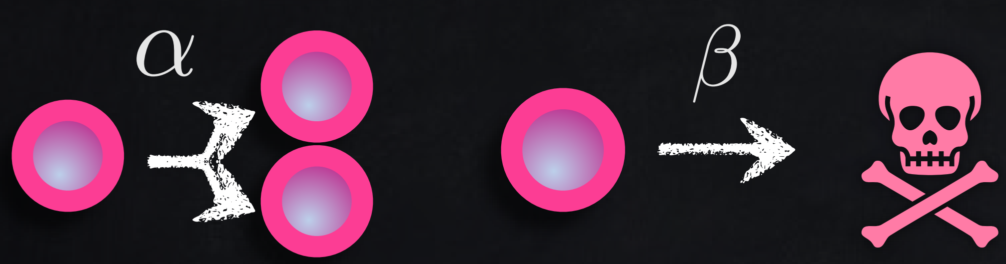
$$p = \frac{\alpha}{\alpha + \beta}$$

N experiments

N_α proliferations

$$P(p|N_\alpha) = \frac{P(N_\alpha|p)P(p)}{P(N_\alpha)}$$

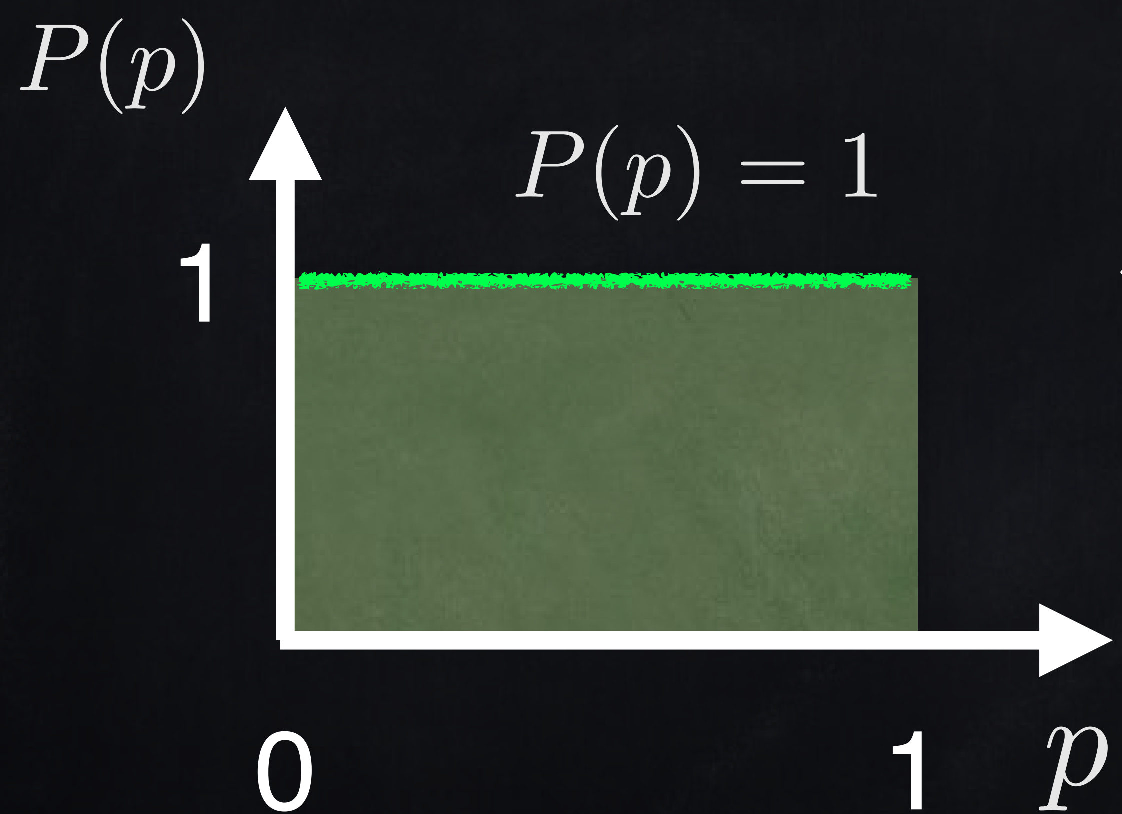
Example 1.1: Determining a proportion



$$p = \frac{\alpha}{\alpha + \beta}$$

N experiments

N_α proliferations



Prior

$$P(p|N_\alpha) = \frac{P(N_\alpha|p)P(p)}{P(N_\alpha)}$$

Example 1.1: Determining a proportion

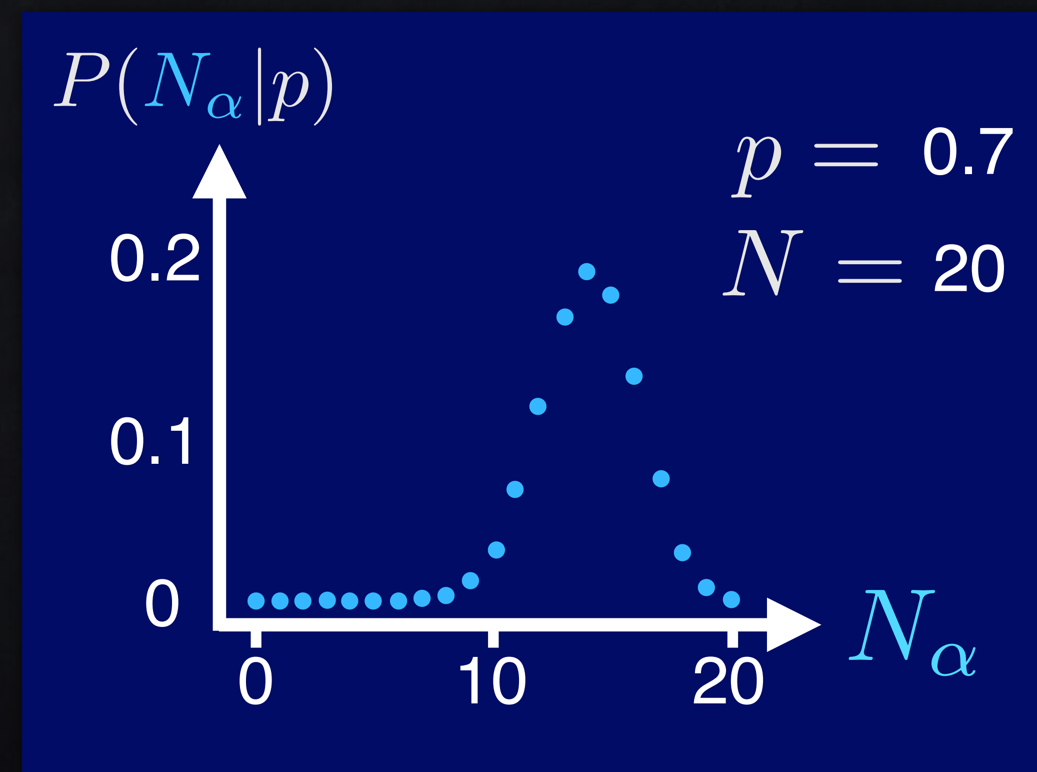
$$p = \frac{\alpha}{\alpha + \beta}$$

N experiments

N_α proliferations

Likelihood follows a binomial distribution

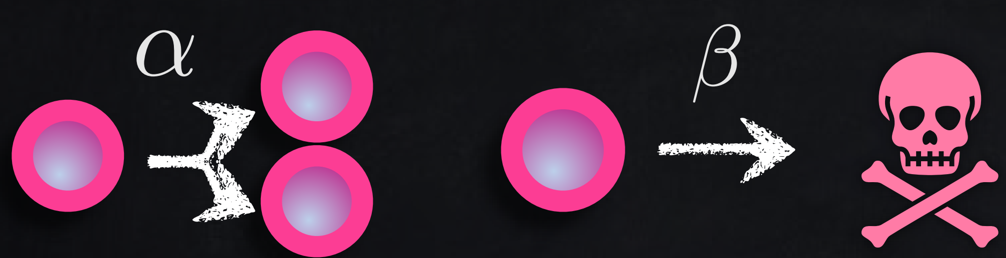
$$P(N_\alpha | p) = \binom{N}{N_\alpha} p^{N_\alpha} (1-p)^{N - N_\alpha}$$



Likelihood

$$P(p|N_\alpha) = \frac{P(N_\alpha|p)P(p)}{P(N_\alpha)} \qquad P(p) = 1$$

Example 1.1: Determining a proportion



$$p = \frac{\alpha}{\alpha + \beta}$$

N experiments

N_α proliferations

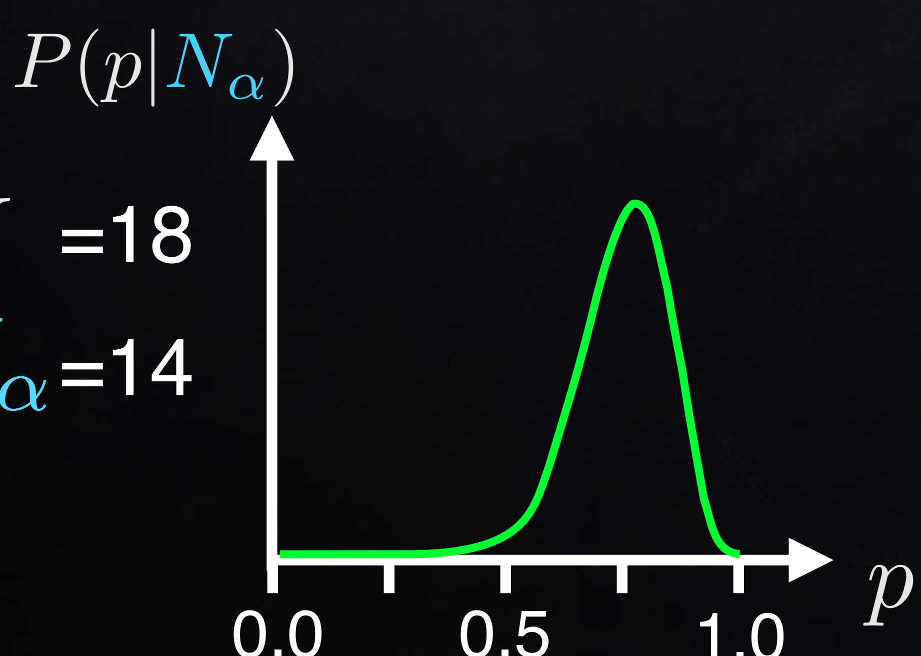
Posterior distribution

$$P(p|N_\alpha) = \frac{1}{P(N_\alpha)} \binom{N}{N_\alpha} p^{N_\alpha} (1-p)^{N-N_\alpha}$$

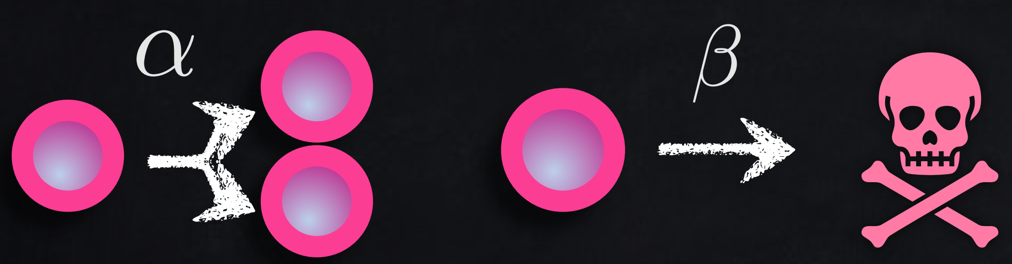
$$P(p|N_\alpha) \propto p^{N_\alpha} (1-p)^{N-N_\alpha}$$

N =18

N_α=14



Example 1.1: Determining a proportion



$$p = \frac{\alpha}{\alpha + \beta}$$

N experiments

N_α proliferations

Posterior distribution

$$P(p|N_\alpha) = \frac{1}{P(N_\alpha)} \binom{N}{N_\alpha} p^{N_\alpha} (1-p)^{N-N_\alpha}$$

$$P(p|N_\alpha) \propto p^{N_\alpha} (1-p)^{N-N_\alpha}$$

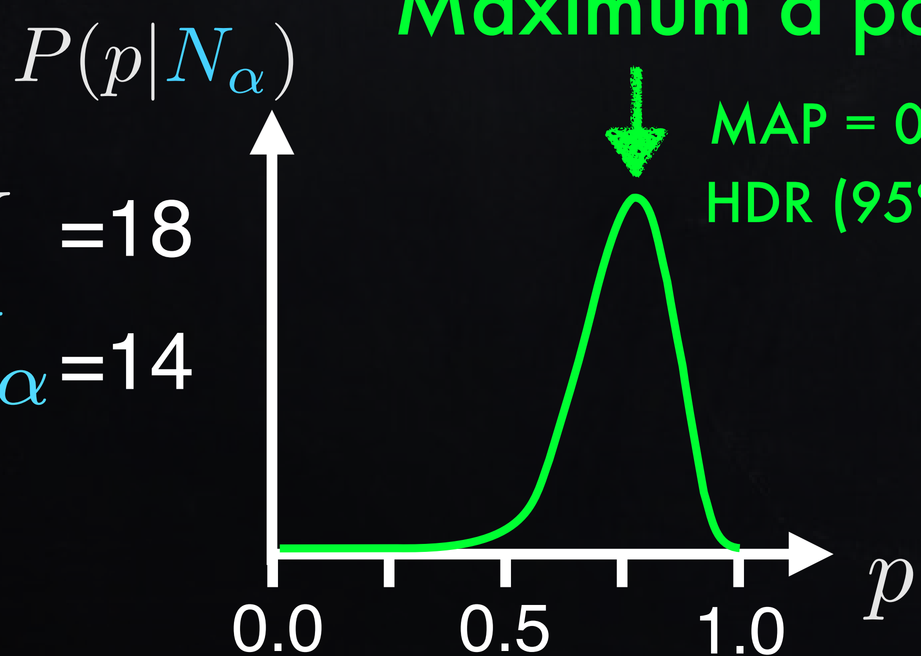
Maximum a posteriori (MAP)

MAP = 0.78

HDR (95%) = [0.56,0.92]

N =18

N_α=14



Example 1.1: Determining a proportion

$$p = \frac{\alpha}{\alpha + \beta}$$

N experiments

N_α proliferations

Posterior distribution

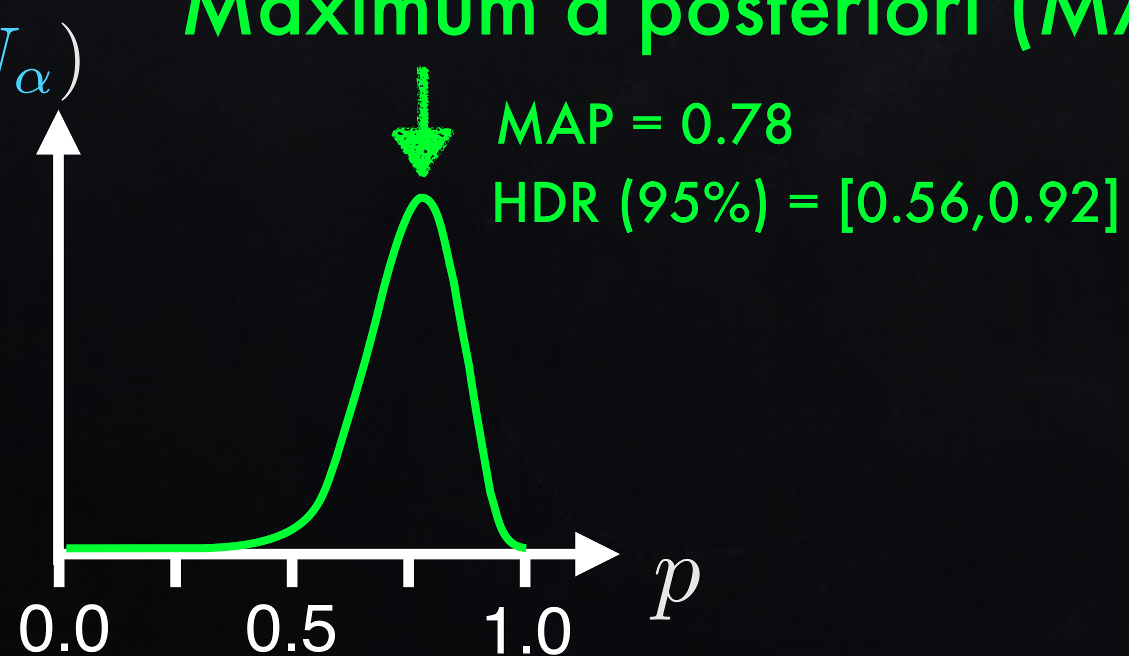
$$P(p|N_\alpha) = \frac{1}{P(N_\alpha)} \binom{N}{N_\alpha} p^{N_\alpha} (1-p)^{N-N_\alpha}$$

$$P(p|N_\alpha) \propto p^{N_\alpha} (1-p)^{N-N_\alpha}$$

Maximum a posteriori (MAP)

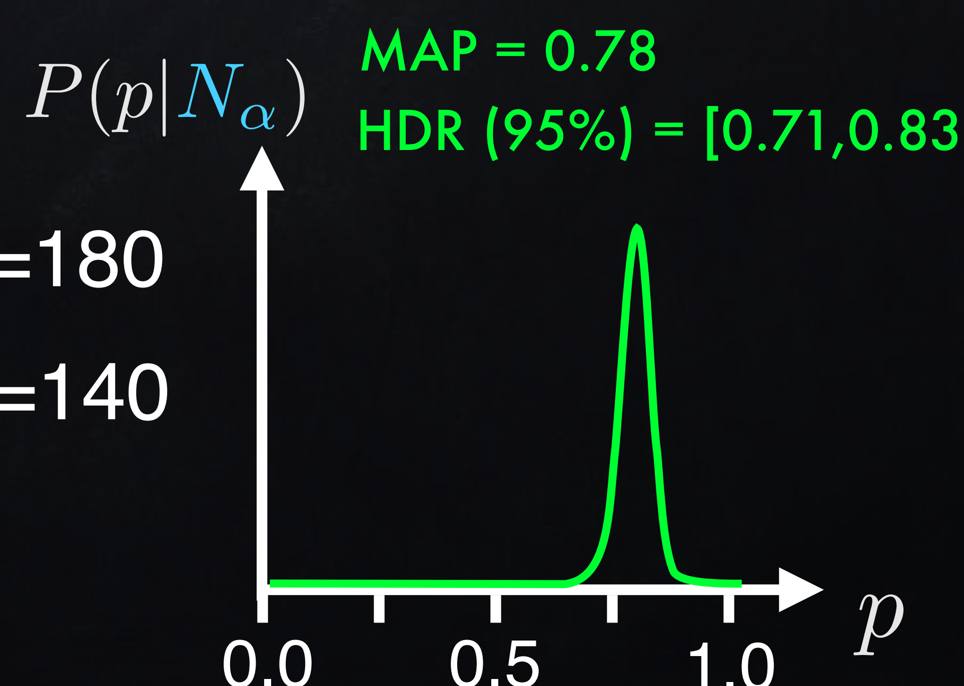
N =18

N_α=14

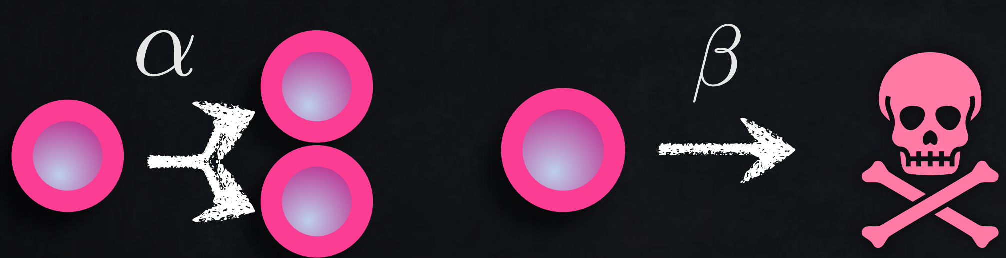


N =180

N_α =140



Example 1.1: Determining a proportion



$$p = \frac{\alpha}{\alpha + \beta}$$

N experiments

N_α proliferations

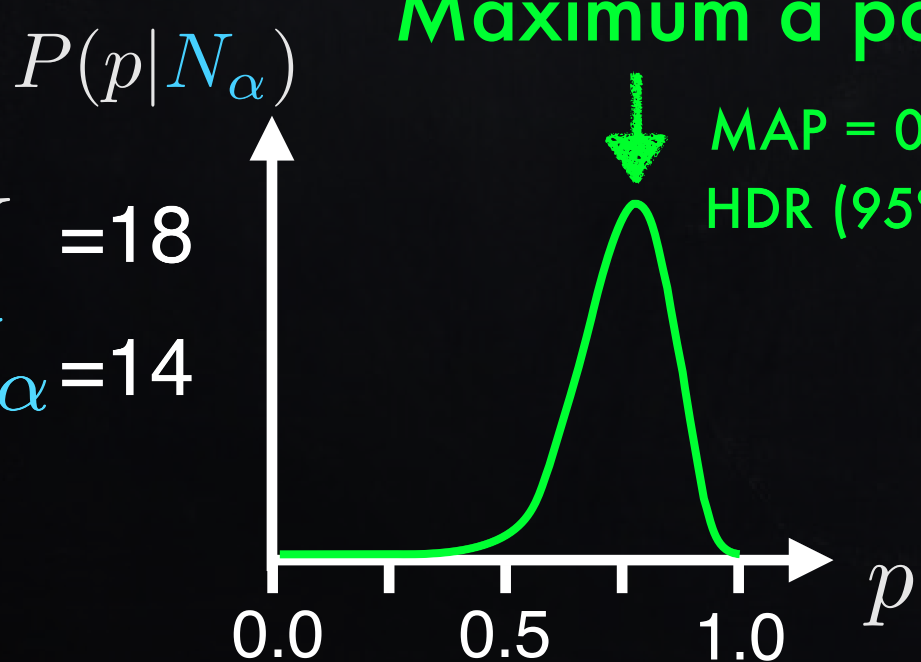
$$P(p|N_\alpha) \propto p^{N_\alpha}(1-p)^{N-N_\alpha}$$

Bayesian inference automatically incorporates the replicate information

Sometimes posteriors can be analysed analytically

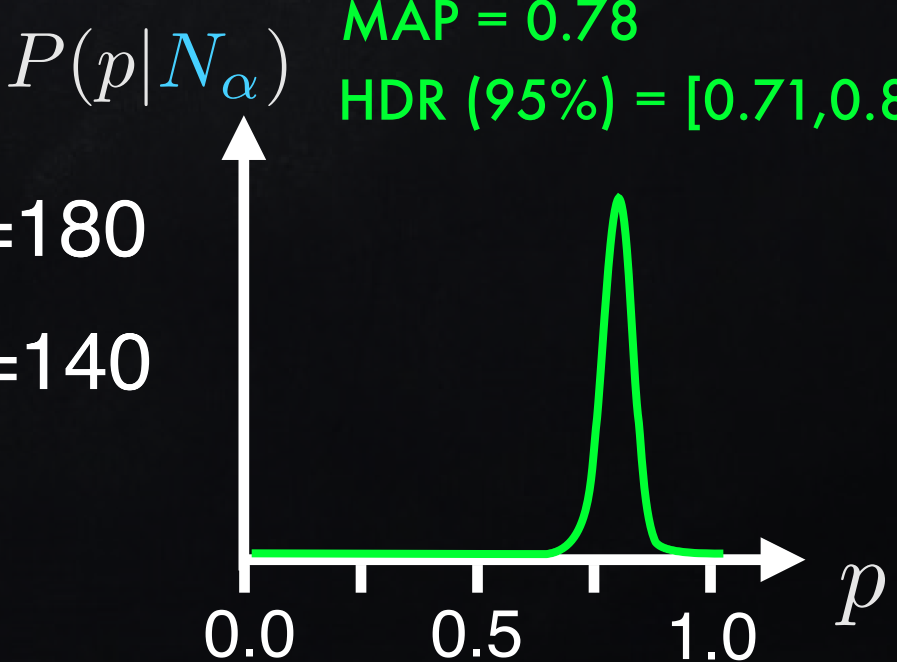
Maximum a posteriori (MAP)

N =18
N_α =14



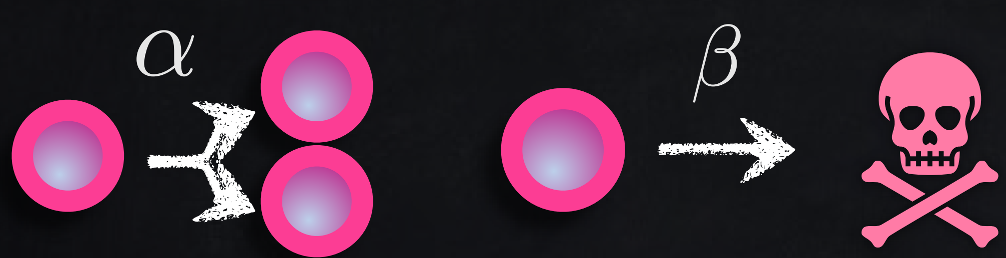
MAP = 0.78
HDR (95%) = [0.56,0.92]

N =180
N_α =140



MAP = 0.78
HDR (95%) = [0.71,0.83]

Example 1.1: Determining a proportion



$$p = \frac{\alpha}{\alpha + \beta}$$

N experiments

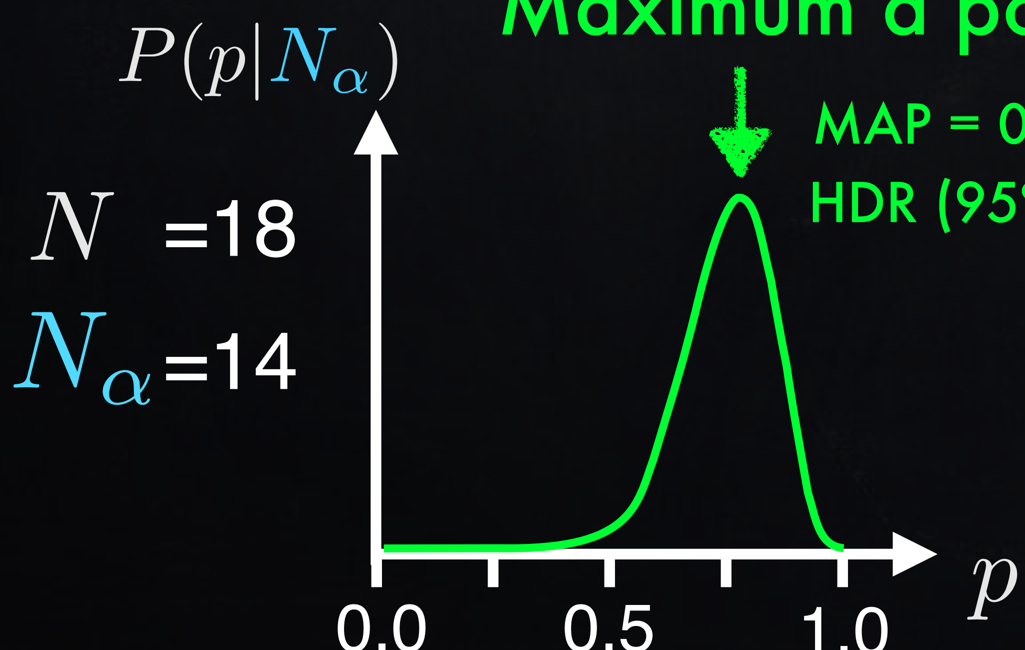
N_α proliferations

$$P(p|N_\alpha) \propto p^{N_\alpha}(1-p)^{N-N_\alpha}$$

Bayesian inference automatically incorporates the replicate information

Sometimes posteriors can be analysed analytically

Maximum a posteriori (MAP)



MAP = 0.78

Question:

Show that the MAP of the binomial inference will always be N_α/N

Question:

Somebody before us has done the same experiment and has observed 10 proliferations in 30 experiments.

- How would you include this as a prior distribution?

- Will you obtain the same posterior distribution by including this previous experiment as a prior distribution than adding the information of both experiments in the likelihood and assuming a uniform prior?