Why doing Bayesian analysis? 🤔

Why doing Bayesian analysis? 🤔

- It's the easiest way to jointly estimate many parameters at once

Why doing Bayesian analysis? 🤔

- It's the easiest way to jointly estimate many parameters at once

- It's the easiest way to get real credible intervals around your parameter estimates

Why doing Bayesian analysis? 🤔

- It's the easiest way to jointly estimate many parameters at once
- It's the easiest way to get real credible intervals around your parameter estimates
- In some cases it's the only way to get the right answer

Why doing Bayesian analysis? 🤔

- It's the easiest way to jointly estimate many parameters at once
- It's the easiest way to get real credible intervals around your parameter estimates
- In some cases it's the only way to get the right answer
- It's much easier to implement MCMC than maximum likelihood optimizers

Why doing Bayesian analysis? 🤔

- It's the easiest way to jointly estimate many parameters at once
- It's the easiest way to get real credible intervals around your parameter estimates
- In some cases it's the only way to get the right answer
- It's much easier to implement MCMC than maximum likelihood optimizers
- It allows hierarchical models, which in turn simplify model testing

Why doing Bayesian analysis? 🤔

- It's the easiest way to jointly estimate many parameters at once
- It's the easiest way to get real credible intervals around your parameter estimates
- In some cases it's the only way to get the right answer
- It's much easier to implement MCMC than maximum likelihood optimizers
- It allows hierarchical models, which in turn simplify model testing
- It's the best way to accommodate uncertainties in the data and parameter estimates

Why doing Bayesian analysis? 🤔

- It's the **easiest** way to jointly estimate many parameters at once
- It's the **easiest** way to get real credible intervals around your parameter estimates
- In some cases it's the only way to get the **right** answer
- It's much **easier** to implement MCMC than maximum likelihood optimizers
- It allows hierarchical models, which in turn **simplify** model testing
- It's the **best** way to accommodate uncertainties in the data and parameter estimates

Why doing Bayesian analysis? 🤔

- It's the **easiest** way to jointly estimate
- It's the **easiest** way to get real credib parameter estimates
- In some cases it's the only way to get
- It's much **easier** to implement MCMC optimizers
- It allows hierarchical models, which i
- It's the **best** way to accommodate un parameter estimates

Basic probability rules

P(X and Y) = P(X) x P(Y)

P(A or Y) = P(X) + P(Y)

(assuming X and Y are independent)

Basic probability rules

P(X and Y) = P(X) x P(Y)

P(A or Y) = P(X) + P(Y)

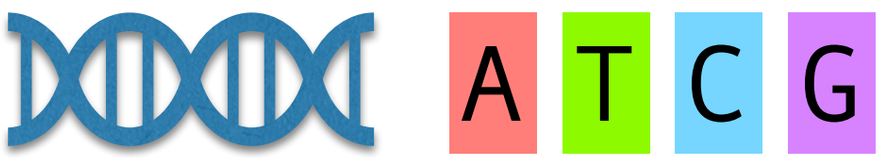
(assuming X and Y are independent)

Basic probability rules

P(X and Y) = P(X) x P(Y)

P(A or Y) = P(X) + P(Y)

(assuming X and Y are independent)



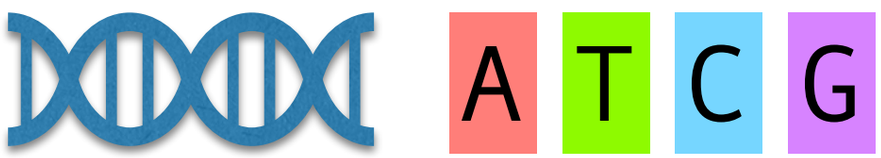
P(A) = 0.25

Basic probability rules

P(X and Y) = P(X) x P(Y)

P(A or Y) = P(X) + P(Y)

(assuming X and Y are independent)



P(A) = 0.25

P(AC) =

Basic probability rules

P(X and Y) = P(X) x P(Y)

P(A or Y) = P(X) + P(Y)

(assuming X and Y are independent)

P(A) = 0.25

P(AC) =

P(purine) =

Basic probability rules

P(X and Y) = P(X) x P(Y)

P(A or Y) = P(X) + P(Y)

(assuming X and Y are independent)

P(A) = 0.25

P(AC) = 0.25 x 0.25 = 0.0625

P(purine) =

Basic probability rules

P(X and Y) = P(X) x P(Y)

P(A or Y) = P(X) + P(Y)

(assuming X and Y are independent)

P(A) = 0.25

P(AC) = 0.25 x 0.25 = 0.0625

P(purine) = 0.25 + 0.25 = 0.5

Bayes theorem

Likelihood Prior probability

Posterior probability

$$P(\alpha|x) = \frac{P(x|\alpha) \times P(\alpha)}{P(x)}$$

probability of the data a.k.a.
marginal likelihood, normalizing
constant or evidence

alpha = unknown parameters of a model *M* (alpha is usually what we are after)

x = data (e.g. a set of measurements, molecular sequences, etc.)

Bayes theorem

Likelihood Prior probability

Posterior probability

$$P(\alpha|x, M) = \frac{P(x|\alpha, M) \times P(\alpha|M)}{P(x|M)}$$

probability of the data a.k.a. marginal likelihood, normalizing constant or evidence

alpha = unknown parameters of a model *M* (alpha is usually what we are after)

x = data (e.g. a set of measurements, molecular sequences, etc.)

M = model (more about this later)

Bayes theorem

Likelihood Prior probability

Posterior probability

$$P(\alpha|x, M) = \frac{P(x|\alpha, M) \times P(\alpha|M)}{P(x|M)}$$

probability of the data a.k.a. marginal likelihood, normalizing constant or evidence

Likelihood: probability of observing the data you observe given a model and its parameters

Prior: probability of the parameter values based on some prior knowledge about them

Marginal likelihood: probability of observing the data you observe given a model and all possible parameter values

Bayes theorem

Likelihood Prior probability

Posterior probability

$$P(\alpha|x, M) = \frac{P(x|\alpha, M) \times P(\alpha|M)}{P(x|M)}$$

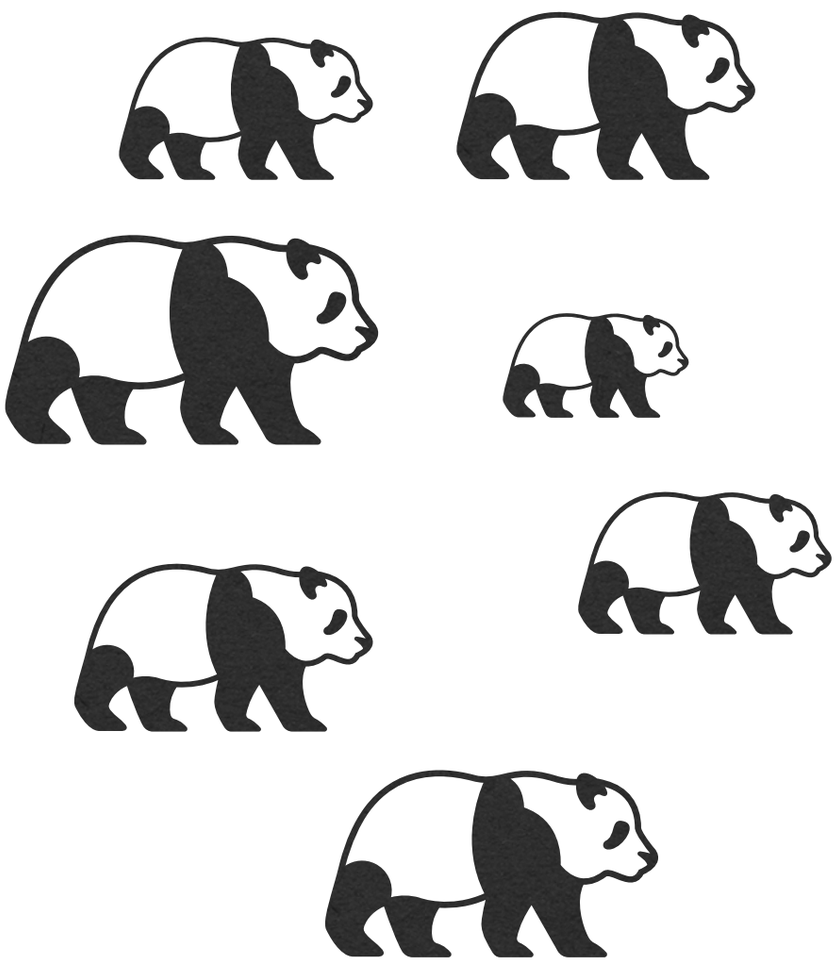
probability of the data a.k.a. marginal likelihood, normalizing constant or evidence

the three components:

data **model** **parameters**

Data

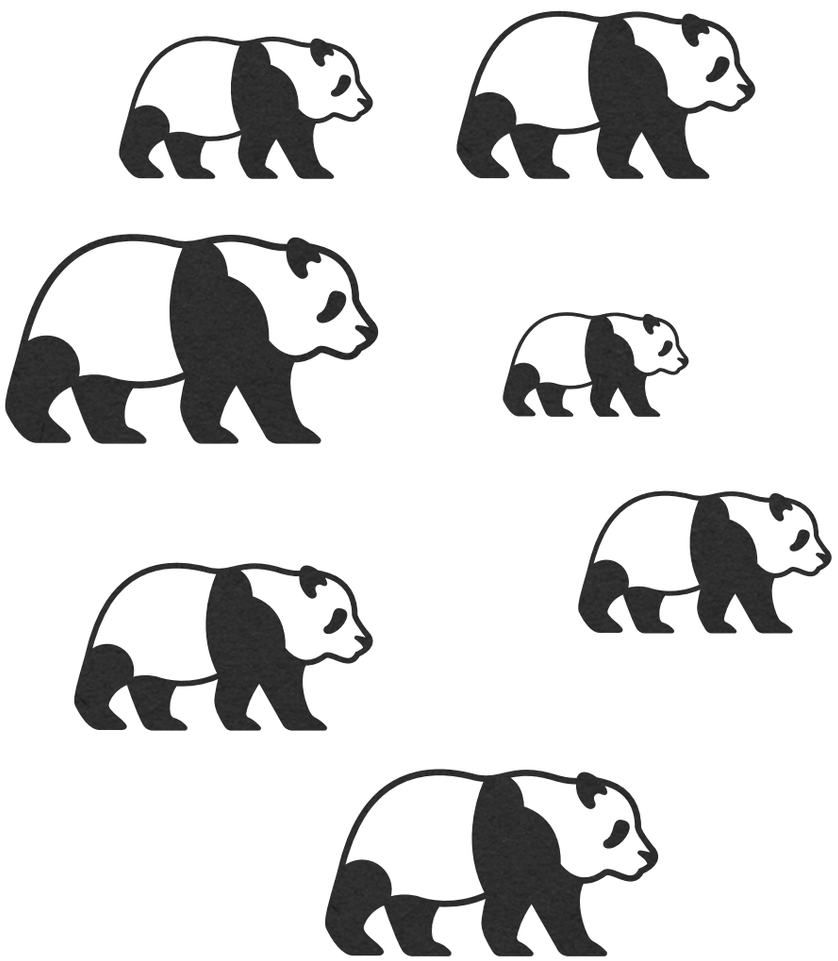
a panda population: several individuals of different sizes



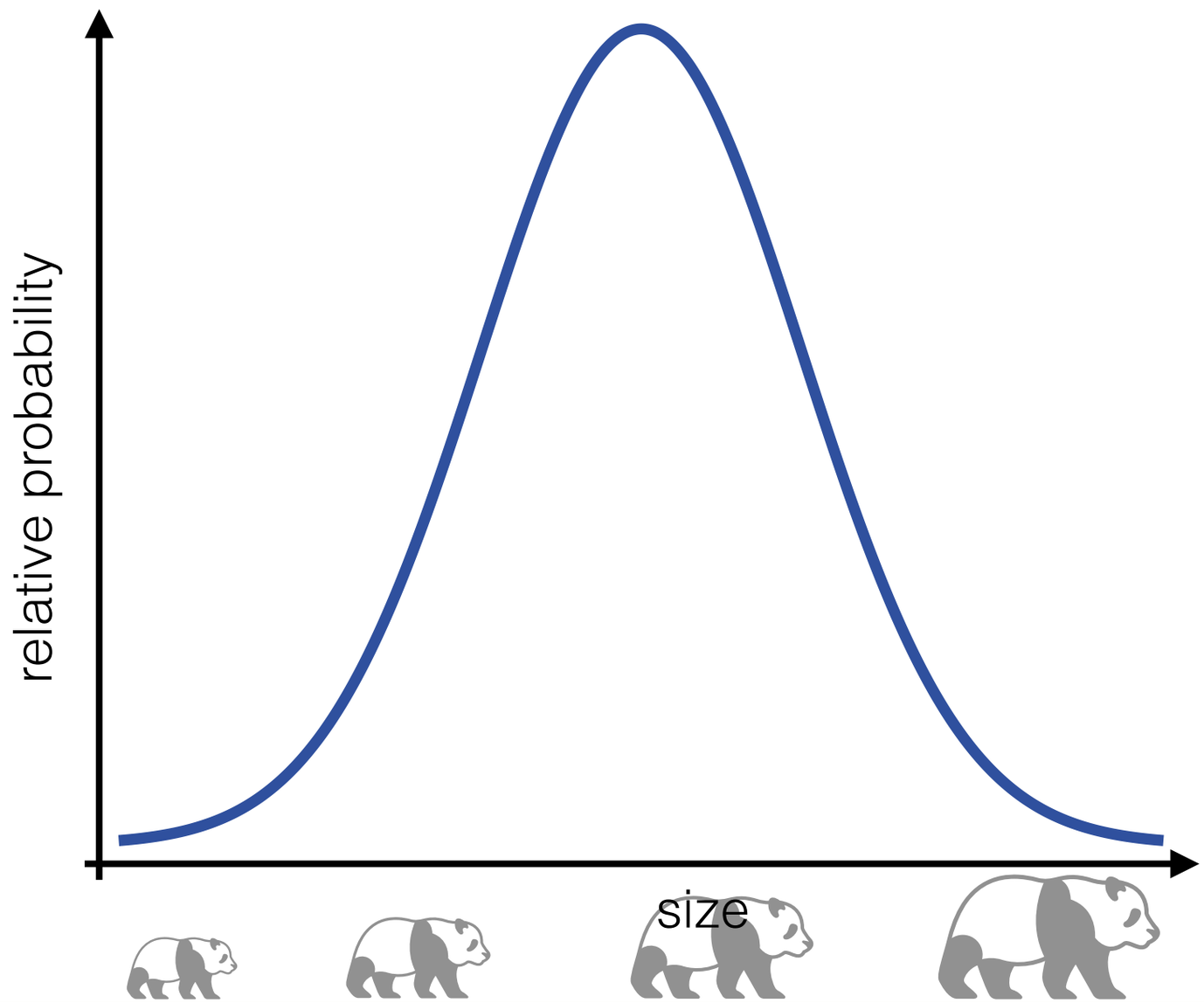
Model

The model should reflect the question we are after.

For instance: What are mean and variance of panda's sizes in a population?

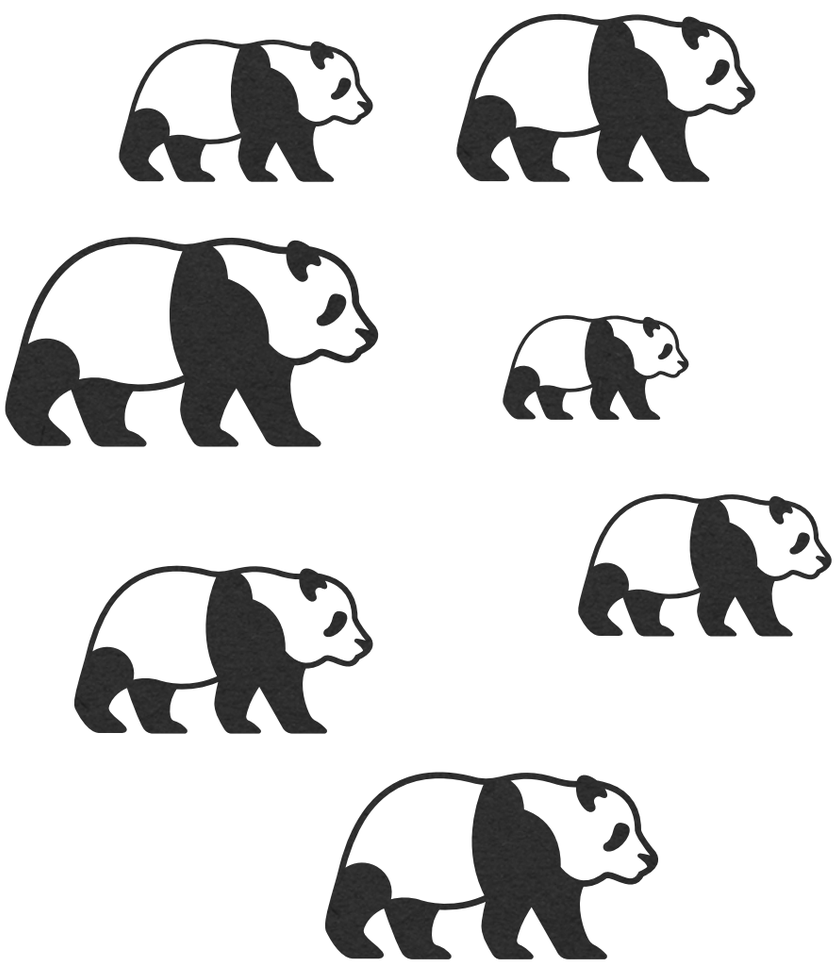


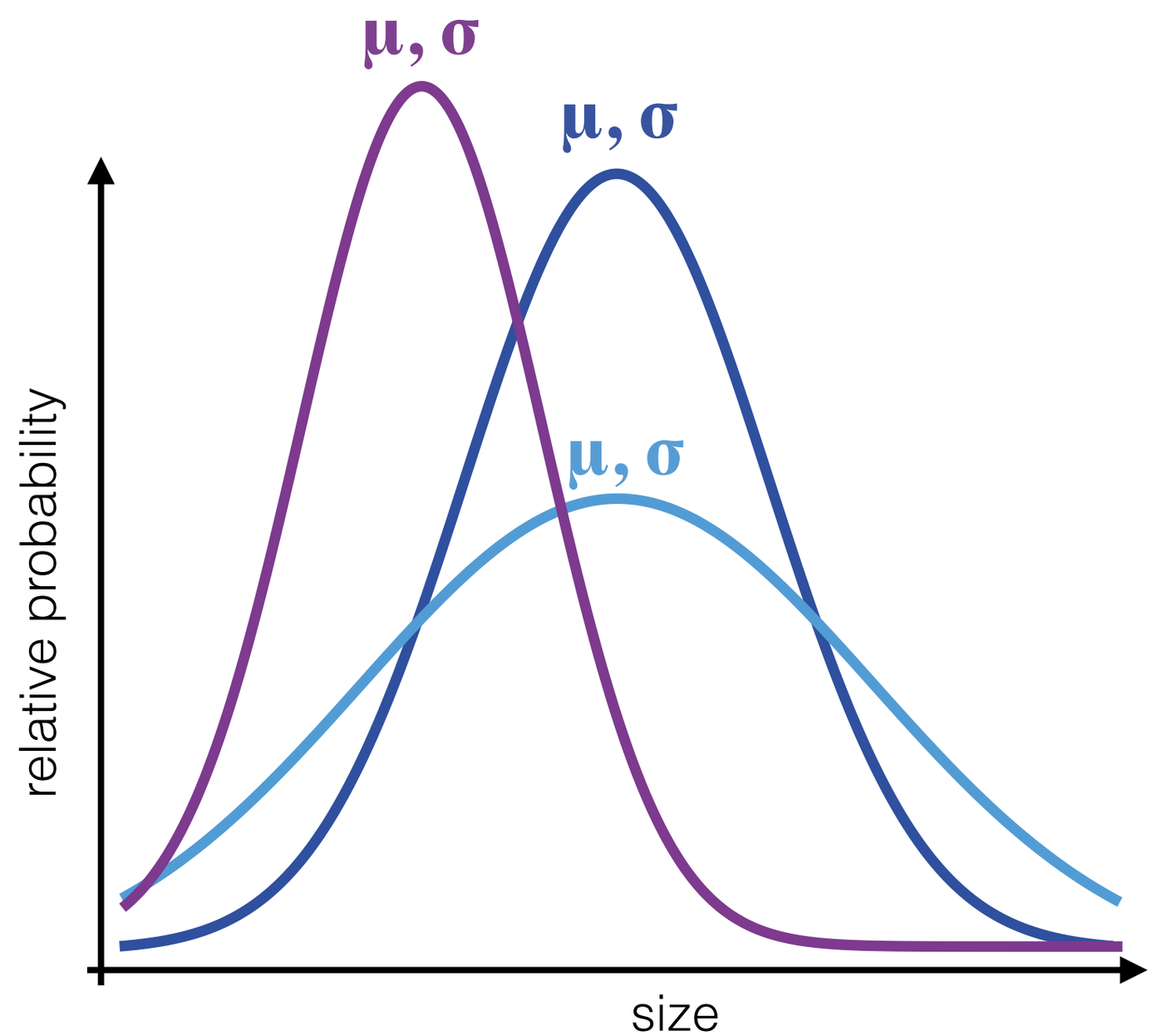
A simple example of a model is a distribution (e.g. a normal distribution)



Parameters

A model is defined by its parameters, e.g. mean and variance of a normal distribution





Bayes theorem for pandas

$$P(\alpha|x, M) = \frac{P(x|\alpha, M) \times P(\alpha|M)}{P(x|M)}$$

alpha = parameters

x = data

M = model

Posterior probability

Likelihood

Prior probability

$$P(\mu, \sigma|🐼, \mathcal{N}) = \frac{P(🐼|\mu, \sigma, \mathcal{N}) \times P(\mu, \sigma)}{P(🐼|\mathcal{N})}$$

Marginal likelihood

Bayes theorem for pandas

$$P(\alpha|x, M) = \frac{P(x|\alpha, M) \times P(\alpha|M)}{P(x|M)}$$

alpha = parameters

x = data

M = model

Posterior probability

Likelihood

Prior probability

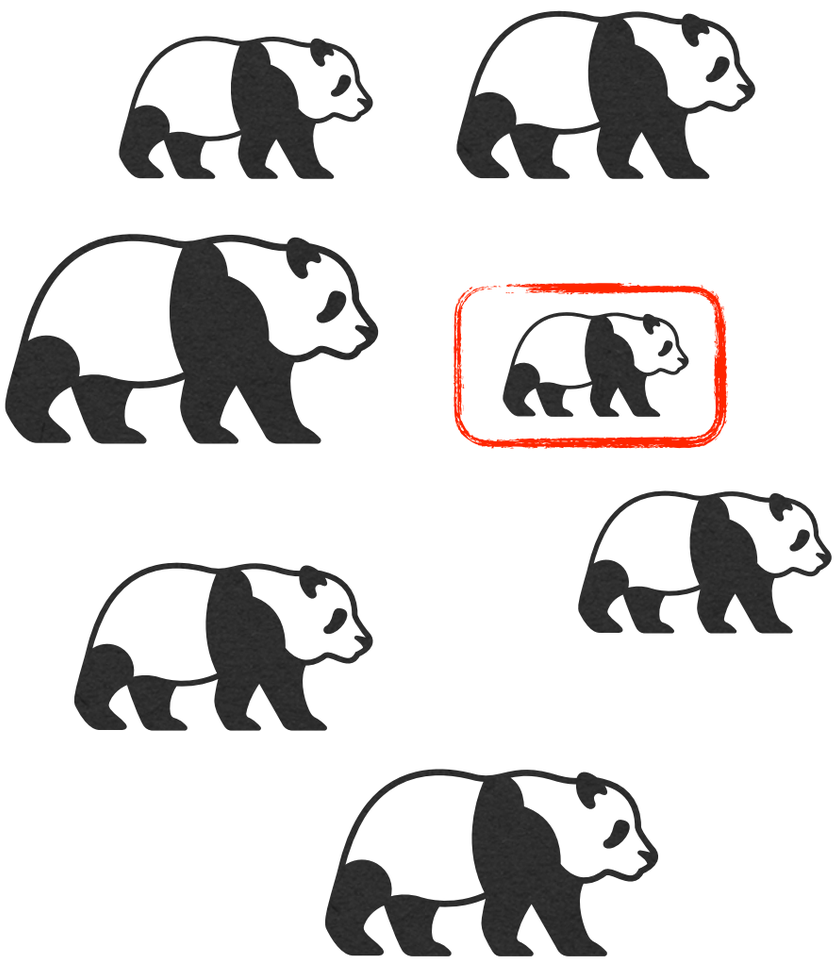
$$P(\mu, \sigma|🐼, \mathcal{N}) = \frac{P(🐼|\mu, \sigma, \mathcal{N}) \times P(\mu, \sigma)}{P(🐼|\mathcal{N})}$$

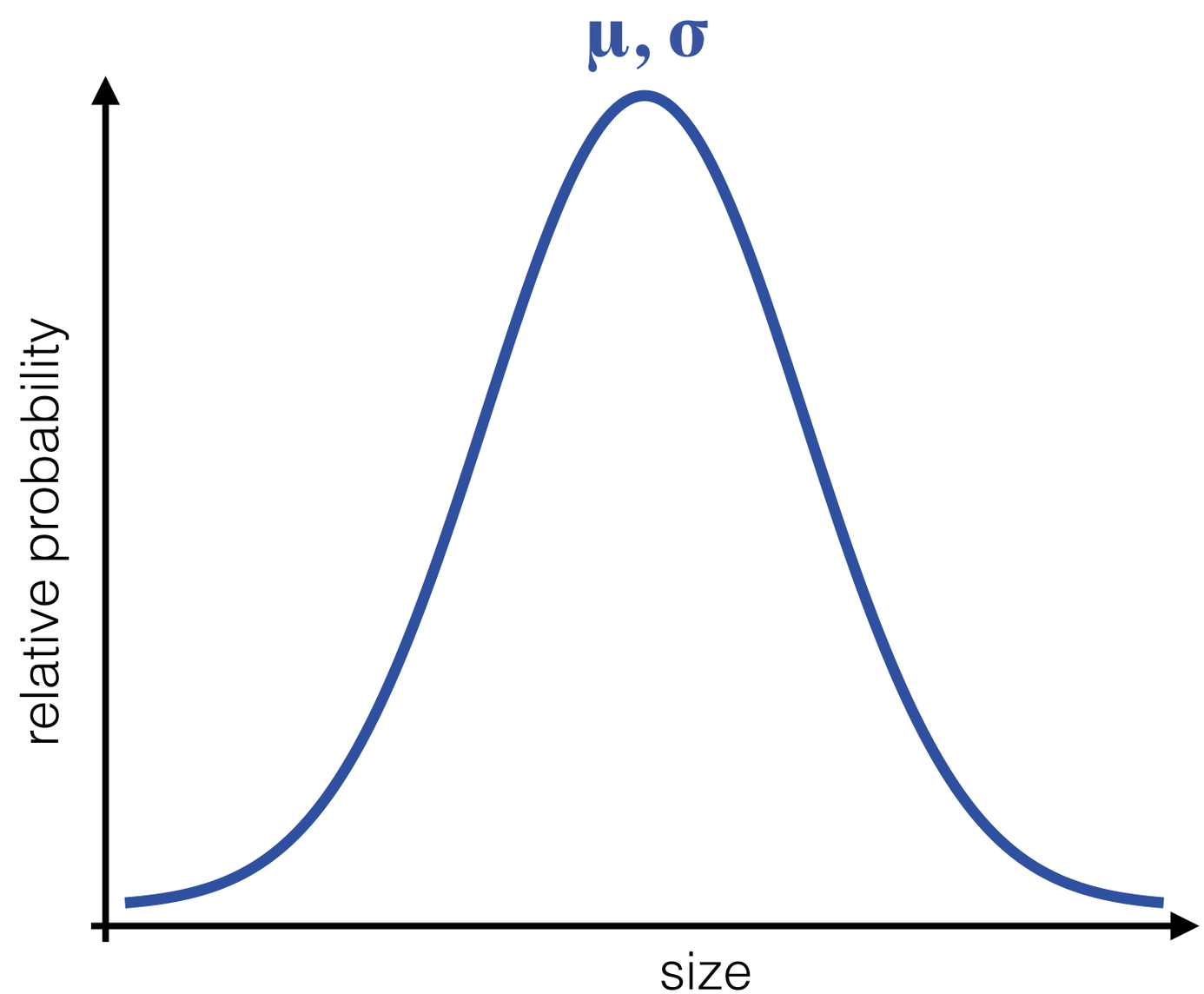
Marginal likelihood

Warning: always avoid emojis in equations, if you can.

Calculating the likelihood: $P(🐼|\mu, \sigma, \mathcal{N})$

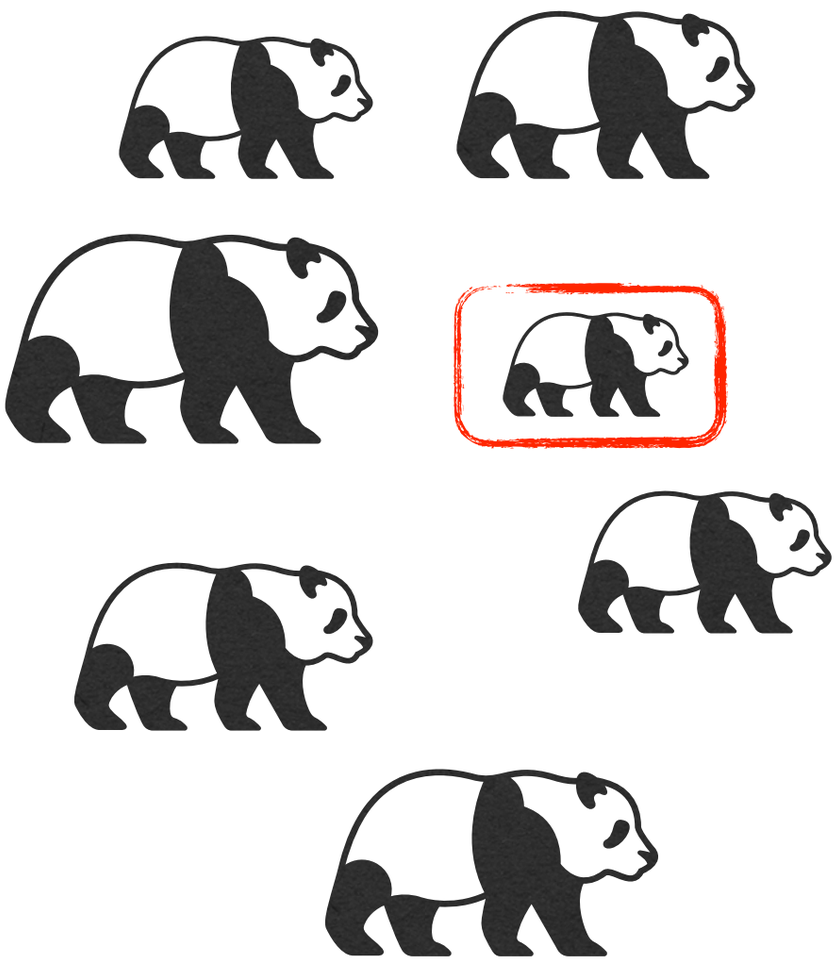
Likelihood: probability of observing the data you observe given a model and its parameters

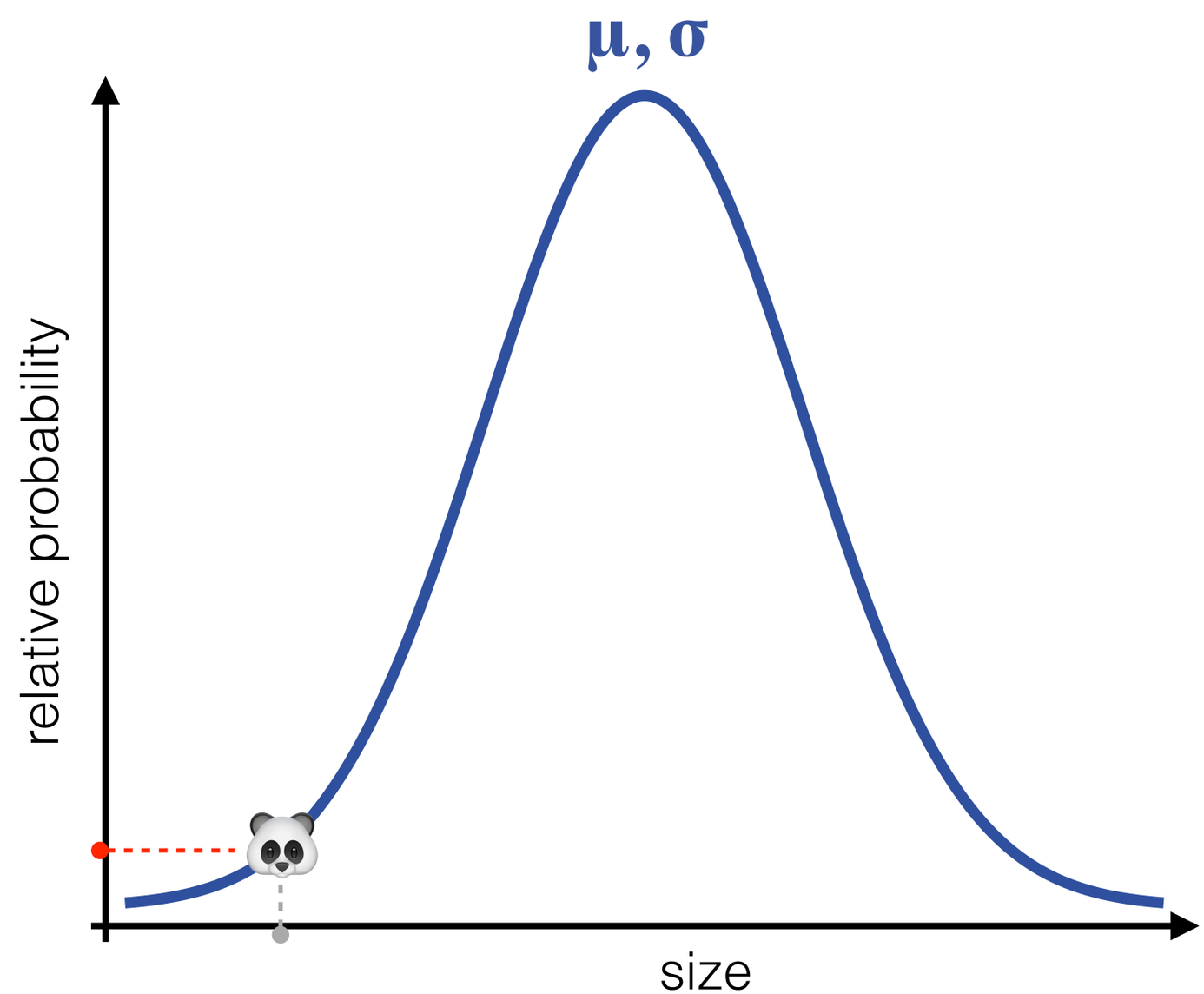




Calculating the likelihood: $P(🐼|\mu, \sigma, \mathcal{N})$

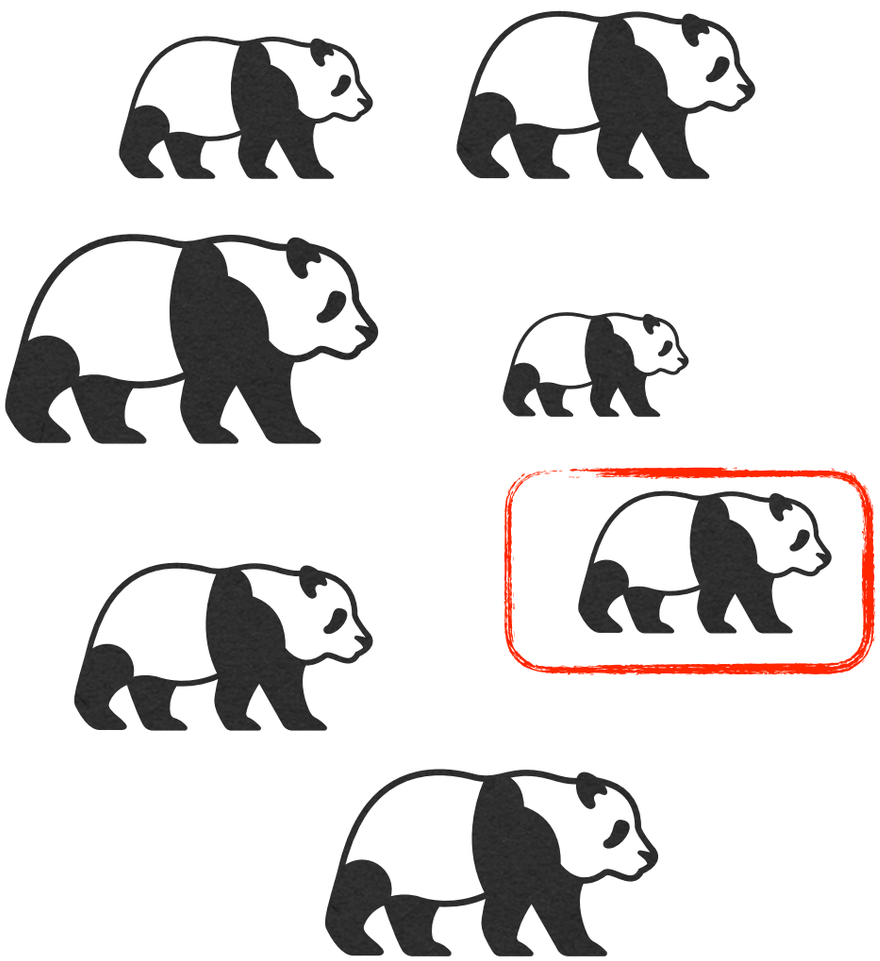
Likelihood: probability of observing the data you observe given a model and its parameters

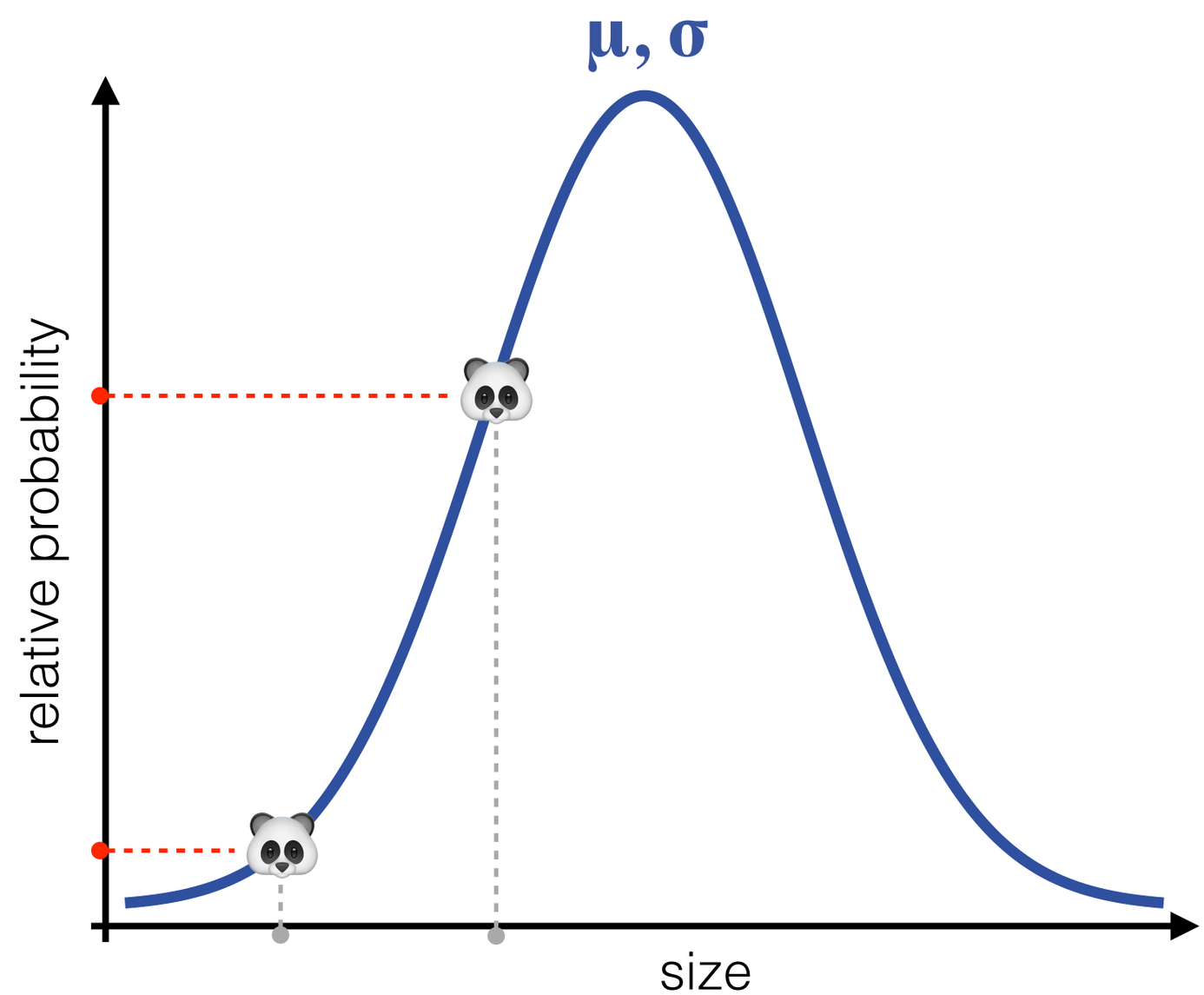




Calculating the likelihood: $P(🐼|\mu, \sigma, \mathcal{N})$

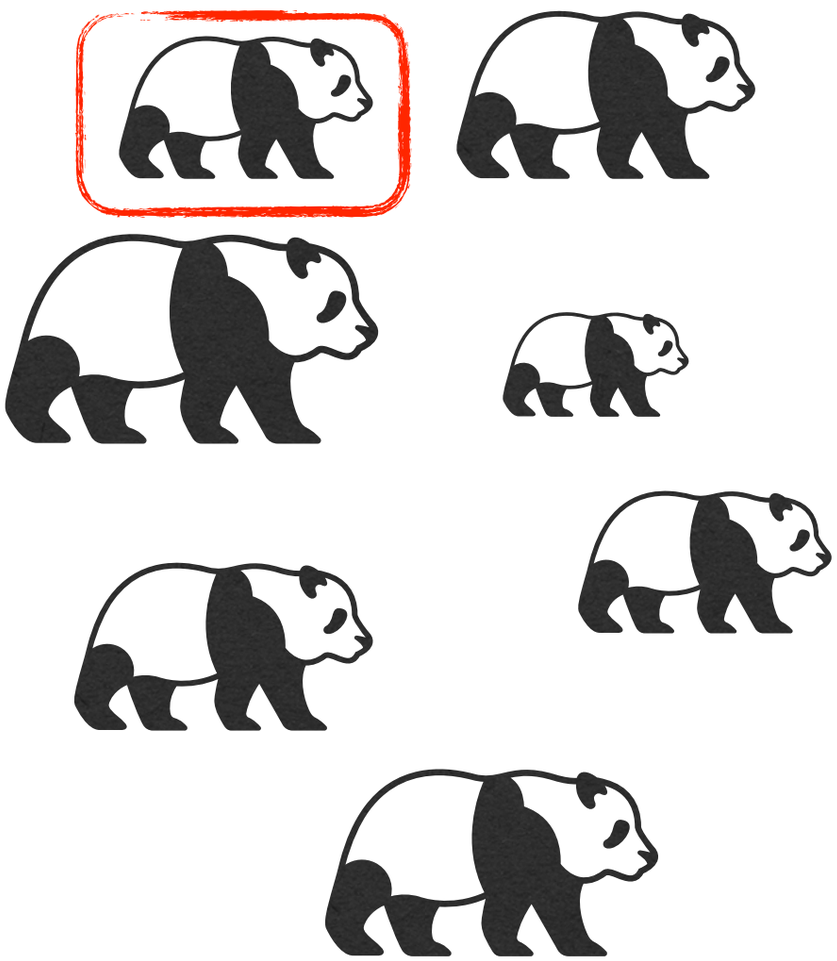
Likelihood: probability of observing the data you observe given a model and its parameters

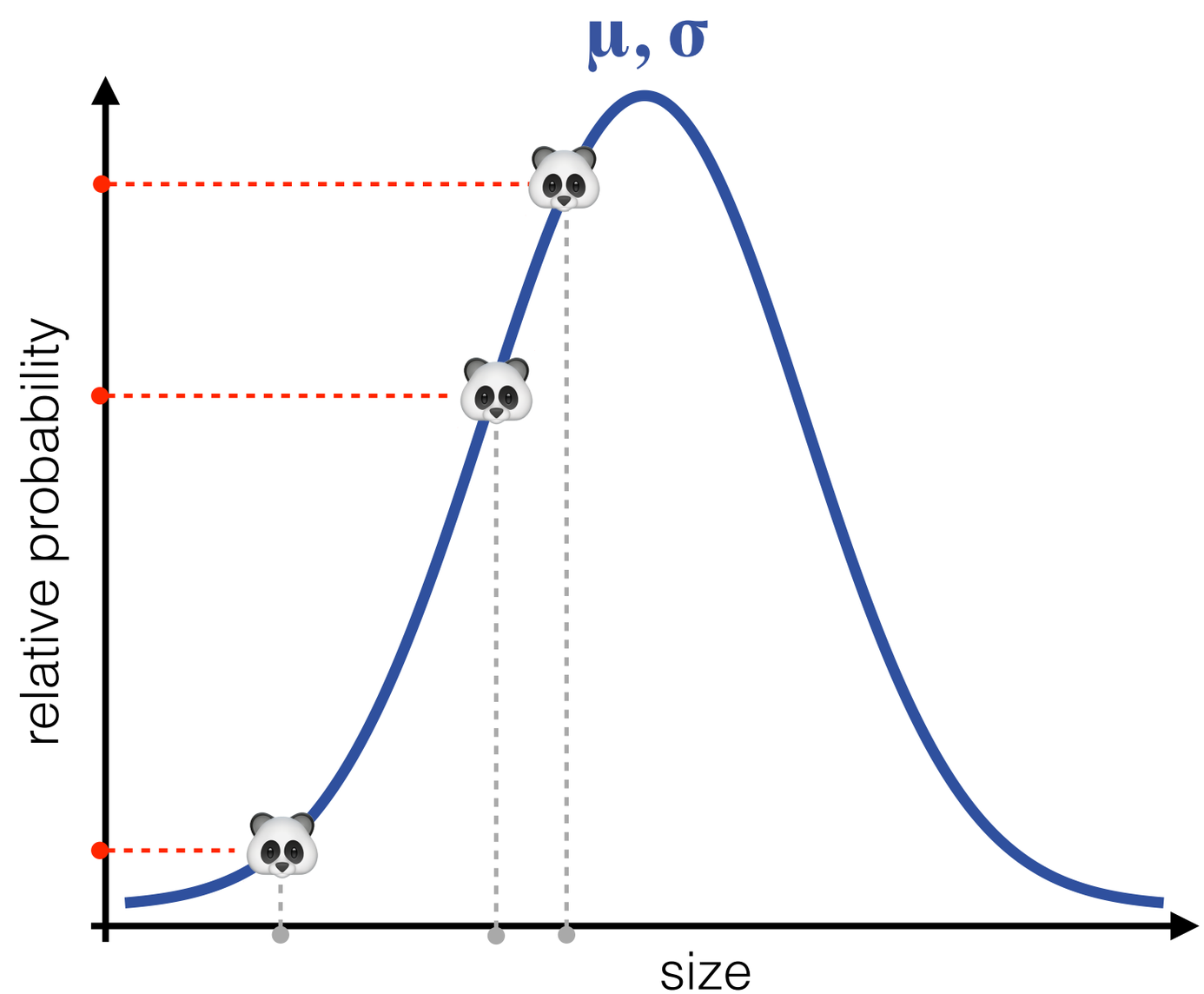




Calculating the likelihood: $P(🐼|\mu, \sigma, \mathcal{N})$

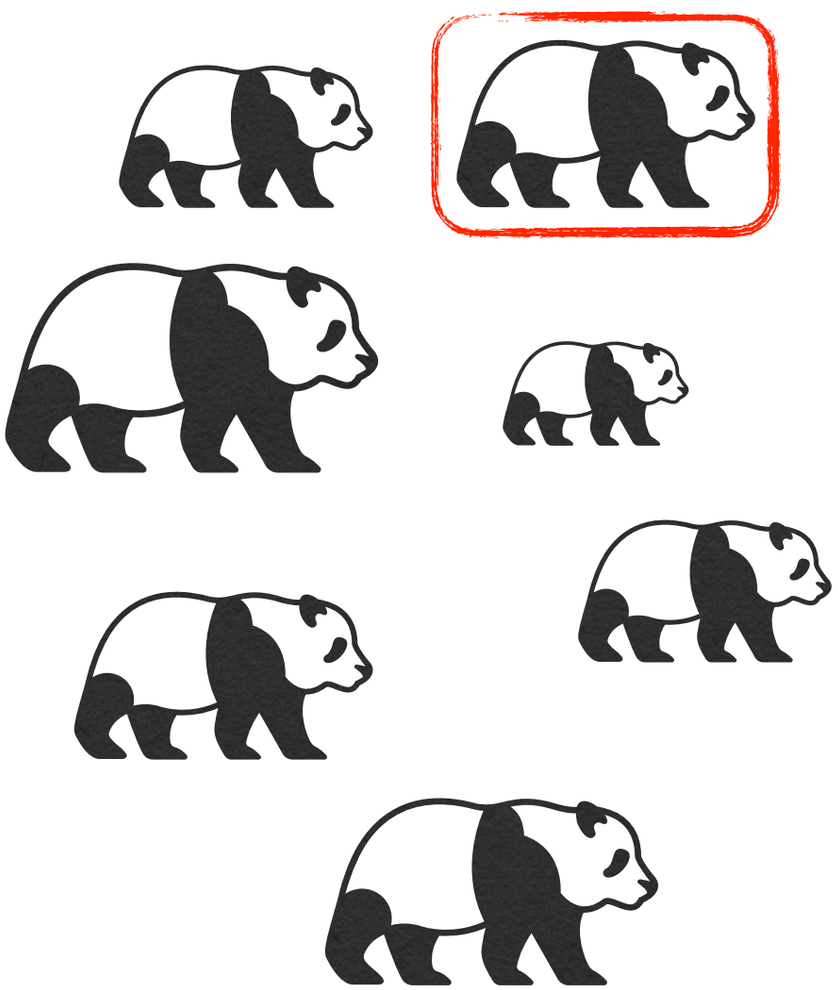
Likelihood: probability of observing the data you observe given a model and its parameters

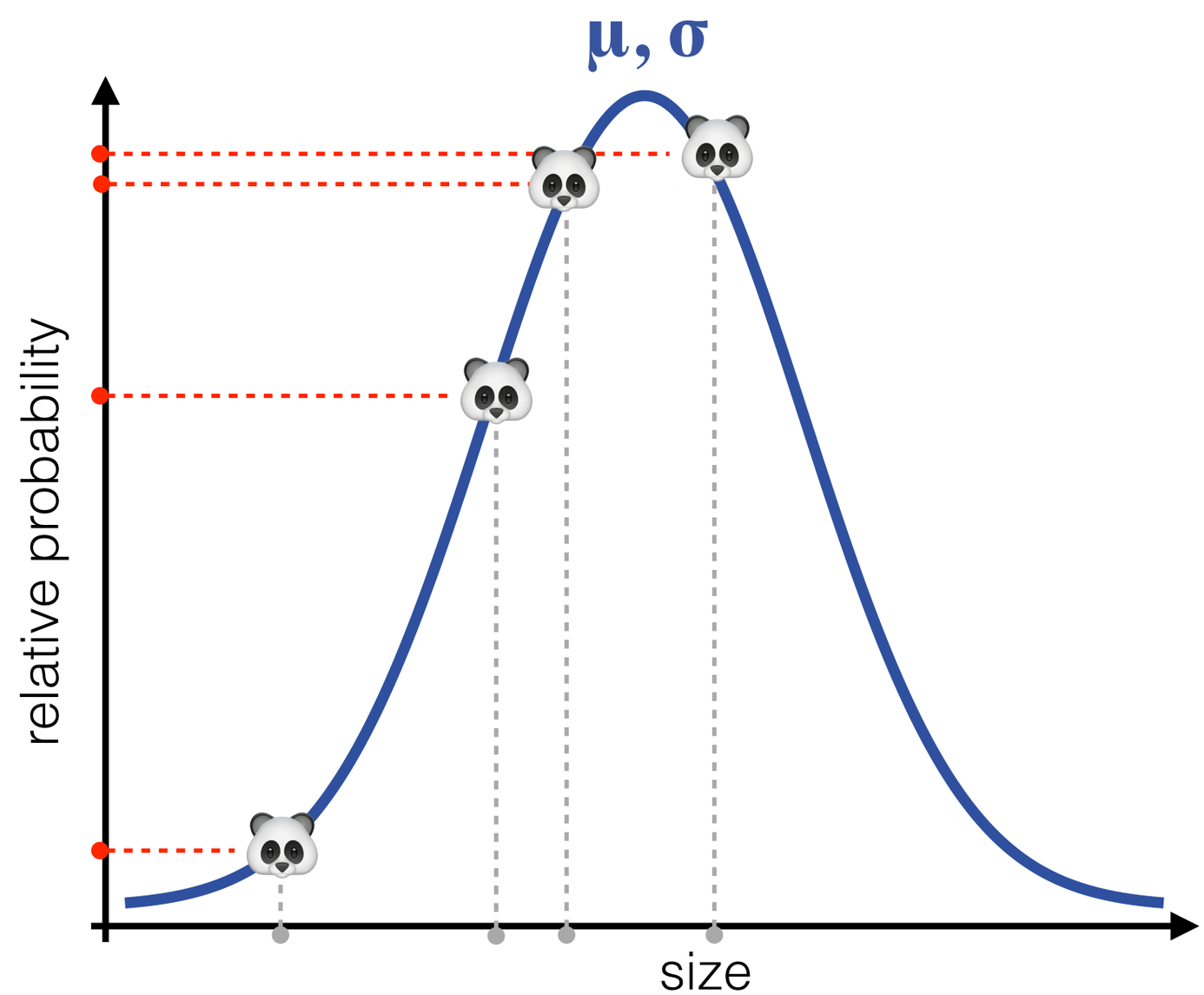




Calculating the likelihood: $P(🐼|\mu, \sigma, \mathcal{N})$

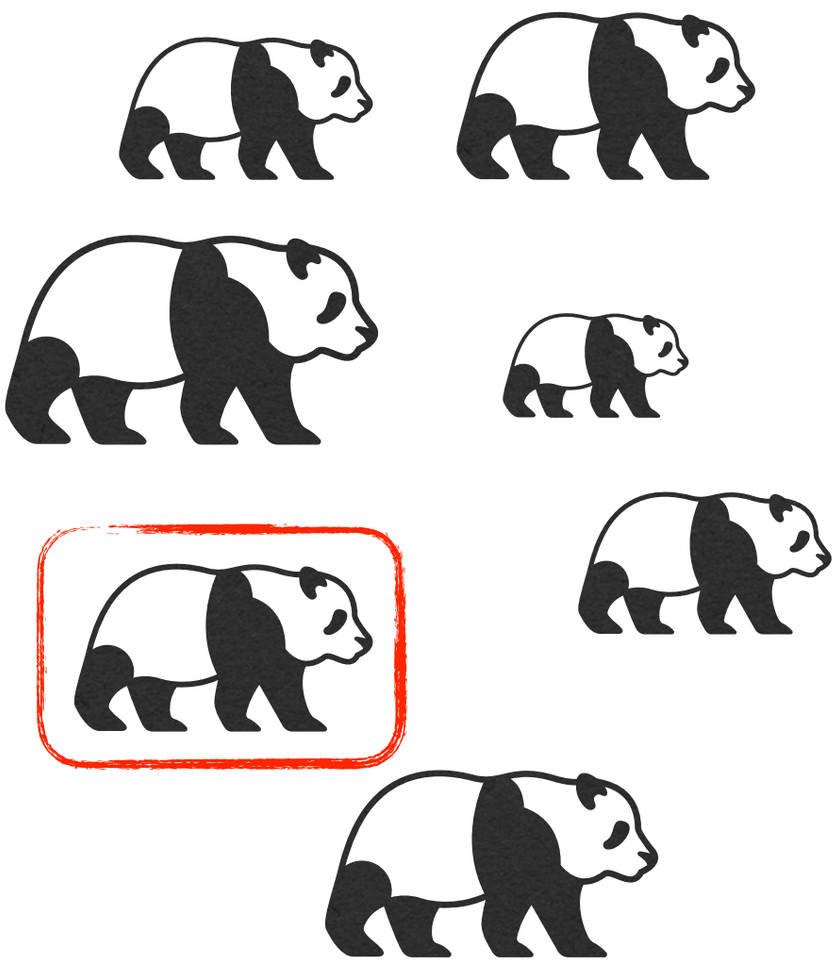
Likelihood: probability of observing the data you observe given a model and its parameters

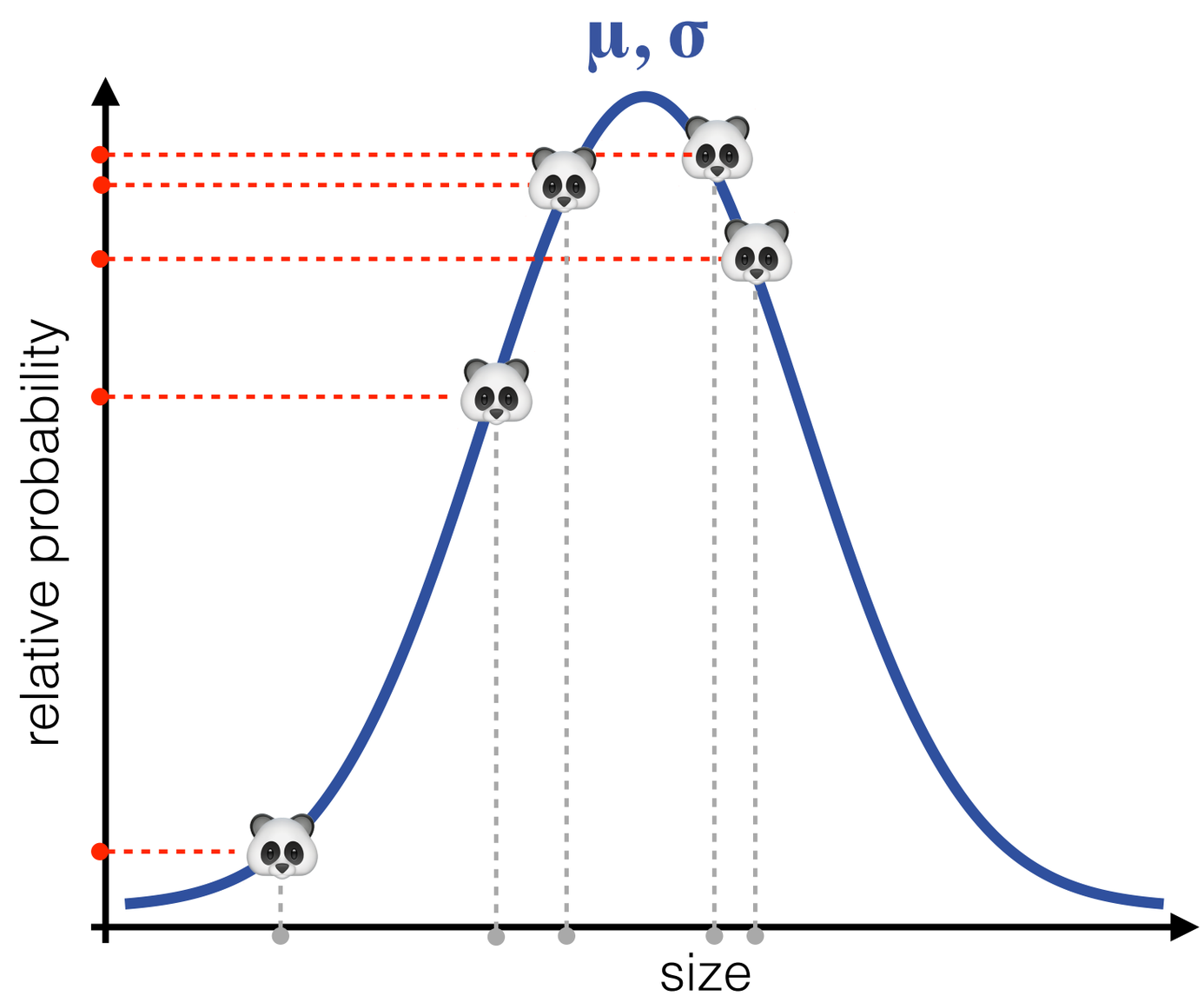




Calculating the likelihood: $P(🐼|\mu, \sigma, \mathcal{N})$

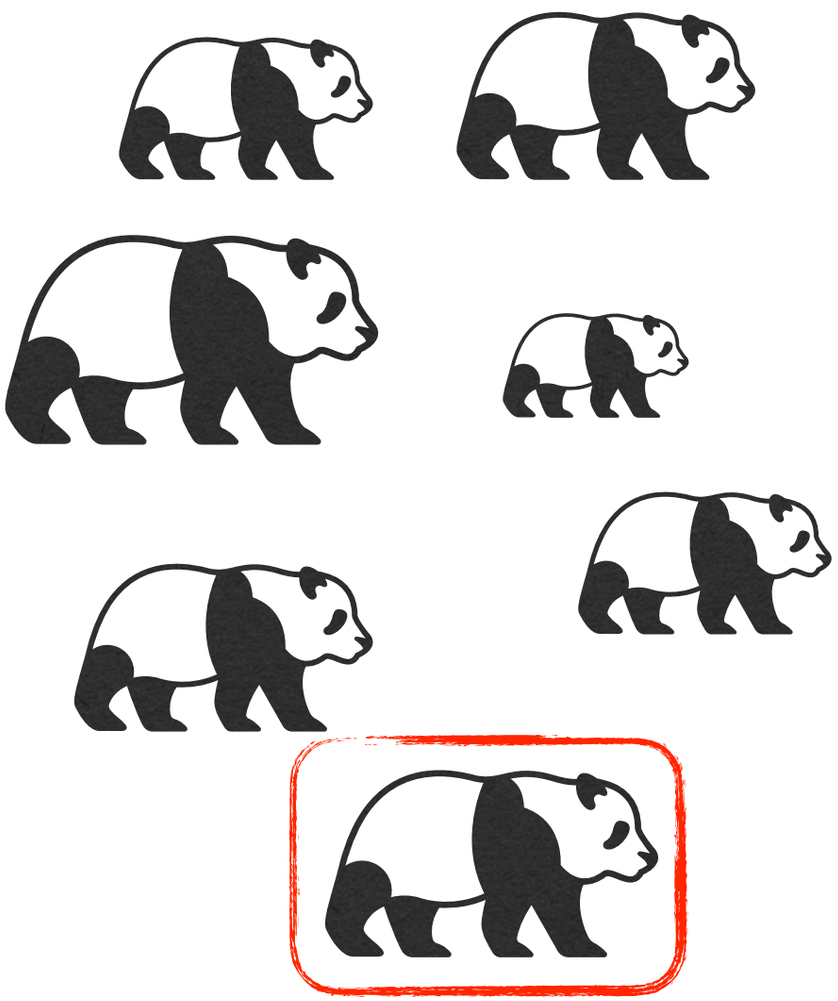
Likelihood: probability of observing the data you observe given a model and its parameters

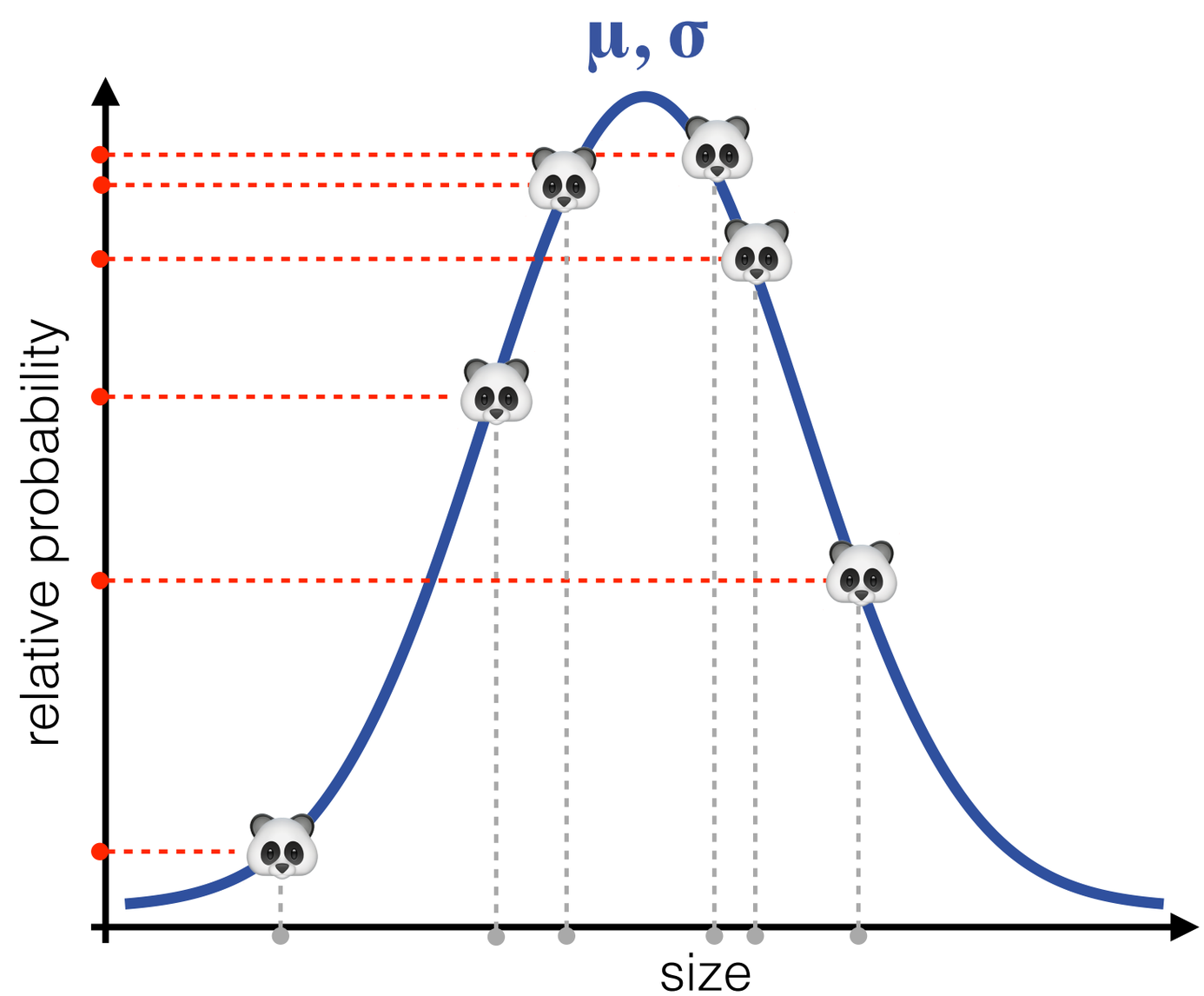




Calculating the likelihood: $P(🐼|\mu, \sigma, \mathcal{N})$

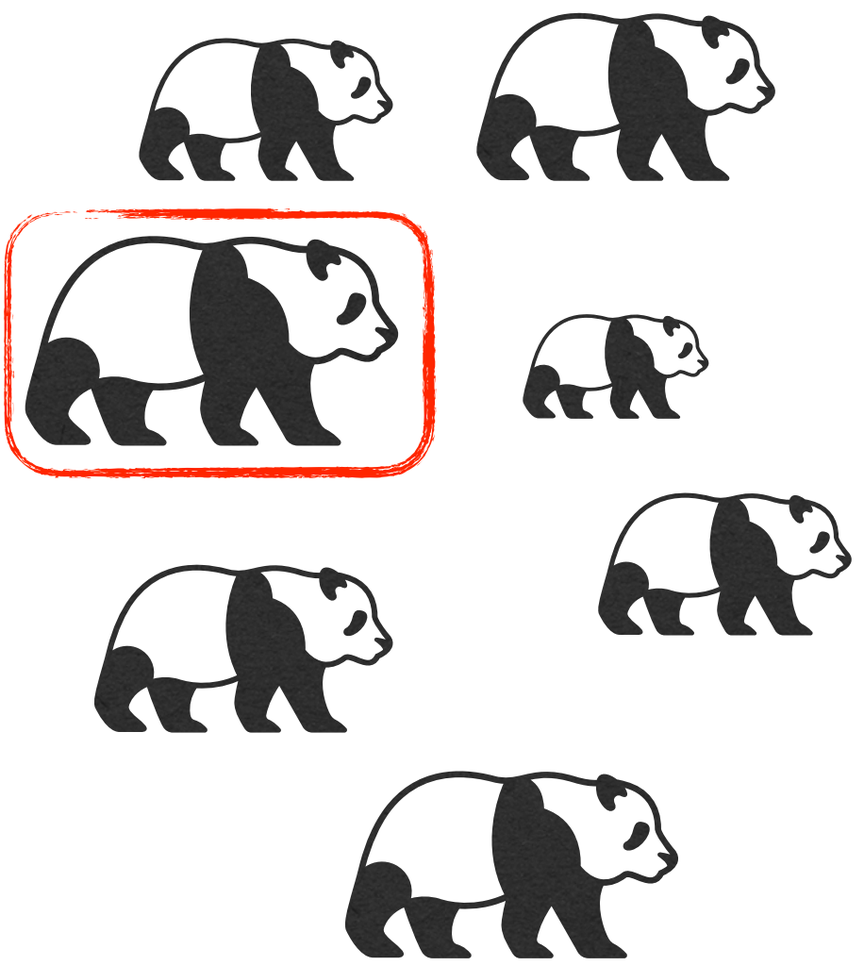
Likelihood: probability of observing the data you observe given a model and its parameters

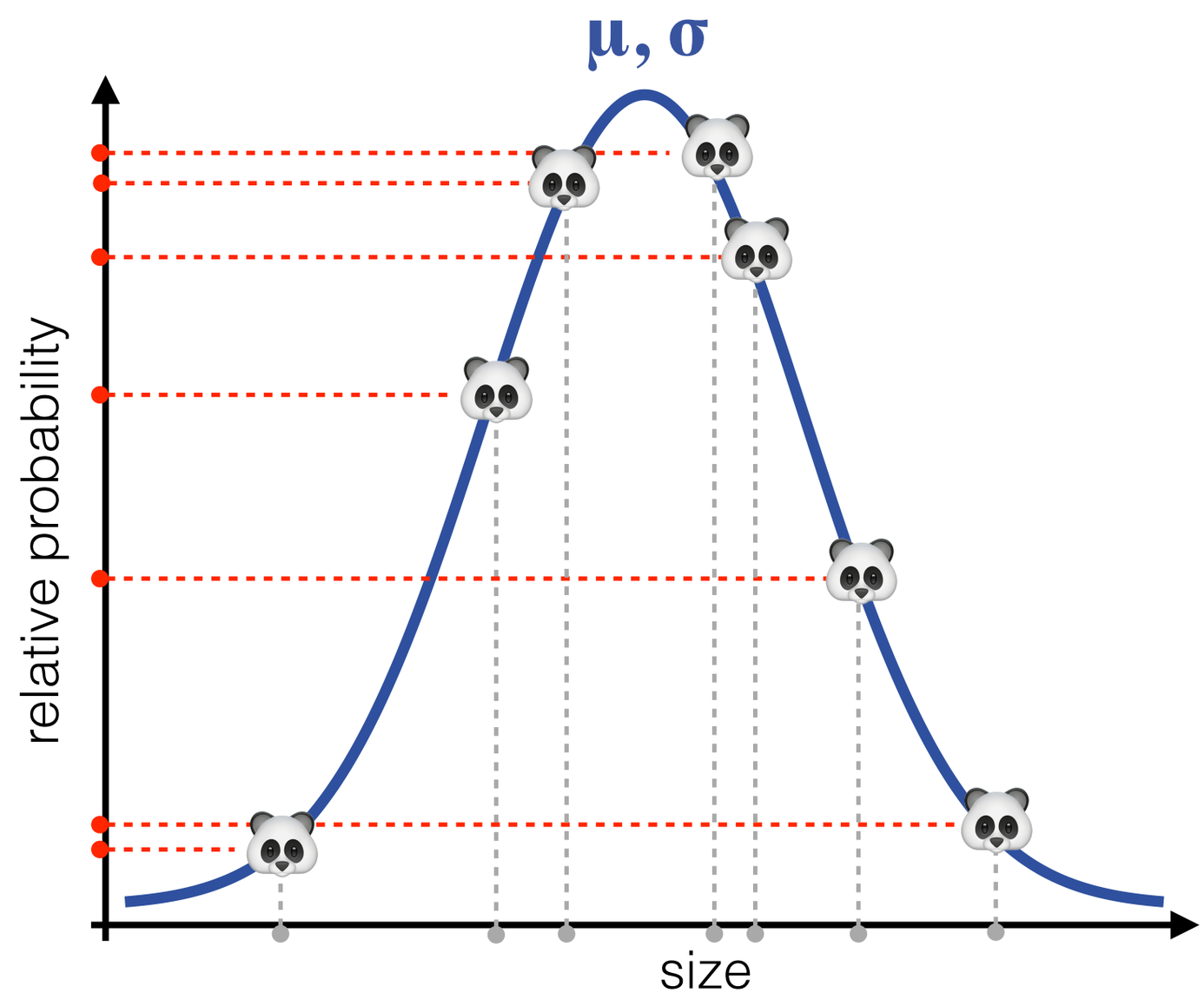




Calculating the likelihood: $P(🐼|\mu, \sigma, \mathcal{N})$

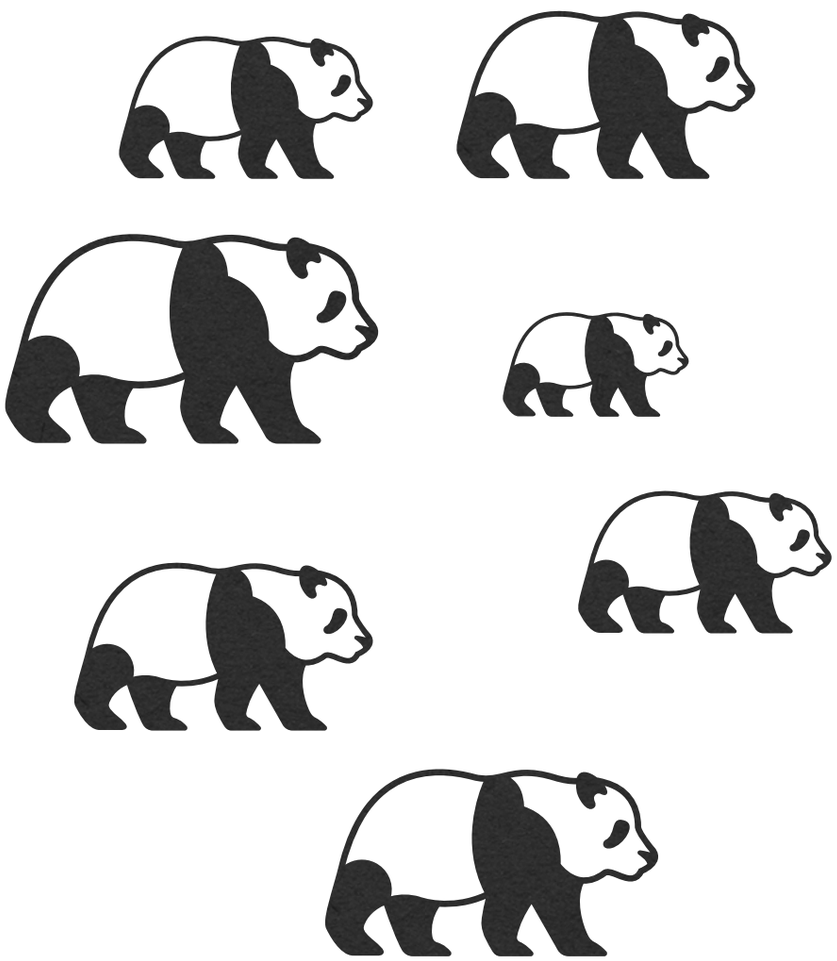
Likelihood: probability of observing the data you observe given a model and its parameters

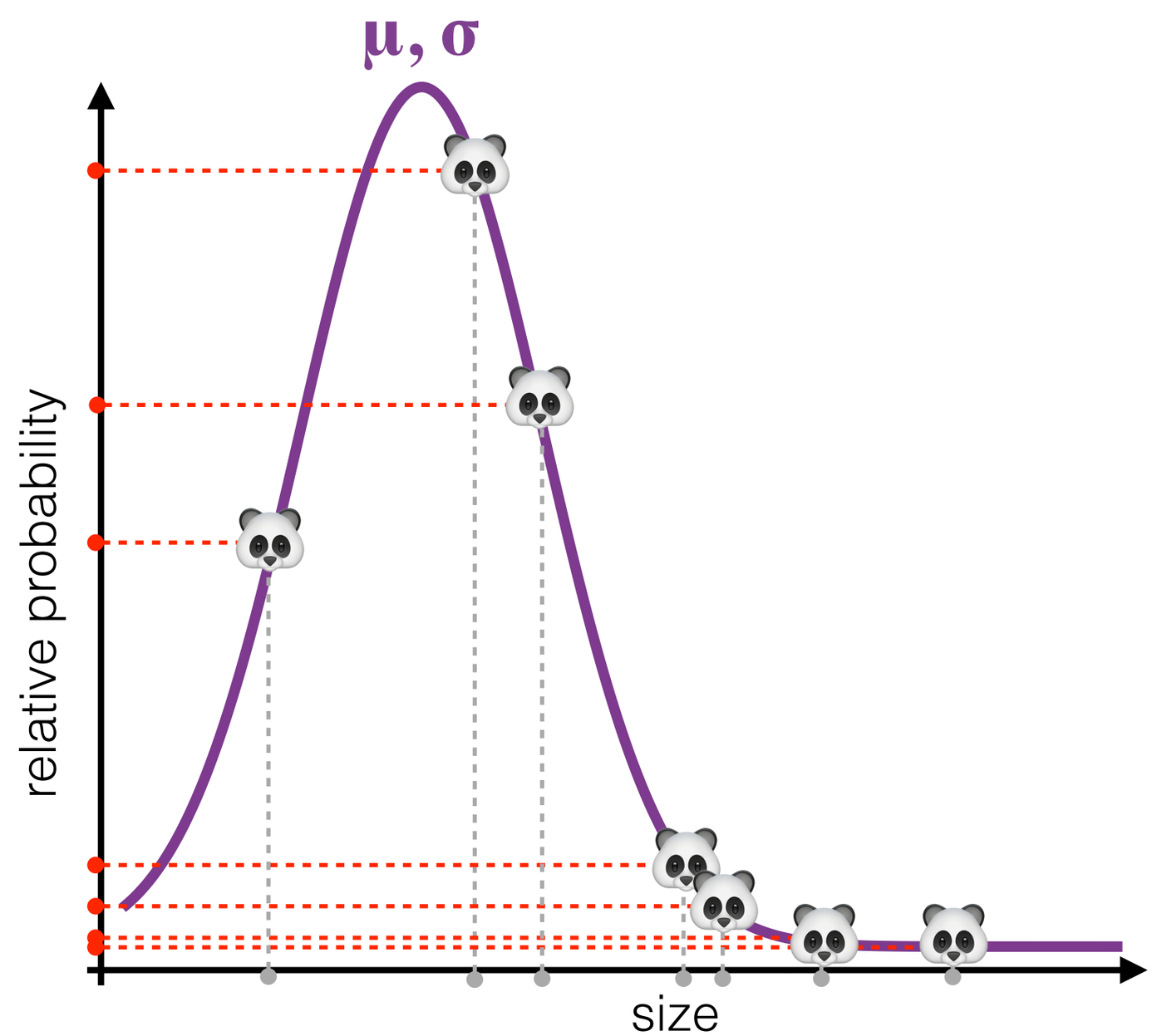




Calculating the likelihood: $P(🐼|\mu, \sigma, \mathcal{N})$

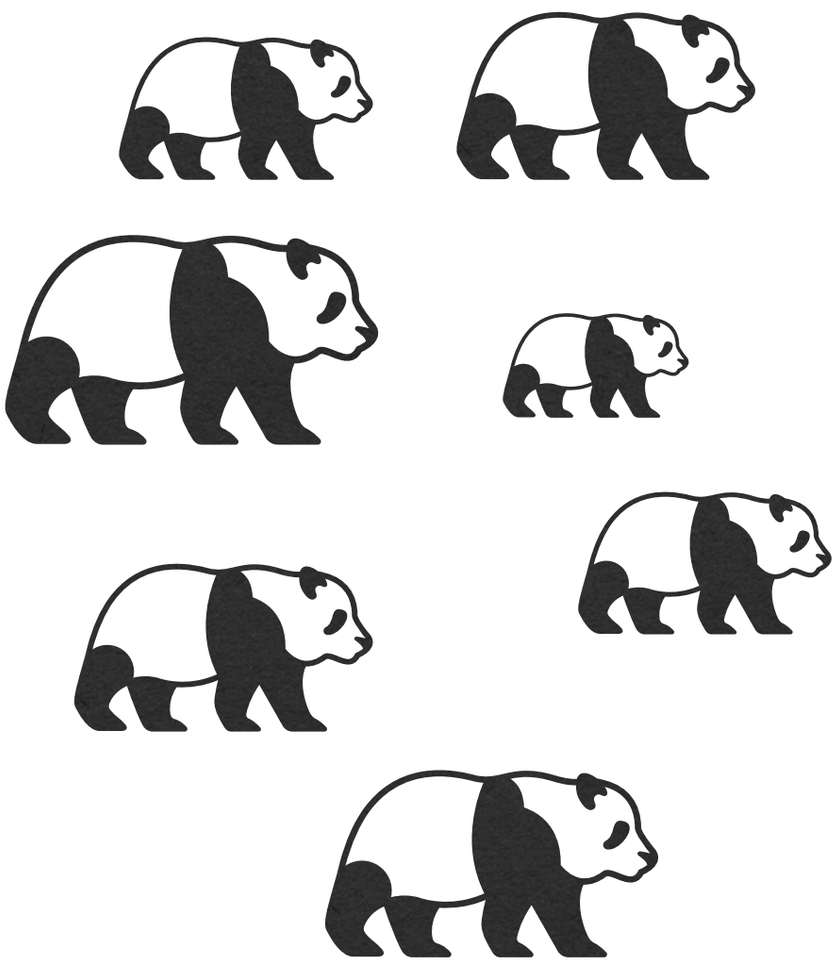
Likelihood: probability of observing the data you observe given a model and its parameters



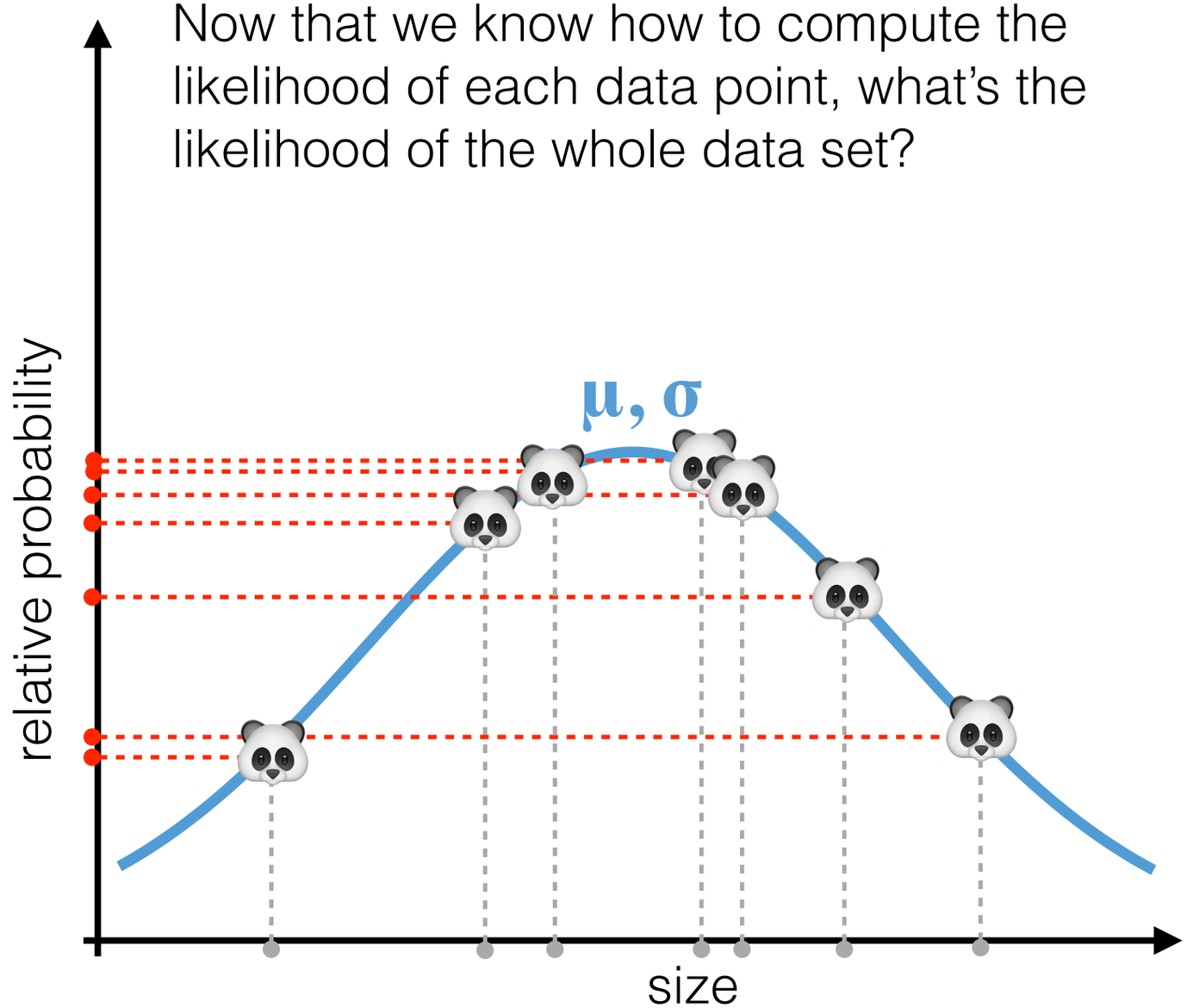


Calculating the likelihood: $P(🐼|\mu, \sigma, \mathcal{N})$

Likelihood: probability of observing the data you observe given a model and its parameters

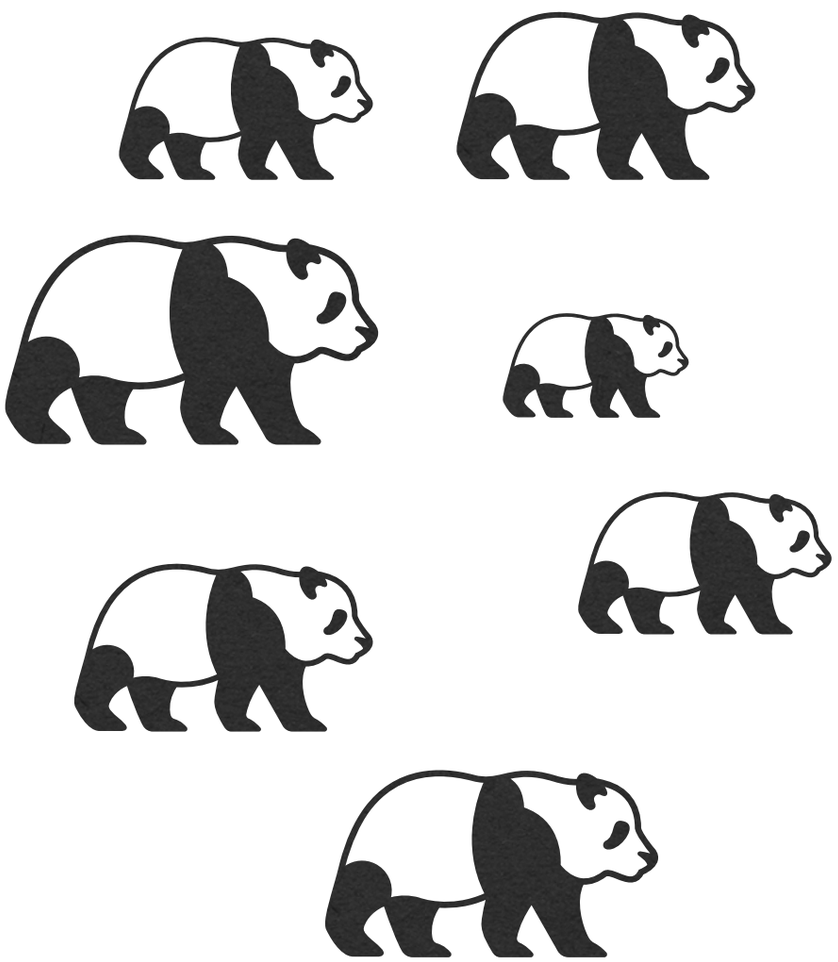


Now that we know how to compute the likelihood of each data point, what's the likelihood of the whole data set?

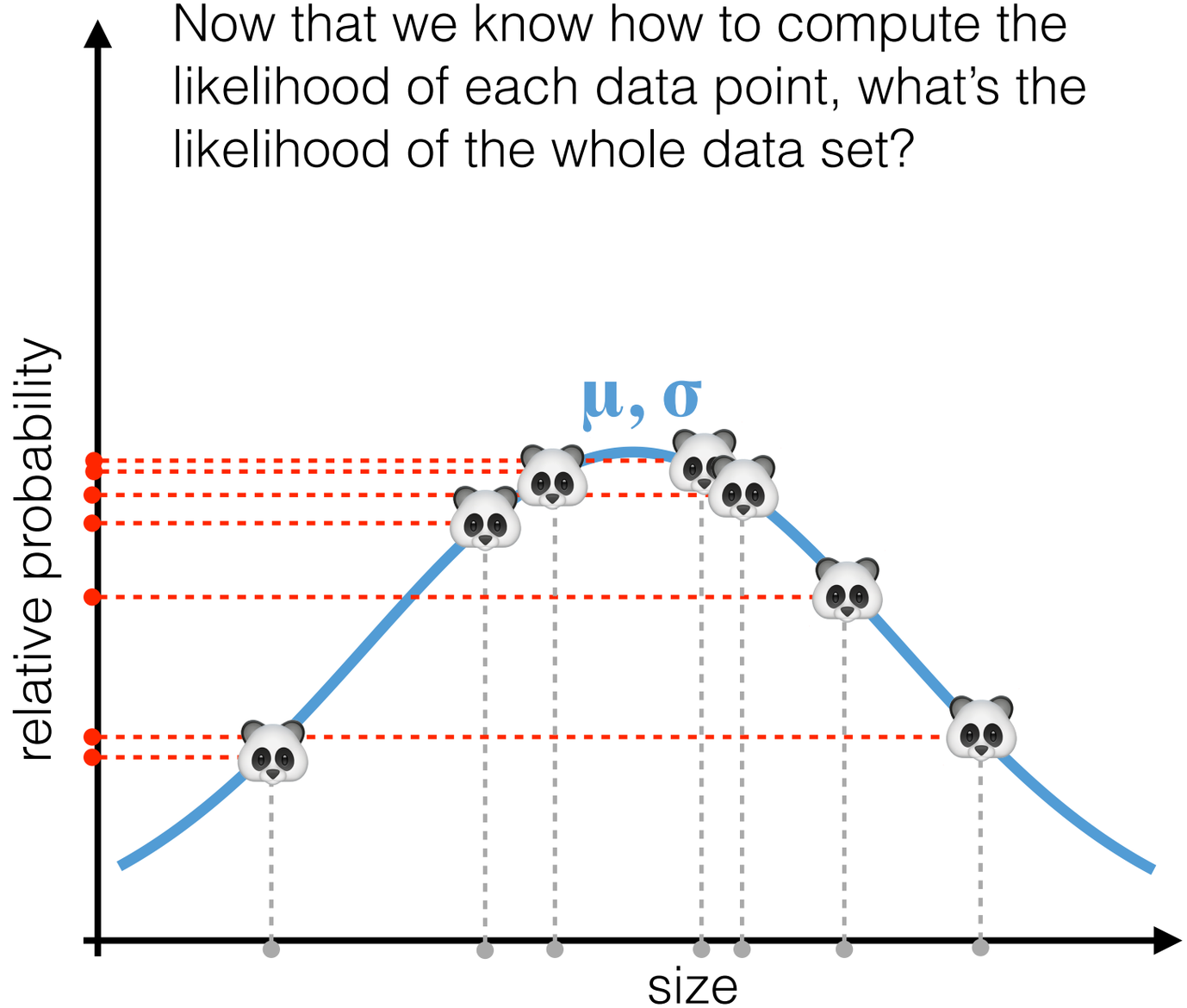


Calculating the likelihood: $P(🐼|\mu, \sigma, \mathcal{N})$

$$P(🐼_1, \ldots, 🐼_7|\mu, \sigma, \mathcal{N}) = \prod_{i=1}^{7} P(🐼_i|\mu, \sigma, \mathcal{N})$$



Now that we know how to compute the likelihood of each data point, what's the likelihood of the whole data set?



Bayes theorem

Likelihood Prior probability

Posterior probability

$$P(\mu, \sigma | 🐼, \mathcal{N}) = \frac{P(🐼 | \mu, \sigma, \mathcal{N}) \times P(\mu, \sigma)}{P(🐼 | \mathcal{N})}$$

Marginal likelihood

Likelihood: probability of observing the data you observe given a model and its parameters

Prior: probability of the parameter values based on some prior knowledge about them

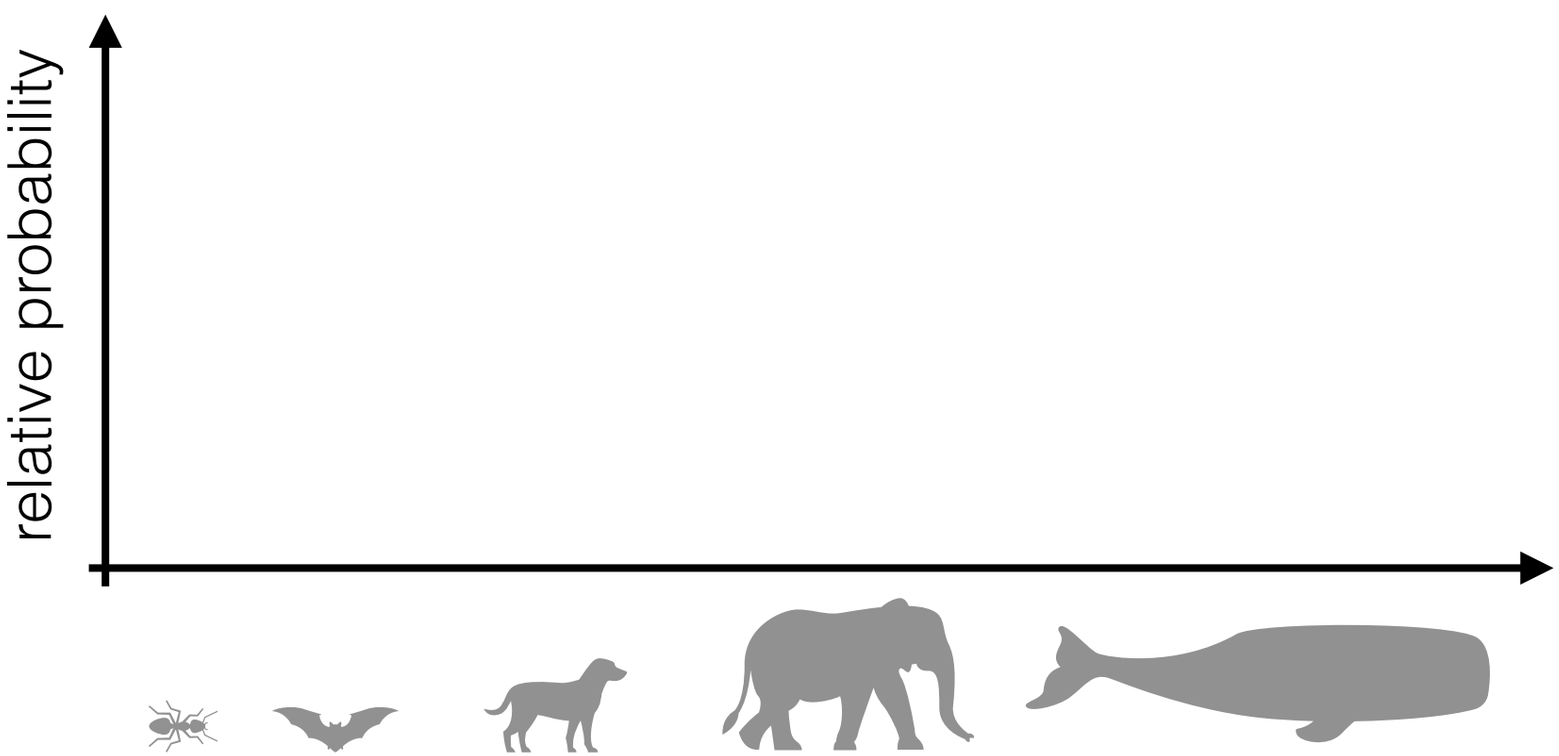
Marginal likelihood: probability of observing the data you observe given a model and all possible parameter values

Calculating the prior: $P(\mu, \sigma)$

Prior: probability of the estimated parameter values μ and σ

$$P(\mu) = ?$$

What do we know a priori about
the parameter μ ?

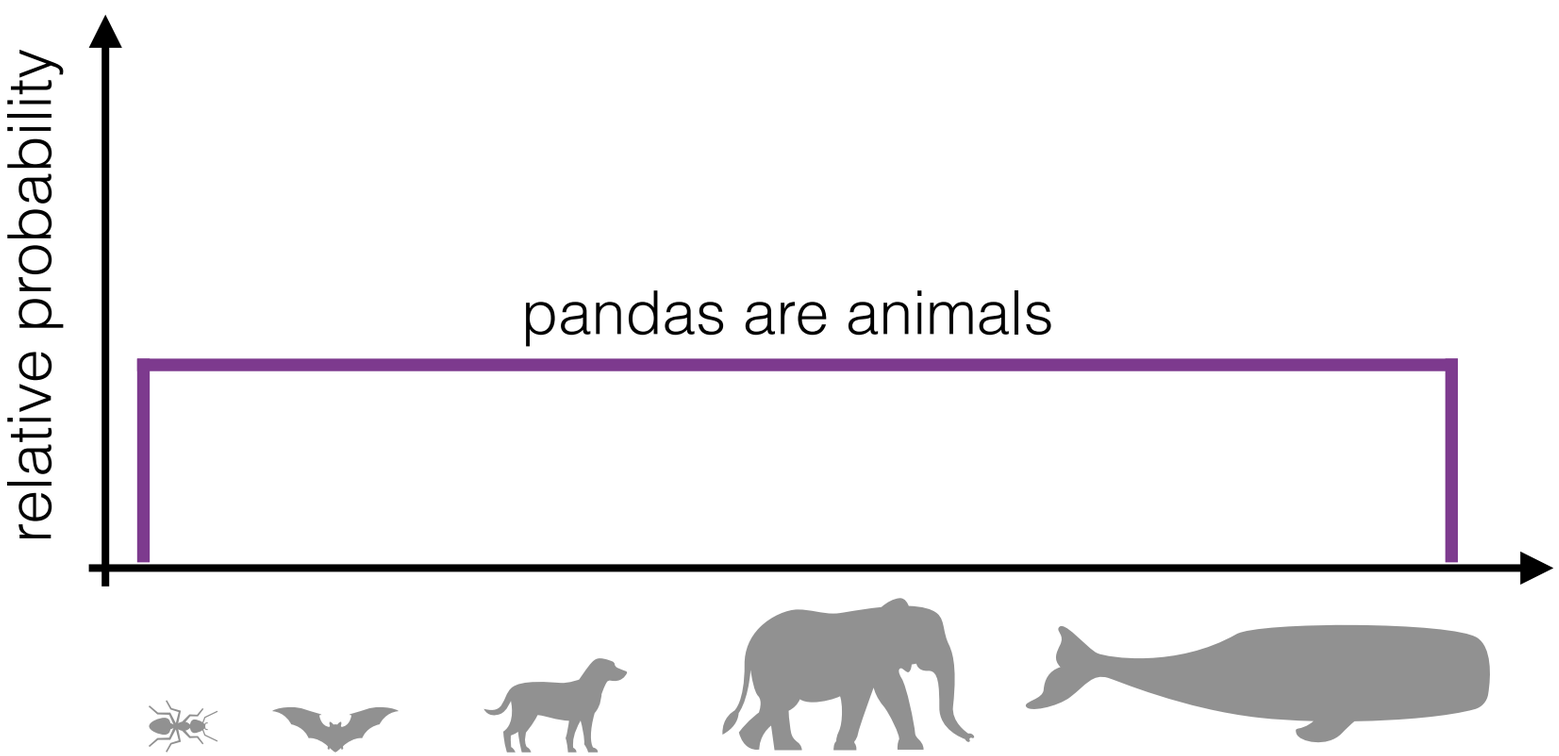


Calculating the prior: $P(\mu, \sigma)$

Prior: probability of the estimated parameter values μ and σ

$$P(\mu) \sim \mathcal{U}(🐜, 🐳)$$

What do we know a priori about
the parameter μ ?

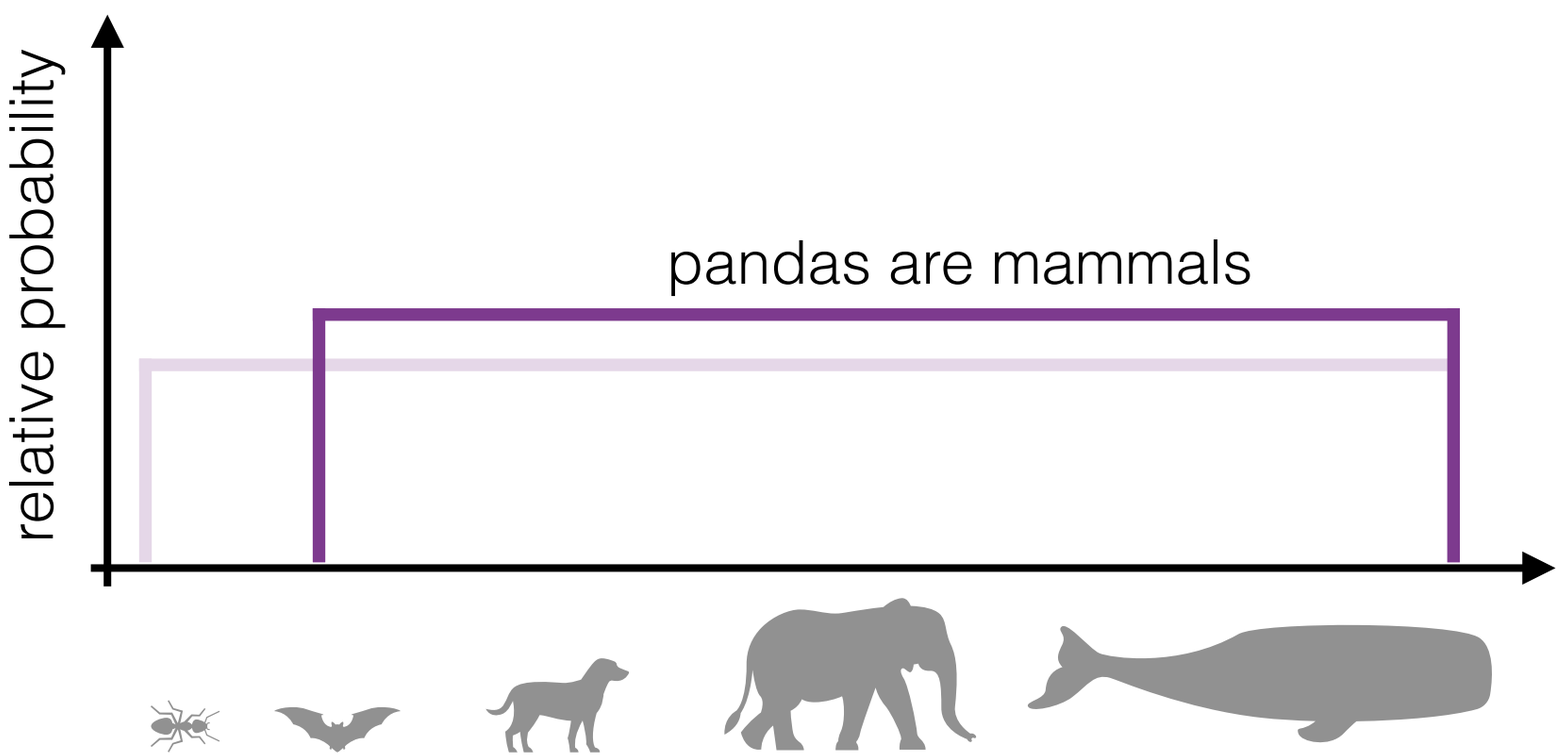


Calculating the prior: $P(\mu, \sigma)$

Prior: probability of the estimated parameter values μ and σ

$$P(\mu) \sim \mathcal{U}(🦇, 🐳)$$

What do we know a priori about
the parameter μ ?

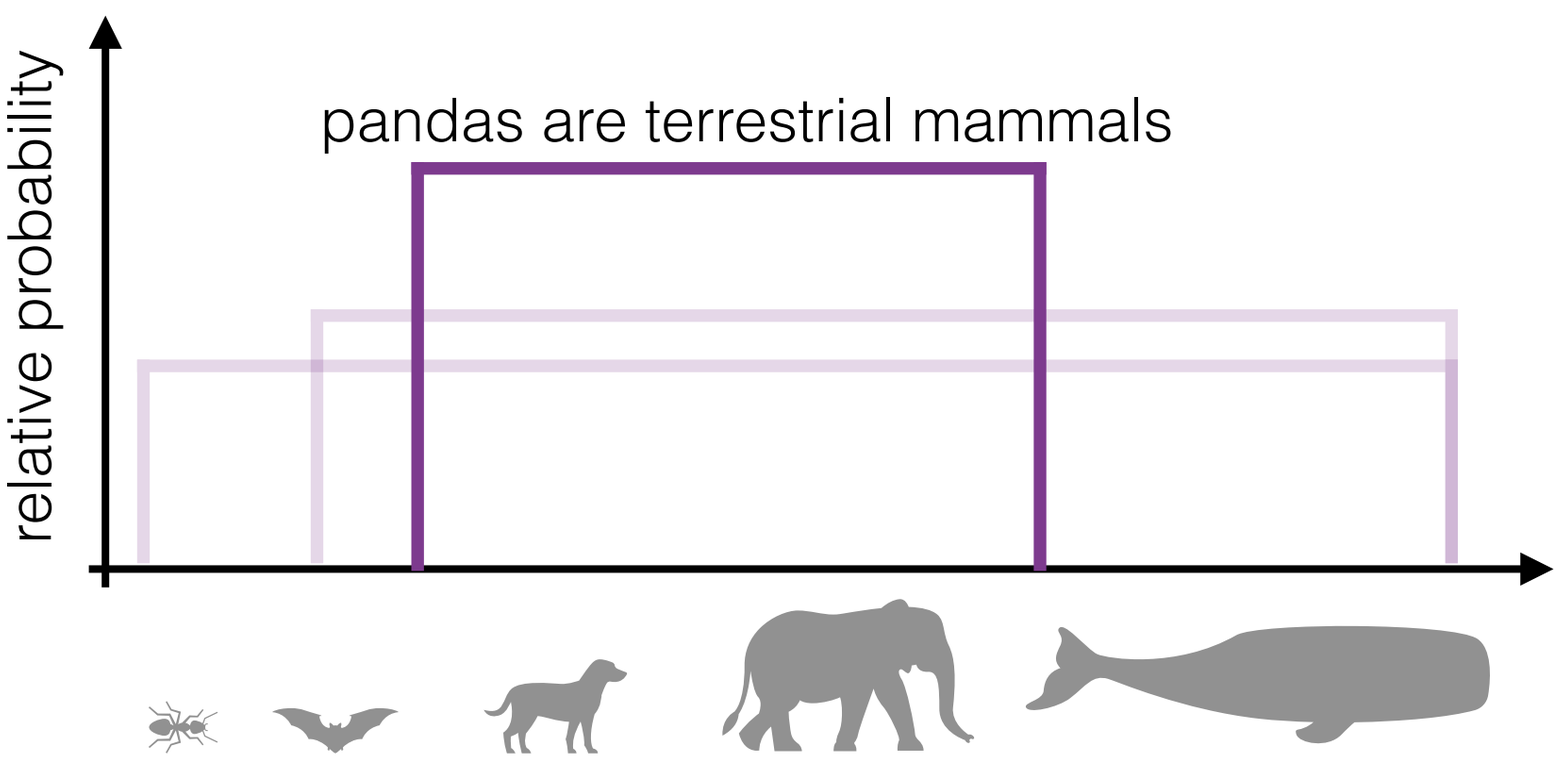


Calculating the prior: $P(\mu, \sigma)$

Prior: probability of the estimated parameter values μ and σ

$$P(\mu) \sim \mathcal{U}(🐭, 🐘)$$

What do we know a priori about
the parameter μ ?



Calculating the prior: $P(\mu, \sigma)$

Prior: probability of the estimated parameter values μ and σ

$$P(\mu) \sim \mathcal{U}(🐭, 🐘)$$

$$P(\sigma) = ?$$

What do we know a priori about the parameter σ ?

Calculating the prior: $P(\mu, \sigma)$

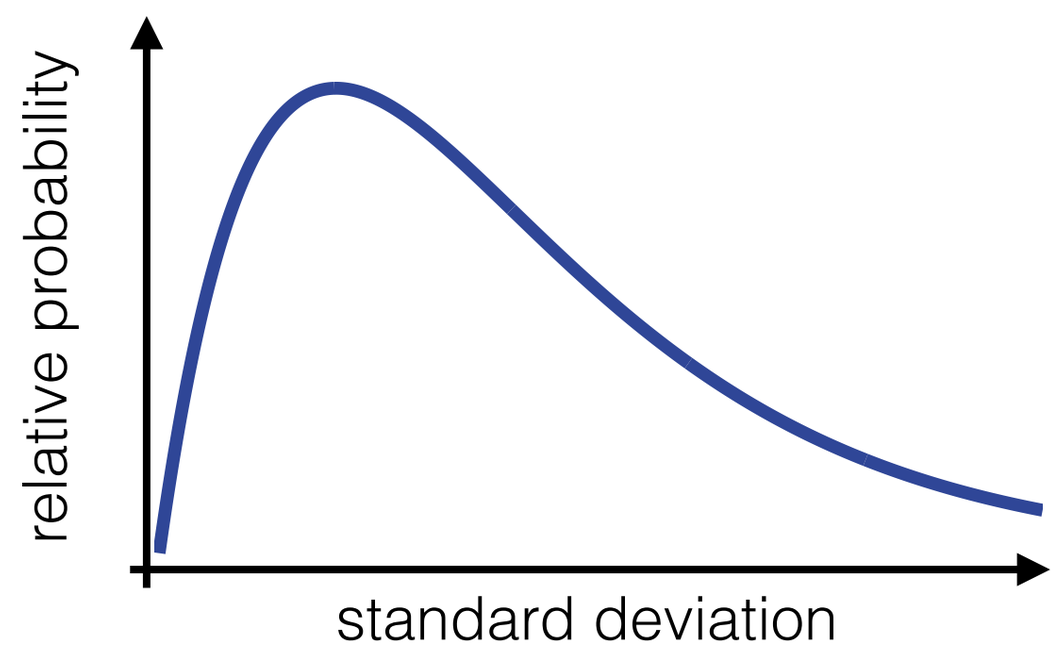
Prior: probability of the estimated parameter values μ and σ

$P(\mu) \sim \mathcal{U}($🐭, 🐘$)$

$P(\sigma) \sim \Gamma(\alpha, \beta)$

What do we know a priori about the parameter σ ?

it must be positive, it is unlikely to be infinite

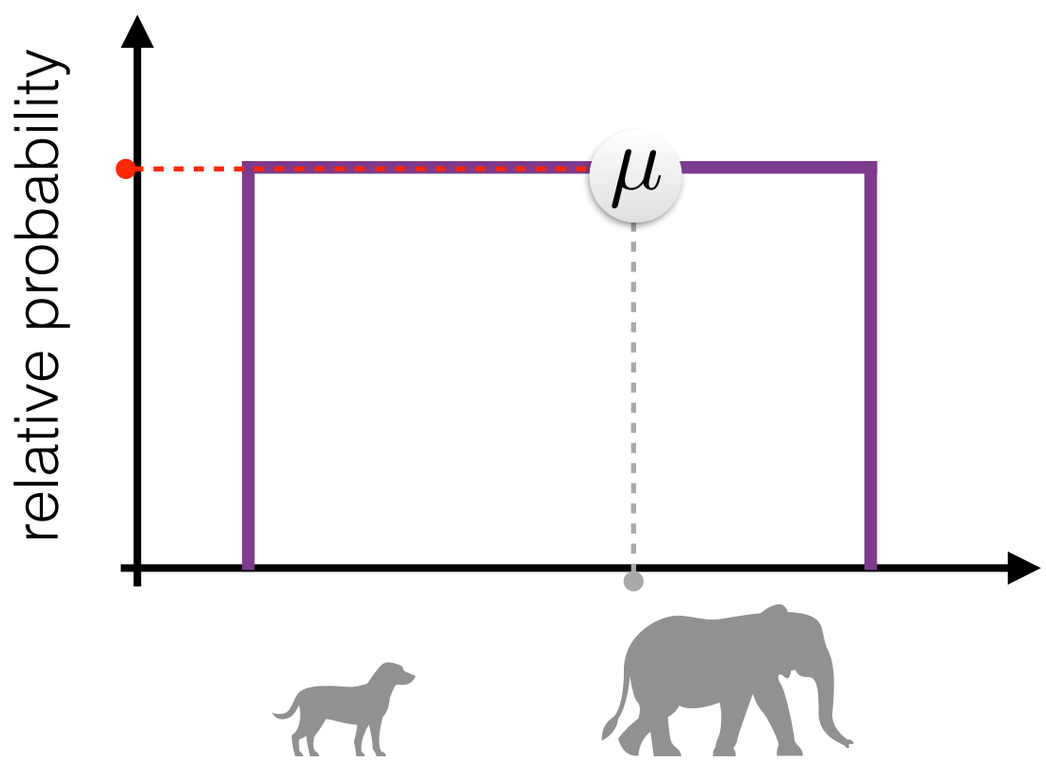


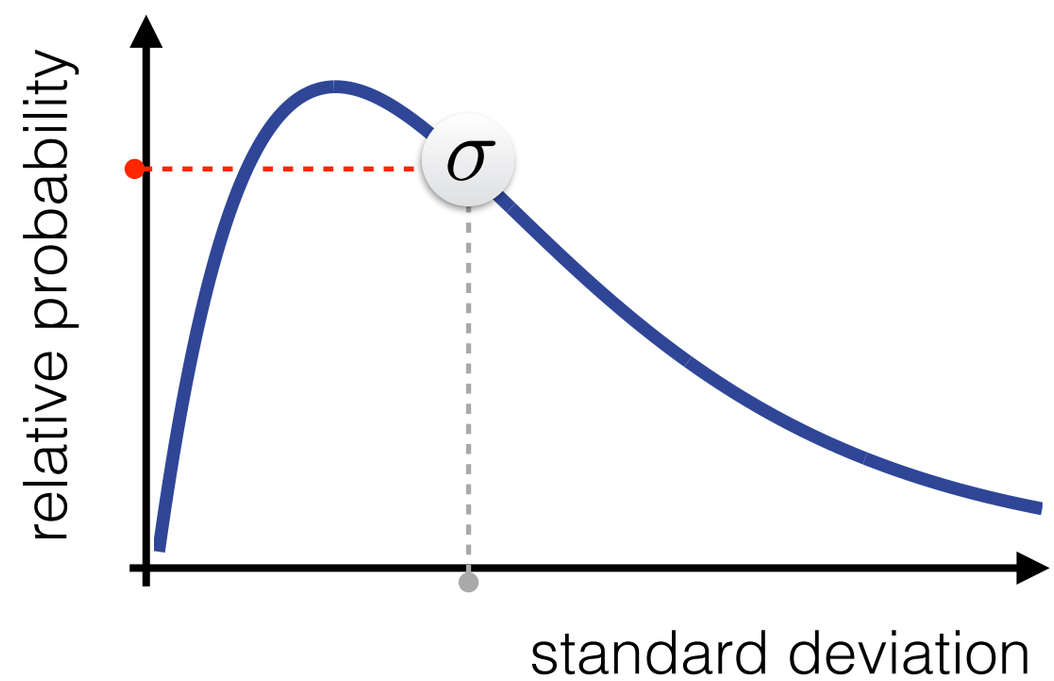
Calculating the prior: $P(\mu, \sigma)$

Prior: probability of the estimated parameter values μ and σ

$$P(\mu) \sim \mathcal{U}(🐭, 🐘) \qquad P(\sigma) \sim \Gamma(\alpha, \beta)$$

$$P(\mu, \sigma) = P(\mu | 🐭, 🐘, \mathcal{U}) \times P(\sigma | \alpha, \beta, \Gamma)$$





Bayes theorem

Likelihood Prior probability

Posterior probability

$$P(\mu, \sigma | 🐼, \mathcal{N}) = \frac{P(🐼 | \mu, \sigma, \mathcal{N}) \times P(\mu, \sigma)}{P(🐼 | \mathcal{N})}$$

Marginal likelihood

Likelihood: probability of observing the data you observe given a model and its parameters

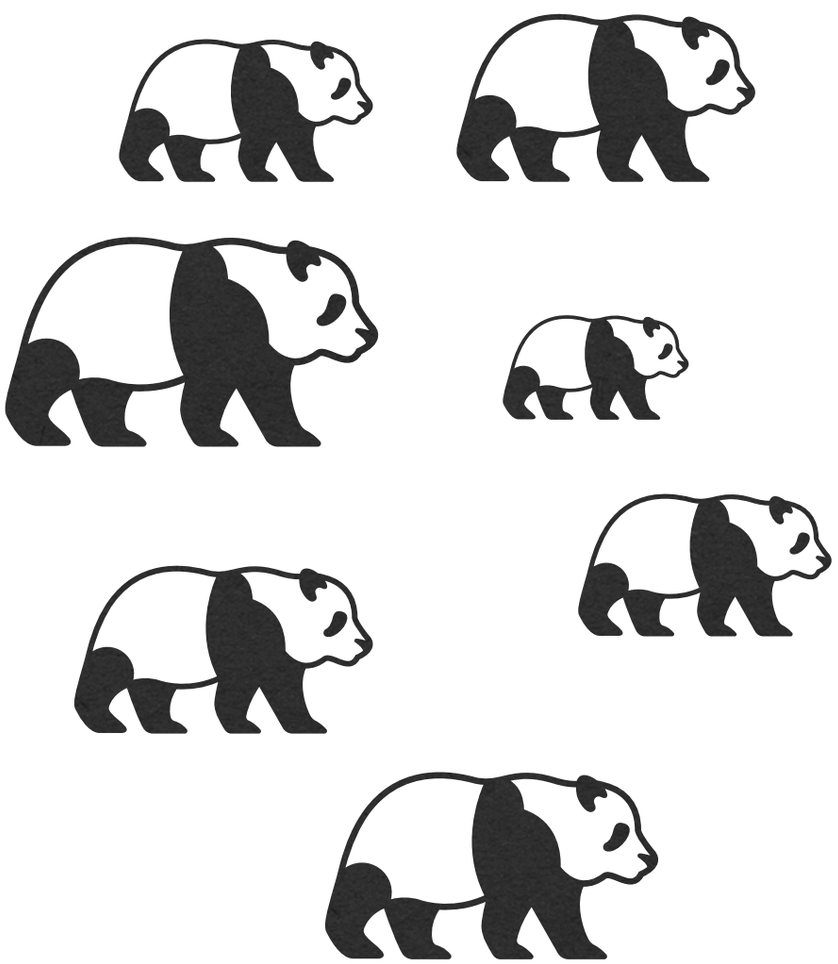
Prior: probability of the parameter values based on some prior knowledge about them

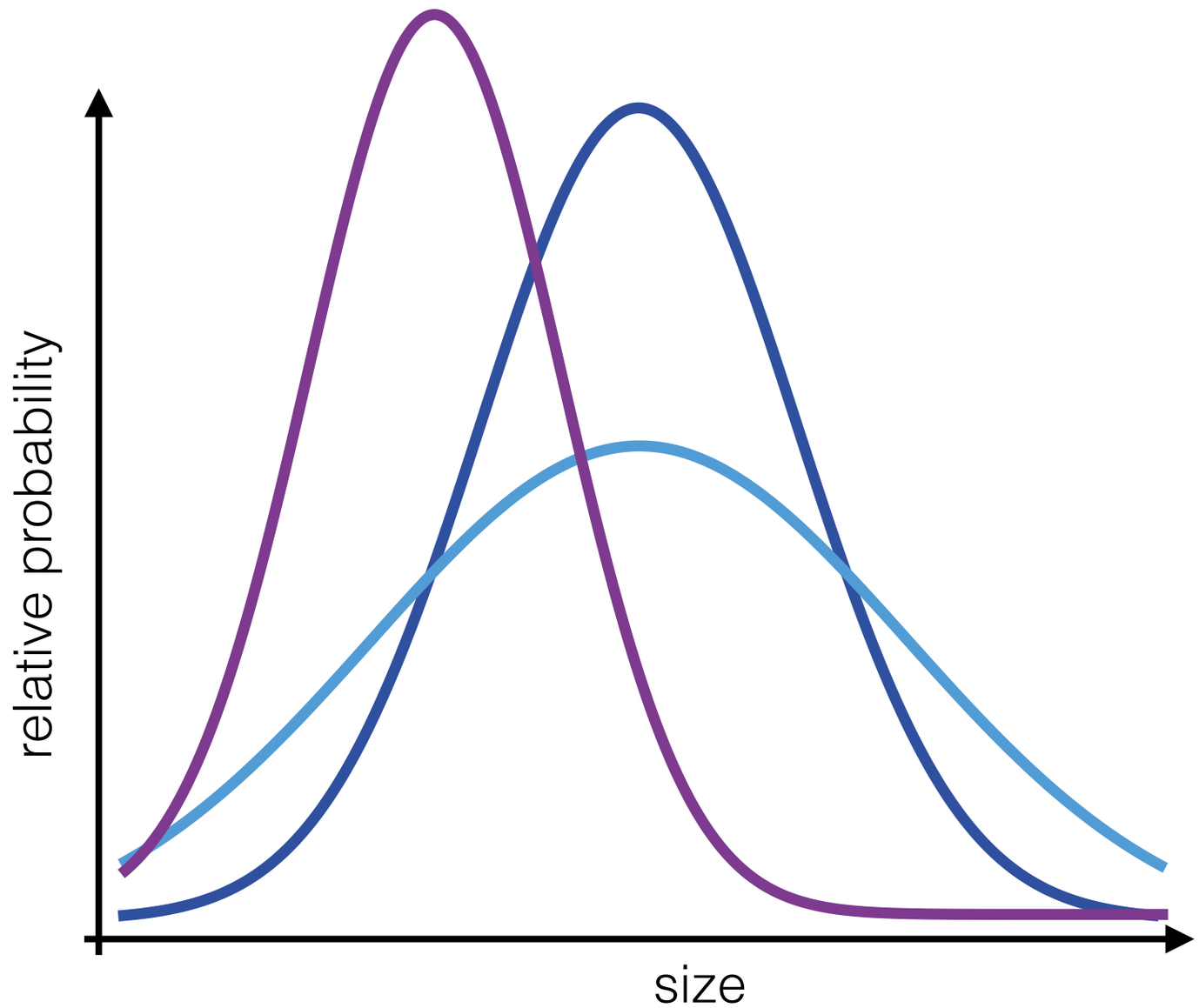
Marginal likelihood: probability of observing the data you observe given a model and all possible parameter values

Probability of the data: $P(🐼|\mathcal{N})$

Probability of the data given the model, integrating all possible parameter values

$$P(🐼|\mathcal{N}) = \int_\mu \int_\sigma P(🐼|\mu, \sigma, \mathcal{N}) \times P(\mu, \sigma|\mathcal{N})\ d\mu\ d\sigma$$

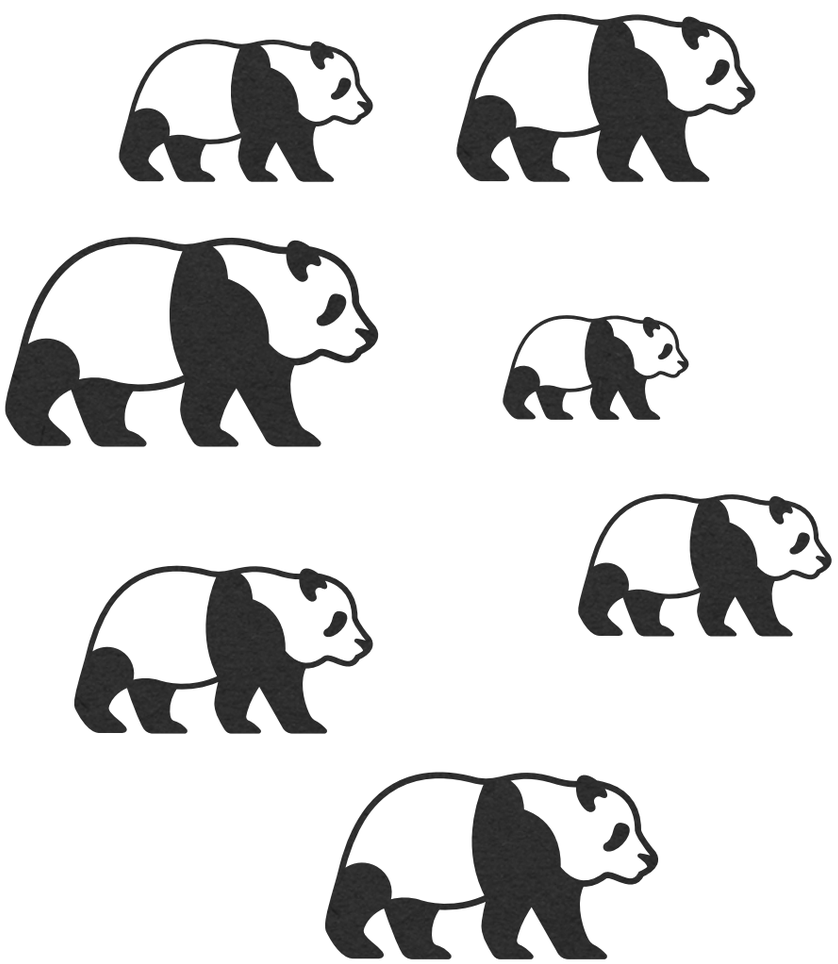


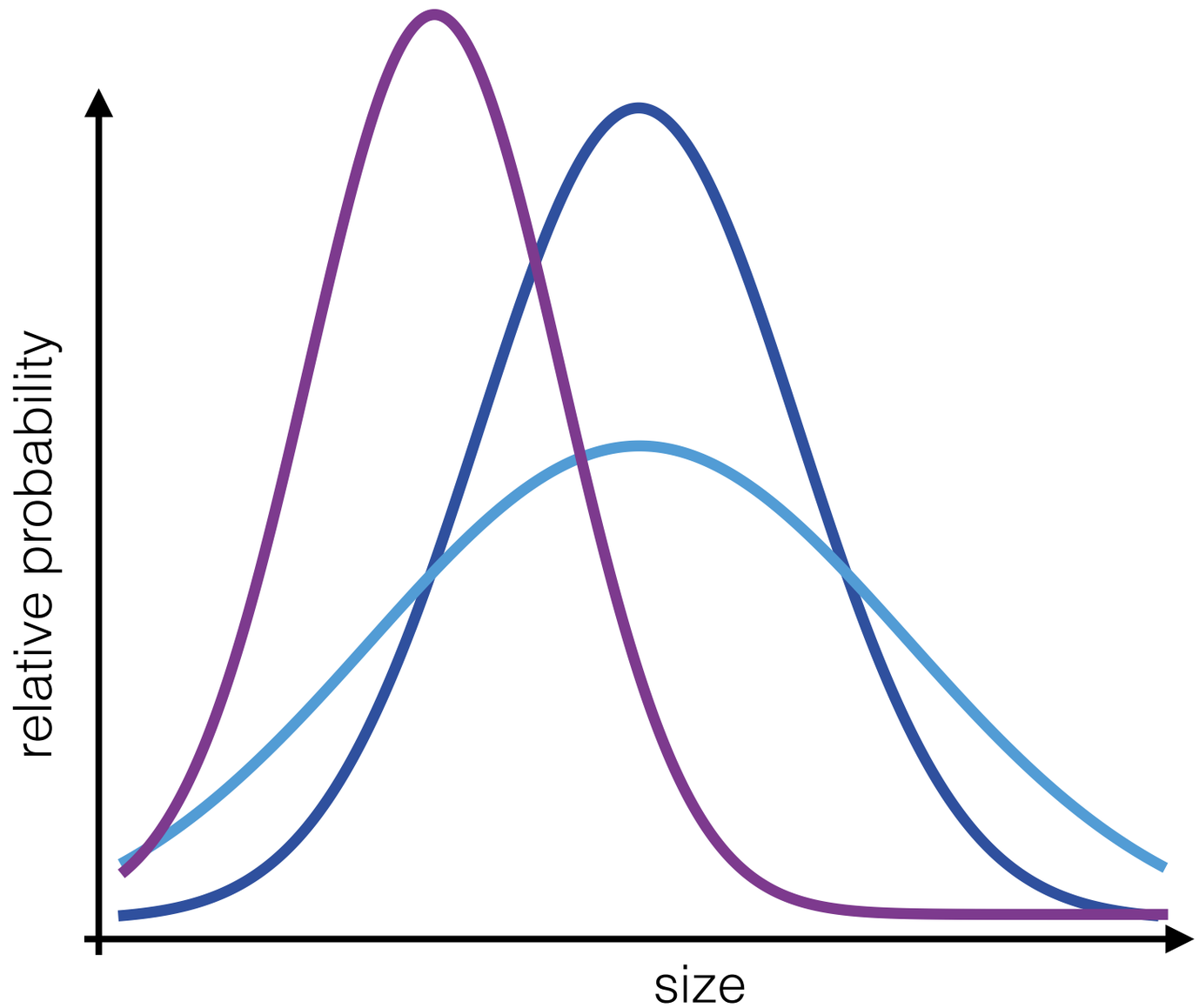


Probability of the data: $P(🐼|\mathcal{N})$

Probability of the data given the model, integrating all possible parameter values

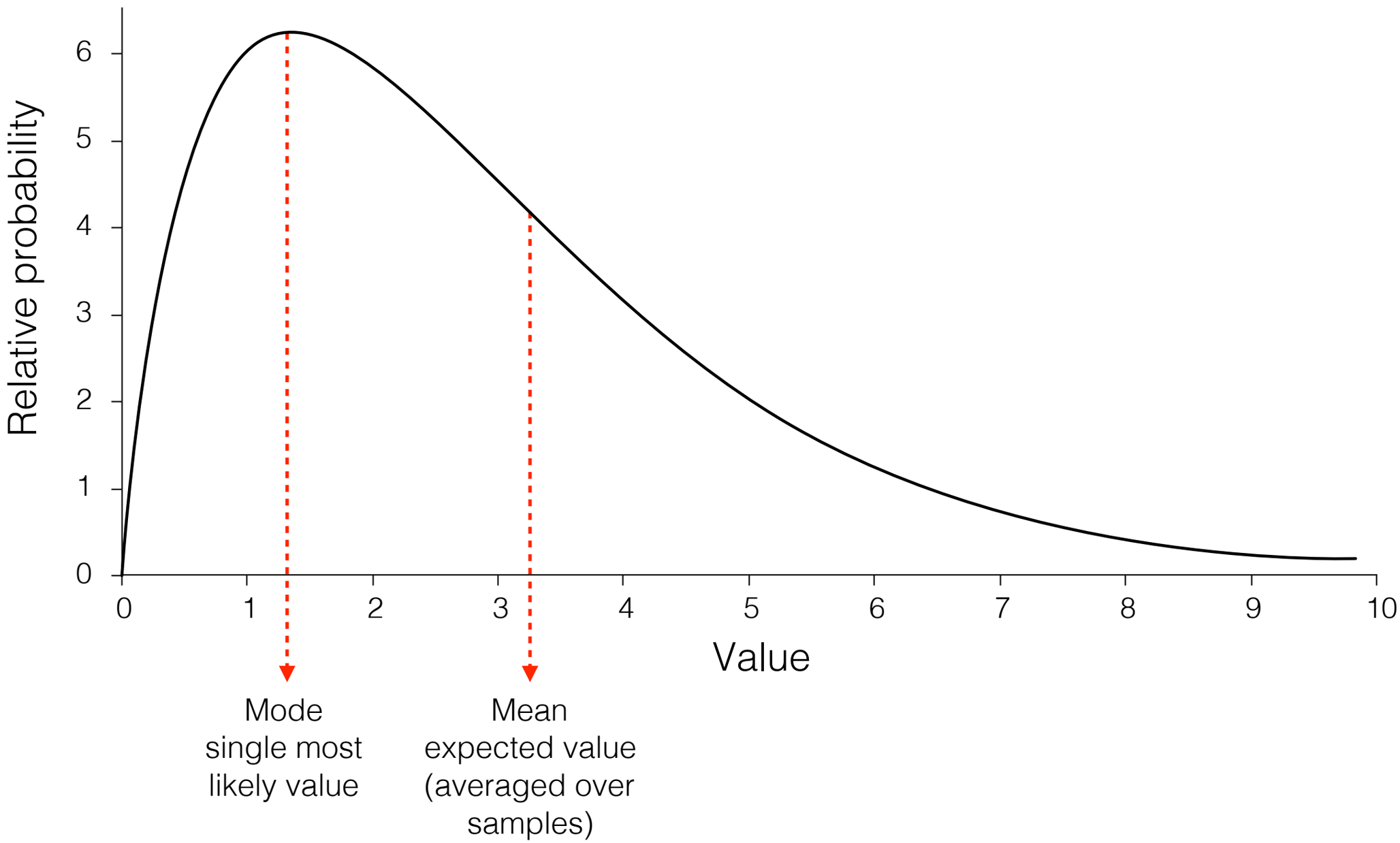
$$P(🐼|\mathcal{N}) = \text{constant}$$



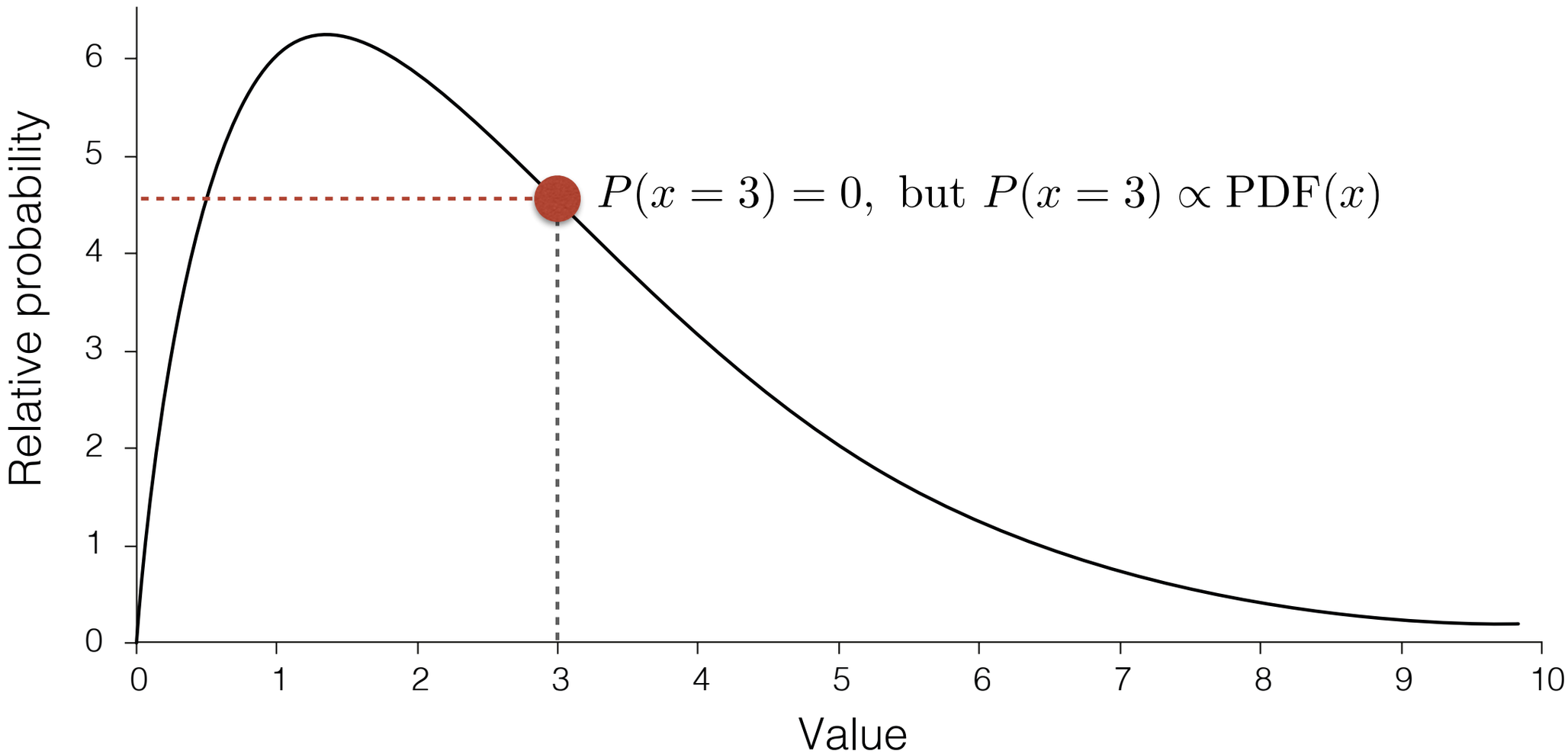


Probability density functions

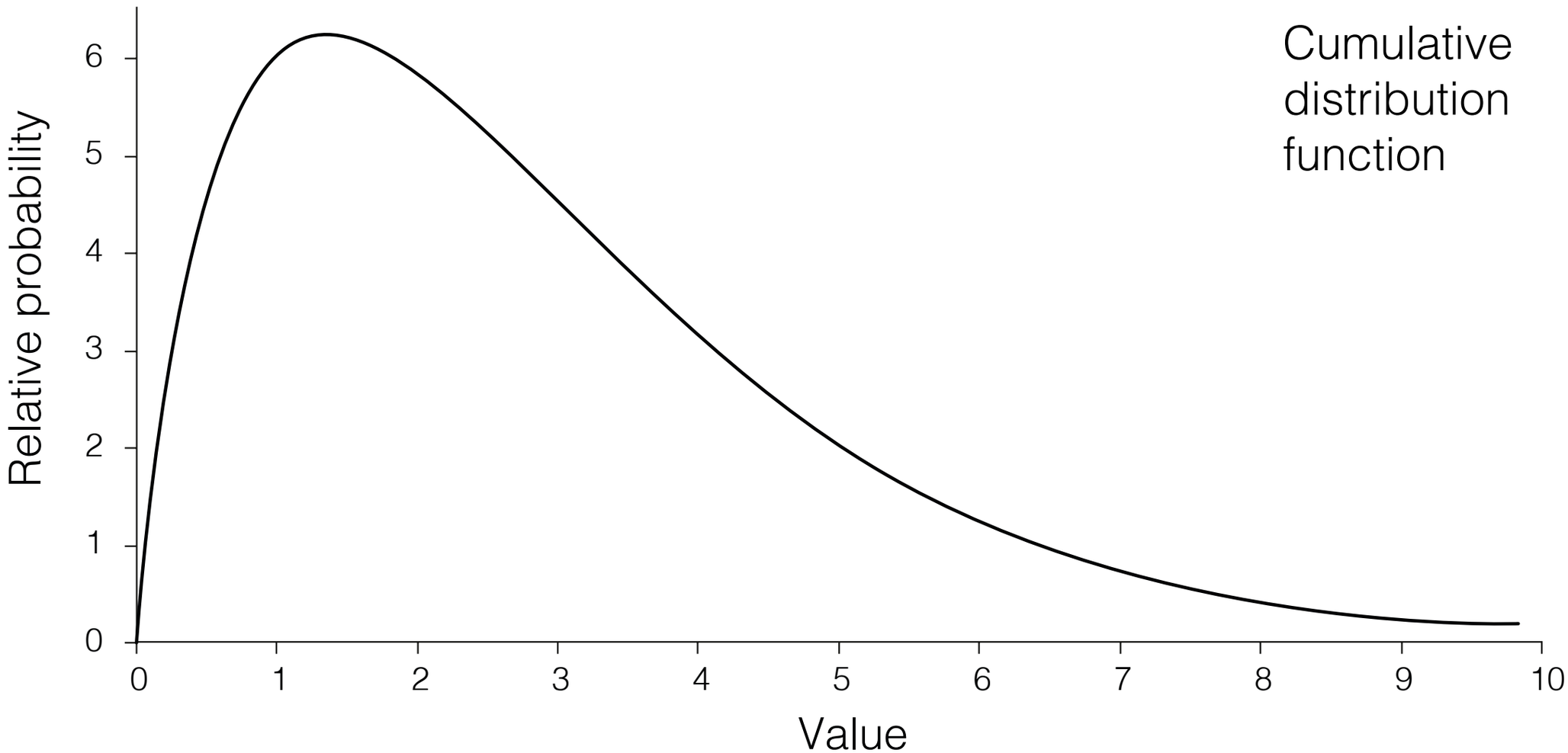
PDF $\neq$



Properties of a PDF



Properties of a PDF

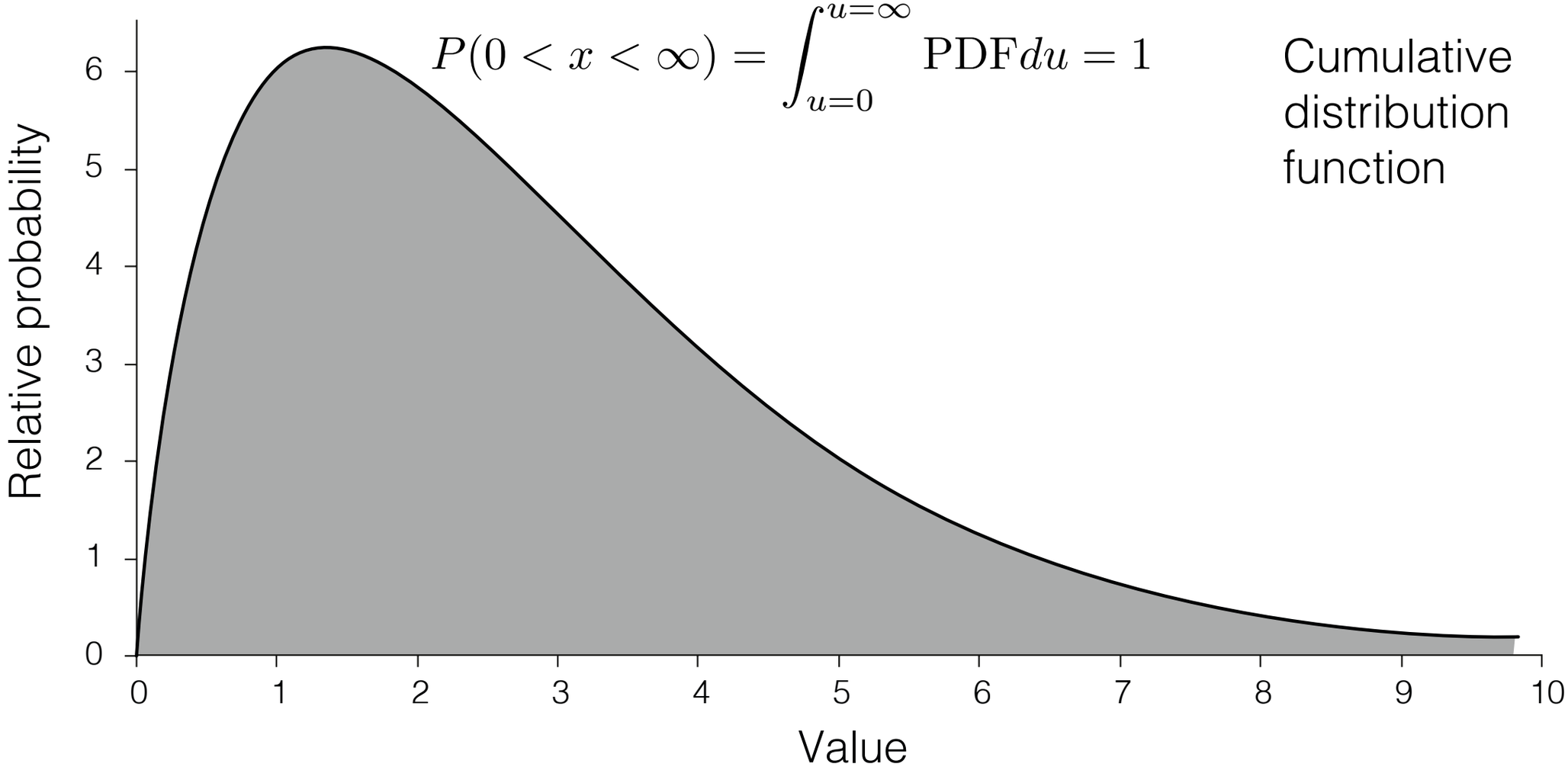


Cumulative distribution function

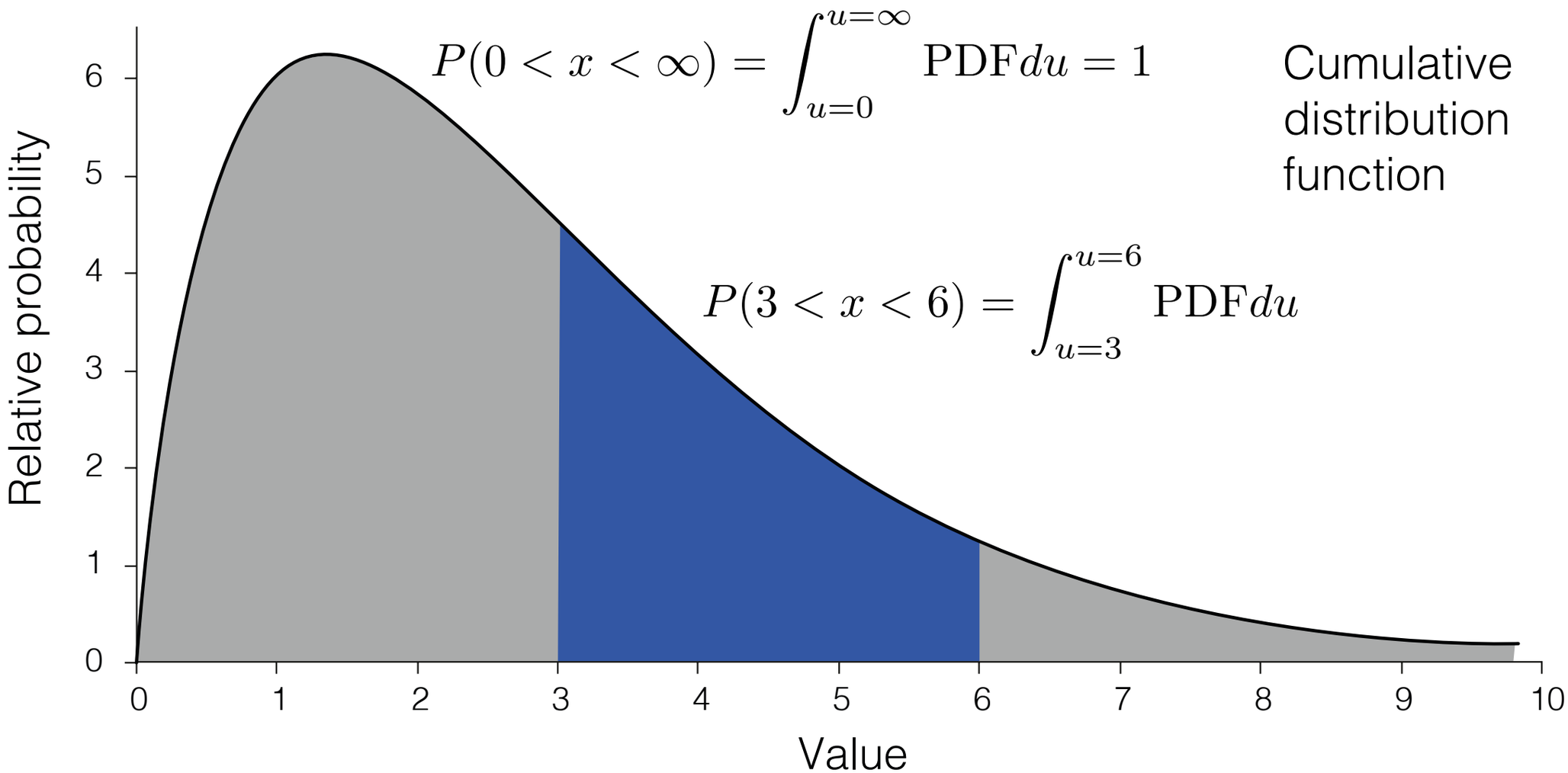
Properties of a PDF

$$P(0 < x < \infty) = \int_{u=0}^{u=\infty} \mathrm{PDF} du = 1$$

Cumulative distribution function



Properties of a PDF



Cumulative distribution function

Why using log-probabilities and log-likelihoods?

$P(A) = 0.25$

Why using log-probabilities and log-likelihoods?

P(A) = 0.25

P(ACGTT)= ?

Why using log-probabilities and log-likelihoods?

$$P(A) = 0.25$$

$$P(ACGTT) = P(A)P(C)P(G)P(T)P(T) = 0.001$$

Why using log-probabilities and log-likelihoods?

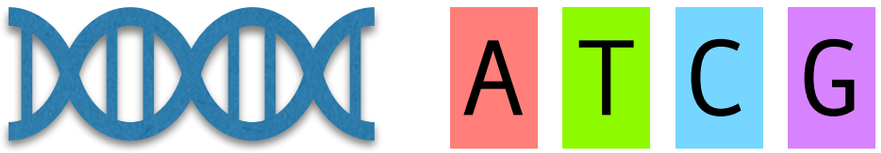
$P(A) = 0.25$

$P(ACGTT) = P(A)P(C)P(G)P(T)P(T) = 0.001$

$P($ $) = 0$

1,000 b

Why using log-probabilities and log-likelihoods?



$P(A) = 0.25$

$P(ACGTT) = P(A)P(C)P(G)P(T)P(T) = 0.001$

$P($ $) =$

1,000 b

Why using log-probabilities and log-likelihoods?

$P(A) = 0.25$

$P(ACGTT) = P(A)P(C)P(G)P(T)P(T) = 0.001$

$P($ $) =$

1,000 b

$\log P($ $) = ?$

1,000 b

Why using log-probabilities and log-likelihoods?

P(A) = 0.25

P(ACGTT) = P(A)P(C)P(G)P(T)P(T) = 0.001

P( ) =

1,000 b

logP( ) = -1386

1,000 b

Exercise: log-transform these operations

```
a+b
a-b
a*b
a/b
sqrt(a)
a^b
a!
exp(a)
```

Exercise: log-transform these operations

`a+b`	`log(a+b)`
`a-b`	`log(a-b)`
`a*b`	`log(a) + log(b)`
`a/b`	`log(a) - log(b)`
`sqrt(a)`	`0.5 * log(a)`
`a^b`	`b * log(a)`
`a!`	`sum(log(range(1,a)))`
`exp(a)`	`a`

Bayesian inference

Likelihood Prior probability

Posterior probability

$$P(\alpha|x, M) = \frac{P(x|\alpha, M) \times P(\alpha|M)}{P(x|M)}$$

Probability of the data

Bayesian inference

Likelihood Prior probability

Posterior probability

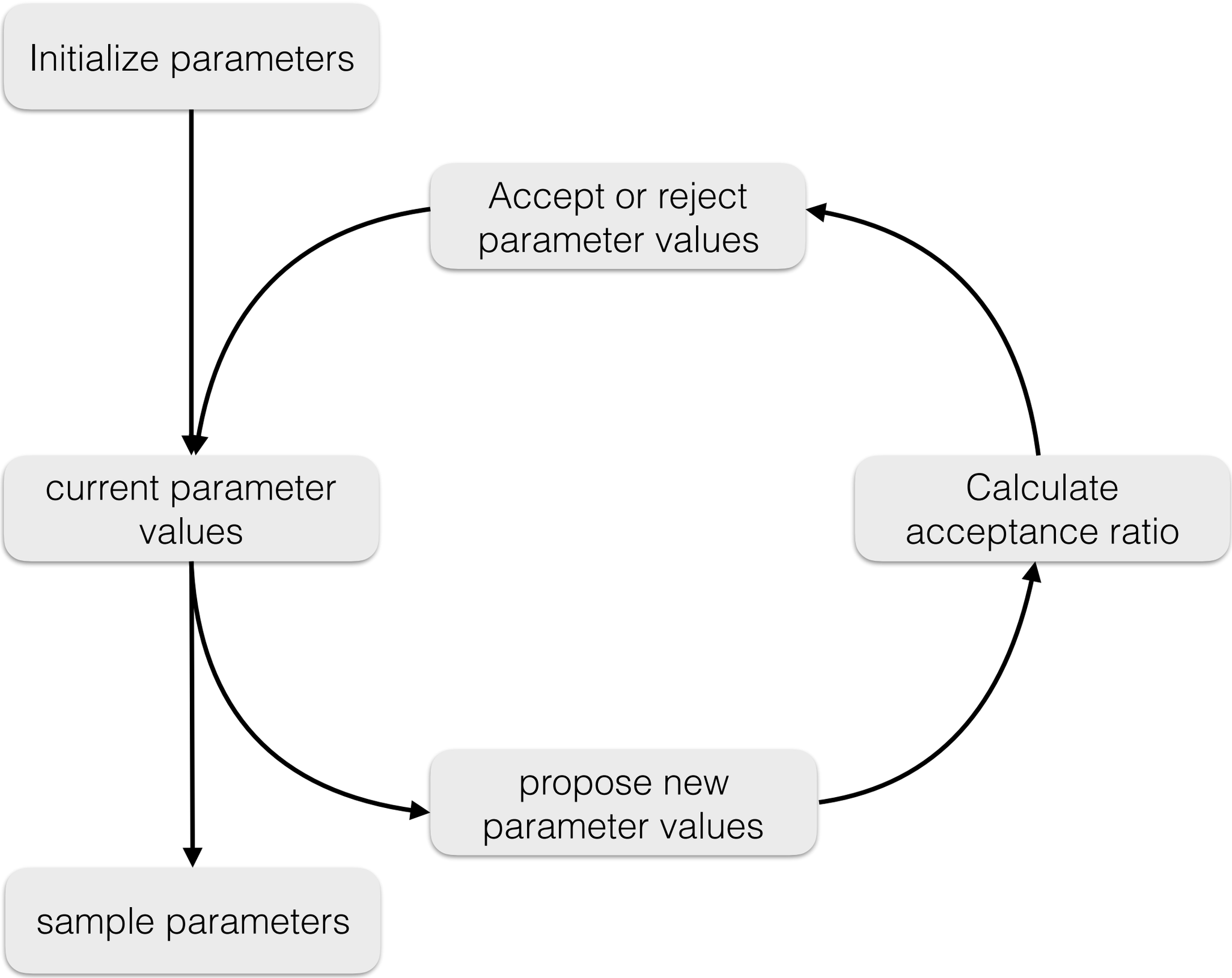
$$P(\alpha|x, M) = \frac{P(x|\alpha, M) \times P(\alpha|M)}{P(x|M)}$$

Probability of the data

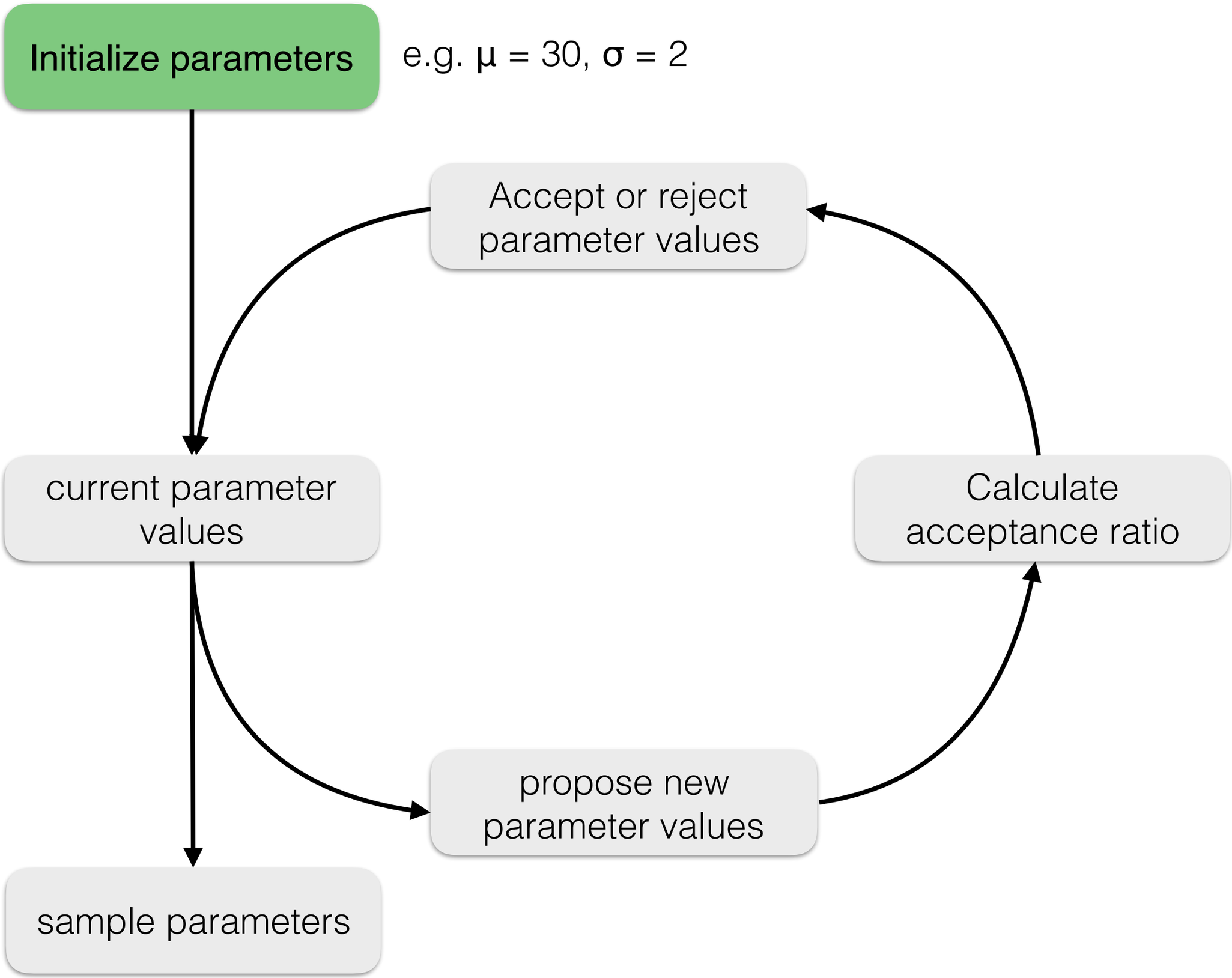
$$P(\alpha|x, M) \propto P(x|\alpha, M) \times P(\alpha|M)$$

Posterior probability is proportional to the product of likelihood and prior

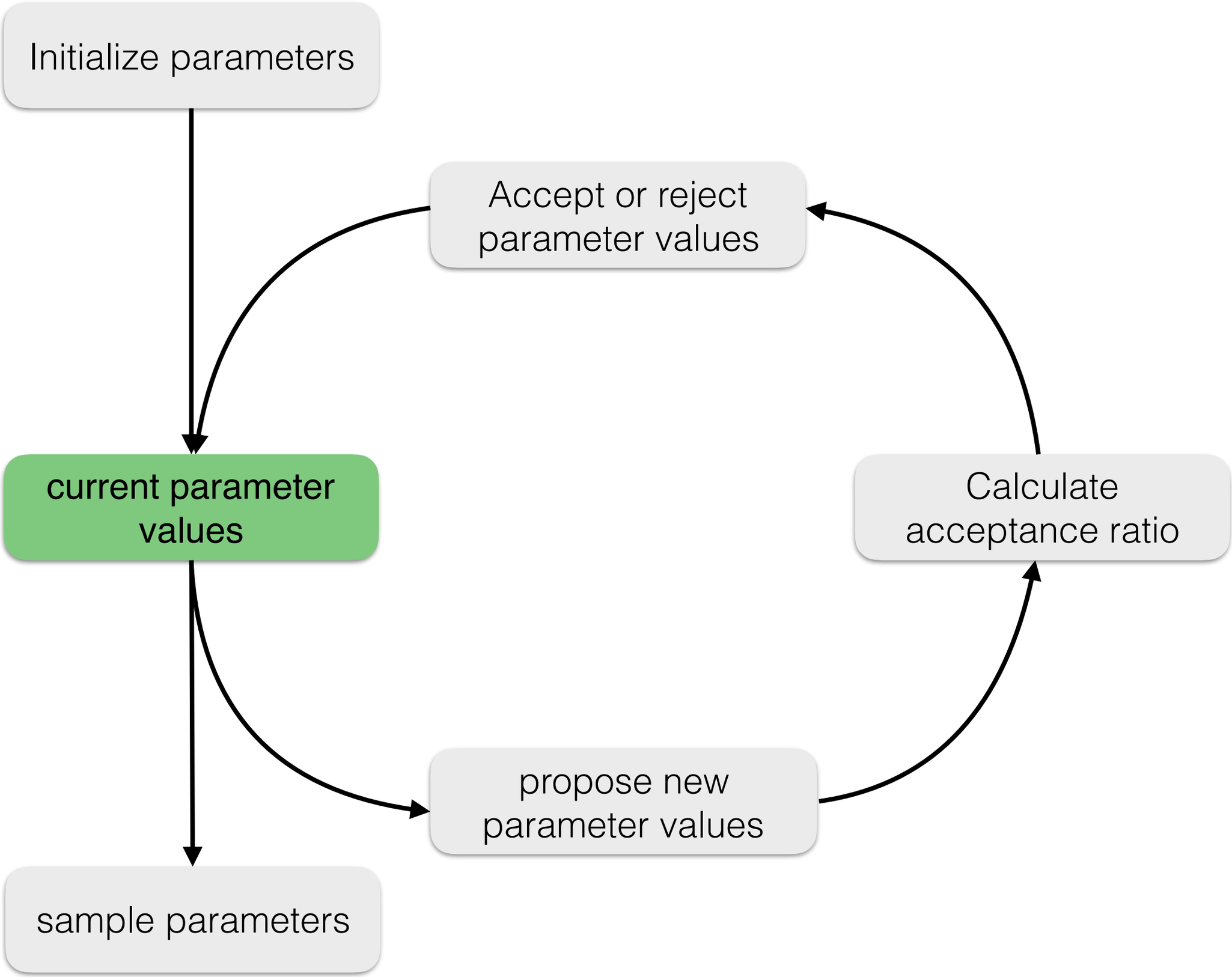
Metropolis-Hastings MCMC



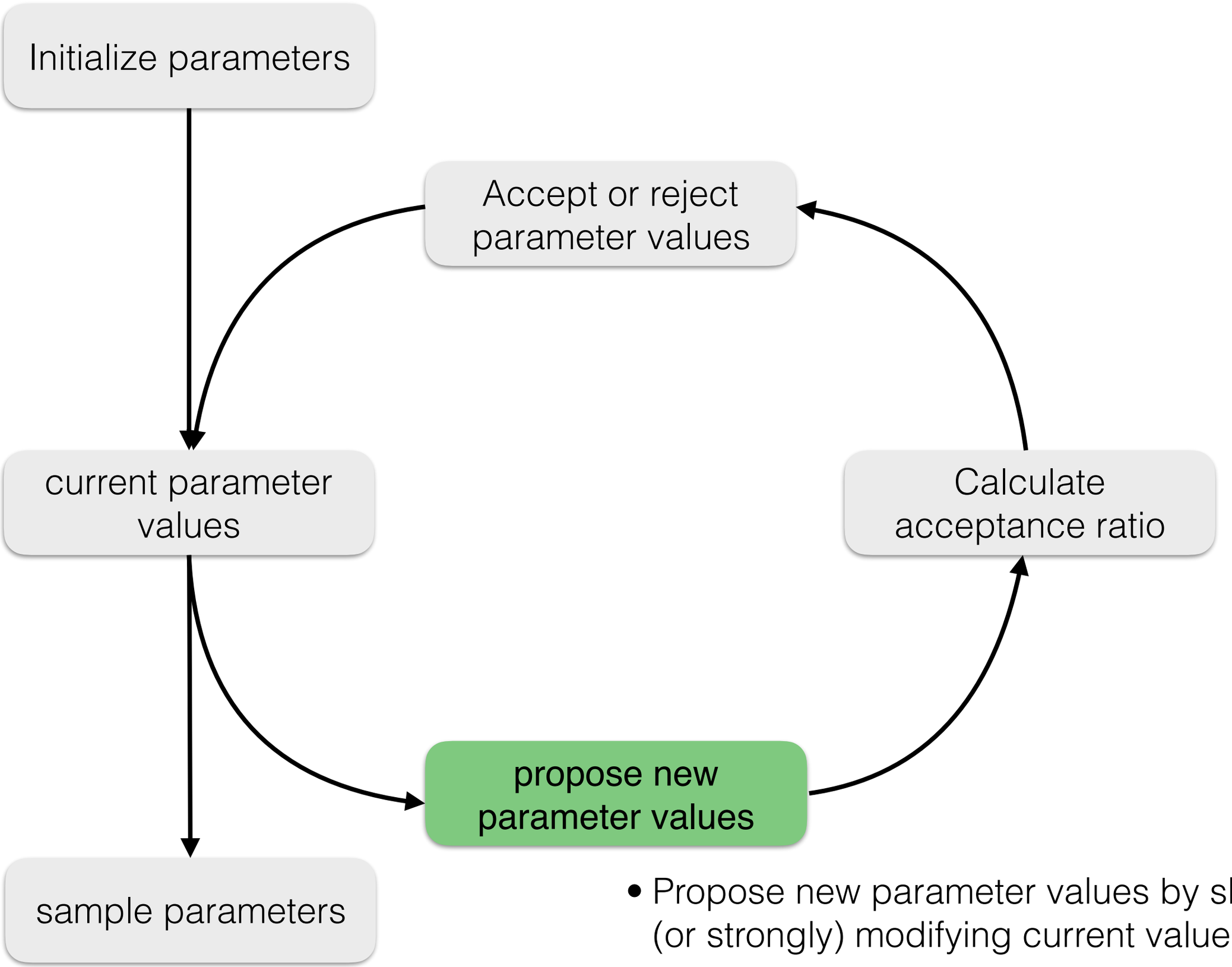
Metropolis-Hastings MCMC



Metropolis-Hastings MCMC



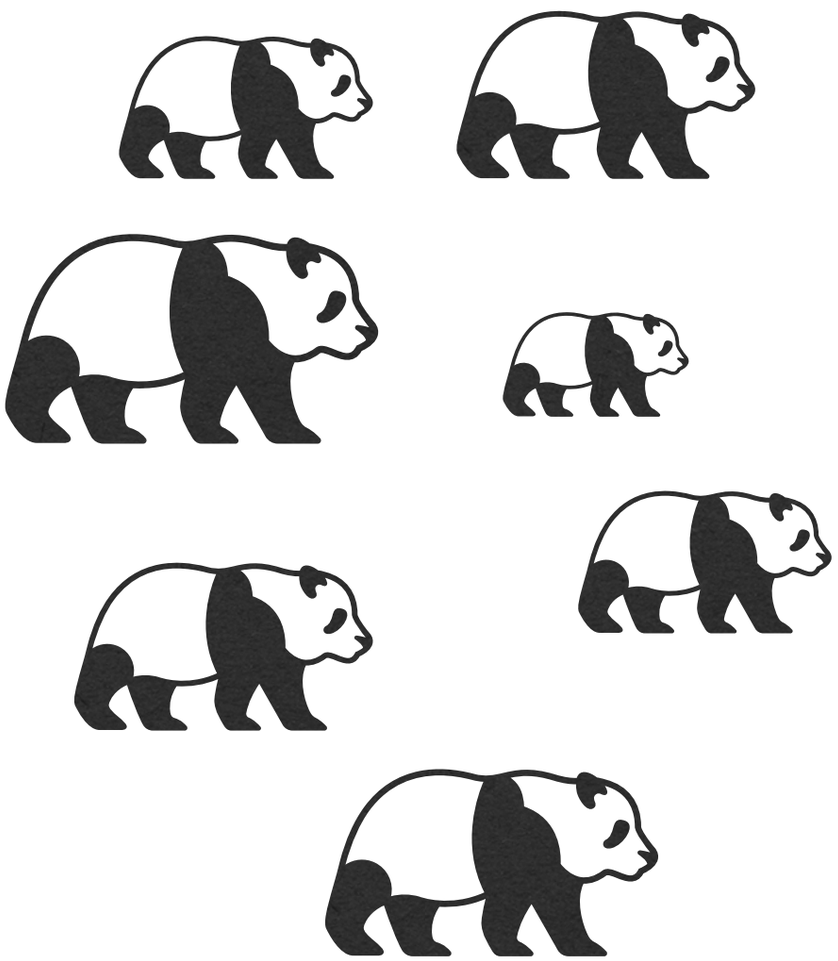
Metropolis-Hastings MCMC

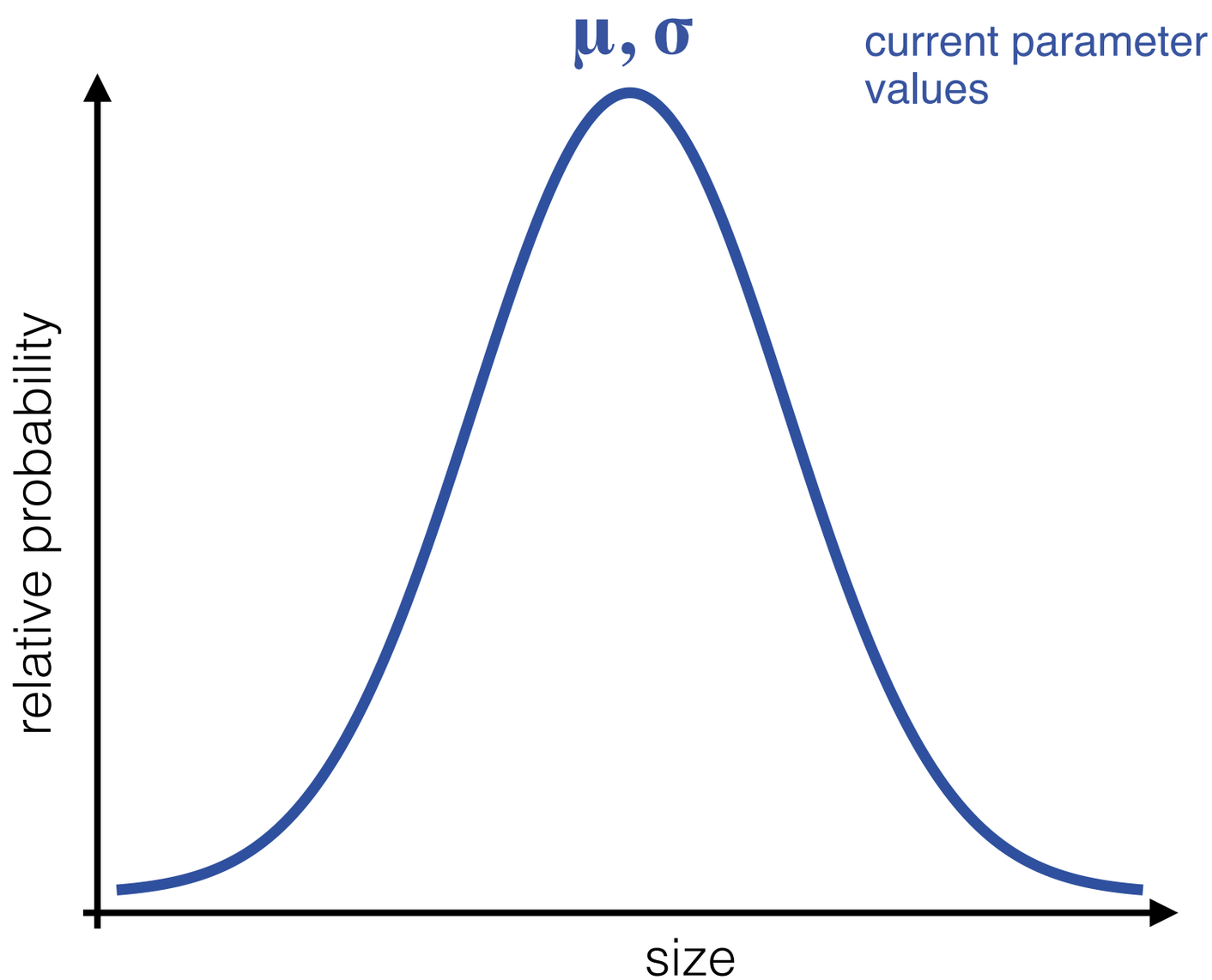


- Propose new parameter values by slightly (or strongly) modifying current values

Propose new parameter values

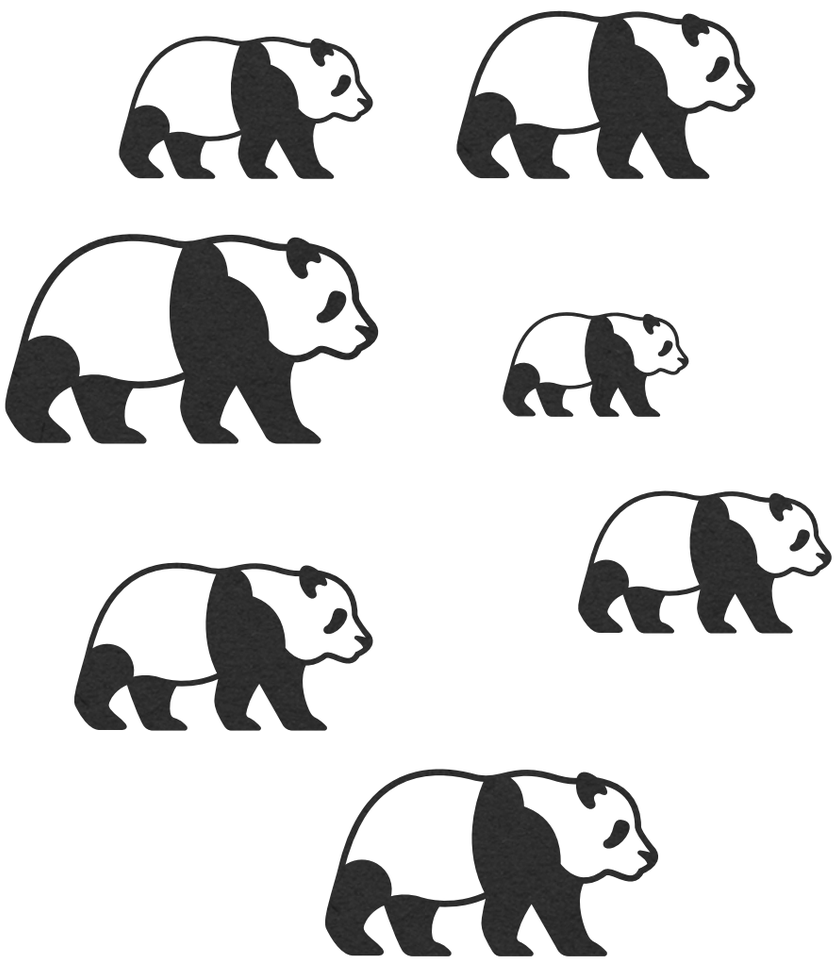
E.g. propose a new standard deviation

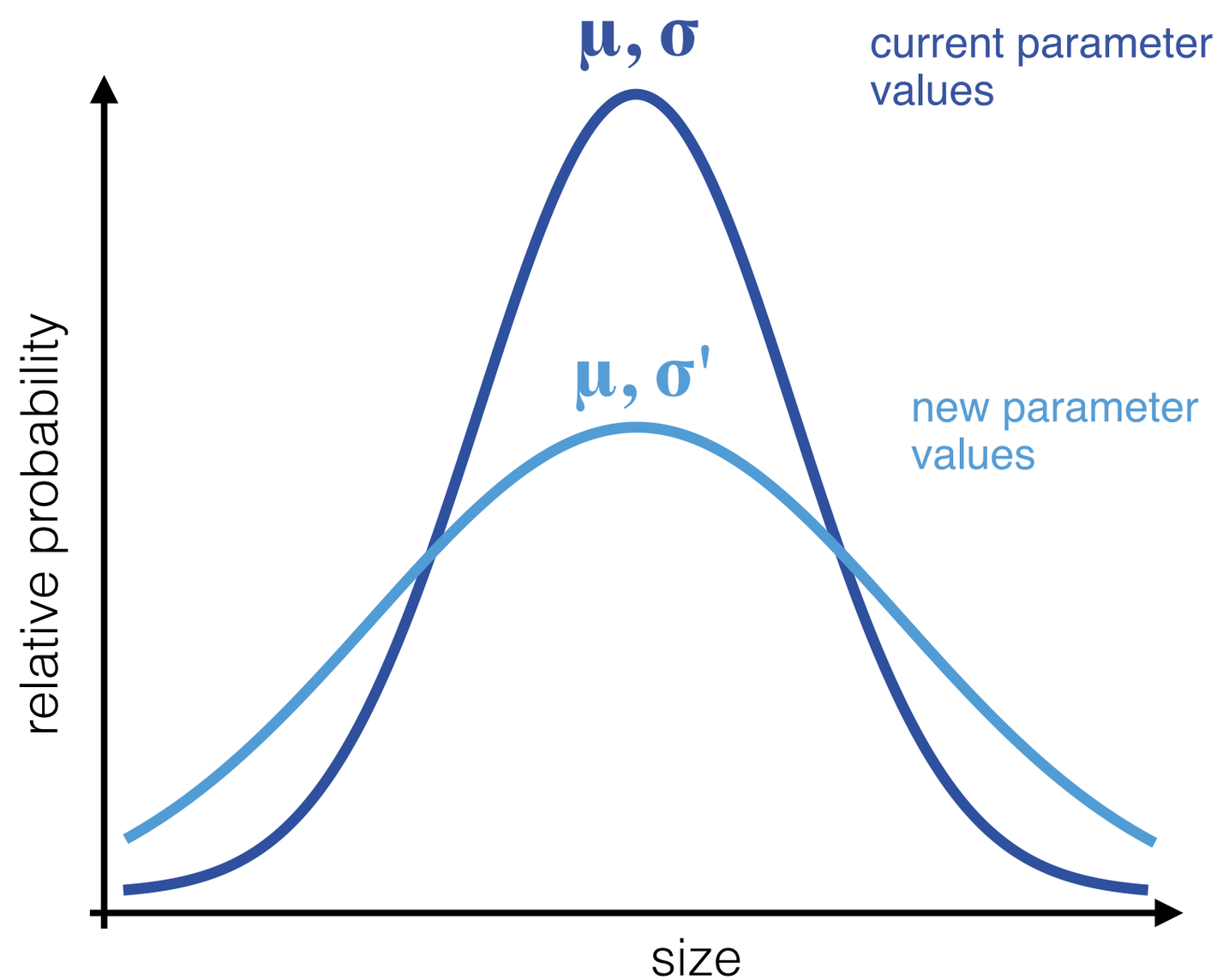




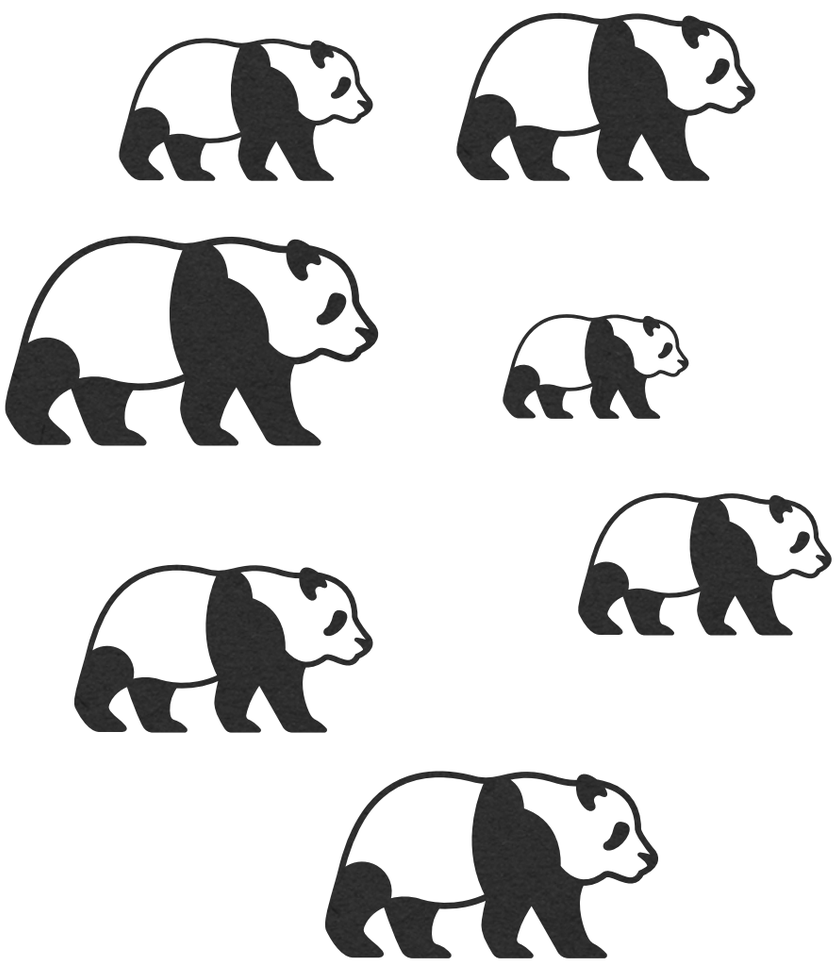
Propose new parameter values

E.g. propose a new standard deviation

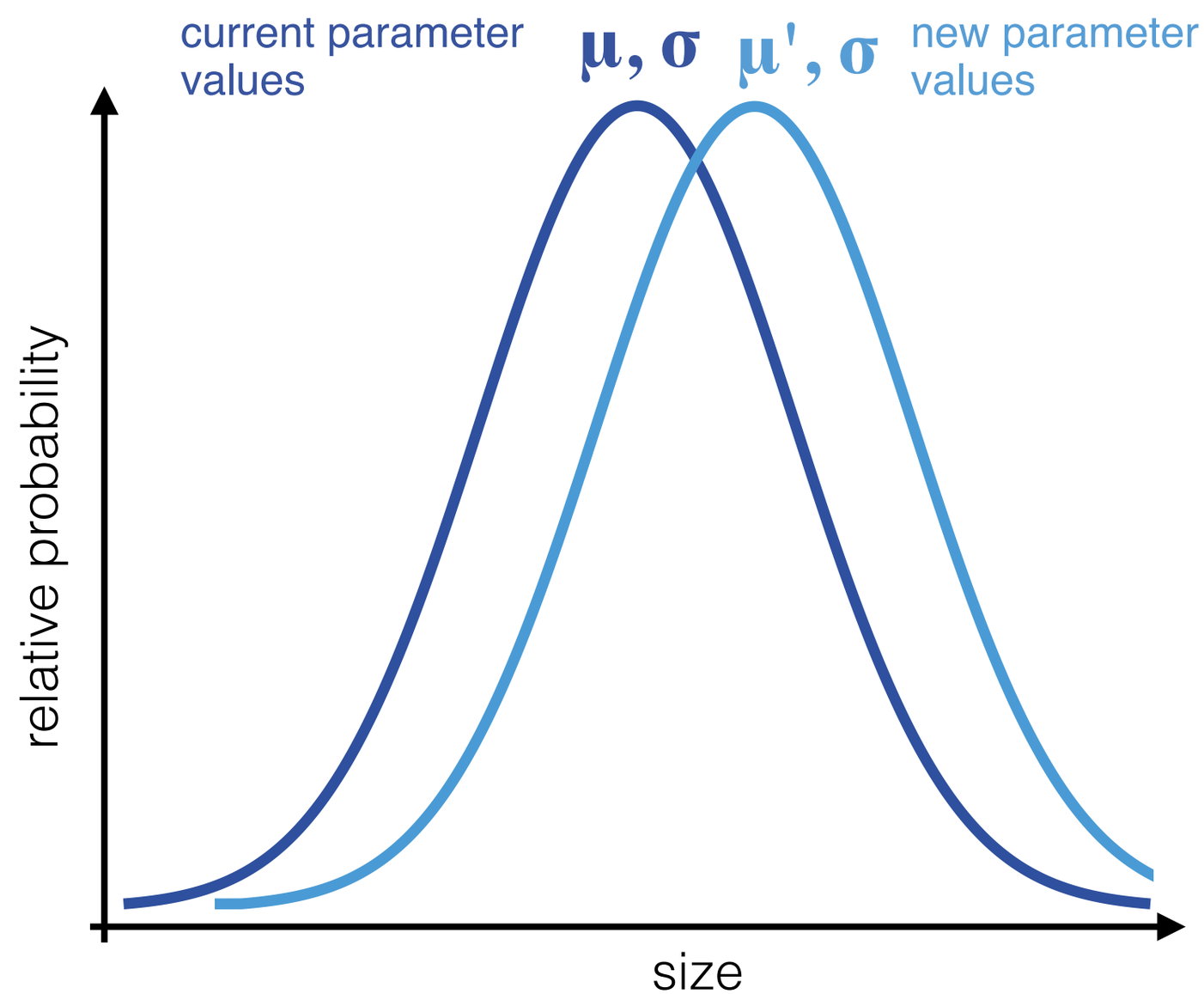




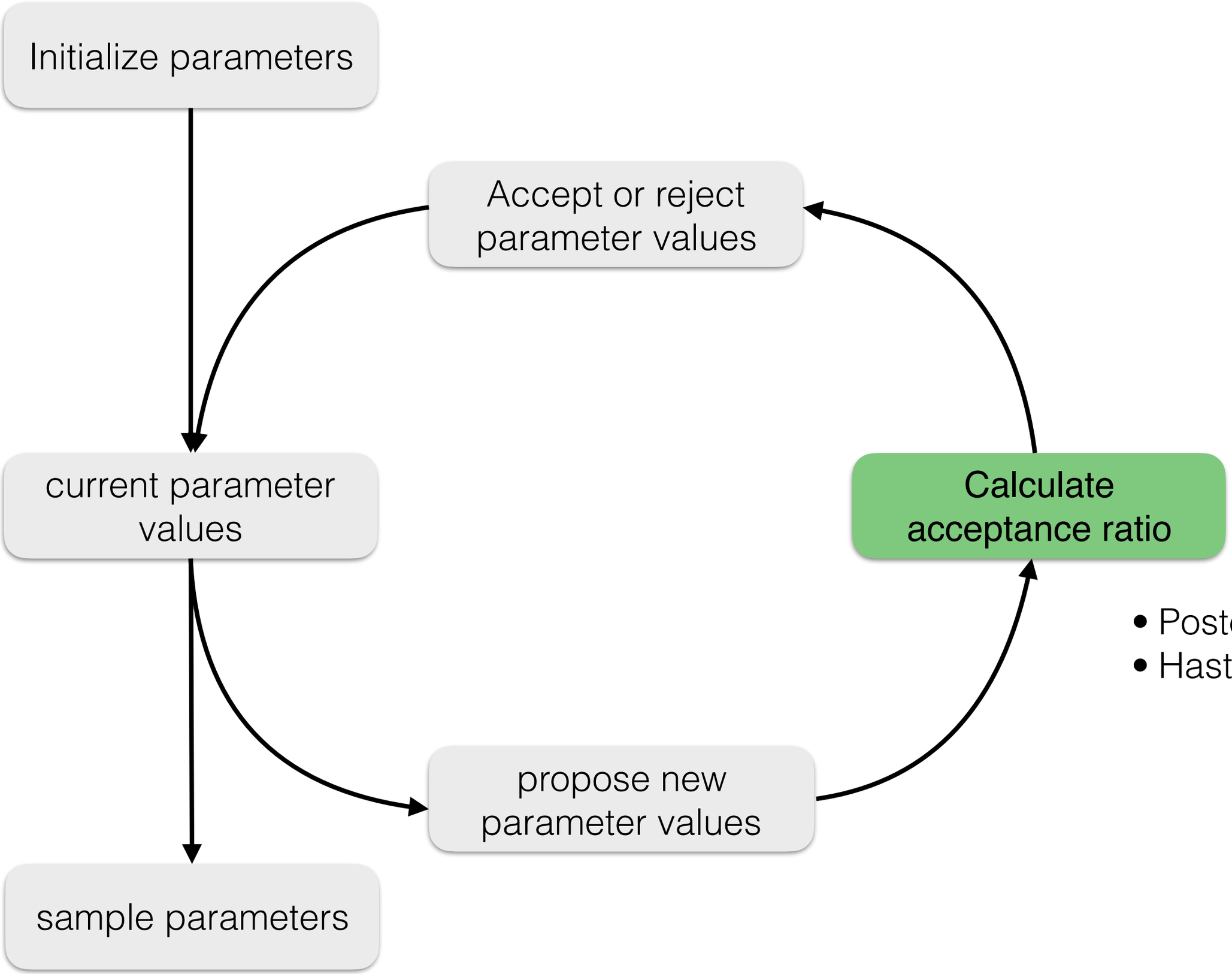
Propose new parameter values



E.g. propose a new mean



Metropolis-Hastings MCMC



- Posterior ratio
- Hastings ratio

Acceptance ratio

Comparing the new state against the old state

Posterior ratio

$$\frac{P(\alpha'|x, M)}{P(\alpha|x, M)}$$

Hastings ratio

Acceptance ratio

Comparing the new state against the old state

Posterior ratio

$$\frac{P(\alpha'|x,M)}{P(\alpha|x,M)} = \frac{\dfrac{P(x|\alpha',M)\times P(\alpha'|M)}{P(x|M)}}{\dfrac{P(x|\alpha,M)\times P(\alpha|M)}{P(x|M)}}$$

Hastings ratio

Acceptance ratio

Comparing the new state against the old state

Posterior ratio

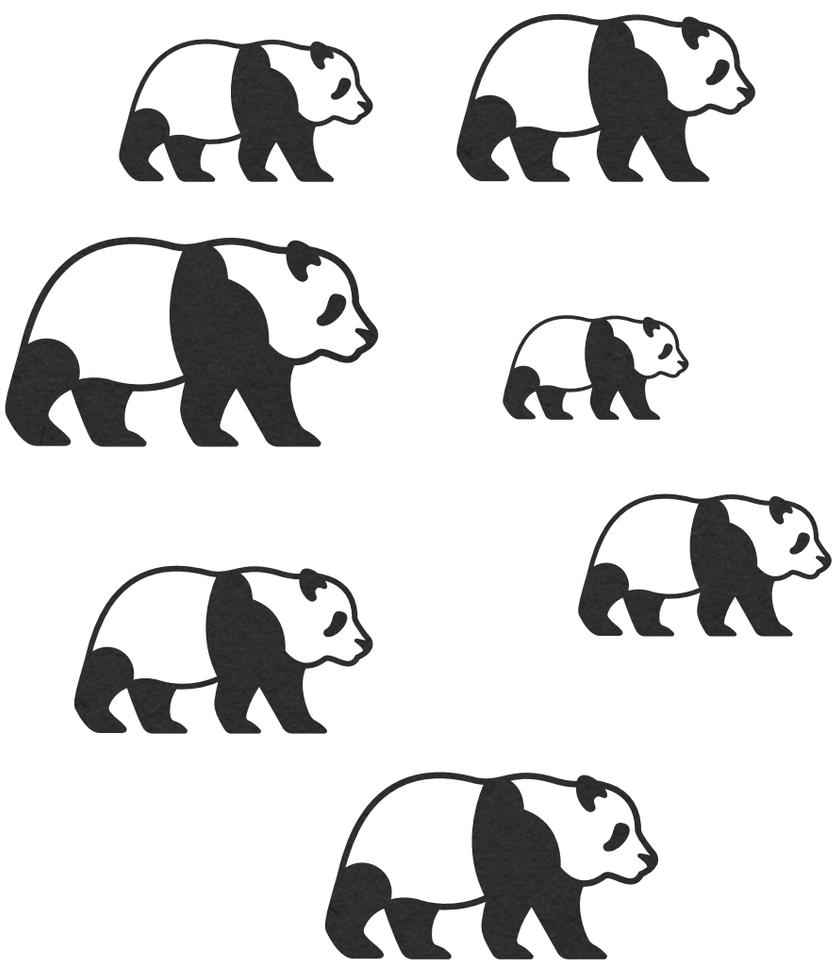
$$\frac{P(\alpha'|x,M)}{P(\alpha|x,M)} = \frac{\dfrac{P(x|\alpha',M)\times P(\alpha'|M)}{P(x|M)}}{\dfrac{P(x|\alpha,M)\times P(\alpha|M)}{P(x|M)}}$$

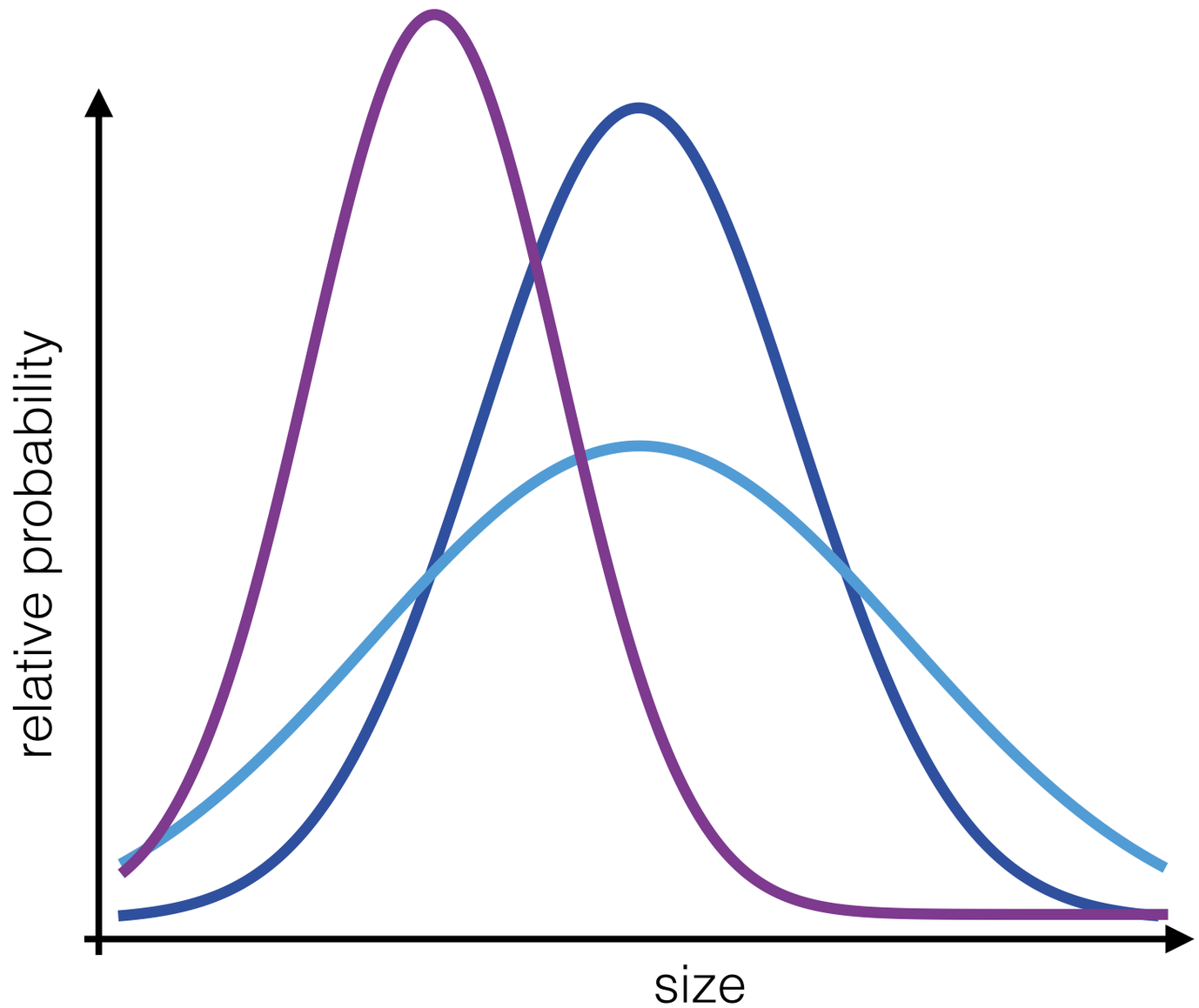
Hastings ratio

Probability of the data: $P(🐼|\mathcal{N})$

Probability of the data given the model

$$P(🐼|\mathcal{N}) = \text{constant}$$





Acceptance ratio

Comparing the new state against the old state

Posterior ratio

$$\frac{P(\alpha'|x,M)}{P(\alpha|x,M)} = \frac{\dfrac{P(x|\alpha',M)\times P(\alpha'|M)}{\cancel{P(x|M)}}}{\dfrac{P(x|\alpha,M)\times P(\alpha|M)}{\cancel{P(x|M)}}}$$

Hastings ratio

Acceptance ratio

Comparing the new state against the old state

Posterior ratio

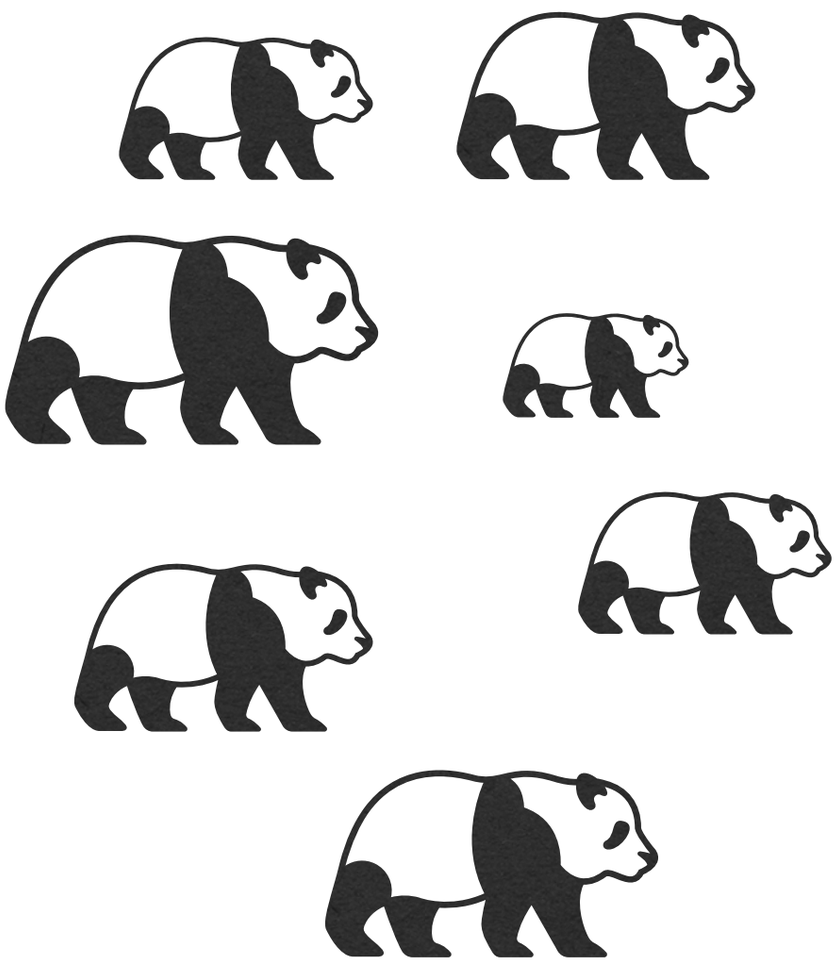
$$\frac{P(\alpha'|x,M)}{P(\alpha|x,M)} = \frac{\dfrac{P(x|\alpha',M)\times P(\alpha'|M)}{\cancel{P(x|M)}}}{\dfrac{P(x|\alpha,M)\times P(\alpha|M)}{\cancel{P(x|M)}}} = \underbrace{\frac{P(x|\alpha',M)}{P(x|\alpha,M)}}_{\text{Likelihood ratio}} \times \underbrace{\frac{P(\alpha'|M)}{P(\alpha|M)}}_{\text{Prior ratio}}$$

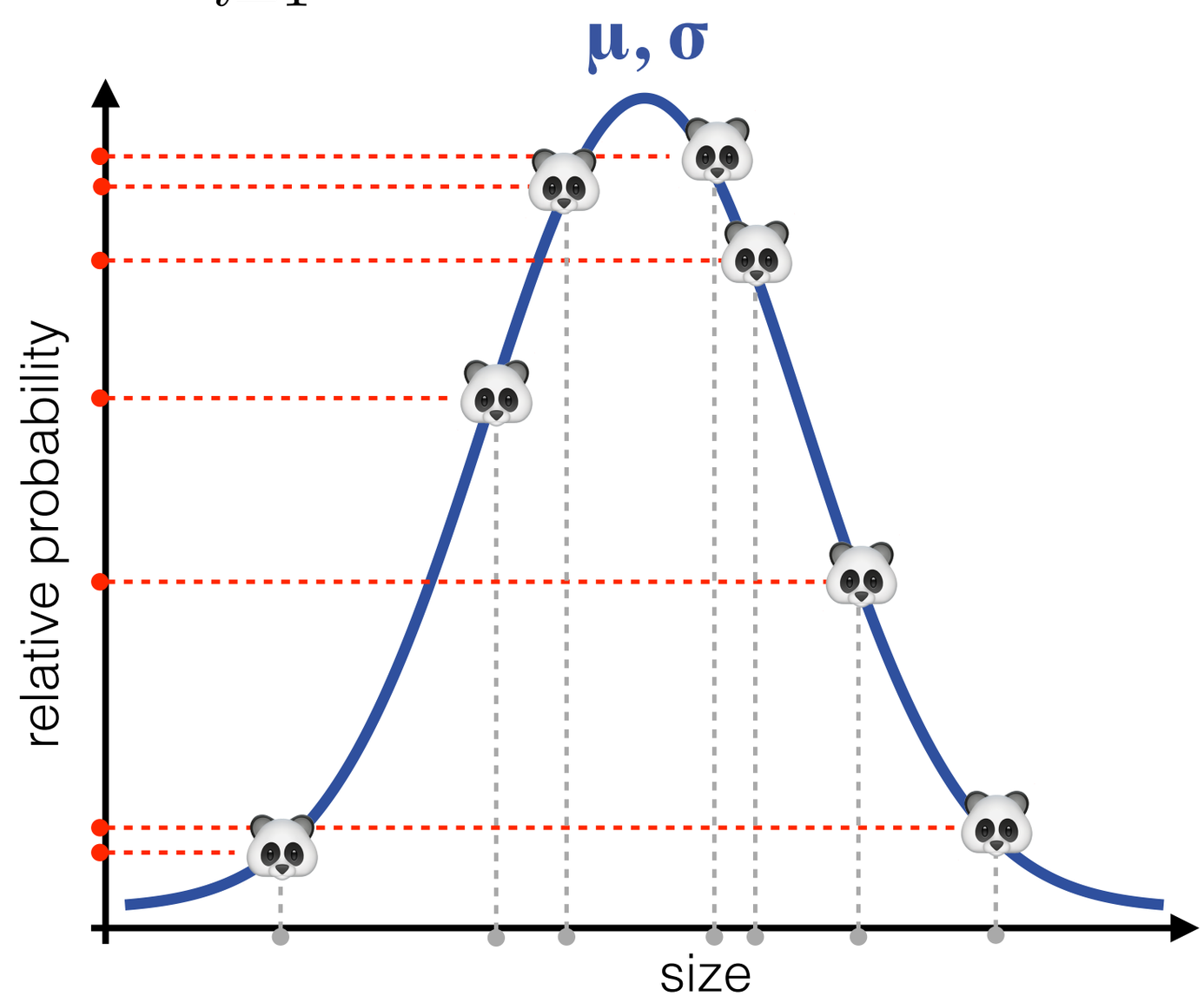
Hastings ratio

Calculating the likelihood: $P(🐼|\mu, \sigma, \mathcal{N})$

Likelihood: probability of observing the data you observe given a model and its parameters

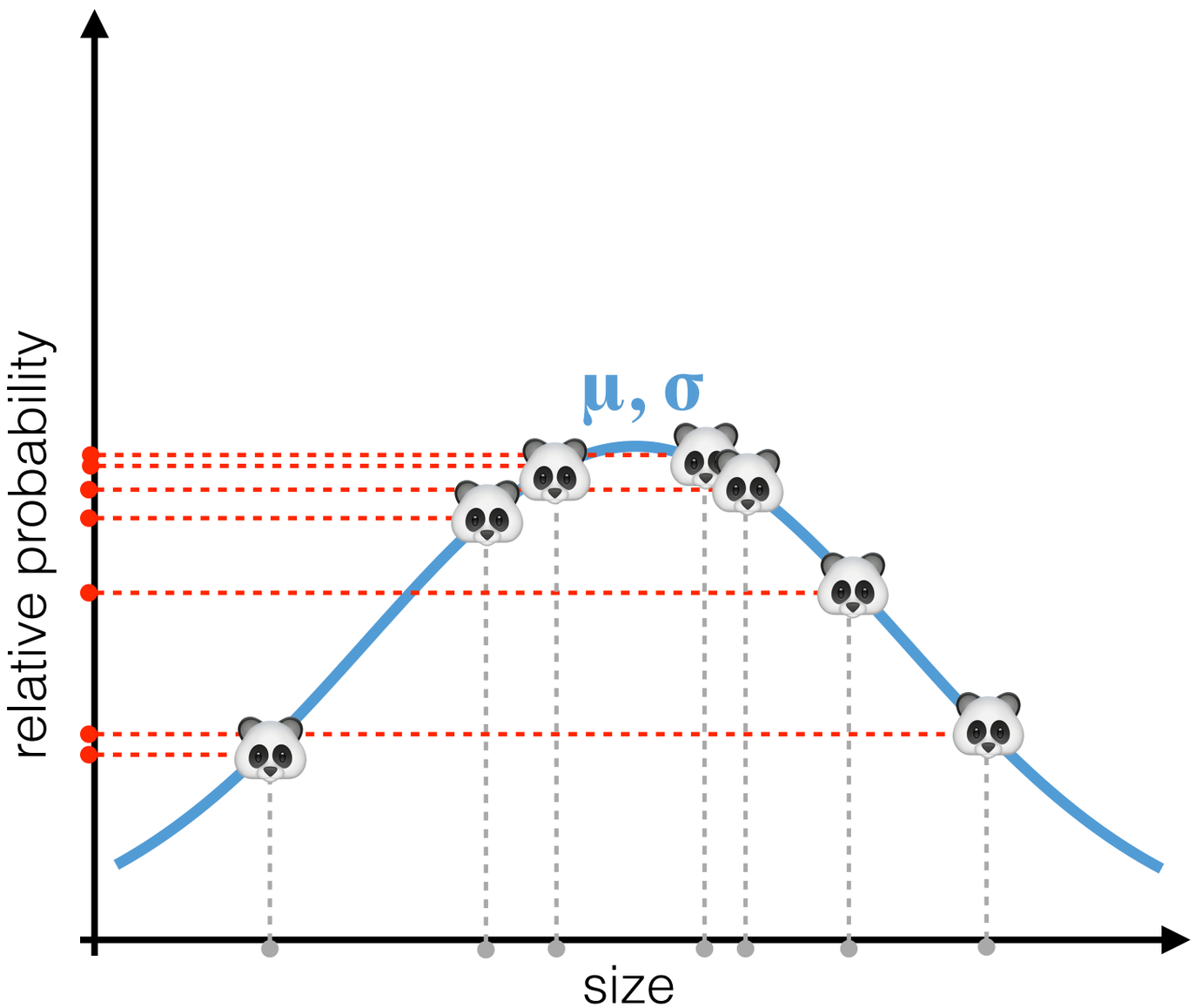
$$P(🐼_1, ..., 🐼_7|\mu, \sigma, \mathcal{N}) = \prod_{i=1}^{7} P(🐼_i|\mu, \sigma, \mathcal{N})$$



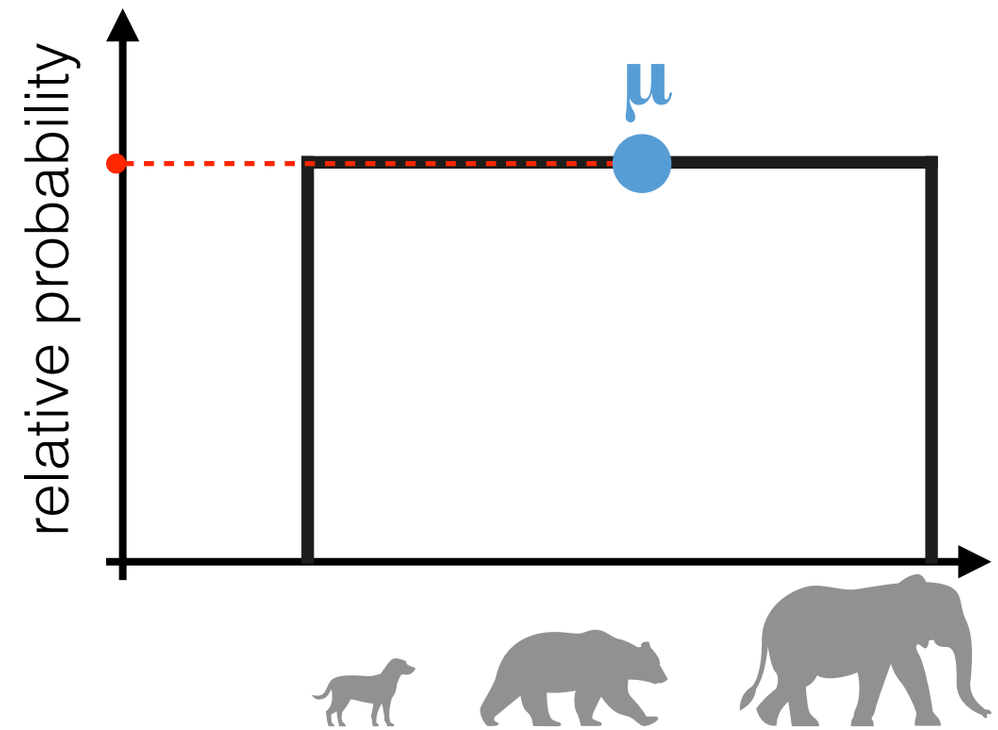


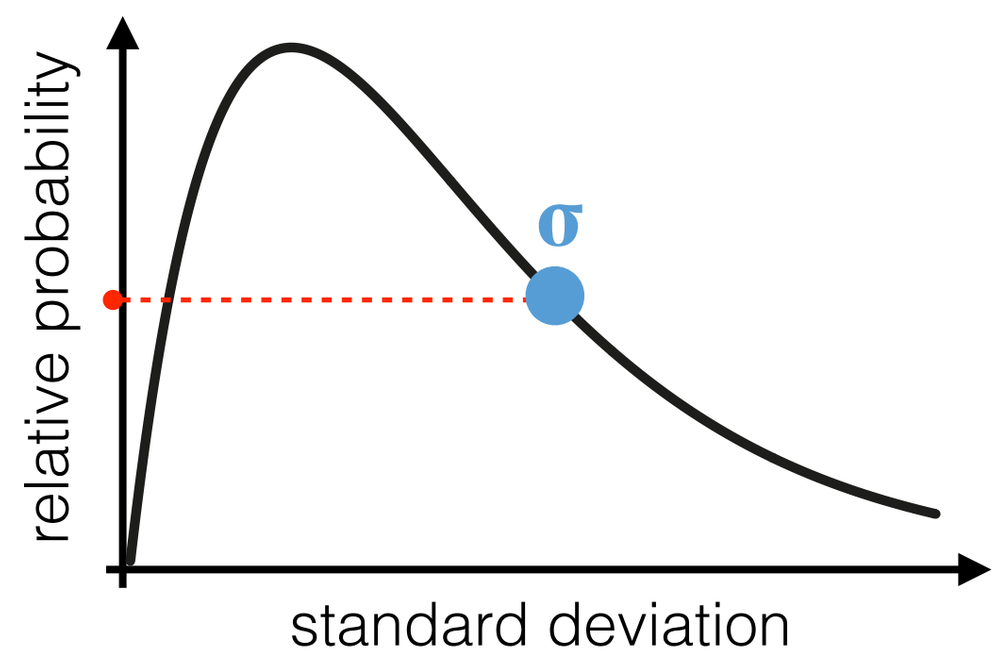
Calculating the posterior: $P(\mu, \sigma | 🐼, \mathcal{N})$

Likelihood



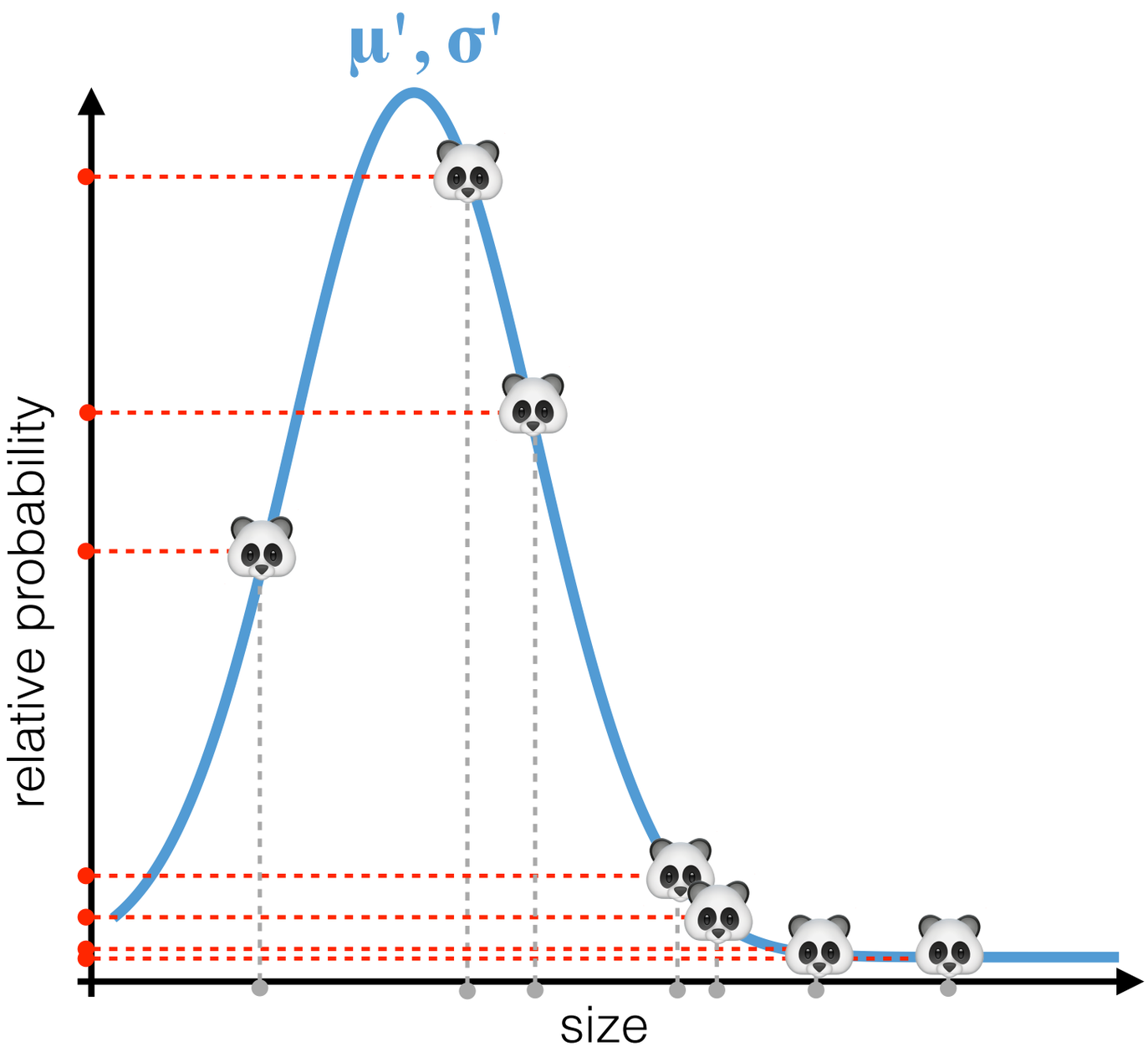
Priors



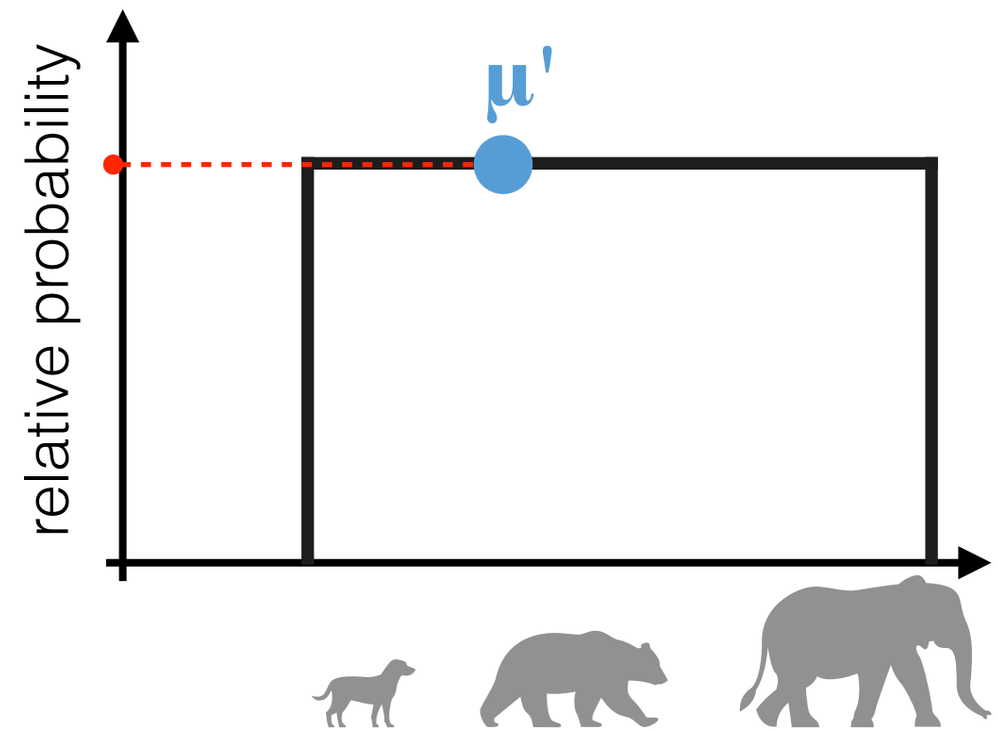


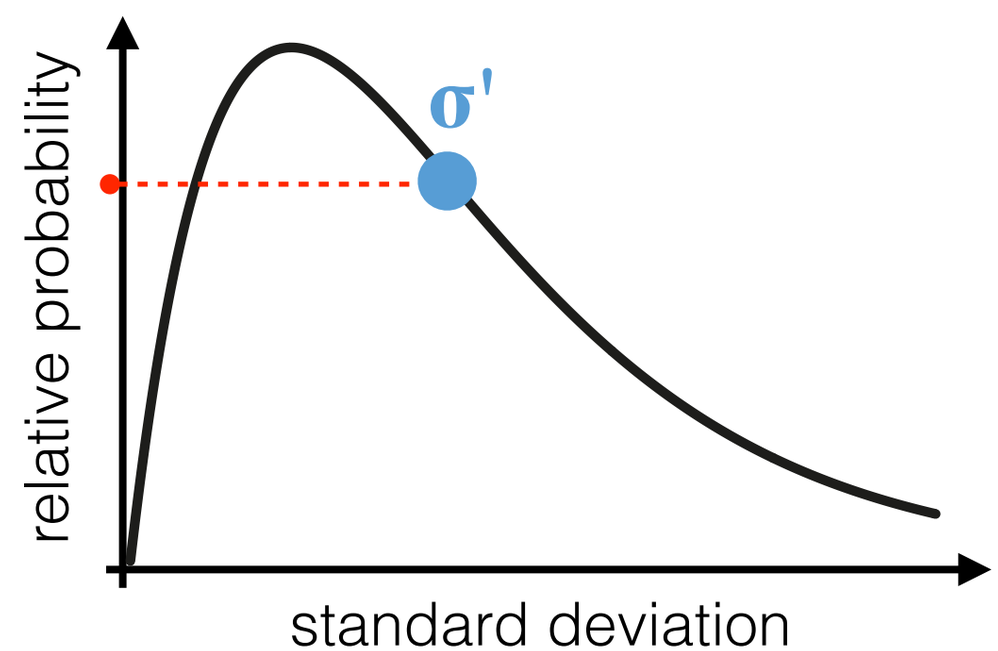
Calculating the posterior: $P(\mu', \sigma' | 🐼, \mathcal{N})$

Likelihood



Priors





Acceptance ratio

Comparing the new state against the old state

Posterior ratio

$$\frac{P(\alpha'|x,M)}{P(\alpha|x,M)} = \underbrace{\frac{P(x|\alpha',M)}{P(x|\alpha,M)}}_{\text{Likelihood ratio}} \times \underbrace{\frac{P(\alpha'|M)}{P(\alpha|M)}}_{\text{Prior ratio}}$$

Hastings ratio

$$\frac{P(\alpha|\alpha')}{P(\alpha'|\alpha)}$$

Probability of going back to previous state

Probability of proposing new state from previous

Acceptance ratio

Comparing the new state against the old state

Posterior ratio

$$\frac{P(\alpha'|x,M)}{P(\alpha|x,M)} = \underbrace{\frac{P(x|\alpha',M)}{P(x|\alpha,M)}}_{\text{Likelihood ratio}} \times \underbrace{\frac{P(\alpha'|M)}{P(\alpha|M)}}_{\text{Prior ratio}}$$

Hastings ratio

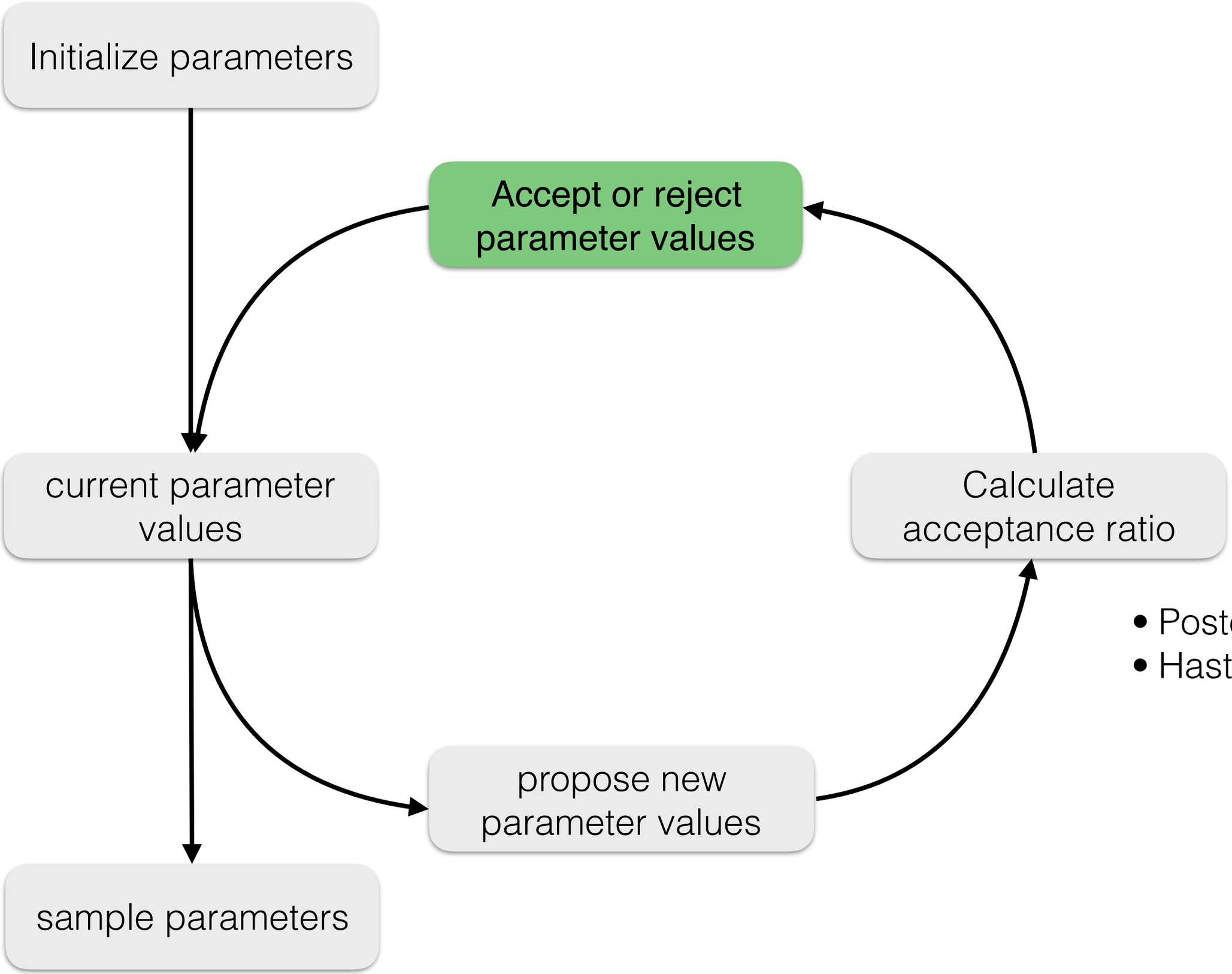
$$\frac{\cancel{P(\alpha|\alpha')}}{\cancel{P(\alpha'|\alpha)}}$$

Probability of going back to previous state

Probability of proposing new state from previous

For now let's assume it's constant (it is in most cases)

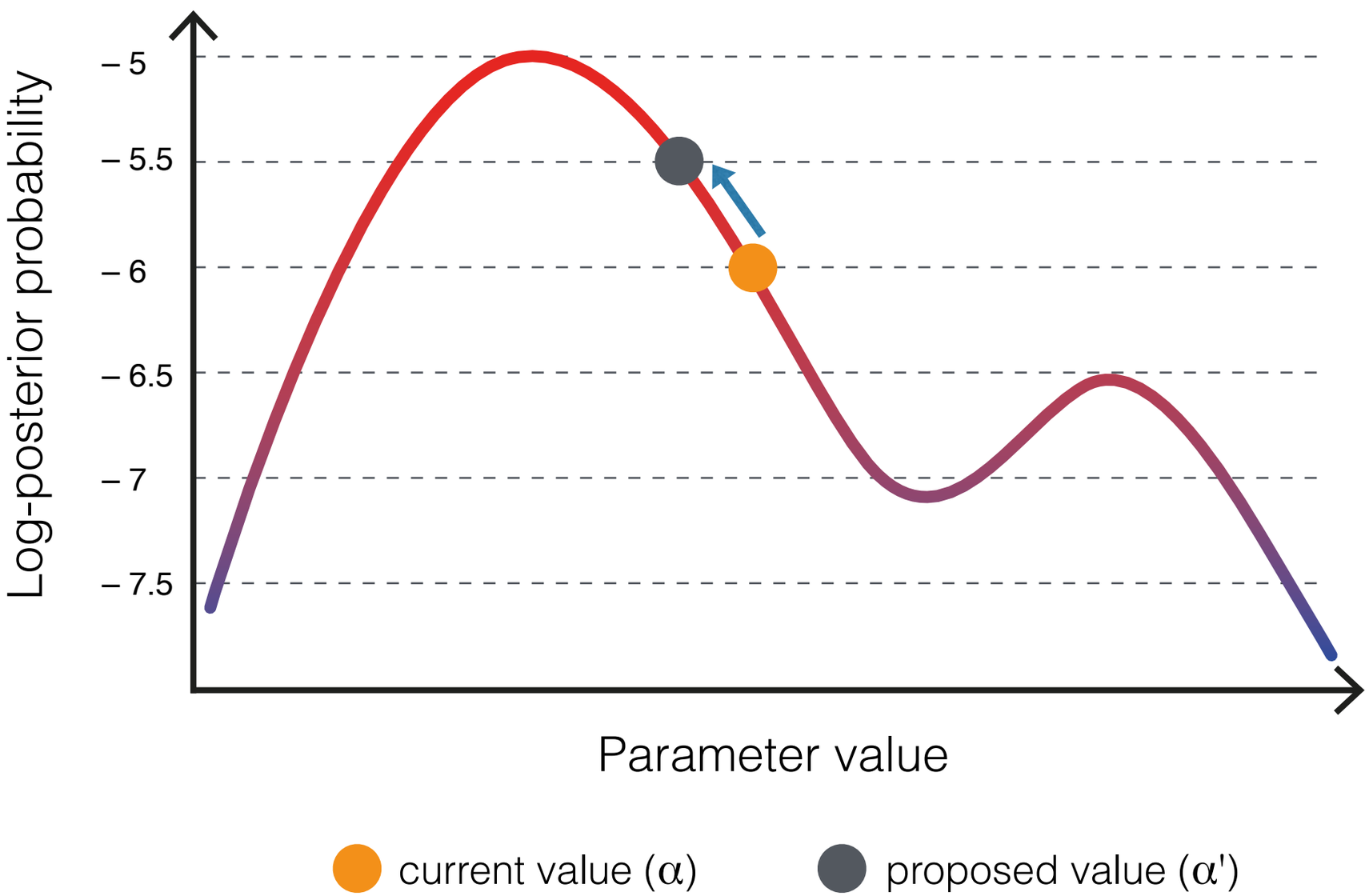
Metropolis-Hastings MCMC



- Posterior ratio
- Hastings ratio

Acceptance probability

$$r = \min\left\{1, \underbrace{\frac{P(\alpha'|x,M)}{P(\alpha|x,M)}}_{\text{Posterior ratio}} \times \underbrace{\frac{\cancel{P(\alpha|\alpha')}}{\cancel{P(\alpha'|\alpha)}}}_{\text{Hastings ratio}}\right\}$$

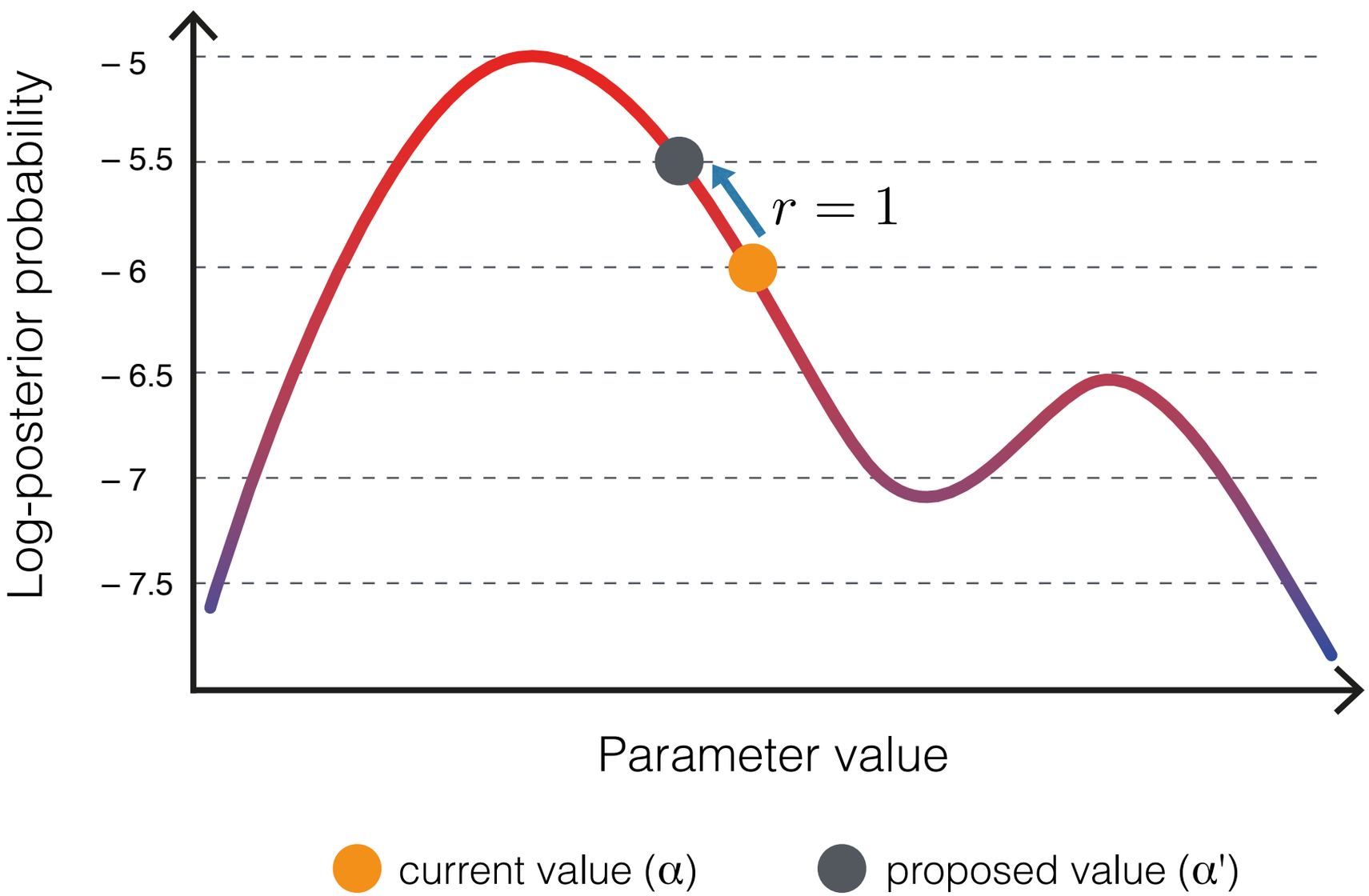


$\log P(\bullet|x,M) = -5.5$

$\log P(\bullet|x,M) = -6$

Acceptance probability

$$r = \min\left\{1, \underbrace{\frac{P(\alpha'|x,M)}{P(\alpha|x,M)}}_{\text{Posterior ratio}} \times \underbrace{\frac{\cancel{P(\alpha|\alpha')}}{\cancel{P(\alpha'|\alpha)}}}_{\text{Hastings ratio}}\right\}$$

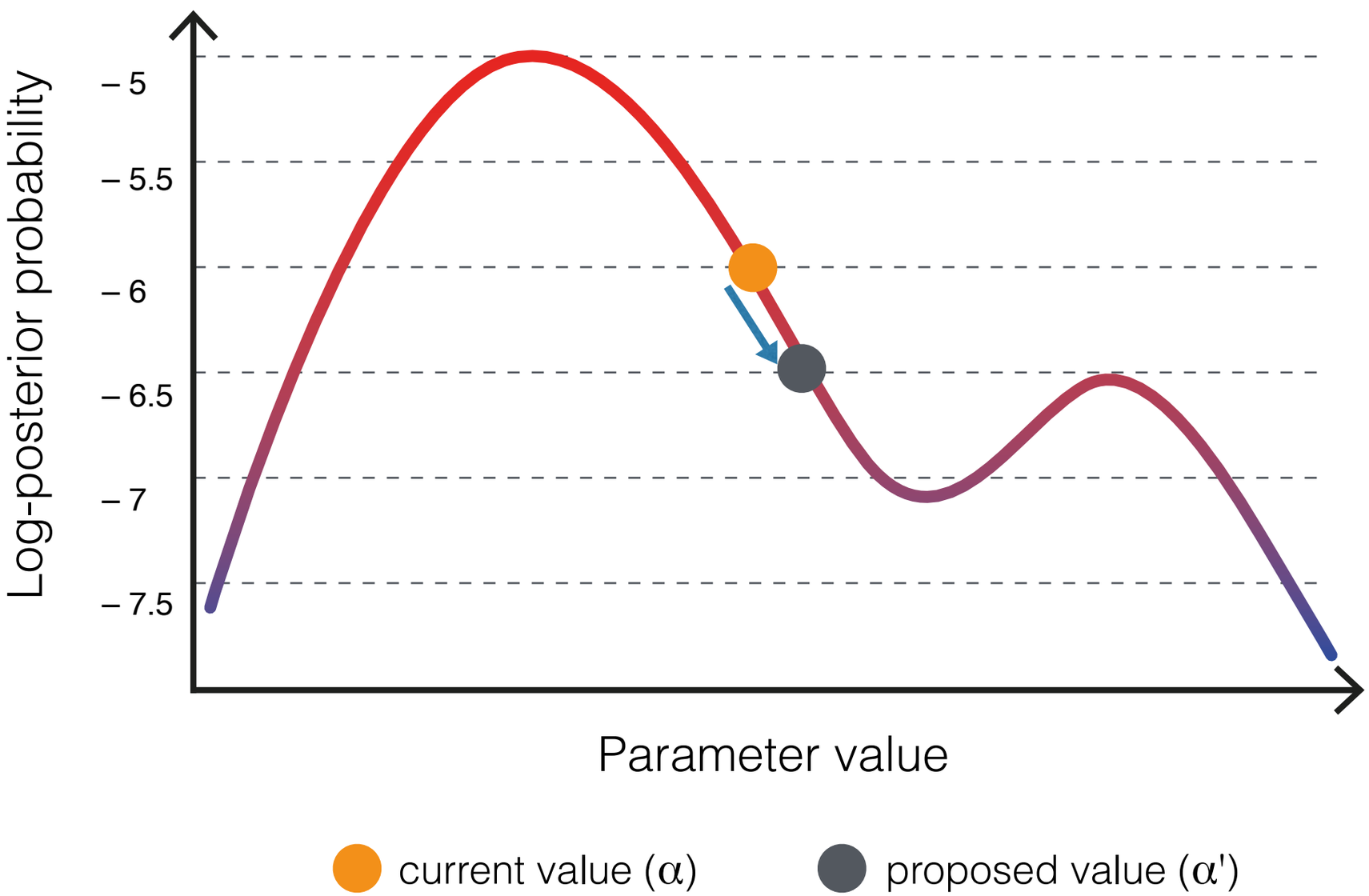


$\log P(\bullet|x, M) = -5.5$

$\log P(\bullet|x, M) = -6$

Acceptance probability

$$r = \min\left\{1, \underbrace{\frac{P(\alpha'|x,M)}{P(\alpha|x,M)}}_{\text{Posterior ratio}} \times \underbrace{\frac{\cancel{P(\alpha|\alpha')}}{\cancel{P(\alpha'|\alpha)}}}_{\text{Hastings ratio}}\right\}$$

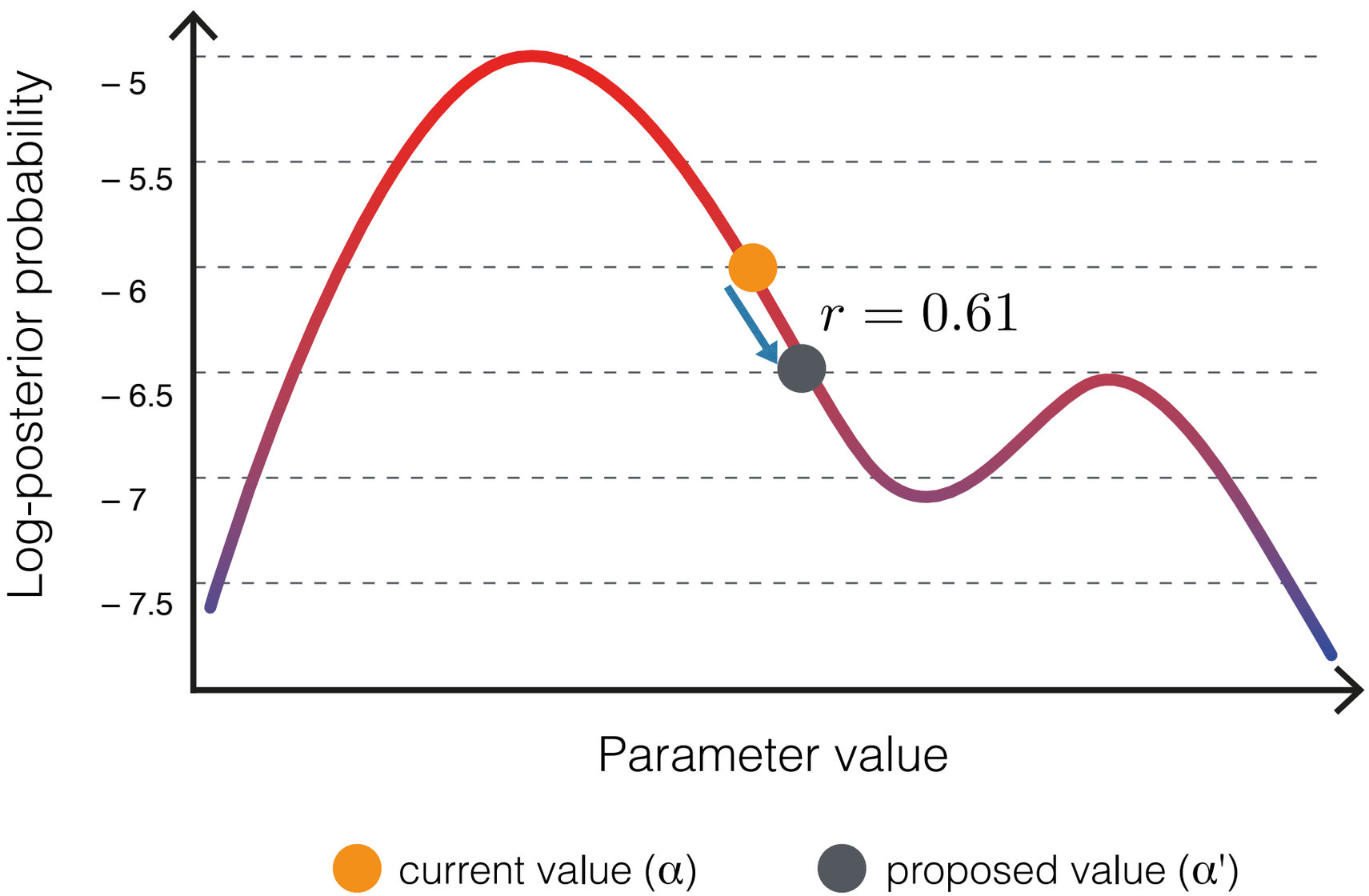


$\log P(\bullet|x,M) = -6$

$\log P(\bullet|x,M) = -6.5$

Acceptance probability

$$r = \min\left\{1, \underbrace{\frac{P(\alpha'|x,M)}{P(\alpha|x,M)}}_{\text{Posterior ratio}} \times \underbrace{\frac{\cancel{P(\alpha|\alpha')}}{\cancel{P(\alpha'|\alpha)}}}_{\text{Hastings ratio}}\right\}$$

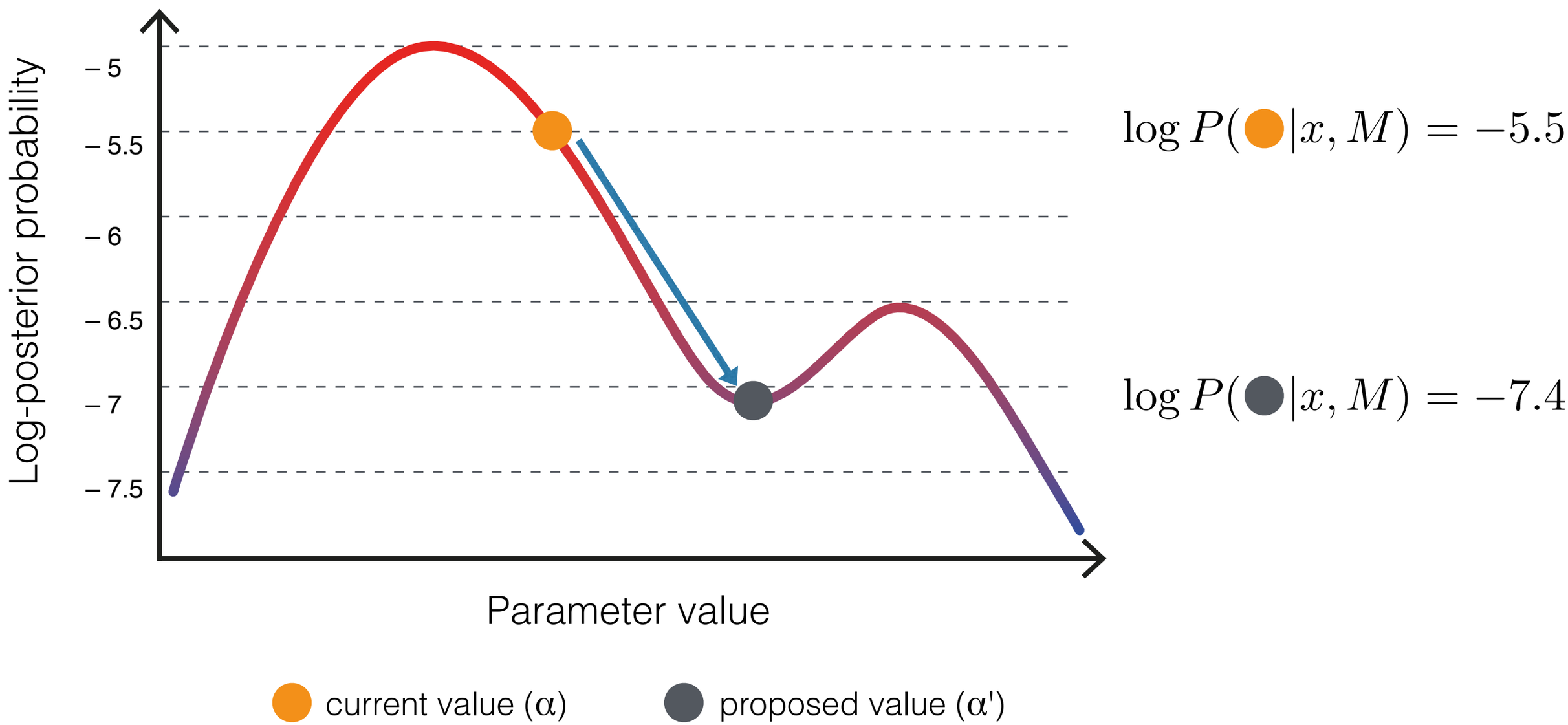


$\log P(\bullet|x,M) = -6$

$\log P(\bullet|x,M) = -6.5$

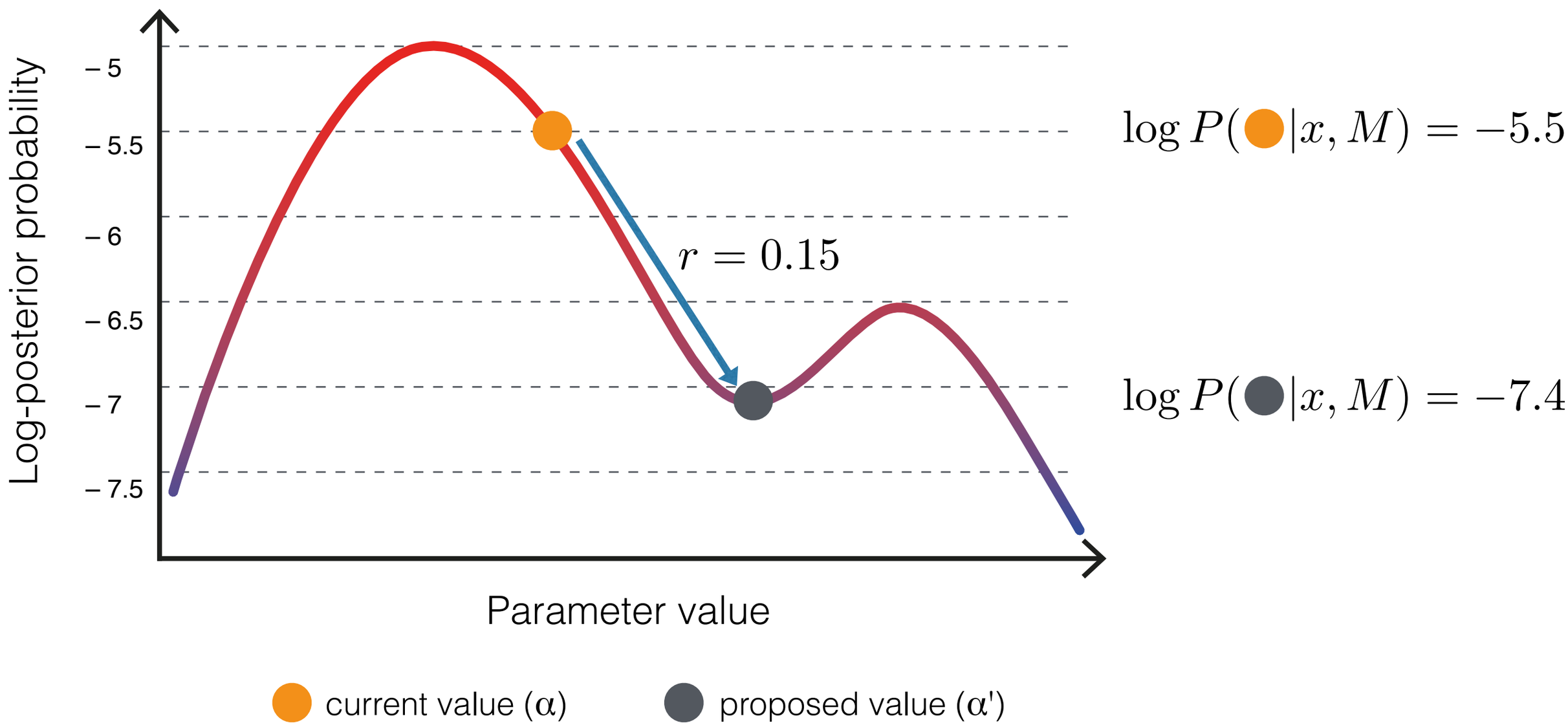
Acceptance probability

$$r = \min\left\{1, \underbrace{\frac{P(\alpha'|x,M)}{P(\alpha|x,M)}}_{\text{Posterior ratio}} \times \underbrace{\frac{\cancel{P(\alpha|\alpha')}}{\cancel{P(\alpha'|\alpha)}}}_{\text{Hastings ratio}}\right\}$$

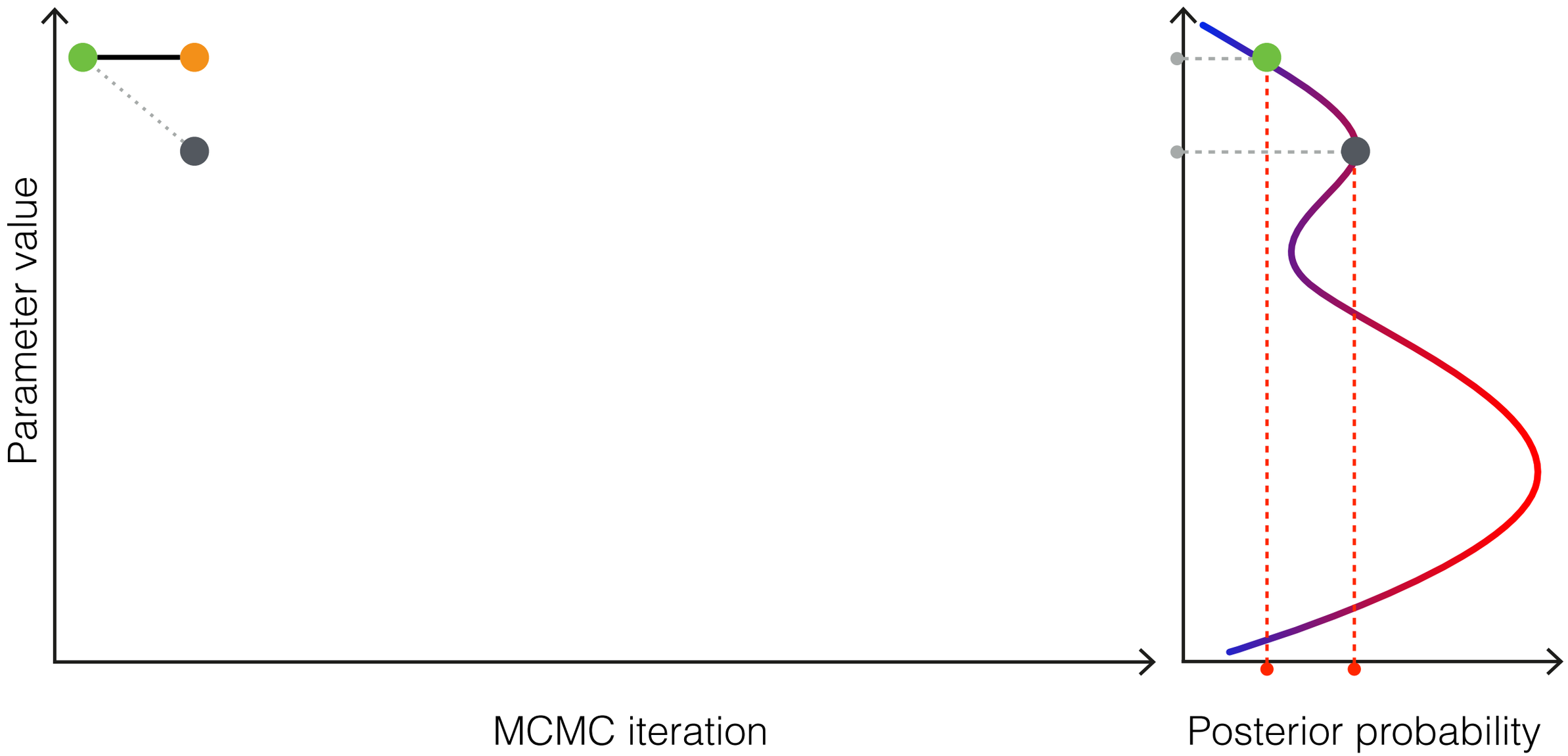


Acceptance probability

$$r = \min\left\{1, \underbrace{\frac{P(\alpha'|x,M)}{P(\alpha|x,M)}}_{\text{Posterior ratio}} \times \underbrace{\frac{\cancel{P(\alpha|\alpha')}}{\cancel{P(\alpha'|\alpha)}}}_{\text{Hastings ratio}}\right\}$$

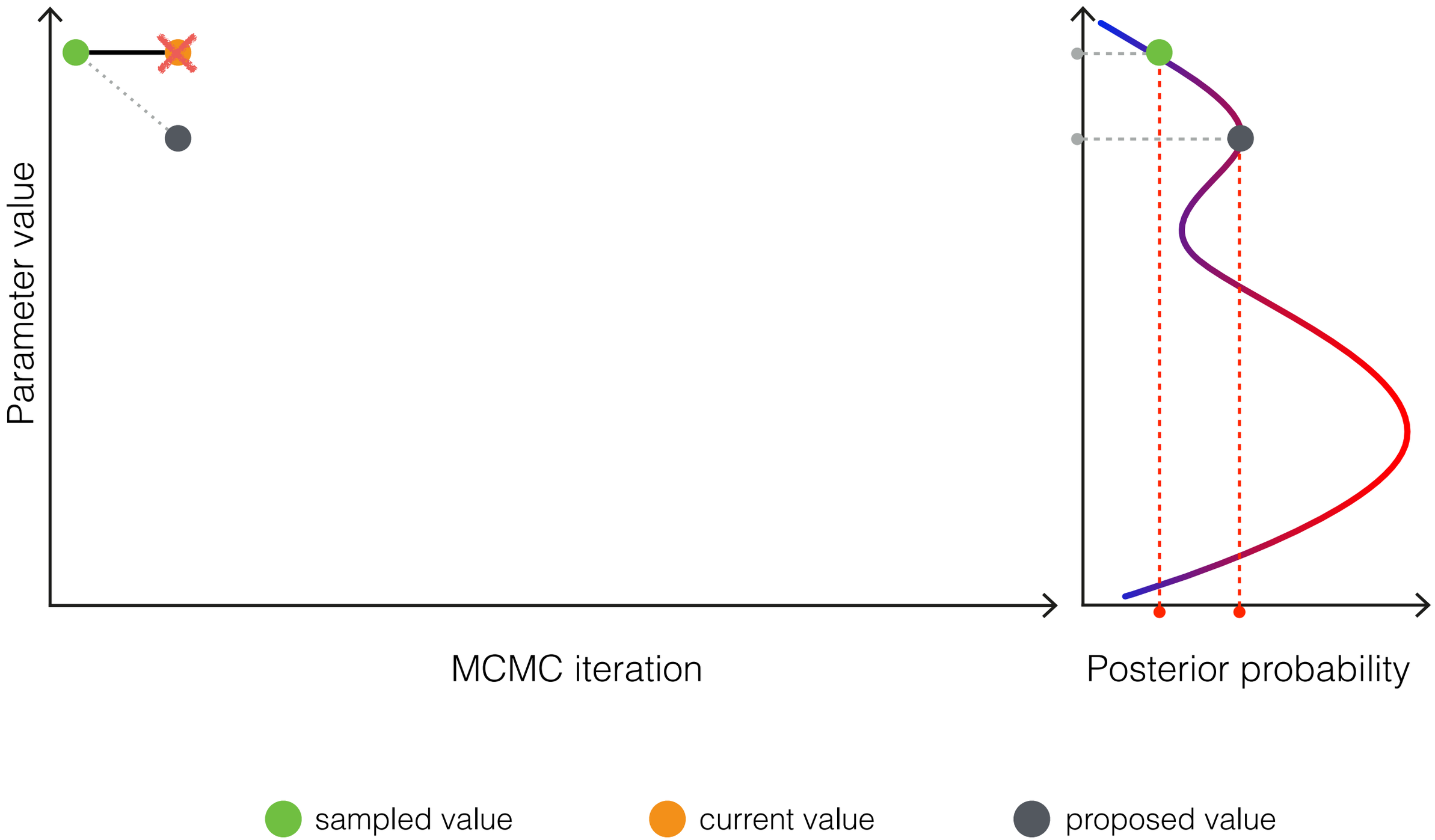


Sampling from the posterior using MCMC

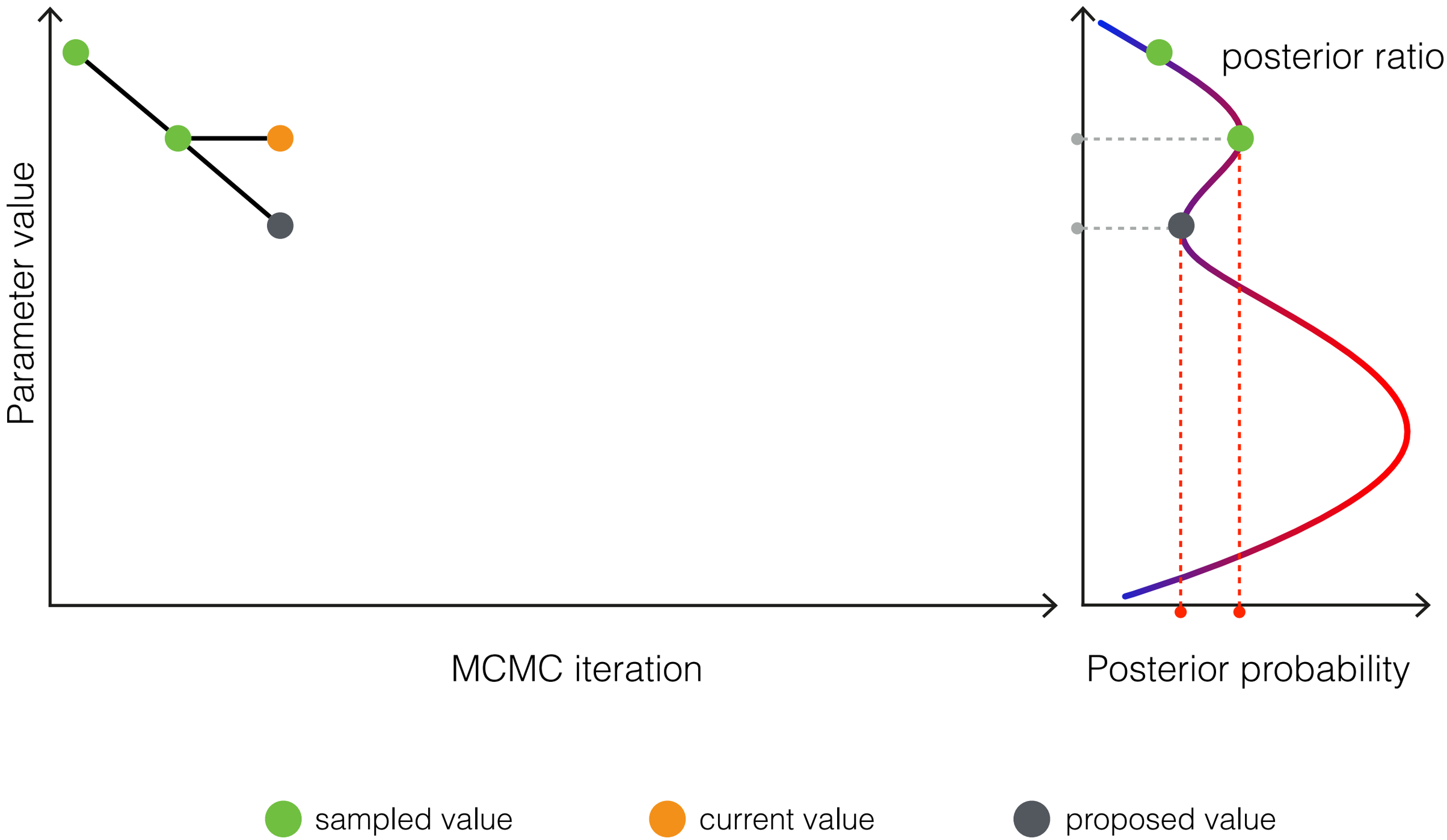


sampled value current value proposed value

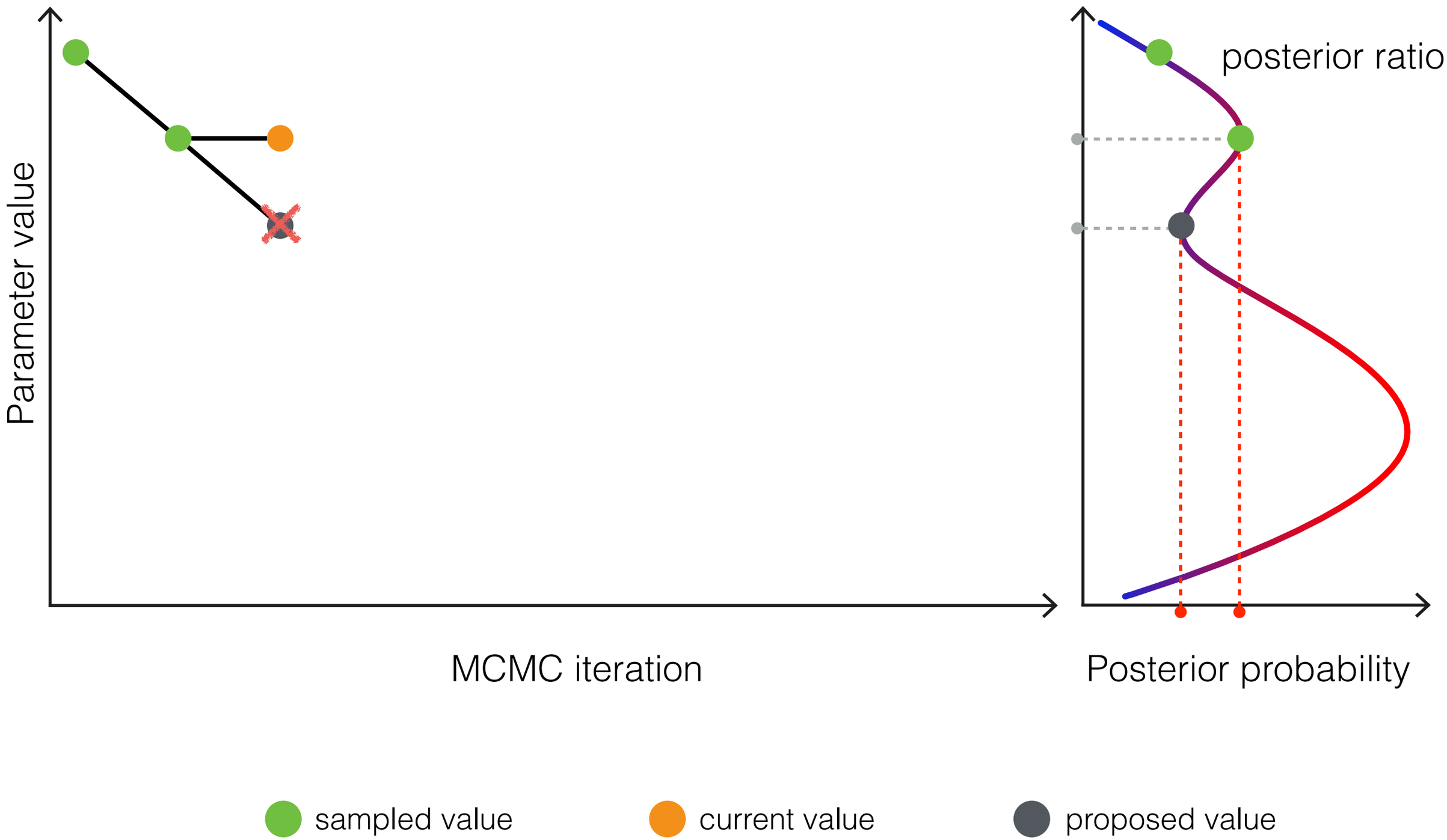
Sampling from the posterior using MCMC



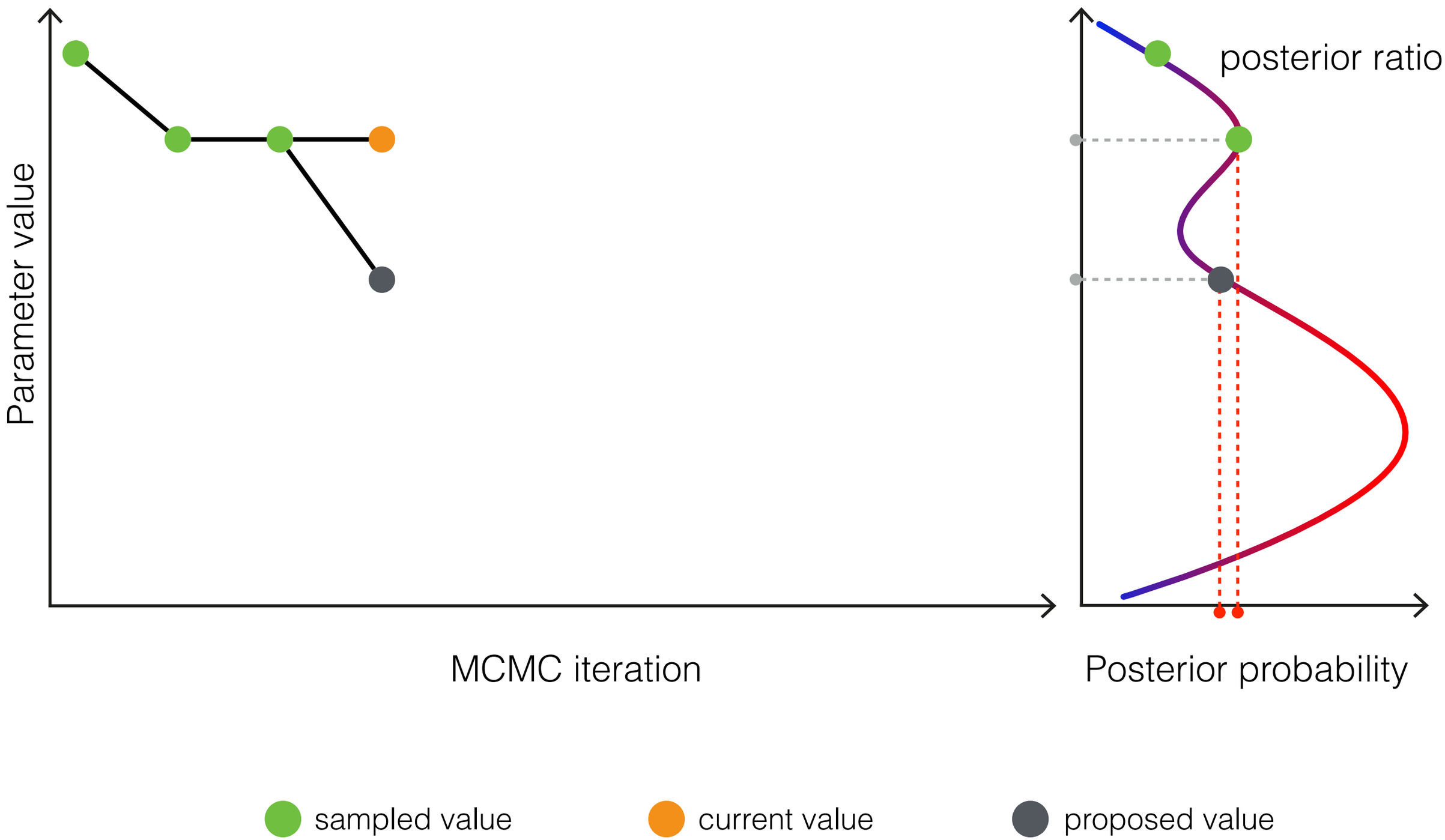
Sampling from the posterior using MCMC



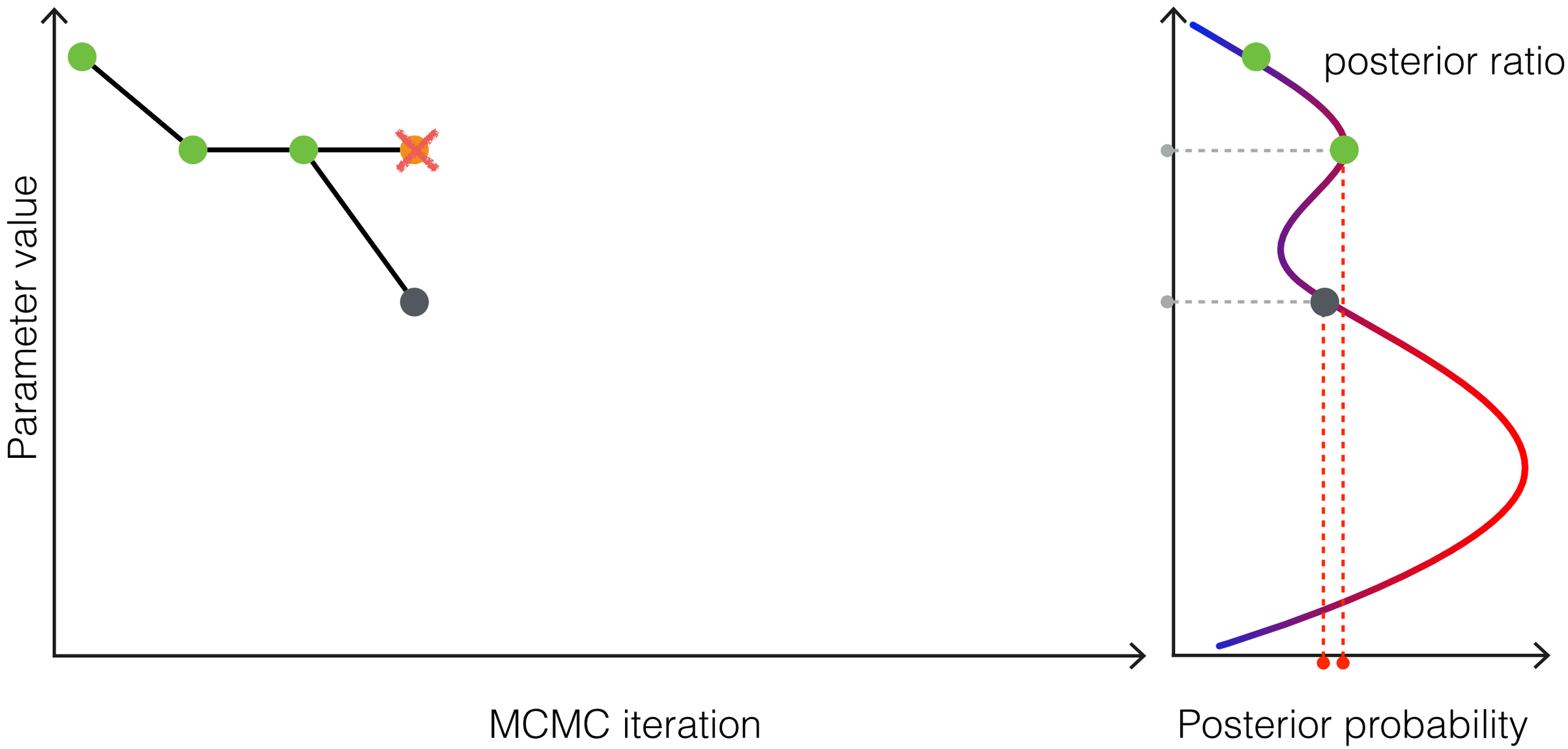
Sampling from the posterior using MCMC



Sampling from the posterior using MCMC

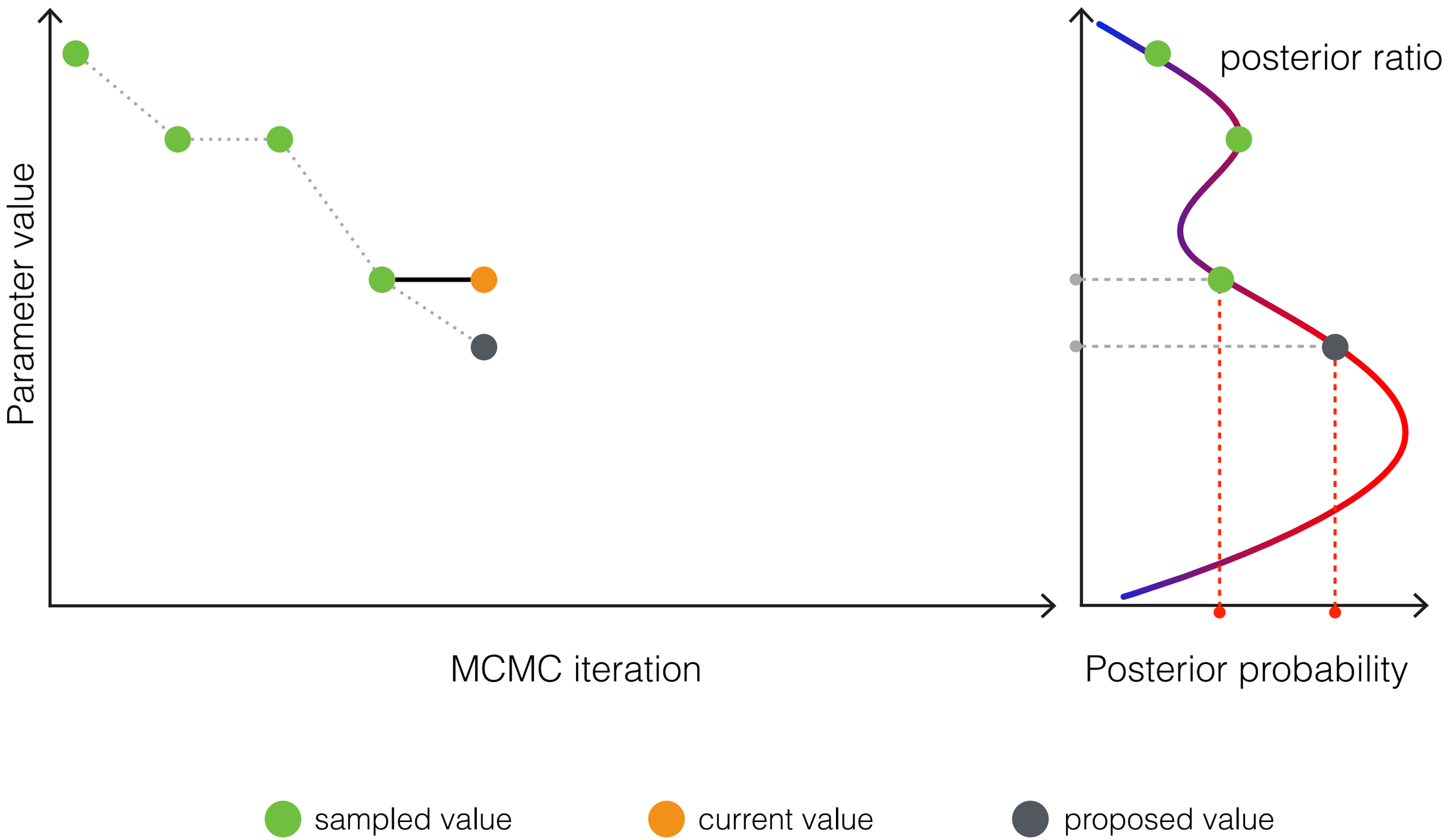


Sampling from the posterior using MCMC

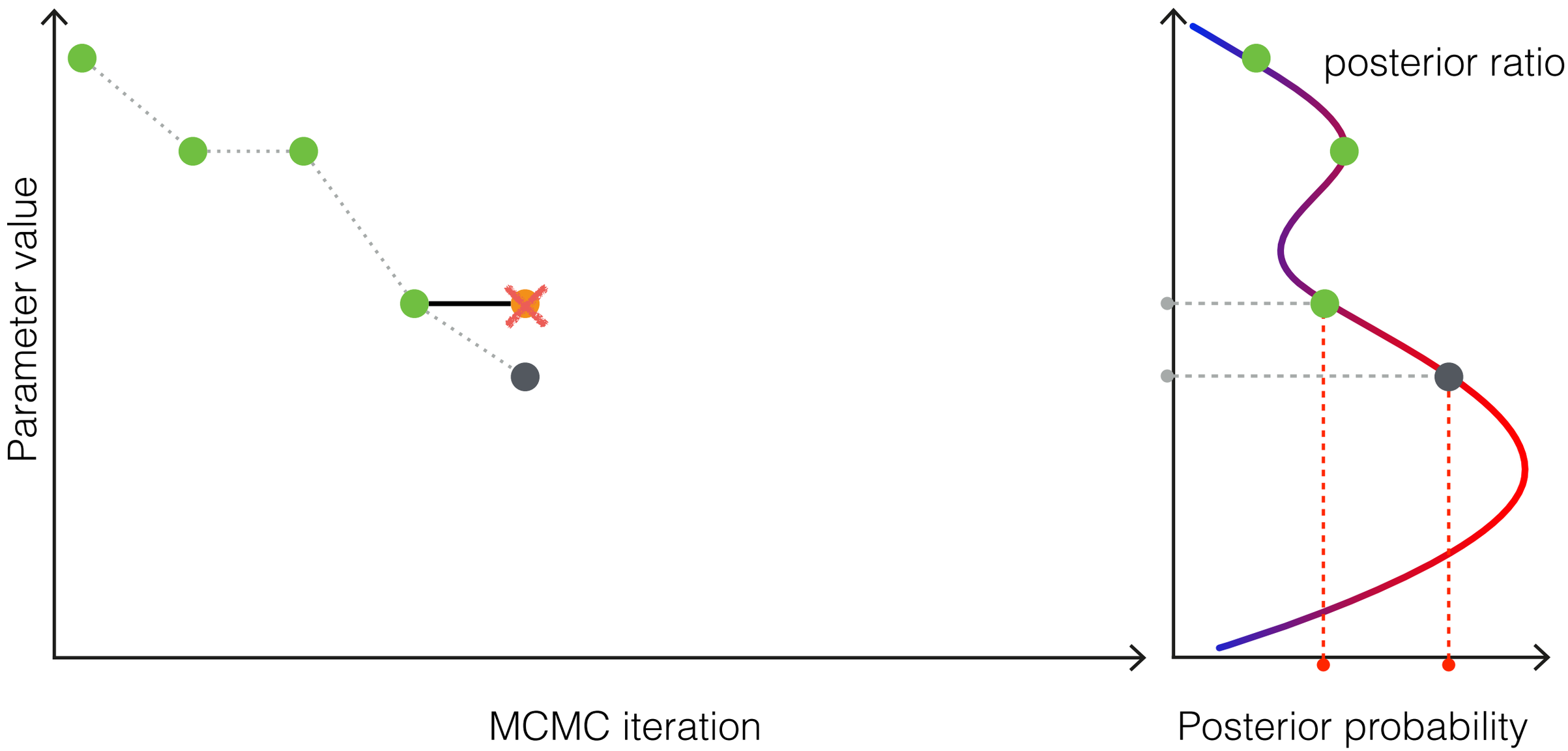


sampled value current value proposed value

Sampling from the posterior using MCMC

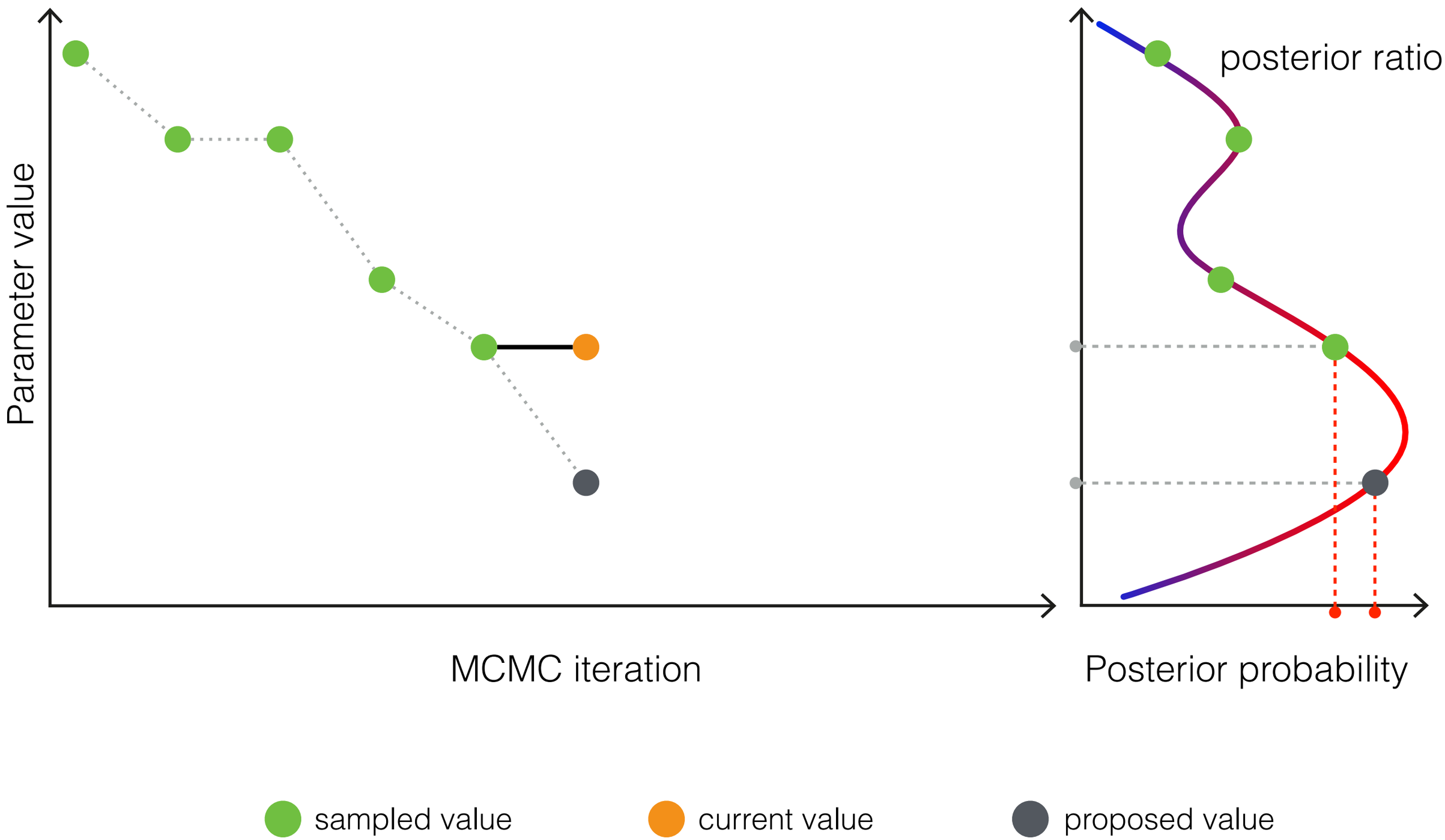


Sampling from the posterior using MCMC

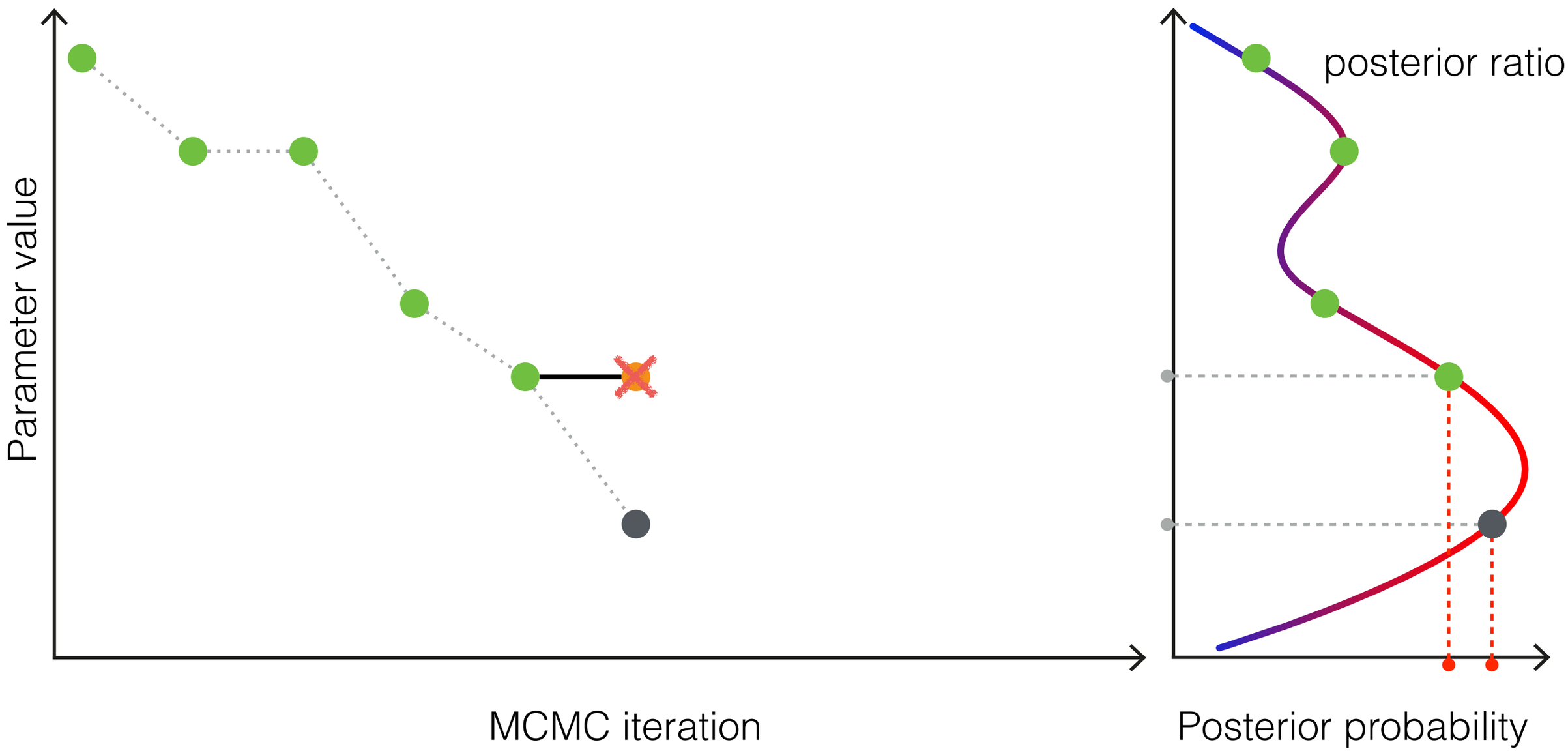


sampled value current value proposed value

Sampling from the posterior using MCMC

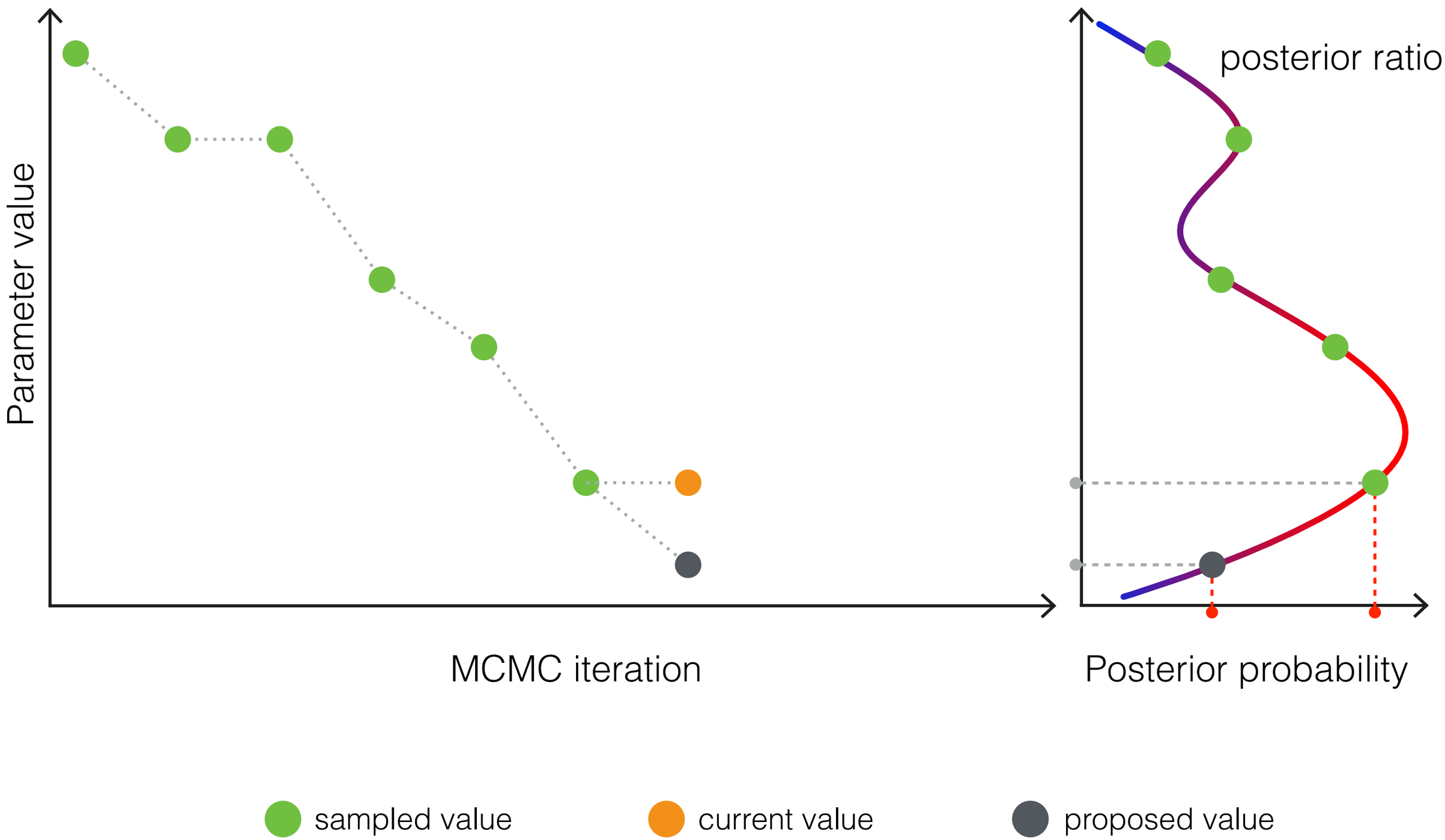


Sampling from the posterior using MCMC

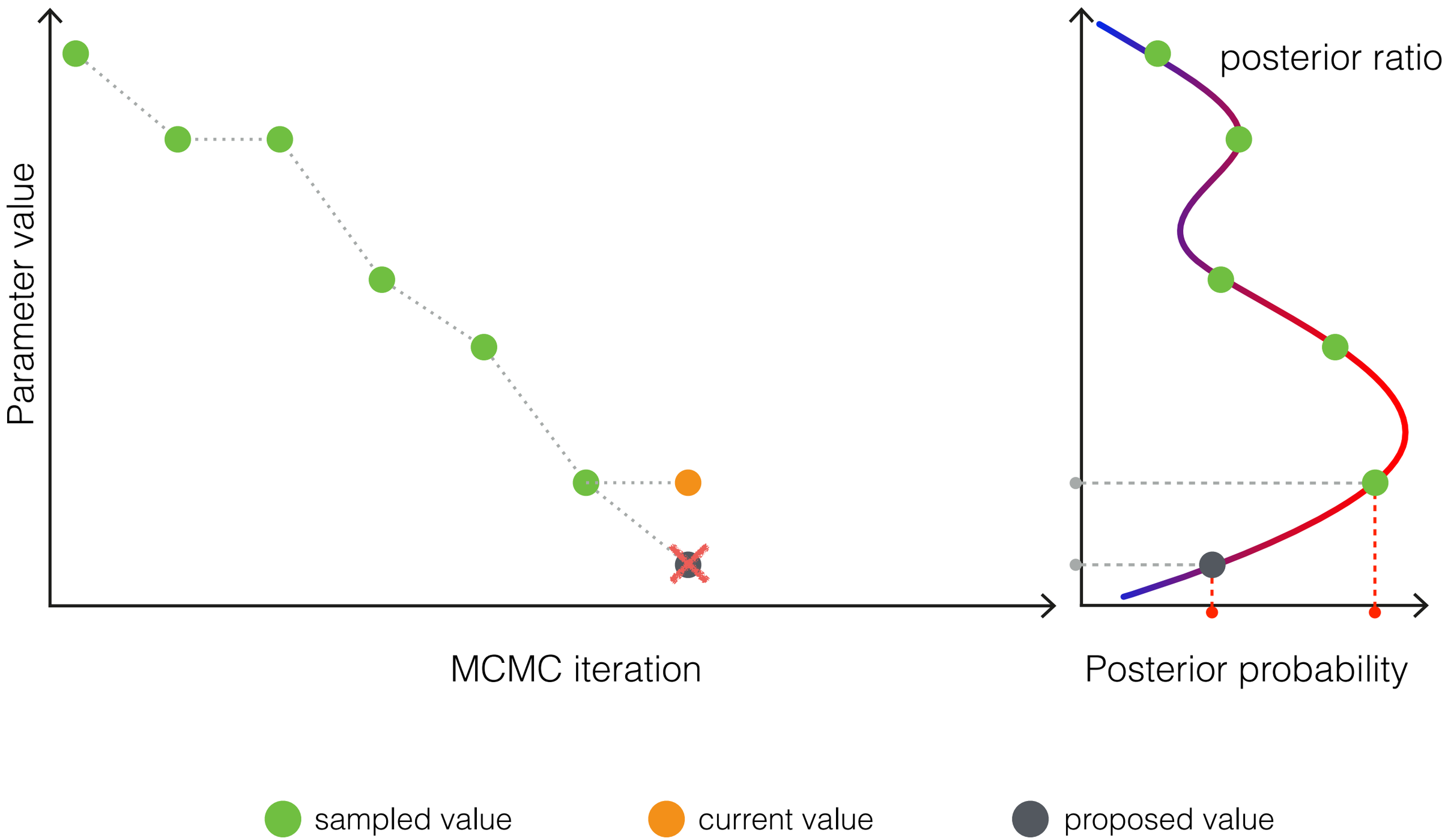


sampled value current value proposed value

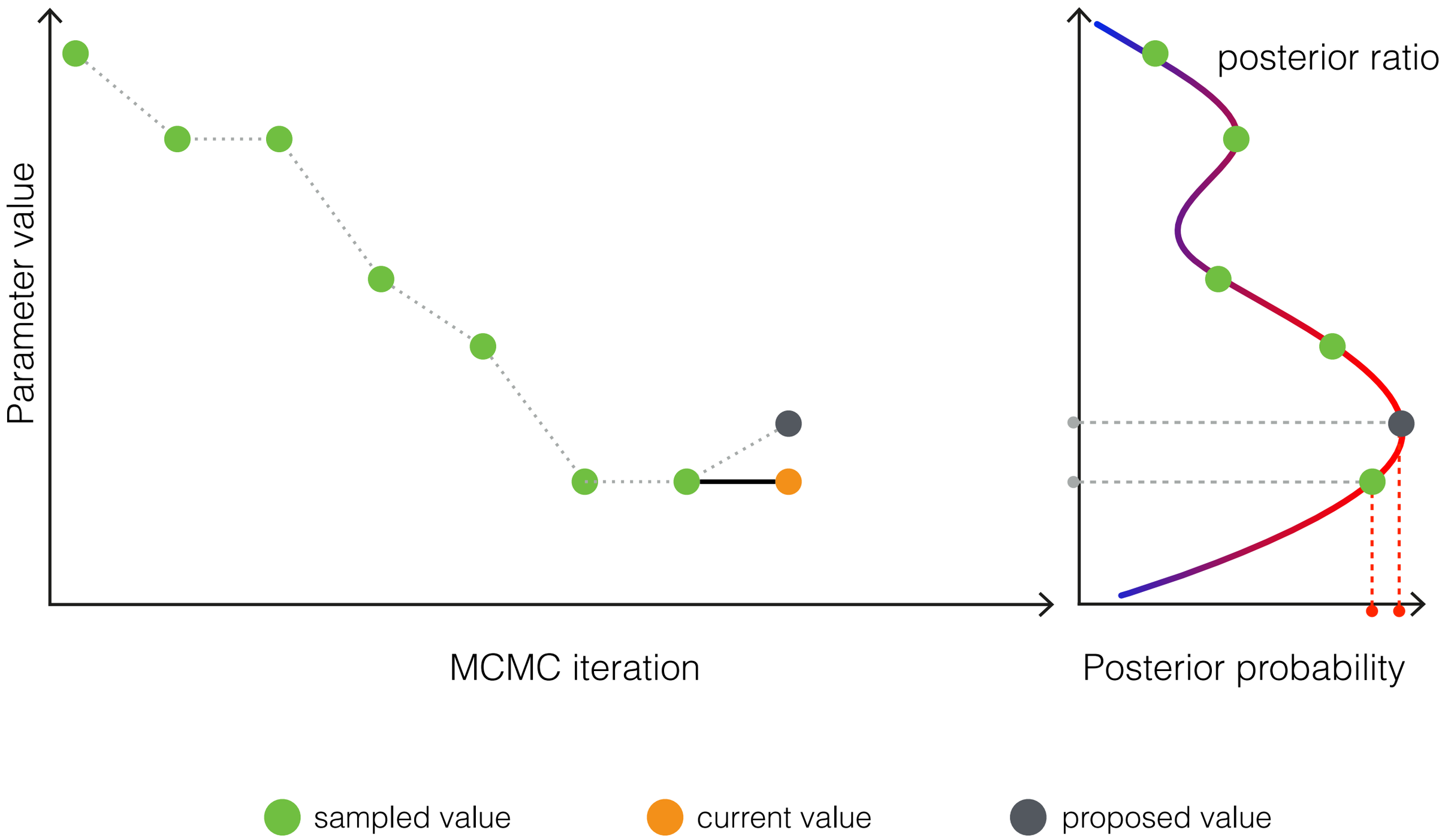
Sampling from the posterior using MCMC



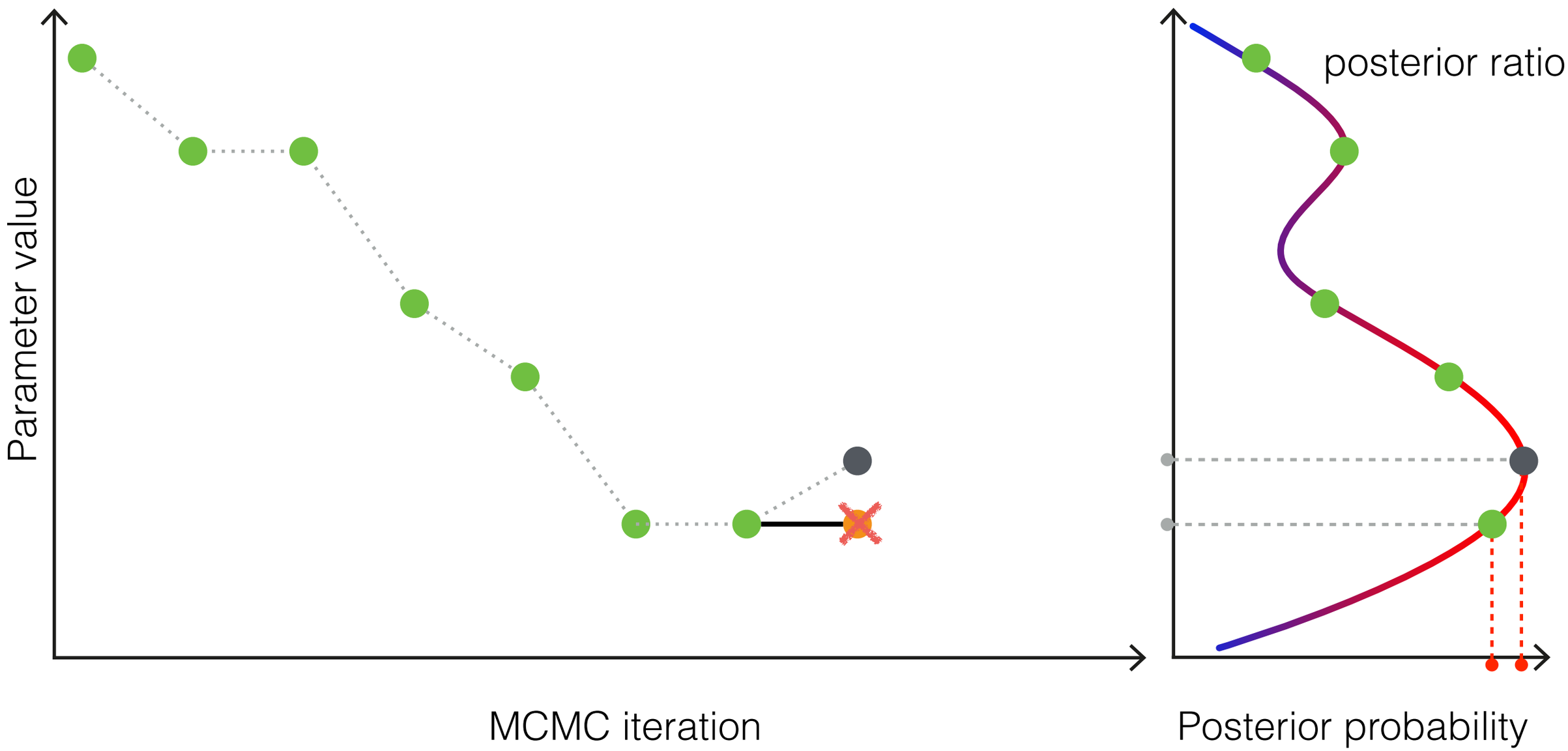
Sampling from the posterior using MCMC



Sampling from the posterior using MCMC

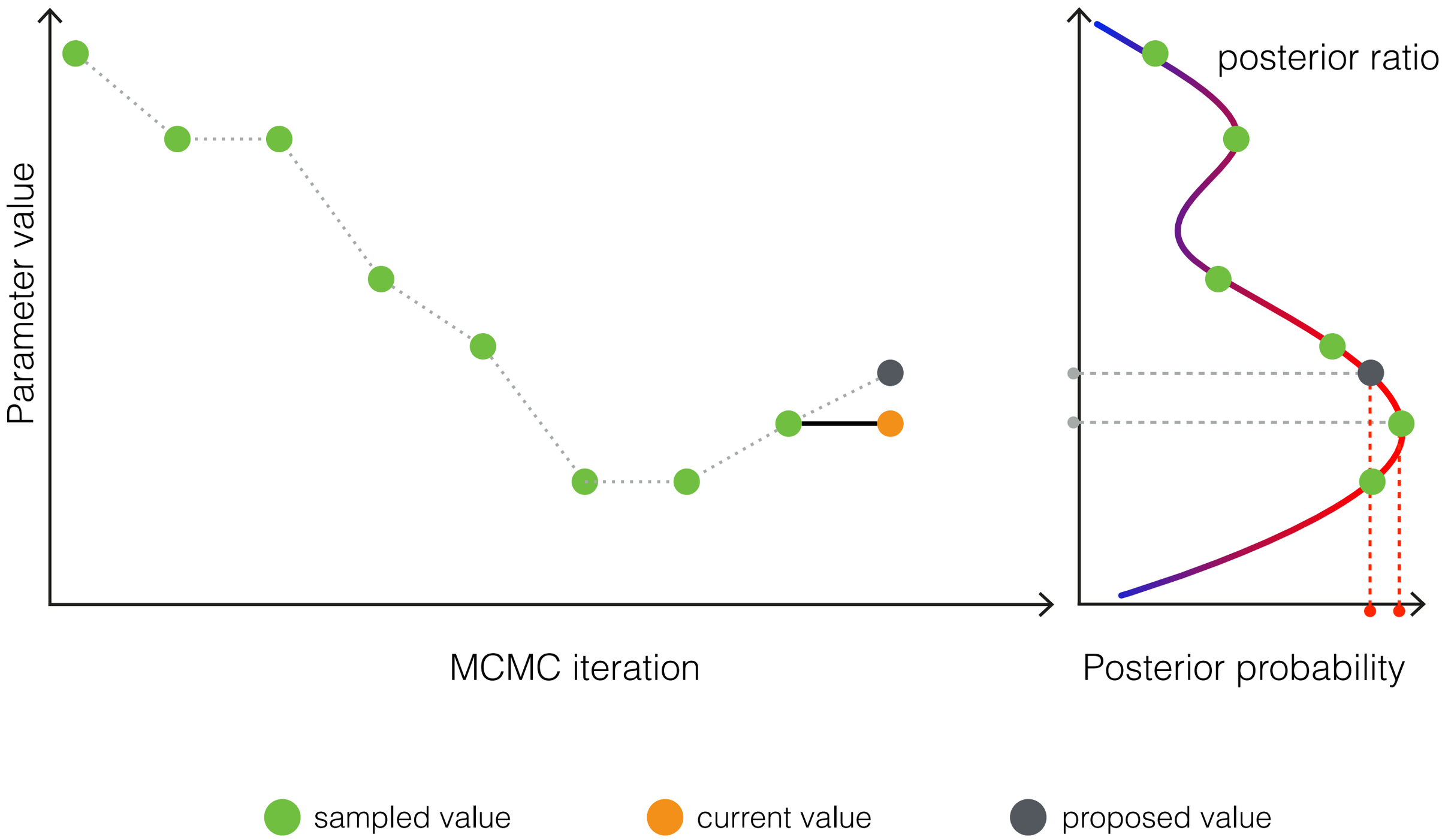


Sampling from the posterior using MCMC

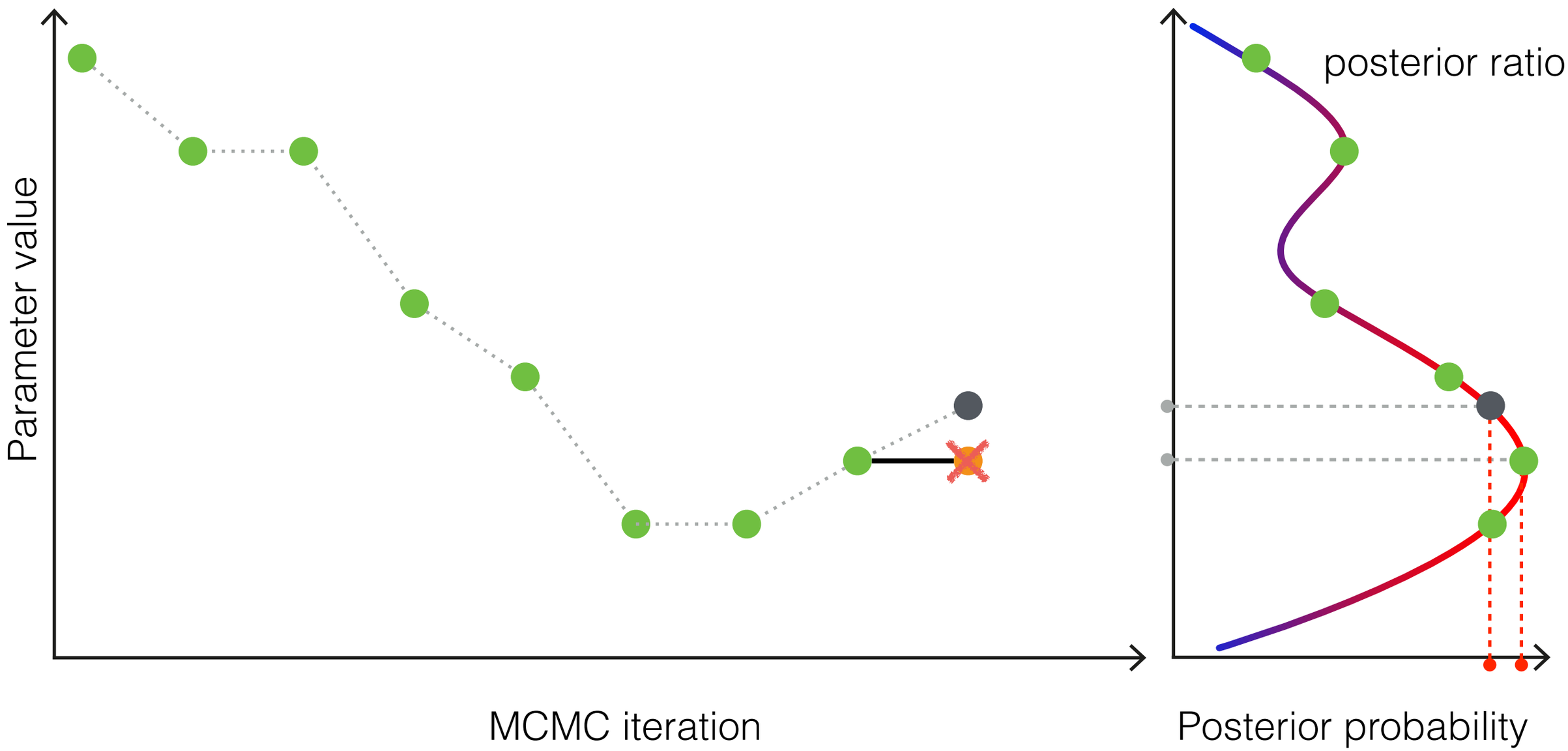


sampled value current value proposed value

Sampling from the posterior using MCMC

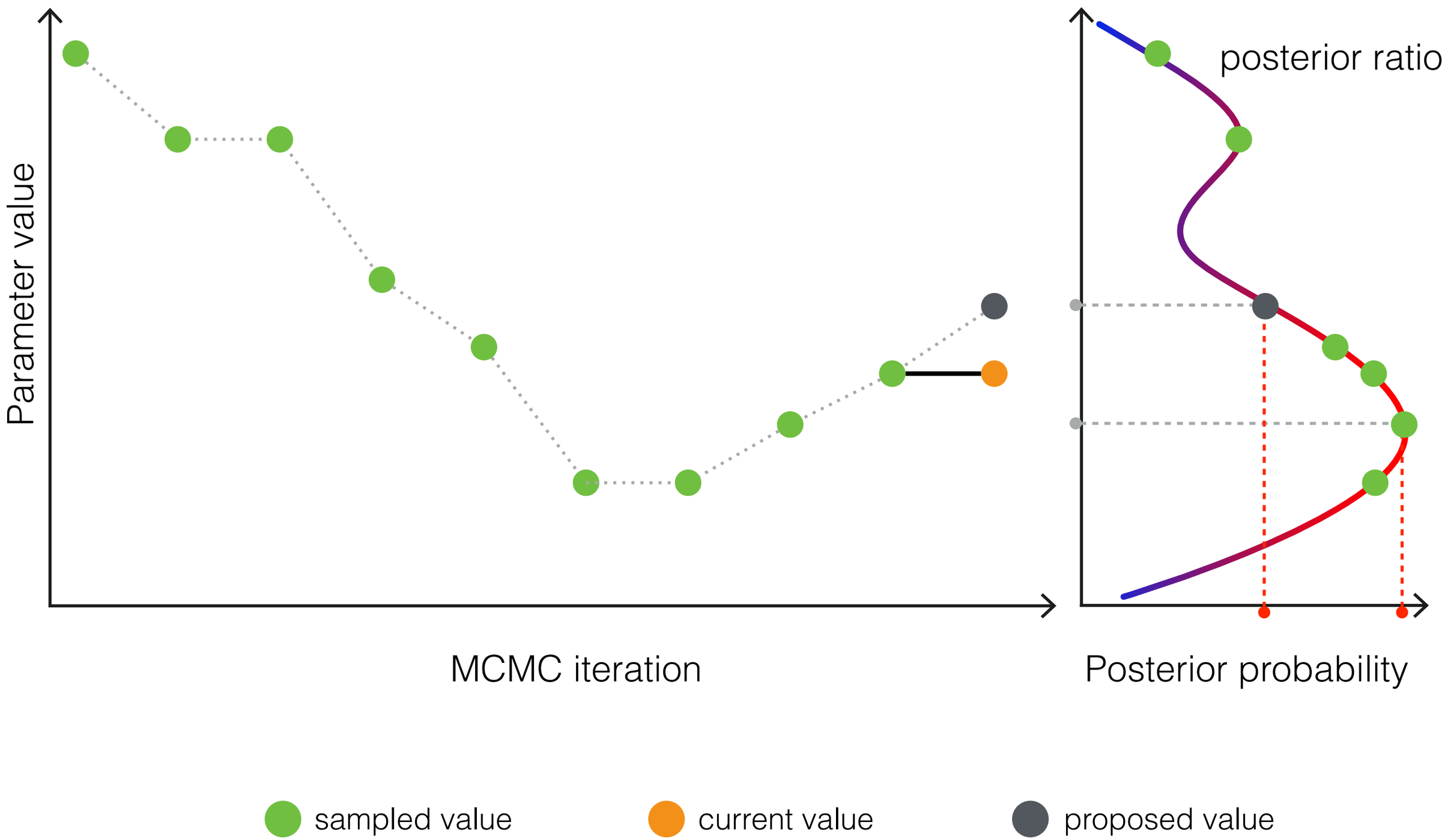


Sampling from the posterior using MCMC

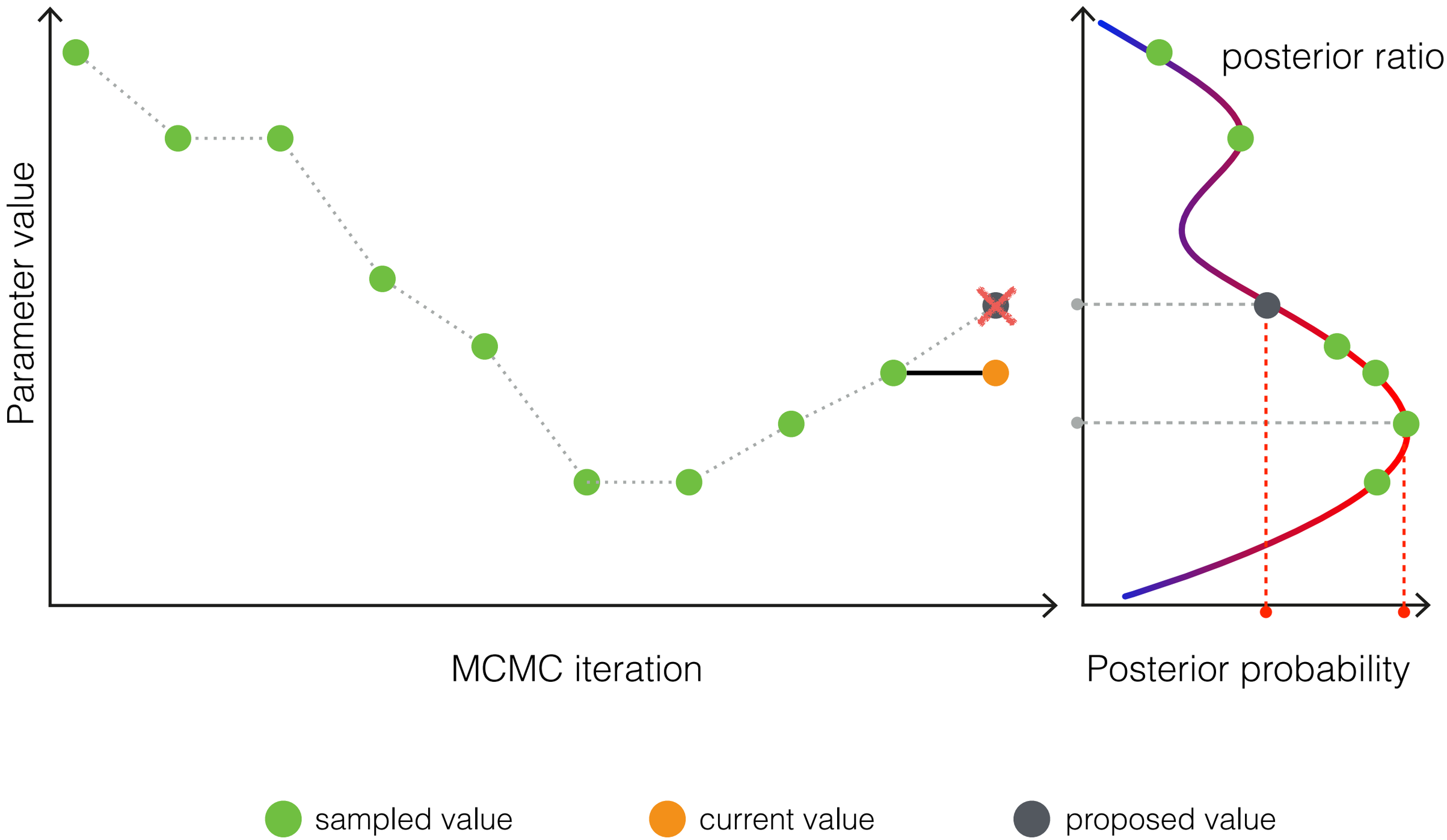


sampled value current value proposed value

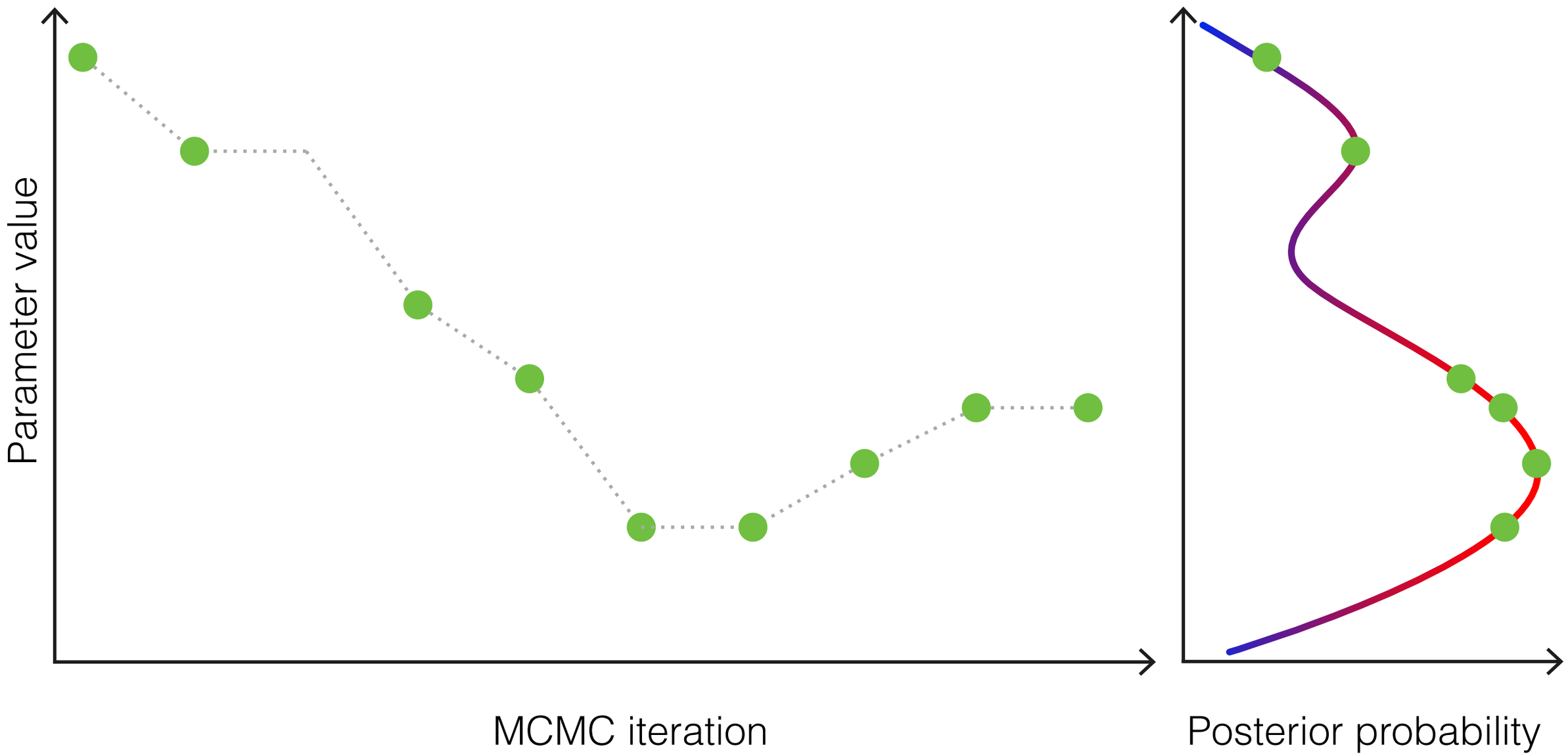
Sampling from the posterior using MCMC



Sampling from the posterior using MCMC

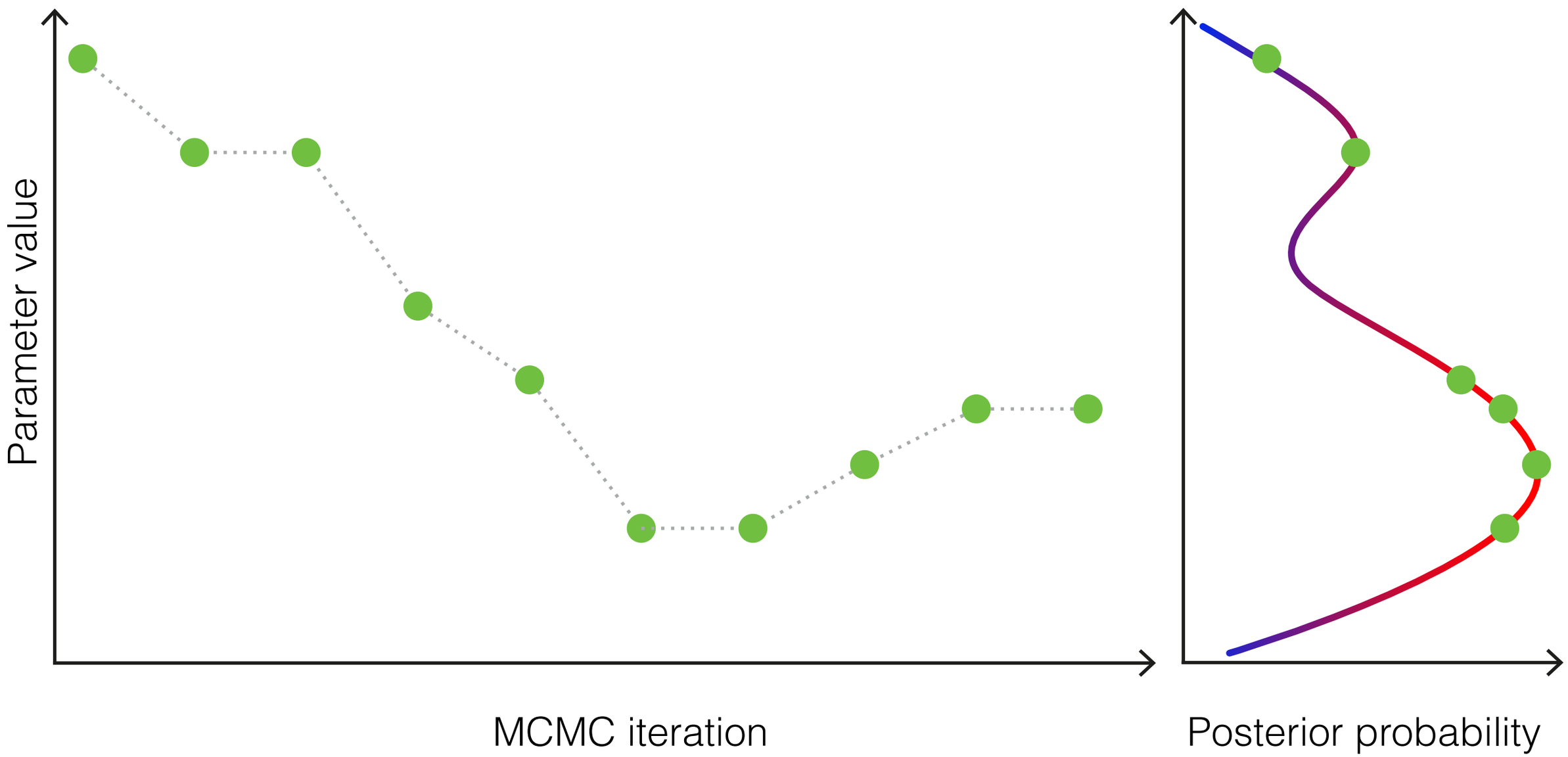


Sampling from the posterior using MCMC



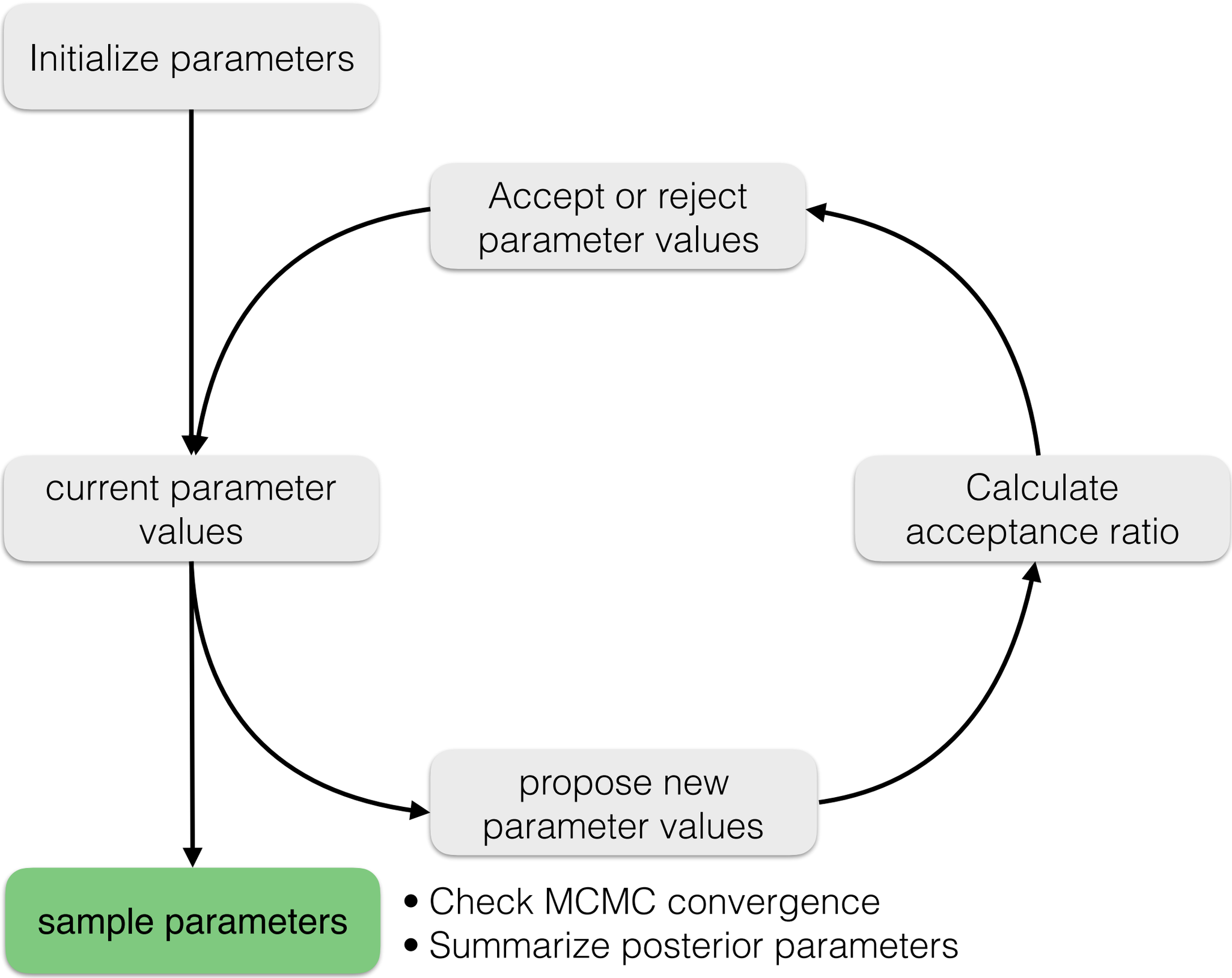
sampled value current value proposed value

Sampling from the posterior using MCMC



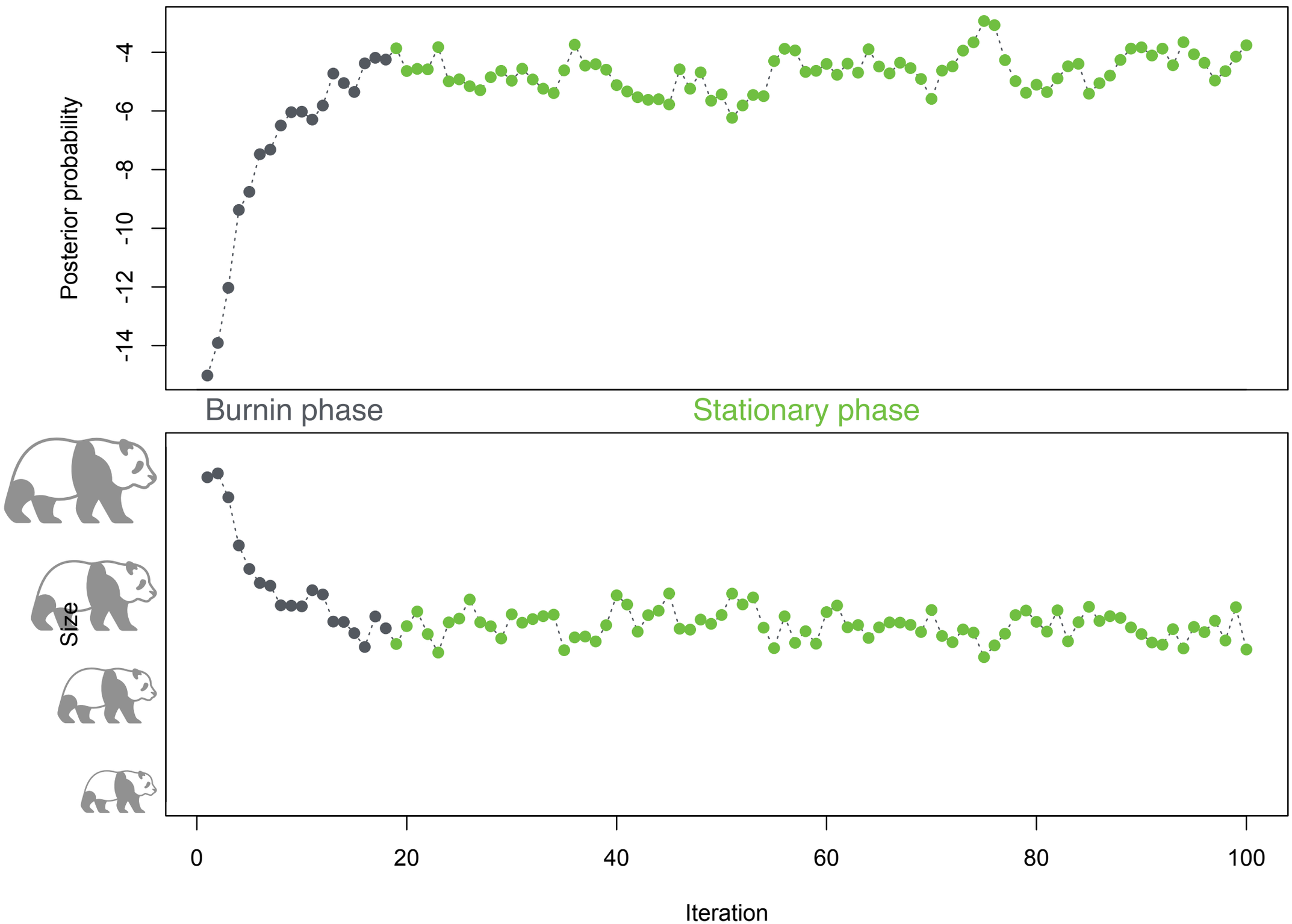
sampled value current value proposed value

Metropolis-Hastings MCMC

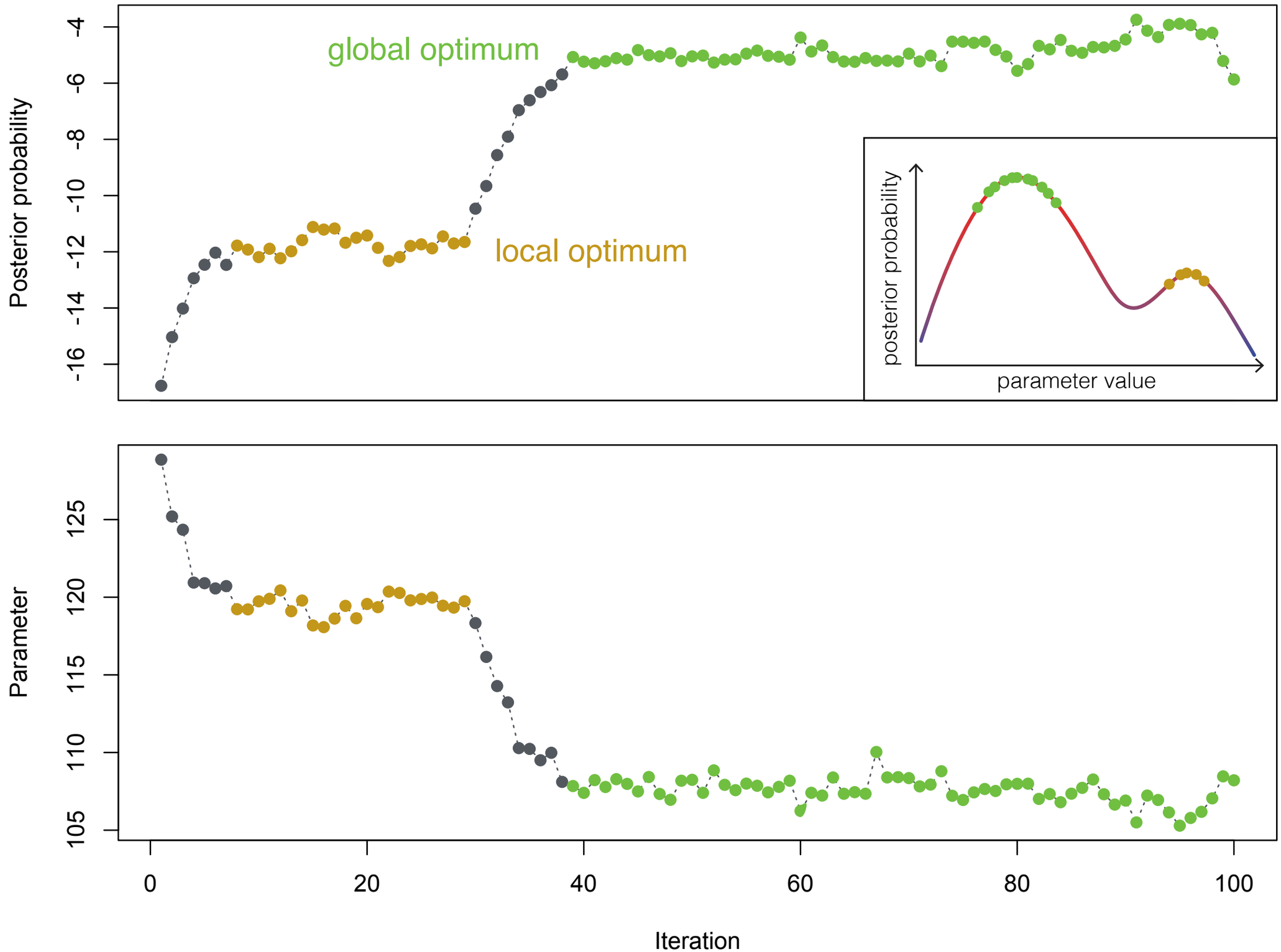


- Check MCMC convergence
- Summarize posterior parameters

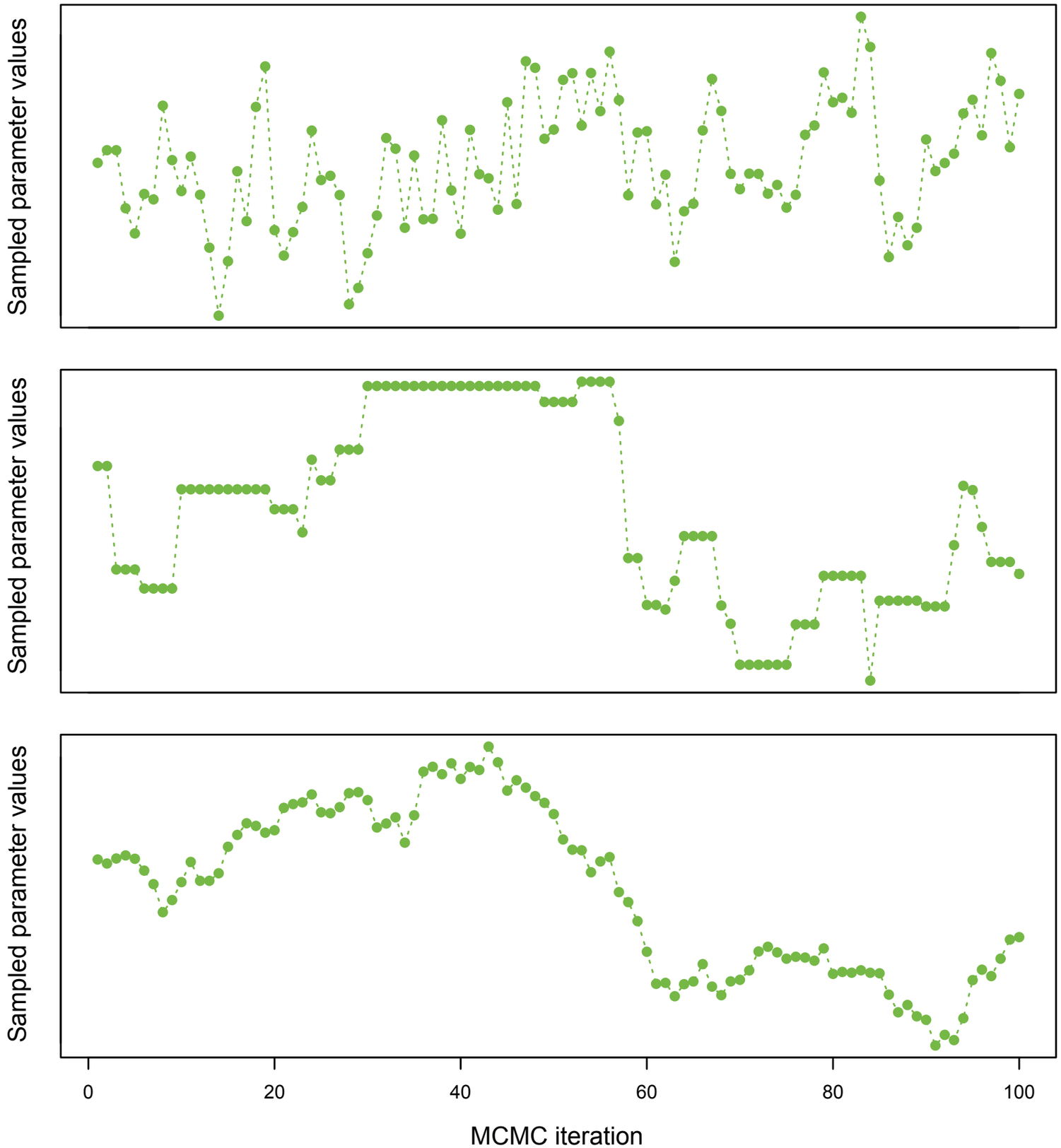
Burn-in and MCMC convergence



Burn-in and MCMC convergence



Posterior samples: Effective Sample Size (ESS)

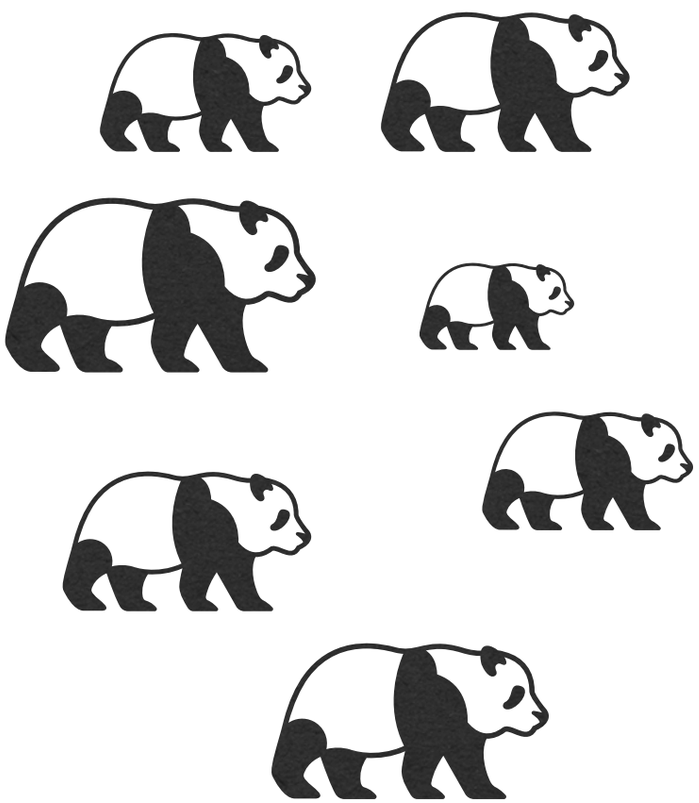


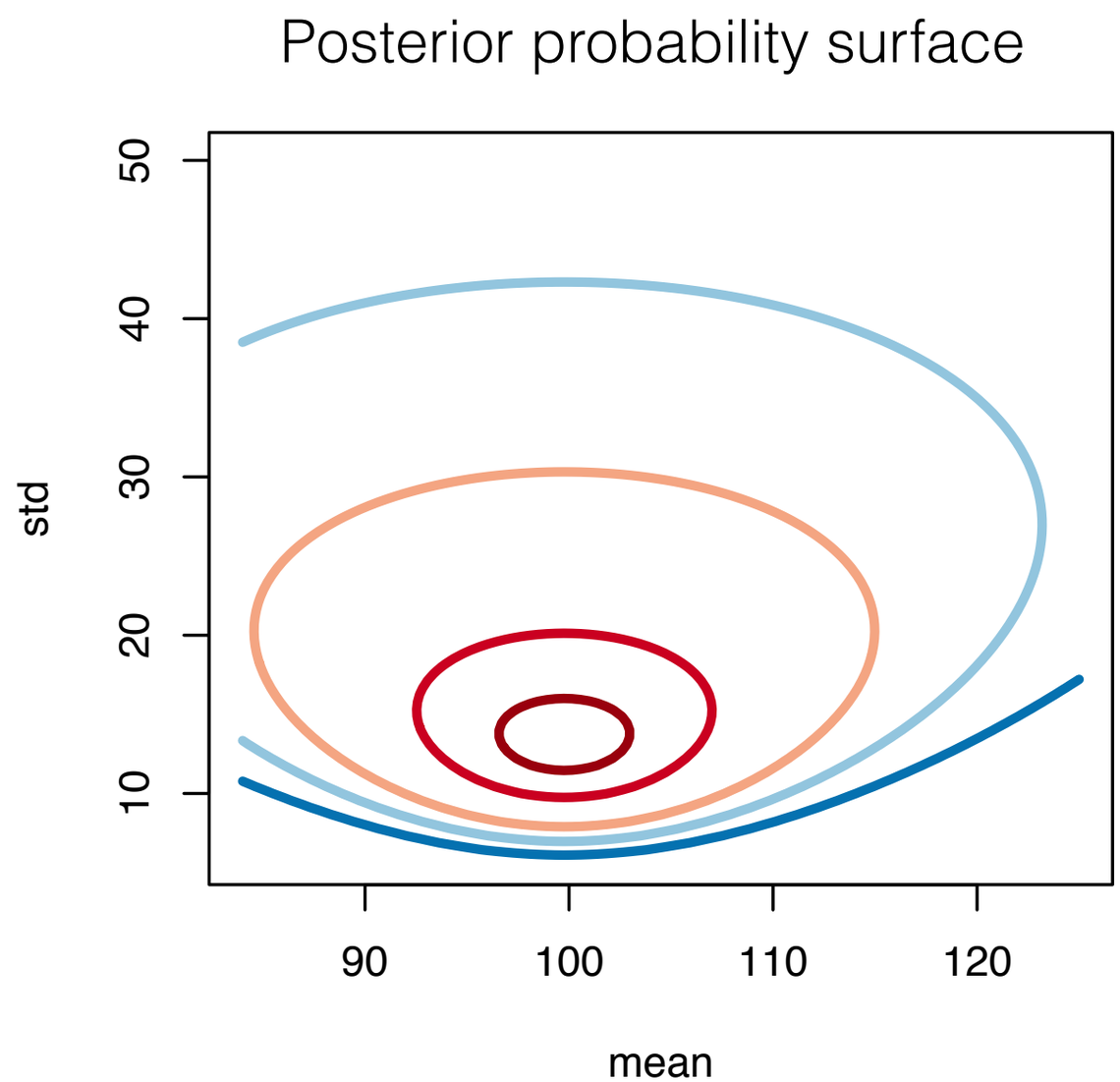
High ESS

Low ESS

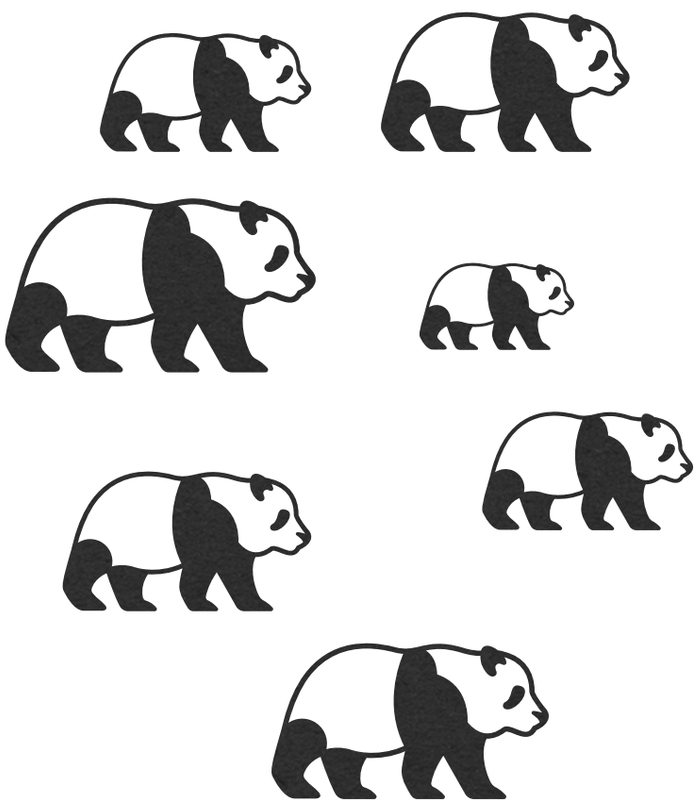
Low ESS

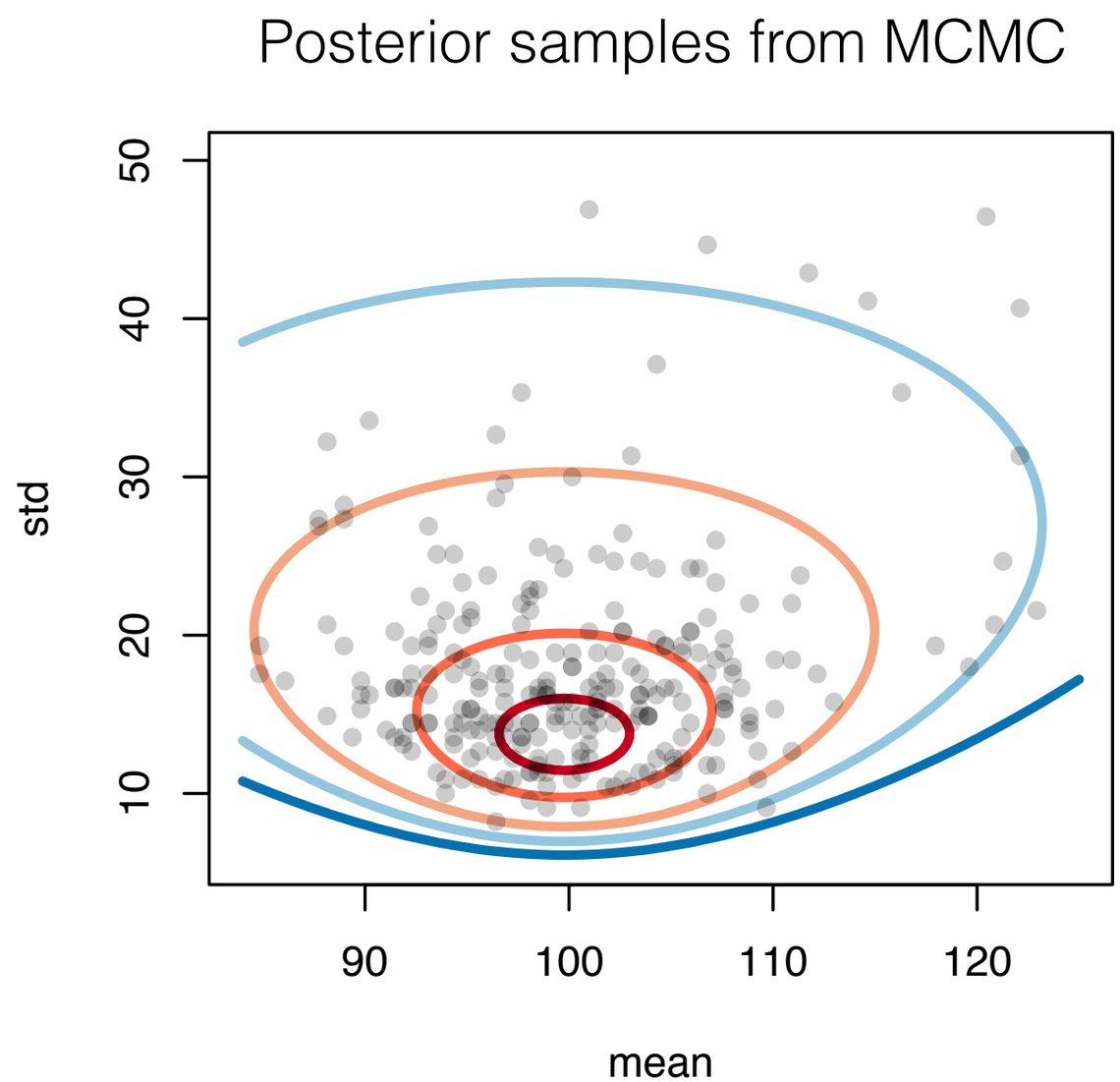
Summarizing MCMC samples



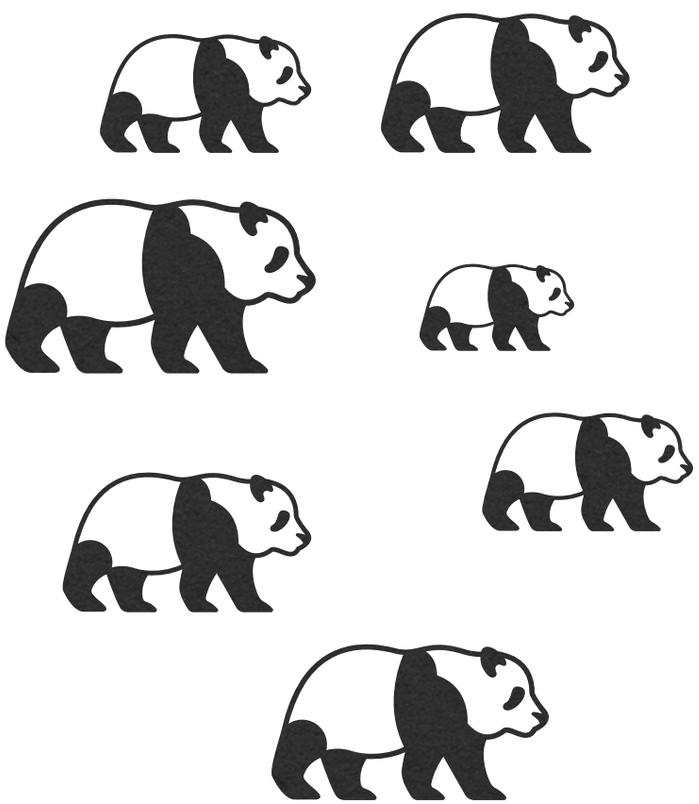


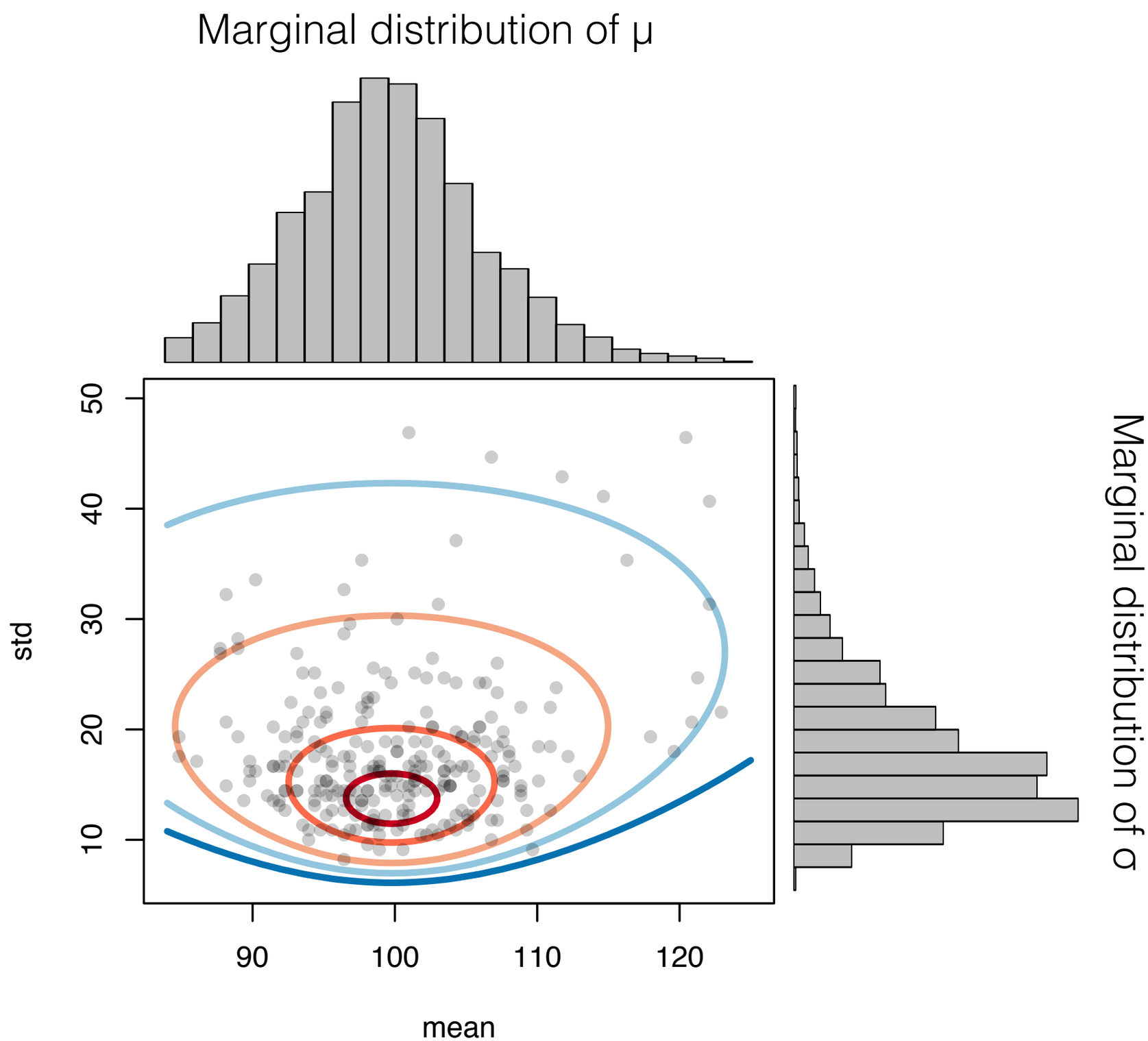
Summarizing MCMC samples



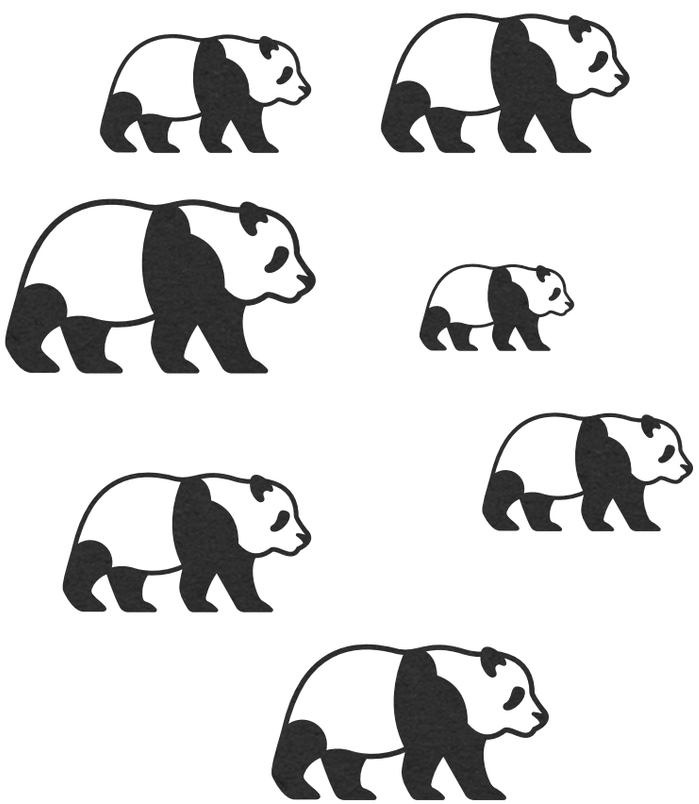
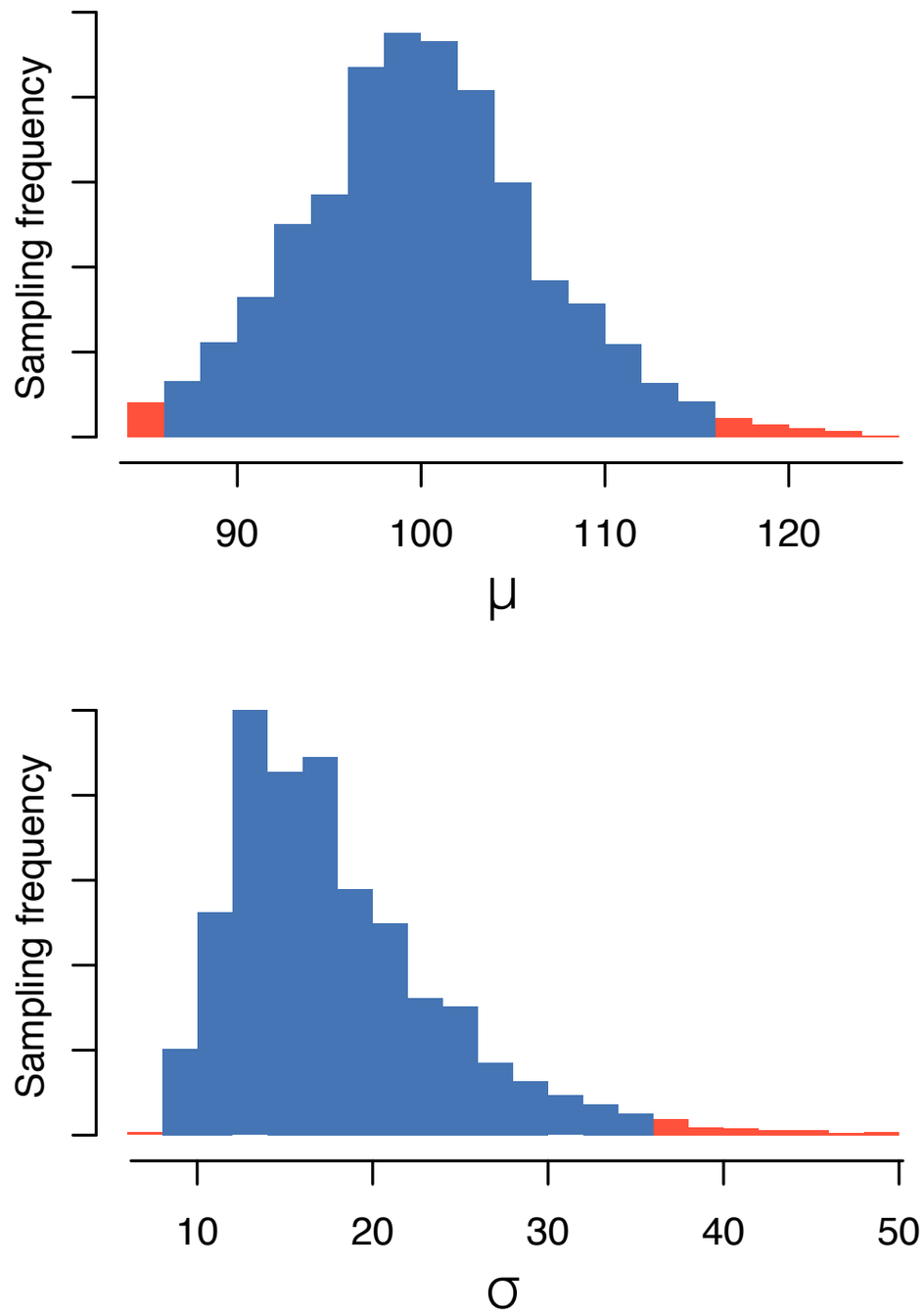


Summarizing MCMC samples





Summarizing MCMC samples

Posterior estimate: mean of MCMC samples

95% Credible intervals