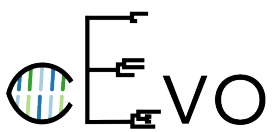
Computational Biology

Lecturers:
Tanja Stadler, Carsten Magnus & Tim Vaughan

Teaching Assistants:
Jūlija Pečerska, Jérémie Sciré,
Sarah Nadeau & Marc Manceau

Computational Evolution
Department of Biosystems Science and Engineering

HS 2019



Phylodynamic coalescent models: Questions

- Under the Wright-Fisher model, how many generations do we have to go back before we find the common ancestor of a pair of genes sampled from a haploid population of size N?

Phylodynamic coalescent models: Questions

- Under the Wright-Fisher model, how many generations do we have to go back before we find the common ancestor of a pair of genes sampled from a haploid population of size N?
- Suppose you had a tree inferred using present-day samples from a population that experienced a severe bottleneck in its recent past. How and why would this bottleneck likely affect our ability to infer ancestral population dynamics?

Phylodynamic coalescent models: Questions

- Under the Wright-Fisher model, how many generations do we have to go back before we find the common ancestor of a pair of genes sampled from a haploid population of size N?
- Suppose you had a tree inferred using present-day samples from a population that experienced a severe bottleneck in its recent past. How and why would this bottleneck likely affect our ability to infer ancestral population dynamics?
- Imagine spreading a Wright-Fisher population across islands in an archipelago, so that movement between the islands is restricted but within each island the population is "well-mixed". Qualitatively, how would you expect this population structure to influence estimates of the effective population size?

One of us has a birthday tomorrow. . . who?





	Jérémie	Jūlija	Tanja	Tim	Carsten
Prior					
Posterior					

We assign a probability to each column such that the row adds up to 1!

Lessons learned from the birthday guessing

- Each one of you had some knowledge BEFORE looking at the data.

Lessons learned from the birthday guessing

- Each one of you had some knowledge BEFORE looking at the data.
- You obtained data from me.

Lessons learned from the birthday guessing

- Each one of you had some knowledge BEFORE looking at the data.
- You obtained data from me.
- You had an implicit "model" in your head. Using this model and the data you updated your starting assumptions to obtain your final belief.

Lessons learned from the birthday guessing

- Each one of you had some knowledge BEFORE looking at the data.
- You obtained data from me.
- You had an implicit "model" in your head. Using this model and the data you updated your starting assumptions to obtain your final belief.
- Everyone received the same facts about each of us, i.e. gets the same data. However your initial assumptions and the models may differ, thus your final beliefs differ.

How would we apply this to phylogenetics?

- We could assume a model for evolution of sequences (e.g. Jukes-Cantor) and for population dynamics (e.g. birth-death model).

How would we apply this to phylogenetics?

- We could assume a model for evolution of sequences (e.g. Jukes-Cantor) and for population dynamics (e.g. birth-death model).
- We could assume some starting knowledge of the parameters (substitution rate, birth rate etc.) BEFORE looking at the data.

How would we apply this to phylogenetics?

- We could assume a model for evolution of sequences (e.g. Jukes-Cantor) and for population dynamics (e.g. birth-death model).
- We could assume some starting knowledge of the parameters (substitution rate, birth rate etc.) BEFORE looking at the data.
- We would then obtain and analyze sequencing data leading to a posterior distribution of trees & model parameters.

How would we apply this to phylogenetics?

- We could assume a model for evolution of sequences (e.g. Jukes-Cantor) and for population dynamics (e.g. birth-death model).
- We could assume some starting knowledge of the parameters (substitution rate, birth rate etc.) BEFORE looking at the data.
- We would then obtain and analyze sequencing data leading to a posterior distribution of trees & model parameters.
- If we received more data, we could use the knowledge obtained from the first analysis as the starting point for the analysis of the new data.

How can we make this kind of procedure quantitative?

What do we mean by "probability"?

Frequency Interpretation

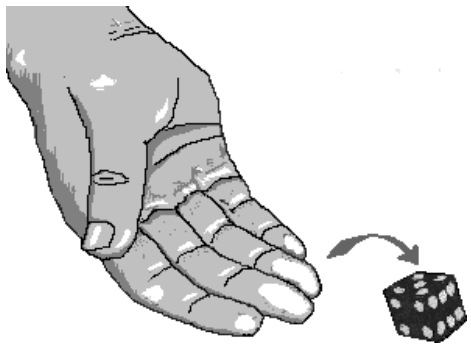
For this interpretation, probabilities are **relative frequencies of outcomes of repeatable random experiments**.

Frequency Interpretation

For this interpretation, probabilities are **relative frequencies of outcomes of repeatable random experiments**.

For example, consider a dice rolling experiment.

- Let N be the total number of rolls.
- Let n_5 be the total number of 5s rolled.



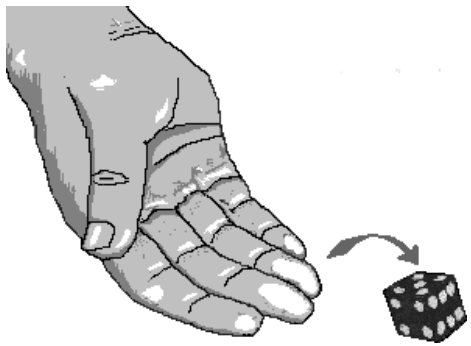
The probability of rolling a 5
$\mathrm{P}(\mathrm{d} = 5) = \lim_{\mathrm{N} \to \infty} \frac{\mathrm{n}_5}{\mathrm{N}}$.

Frequency Interpretation

For this interpretation, probabilities are **relative frequencies of outcomes of repeatable random experiments**.

For example, consider a dice rolling experiment.

- Let N be the total number of rolls.
- Let n_5 be the total number of 5s rolled.



The probability of rolling a 5
$\mathrm{P}(\mathrm{d} = 5) = \lim_{\mathrm{N}\to\infty} \frac{\mathrm{n}_5}{\mathrm{N}}$.

Characteristics of this view

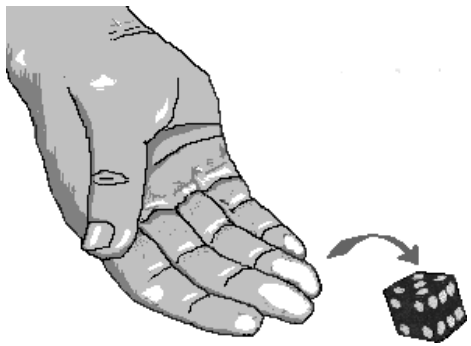
- Probabilities only assignable to outcomes of repeatable experiments (i.e. data).

Frequency Interpretation

For this interpretation, probabilities are **relative frequencies of outcomes of repeatable random experiments**.

For example, consider a dice rolling experiment.

- Let N be the total number of rolls.
- Let n_5 be the total number of 5s rolled.



The probability of rolling a 5
$\mathrm{P}(\mathrm{d} = 5) = \lim_{\mathrm{N}\to\infty} \frac{\mathrm{n}_5}{\mathrm{N}}$.

Characteristics of this view

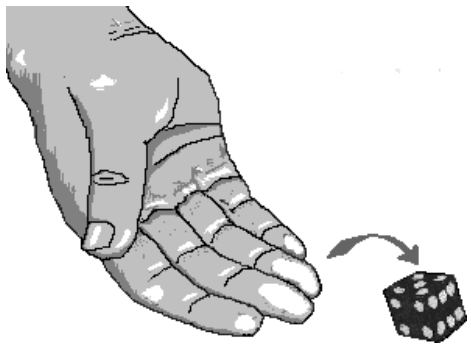
- Probabilities only assignable to outcomes of repeatable experiments (i.e. data).
- Probabilities treated as an intrinsic property of the system.

Frequency Interpretation

For this interpretation, probabilities are **relative frequencies of outcomes of repeatable random experiments**.

For example, consider a dice rolling experiment.

- Let N be the total number of rolls.
- Let n_5 be the total number of 5s rolled.



The probability of rolling a 5
$\mathrm{P}(\mathrm{d} = 5) = \lim_{\mathrm{N}\to\infty} \frac{\mathrm{n}_5}{\mathrm{N}}$.

Characteristics of this view

- Probabilities only assignable to outcomes of repeatable experiments (i.e. data).
- Probabilities treated as an intrinsic property of the system.
- Inference of model parameters is treated as a fundamentally distinct problem to the prediction of outcomes.

Bayesian Interpretation

Bayesian probabilities are the **plausibilities of propositions conditional on available information**.

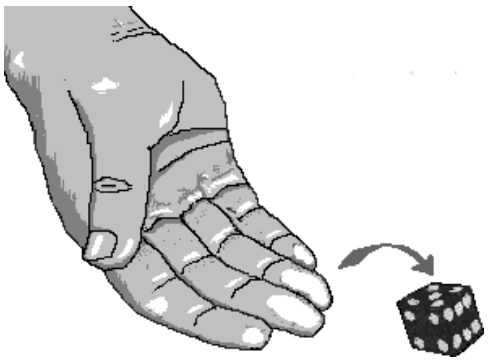
The probability of a given proposition (e.g. the next roll of a die will yield 5) depends on the information available.

Bayesian Interpretation

Bayesian probabilities are the **plausibilities of propositions conditional on available information**.

The probability of a given proposition (e.g. the next roll of a die will yield 5) depends on the information available.

$P(d=5|I_{Einstein})=1$



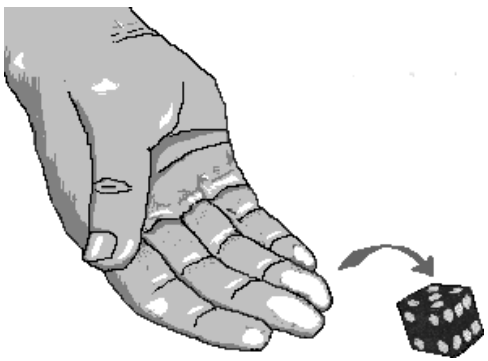
$P(d=5|I_{Homer})=1/6$

Bayesian Interpretation

Bayesian probabilities are the **plausibilities of propositions conditional on available information**.

The probability of a given proposition (e.g. the next roll of a die will yield 5) depends on the information available.

$P(d=5|I_{Einstein})=1$



$P(d=5|I_{Homer})=1/6$

Characteristics of this view

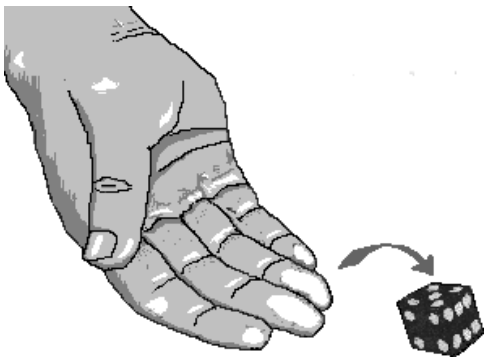
- Probabilities assignable to any unambiguous proposition.

Bayesian Interpretation

Bayesian probabilities are the **plausibilities of propositions conditional on available information**.

The probability of a given proposition (e.g. the next roll of a die will yield 5) depends on the information available.

$P(d=5|I_{Einstein})=1$



$P(d=5|I_{Homer})=1/6$

Characteristics of this view

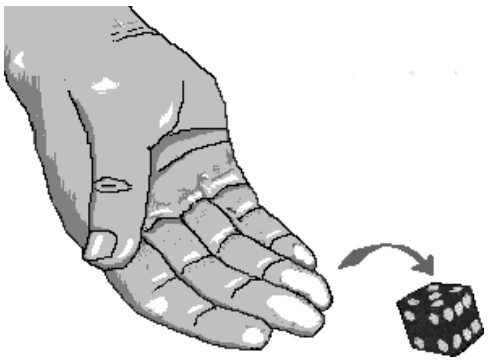
- Probabilities assignable to any unambiguous proposition.
- Probabilities represent lack of information to predict with complete certainty.

Bayesian Interpretation

Bayesian probabilities are the **plausibilities of propositions conditional on available information**.

The probability of a given proposition (e.g. the next roll of a die will yield 5) depends on the information available.

$P(d=5|I_{Einstein})=1$



$P(d=5|I_{Homer})=1/6$

Characteristics of this view

- Probabilities assignable to any unambiguous proposition.
- Probabilities represent lack of information to predict with complete certainty.
- Inference of model parameters is treated in the same way as prediction of outcomes.

Aside 1: A word about notation

One way to think about Bayesian probabilities is that they assigned to propositions, i.e. statements that can either be true or false:

Aside 1: A word about notation

One way to think about Bayesian probabilities is that they assigned to propositions, i.e. statements that can either be true or false:

- Tim is a cat.

Aside 1: A word about notation

One way to think about Bayesian probabilities is that they assigned to propositions, i.e. statements that can either be true or false:

- Tim is a cat.
- $\mathrm{N} = 5$ (where N represents some unknown quantity)

Aside 1: A word about notation

One way to think about Bayesian probabilities is that they assigned to propositions, i.e. statements that can either be true or false:

- Tim is a cat.
- $\mathrm{N} = 5$ (where N represents some unknown quantity)

A statement such as $\mathrm{P}(\mathrm{N})$ is therefore as meaningless as $\mathrm{P}(\mathrm{Tim})$.

Aside 1: A word about notation

One way to think about Bayesian probabilities is that they assigned to propositions, i.e. statements that can either be true or false:

- Tim is a cat.
- $\mathrm{N} = 5$ (where N represents some unknown quantity)

A statement such as $\mathrm{P}(\mathrm{N})$ is therefore as meaningless as $\mathrm{P}(\mathrm{Tim})$.

However, where propositions concern the value of a variable like N, we often use $\mathrm{P}(\mathrm{N})$ as shorthand for $\mathrm{P}(\mathrm{N} = \mathfrak{n})$.

Aside 1: A word about notation

One way to think about Bayesian probabilities is that they assigned to propositions, i.e. statements that can either be true or false:

- Tim is a cat.
- $\mathrm{N} = 5$ (where N represents some unknown quantity)

A statement such as $\mathrm{P}(\mathrm{N})$ is therefore as meaningless as $\mathrm{P}(\mathrm{Tim})$.

However, where propositions concern the value of a variable like N, we often use $\mathrm{P}(\mathrm{N})$ as shorthand for $\mathrm{P}(\mathrm{N} = \mathrm{n})$.

In general, this shorthand is okay, but take care that it does not lead to confusion.

Aside 1: A word about notation

One way to think about Bayesian probabilities is that they assigned to propositions, i.e. statements that can either be true or false:

- Tim is a cat.
- $\mathrm{N} = 5$ (where N represents some unknown quantity)

A statement such as $\mathrm{P}(\mathrm{N})$ is therefore as meaningless as $\mathrm{P}(\mathrm{Tim})$.

However, where propositions concern the value of a variable like N, we often use $\mathrm{P}(\mathrm{N})$ as shorthand for $\mathrm{P}(\mathrm{N} = \mathrm{n})$.

In general, this shorthand is okay, but take care that it does not lead to confusion.

Example confusing statement: $\mathrm{P}(\mathrm{N}|\mathrm{N} > 0.5)$. How could this be written more carefully?

Aside 2: Continuous variables

Propositions regarding continuous variables require special treatment.

Aside 2: Continuous variables

Propositions regarding continuous variables require special treatment.

- Suppose X may take any real value between 0 and 10.

Aside 2: Continuous variables

Propositions regarding continuous variables require special treatment.

- Suppose X may take any real value between 0 and 10.
- The probability $P(X = x)$ will usually be zero!

Aside 2: Continuous variables

Propositions regarding continuous variables require special treatment.

- Suppose X may take any real value between 0 and 10.
- The probability $P(X = x)$ will usually be zero!
- Instead, define

$$f(x) := \lim_{\delta \to 0} \frac{P(X \in [x, x + \delta])}{\delta}$$

Aside 2: Continuous variables

Propositions regarding continuous variables require special treatment.

- Suppose X may take any real value between 0 and 10.
- The probability $P(X = x)$ will usually be zero!
- Instead, define

$$f(x) := \lim_{\delta \to 0} \frac{P(X \in [x, x + \delta])}{\delta}$$

The function $f(x)$ is a probability *density* function (PDF) and satisfies the following rules:

Aside 2: Continuous variables

Propositions regarding continuous variables require special treatment.

- Suppose X may take any real value between 0 and 10.
- The probability $P(X = x)$ will usually be zero!
- Instead, define

$$f(x) := \lim_{\delta \to 0} \frac{P(X \in [x, x + \delta])}{\delta}$$

The function $f(x)$ is a probability *density* function (PDF) and satisfies the following rules:

- It is normalized: $\int_0^{10} f(x)dx = 1$

Aside 2: Continuous variables

Propositions regarding continuous variables require special treatment.

- Suppose X may take any real value between 0 and 10.
- The probability $P(X = x)$ will usually be zero!
- Instead, define

$$f(x) := \lim_{\delta \to 0} \frac{P(X \in [x, x + \delta])}{\delta}$$

The function $f(x)$ is a probability *density* function (PDF) and satisfies the following rules:

- It is normalized: $\int_0^{10} f(x)dx = 1$
- It is positive: $f(x) \geqslant 0$

Aside 2: Continuous variables

Propositions regarding continuous variables require special treatment.

- Suppose X may take any real value between 0 and 10.
- The probability $P(X = x)$ will usually be zero!
- Instead, define

$$f(x) := \lim_{\delta \to 0} \frac{P(X \in [x, x + \delta])}{\delta}$$

The function $f(x)$ is a probability *density* function (PDF) and satisfies the following rules:

- It is normalized: $\int_0^{10} f(x)dx = 1$
- It is positive: $f(x) \geqslant 0$

However, note that $f(x)$ may exceed 1!

Example: Inference of genetic distance

These two sequences are separated by an unknown genetic distance, $\mathbf{d}$:

Sequence 1	A A T C T G T G T G
Sequence 2	A G C C T G G G T A

Example: Inference of genetic distance

These two sequences are separated by an unknown genetic distance, $\mathbf{d}$:

Sequence 1	A A T C T G T G T G
Sequence 2	A G C C T G G G T A

The Jukes-Cantor transition probabilities for each site are a function of the random variable $\mathbf{d}$:

$$\mathbf{p}_{ij}(\mathbf{d}) = \begin{cases} \frac{1}{4} + \frac{3}{4}\exp\left(-\frac{4}{3}\mathbf{d}\right) & \text{if } \mathbf{i} = \mathbf{j} \\ \frac{1}{4} - \frac{1}{4}\exp\left(-\frac{4}{3}\mathbf{d}\right) & \text{if } \mathbf{i} \neq \mathbf{j} \end{cases},$$

Example: Inference of genetic distance

These two sequences are separated by an unknown genetic distance, d:

Sequence 1	A A T C T G T G T G
Sequence 2	A G C C T G G G T A

The Jukes-Cantor transition probabilities for each site are a function of the random variable d:

$$\mathrm{p_{ij}}(\mathrm{d}) = \begin{cases} \frac{1}{4} + \frac{3}{4}\exp\left(-\frac{4}{3}\mathrm{d}\right) & \text{if } \mathrm{i} = \mathrm{j} \\ \frac{1}{4} - \frac{1}{4}\exp\left(-\frac{4}{3}\mathrm{d}\right) & \text{if } \mathrm{i} \neq \mathrm{j} \end{cases},$$

The number of substitutions is $\mathrm{S} = 4$ and the total number of sites is $\mathrm{L} = 10$, so the likelihood for the pairwise alignment is:

Example: Inference of genetic distance

These two sequences are separated by an unknown genetic distance, d:

```
Sequence 1    A A T C T G T G T G
Sequence 2    A G C C T G G G T A
```

The Jukes-Cantor transition probabilities for each site are a function of the random variable d:

$$\mathrm{p_{ij}}(\mathrm{d}) = \begin{cases} \frac{1}{4} + \frac{3}{4}\exp\left(-\frac{4}{3}\mathrm{d}\right) & \text{if } \mathrm{i} = \mathrm{j} \\ \frac{1}{4} - \frac{1}{4}\exp\left(-\frac{4}{3}\mathrm{d}\right) & \text{if } \mathrm{i} \neq \mathrm{j} \end{cases},$$

The number of substitutions is $S = 4$ and the total number of sites is $\mathrm{L} = 10$, so the likelihood for the pairwise alignment is:

$$\mathrm{P}[\mathrm{S}|\mathrm{d}, \mathrm{L}] = \left[\frac{1}{4} + \frac{3}{4}\exp\left(-\frac{4}{3}\mathrm{d}\right)\right]^{\mathrm{L}-S} \times \left[\frac{1}{4} - \frac{1}{4}\exp\left(-\frac{4}{3}\mathrm{d}\right)\right]^{S}$$

Example: Inference of genetic distance

- Our model M has provided us with the probability of the number of segregating sites S given the genetic distance d and the length L of the sequences: $\mathrm{P}(\mathrm{S}|\mathrm{d}, \mathrm{L}, \mathrm{M})$.

Example: Inference of genetic distance

- Our model M has provided us with the probability of the number of segregating sites S given the genetic distance d and the length L of the sequences: $\mathrm{P}(\mathrm{S}|\mathrm{d}, \mathrm{L}, \mathrm{M})$.
- Our Bayesian interpretation of probabilities means that it is sensible to talk about the probability of d given S and L: $\mathrm{P}(\mathrm{d}|\mathrm{S}, \mathrm{L}, \mathrm{M})$.

Example: Inference of genetic distance

- Our model M has provided us with the probability of the number of segregating sites S given the genetic distance d and the length L of the sequences: $\mathrm{P}(\mathrm{S}|\mathrm{d}, \mathrm{L}, \mathrm{M})$.
- Our Bayesian interpretation of probabilities means that it is sensible to talk about the probability of d given S and L: $\mathrm{P}(\mathrm{d}|\mathrm{S}, \mathrm{L}, \mathrm{M})$.
- This distribution quantifies our state of knowledge regarding d once the observed S is taken into account.

Example: Inference of genetic distance

- Our model M has provided us with the probability of the number of segregating sites S given the genetic distance d and the length L of the sequences: $\mathrm{P}(\mathrm{S}|\mathrm{d}, \mathrm{L}, \mathrm{M})$.
- Our Bayesian interpretation of probabilities means that it is sensible to talk about the probability of d given S and L: $\mathrm{P}(\mathrm{d}|\mathrm{S}, \mathrm{L}, \mathrm{M})$.
- This distribution quantifies our state of knowledge regarding d once the observed S is taken into account.

This is precisely what we want to know!

Example: Inference of genetic distance

How do we obtain $\mathrm{P}(\mathrm{d}|S, \mathrm{L}, \mathrm{M})$? Use the product rule!

Example: Inference of genetic distance

How do we obtain $P(d|S, L, M)$? Use the product rule!

$$P(d|S, L, M) = \frac{P(S, d|L, M)}{P(S|L, M)}$$
$$= \frac{P(S|d, L, M)P(d|L, M)}{P(S|L, M)}$$

Example: Inference of genetic distance

How do we obtain $\mathrm{P}(\mathbf{d}|S, \mathrm{L}, \mathrm{M})$? Use the product rule!

$$\begin{aligned}\mathrm{P}(\mathbf{d}|S, \mathrm{L}, \mathrm{M}) &= \frac{\mathrm{P}(S, \mathbf{d}|\mathrm{L}, \mathrm{M})}{\mathrm{P}(S|\mathrm{L}, \mathrm{M})} \\ &= \frac{\mathrm{P}(S|\mathbf{d}, \mathrm{L}, \mathrm{M})\mathrm{P}(\mathbf{d}|\mathrm{L}, \mathrm{M})}{\mathrm{P}(S|\mathrm{L}, \mathrm{M})}\end{aligned}$$

$\mathrm{P}(\mathbf{d}|\mathrm{L}, \mathrm{M}) = \mathrm{P}(\mathbf{d}|\mathrm{M})$ quantifies knowledge of $\mathbf{d}$ in the absence of the observation, while $\mathrm{P}(S|\mathrm{L}, \mathrm{M})$ is the distribution over possible numbers of segregating sites given the JC69 model and any independent knowledge of $\mathbf{d}$.

Example: Inference of genetic distance

How do we obtain $\mathrm{P}(\mathrm{d}|S, \mathrm{L}, \mathrm{M})$? Use the product rule!

$$\begin{aligned}\mathrm{P}(\mathrm{d}|S, \mathrm{L}, \mathrm{M}) &= \frac{\mathrm{P}(S, \mathrm{d}|\mathrm{L}, \mathrm{M})}{\mathrm{P}(S|\mathrm{L}, \mathrm{M})} \\ &= \frac{\mathrm{P}(S|\mathrm{d}, \mathrm{L}, \mathrm{M})\mathrm{P}(\mathrm{d}|\mathrm{L}, \mathrm{M})}{\mathrm{P}(S|\mathrm{L}, \mathrm{M})}\end{aligned}$$

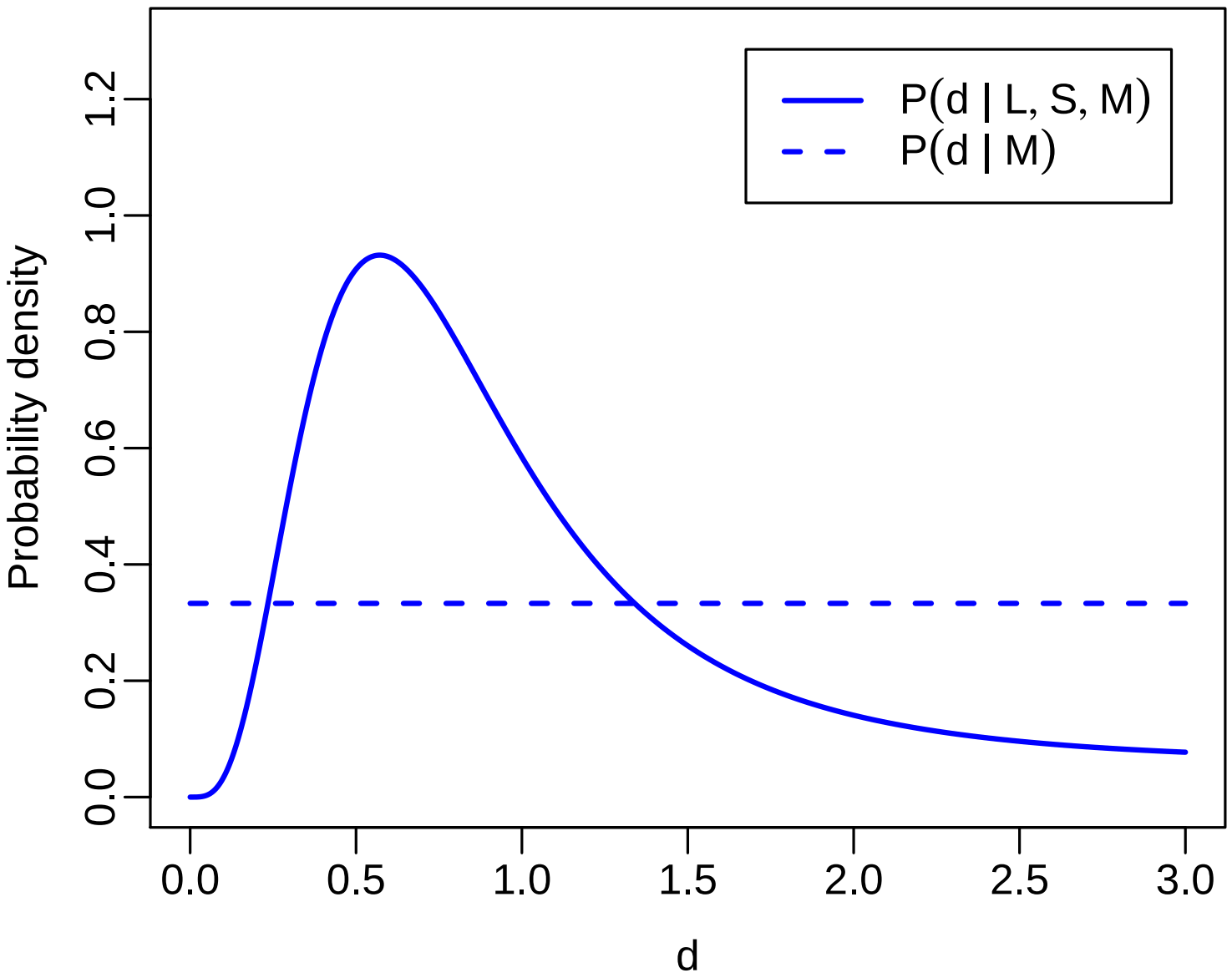
$\mathrm{P}(\mathrm{d}|\mathrm{L}, \mathrm{M}) = \mathrm{P}(\mathrm{d}|\mathrm{M})$ quantifies knowledge of d in the absence of the observation, while $\mathrm{P}(S|\mathrm{L}, \mathrm{M})$ is the distribution over possible numbers of segregating sites given the JC69 model and any independent knowledge of d.

Here we assume that our prior information is only that $0 \leqslant \mathrm{d} \leqslant 3$, so we take

$$\mathrm{P}(\mathrm{d}|\mathrm{M}) = \begin{cases} \frac{1}{3} & \text{for } 0 \leqslant \mathrm{d} \leqslant 3 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

(i.e. a uniform distribution between 0 and 3).

Example: Inference of genetic distance



Bayes theorem

In answering this question we have *accidentally* discovered Bayes theorem:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

Bayes theorem

In answering this question we have *accidentally* discovered Bayes theorem:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

Here θ_M are parameters of some model M, while D are data assumed to be generated by that same model.

Bayes theorem

In answering this question we have *accidentally* discovered Bayes theorem:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

Here θ_M are parameters of some model M, while D are data assumed to be generated by that same model.

The components of the theorem have the following names:

Bayes theorem

In answering this question we have *accidentally* discovered Bayes theorem:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

Here θ_M are parameters of some model M, while D are data assumed to be generated by that same model.

The components of the theorem have the following names:

- $P(\theta_M|M)$ is the prior for the model parameters,

Bayes theorem

In answering this question we have *accidentally* discovered Bayes theorem:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

Here θ_M are parameters of some model M, while D are data assumed to be generated by that same model.

The components of the theorem have the following names:

- $P(\theta_M|M)$ is the prior for the model parameters,
- $P(D|\theta_M, M)$ is the likelihood of the parameters given the data (as discussed in Lecture 4),

Bayes theorem

In answering this question we have *accidentally* discovered Bayes theorem:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

Here θ_M are parameters of some model M, while D are data assumed to be generated by that same model.

The components of the theorem have the following names:

- $P(\theta_M|M)$ is the prior for the model parameters,
- $P(D|\theta_M, M)$ is the likelihood of the parameters given the data (as discussed in Lecture 4),
- $P(D|M)$ is the marginal likelihood of (or evidence for) the model, and

Bayes theorem

In answering this question we have *accidentally* discovered Bayes theorem:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

Here θ_M are parameters of some model M, while D are data assumed to be generated by that same model.

The components of the theorem have the following names:

- $P(\theta_M|M)$ is the prior for the model parameters,
- $P(D|\theta_M, M)$ is the likelihood of the parameters given the data (as discussed in Lecture 4),
- $P(D|M)$ is the marginal likelihood of (or evidence for) the model, and
- $P(\theta_M|D, M)$ is the posterior of the model parameters given the model and the data.

Rev. Thomas Bayes (1701–1761)

- 18th century Presbyterian minister.
- Derived special case of what we now know as Bayes theorem.
- Published in "An Essay towards solving a Problem in the Doctrine of Chances (1763)".

Bayesian updating: including more data

Suppose we acquired sequence data for an additional 90 sites from the same pair of genomes as the original 10 sites. This new alignment has length $\mathrm{L}' = 90$ and differs at $S' = 48$ sites.

Bayesian updating: including more data

Suppose we acquired sequence data for an additional 90 sites from the same pair of genomes as the original 10 sites. This new alignment has length $\mathrm{L}' = 90$ and differs at $S' = 48$ sites.

Question

How can we update our estimate for d?

Bayesian updating: including more data

Suppose we acquired sequence data for an additional 90 sites from the same pair of genomes as the original 10 sites. This new alignment has length $\mathrm{L}' = 90$ and differs at $\mathrm{S}' = 48$ sites.

Question

How can we update our estimate for d?

Answer

Simply apply Bayes theorem with the posterior of the previous analysis as the prior for the next analysis.

Bayesian updating: including more data

Suppose we acquired sequence data for an additional 90 sites from the same pair of genomes as the original 10 sites. This new alignment has length $L' = 90$ and differs at $S' = 48$ sites.

Question

How can we update our estimate for d?

Answer

Simply apply Bayes theorem with the posterior of the previous analysis as the prior for the next analysis.

$$\begin{aligned} P(d|S', S, L', L, M) &= \frac{P(S'|d, L', M)P(d|S, L, M)}{P(S'|S, L, L', M)} \\ &= \frac{P(S'|d, L', M)P(S|d, L, M)P(d|M)}{P(S'|S, L, L', M)P(S|L, M)} \\ &= \frac{P(S', S'|d, L, L', M)P(d|M)}{P(S', S|L, L', M)} \end{aligned}$$

Bayesian updating: including more data

Suppose we acquired sequence data for an additional 90 sites from the same pair of genomes as the original 10 sites. This new alignment has length $\mathrm{L}' = 90$ and differs at $S' = 48$ sites.

Question

How can we update our estimate for d?

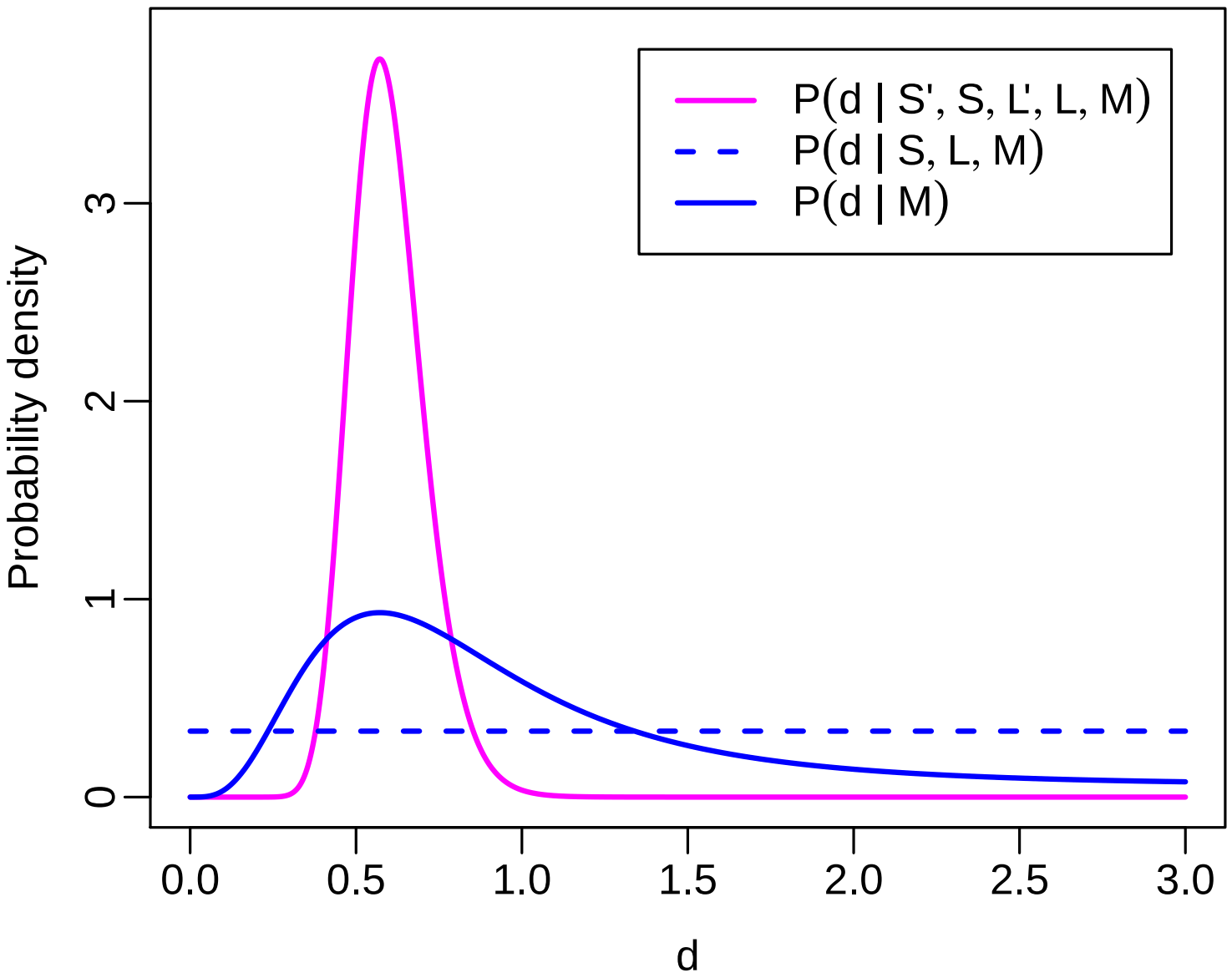
Answer

Simply apply Bayes theorem with the posterior of the previous analysis as the prior for the next analysis.

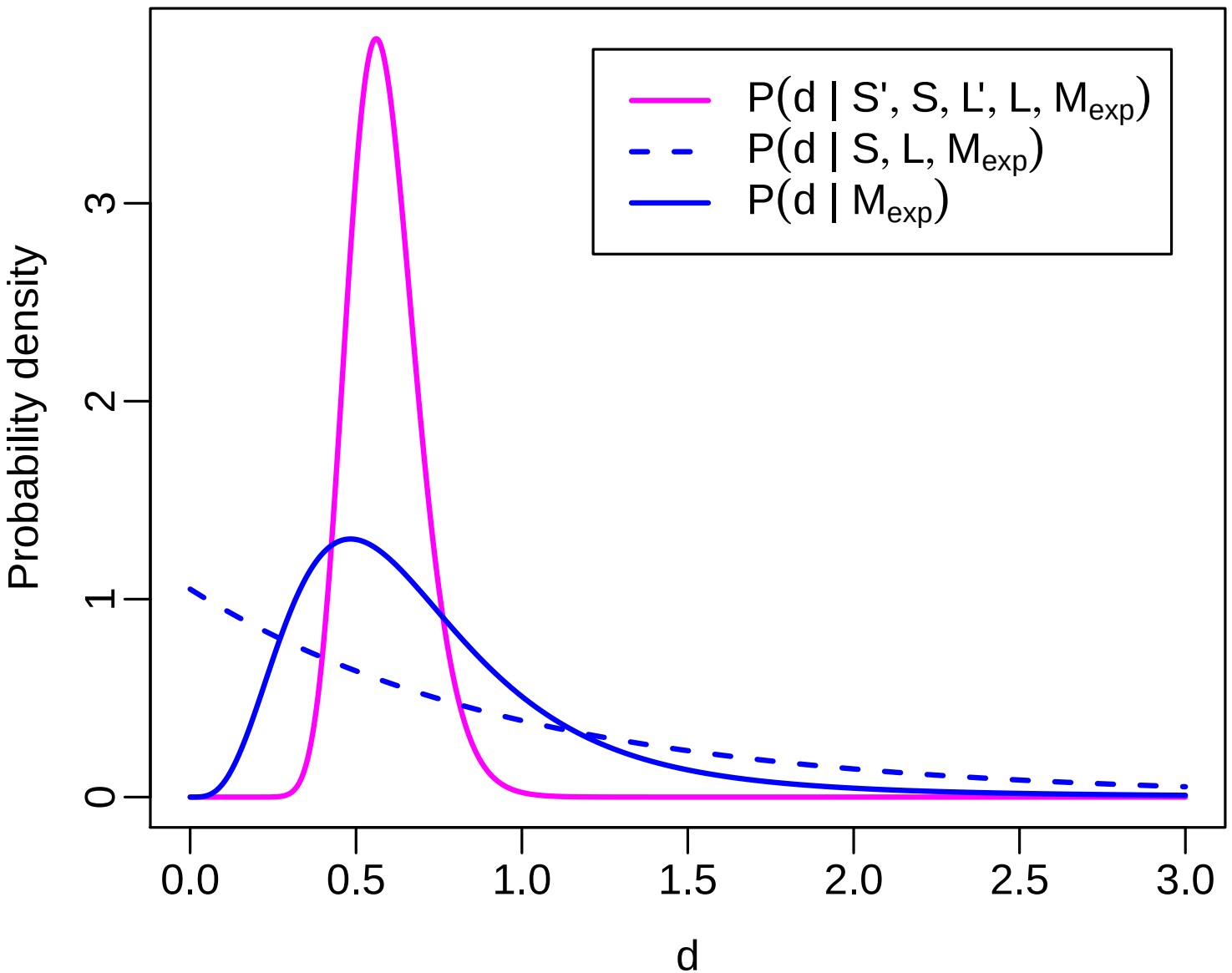
$$\begin{aligned}
\mathrm{P}(\mathrm{d}|S', S, \mathrm{L}', \mathrm{L}, \mathrm{M}) &= \frac{\mathrm{P}(S'|\mathrm{d}, \mathrm{L}', \mathrm{M})\mathrm{P}(\mathrm{d}|S, \mathrm{L}, \mathrm{M})}{\mathrm{P}(S'|S, \mathrm{L}, \mathrm{L}', \mathrm{M})} \\
&= \frac{\mathrm{P}(S'|\mathrm{d}, \mathrm{L}', \mathrm{M})\mathrm{P}(S|\mathrm{d}, \mathrm{L}, \mathrm{M})\mathrm{P}(\mathrm{d}|\mathrm{M})}{\mathrm{P}(S'|S, \mathrm{L}, \mathrm{L}', \mathrm{M})\mathrm{P}(S|\mathrm{L}, \mathrm{M})} \\
&= \frac{\mathrm{P}(S', S|\mathrm{d}, \mathrm{L}, \mathrm{L}', \mathrm{M})\mathrm{P}(\mathrm{d}|\mathrm{M})}{\mathrm{P}(S', S|\mathrm{L}, \mathrm{L}', \mathrm{M})}
\end{aligned}$$

Equivalent to inferring d *from both data sets simultaneously!*

Bayesian updating: including more data



Using different prior info.: $P(d|M_{exp}) = e^{-d}$



What is a prior probability distribution?

A prior probability distribution is:

What is a prior probability distribution?

A prior probability distribution is:

- a probability distribution!

What is a prior probability distribution?

A prior probability distribution is:

- a probability distribution!
  - I.e. a quantification of knowledge.

What is a prior probability distribution?

A prior probability distribution is:

- a probability distribution!
  - I.e. a quantification of knowledge.
- Your knowledge about some variable in the absence of the data included in the likelihood term.

What is a prior probability distribution?

A prior probability distribution is:

- a probability distribution!
  - I.e. a quantification of knowledge.
- Your knowledge about some variable in the absence of the data included in the likelihood term.
  - Your prior for the analysis of *this data* may be informed by *other data*.

What is a prior probability distribution?

A prior probability distribution is:

- a probability distribution!
  - I.e. a quantification of knowledge.
- Your knowledge about some variable in the absence of the data included in the likelihood term.
  - Your prior for the analysis of *this data* may be informed by *other data*.

In principle, any two (rational) people/computers with access to precisely the same background information should specify exactly the same prior.

What is a prior probability distribution?

A prior probability distribution is:

- a probability distribution!
  - I.e. a quantification of knowledge.
- Your knowledge about some variable in the absence of the data included in the likelihood term.
  - Your prior for the analysis of *this data* may be informed by *other data*.

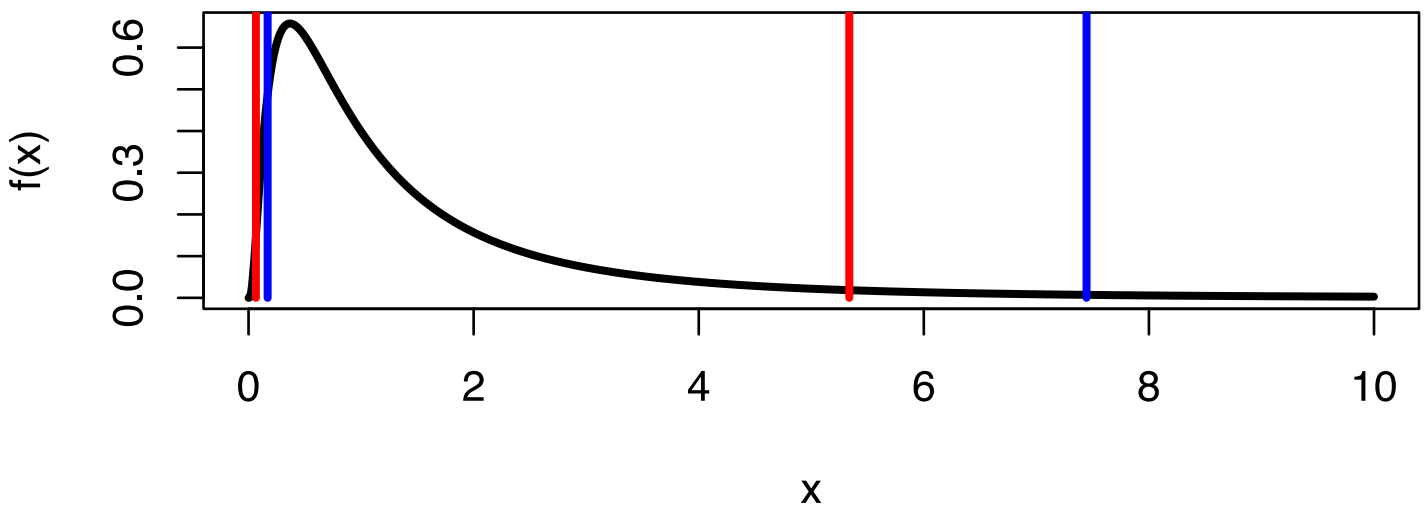
In principle, any two (rational) people/computers with access to precisely the same background information should specify exactly the same prior.

In practice, truly objective prior selection is difficult to achieve.

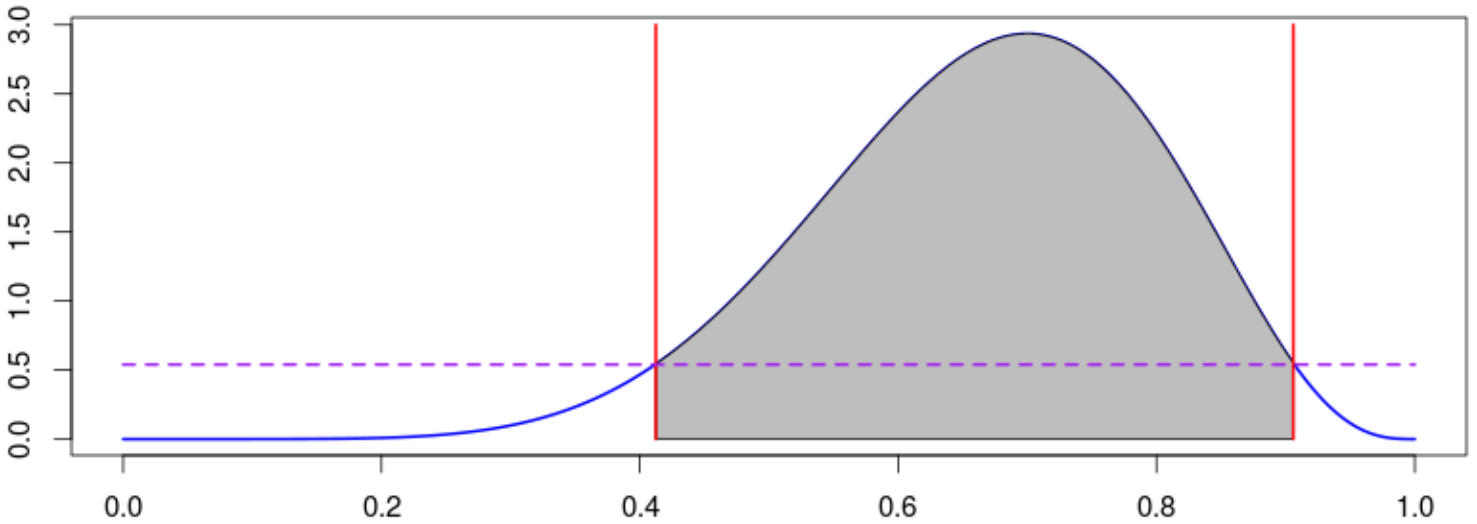
Credible Intervals

The 95 % credible interval is an interval of the posterior distribution containing 95 % of the probability.

- One can neglect 2.5 % of the samples on both ends of the posterior distribution (blue below).
- Often, the smallest interval spanned by 95 % of the samples is chosen (also called highest posterior density (HPD), red below).

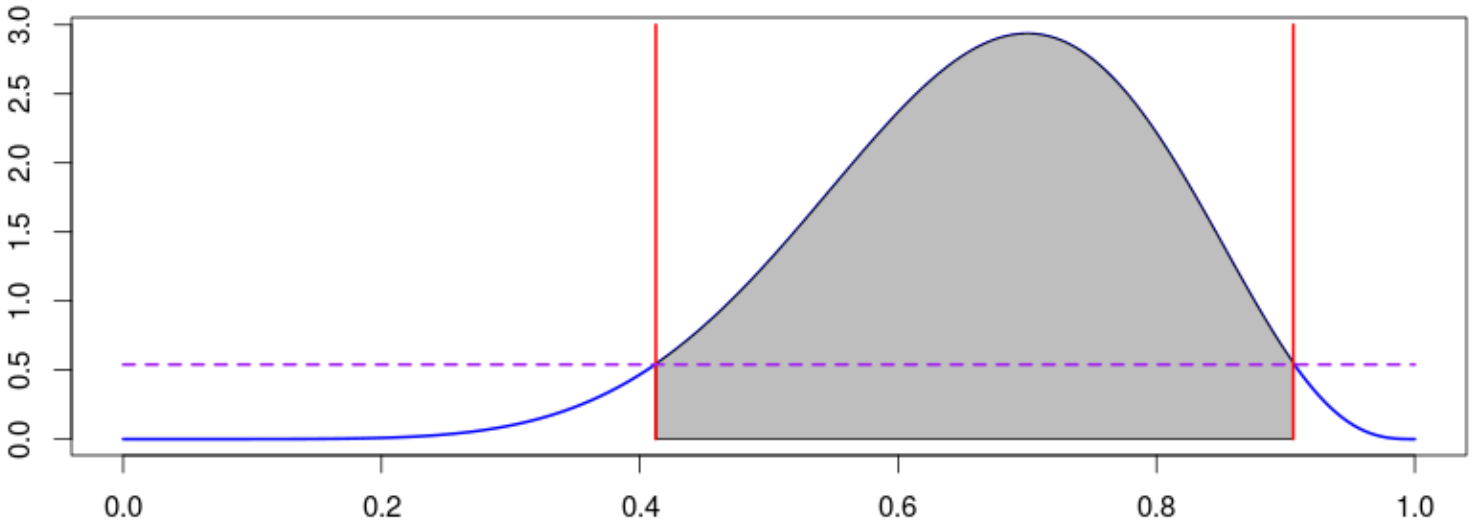


Highest posterior density (HPD intervals



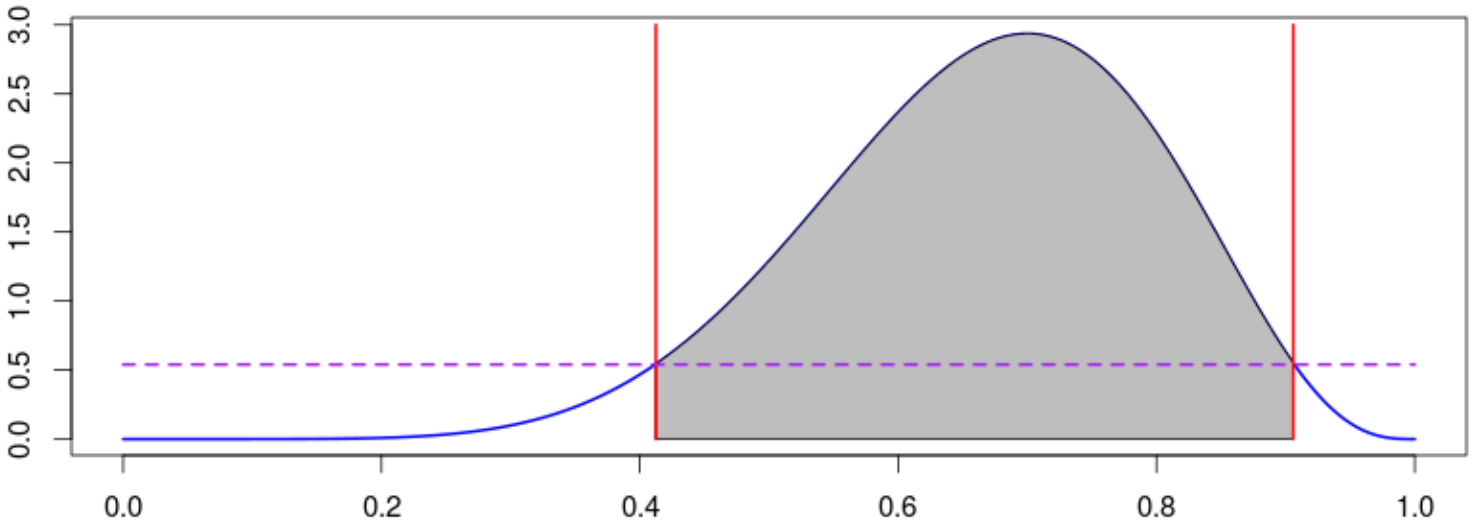
- The 95% HPD interval can also be found by lowering a threshold density until the area under the curve where the density exceeds the threshold is 0.95.

Highest posterior density (HPD intervals



- The 95% HPD interval can also be found by lowering a threshold density until the area under the curve where the density exceeds the threshold is 0.95.
- The meaning of this interval is simply: the probability of the unknown value falling in this region is 95% **given the observed data**.

Highest posterior density (HPD intervals



- The 95% HPD interval can also be found by lowering a threshold density until the area under the curve where the density exceeds the threshold is 0.95.
- The meaning of this interval is simply: the probability of the unknown value falling in this region is 95% **given the observed data**.
- This is different to a 95% *confidence* interval, which is instead an interval produced by a method which generates truth-containing intervals 95% of the time when averaging over **all possible data sets**.

What is so difficult about Bayesian inference?

What is so difficult about Bayesian inference?

INTEGRATION

What is so difficult about Bayesian inference?

INTEGRATION

Bayes' theorem has a troublesome denominator:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

What is so difficult about Bayesian inference?

INTEGRATION

Bayes' theorem has a troublesome denominator:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

The marginal likelihood $P(D|M)$ can be considered a normalizing constant for the posterior distribution, and can be expanded as follows:

$$P(D|M) = \int P(D|\theta_M, M)P(\theta_M|M)d\theta_M$$

What is so difficult about Bayesian inference?

INTEGRATION

Bayes' theorem has a troublesome denominator:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

The marginal likelihood $P(D|M)$ can be considered a normalizing constant for the posterior distribution, and can be expanded as follows:

$$P(D|M) = \int P(D|\theta_M, M)P(\theta_M|M)d\theta_M$$

- Unless you are *very* lucky, this integral can't be solved with pen and paper.

What is so difficult about Bayesian inference?

INTEGRATION

Bayes' theorem has a troublesome denominator:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

The marginal likelihood $P(D|M)$ can be considered a normalizing constant for the posterior distribution, and can be expanded as follows:

$$P(D|M) = \int P(D|\theta_M, M)P(\theta_M|M)d\theta_M$$

- Unless you are *very* lucky, this integral can't be solved with pen and paper.
- If θ_M has many dimensions, you won't even be able to directly integrate this using a computer.

What is so difficult about Bayesian inference?

INTEGRATION

Bayes' theorem has a troublesome denominator:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

The marginal likelihood $P(D|M)$ can be considered a normalizing constant for the posterior distribution, and can be expanded as follows:

$$P(D|M) = \int P(D|\theta_M, M)P(\theta_M|M)d\theta_M$$

- Unless you are *very* lucky, this integral can't be solved with pen and paper.
- If θ_M has many dimensions, you won't even be able to directly integrate this using a computer.

What is so difficult about Bayesian inference?

INTEGRATION

Bayes' theorem has a troublesome denominator:

$$P(\theta_M|D, M) = \frac{P(D|\theta_M, M)P(\theta_M|M)}{P(D|M)}$$

The marginal likelihood $P(D|M)$ can be considered a normalizing constant for the posterior distribution, and can be expanded as follows:

$$P(D|M) = \int P(D|\theta_M, M)P(\theta_M|M)d\theta_M$$

- Unless you are *very* lucky, this integral can't be solved with pen and paper.
- If θ_M has many dimensions, you won't even be able to directly integrate this using a computer.

This is true for most phylogenetic and phylodynamic problems.

Monte Carlo methods

- In our context, Monte Carlo methods are algorithms which produce random samples of values in order to characterize a probability distribution.

Monte Carlo methods

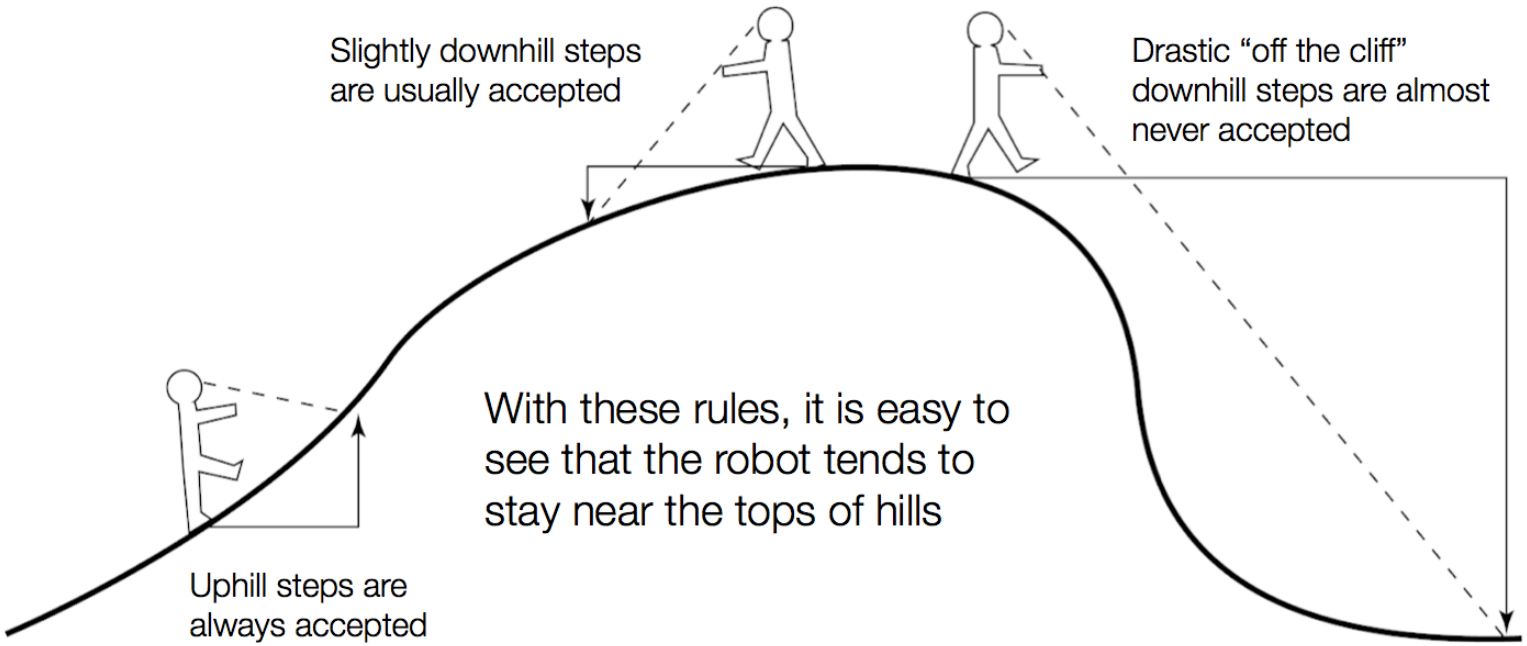
- In our context, Monte Carlo methods are algorithms which produce random samples of values in order to characterize a probability distribution.
- Usually, the algorithms we deal with seek to produce an arbitrary number of independent samples of possible parameter values θ_M drawn from the posterior distribution $P(\theta_M|D, M)$.

Monte Carlo methods

- In our context, Monte Carlo methods are algorithms which produce random samples of values in order to characterize a probability distribution.
- Usually, the algorithms we deal with seek to produce an arbitrary number of independent samples of possible parameter values θ_M drawn from the posterior distribution $P(\theta_M|D, M)$.
- Markov Chain Monte Carlo is an example of such an algorithm which is extremely popular in Bayesian phylogenetics and phylodynamics.

Markov chain Monte Carlo (MCMC) robot

[courtesy of Paul O Lewis]



The MCMC robot produces a carefully constructed random walk on the domain of the target distribution.

In Bayesian MCMC the target distribution is the posterior distribution $\mathrm{P}(\theta_{\mathrm{M}}|\mathrm{D},\mathrm{M})$.

Markov chain Monte Carlo (MCMC) robot

Let $\theta_{\mathcal{M}}$ be the current state. Let $\theta'_{\mathcal{M}}$ be a proposed parameter set for the new state.

Markov chain Monte Carlo (MCMC) robot

Let θ_M be the current state. Let θ'_M be a proposed parameter set for the new state.

- For the proposed parameter set, we calculate

$$R = \frac{P(\theta'_M|D, M)}{P(\theta_M|D, M)} = \frac{P(D|\theta'_M, M)P(\theta'_M|M)}{P(D|\theta_M, M)P(\theta_M|M)}.$$

Markov chain Monte Carlo (MCMC) robot

Let θ_M be the current state. Let θ'_M be a proposed parameter set for the new state.

- For the proposed parameter set, we calculate

$$R = \frac{P(\theta'_M|D, M)}{P(\theta_M|D, M)} = \frac{P(D|\theta'_M, M)P(\theta'_M|M)}{P(D|\theta_M, M)P(\theta_M|M)}.$$

- We draw a uniform number u on $(0, 1)$. We accept the proposed step if $u \leqslant R$.

Markov chain Monte Carlo (MCMC) robot

Let θ_M be the current state. Let θ'_M be a proposed parameter set for the new state.

- For the proposed parameter set, we calculate

$$R = \frac{P(\theta'_M|D, M)}{P(\theta_M|D, M)} = \frac{P(D|\theta'_M, M)P(\theta'_M|M)}{P(D|\theta_M, M)P(\theta_M|M)}.$$

- We draw a uniform number u on $(0, 1)$. We accept the proposed step if $u \leqslant R$.

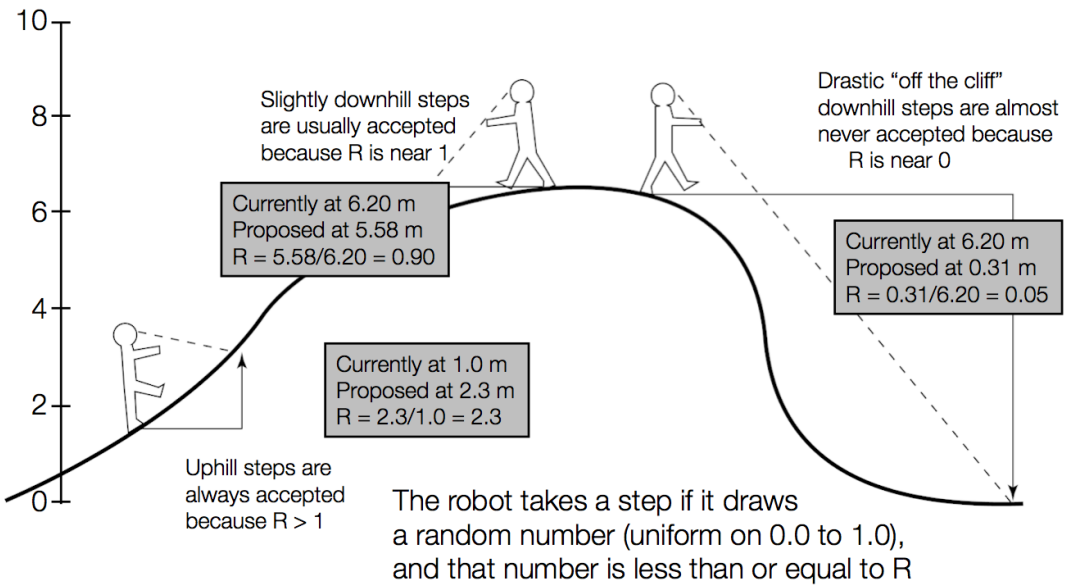


Figure adapted from PO Lewis.

Metropolis-Hastings algorithm

- The introduced MCMC robot implements the "Metropolis algorithm".

Metropolis-Hastings algorithm

- The introduced MCMC robot implements the "Metropolis algorithm".
- The robot will produce a sample from the posterior distribution.

Metropolis-Hastings algorithm

- The introduced MCMC robot implements the "Metropolis algorithm".
- The robot will produce a sample from the posterior distribution.
- The Metropolis algorithm requires that the proposal probability q for a new state θ'_M given θ_M satisfies $q(\theta'_M|\theta_M) = q(\theta_M|\theta'_M)$.

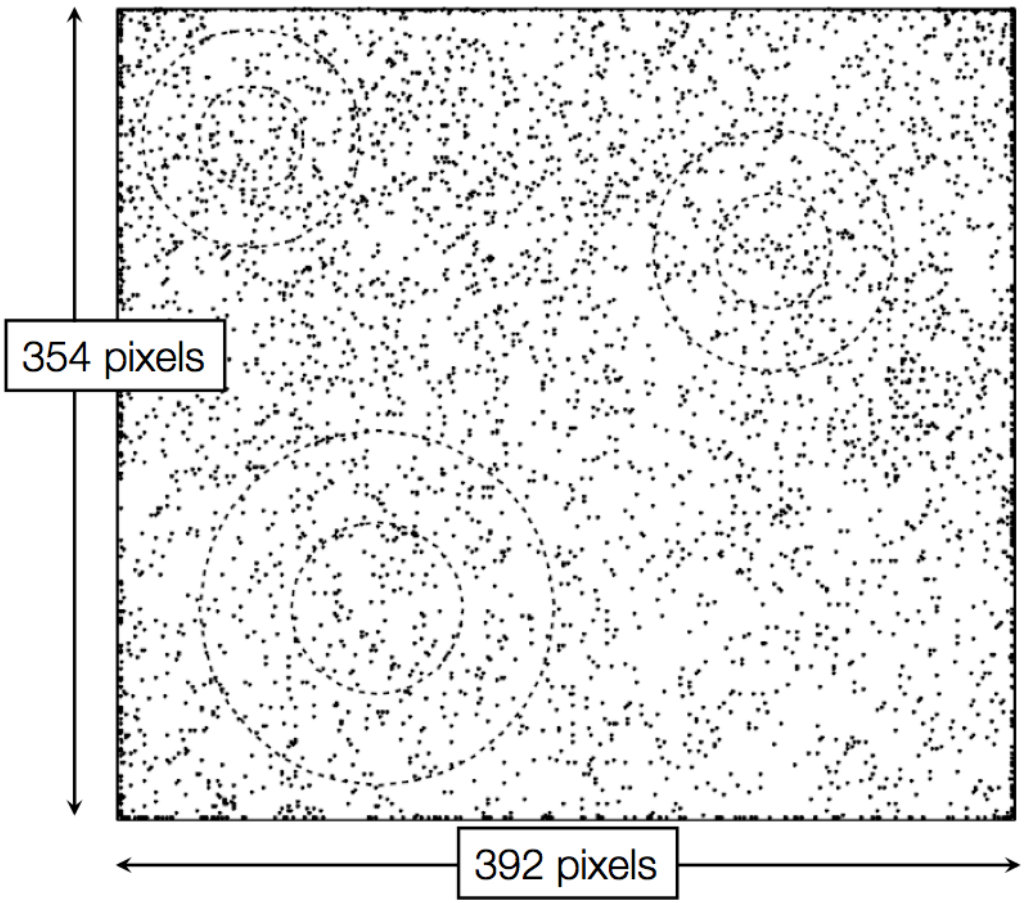
Metropolis-Hastings algorithm

- The introduced MCMC robot implements the "Metropolis algorithm".
- The robot will produce a sample from the posterior distribution.
- The Metropolis algorithm requires that the proposal probability q for a new state θ'_M given θ_M satisfies $q(\theta'_M|\theta_M) = q(\theta_M|\theta'_M)$.
- The Metropolis-Hastings algorithm allows for non-symmetric proposals by using

$$R = \frac{P(\theta'_M|D, M)q(\theta_M|\theta'_M)}{P(\theta_M|D, M)q(\theta'_M|\theta_M)}.$$

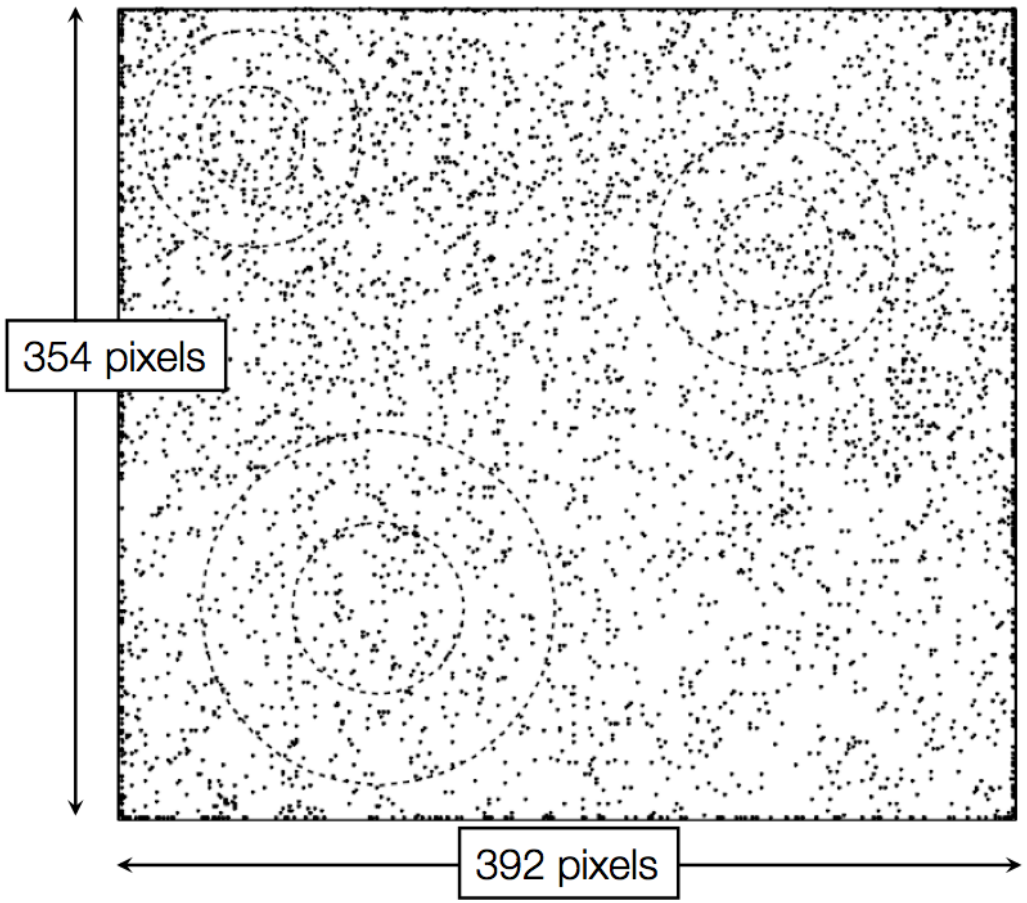
Pure Random Walk

[courtesy of Paul O Lewis]



Pure Random Walk

[courtesy of Paul O Lewis]

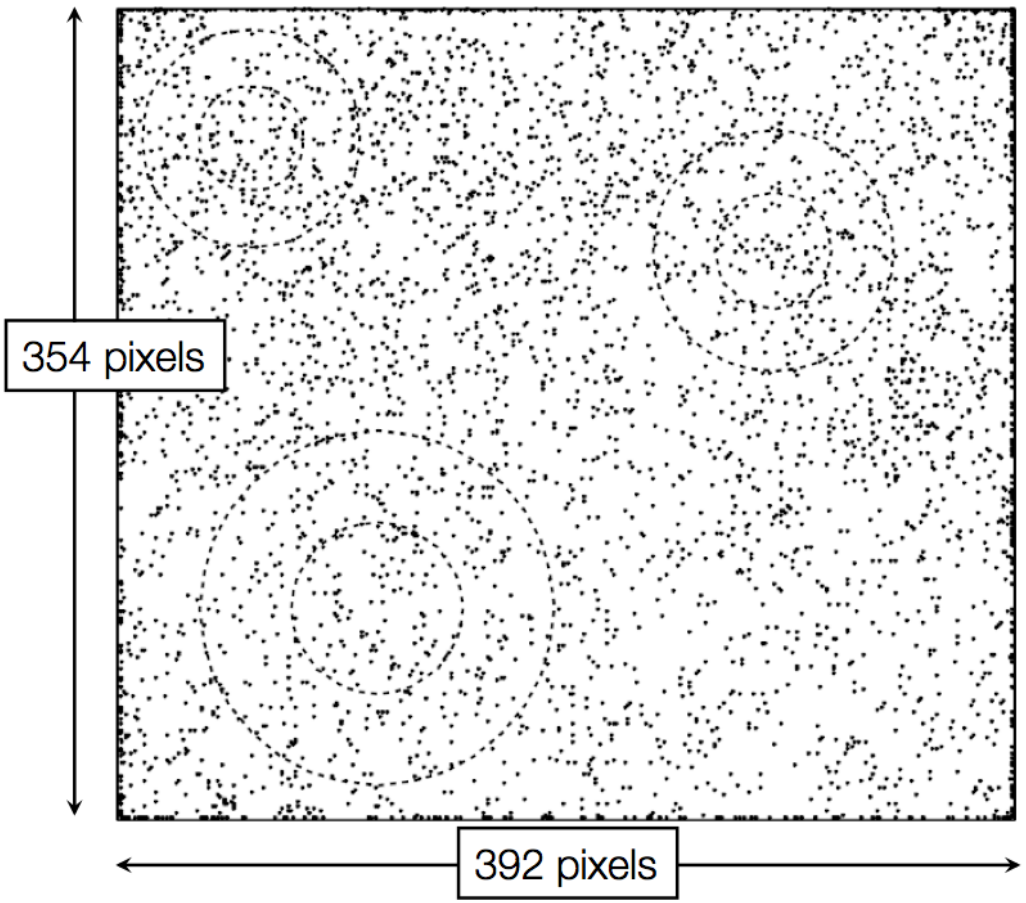


Proposal scheme:

- random direction
- gamma-distributed step length (mean 45 pixels, s.d. 40 pixels)
- reflection at edges

Pure Random Walk

[courtesy of Paul O Lewis]



Proposal scheme:

- random direction
- gamma-distributed step length (mean 45 pixels, s.d. 40 pixels)
- reflection at edges

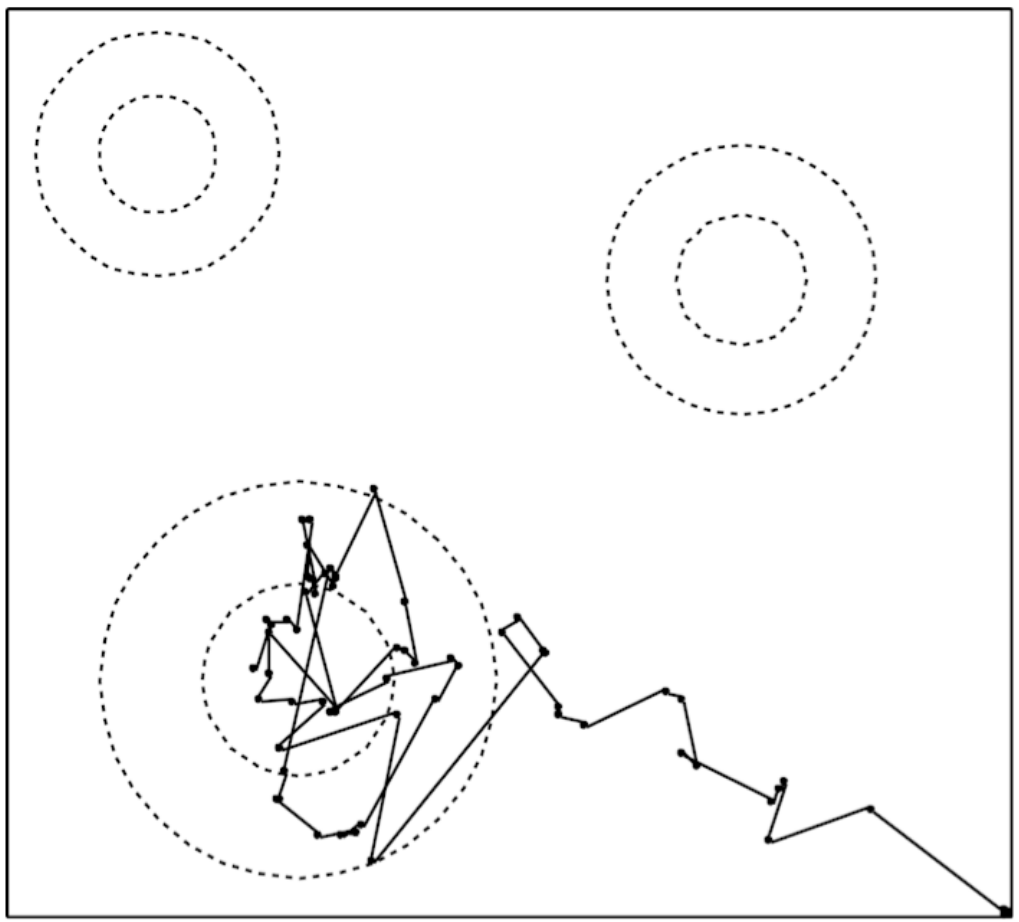
Target distribution:

- equal mixture of 3 bivariate normal "hills"
- inner contours: 50%
- outer contours: 95%

In this case the robot is accepting every step and 5000 steps are shown.

Burn In

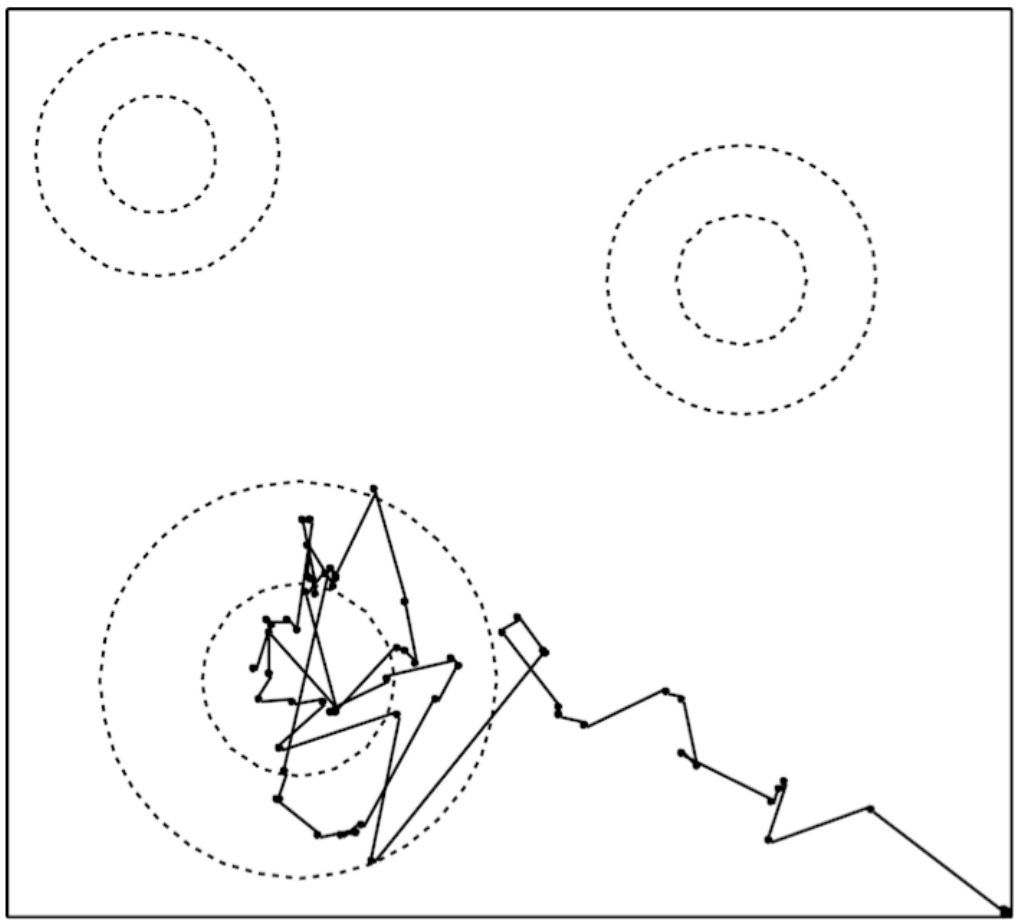
[courtesy of Paul O Lewis]



Robot is now following the Metropolis rules and thus quickly finds one of the three hills.

Burn In

[courtesy of Paul O Lewis]

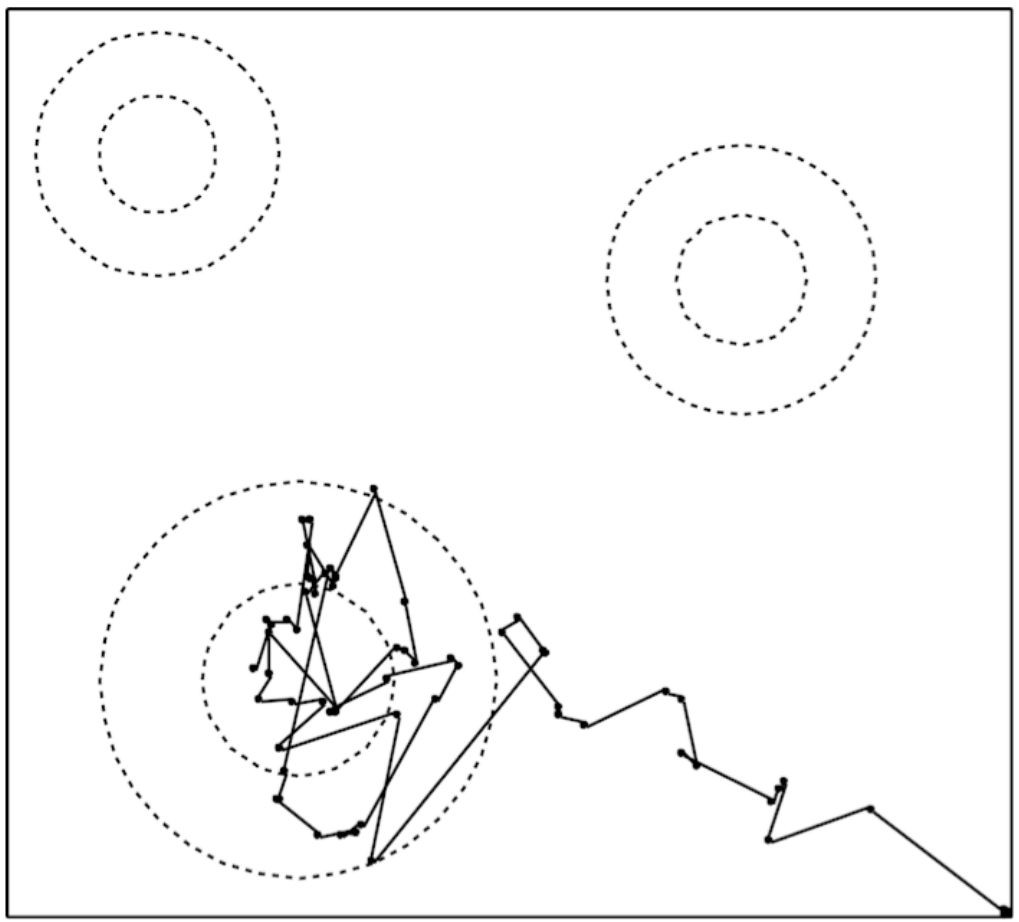


Robot is now following the Metropolis rules and thus quickly finds one of the three hills.

Note that first few steps are not at all representative of the distribution. These steps are called "burn in" and are eliminated from the chain.

Burn In

[courtesy of Paul O Lewis]



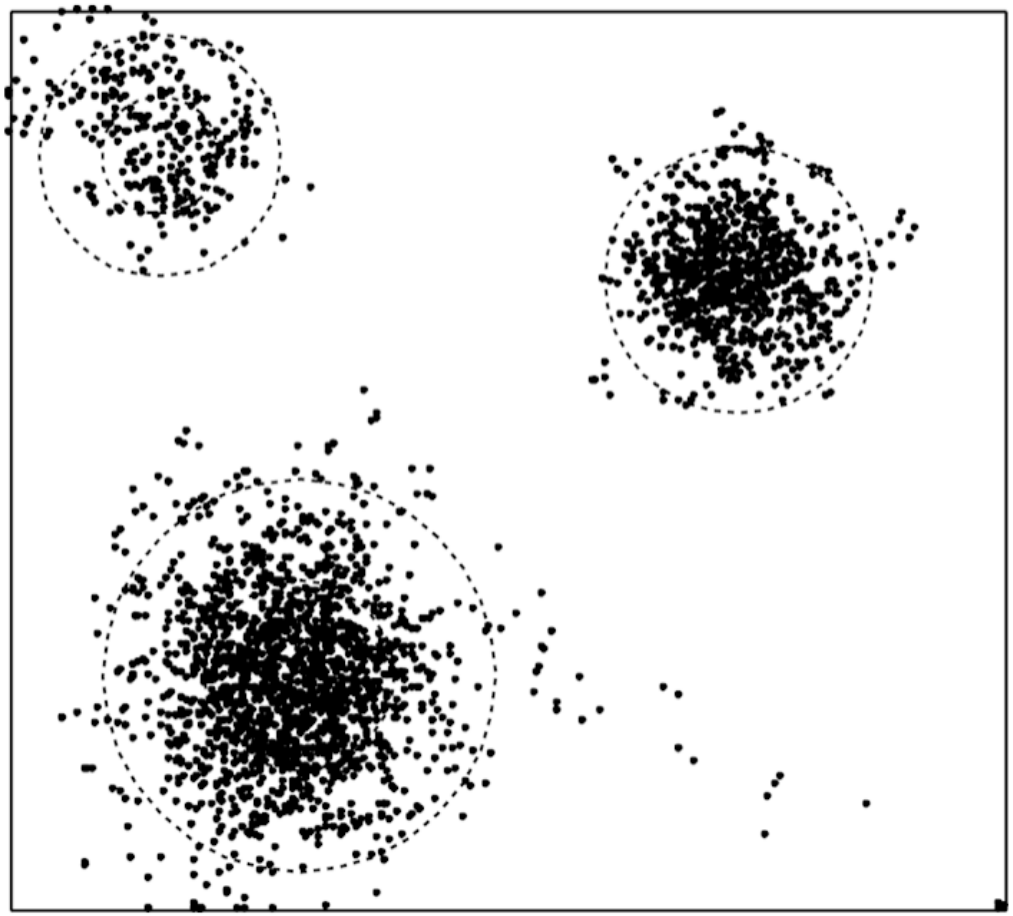
Robot is now following the Metropolis rules and thus quickly finds one of the three hills.

Note that first few steps are not at all representative of the distribution. These steps are called "burn in" and are eliminated from the chain.

100 steps have been taken since the starting point.

Posterior Distribution Approximation

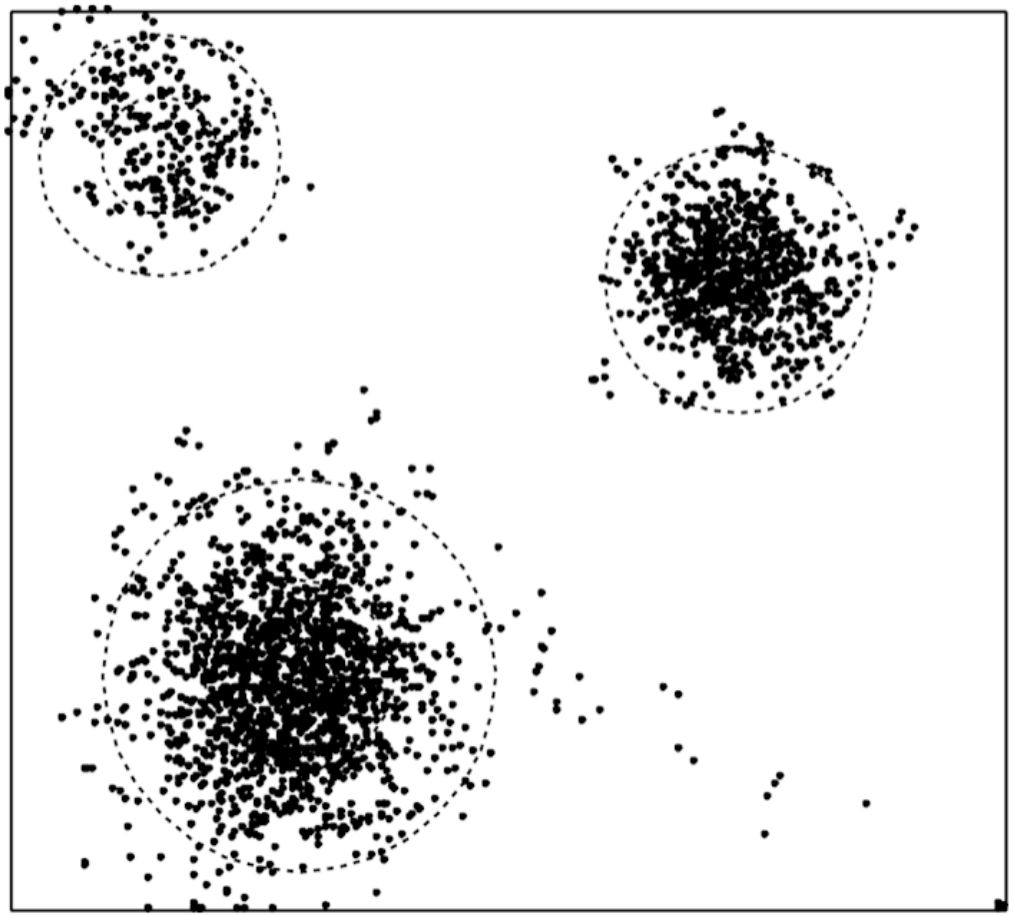
[courtesy of Paul O Lewis]



How good is the MCMC approximation?

Posterior Distribution Approximation

[courtesy of Paul O Lewis]

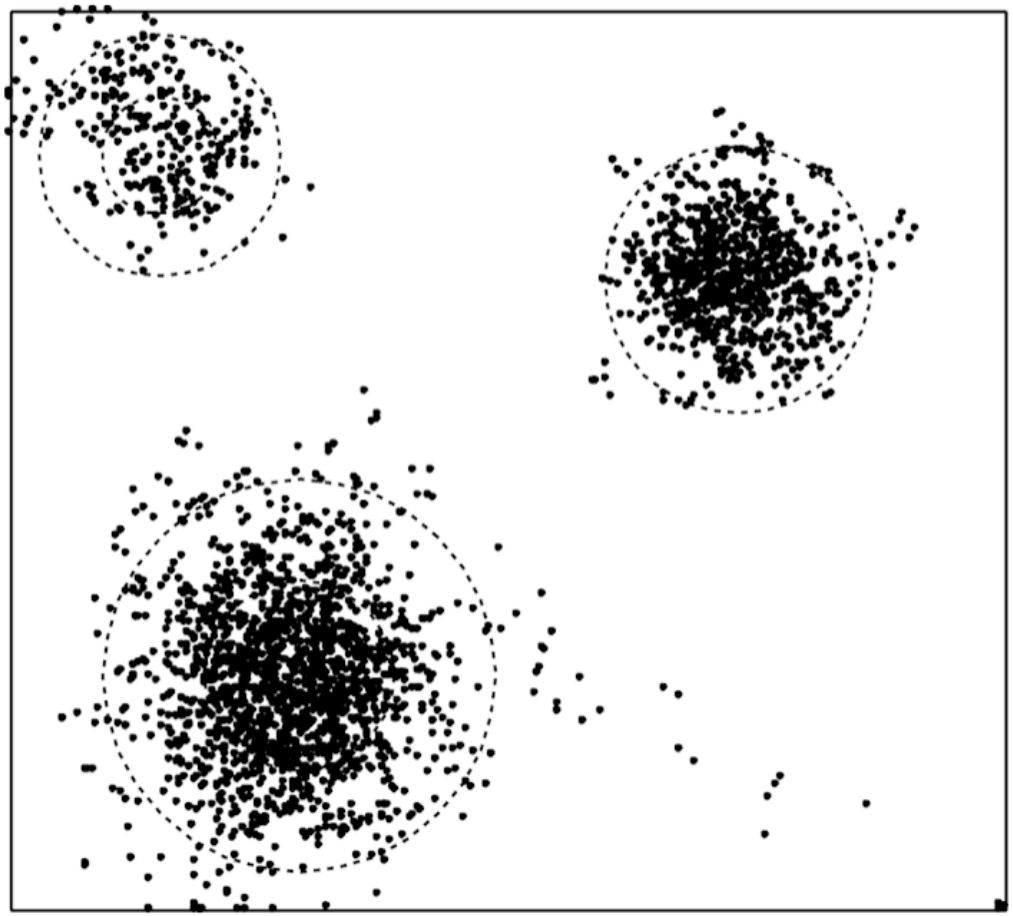


How good is the MCMC approximation?

- 51.2% of points are inside inner contours (cf. 50% actual).
- 93.6% of points are inside outer contours (cf. 95% actual).

Posterior Distribution Approximation

[courtesy of Paul O Lewis]



How good is the MCMC approximation?

- 51.2% of points are inside inner contours (cf. 50% actual).
- 93.6% of points are inside outer contours (cf. 95% actual).

Approximation gets better the longer the chain is allowed to run.

(here 5000 steps have been taken).

The Phylogenetic Likelihood

$$P(A|\tau, Q)$$

- A is a sequence alignment.
- τ is a phylogenetic tree.
- Q is the substitution rate matrix (and possibly other substitution model parameters).

The Phylogenetic Likelihood

$$P(A|\tau, Q)$$

- A is a sequence alignment.
- τ is a phylogenetic tree.
- Q is the substitution rate matrix (and possibly other substitution model parameters).

We are doing Bayesian inference though: we need a probability distribution for τ!

The Phylogenetic Posterior

$$P(\tau, Q, \eta | A) = \frac{1}{P(A)} P(A|\tau, Q) P(\tau|\eta) P(Q, \eta)$$

Here η are the phylodynamic model parameters.

The Phylogenetic Posterior

$$P(\tau, Q, \eta | A) = \frac{1}{P(A)} P(A|\tau, Q) P(\tau|\eta) P(Q, \eta)$$

Here η are the phylodynamic model parameters.

- $P(\tau|\eta)$ is the "tree prior" or "phylodynamic likelihood".

The Phylogenetic Posterior

$$P(\tau, Q, \eta|A) = \frac{1}{P(A)} P(A|\tau, Q) P(\tau|\eta) P(Q, \eta)$$

Here η are the phylodynamic model parameters.

- $P(\tau|\eta)$ is the "tree prior" or "phylodynamic likelihood".
- $P(Q, \eta) = P(Q)P(\eta)$ are the parameter prior distributions.

The Phylogenetic Posterior

$$P(\tau, Q, \eta|A) = \frac{1}{P(A)} P(A|\tau, Q) P(\tau|\eta) P(Q, \eta)$$

Here η are the phylodynamic model parameters.

- $P(\tau|\eta)$ is the "tree prior" or "phylodynamic likelihood".
- $P(Q, \eta) = P(Q)P(\eta)$ are the parameter prior distributions.

Questions

- Is the tree prior really a prior? (Does it depend on data?)

The Phylogenetic Posterior

$$P(\tau, Q, \eta|A) = \frac{1}{P(A)} P(A|\tau, Q) P(\tau|\eta) P(Q, \eta)$$

Here η are the phylodynamic model parameters.

- $P(\tau|\eta)$ is the "tree prior" or "phylodynamic likelihood".
- $P(Q, \eta) = P(Q)P(\eta)$ are the parameter prior distributions.

Questions

- Is the tree prior really a prior? (Does it depend on data?)
- What is $P(A)$?

The Phylogenetic Posterior

$$P(\tau, Q, \eta|A) = \frac{1}{P(A)} P(A|\tau, Q) P(\tau|\eta) P(Q, \eta)$$

Here η are the phylodynamic model parameters.

- $P(\tau|\eta)$ is the "tree prior" or "phylodynamic likelihood".
- $P(Q, \eta) = P(Q)P(\eta)$ are the parameter prior distributions.

Questions

- Is the tree prior really a prior? (Does it depend on data?)
- What is $P(A)$?
- Is $P(A)$ feasible to calculate directly?

Features of Bayesian Phylogenetic Inference

Some of the practical characteristics of the Bayesian approach include:

Features of Bayesian Phylogenetic Inference

Some of the practical characteristics of the Bayesian approach include:

- The approach jointly infers the phylogenetic tree, the substitution model parameters and the phylodynamic model parameters.

Features of Bayesian Phylogenetic Inference

Some of the practical characteristics of the Bayesian approach include:

- The approach jointly infers the phylogenetic tree, the substitution model parameters and the phylodynamic model parameters.
- The approach correctly accounts for uncertainty both in the phylogenetic tree itself (due to our stochastic models of sequence evolution) and in the model parameters.

Features of Bayesian Phylogenetic Inference

Some of the practical characteristics of the Bayesian approach include:

- The approach jointly infers the phylogenetic tree, the substitution model parameters and the phylodynamic model parameters.
- The approach correctly accounts for uncertainty both in the phylogenetic tree itself (due to our stochastic models of sequence evolution) and in the model parameters.
- Additional sources of information are straight-forward to include. (E.g. prior information about parameter values, constraints on tree topology, etc.)

Features of Bayesian Phylogenetic Inference

Some of the practical characteristics of the Bayesian approach include:

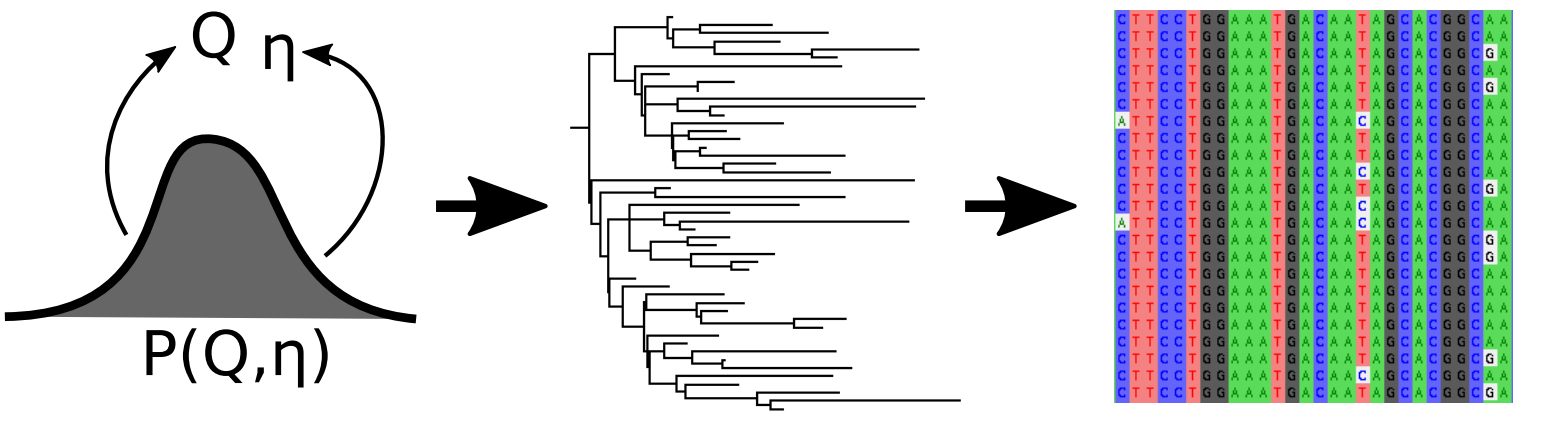
- The approach jointly infers the phylogenetic tree, the substitution model parameters and the phylodynamic model parameters.
- The approach correctly accounts for uncertainty both in the phylogenetic tree itself (due to our stochastic models of sequence evolution) and in the model parameters.
- Additional sources of information are straight-forward to include. (E.g. prior information about parameter values, constraints on tree topology, etc.)
- Resulting posterior distributions naturally include the uncertainty in the inference results.

Aside: neutrality assumption

Because of the way we have factorized the joint probability for the tree and model parameters, we are implicitly assuming our alignment could have been produced in the following fashion:

Aside: neutrality assumption

Because of the way we have factorized the joint probability for the tree and model parameters, we are implicitly assuming our alignment could have been produced in the following fashion:



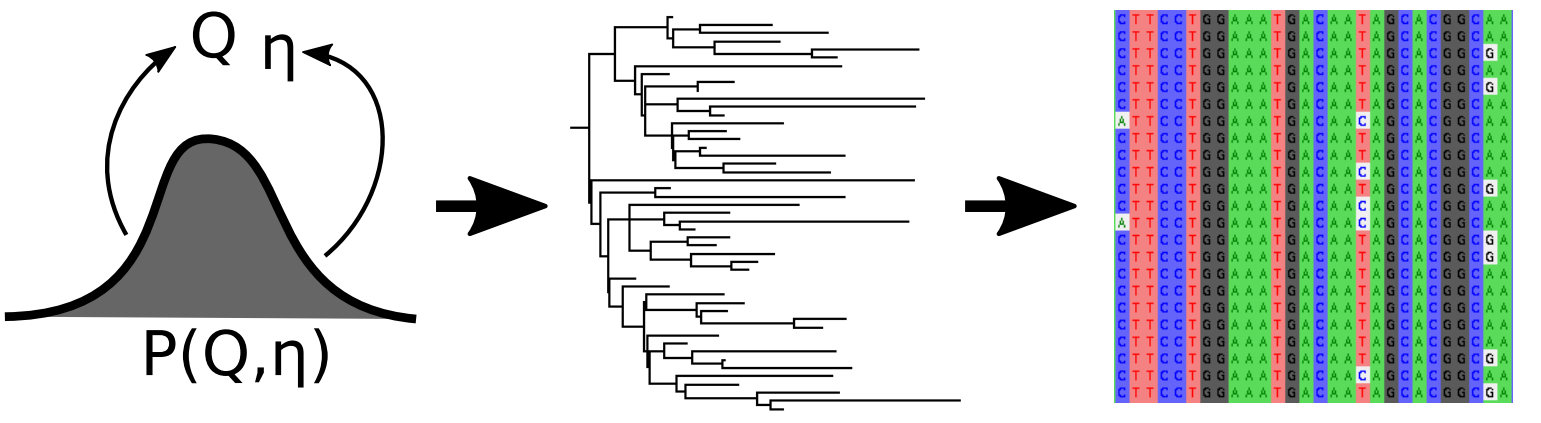
Sample parameters from prior

Sample tree from prior

Simulate sequence alignment

Aside: neutrality assumption

Because of the way we have factorized the joint probability for the tree and model parameters, we are implicitly assuming our alignment could have been produced in the following fashion:



Sample parameters from prior

Sample tree from prior

Simulate sequence alignment

Separating the process of tree generation from that of sequence evolution implies the sequence evolution is effectively neutral.

MCMC with Metropolis-Hastings algorithm

$$P(\tau, Q, \eta|A) = \frac{1}{P(A)} P(A|\tau, Q) P(\tau|\eta) P(Q, \eta)$$

- The MCMC algorithm proposes new state τ', θ', η' based on state τ, θ, η and evaluates the numerator of Bayes formula.

MCMC with Metropolis-Hastings algorithm

$$P(\tau, Q, \eta|A) = \frac{1}{P(A)} P(A|\tau, Q) P(\tau|\eta) P(Q, \eta)$$

- The MCMC algorithm proposes new state τ', θ', η' based on state τ, θ, η and evaluates the numerator of Bayes formula.
  - New phylogenetic tree τ' is proposed using specialized tree-space proposal distributions.

MCMC with Metropolis-Hastings algorithm

$$P(\tau, Q, \eta|A) = \frac{1}{P(A)} P(A|\tau, Q) P(\tau|\eta) P(Q, \eta)$$

- The MCMC algorithm proposes new state τ', θ', η' based on state τ, θ, η and evaluates the numerator of Bayes formula.
  - New phylogenetic tree τ' is proposed using specialized tree-space proposal distributions.
  - Other parameters are real scalar variables and new states can be proposed via random scaling, uniform random walks, etc.

MCMC with Metropolis-Hastings algorithm

$$P(\tau, Q, \eta|A) = \frac{1}{P(A)} P(A|\tau, Q) P(\tau|\eta) P(Q, \eta)$$

- The MCMC algorithm proposes new state τ', θ', η' based on state τ, θ, η and evaluates the numerator of Bayes formula.
  - New phylogenetic tree τ' is proposed using specialized tree-space proposal distributions.
  - Other parameters are real scalar variables and new states can be proposed via random scaling, uniform random walks, etc.
- Acceptance/rejection of the new state leads eventually to a set of accepted states which is a sample from the posterior distribution $P(\tau, Q, \eta|D)$.

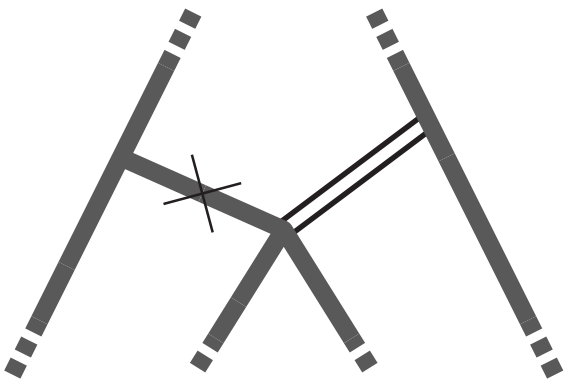
Proposal distributions for tree space

Require a set of proposal distributions $\mathbf{q_i}(\tau'|\tau)$ where τ is a point in the space of rooted time trees.

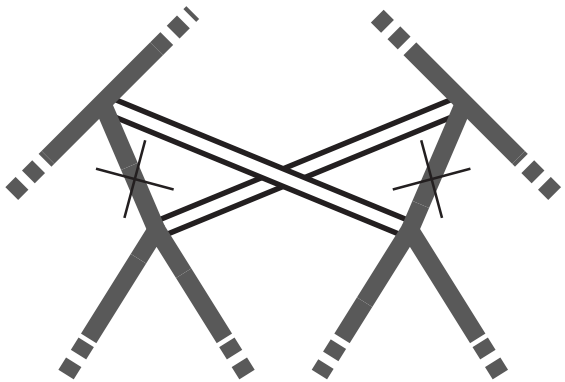
Proposal distributions for tree space

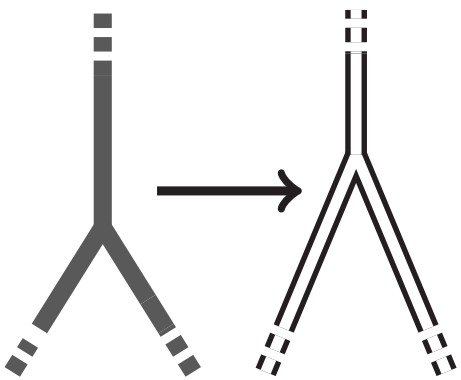
Require a set of proposal distributions $\mathbf{q_i}(\tau'|\tau)$ where τ is a point in the space of rooted time trees.

Wilson-Balding

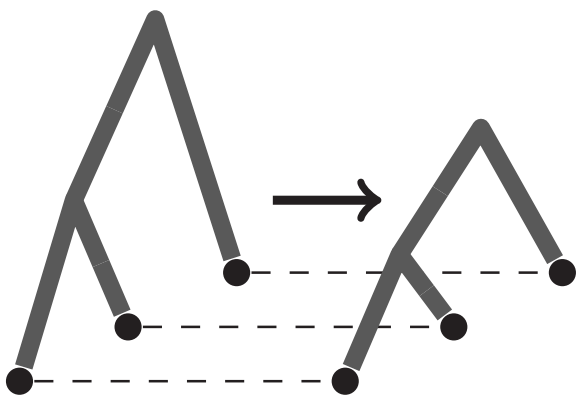


Subtree Exchange





Uniform node height



Tree Scaling

Bayesian phylogenetic inference software

Popular Bayesian phylogenetic inference software:

MrBayes [Huelsenbeck et al., 2001]
Early command-line program for phylogenetic inference.

RevBayes [Höhna et al., 2015]
R-like syntax for specifying phylogenetic models.

BEAST/BEAST2 [Drummond and Rambaut, 2007, Bouckaert et al., 2014, Bouckaert et al., 2019]
XML specification of phylogenetic models.

Bayesian phylogenetic inference software

Popular Bayesian phylogenetic inference software:

MrBayes [Huelsenbeck et al., 2001]
Early command-line program for phylogenetic inference.

RevBayes [Höhna et al., 2015]
R-like syntax for specifying phylogenetic models.

BEAST/BEAST2 [Drummond and Rambaut, 2007, Bouckaert et al., 2014, Bouckaert et al., 2019]
XML specification of phylogenetic models.

Some software implementing special models:

MIGRATE [Beerli and Felsenstein, 2001]
Performs inference under the Structured Coalescent (accounts for population structure).

ClonalFrame/ClonalOrigin
[Didelot and Falush, 2007, Didelot et al., 2010]
Infers bacterial Ancestral Recombination Graphs (generalization of phylogenetic trees when recombination is present).

Bayesian phylogenetics in practice

- Based on sequencing data (and potentially fossils), dated trees together with the evolutionary and population dynamic parameters are inferred.

Bayesian phylogenetics in practice

- Based on sequencing data (and potentially fossils), dated trees together with the evolutionary and population dynamic parameters are inferred.
- We obtain a set of trees and parameters – a sample from the posterior distribution – which naturally allows us to assess the uncertainty (i.e. variance in the parameter estimates).

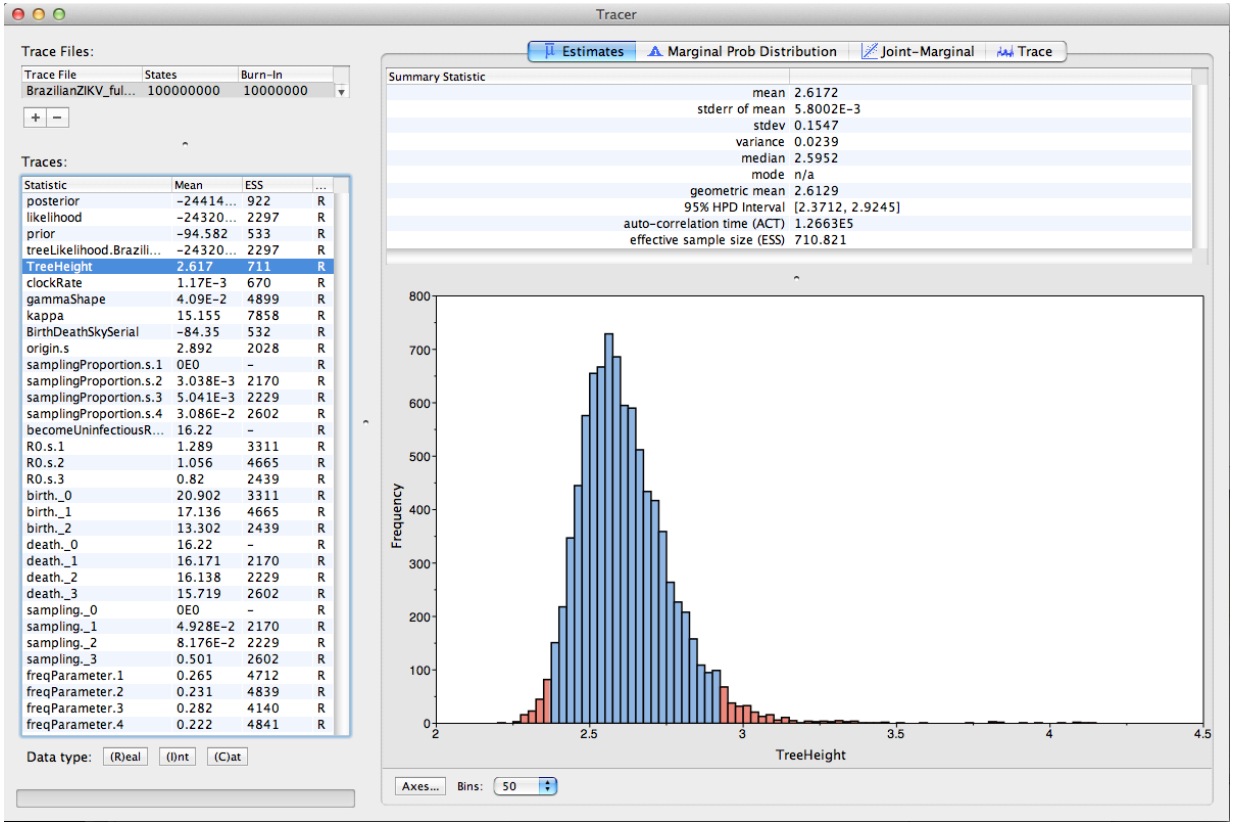
Bayesian phylogenetics in practice

- Based on sequencing data (and potentially fossils), dated trees together with the evolutionary and population dynamic parameters are inferred.
- We obtain a set of trees and parameters – a sample from the posterior distribution – which naturally allows us to assess the uncertainty (i.e. variance in the parameter estimates).
- The samples visited by the MCMC algorithm (in principle draws from the posterior distribution) are recorded.

Bayesian phylogenetics in practice

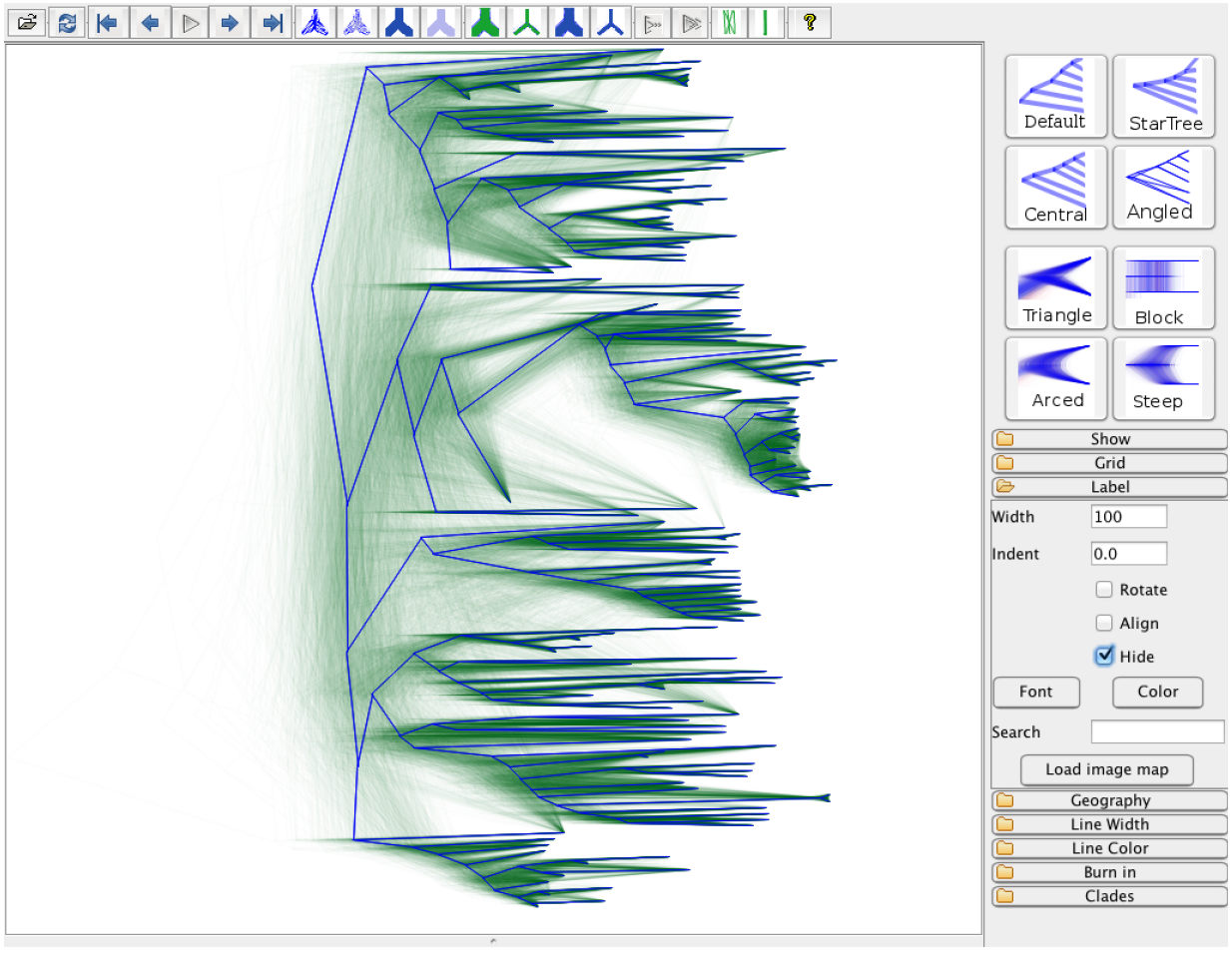
- Based on sequencing data (and potentially fossils), dated trees together with the evolutionary and population dynamic parameters are inferred.
- We obtain a set of trees and parameters – a sample from the posterior distribution – which naturally allows us to assess the uncertainty (i.e. variance in the parameter estimates).
- The samples visited by the MCMC algorithm (in principle draws from the posterior distribution) are recorded.
  - A chain of length N will result in N trees and N values for each parameter.

Summarizing the tree height marginal posterior



Posterior distribution of the tree height based on the Zika samples in South America, visualized in Tracer; tree height is the time between the first branching event and the most recent sample.

Summarizing the phylogenetic tree posterior



Posterior distribution of a Zika virus phylogeny , visualized in DensiTree.

Application: quantifying the spread of Ebola

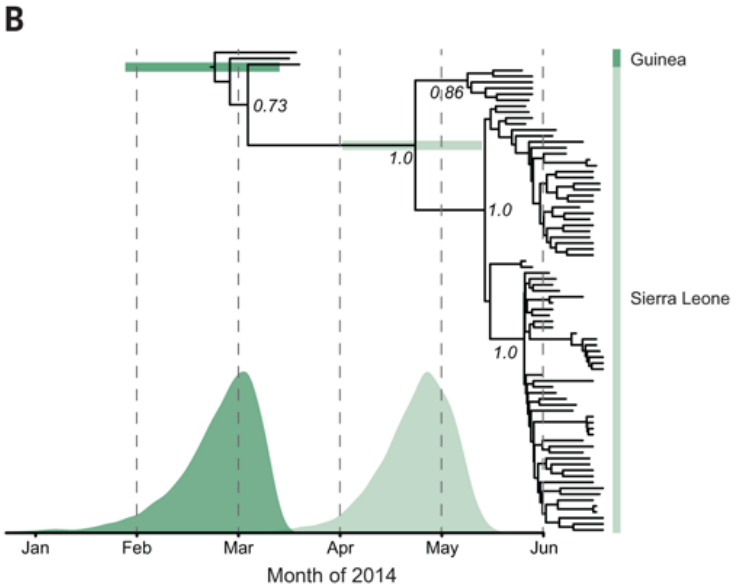


Figure adapted from [Gire et al., 2014]

Phylogenetic tree of Ebola in West Africa.

- [Stadler et al., 2014] obtained the maximum likelihood estimates $\hat{\beta}$ and $\hat{\delta}$ using the phylodynamic likelihood, and estimated $R_0 = \hat{\beta}/\hat{\delta}$ (1.34 (1.12 - 1.55)).

Application: quantifying the spread of Ebola

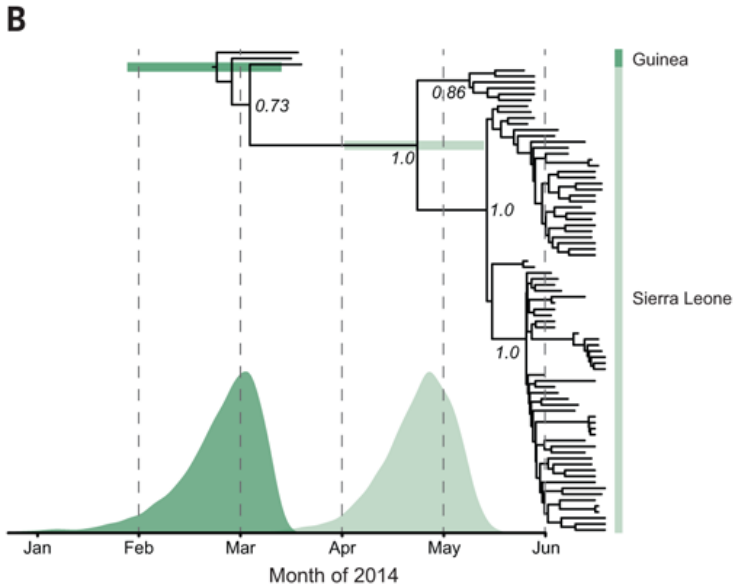


Figure adapted from [Gire et al., 2014]

Phylogenetic tree of Ebola in West Africa.

- [Stadler et al., 2014] obtained the maximum likelihood estimates $\hat{\beta}$ and $\hat{\delta}$ using the phylodynamic likelihood, and estimated $R_0 = \hat{\beta}/\hat{\delta}$ (1.34 (1.12 - 1.55)).
- However, uncertainty in phylogenetic tree should contribute to uncertainty in this ML estimate of the parameters and R_0.

Application: quantifying the spread of Ebola

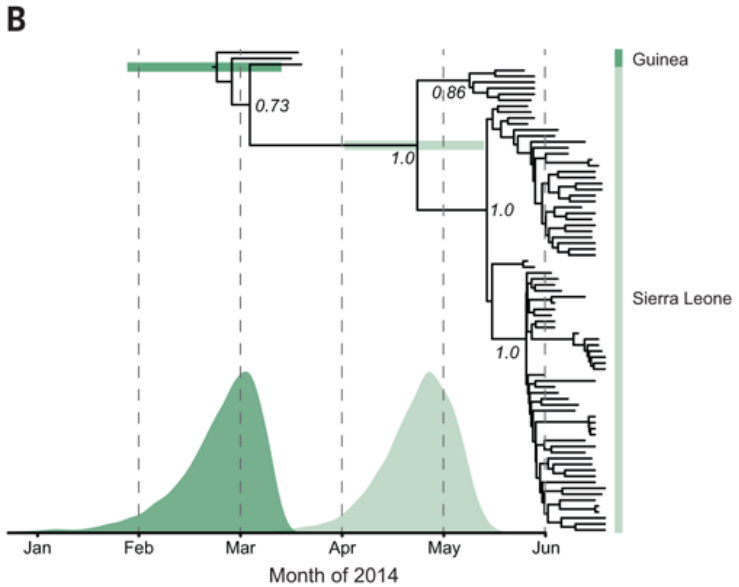
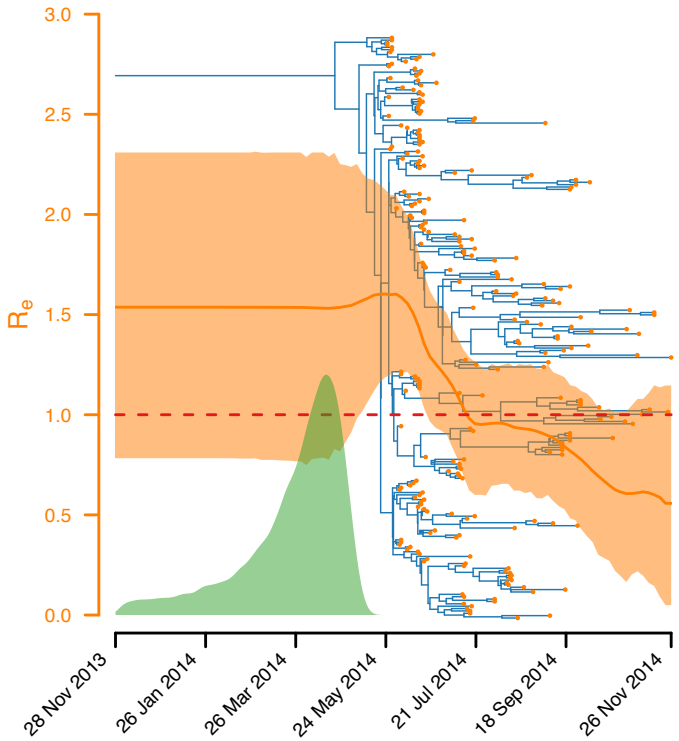


Figure adapted from [Gire et al., 2014]

Phylogenetic tree of Ebola in West Africa.

- [Stadler et al., 2014] obtained the maximum likelihood estimates $\hat{\beta}$ and $\hat{\delta}$ using the phylodynamic likelihood, and estimated $R_0 = \hat{\beta}/\hat{\delta}$ (1.34 (1.12 - 1.55)).
- However, uncertainty in phylogenetic tree should contribute to uncertainty in this ML estimate of the parameters and R_0.
- Also, ML result does not incorporate prior knowledge of these parameters.

Application: quantifying the spread of Ebola



Sierra Leone was first affected by Ebola in the South-East. Here we plot one tree, the posterior distribution of the start of the epidemic, and the median with 95% HPDs for the reproductive number, all obtained from a Bayesian analysis in BEAST2. Analysis performed by Louis du Plessis.

Application: dating the penguin phylogeny

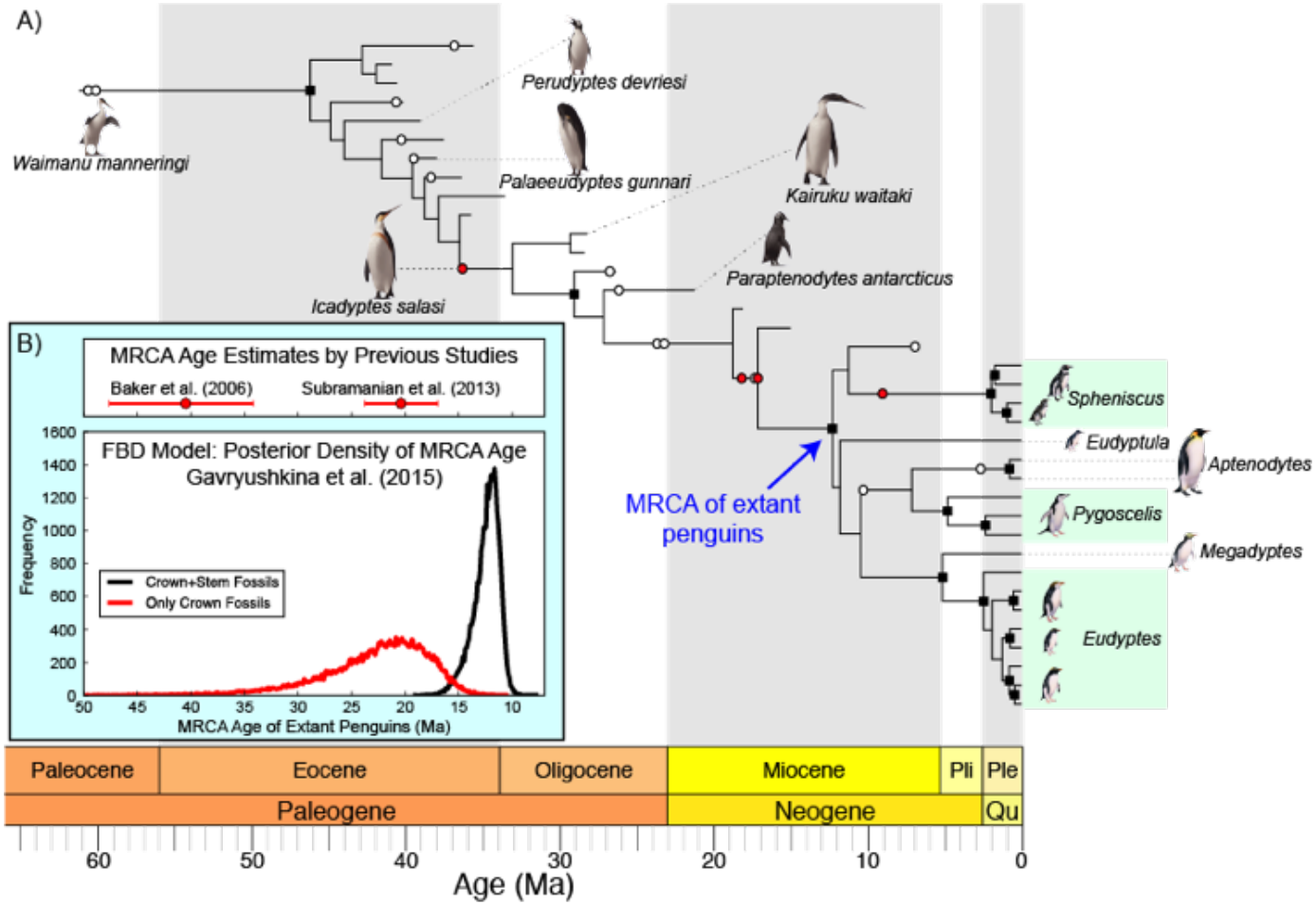


Figure adapted from [Gavryushkina et al., 2017]

Penguin phylogeny obtained from fossil dates, morphology, and extant species sequencing data, again using BEAST2.

Summary

- Bayesian probability distributions quantify states of knowledge.

Summary

- Bayesian probability distributions quantify states of knowledge.
- Such probabilities apply equally well to data and parameters.

Summary

- Bayesian probability distributions quantify states of knowledge.
- Such probabilities apply equally well to data and parameters.
- Bayes theorem provides a natural way to quantitatively combine new data with prior knowledge of model parameters. The result (the posterior distribution for the parameters) provides a comprehensive representation of the final state of knowledge.

Summary

- Bayesian probability distributions quantify states of knowledge.
- Such probabilities apply equally well to data and parameters.
- Bayes theorem provides a natural way to quantitatively combine new data with prior knowledge of model parameters. The result (the posterior distribution for the parameters) provides a comprehensive representation of the final state of knowledge.
- If priors do not exclude the truth, then two practitioners with the same likelihood functions will converge on the same inference with increasing amounts of data.

Summary

- Bayesian probability distributions quantify states of knowledge.
- Such probabilities apply equally well to data and parameters.
- Bayes theorem provides a natural way to quantitatively combine new data with prior knowledge of model parameters. The result (the posterior distribution for the parameters) provides a comprehensive representation of the final state of knowledge.
- If priors do not exclude the truth, then two practitioners with the same likelihood functions will converge on the same inference with increasing amounts of data.
- Markov chain Monte Carlo is a simple algorithm that can be employed to study large problems.

Bayesian Inference: Questions

- Does a Bayesian phylogenetic analysis of the kind described here allow one to directly infer ancestral sequences? Why/Why not?
- How might we test to see whether a Bayesian MCMC analysis has explored the full state space supported by the posterior?
- Suppose you have conducted a Bayesian phylodynamic analysis and recovered a 95% HPD interval for the birth rate parameter. If you take this result and use it to construct a new prior for this parameter and use this prior to analyze the *same data*, would the resulting second posterior be valid?

References I

- Beerli, P. and Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc Natl Acad Sci U S A*, 98(8):4563–4568.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., and Drummond, A. J. (2014). Beast 2: a software platform for bayesian evolutionary analysis. *PLoS Comput Biol*, 10(4):e1003537.
- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., Maio, N. D., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M. A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., and Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 15(4):e1006650.
- Didelot, X. and Falush, D. (2007). Inference of bacterial microevolution using multilocus sequence data. *Genetics*, 175:1251.
- Didelot, X., Lawson, D., Daarling, A., and Falush, D. (2010). Inference of homologous recombination in bacteria using whole-genome sequences. *Genetics*, 186:1435.
- Drummond, A. J. and Rambaut, A. (2007). Beast: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7(1):1.
- Gavryushkina, A., Heath, T. A., Ksepka, D. T., Stadler, T., Welch, D., and Drummond, A. J. (2017). Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic biology*, 66(1):57–73.
- Gire, S. K., Goba, A., Andersen, K. G., Sealfon, R. S., Park, D. J., Kanneh, L., Jalloh, S., Momoh, M., Fullah, M., Dudas, G., et al. (2014). Genomic surveillance elucidates ebola virus origin and transmission during the 2014 outbreak. *Science*, 345(6202):1369–1372.
- Höhna, S., Landis, M., Heath, T., Boussau, B., Lartillot, N., Moore, B., Huelsenbeck, J., and Ronquist, F. (2015). Revbayes: A flexible framework for bayesian inference of phylogeny. *Systematic Biology*, 64(5).
- Huelsenbeck, J. P., Ronquist, F., et al. (2001). Mrbayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8):754–755.
- Stadler, T., Kühnert, D., Rasmussen, D. A., and du Plessis, L. (2014). Insights into the early epidemic spread of ebola in sierra leone provided by viral sequence data. *PLoS currents*, 6.