Beta diffusion trees

Joint work with

David A. Knowles (Stanford)
Zoubin Ghahramani (Cambridge)

Outline

The Dirichlet diffusion tree (DDT; Neal, 2003) is a random tree structure over non-overlapping clusters of objects (aka partitions).

Outline

The Dirichlet diffusion tree (DDT; Neal, 2003) is a random tree structure over non-overlapping clusters of objects (aka partitions).

We are interested in random tree structures on overlapping clusters (aka feature allocations).

Outline

1. (Very quickly) review the Dirichlet diffusion tree and how it provides a tree over partitions.

Outline

1. (Very quickly) review the Dirichlet diffusion tree and how it provides a tree over partitions.
2. Describe a very similar process (the beta diffusion tree) that leads to trees over feature allocations.

Outline

1. (Very quickly) review the Dirichlet diffusion tree and how it provides a tree over partitions.
2. Describe a very similar process (the beta diffusion tree) that leads to trees over feature allocations.
3. Provide some alternative perspectives on this tree structure, which will help us to understand its behavior.

Outline

1. (Very quickly) review the Dirichlet diffusion tree and how it provides a tree over partitions.
2. Describe a very similar process (the beta diffusion tree) that leads to trees over feature allocations.
3. Provide some alternative perspectives on this tree structure, which will help us to understand its behavior.
4. Present some inference techniques over the tree structure and some experimental results.

The Dirichlet diffusion tree (A very quick review)

Let $\alpha > 0$. Consider a sequence of N particles diffusing in a continuous space $\mathcal{X}$ over the unit time interval $[0, 1]$.

The Dirichlet diffusion tree (A very quick review)

Let $\alpha > 0$. Consider a sequence of N particles diffusing in a continuous space $\mathcal{X}$ over the unit time interval $[0, 1]$.

1. Particles begin by following the path of previous particles, possibly diverging from the path at random times.

The Dirichlet diffusion tree (A very quick review)

Let $\alpha > 0$. Consider a sequence of N particles diffusing in a continuous space $\mathcal{X}$ over the unit time interval $[0, 1]$.

1. Particles begin by following the path of previous particles, possibly diverging from the path at random times.
2. A particle on a path may diverge from the path after

   $$\exp(\alpha/m)$$

   time, where m is the # particles previously down the path.

The Dirichlet diffusion tree (A very quick review)

Let $\alpha > 0$. Consider a sequence of N particles diffusing in a continuous space $\mathcal{X}$ over the unit time interval $[0, 1]$.

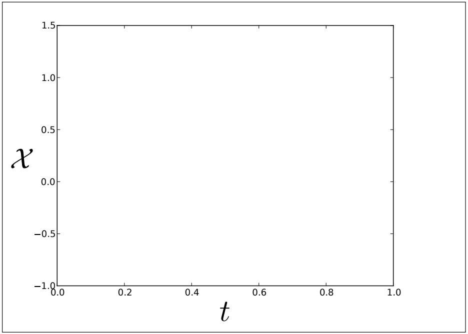
1. Particles begin by following the path of previous particles, possibly diverging from the path at random times.
2. A particle on a path may diverge from the path after
   $$\exp(\alpha/m)$$
   time, where m is the # particles previously down the path.
3. At previous divergence points, a particle chooses a path with prob. proportional to the number previously down the path.

The Dirichlet diffusion tree (A very quick review)

Let $\alpha > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.

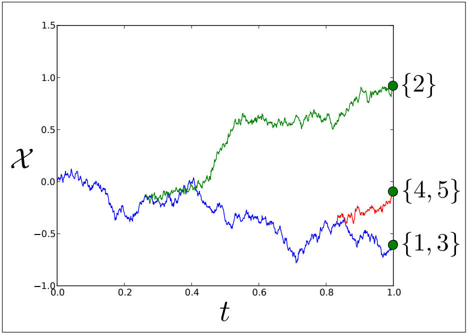
The Dirichlet diffusion tree (A very quick review)

Let $\alpha > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



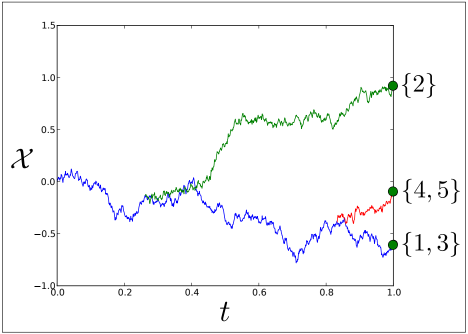
The Dirichlet diffusion tree (A very quick review)

Let $\alpha > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



The Dirichlet diffusion tree (A very quick review)

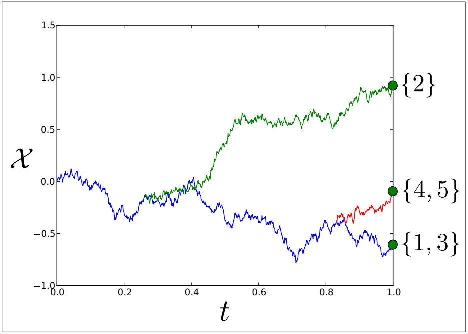
Let $\alpha > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



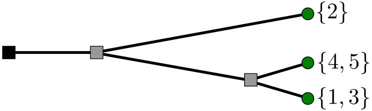
1. Follow previous particles.
2. Possibly diverge after an $\exp(\alpha/m)$ amount of time.
3. Choose branches w.p. $\propto$ the number of particles down the branch.

The Dirichlet diffusion tree (A very quick review)

Let $\alpha > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.

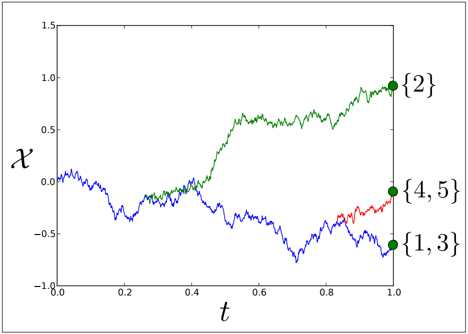


The corresponding tree structure:

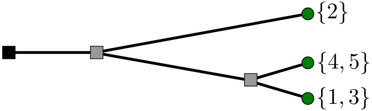


The Dirichlet diffusion tree (A very quick review)

Let $\alpha > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



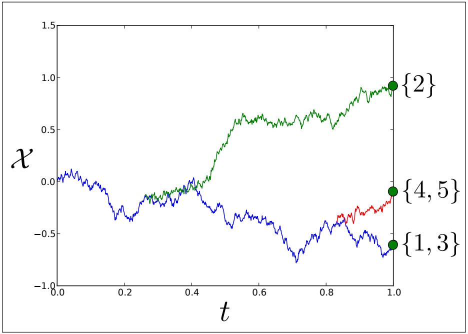
The corresponding **tree structure**:



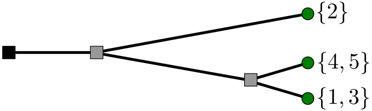
- Each particle represents an object.

The Dirichlet diffusion tree (A very quick review)

Let $\alpha > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



The corresponding tree structure:



- Each particle represents an object.
- The leaves form a partition of $[5] := \{1, \ldots, 5\}$.

The beta diffusion tree

The beta diffusion tree will proceed very similarly, except that copies of a particle may be created or destroyed at random times.

The beta diffusion tree

The beta diffusion tree will proceed very similarly, except that copies of a particle may be created or destroyed at random times.

Multiple copies of each particle (representing the objects) may exist and follow multiple paths down the tree.

The beta diffusion tree

Let $\lambda_r, \lambda_s, \theta_r, \theta_s > 0$. Proceed exactly as in the DDT, except:

The beta diffusion tree

Let $\lambda_r, \lambda_s, \theta_r, \theta_s > 0$. Proceed exactly as in the DDT, except:

1. A particle on a path may replicate (make a copy of itself) after
$$\exp\left(\frac{\theta_r}{\theta_r + m}\lambda_r\right) \qquad \text{time.}$$

The beta diffusion tree

Let $\lambda_r, \lambda_s, \theta_r, \theta_s > 0$. Proceed exactly as in the DDT, except:

1. A particle on a path may replicate (make a copy of itself) after

$$\exp\left(\frac{\theta_r}{\theta_r + m}\lambda_r\right) \qquad \text{time.}$$

2. A particle on a path may stop diffusing after

$$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right) \qquad \text{time.}$$

The beta diffusion tree

Let $\lambda_r, \lambda_s, \theta_r, \theta_s > 0$. Proceed exactly as in the DDT, except:

1. A particle on a path may replicate (make a copy of itself) after
$$\exp\left(\frac{\theta_r}{\theta_r + m}\lambda_r\right) \qquad \text{time.}$$
2. A particle on a path may stop diffusing after
$$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right) \qquad \text{time.}$$
3. At a replicate point, make a copy of the particle with prob.
$$\frac{n_r}{\theta_r + m}, \qquad n_r \text{ is \# that did so previously.}$$

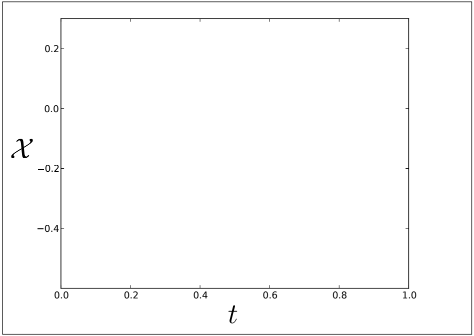
The beta diffusion tree

Let $\lambda_r, \lambda_s, \theta_r, \theta_s > 0$. Proceed exactly as in the DDT, except:

1. A particle on a path may replicate (make a copy of itself) after
$$\exp\left(\frac{\theta_r}{\theta_r + m}\lambda_r\right) \qquad \text{time.}$$
2. A particle on a path may stop diffusing after
$$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right) \qquad \text{time.}$$
3. At a replicate point, make a copy of the particle with prob.
$$\frac{n_r}{\theta_r + m}, \qquad n_r \text{ is \# that did so previously.}$$
4. At a stop point, also stop with prob.
$$\frac{n_s}{\theta_s + m}, \qquad n_s \text{ is \# that did so previously.}$$

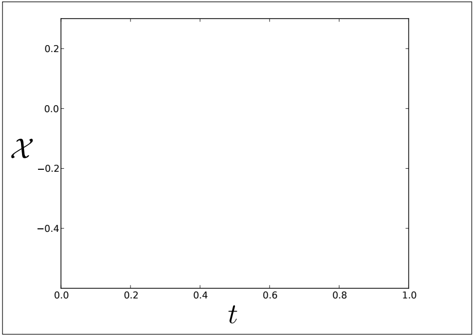
The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



The beta diffusion tree

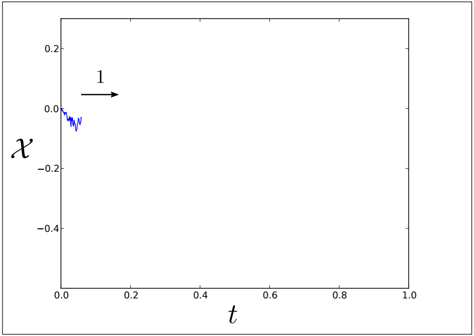
Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



Consider the first object.

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.

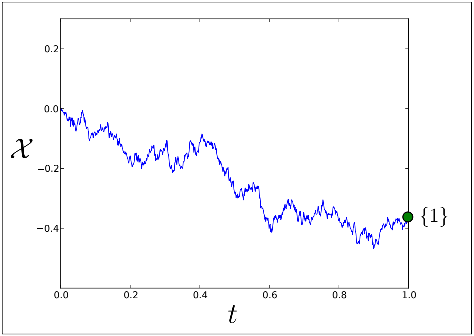


1. The first particle starts at the origin.
2. May replicate or stop after $\exp(\lambda_r)$ or $\exp(\lambda_s)$.

Consider the first object.

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.

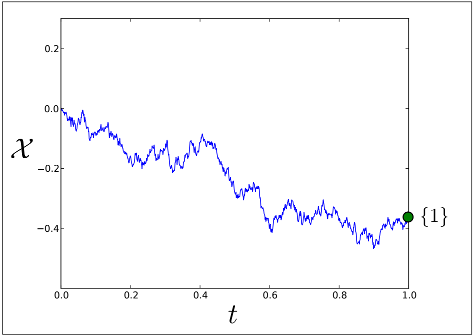


1. The first particle starts at the origin.
2. May replicate or stop after $\exp(\lambda_r)$ or $\exp(\lambda_s)$.
3. Diffuses as Brownian motion until $t = 1$.

Consider the first object.

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



1. Follow previous particles.
2. May replicate or stop after

   $$\exp\Big(\frac{\theta_r}{\theta_r + m}\lambda_r\Big)$$

   or

   $$\exp\Big(\frac{\theta_s}{\theta_s + m}\lambda_s\Big)$$

   amount of time.
3. Replicate and stop w.p.

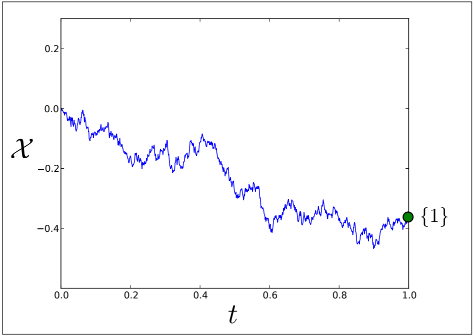
   $$\frac{n_r}{\theta_r + m}$$

   and

   $$\frac{n_s}{\theta_s + m}$$

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



1. Follow previous particles.
2. May replicate or stop after

$$\exp\left(\frac{\theta_r}{\theta_r + m}\lambda_r\right)$$

or

$$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right)$$

amount of time.

3. Replicate and stop w.p.

$$\frac{n_r}{\theta_r + m}$$

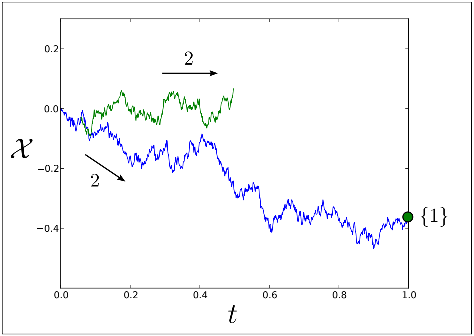
and

$$\frac{n_s}{\theta_s + m}$$

Consider the second object.

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



Consider the second object.

1. Follow previous particles.
2. May replicate or stop after

   $$\exp\left(\frac{\theta_r}{\theta_r + m}\lambda_r\right)$$

   or

   $$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right)$$

   amount of time.
3. Replicate and stop w.p.

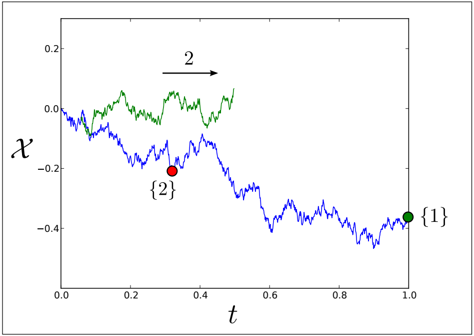
   $$\frac{n_r}{\theta_r + m}$$

   and

   $$\frac{n_s}{\theta_s + m}$$

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



1. Follow previous particles.
2. May replicate or stop after

$$\exp\Big(\frac{\theta_r}{\theta_r + m}\lambda_r\Big)$$

or

$$\exp\Big(\frac{\theta_s}{\theta_s + m}\lambda_s\Big)$$

amount of time.
3. Replicate and stop w.p.

$$\frac{n_r}{\theta_r + m}$$

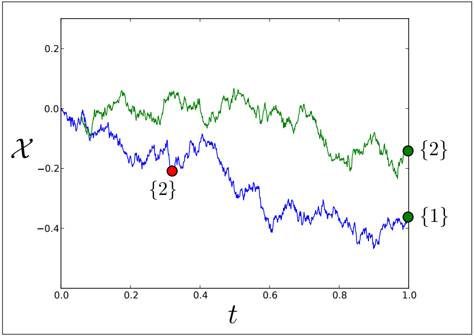
and

$$\frac{n_s}{\theta_s + m}$$

Consider the second object.

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



1. Follow previous particles.
2. May replicate or stop after

   $$\exp\left(\frac{\theta_r}{\theta_r + m}\lambda_r\right)$$

   or

   $$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right)$$

   amount of time.
3. Replicate and stop w.p.

   $$\frac{n_r}{\theta_r + m}$$

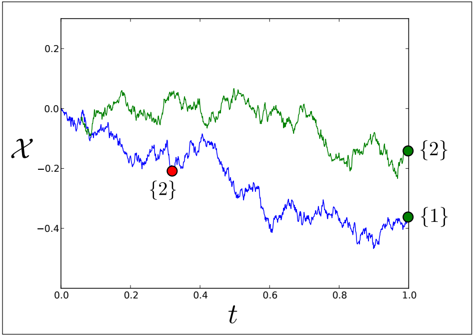
   and

   $$\frac{n_s}{\theta_s + m}$$

Consider the second object.

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



1. Follow previous particles.
2. May replicate or stop after

   $$\exp\Big(\frac{\theta_r}{\theta_r + m}\lambda_r\Big)$$

   or

   $$\exp\Big(\frac{\theta_s}{\theta_s + m}\lambda_s\Big)$$

   amount of time.
3. Replicate and stop w.p.

   $$\frac{n_r}{\theta_r + m}$$

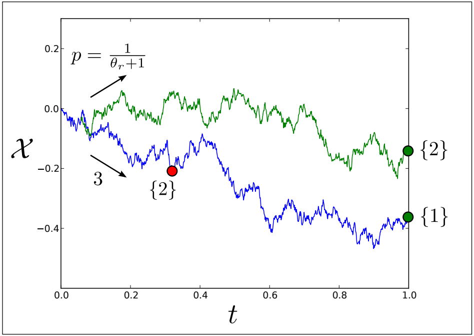
   and

   $$\frac{n_s}{\theta_s + m}$$

Consider the third object.

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



Consider the third object.

1. Follow previous particles.
2. May replicate or stop after

$$\exp\left(\frac{\theta_r}{\theta_r + m}\lambda_r\right)$$

or

$$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right)$$

amount of time.

3. Replicate and stop w.p.

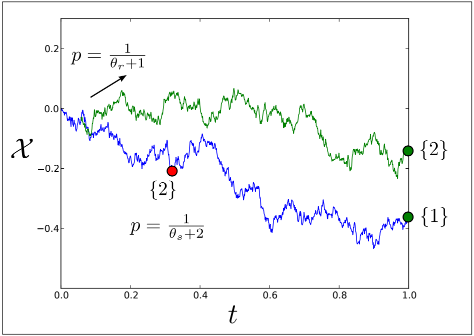
$$\frac{n_r}{\theta_r + m}$$

and

$$\frac{n_s}{\theta_s + m}$$

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



1. Follow previous particles.
2. May replicate or stop after

$$\exp\left(\frac{\theta_r}{\theta_r + m}\lambda_r\right)$$

or

$$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right)$$

amount of time.

3. Replicate and stop w.p.

$$\frac{n_r}{\theta_r + m}$$

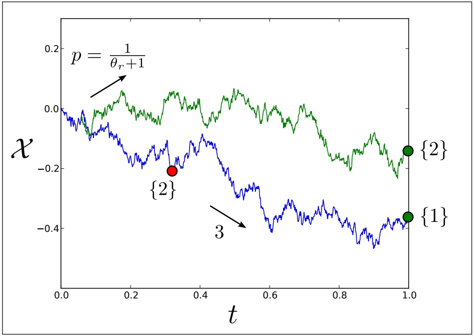
and

$$\frac{n_s}{\theta_s + m}$$

Consider the third object.

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



Consider the third object.

1. Follow previous particles.
2. May replicate or stop after

   $$\exp\left(\frac{\theta_r}{\theta_r + m}\lambda_r\right)$$

   or

   $$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right)$$

   amount of time.
3. Replicate and stop w.p.

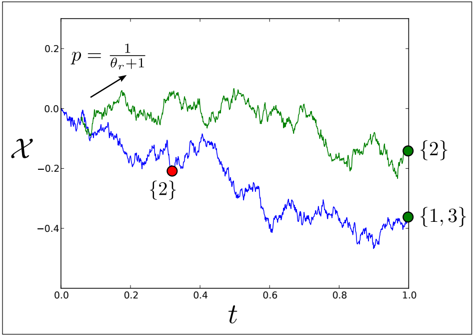
   $$\frac{n_r}{\theta_r + m}$$

   and

   $$\frac{n_s}{\theta_s + m}$$

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



1. Follow previous particles.
2. May replicate or stop after

$$\exp\left(\frac{\theta_r}{\theta_r + m}\lambda_r\right)$$

or

$$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right)$$

amount of time.

3. Replicate and stop w.p.

$$\frac{n_r}{\theta_r + m}$$

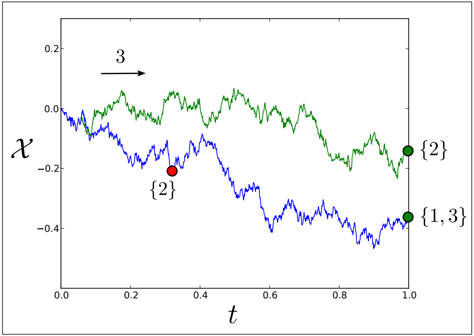
and

$$\frac{n_s}{\theta_s + m}$$

Consider the third object.

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



1. Follow previous particles.
2. May replicate or stop after

   $$\exp\left(\frac{\theta_r}{\theta_r + m}\lambda_r\right)$$

   or

   $$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right)$$

   amount of time.
3. Replicate and stop w.p.

   $$\frac{n_r}{\theta_r + m}$$

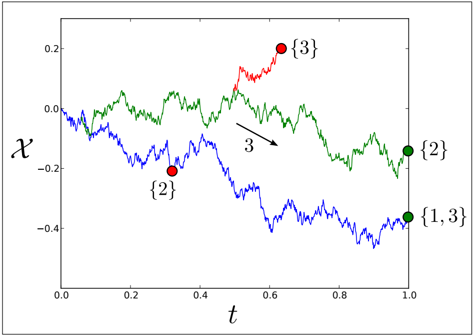
   and

   $$\frac{n_s}{\theta_s + m}$$

Consider the third object.

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



Consider the third object.

1. Follow previous particles.
2. May replicate or stop after

   $$\exp\Big(\frac{\theta_r}{\theta_r + m}\lambda_r\Big)$$

   or

   $$\exp\Big(\frac{\theta_s}{\theta_s + m}\lambda_s\Big)$$

   amount of time.
3. Replicate and stop w.p.

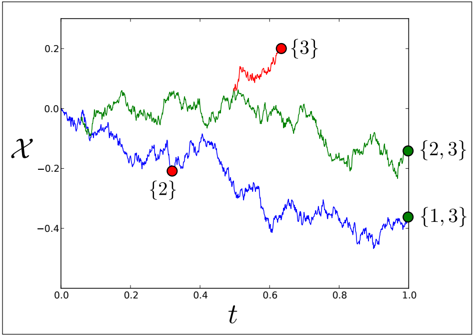
   $$\frac{n_r}{\theta_r + m}$$

   and

   $$\frac{n_s}{\theta_s + m}$$

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



Consider the third object.

1. Follow previous particles.
2. May replicate or stop after

$$\exp\Big(\frac{\theta_r}{\theta_r + m}\lambda_r\Big)$$

or

$$\exp\Big(\frac{\theta_s}{\theta_s + m}\lambda_s\Big)$$

amount of time.

3. Replicate and stop w.p.

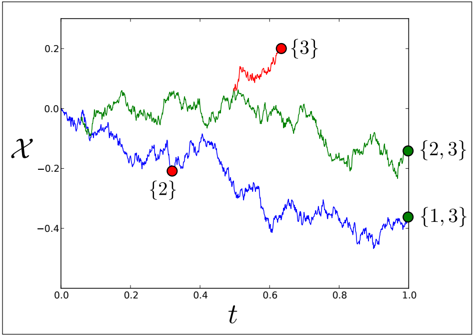
$$\frac{n_r}{\theta_r + m}$$

and

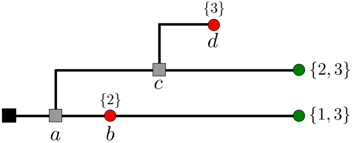
$$\frac{n_s}{\theta_s + m}$$

The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.

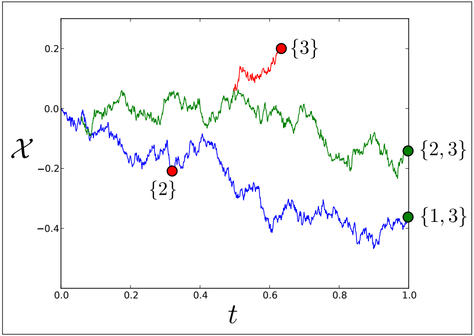


The corresponding tree structure:

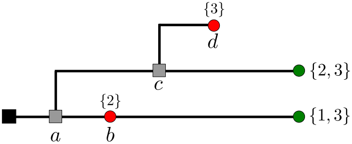


The beta diffusion tree

Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



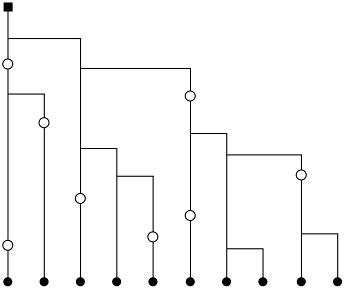
The corresponding tree structure:



The leaves form a feature allocation of [3].

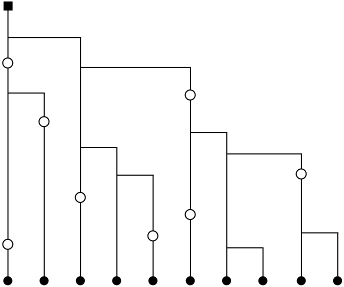
The beta diffusion tree

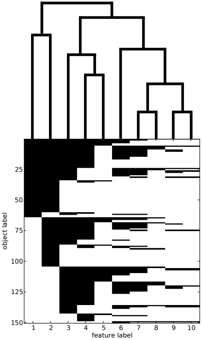
Consider the following larger example:



The beta diffusion tree

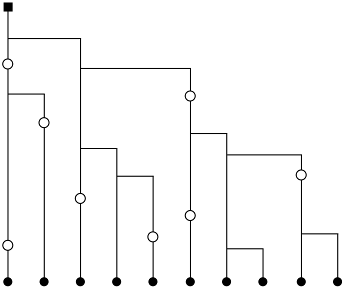
Consider the following larger example:

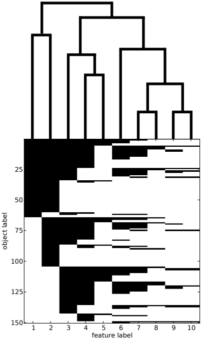




The beta diffusion tree

Consider the following larger example:





A tree structure over the columns of an infinite binary matrix.

Connections to the IBP should be clear.

The beta diffusion tree

Hierarchies of features are a natural assumption:

The beta diffusion tree

Hierarchies of features are a natural assumption:

For image labels in scene analyses:

“image contains a chair”

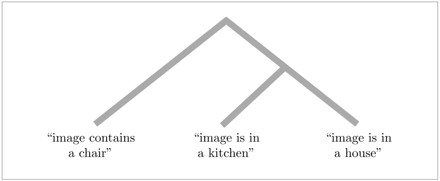
“image is in a kitchen”

“image is in a house”

The beta diffusion tree

Hierarchies of features are a natural assumption:

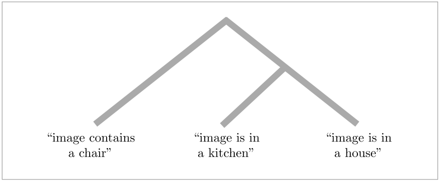
For image labels in scene analyses:



The beta diffusion tree

Hierarchies of features are a natural assumption:

For image labels in scene analyses:



"image contains a chair"

"image is in a kitchen"

"image is in a house"

For admixture in genotype analyses:

"European"

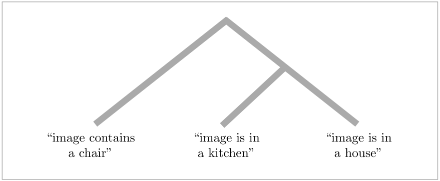
"African"

"North-African"

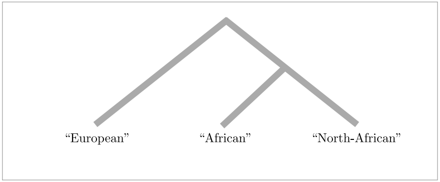
The beta diffusion tree

Hierarchies of features are a natural assumption:

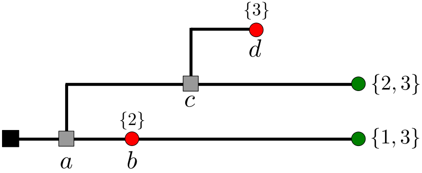
For image labels in scene analyses:



For admixture in genotype analyses:

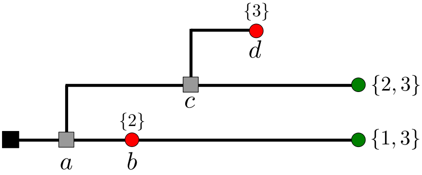


The beta diffusion tree



We defined the tree in terms of the ordered sequence of objects, however:

The beta diffusion tree



We defined the tree in terms of the ordered sequence of objects, however:

Theorem (exchangeability)

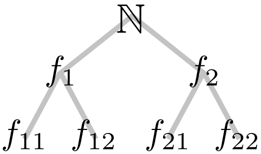
The beta diffusion tree is exchangeable with respect to the ordering of the objects.

The beta diffusion tree

This exchangeability result is non-trivial, however, it is made more obvious from the alternative perspective of a nested feature allocation scheme.

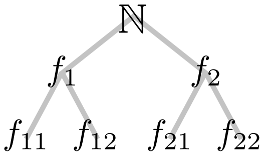
Nested feature allocation scheme

Consider the nested feature allocation scheme with L levels:



Nested feature allocation scheme

Consider the nested feature allocation scheme with L levels:

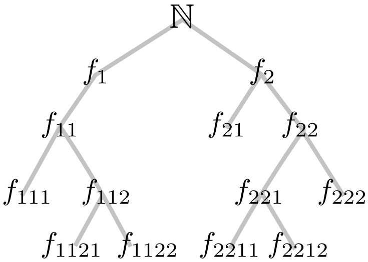


where the objects in each feature at level $\ell - 1$ are allocated to two features at level ℓ with (independent) probabilities:

$$\begin{aligned} p_1^{(\ell)} &\sim \text{beta}\Big(\theta_s\Big(1-\frac{\lambda_s}{L}\Big),\, \theta_s\frac{\lambda_s}{L}\Big), \\ p_2^{(\ell)} &\sim \text{beta}\Big(\theta_r\frac{\lambda_r}{L},\, \theta_r\Big(1-\frac{\lambda_r}{L}\Big)\Big). \end{aligned} \tag{1}$$

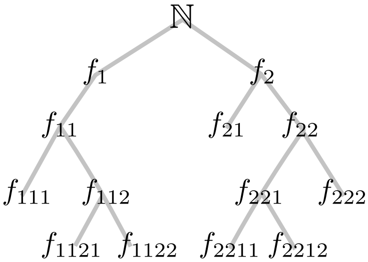
Nested feature allocation scheme

Let the non-empty features be nodes in a tree structure:



Nested feature allocation scheme

Let the non-empty features be nodes in a tree structure:



Theorem (continuum limit)

In the limit $L \to \infty$, we obtain the tree structure of the beta diffusion tree.

Nested feature allocation scheme

Some high-level intuition:

$$\begin{aligned} p_1^{(\ell)} &\sim \text{beta}\Big(\theta_s \Big(1 - \frac{\lambda_s}{L}\Big), \theta_s \frac{\lambda_s}{L} \Big), \\ p_2^{(\ell)} &\sim \text{beta}\Big(\theta_r \frac{\lambda_r}{L}, \theta_r \Big(1 - \frac{\lambda_r}{L}\Big) \Big). \end{aligned} \tag{2}$$

Nested feature allocation scheme

Some high-level intuition:

$$\begin{aligned} p_1^{(\ell)} &\sim \text{beta}\Big(\theta_s \Big(1 - \frac{\lambda_s}{L}\Big), \theta_s \frac{\lambda_s}{L} \Big), \\ p_2^{(\ell)} &\sim \text{beta}\Big(\theta_r \frac{\lambda_r}{L}, \theta_r \Big(1 - \frac{\lambda_r}{L}\Big) \Big). \end{aligned} \tag{2}$$

So as $L \to \infty$, we have $p_1^{(\ell)} \to 1$ and $p_2^{(\ell)} \to 0$. Then:

Nested feature allocation scheme

Some high-level intuition:

$$\begin{aligned} p_1^{(\ell)} &\sim \text{beta}\Big(\theta_s\Big(1-\frac{\lambda_s}{L}\Big), \theta_s\frac{\lambda_s}{L}\Big), \\ p_2^{(\ell)} &\sim \text{beta}\Big(\theta_r\frac{\lambda_r}{L}, \theta_r\Big(1-\frac{\lambda_r}{L}\Big)\Big). \end{aligned} \tag{2}$$

So as $L \to \infty$, we have $p_1^{(\ell)} \to 1$ and $p_2^{(\ell)} \to 0$. Then:

1. NOT allocating to the first feature (i.e., stopping), and

Nested feature allocation scheme

Some high-level intuition:

$$\begin{aligned} p_1^{(\ell)} &\sim \text{beta}\Big(\theta_s\Big(1-\frac{\lambda_s}{L}\Big), \theta_s\frac{\lambda_s}{L}\Big), \\ p_2^{(\ell)} &\sim \text{beta}\Big(\theta_r\frac{\lambda_r}{L}, \theta_r\Big(1-\frac{\lambda_r}{L}\Big)\Big). \end{aligned} \tag{2}$$

So as $L \to \infty$, we have $p_1^{(\ell)} \to 1$ and $p_2^{(\ell)} \to 0$. Then:

1. NOT allocating to the first feature (i.e., stopping), and
2. allocating to the second feature (i.e., replicating)

Nested feature allocation scheme

Some high-level intuition:

$$\begin{aligned} p_1^{(\ell)} &\sim \text{beta}\Big(\theta_s\Big(1-\frac{\lambda_s}{L}\Big), \theta_s\frac{\lambda_s}{L}\Big), \\ p_2^{(\ell)} &\sim \text{beta}\Big(\theta_r\frac{\lambda_r}{L}, \theta_r\Big(1-\frac{\lambda_r}{L}\Big)\Big). \end{aligned} \tag{2}$$

So as $L \to \infty$, we have $p_1^{(\ell)} \to 1$ and $p_2^{(\ell)} \to 0$. Then:

1. NOT allocating to the first feature (i.e., stopping), and
2. allocating to the second feature (i.e., replicating)

become rare events, i.e., the times until these events are exponentially distributed.

Nested feature allocation scheme

Some high-level intuition:

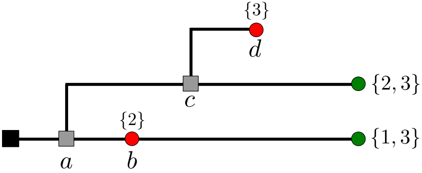
$$\begin{aligned} p_1^{(\ell)} &\sim \operatorname{beta}\Big(\theta_s \Big(1 - \frac{\lambda_s}{L}\Big), \, \theta_s \frac{\lambda_s}{L} \Big), \\ p_2^{(\ell)} &\sim \operatorname{beta}\Big(\theta_r \frac{\lambda_r}{L}, \, \theta_r \Big(1 - \frac{\lambda_r}{L}\Big) \Big). \end{aligned} \tag{3}$$

Then the collection

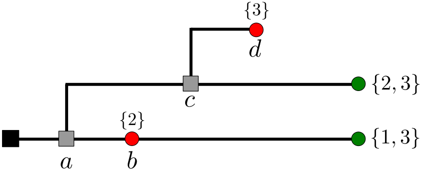
$$\mathcal{F} := \{(p_1^{(1)}, p_2^{(1)}), (p_1^{(2)}, p_2^{(2)}), \dots \}$$

characterizes the de Finetti measure of the beta diffusion tree.

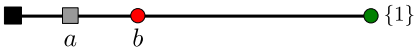
de Finetti measure



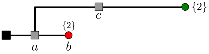
de Finetti measure



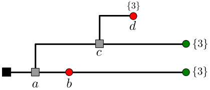
$=$



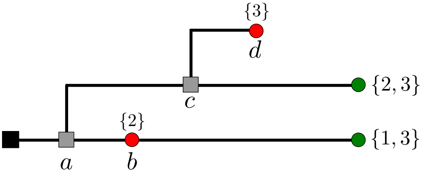
$\bigcup$

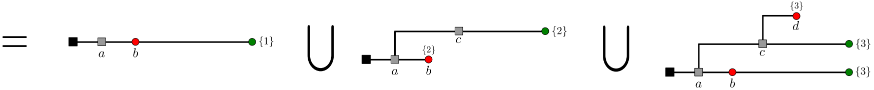


$\bigcup$

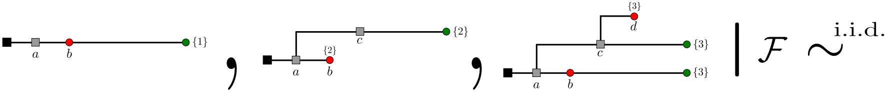


de Finetti measure





and so

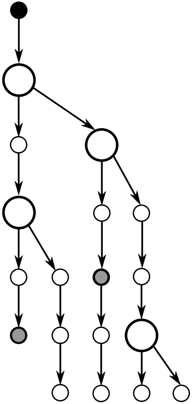


de Finetti measure

This characterization of $\mathcal{F}$ leads us to consider one last perspective on the beta diffusion tree...

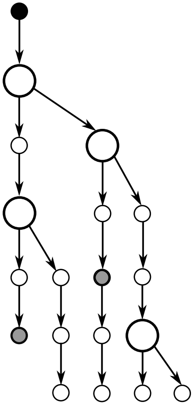
Branching processes

Reconsider the <u>nested feature allocation scheme</u>:



Branching processes

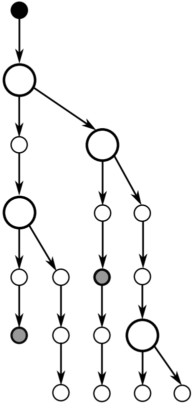
Reconsider the nested feature allocation scheme:



- Nodes are non-empty features
- Large nodes are replicate nodes
- Shaded nodes are stop nodes
- The number of objects allocated to a feature determines its type

Branching processes

Reconsider the nested feature allocation scheme:

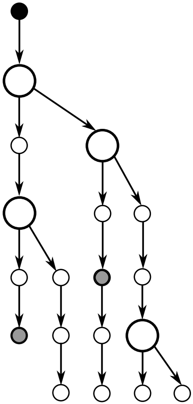


New perspective:

- Chains are individuals in a population
- Individual gives birth at replicate nodes
- At stop nodes:
    - the individual changes type
    - or the individual dies

Branching processes

Reconsider the nested feature allocation scheme:



Branching processes:

- The nested feature allocation scheme is a multitype Galton–Watson process
- The continuum limit (beta diffusion tree) is a multitype continuous-time branching process
- With $\mathcal{F}$, we can find the child distributions.

Branching processes

For example, consider a feature at level $\ell - 1$ with m objects.

Branching processes

For example, consider a feature at level $\ell - 1$ with m objects.

Assigning k objects to the first feature at level ℓ is distributed as

$$\text{binomial}(k; m, p_1^{(\ell)}) = k^{p_1^{(\ell)}} (m-k)^{1-p_1^{(\ell)}} \tag{4}$$

Branching processes

For example, consider a feature at level $\ell - 1$ with m objects.

Assigning k objects to the first feature at level ℓ is distributed as

$$\mathrm{binomial}(k; m, p_1^{(\ell)}) = k^{p_1^{(\ell)}}(m-k)^{1-p_1^{(\ell)}} \tag{4}$$

Similarly for the second feature.

Branching processes

For example, consider a feature at level $\ell - 1$ with m objects.

Assigning k objects to the first feature at level ℓ is distributed as

$$\text{binomial}(k; m, p_1^{(\ell)}) = k^{p_1^{(\ell)}} (m - k)^{1-p_1^{(\ell)}} \tag{4}$$

Similarly for the second feature.

And we can integrate out $p_1^{(\ell)}$ and $p_2^{(\ell)}$.

Etc. etc. See the journal version of the paper.

Branching processes

One immediately useful result that follows:

Theorem

If $N, \lambda_s, \lambda_r, \theta_s, \theta_r < \infty$*. Then the number of leaves in the beta diffusion tree is almost surely finite.*

Branching processes

One immediately useful result that follows:

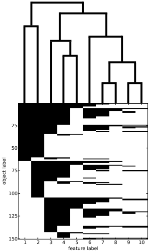
Theorem

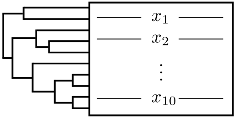
If $N, \lambda_s, \lambda_r, \theta_s, \theta_r < \infty$*. Then the number of leaves in the beta diffusion tree is almost surely finite.*

We also characterize the expected number of leaves. Probably more.

Example: A hierarchical factor analysis model

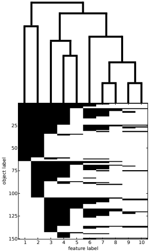
Example: A hierarchical factor analysis model

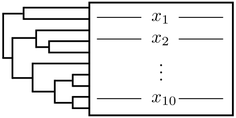




$$\underbrace{\underset{N \times D}{Y}}_{\text{Observations}} = \underbrace{\underset{N \times K}{Z}}_{\substack{\text{factors}\\ = \text{feature allocation}}} \bullet \underbrace{\underset{K \times D}{X}}_{\substack{\text{factor loadings}\\ = \text{leaves in } \mathcal{X}}} + \underbrace{\underset{N \times D}{\varepsilon}}_{\substack{\text{i.i.d. Gaussian}\\ \text{noise}}}$$

Example: A hierarchical factor analysis model





$$\underbrace{\underset{N \times D}{Y}}_{\text{Observations}} = \underbrace{\underset{N \times K}{Z}}_{\substack{\text{factors} \\ = \text{ feature allocation}}} \bullet \underbrace{\underset{K \times D}{X}}_{\substack{\text{factor loadings} \\ = \text{ leaves in } \mathcal{X}}} + \underbrace{\underset{N \times D}{\varepsilon}}_{\substack{\text{i.i.d. Gaussian} \\ \text{noise}}}$$

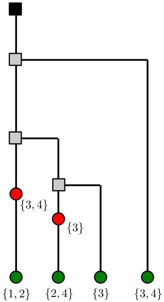
Both the factors and factor loadings are hierarchical structured!

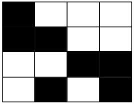
Inference

We perform a series of Metropolis–Hastings proposals to integrate over the random tree structures.

Inference

Consider resampling an internal subtree:

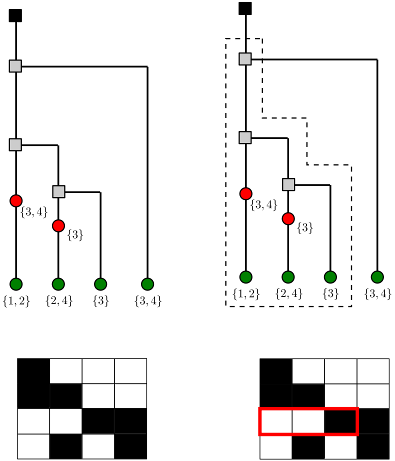




■	□	□	□
■	■	□	□
□	□	■	■
□	■	□	■

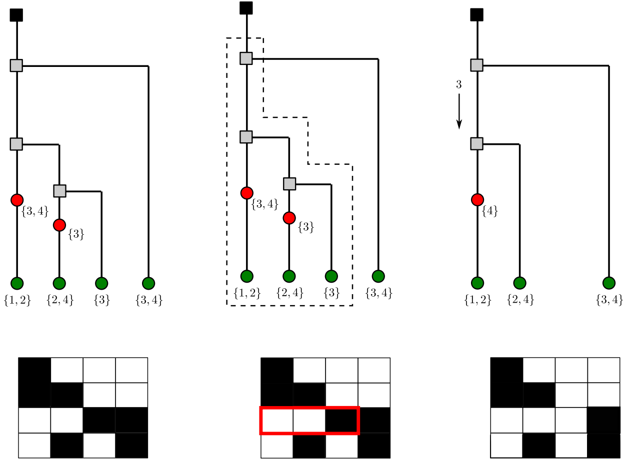
Inference

Consider resampling an internal subtree:



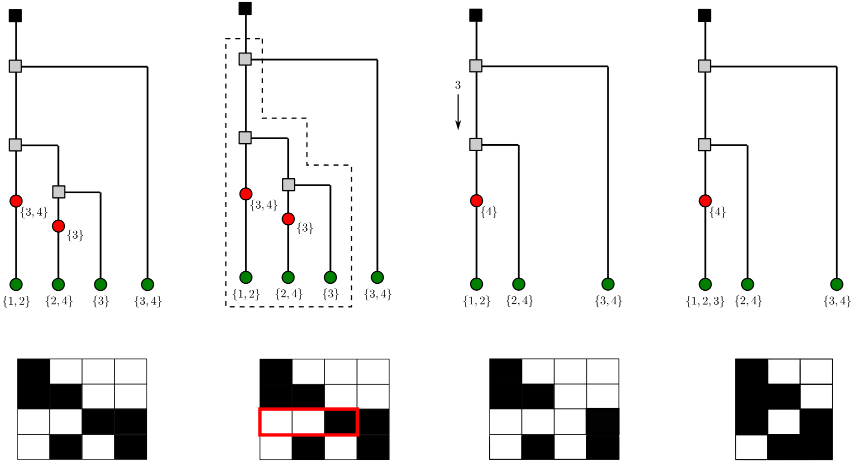
Inference

Consider resampling an internal subtree:



Inference

Consider resampling an internal subtree:



More inference moves

1. Additionally resample more than one particle at a time.

More inference moves

1. Additionally resample more than one particle at a time.
2. Propose removing an internal (either replicate or stop) node. Conversely, propose a node to a branch.

More inference moves

1. Additionally resample more than one particle at a time.
2. Propose removing an internal (either replicate or stop) node. Conversely, propose a node to a branch.
3. Propose changing the decisions that particles take at internal nodes (i.e., the decisions to either replicate or stop).

More inference moves

1. Additionally resample more than one particle at a time.
2. Propose removing an internal (either replicate or stop) node. Conversely, propose a node to a branch.
3. Propose changing the decisions that particles take at internal nodes (i.e., the decisions to either replicate or stop).
4. Propose removing "unpopular" replicate or stop nodes (this is a heuristic).

More inference moves

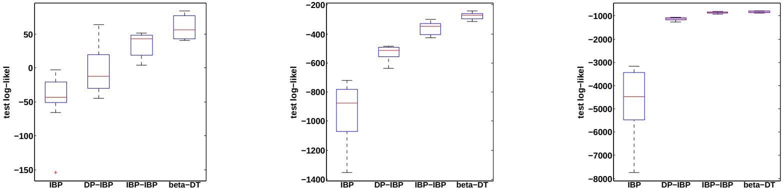
1. Additionally resample more than one particle at a time.
2. Propose removing an internal (either replicate or stop) node. Conversely, propose a node to a branch.
3. Propose changing the decisions that particles take at internal nodes (i.e., the decisions to either replicate or stop).
4. Propose removing "unpopular" replicate or stop nodes (this is a heuristic).

Each is accepted or rejected with a Metropolis–Hastings step.

Moves verified with joint distribution tests (Geweke, 2004).

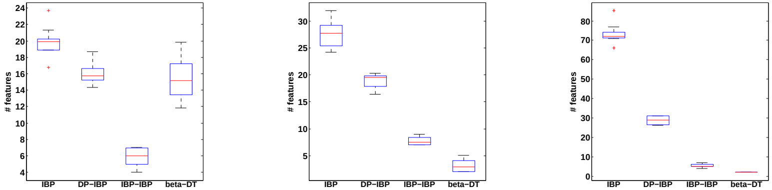
Some numerical results

Test log-likelihoods:



(a) *E. Coli* (b) UN (c) India

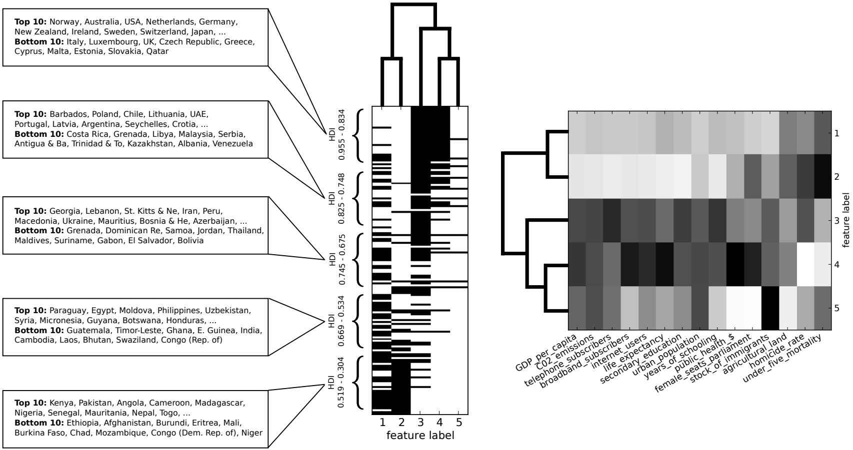
Average number of inferred features:



(d) *E. Coli* (e) UN (f) India

An analysis of UN human development statistics

We can extend the analysis by Doshi–Velez & Ghahramani (2009):



Top 10: Norway, Australia, USA, Netherlands, Germany, New Zealand, Ireland, Sweden, Switzerland, Japan, ...
Bottom 10: Italy, Luxembourg, UK, Czech Republic, Greece, Cyprus, Malta, Estonia, Slovakia, Qatar

Top 10: Barbados, Poland, Chile, Lithuania, UAE, Portugal, Latvia, Argentina, Seychelles, Crotia, ...
Bottom 10: Costa Rica, Grenada, Libya, Malaysia, Serbia, Antigua & Ba, Trinidad & To, Kazakhstan, Albania, Venezuela

Top 10: Georgia, Lebanon, St. Kitts & Ne, Iran, Peru, Macedonia, Ukraine, Mauritius, Bosnia & He, Azerbaijan, ...
Bottom 10: Grenada, Dominican Re, Samoa, Jordan, Thailand, Maldives, Suriname, Gabon, El Salvador, Bolivia

Top 10: Paraguay, Egypt, Moldova, Philippines, Uzbekistan, Syria, Micronesia, Guyana, Botswana, Honduras, ...
Bottom 10: Guatemala, Timor-Leste, Ghana, E. Guinea, India, Cambodia, Laos, Bhutan, Swaziland, Congo (Rep. of)

Top 10: Kenya, Pakistan, Angola, Cameroon, Madagascar, Nigeria, Senegal, Mauritania, Nepal, Togo, ...
Bottom 10: Ethiopia, Afghanistan, Burundi, Eritrea, Mali, Burkina Faso, Chad, Mozambique, Congo (Dem. Rep. of), Niger