Beta diffusion trees

Creighton Heaukulani
University of Cambridge

Joint work with

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

ICML 2014, Beijing
June 2014

Outline

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

Outline

1. Briefly review feature allocations.

Outline

1. Briefly review feature allocations.
2. Describe a stochastic process (the beta diffusion tree) that leads to trees over feature allocations.

Outline

1. Briefly review feature allocations.

2. Describe a stochastic process (the beta diffusion tree) that leads to trees over feature allocations.

3. Describe a hierarchically-clustered factor analysis model.

Outline

1. Briefly review feature allocations.
2. Describe a stochastic process (the beta diffusion tree) that leads to trees over feature allocations.
3. Describe a hierarchically-clustered factor analysis model.
4. Present some inference techniques over the tree structure and some experimental results.

Feature allocations

Feature allocations

Feature allocations = Overlapping subsets = Overlapping clusters

Feature allocations

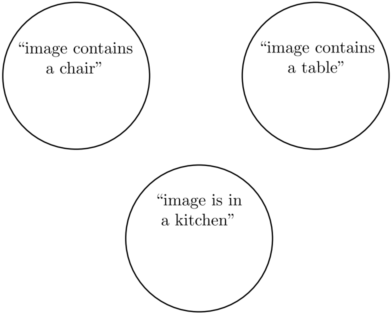
Feature allocations = Overlapping subsets = Overlapping clusters

Consider an image labelling application:

Feature allocations

Feature allocations = Overlapping subsets = Overlapping clusters

Consider an image labelling application:



Feature allocations

Feature allocations = Overlapping subsets = Overlapping clusters

Consider an image labelling application:

"image contains a chair"
$\{1\}$

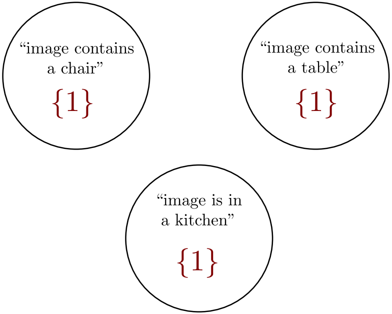
"image contains a table"
$\{1\}$

"image is in a kitchen"
$\{1\}$

Feature allocations

Feature allocations = Overlapping subsets = Overlapping clusters

Consider an image labelling application:



Feature allocations

Feature allocations = Overlapping subsets = Overlapping clusters

Consider an image labelling application:

"image contains a chair" $\{1, 2\}$

"image contains a table" $\{1, 2\}$

"image is in a kitchen" $\{1\}$

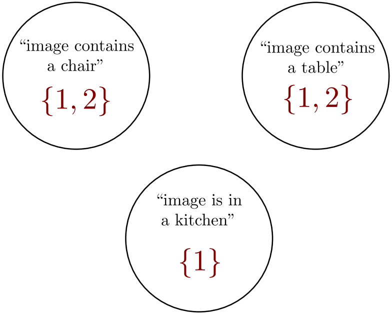
Feature allocations

Feature allocations = Overlapping subsets = Overlapping clusters

Consider an image labelling application:

Feature allocation:

$\{\{1,2\},\{1,2\},\{1\}\}$



Feature allocations

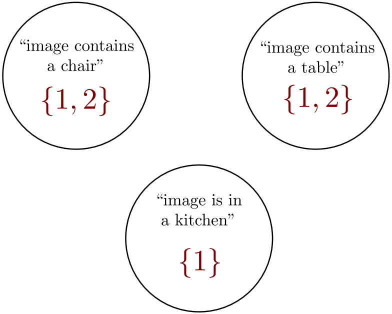
Feature allocations = Overlapping subsets = Overlapping clusters

Consider an image labelling application:

Feature allocation:

$$\{\{1,2\},\{1,2\},\{1\}\}$$

The clusters/subsets are called <u>features</u>



Hierarchies on features

Hierarchies on features

Hierarchies of features are a natural assumption:

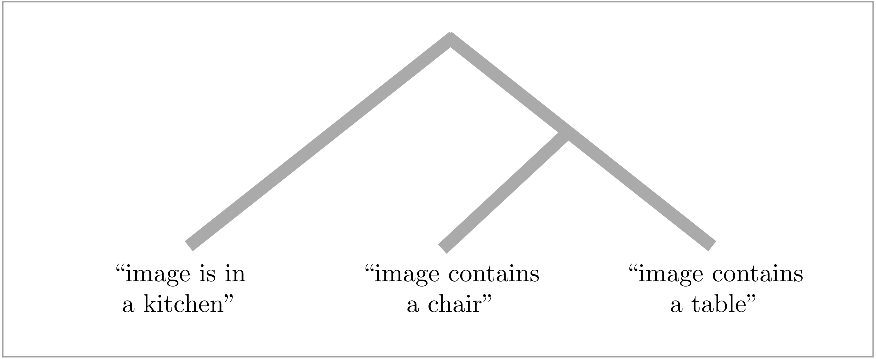
“image is in a kitchen”

“image contains a chair”

“image contains a table”

Hierarchies on features

Hierarchies of features are a natural assumption:



The beta diffusion tree

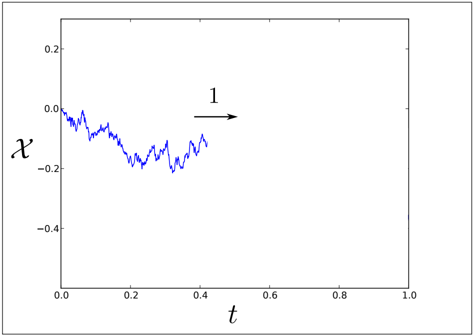
The beta diffusion tree

Consider a sequence of particles, labeled with one of N objects, diffusing in a continuous space $\mathcal{X}$ over the unit time interval $[0, 1]$.

The beta diffusion tree

Consider a sequence of particles, labeled with one of N objects, diffusing in a continuous space $\mathcal{X}$ over the unit time interval $[0, 1]$.

For example:



- $\mathcal{X} = \mathbb{R}$.
- Diffusion is Brownian motion.

The beta diffusion tree

Consider a sequence of particles, labeled with one of N objects, diffusing in a continuous space $\mathcal{X}$ over the unit time interval $[0, 1]$.

The beta diffusion tree

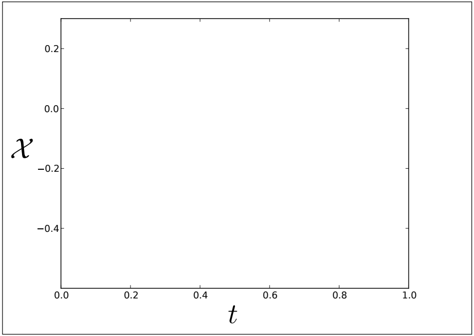
Consider a sequence of particles, labeled with one of N objects, diffusing in a continuous space $\mathcal{X}$ over the unit time interval $[0, 1]$.

At random times:

- **Replicate**: Copies of particles may be created.
- **Stop**: Particles are destroyed.

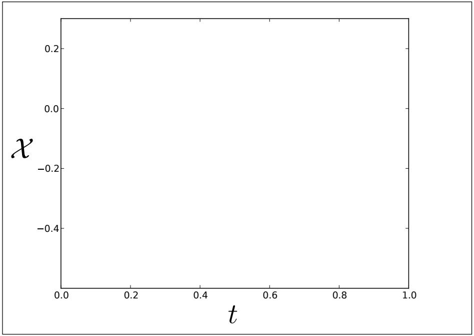
The beta diffusion tree

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



The beta diffusion tree

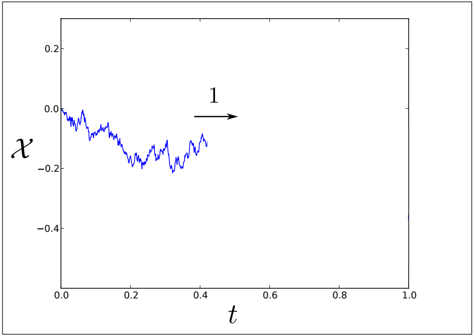
Let $\lambda_s, \lambda_r, \theta_s, \theta_r > 0$ and $\mathcal{X} = \mathbb{R}$. Diffusion is Brownian motion.



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.



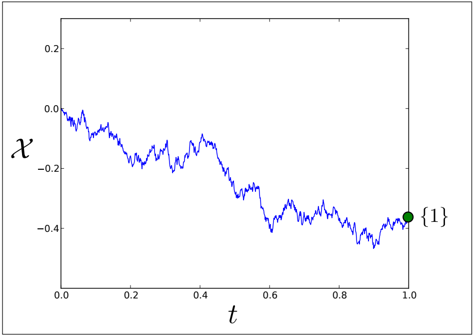
1. The first particle starts at the origin.

2. May replicate or stop after $\exp(\lambda_r)$ or $\exp(\lambda_s)$ time.

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.

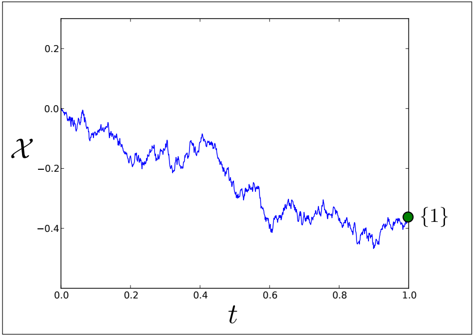


1. The first particle starts at the origin.
2. May replicate or stop after $\exp(\lambda_r)$ or $\exp(\lambda_s)$ time.
3. Diffuses as Brownian motion until $t = 1$.

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.

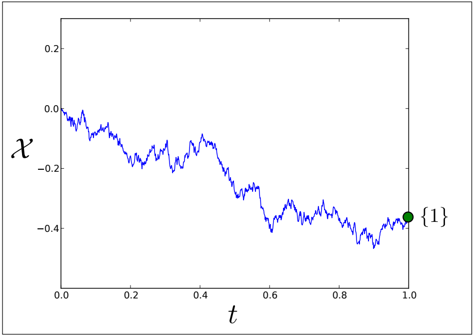


1. Follow previous particles.

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

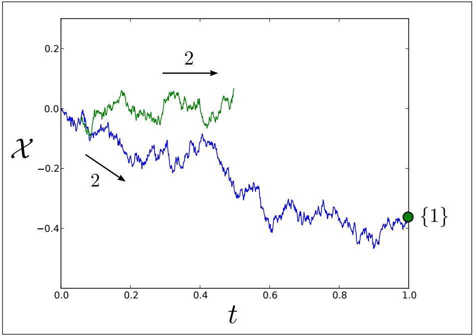
$$\exp\Big(\frac{\theta_s}{\theta_s + m}\lambda_s\Big)$$

amount of time.

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

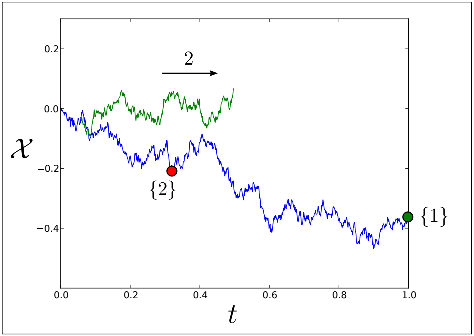
$$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right)$$

amount of time.

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

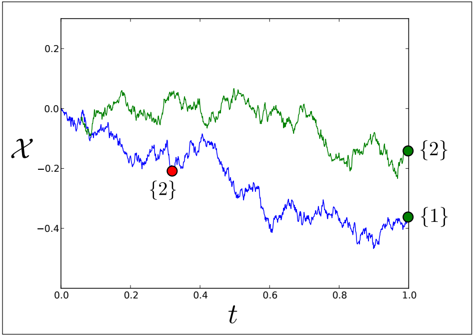
$$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right)$$

amount of time.

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

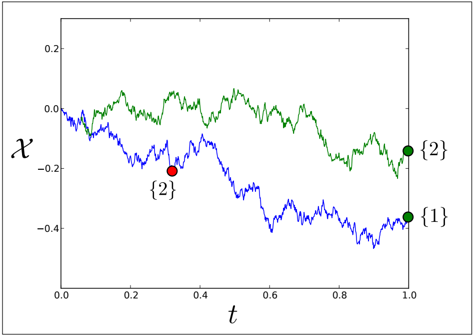
$$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right)$$

amount of time.

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

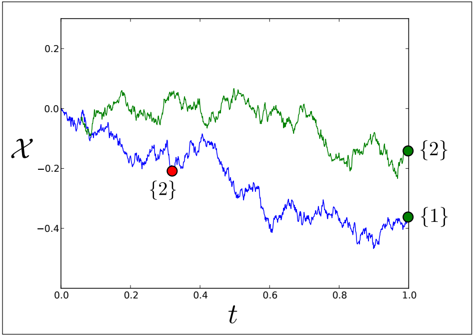
$$\exp\left(\frac{\theta_s}{\theta_s + m}\lambda_s\right)$$

amount of time.

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.

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

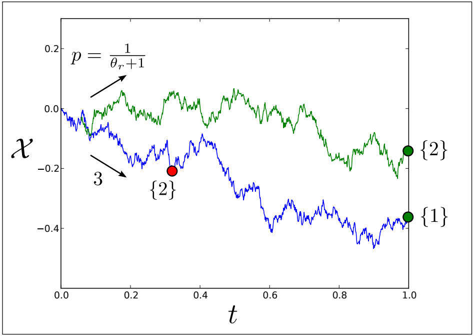
and

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

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.

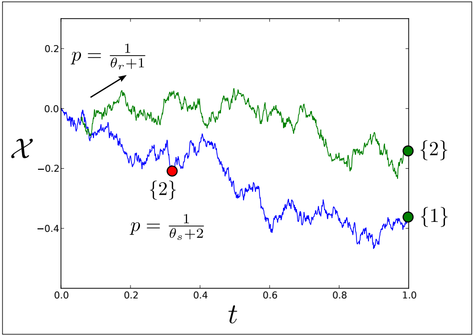


1. Follow previous particles.
2. May replicate or stop after
$$\exp\Big(\frac{\theta_r}{\theta_r + m}\lambda_r\Big)$$
or
$$\exp\Big(\frac{\theta_s}{\theta_s + m}\lambda_s\Big)$$
amount of time.
3. Replicate and stop w.p.
$$\frac{n_r}{\theta_r + m}$$
and
$$\frac{n_s}{\theta_s + m}$$

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.

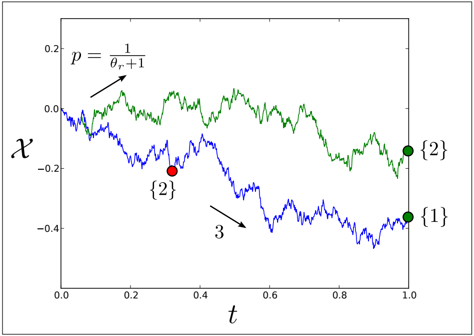
$$\frac{n_r}{\theta_r + m}$$

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.

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

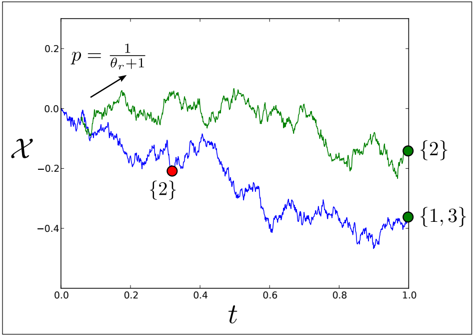
and

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

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.



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.

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

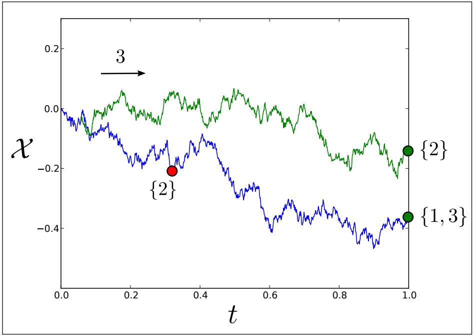
and

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

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.

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

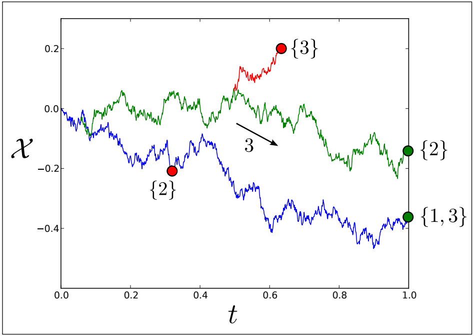
and

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

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.



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.

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

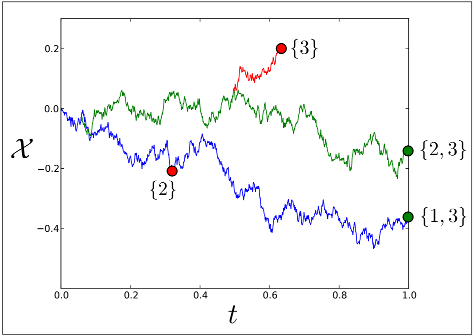
and

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

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.

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

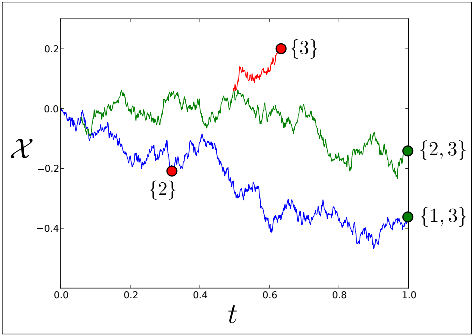
and

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

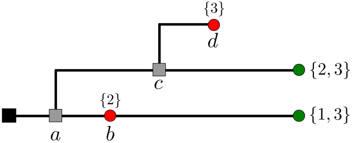
Consider the third object.

The beta diffusion tree

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

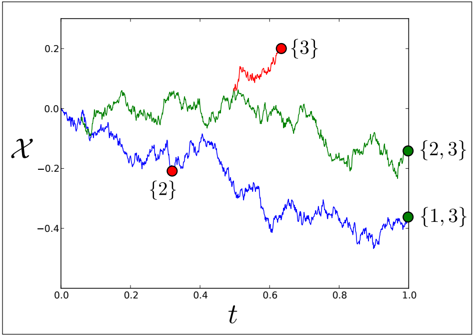


The corresponding tree structure:

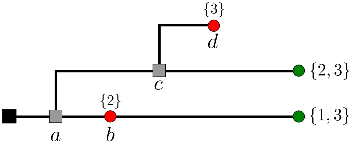


The beta diffusion tree

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]:

$$\{\{1,3\},\{2,3\}\}$$

Related work

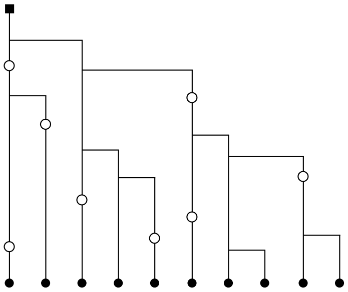
As opposed to...

Random tree structures over non-overlapping clusters (aka partitions):

- Dirichlet diffusion tree (Neal, 2003 [6])
- Pitman–Yor diffusion tree (Knowles & Ghahramani, 2011 [5])
- Kingman's coalescent (Kingman, 1982 [4])

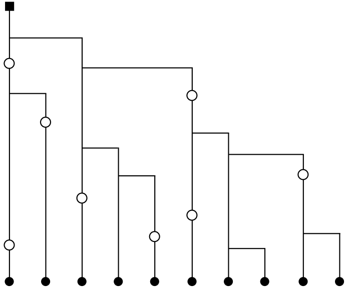
The beta diffusion tree

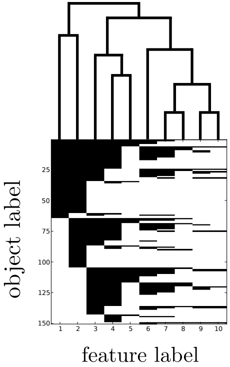
Consider the following larger example:



The beta diffusion tree

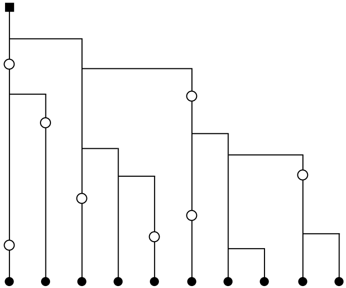
Consider the following larger example:

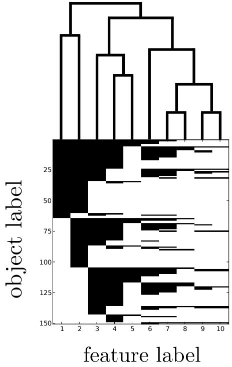




The beta diffusion tree

Consider the following larger example:

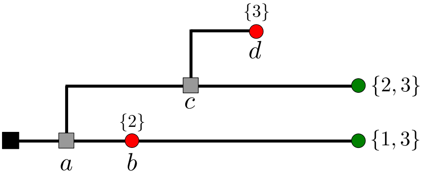




A tree structure over the columns of a binary matrix.

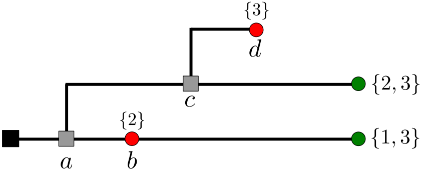
Connections to the Indian buffet process
(Ghahramani, Griffiths, & Sollich, 2007 [3]).

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)

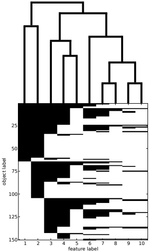
The beta diffusion tree is exchangeable with respect to the ordering of the objects.

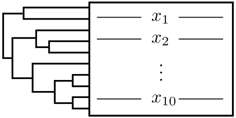
Other properties

- The beta diffusion tree is the continuum limit of a nested feature allocation scheme
- The beta diffusion tree is a continuous-time branching process

Example: A hierarchical factor analysis model

Example: A hierarchical factor analysis model

Y = 

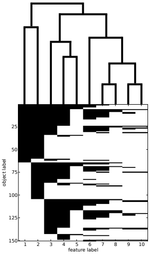
$\bullet$ 

$+ \ \varepsilon$

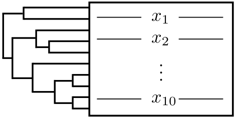
$$\underbrace{N \times D}_{\text{Observations}} \qquad \underbrace{\underset{N \times K}{Z}}_{\substack{\text{factors} \\ = \text{feature allocation}}} \qquad \underbrace{\underset{K \times D}{X}}_{\substack{\text{factor loadings} \\ = \text{leaves in } \mathcal{X}}} \qquad \underbrace{N \times D}_{\text{i.i.d.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

Example: A hierarchical factor analysis model

Y $=$



$\bullet$



$+$ ε

$\underbrace{N \times D}_{}$

Observations

Z

$\underbrace{N \times K}_{}$

factors
= feature allocation

X

$\underbrace{K \times D}_{}$

factor loadings
= leaves in $\mathcal{X}$

$\underbrace{N \times D}_{}$

i.i.d. Gaussian noise

Both the factors and factor loadings are hierarchically structured

Inference

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

Inference

1. Resample the paths of particles down the tree.

Inference

1. Resample the paths of particles down the tree.

2. Propose removing an internal (either replicate or stop) node. Conversely, propose a node to a branch.

Inference

1. Resample the paths of particles down the tree.

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).

Inference

1. Resample the paths of particles down the tree.

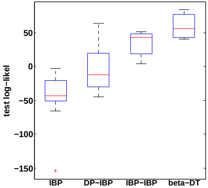
2. Propose removing an internal (either replicate or stop) node. Conversely, propose a node to a branch.

3. Propose changing the decisions that particles take at internal nodes (i.e., the decisions to either replicate or stop).

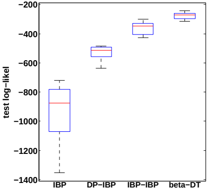
Each is accepted or rejected with a Metropolis–Hastings step.

Some numerical results

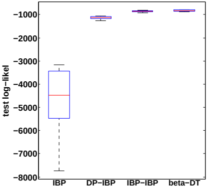
Test log-likelihoods:



(a) *E. Coli*



(b) UN



(c) India

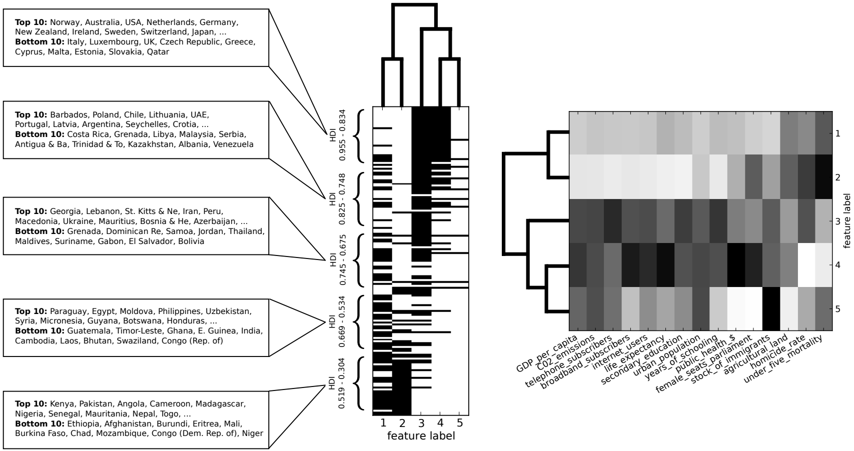
(a) A gene expression dataset

(b) International development statistics

(c) Household socioeconomic indicators

An analysis of UN human development statistics

We can extend the analysis by Doshi–Velez & Ghahramani (2009) [1]:



[1] F. Doshi-Velez and Z. Ghahramani. Correlated non-parametric latent feature models. In Proc. UAI, 2009.

[2] J. Geweke. Getting it right: Joint distribution tests of posterior simulators. Journal of the American Statistical Association, 99(467):799–804, 2004.

[3] Z. Ghahramani, T. L. Griffiths, and P. Sollich. Bayesian nonparametric latent feature models. Bayesian Statistics, 8:201–226, 2007. See also the discussion and rejoinder.

[4] J. F. C. Kingman. The coalescent. Stochastic Processes and Their Applications, 13(3):235–248, 1982.

[5] D. Knowles and Z. Ghahramani. Pitman-Yor diffusion trees. In Proc. UAI, 2011.

[6] R. M. Neal. Density modeling and clustering using Dirichlet diffusion trees. Bayesian Statistics, 7, 2003.