Random Coagulations and Fragmentations
(Chinese restaurants through time)

The plan

- Why?
- Two generative tree-structured stories:
  - Top-down (fragmentation)
  - Bottom-up (coagulation)
- An equivalence of these stories.
- Processes that both fragment and coagulate.

Applications of hierarchies
[Neal, 2001]

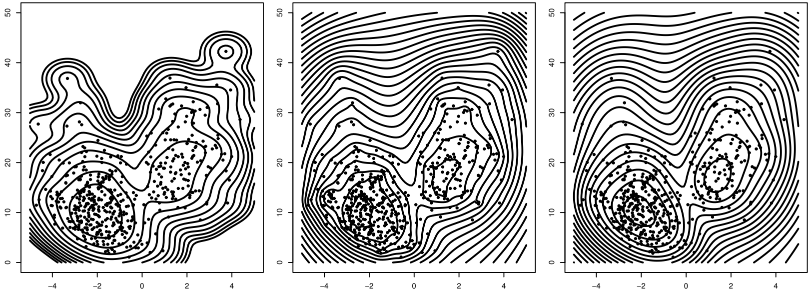
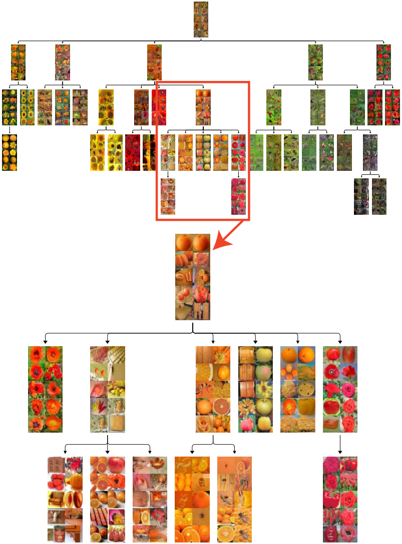
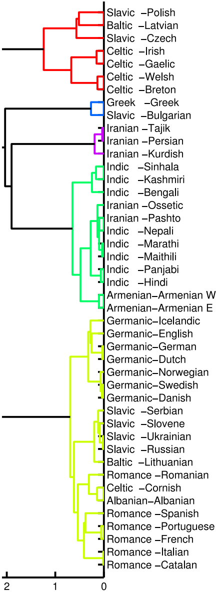


Figure 6. *The two-dimensional density modeling example. The plots show the data points along with contours of the natural log of the estimated density, spaced 0.5 apart. The left plot shows a simple kernel density estimate; the middle and right plots show Dirichlet diffusion tree estimates with different divergence functions.*

Applications of hierarchies

[Adams, Ghahramani, Jordan, 2010; Görür, Teh, 2008]





Applications of hierarchies

[Görür, Boyles, Welling, 2012]

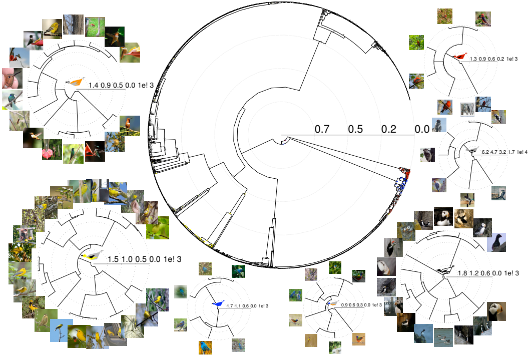
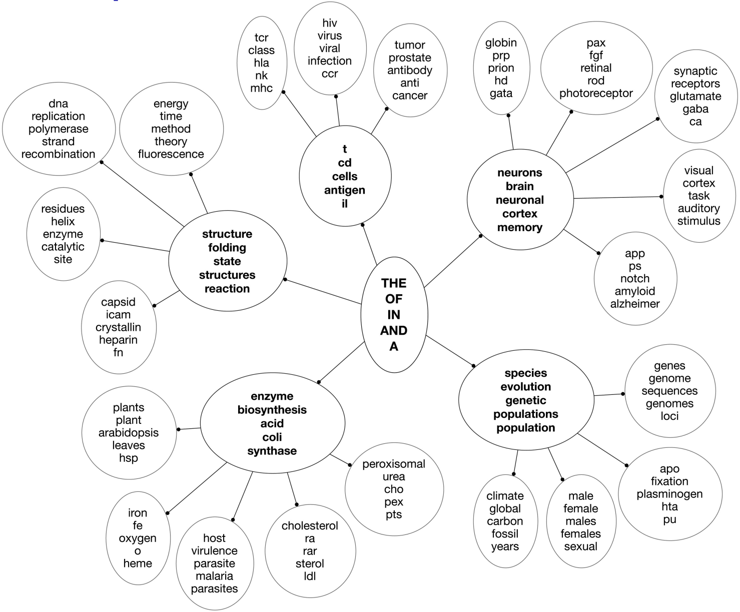


Figure 1: A sample tree from our SMC algorithm run on the Birds200 dataset (center), and several subtrees. Bird caricatures are displayed on the nodes of the tree that summarize the color properties of the descendants of that node. (Zoom in for better visibility.) Five out of 15 features are used in the caricatures for interpretability. Light gray indicates "uncertain," and all other colors denote the color of the bird's body, belly, tail, wing, crown, and eyes. The subtrees show the images corresponding to the data at the leaves along with the corresponding common ancestor caricature in the middle.

Application of hierarchies

[Blei et al, 2004]



Partition-valued Processes

- Markov processes whose states are partitions of the natural numbers.
- In the continuous case, these will be Markov jump processes.
- Sequences of partitions provides structure for a likelihood model.
- Most of the processes we examine will be constructed using the Chinese restaurant process.

Partitions: Chinese restaraunt process

Let π be a partition of $\mathbb{N}$, generated by a CRP.

- i.e., π is a set of disjoint sets of $\mathbb{N}$ such that the union of these sets is $\mathbb{N}$.

π is generated iteratively, by considering each natural number in turn.

- At step 1, there is 1 cluster.
- At step $n > 1$, there are K clusters.
  - $n + 1$ joins cluster i with probability

$$\frac{n_i - \alpha}{n + \theta}$$

  - $n + 1$ joins a new cluster with probability

$$\frac{\theta + K\alpha}{n + \theta}$$

Partitions: Chinese restaraunt process

- Let π_n be the restriction of π to a partition of $\{1, 2, \ldots, n\}$.
- The density of π_n, with K clusters, is

$$p(\pi_n) = \frac{(\theta+\alpha)_{K-1\uparrow\alpha}}{(\theta+1)_{n-1\uparrow 1}} \prod_{i=1}^{K} (1-\alpha)_{n_i-1\uparrow 1}$$

where

$$(x)_{n\uparrow\alpha} = \prod_{i=0}^{n-1} (x + i\alpha)$$

Fragmenting one partition with several others

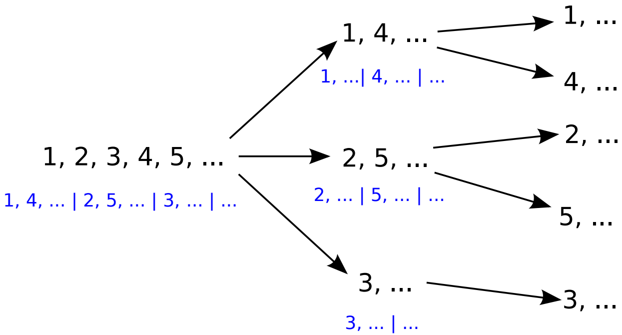
Top-down hierarchies: fragmentation

A Markov kernel on partitions that can *increase* the number of populated clusters in a partition.

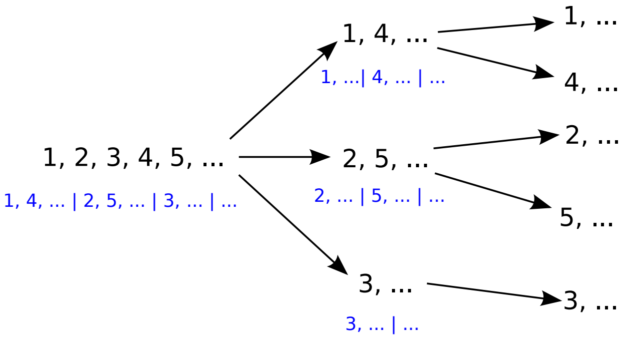
- Let π be a partition of $\mathbb{N}$ with clusters $\{A_i \subseteq \mathbb{N}\}_{i=1}^{\infty}$.
- Let $\{\rho^{(i)}\}_{i=1}^{\infty}$ be random partitions i.i.d. from $\mathrm{CRP}(\alpha, \theta)$
- Then π' is the fragmentation of π if it is a refinement of π where
  - Replace each A_i with the clusters of $\rho_{A_i}^{(i)}$.
  - i.e., each cluster i of π is split according to its corresponding random partition $\rho^{(i)}$.
- Notation: $\pi' | \pi \sim \mathrm{Frag}_{\alpha,\theta}(\pi)$.

Top-down hierarchies: nested Chinese restaurant process
[Blei, Griffiths, Jordan, 2003]

- Let $\pi(0)$ be the partition with just one cluster.
- $\pi(t+1)|\pi(t) \sim \mathrm{Frag}_{0,\theta}(\pi(t))$.



Top-down hierarchies: nested Chinese restaurant process



$$p(\pi_n(t+1)|\pi_n(t)) = \prod_i p(\rho^{(i)}_{A_i})$$

$$p(\rho^{(i)}_{A_i}) = \frac{(\theta)^{K_i-1}}{(\theta+1)_{n_i-1\uparrow 1}} \prod_{j=1}^{K_i} \Gamma(n_{ij})$$

Top-down hierarchies: Dirichlet diffusion trees

[Neal, 2001]

- Continuous-indexed analogue of nCRP:
  - $\pi(0)$ is the partition with just one cluster.
  - $\pi(t + dt) | \pi(t) \sim \mathrm{Frag}_{0, a(t)dt}(\pi(t))$

Top-down hierarchies: Dirichlet diffusion trees
[Neal, 2001]

- Continuous-indexed analogue of nCRP:
  - $\pi(0)$ is the partition with just one cluster.
  - $\pi(t+dt)|\pi(t) \sim \mathrm{Frag}_{0,a(t)dt}(\pi(t))$
- For simplicity, let $a(t) = \theta$. Consider the restriction of the process to n items as $dt \to 0$. Density:

$$p(\pi_n(t+dt)|\pi_n(t)) = \prod_i p(\rho^{(i)}_{n_i})$$

$$p(\rho^{(i)}_{n_i}) = \frac{(\theta dt)^{K_i - 1}}{(\theta dt + 1)_{n_i - 1 \uparrow 1}} \prod_{j=1}^{K_i} \Gamma(n_{ij})$$

Top-down hierarchies: Dirichlet diffusion trees

- Consider a single cluster i:

$$
\begin{aligned}
\lim_{dt \to 0} p(\rho^{(i)}_{n_i}) &= \lim_{dt \to 0} \frac{(\theta dt)^{K_i - 1}}{(\theta dt + 1)_{n_i - 1 \uparrow 1}} \prod_{j=1}^{K_i} \Gamma(n_{ij}) \\
&= \frac{\theta^{K_i - 1}}{\Gamma(n_i)} \left[\prod_{j=1}^{K_i} \Gamma(n_{ij}) \right] \lim_{dt \to 0} dt^{K_i - 1} \\
&= \begin{cases} O(1) & \text{if } K_i = 1, \\ O(dt) & \text{if } K_i = 2, \\ O(dt^2) & \text{if } K_i > 2. \end{cases}
\end{aligned}
$$

Top-down hierarchies: Dirichlet diffusion trees

- Consider a single cluster i:

$$\lim_{dt\to 0} p(\rho^{(i)}_{n_i}) = \lim_{dt\to 0} \frac{(\theta dt)^{K_i - 1}}{(\theta dt + 1)_{n_i - 1\uparrow 1}} \prod_{j=1}^{K_i} \Gamma(n_{ij})$$
$$= \frac{\theta^{K_i-1}}{\Gamma(n_i)} \left[\prod_{j=1}^{K_i} \Gamma(n_{ij})\right] \lim_{dt\to 0} dt^{K_i - 1}$$
$$= \begin{cases} O(1) & \text{if } K_i = 1, \\ O(dt) & \text{if } K_i = 2, \\ O(dt^2) & \text{if } K_i > 2. \end{cases}$$

- Hence the rate at which a cluster of size n in $\pi(t)$ fragments into clusters of sizes n_1 and n_2 in $\pi(t + dt)$ is:

$$\theta \frac{\Gamma(n_1)\Gamma(n_2)}{\Gamma(n)}$$

Top-down hierarchies: Dirichlet diffusion trees

- Consider all clusters:

$$\lim_{dt \to 0} p(\pi(t+dt)|\pi(t)) = \prod_i \frac{\theta^{K_i-1}}{\Gamma(n_i)} \left[\prod_{j=1}^{K_i} \Gamma(n_{ij}) \right] \lim_{dt \to 0} dt^{K_i-1}$$

$$= \begin{cases} 1 & \text{if all } K_i = 1, \\ O(dt) & \text{iff exactly one } K_i = 2, \\ O(dt^2) & \text{if more than on } K_i > 1. \end{cases}$$

Top-down hierarchies: Dirichlet diffusion trees

- Consider all clusters:

$$\lim_{dt\to 0} p(\pi(t+dt)|\pi(t)) = \prod_i \frac{\theta^{K_i-1}}{\Gamma(n_i)} \left[\prod_{j=1}^{K_i} \Gamma(n_{ij})\right] \lim_{dt\to 0} dt^{K_i-1}$$

$$= \begin{cases} 1 & \text{if all } K_i = 1, \\ O(dt) & \text{iff exactly one } K_i = 2, \\ O(dt^2) & \text{if more than on } K_i > 1. \end{cases}$$

- Each cluster fragments independently, with total rate:

$$\sum_{n_1,n_2} \theta \frac{\Gamma(n_1)\Gamma(n_2)}{\Gamma(n)} = \theta H_{n-1}$$

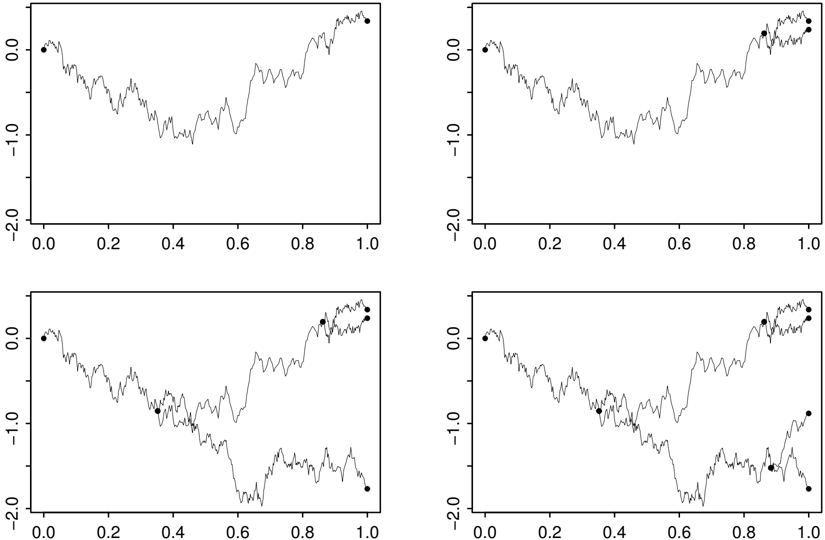
$$\approx \theta \log(n-1) + O(1)$$

- A curious aside: it's also possible to derive DDT under another limit:
$\lim_{dt\to 0} p(\mathrm{Frag}_{\alpha dt,0}(\pi) = \xi) = \lim_{dt\to 0} p(\mathrm{Frag}_{0,\alpha dt}(\pi) = \xi)$

Top-down hierarchies: Dirichlet diffusion trees

- Partition structure describes a tree.
- Diffusion likelihood is a tree-structured Gaussian graphical model consistent with the partitioning structure.
- Diffusion likelihood:
  - Location at root is distributed according to $\mathcal{N}(0, \sigma^2)$
  - Cluster evolves according to $\mathcal{N}(\mu, \sigma^2 dt)$ where μ is initial location of cluster.

Top-down hierarchies: Dirichlet diffusion trees



Gibbs fragmentation trees and Pitman-Yor Diffusion trees

[McCullagh, Pitman, Winkel 2008; Knowles, Ghahramani, 2011]

- A generalisation of fragmentation beyond a one-parameter CRP used for nCRP and DDT.
- [McCullagh et al, 2008] show that if the distribution of $\rho^{(i)}$ is of the form:

$$\frac{1}{Z}\frac{a(k)}{c(K)}\prod_{i=1}^{K} w(n_i)$$

  and the fragmentation is to be projective, then p is necessarily the two-parameter CRP. (Exists similar result by Kerov for random partitions).
- [Knowles, Ghahramani, 2011] examine posterior inference in this model with a diffusion likelihood as in DDT.

Coagulating one partition with another

Bottom-up hierarchies: coagulation

A Markov kernel on partitions that can *decrease* the number of populated clusters in a partition.

- Let π be a partition of $\mathbb{N}$ with clusters $\{A_i \subseteq \mathbb{N}\}_{i=1}^{\infty}$.
- Let ϕ be a random partition i.i.d. from CRP(α, θ)
- Then π' is the coagulation of π if is a coarsening of π where
  - the clusters $\pi' = \{C_j\}_{j=1}^{\infty}$ are:

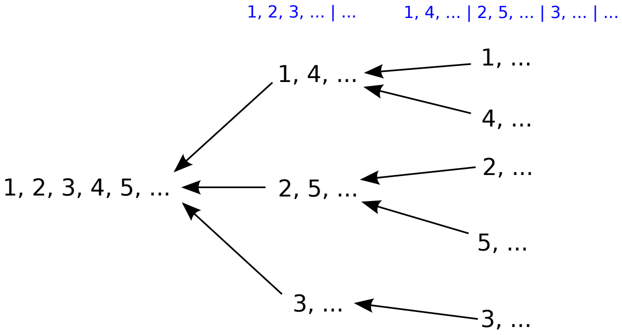
    $$C_j = \bigcup_{i \in B_j} A_i$$

    where $\{B_j\}_{j=1}^{\infty}$ are the clusters of ϕ.
  - i.e., each cluster of π' the union of one or more clusters of π, indexed by ϕ. B_j tells us the indices of the clusters in π to merge to form cluster j in π'.
- Notation: $\pi'|\pi \sim \text{Coag}_{\alpha,\theta}(\pi)$.

Bottom-up hierarchies: Chinese restaurant franchise

[Teh et al, 2004]

- By simple analogy to the nCRP: construct the hierarchy from the bottom instead of from the top.
- $\pi(0)$ places all natural numbers in their own cluster.
- $\pi(t)|\pi(t-1) \sim \text{Coag}_{0,\theta}(\pi(t-1))$



Bottom-up hierarchies: Kingman's coalescent

[Kingman, 1982]

- Bottom-up analogue of Dirichlet diffusion tree.
- Continuous-indexed version of a discrete coagulation tree.
- $\pi(0)$ is the partition with just one cluster.
- $\pi(t + dt)|\pi(t) \sim \mathrm{Coag}_{0,\theta/dt}(\pi(t))$
- Consider the restriction of ϕ to n, the number of clusters in $\pi_m(t)$.
- Let K be the number of clusters in $\pi_m(t+1)$. As $dt \to 0$, the density is:

$$
\begin{aligned}
p(\pi_m(t + dt)|\pi_m(t)) &= p(\phi_n) \\
&= \frac{(\theta/dt)^{K-1}}{(\theta/dt + 1)_{n-1\uparrow 1}} \prod_{i=1}^{K} \Gamma(n_i)
\end{aligned}
$$

Bottom-up hierarchies: Kingman's coalescent

K clusters in $\pi(t+dt)$, n clusters in $\pi(t)$

$$
\begin{aligned}
\lim_{dt\to 0} p(\pi_m(t+dt)|\pi_m(t)) &= \lim_{dt\to 0} \frac{(\theta/dt)^{K-1}}{(\theta/dt+1)_{n-1\uparrow 1}} \prod_{i=1}^{K} \Gamma(n_i) \\
&= \theta^{K-n} \left[\prod_{i=1}^{K} \Gamma(n_i)\right] \lim_{dt\to 0} \frac{dt^{-K+1}}{dt^{-n+1}} \\
&= \begin{cases} 1 & \text{if } n = K, \\ O(dt) & \text{if } K = n-1, \\ O(dt^2) & \text{if } n > K+1. \end{cases}
\end{aligned}
$$

Bottom-up hierarchies: Kingman's coalescent

K clusters in $\pi(t+dt)$, n clusters in $\pi(t)$

$$\lim_{dt\to 0} p(\pi_m(t+dt)|\pi_m(t)) = \lim_{dt\to 0} \frac{(\theta/dt)^{K-1}}{(\theta/dt+1)_{n-1\uparrow 1}} \prod_{i=1}^{K} \Gamma(n_i)$$
$$= \theta^{K-n}\left[\prod_{i=1}^{K}\Gamma(n_i)\right] \lim_{dt\to 0} \frac{dt^{-K+1}}{dt^{-n+1}}$$
$$= \begin{cases} 1 & \text{if } n = K, \\ O(dt) & \text{if } K = n-1, \\ O(dt^2) & \text{if } n > K+1. \end{cases}$$

- The rate at which a cluster in $\pi(t)$ coagulates is $1/\theta$.
- The rate at which any cluster in $\pi(t)$ coagulates is $\frac{K(K-1)}{2}\frac{1}{\theta}$

Bottom-up hierarchies: Λ-coalescent

[Pitman, 1999]

Suppose $\pi_n(t)$ has b blocks. The rate at which any k of these blocks merge into one block is:

$$\lambda_{b,k} = \int_0^1 x^{k-2}(1-x)^{b-k}\Lambda(dx)$$

where Λ is a finite measure on $[0,1]$. When $\Lambda = \frac{1}{\theta}\delta_0$, Kingman's coalescent is recovered.

This gives rise to a broader class of EPPFs: not necessarily of simple form (e.g., Gibbs type) or stationary.

Duality

Pitman (1999; Theorem 12): For all $0 < \alpha < 1$, $0 \leq \beta < 1$, and $\theta > -\alpha\beta$:

$$\begin{array}{c} \pi \sim \mathrm{CRP}(\alpha\beta, \theta) \\ \xi | \pi \sim \mathrm{Frag}_{\alpha, -\alpha\beta}(\pi) \end{array} \quad \Longleftrightarrow \quad \begin{array}{c} \xi \sim \mathrm{CRP}(\alpha, \theta) \\ \pi | \xi \sim \mathrm{Coag}_{\beta, \theta/\alpha}(\xi) \end{array}$$

Duality, marginalisation: Sequence Memoizer

[Wood et al, 2011]

Consider a particular case of the Chinese restaurant franchise where $\theta = 0$:

$$\pi(0) \sim \mathrm{CRP}(\alpha(0), 0)$$
$$\pi(t+1)|\pi(t) \sim \mathrm{Coag}_{\alpha(t+1),0}(\pi(t))$$

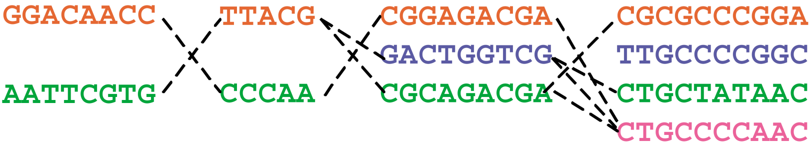
Inductively can marginalize out any $\pi(t+1)$ as, from duality:

$$\pi(t+2)|\pi(t) \sim \mathrm{Coag}_{\alpha(t+2)\alpha(t+1),0}(\pi(t))$$

Since:

$$\begin{array}{c} \xi \sim \mathrm{CRP}(\alpha, 0) \\ \pi|\xi \sim \mathrm{Coag}_{\beta,0}(\xi) \end{array} \iff \begin{array}{c} \pi \sim \mathrm{CRP}(\alpha\beta, 0) \\ \xi|\pi \sim \mathrm{Frag}_{\alpha,-\alpha\beta}(\pi) \end{array}$$

Duality: Discrete Fragmentation Coagulation Processes



$$
\begin{aligned}
\pi(0) &\sim \mathrm{CRP}(0,\theta) \\
\xi(t)|\pi(t) &\sim \mathrm{Frag}_{\alpha,0}(\pi_t) \\
\pi(t+1)|\xi(t) &\sim \mathrm{Coag}_{0,\theta/\alpha}(\xi_t) \\
\implies \pi(t) &\sim \mathrm{CRP}(0,\theta) \\
\xi(t) &\sim \mathrm{CRP}(\alpha,\theta)
\end{aligned}
$$

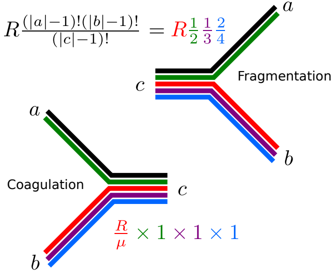
Since from duality:

$$
\begin{aligned}
\pi &\sim \mathrm{CRP}(0,\theta) \\
\xi|\pi &\sim \mathrm{Frag}_{\alpha,0}(\pi)
\end{aligned}
\quad \iff \quad
\begin{aligned}
\xi &\sim \mathrm{CRP}(\alpha,\theta) \\
\pi|\xi &\sim \mathrm{Coag}_{0,\theta/\alpha}(\xi)
\end{aligned}
$$

Duality: Continuous Fragmentation Coagulation Processes

[Teh et al, 2012]

Using the same time scaling of the parameters as Dirichlet diffusion trees and Kingman's coalescent:



$$
\begin{aligned}
\pi(0) &\sim \mathrm{CRP}(0, \theta) \\
\xi(t)|\pi(t) &\sim \mathrm{Frag}_{\alpha dt, 0}(\pi_t) \\
\pi(t + dt)|\xi(t) &\sim \mathrm{Coag}_{0, \theta/\alpha dt}(\xi_t) \\
\implies \pi(t) &\sim \mathrm{CRP}(0, \theta) \\
\xi(t) &\sim \mathrm{CRP}(\alpha, \theta)
\end{aligned}
$$

- Rate of fragmentating a single cluster same as Dirichlet diffusion tree. i.e., αH_{n-1}
- Rate of coagulating a single cluster same as Kingman's coalescent. i.e., α/θ

Exchangeable Fragmentation Coagulation Processes

[Bertoin 2001; Berestycki, 2004]

- EFCPs are a generalisation to all possible exchangeable processes with fragmentation of one block at a time.
- Kingman's paintbox construction of a partition π of $\mathbb{N}$:
    1. Let ϑ be a random open subset of $[0, 1]$,
    2. For each $i \in \mathbb{N}$ associate $U_i \sim \text{Uniform}(0, 1)$
    3. If i and j lie in the same interval of ϑ then i and j cohabit the same cluster.
- Fragmentation is any mixture of erosion (dust formation) and fragmentation by Kingman's paintbox.
- Coagulation is any mixture of Kingman's coalescent and coagulation by Kingman's paintbox.

Summary

- Reviewed a few partition-valued processes.
- Constructed from three fundamental building blocks:
  1. Chinese restaurant process
  2. Fragmentation
  3. Coagulation
- Such constructions ensure exchangeability and projectivity of partitions.

	Discrete	Continuous & Binary	Continuous & Multibranching
Fragmentation	Nested CRP	Dirichet Diffusion Tree	Pitman-Yor Diffusion Tree
Coagulation	Chinese Restaurant Franchise	Kingman's Coalescent	Lambda-coalescent
Fragmentation & Coagulation	DFCP	FCP	EFCP Non-CRP EPPFs