Interval Partition Evolutions

Quan Shi

Universität Mannheim

quanshi.math@gmail.com

University of Liverpool, 26.Feb.2021

Interval Partition Evolutions

A visualisation: `http://www.stats.ox.ac.uk/~winkel/5_sim_skewer.gif`

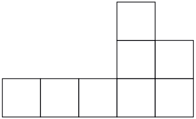
Outline

- Motivation: 3 Problems
- Chinese Restaurant Processes and Interval Partition Evolutions
- Applications

1. Motivation

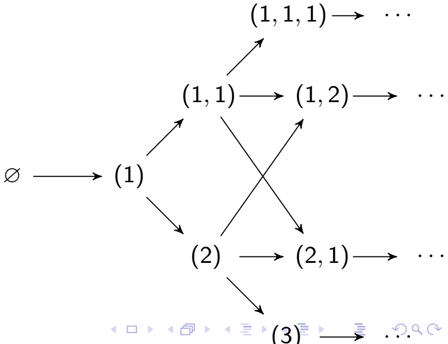
Markov Processes on the Graph of Compositions

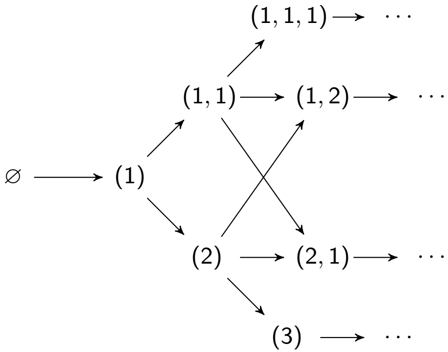
A composition of $n \in \mathbb{N}$ is a tuple $\sigma = (\sigma_1, \dots, \sigma_k)$ of positive integers with $n = \sigma_1 + \cdots + \sigma_k$.

- Keeping track of only the sizes of parts, and not their order: a partition of an integer
- The ranked sequence of a composition: a partition of an integer
- A composition $(1, 1, 1, 3, 2) \Leftrightarrow$ a diagram 

The (directed) graph of compositions

- An edge from σ to λ, denoted by $\sigma \uparrow \lambda$ and $\lambda \downarrow \sigma$: if λ can be obtained from σ by stacking or inserting one box.





- The graph of compositions is an example of branching graphs known in algebraic combinatorics and representation theory.
- Other branching graphs: Young graph of partitions, Pascal triangle
- Scaling limit of Markov chains on Young graph of partitions: Diffusion on the Kingman simplex (Borodin, Olshanski, Fulman, Pertrov)

Question 1: The scaling limit of Markov chains on the graph of compositions?

Labelled infinitely-many-neutral-alleles model with parameter $\theta \geq 0$ (Kimura–Crow, Watterson, Ethier–Kurtz)

- S: the space of allelic types
- Mutations occur with intensity $\theta/2$
- The type of each mutant offspring is chosen independently according to a probability law ν_0 on S
- Basic assumption: every mutant is of a new type, i.e. ν_0 is non-atomic

Characterize the evolution of the relative frequencies of types (Ethier–Kurtz)

- A process $(\mu_t, t \geq 0)$ taking values on $\mathcal{M}_1^a(S)$, the atomic probability measures on S
- Unique stationary distribution: Pitman–Yor distribution $\mathtt{PY}(0, \theta, \nu_0)$
- For $\alpha \in [0, 1)$ and $\theta > -\alpha$, Pitman–Yor distribution on $\mathcal{M}_1^a(S)$:

$$\sum_{i \geq 1} A_i \delta_{U_i} \sim \mathtt{PY}(\alpha, \theta, \nu_0)$$

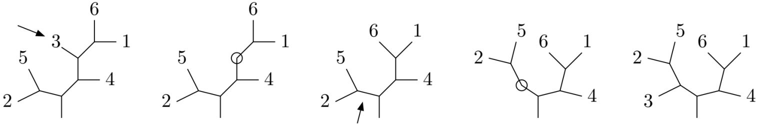
where $(A_i)_{i \geq 1}$ has the Poisson–Dirichlet distribution with parameter (α, θ) on the Kingman simplex and $U_i \sim \nu_0, i \geq 1$, independent of each other.

- Pitman–Yor distributions are widely used in non-parametric Bayesian analysis.

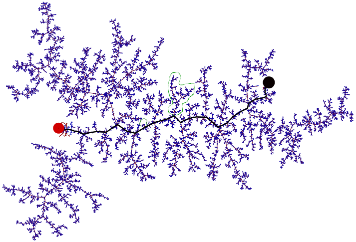
Question 2: generalize the model to a two-parameter family for $\alpha \in [0, 1)$ and $\theta > -\alpha$ (existence conjectured by Feng–Sun)

Continuum-Tree-Valued Diffusions

- For $n \in \mathbb{N}$, a Markov chain on the space of rooted binary labelled trees with n leaves:



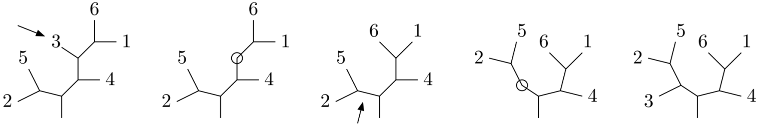
- The law of a uniform binary tree with n leaves is the stationary distribution of this Markov chain.
- As $n \to \infty$, a uniform binary tree with n leaves converges to a Brownian continuum random tree.



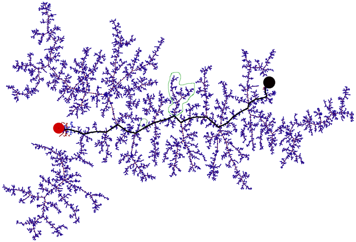
@Kortchemski

Continuum-Tree-Valued Diffusions

- For $n \in \mathbb{N}$, a Markov chain on the space of rooted binary labelled trees with n leaves:



- The law of a uniform binary tree with n leaves is the stationary distribution of this Markov chain.
- As $n \to \infty$, a uniform binary tree with n leaves converges to a Brownian continuum random tree.



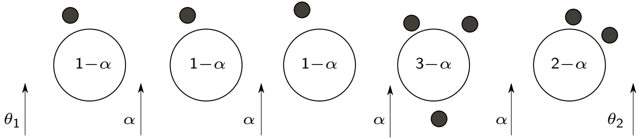
@Kortchemski

- Question 3: As $n \to \infty$, Aldous conjectured a limiting diffusion on the space of continuum trees, with stationary distribution given by the Brownian continuum random tree.

2. Chinese Restaurant Processes and Interval Partition Evolutions

Up-Down Ordered Chinese Restaurant Processes

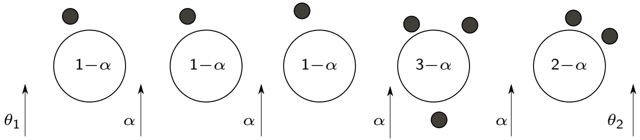
- Tables are ordered in a line.
- Fix $\alpha \in (0,1)$ and $\theta_1, \theta_2 \geq 0$. We construct a continuous-time Markov chain.
- Arriving (up-step):
  - For each occupied table, say there are $m \in \mathbb{N}$ customers, a new customer comes to join this table at rate $m - \alpha$
  - At rate θ_1 and θ_2 respectively, a new customer enters to start a new table at the leftmost and the rightmost positions.
  - Between each pair of two neighbouring occupied tables, a new customer enters and begins a new table there at rate α;



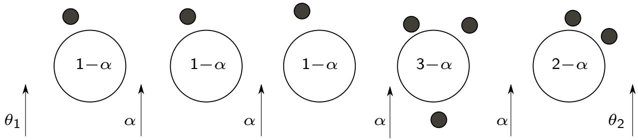
- This is an ordered version of a Chinese Restaurant Process with parameter $\alpha \in (0,1)$ and $\theta = \theta_1 + \theta_2 - \alpha \geq -\alpha$.

Up-Down Ordered Chinese Restaurant Processes

- Tables are ordered in a line.
- Fix $\alpha \in (0,1)$ and $\theta_1, \theta_2 \geq 0$. We construct a continuous-time Markov chain.
- Arriving (up-step):
  - For each occupied table, say there are $m \in \mathbb{N}$ customers, a new customer comes to join this table at rate $m - \alpha$
  - At rate θ_1 and θ_2 respectively, a new customer enters to start a new table at the leftmost and the rightmost positions.
  - Between each pair of two neighbouring occupied tables, a new customer enters and begins a new table there at rate α;



- This is an ordered version of a Chinese Restaurant Process with parameter $\alpha \in (0,1)$ and $\theta = \theta_1 + \theta_2 - \alpha \geq -\alpha$.
- Leaving (down-step): Each customer leaves at rate 1.
- When the restaurant is empty, a new customer arrives at rate $\theta \vee 0$.



- At each time $t \geq 0$, list the numbers of customers of occupied tables, from left to right, by a composition of an integer.
- View the model as a Markov process on the graph of compositions:
  - $(1, 1, 1, 3, 2)$ to $(1, 1, 1, 1, 3, 2)$ at rate $\theta_1 + 3\alpha$
  - $(1, 1, 1, 3, 2)$ to $(1, 1, 3, 2)$ at rate 3.

Definition

The composition-valued process is called a Poissonized up-down ordered Chinese Restaurant Process with parameters $(\alpha, \theta_1, \theta_2)$, $\mathrm{PCRP}(\alpha, \theta_1, \theta_2)$.

Main result

Theorem (S., Winkel)

Let $\alpha \in (0,1)$ *and* $\theta_1, \theta_2 \geq 0$. *For every* $n \in \mathbb{N}$, *let* $(C^{(n)}(t), t \geq 0)$ *be a sequence of* $\mathrm{PCRP}(\alpha, \theta_1, \theta_2)$ *started from* $C^{(n)}(0)$. *Suppose that*

$$n^{-1} C^{(n)}(0) \underset{n\to\infty}{\longrightarrow} \gamma \quad \textit{in the space of interval partitions}$$

Then, under the Skorokhod topology,

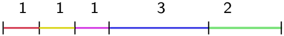
$$(n^{-1} C^{(n)}(2nt), t \geq 0) \underset{n\to\infty}{\longrightarrow} (\beta(t), t \geq 0) \quad \textit{in distribution.}$$

The limiting process $(\beta(t), t \geq 0)$ *is called an* $(\alpha, \theta_1, \theta_2)$*-self-similar interval-partition evolution,* $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$.

The Space of Interval Partitions

$L \geq 0$. We say β is an interval partition of the interval $[0, L]$, if

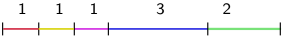
- $\beta = \{(a_i, b_i) \subset (0, L) \colon i \geq 1\}$ a collection of disjoint open intervals
- The total mass (sum of lengths) of β is $\|\beta\| := \sum_{i\geq 1}(b_i - a_i) = L$.
- a composition $(1, 1, 1, 3, 2)$ of integer 8
  an interval partition $\{(0,1), (1,2), (2,3), (3,6), (6,8)\}$ of $[0, 8]$

$(1, 1, 1, 3, 2) \Leftrightarrow$ 

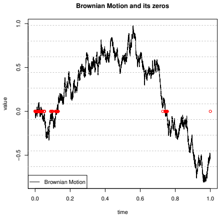
The Space of Interval Partitions

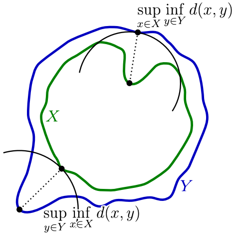
$L \geq 0$. We say β is an interval partition of the interval $[0, L]$, if

- $\beta = \{(a_i, b_i) \subset (0, L) \colon i \geq 1\}$ a collection of disjoint open intervals
- The total mass (sum of lengths) of β is $\|\beta\| := \sum_{i \geq 1}(b_i - a_i) = L$.
- a composition $(1, 1, 1, 3, 2)$ of integer 8
  an interval partition $\{(0, 1), (1, 2), (2, 3), (3, 6), (6, 8)\}$ of $[0, 8]$

  $(1, 1, 1, 3, 2) \Leftrightarrow$ 

- Zero points $\mathcal{Z}$ of a Brownian motion on $(0, 1)$: interval components of the open set $(0, 1) \setminus \mathcal{Z}$ form an interval partition β of $[0, 1]$





- The space $\mathcal{I}$ of all interval partitions is equipped with the Hausdorff metric d_H (between the endpoint sets $[0, L] \setminus \beta$).
- $(\mathcal{I}, d_H)$ is not complete but the induced topological space is Polish.

Main Result

Theorem (S., Winkel)

Let $\alpha \in (0,1)$ *and* $\theta_1, \theta_2 \geq 0$. *For every* $n \in \mathbb{N}$, *let* $(C^{(n)}(t), t \geq 0)$ *be a sequence of* $\mathrm{PCRP}(\alpha, \theta_1, \theta_2)$ *started from* $C^{(n)}(0)$. *Suppose that*

$$n^{-1} C^{(n)}(0) \underset{n\to\infty}{\longrightarrow} \gamma \quad \text{in } (\mathcal{I}, d_H)$$

Then, under the Skorokhod topology,

$$(n^{-1} C^{(n)}(2nt), t \geq 0) \underset{n\to\infty}{\longrightarrow} (\beta(t), t \geq 0) \quad \text{in distribution.}$$

The limiting process $(\beta(t), t \geq 0)$ *is called an* $(\alpha, \theta_1, \theta_2)$*-self-similar interval-partition evolution,* $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$.

Main Result

Theorem (S., Winkel)

Let $\alpha \in (0,1)$ and $\theta_1, \theta_2 \geq 0$. For every $n \in \mathbb{N}$, let $(C^{(n)}(t), t \geq 0)$ be a sequence of $\mathrm{PCRP}(\alpha, \theta_1, \theta_2)$ started from $C^{(n)}(0)$. Suppose that

$$n^{-1} C^{(n)}(0) \underset{n \to \infty}{\longrightarrow} \gamma \quad \text{in } (\mathcal{I}, d_H)$$

Then, under the Skorokhod topology,

$$(n^{-1} C^{(n)}(2nt), t \geq 0) \underset{n \to \infty}{\longrightarrow} (\beta(t), t \geq 0) \quad \text{in distribution.}$$

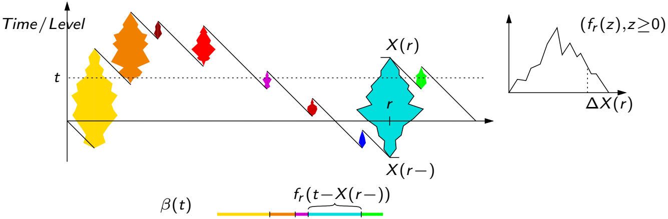
The limiting process $(\beta(t), t \geq 0)$ is called an $(\alpha, \theta_1, \theta_2)$-self-similar interval-partition evolution, $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$.

Comments:

- Scaling limits of Markov chains: on Young graph of partitions (Borodin, Olshanski, Fulman, Pertrov); on the graph of compositions (Rivera-Lopez, Rizzolo)
- Our method is very different from these works.
- We establish a pathwise construction of the limiting process SSIPE.

Construction of SSIPE

- A spectrally positive Lévy process $(X(s), s \geq 0)$ stopped at a random time
- Mark each jump by an excursion $(f_r(z), z \geq 0)$, whose length satisfies $\inf\{z > 0 \colon f_r(z) = 0\} = \Delta X(r) = X(r) - X(r-)$
- A table is add at position r at time/level $X(r-)$, whose size evolves according to f_r
- Skewer at level t: the sizes of ordered tables at level t form an interval partition $\beta(t)$



A simulation: `http://www.stats.ox.ac.uk/~winkel/5_sim_skewer.gif`

Properties of SSIPE

Theorem (Forman, Rizzolo, S., Winkel)

For $\alpha \in (0,1)$ *and* $\theta_1, \theta_2 \geq 0$*, let* $(\beta(t), t \geq 0)$ *be an* $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$ *starting from* $\gamma \in \mathcal{I}$*.*

- *It is a path-continuous Hunt process on* $(\mathcal{I}, d_H)$
- *(Self-similar with index 1) For* $c > 0$*, the space-time rescaled process* $(c\beta(t/c), t \geq 0)$ *is also an* $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$
- *Let* $\theta = \theta_1 + \theta_2 - \alpha$*. There are three phases:*
  - *when* $\theta \geq 1$*, it a.s. never hits* $\emptyset$
  - *when* $\theta \in (0,1)$*, it is reflected at* $\emptyset$
  - *when* $\theta \in [-\alpha, 0]$*, it is absorbed at* $\emptyset$
- *The total mass* $(\|\beta(t)\|, t \geq 0)$ *evolves according to a squared Bessel process of "dimension* 2θ*".*
- *For any* $t > 0$*,* $\beta(t)$ *a.s. has the* α*-diversity property, i.e. the following limit exists for each* $x \geq 0$*:*

$$\mathcal{D}_\alpha(x) := \Gamma(1-\alpha) \lim_{h \downarrow 0} h^\alpha \#\{(a,b) \in \beta(t) : \ |b-a| > h, \ b \leq x\}.$$

Remark: when $\theta_2 = \alpha$, we extend SSIPE to the completion of $(\mathcal{I}, d_H)$

De-Poissonized process and Stationary Distribution

Theorem (Forman, Rizzolo, S., Winkel)

For an $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$ $(\beta(t), t \geq 0)$, *introduce a Lamperti/Shiga-type time-change*

$$\tau(u) := \inf\left\{t \geq 0 \colon \int_0^t \|\beta(r)\| dr > u\right\}, \quad u \geq 0.$$

The de-Poissonized $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$ *(renormalized and time-changed)*

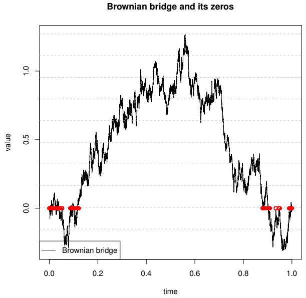
$$\bar{\beta}(u) := \|\beta(\tau(u))\|^{-1} \beta(\tau(u)), \qquad u \geq 0$$

is a continuous Hunt process on the space of unit interval partitions, with stationary distribution denoted by $\mathtt{PDIP}(\alpha, \theta_1, \theta_2)$.

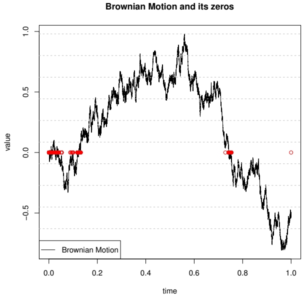
Poisson–Dirichlet Interval Partition $\mathtt{PDIP}(\alpha, \theta_1, \theta_2)$

- The ranked lengths of intervals in a $\mathtt{PDIP}(\alpha, \theta_1, \theta_2)$ has the law of Poisson–Dirichlet distribution (α, θ) on the Kingman simplex with $\theta = \theta_1 + \theta_2 - \alpha$.
- Stick-breaking construction (S., Winkel)
- When $\theta_2 = \alpha$: related to regenerative composition structures (Gnedin–Pitman, Winkel–Pitman)
- Examples:

$\mathtt{PDIP}(1/2, 1/2, 1/2)$: zero points of a Brownian bridge on $[0, 1]$ from zero to zero.



$\mathtt{PDIP}(1/2, 1/2, 0)$: zero points of a Brownian motion on $[0, 1]$.

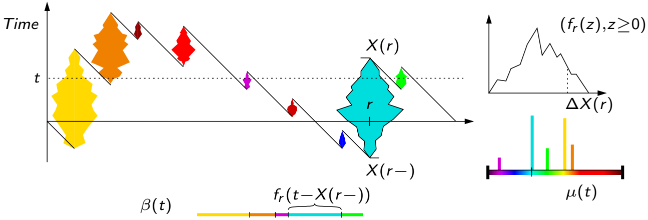


3. Applications

Projection on the Kingman Simplex

- An $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$ is the scaling limit of certain Markov processes on the graph of compositions
- Consider a de-Poissonized $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$ $(\bar{\beta}(u), u \geq 0)$:
  Write $W(u)$ for the ranked interval lengths of $\bar{\beta}(u)$. Then (Forman, Pal, Rizzolo, Winkel) prove that the process $(W(u/2), u \geq 0)$ is an EKP diffusion on the Kingman simplex introduced by (Ethier–Kurtz,Petrov)
- An EKP diffusion is the scaling limit of certain Markov chains on the graph of partitions (Ethier–Kurtz,Petrov).

A Related Population-Genetic Model



(Forman, Rizzolo, S., Winkel)

- A Lévy process marked by a pair (f_r, U_r): an excursion f_r and an independent *allelic type* $U_r \sim \nu_0$ (colour).
- Statistic of alleles: a measure-valued process $(\mu(t), t \geq 0)$ associated with an SSIPE $(\beta(t), t \geq 0)$.
- The de-Poissonized process has a stationary distribution: the Pitman–Yor distribution $\mathtt{PY}(\alpha, \theta, \nu_0)$ with $\alpha \in (0, 1)$ and $\theta = \theta_1 + \theta_2 - \alpha \geq -\alpha$.
- This generalizes the *labelled infinitely-many-neutral-alleles* model ($\alpha = 0$) by (Ethier–Kurtz).

Continuum-Tree-Valued Diffusions

- $\rho \in (1, 2]$, ρ-stable continuum random tree [Aldous, Duquesne, LeGall]
- $\rho = 2$: Brownian Continuum-random tree
- Question: construct a continuum-tree-valued diffusion which is stationary under the law of the ρ-stable continuum random tree?
- In the Brownian case $\rho = 2$: Aldous's conjectured diffusion (Forman–Pal–Rizzolo–Winkel, Löhr–Mytnik–Winter)
- Idea: using the de-Poissonized SSIPE with stationary distribution $\mathtt{PDIP}$ (S.–Winkel in progress)

Further questions:

- Ford's tree growth model (Ford)
- Alpha-gamma model (Chen–Ford–Winkel)
- Continuum fragmentation trees (Haas, Miermont, Pitman, Winkel)

,

PDIPs in Continuum Random Trees

- A ρ-stable tree is a metric space equipped with a mass measure of total mass 1.
  With $\alpha = 1 - 1/\rho$:

masses of spinal bushes $(M_i)_{i\geq 1}$
distances to the root $(\ell_i)_{i\geq 1}$ $\overset{\text{law}}{\Longleftrightarrow}$ $\beta = \{U_i, i \geq 1\} \sim \mathtt{PDIP}(\alpha,\alpha,\alpha)$,
α-diversity $(\mathcal{D}_\alpha(\inf U_i), i \geq 1)$.

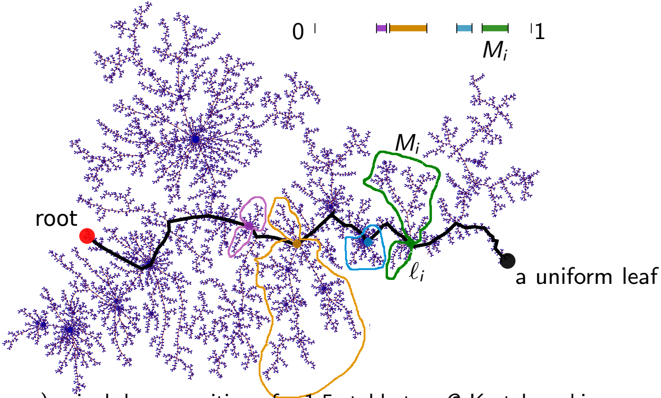


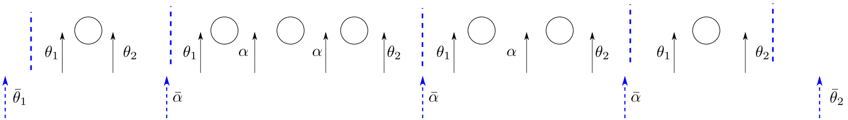
Figure: (coarse) spinal decomposition of a 1.5-stable tree © Kortchemski

- Difficulty in the non-Brownian case: branch point with infinite degree (S., Winkel): nested SSIPE

Scaling limit of Nested PCRP

Nested PCRP: clusters of tables

- Clusters: $\mathrm{PCRP}(\bar{\alpha}, \bar{\theta}_1, \bar{\theta}_2)$
- Tables in each cluster: $\mathrm{PCRP}(\alpha, \theta_1, \theta_2)$
- Consistency: $\theta := \theta_1 + \theta_2 - \alpha = -\bar{\alpha} < 0$



- Nested Chinese restaurant processes are widely applied in non-parametric Bayesian analysis.
- **S., Winkel**: the scaling limit of nested PCRP is nested SSIPE.
- Applications to multifurcating trees: bushes of subtrees
- Key ingredient: $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$ excursion away from $\emptyset$ with $\theta < 0$ (though $\emptyset$ is an absorbing state!)

SSIPE excursions

Theorem (S., Winkel)

Let $(C(t), t \geq 0)$ be a $\mathrm{PCRP}(\alpha, \theta_1, \theta_2)$ starting from the composition of 1 and suppose that $\theta := \theta_1 + \theta_2 - \alpha < 1$. Denote the law of the process $(n^{-1}C(2nt), t \geq 0)$ by $\mathrm{P}^{(n)}$. Then the following convergence holds vaguely:

$$n^{1-\theta} \cdot \mathrm{P}^{(n)} \underset{n \to \infty}{\longrightarrow} \Lambda.$$

The limit Λ is a sigma-finite measure on the space of continuous interval-partition excursions away from $\emptyset$.

Comments:

- When $\theta > 0$, an $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$ is reflected at $\emptyset$ and Λ is the Itô measure.
- When $\theta \leq 0$, an $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$ is absorbed at $\emptyset$, but our description of Λ still makes sense of an $\mathrm{SSIPE}(\alpha, \theta_1, \theta_2)$ excursion measure.

Ideas of the proof

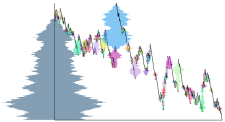
- In a Chinese restaurant process, a new table has one customer initially, and this table is removed when all customers of this table have left.
- the number of customers at a single table evolves according to a birth-death process X, started from 1 and absorbed at zero, with birth rate $P_{i \to i+1} = i - \alpha$ and death rate $P_{i \to i-1} = i$.

Lemma

Denote the law of the process $(\frac{1}{n} X(2nt), t \geq 0)$ by $\pi^{(n)}$. The following convergence holds vaguely:

$$2\alpha n^{1+\alpha} \cdot \pi^{(n)} \underset{n\to\infty}{\longrightarrow} \mu.$$

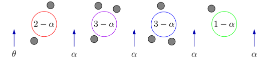
The limit μ is a σ-finite measure of the space of positive continuous excursions: excursion measure of (-2α)-dimensional squared Bessel process (Pitman–Yor)



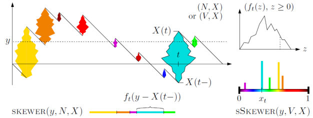
- Define the lifetime of a positive excursion f by $\zeta(f) := \sup\{s \geq 0 \colon f(s) > 0\}$. Then $\mu(\zeta(f) \in \cdot)$ coincides with the Lévy measure of scaffolding stable $(1+\alpha)$ precess.
- Under $\mu(\cdot \mid \zeta(f) = x)$, the excursion is a (-2α)-dimensional squared Bessel process of length x.

Summery

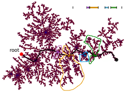
- We have constructed a three-parameter family of interval-partition diffusions
- Scaling limit of Markov chains on the graph of compositions



- Generalized labelled infinitely-many-neutral-alleles model



- Future work: continuum-tree-valued process with stationary distribution



- N. Forman, D. Rizzolo, Q. Shi and M. Winkel.
  A two-parameter family of measure-valued diffusions with Poisson–Dirichlet stationary distributions.
  arXiv:2007.05250.
- N. Forman, D. Rizzolo, Q. Shi and M. Winkel.
  Diffusions on a space of interval partitions: the two-parameter model.
  arXiv:2008.02823.
- Q. Shi and M. Winkel.
  Two-sided immigration, emigration and symmetry properties of self-similar interval partition evolutions.
  arXiv:2011.13378.
- Q. Shi and M. Winkel.
  Up-down ordered Chinese restaurant processes with two-sided immigration, diffusion limits and emigration.
  arXiv:2012.15758.

Thanks!