Markov chains on partitions and their diffusion analogs

Soumik Pal
University of Washington, Seattle

Dec 05, 2016

Where are we heading?

- High-dimensional interacting Markov chains are complicated.
- Sometimes they can be simplified to independent one-dimensional systems.
- This helps conceptual understanding and allows computations.

Where are we heading?

- High-dimensional interacting Markov chains are complicated.
- Sometimes they can be simplified to independent one-dimensional systems.
- This helps conceptual understanding and allows computations.
- Simple example: How to sample from a $(k-1)$ dimensional simplex?
- Generate $X_1, X_2, \ldots, X_k$ iid Exponential(1). $S = X_1 + \ldots + X_k$.

$$\left(\frac{X_1}{S}, \frac{X_2}{S}, \ldots, \frac{X_k}{S} \right).$$

- Will generalize above to Markov chains.

Partitions and compositions

- F - finite set, say $[n] = \{1, 2, \ldots, n\}$.
- A partition of F in k blocks is $\{A_1, \ldots, A_k\}$, disjoint and $F = \cup_i A_i$.
- A composition of n in k blocks is $(n_1, \ldots, n_k)$, $\sum_i n_i = n$.

- Partitions naturally occur in clustering problems.
- Composition = Vector of sizes of clusters.
- $k = \#$ possible labels for the data points.

Exchangeability and random partitions

- Bayesian nonparametric clustering: Ishwaran-James, Griffiths-Gharamani, Gershman-Blei, Broderick-Jordan-Pitman.
- Requires prior distributions on the set of possible partitions.

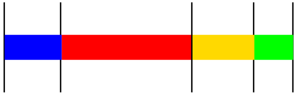
Exchangeability and random partitions

- Bayesian nonparametric clustering: Ishwaran-James, Griffiths-Gharamani, Gershman-Blei, Broderick-Jordan-Pitman.
- Requires prior distributions on the set of possible partitions.

- n data points. Assume exchangeability of data labels.
- Suffices to generate random composition: say, $(N_1, N_2, \ldots, N_k)$.
- How to generate random exchangeable partition?
- Kingman's paintbox representation.

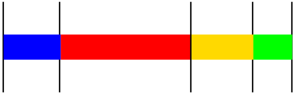
Interval partitions

- Start with random composition $(N_1, N_2, \ldots, N_k)$.
- Partition $[0,1]$ in successive disjoint subintervals of size $(I_1, I_2, \ldots, I_k)$, where $|I_j| = N_j/n$.
- Random interval partition of $[0,1]$.



Interval partitions

- Start with random composition $(N_1, N_2, \ldots, N_k)$.
- Partition $[0,1]$ in successive disjoint subintervals of size $(I_1, I_2, \ldots, I_k)$, where $|I_j| = N_j/n$.
- Random interval partition of $[0,1]$.



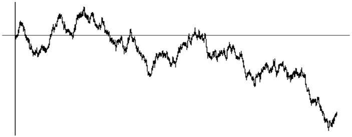
- Sample iid Uni$[0,1]$, $U_1, U_2, \ldots, U_n$.
- If U_j lands in I_k, put data j in cluster k.
- Exchangeable random partition of $[n]$.

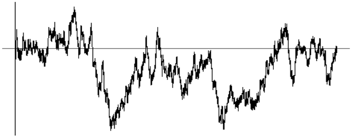
Models for interval partition

- Nice models for interval partitions of $[0, 1]$?
- Multinomial$(n; p_1, \ldots, p_k)$: throw n balls in k ordered bins.
- Dirichlet$(\alpha_1, \ldots, \alpha_k)$: density on $(k-1)$ simplex $\propto \prod_i x_i^{\alpha_i - 1}$.

Models for interval partition

- Nice models for interval partitions of $[0,1]$?
- Multinomial$(n; p_1, \ldots, p_k)$: throw n balls in k ordered bins.
- Dirichlet$(\alpha_1, \ldots, \alpha_k)$: density on $(k-1)$ simplex $\propto \prod_i x_i^{\alpha_i - 1}$.
- Excursion lengths of Brownian motion/ bridge:





Poissonization and de-Poissonization of Markov chains

Up-down chains on partitions

- Composition of n: $(n_1, n_2, \ldots, n_k)$. Interval partition of $[0, 1]$.
- (up-move) Sample a new point from the paintbox.
- (down-move) Delete an existing point at random.
- Large n, repeat many times. How does this Markov chain behave?

Up-down chains on partitions

- Composition of n: $(n_1, n_2, \ldots, n_k)$. Interval partition of $[0, 1]$.
- (up-move) Sample a new point from the paintbox.
- (down-move) Delete an existing point at random.
- Large n, repeat many times. How does this Markov chain behave?
- Consider diffusion limit. As $n \to \infty$:

$$\left(\frac{n_1}{n}, \frac{n_2}{n}, \ldots, \frac{n_k}{n}\right) \text{ at time } \lfloor n^2 t \rfloor$$

  converges to a k-dim diffusion in $(k-1)$-simplex.
- $(X_1, \ldots, X_k)(t)$, $t \geq 0$. Wright-Fisher diffusion.
- Easiest way to understand sample paths is to Poissonize.

Poissonization

- Make time continuous. Break dependency between blocks.
- Each block i adds a point at rate n_i. Each existing data point vanishes at rate 1.
- Total number of data points fluctuates. Every block evolves independently.
- Diffusion limit of the size of each block as $n \to \infty$?

Poissonization

- Make time continuous. Break dependency between blocks.
- Each block i adds a point at rate n_i. Each existing data point vanishes at rate 1.
- Total number of data points fluctuates. Every block evolves independently.
- Diffusion limit of the size of each block as $n \to \infty$?
- The $\texttt{BESQ}(0)$ process is given by

$$f(t) = y_0 + 2\int_0^t \sqrt{f(s)}dB(s), \quad y_0 > 0.$$

- When it hits zero, stop.

De-Poissonization

- When Poissonized, each block is a $\mathtt{BESQ}(0)$ processes $Y_1, \ldots, Y_k$.
- Moreover, the original can be recovered by de-Poissonization.

De-Poissonization

- When Poissonized, each block is a $\mathtt{BESQ}(0)$ processes $Y_1, \ldots, Y_k$.
- Moreover, the original can be recovered by de-Poissonization.
- Let $S(t) = Y_1(t) + \ldots Y_k(t)$. Then (see Warren-Yor, P.)

$$\{(X_1, \ldots, X_k)(t),\ t \geq 0\} \stackrel{d}{=} \left\{ \left(\frac{Y_1}{S}, \ldots, \frac{Y_k}{S} \right)(\tau_t),\ t \geq 0 \right\}.$$

- Here τ_t is the Bessel clock time-reparametrization:

$$\tau_t = \inf \left\{ s \geq 0 : \int_0^s \frac{du}{S_u} > t \right\}.$$

De-Poissonization

- When Poissonized, each block is a $\texttt{BESQ}(0)$ processes $Y_1, \ldots, Y_k$.
- Moreover, the original can be recovered by de-Poissonization.
- Let $S(t) = Y_1(t) + \ldots Y_k(t)$. Then (see Warren-Yor, P.)

$$\{(X_1, \ldots, X_k)(t),\ t \geq 0\} \stackrel{d}{=} \left\{ \left(\frac{Y_1}{S}, \ldots, \frac{Y_k}{S} \right) (\tau_t),\ t \geq 0 \right\}.$$

- Here τ_t is the Bessel clock time-reparametrization:

$$\tau_t = \inf \left\{ s \geq 0 : \int_0^s \frac{du}{S_u} > t \right\}.$$

- Dirichlet distributions are ratios of independent Gamma r.v.s.

Some remarks

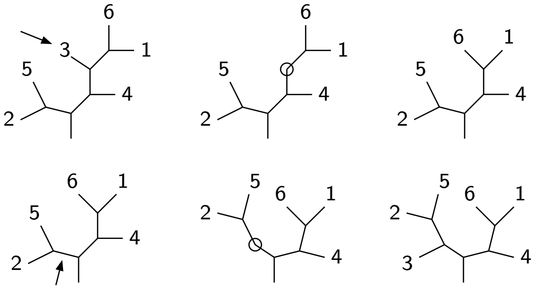
- Poissonization reduces a multi-dim diffusion problem to one dimension.
- Time-change (reparametrization) does not affect path properties.
- Exact computations are now easy.
- E.g., only one block survives. Distribution when each $Y_i = 0$ is inverse-Gamma. All independent.

Some remarks

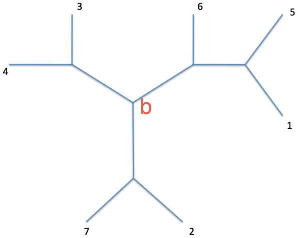
- Poissonization reduces a multi-dim diffusion problem to one dimension.
- Time-change (reparametrization) does not affect path properties.
- Exact computations are now easy.
- E.g., only one block survives. Distribution when each $Y_i = 0$ is inverse-Gamma. All independent.
- Does not work exactly for Markov chains. Only in diffusion limits.

Up-down chains on nested partitions and trees

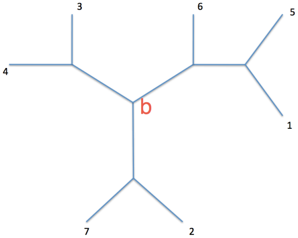
Binary trees



- Markov chain on rooted *n*-leaf binary trees. Equivalently, nested partitions.
- Down-move: delete uniform random leaf, contract edge.
- Up-move: select uniform random edge, insert branch point and a new leaf-edge.
- Stationary w.r.t. uniform distribution. (Aldous '01)



- Fix an internal branch point.
- $X_i = \#$ leaves for ith subtree. $(X_1/n, X_2/n, X_3/n)$ - Markov chain.
- Take n large, run time fast by n^2. What do we observe?



- Fix an internal branch point.
- $X_i = \#$ leaves for ith subtree. $(X_1/n, X_2/n, X_3/n)$ - Markov chain.
- Take n large, run time fast by n^2. What do we observe?
- Poissonize: Each leaf dies at rate two; each edge gives birth at rate one.
- Each leaf count now becomes independent.

- The squared Bessel processes $\texttt{BESQ}(-1)$ is limit of Galton-Watson branching processes with emigration rate 1.
- $\texttt{BESQ}(-1)$ is a diffusion, and solves

$$f(t) = y_0 - t + 2\int_0^t \sqrt{f(s)}dB_s, \quad y_0 > 0.$$

- The squared Bessel processes $\mathtt{BESQ}(-1)$ is limit of Galton-Watson branching processes with emigration rate 1.
- $\mathtt{BESQ}(-1)$ is a diffusion, and solves

$$f(t) = y_0 - t + 2\int_0^t \sqrt{f(s)}dB_s, \quad y_0 > 0.$$

- Generate independent $\mathtt{BESQ}(-1)$ processes (Y_1, Y_2, Y_3). Let $S(t) = Y_1(t) + Y_2(t) + Y_3(t)$.
- De-Poissoniation: original Markov chain converges to

$$\left(\frac{Y_1}{S}, \frac{Y_2}{S}, \frac{Y_3}{S}\right), \quad \text{up to time reparametrization.}$$

Here τ_t is the Bessel clock time-reparametrization:

$$\tau_t = \inf\left\{s \geq 0 : \int_0^s \frac{du}{S_u} > t\right\}.$$

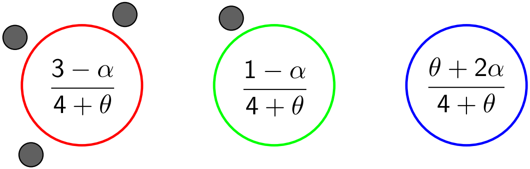
Up-down Markov chains on Chinese restaurants

The (α, θ) Chinese restaurant process

$\alpha \in [0, 1)$ and $\theta \geq 0$. Customer $(n+1)$ chooses according to the rule:

- Suppose there are k occupied tables.
- join a table i with n_i other customers with probability $(n_i - \alpha)/(\theta + n)$, or
- sit at an empty table with probability $(\theta + k\alpha)/(\theta + n)$.

Probabilities of customer 5 joining each table

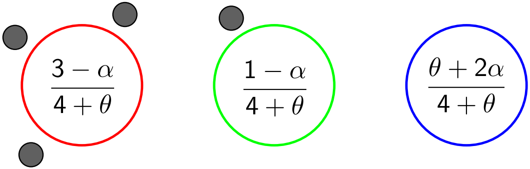


The (α, θ) Chinese restaurant process

$\alpha \in [0, 1)$ and $\theta \geq 0$. Customer $(n+1)$ chooses according to the rule:

- Suppose there are k occupied tables.
- join a table i with n_i other customers with probability $(n_i - \alpha)/(\theta + n)$, or
- sit at an empty table with probability $(\theta + k\alpha)/(\theta + n)$.

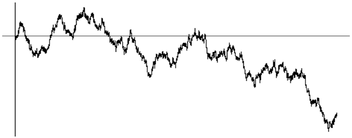
Probabilities of customer 5 joining each table

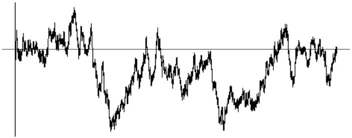


Due to Dubins-Pitman. See Kingman '75; Perman-Pitman-Yor '92; Pitman-Yor '97 Pitman, *Combinatorial Stochastic Processes*, '03.

Poisson-Dirichlet distributions $\mathtt{PD}(\alpha, \theta)$

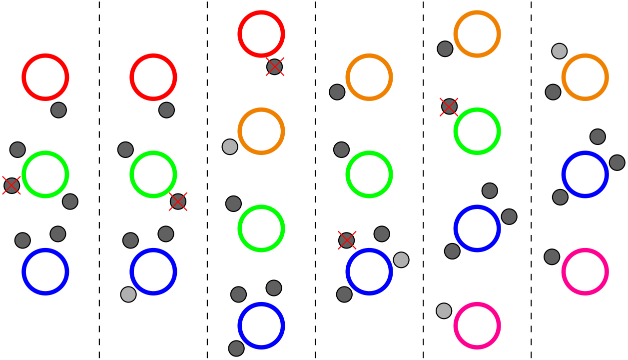
- Kingman simplex: $\nabla_\infty = \{x_1 \geq x_2 \geq \ldots \geq 0, \ \sum_{i \in \mathbb{N}} x_i = 1\}$.
- As $n \to \infty$, ranked proportions $\left(\frac{n_{(1)}}{n} \geq \frac{n_{(2)}}{n} \geq \ldots\right)$ in $\mathtt{CRP}(\alpha, \theta)$ is $\mathtt{PD}(\alpha, \theta)$.
- Ranked lengths of excursion intervals of BM (BB) have law $\mathtt{PD}\left(\frac{1}{2}, 0\right)$ (resp. $\mathtt{PD}\left(\frac{1}{2}, \frac{1}{2}\right)$).





A Chinese restaurant with re-seating

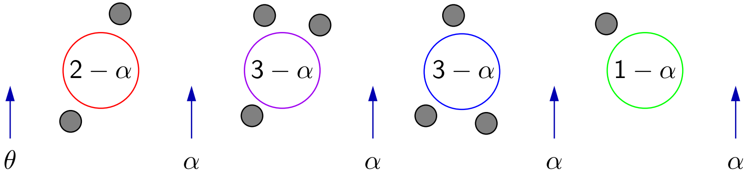
- Up-down chain on compositions of $[n]$.
- Down move: uniform random customer leaves. Up-move: re-enters according to $\texttt{CRP}(\alpha, \theta)$ seating rule.



- See Petrov '09; Borodin-Olshanski '05, '09; Fulman '05.

Poissonized, ordered CRP with re-seating

- Diffusion limit of up-down chain? A special ordering is required.
- Poissonization:



- For table w/ $m > 0$ customers, new people join with rate $m - \alpha$.
- For table w/ $m > 0$ customers, people enter to begin new tables immediately to the right w/ rate α.
- Customers begin new tables at far left end with rate θ.
- Each customer leaves w/ rate 1.

Continuum limit of Poissonized ordered CRP with re-seating

- What is the continuum limit of the Markov chain above?
- Notice that, Poissonization changes the total number of customers.
- How can we obtain the diffusion limit of the original (non-Poissonized) chain? De-Poissonization.

Continuum limit of Poissonized ordered CRP with re-seating

- What is the continuum limit of the Markov chain above?
- Notice that, Poissonization changes the total number of customers.
- How can we obtain the diffusion limit of the original (non-Poissonized) chain? De-Poissonization.
- Will focus on $(\alpha, \theta) = (1/2, 0)$.
- Result: Diffusion on interval partitions stationary with respect to the law of Brownian excursion intervals.
- When intervals lengths are ranked, gives back $\mathtt{PD}\,(1/2, 0)$.

Thank you very much.