

IMI共同利用研究 公開ワークショップ 「実社会に見られる複雑なネットワークと無限粒子系の交差点Ⅱ」

本研究課題の目標：

「無限粒子系と実問題の共通部分を見つける」

ワークショップの目的：

問題意識の共有と話題提供

問題意識：

昨年度の課題の継続・発展→最初の講演で話したいこと

IMI共同利用研究

「実社会に見られる複雑なネットワークと無限粒子系の交差点Ⅱー複雑ネットワーク上の情報流ー」

(組織委員)

森 隆大 (京都大学 数理解析研究所)

林 晃平 (東京大学大学院数理科学研究科)

須田 颯 (慶應義塾大学 理工学研究科)

上島 芳倫 (北海道大学 数理・データサイエンス教育研究センター)

長田 翔太 (九州大学マス・フォア・インダストリ研究所)

新井 裕太 (千葉商科大学 基盤教育機構)

早川 知志 (Mathematical Institute, University of Oxford)

(アドバイザー)

恐神 貴行 (IBM東京基礎研究所)

IMI共同利用研究 公開ワークショップ 「実社会に見られる複雑なネットワークと無限粒子系の交差点Ⅱ」

ワークショップのキーワード：

行列式点過程, サンプルング, 機械学習, etc...

今日の講演：

早川 知志 (University of Oxford・大学院生)

数値積分のための点過程とサブサンプリング

平尾 将剛 (愛知県立大学・准教授)

球面上の行列式点過程とその周辺

中村 ちから (株式会社ニコン 研究開発本部 数理技術研究所)

機械学習手法による極値統計学のハイパーパラメータ選定について

恐神 貴行 (IBM東京基礎研究所)

強化学習における多様性

IMI共同利用研究 公開ワークショップ 「実社会に見られる複雑なネットワークと無限粒子系の交差点Ⅱ」

参加者の方へのお願い

- ZOOM表示名とGoogle form登録名の統一をお願いします。
- スクリーンショットや録画等のご遠慮ください。
- 本ワークショップは録画しています。
- 講演中はマイクをオフにしてください。

- 進 行：講演30分~40分+ディスカッション（共同ホスト）+休憩10~20分。
- 質疑応答：共同ホストは随時チャット・音声。
参加者は・講演中→チャット ・講演後→チャット・挙手

IMI共同利用研究

「実社会に見られる複雑なネットワークと無限粒子系の交差点Ⅱー複雑ネットワーク上の情報流ー」

（組織委員）

森 隆大（京都大学 数理解析研究所）

林 晃平（東京大学大学院数理科学研究科）

須田 颯（慶應義塾大学 理工学研究科）

上島 芳倫（北海道大学 数理数理・データサイエンス教育研究センター）

長田 翔太（九州大学マス・フォア・インダストリ研究所）

新井 裕太（千葉商科大学 基盤教育機構）

早川 知志（Mathematical Institute, University of Oxford）

（アドバイザー）

恐神 貴行（IBM東京基礎研究所）

実社会に見られる複雑なネットワークと無限粒子系の交差点

九州大学マス・フォア・インダストリ研究所

長田翔太

2021年8月19日（木）

「実社会に見られる複雑なネットワークと無限粒子系の交差点Ⅱ」

昨年度の振り返り

昨年度の目標：

- ・無限粒子系：ポイントクラウド, 信号, ランダムウォーク. . .
 - ・複雑なネットワーク：複雑ネットワーク, ニューラルネットワーク. . .
- これらのインターセクションを見つける.

昨年度の成果：

- ・行列式点過程を用いたモンテカルロ法による数値積分.
- 離散化・カップリングによるコストと精度のトレードオフを提案.

今年度の方針：

- ・昨年度の議論を継続・発展させる.
- 具体的課題の設定・広がりの開拓.

中心的トピック：行列式点過程, 機械学習

目次

内容：昨年度の振り返り（MC with DPP, コストと精度のトレードオフ）

◆行列式点過程（Determinantal Point Process）

◆Monte Carlo with DPP

◆DPPの離散化

◆DPPのカップリング

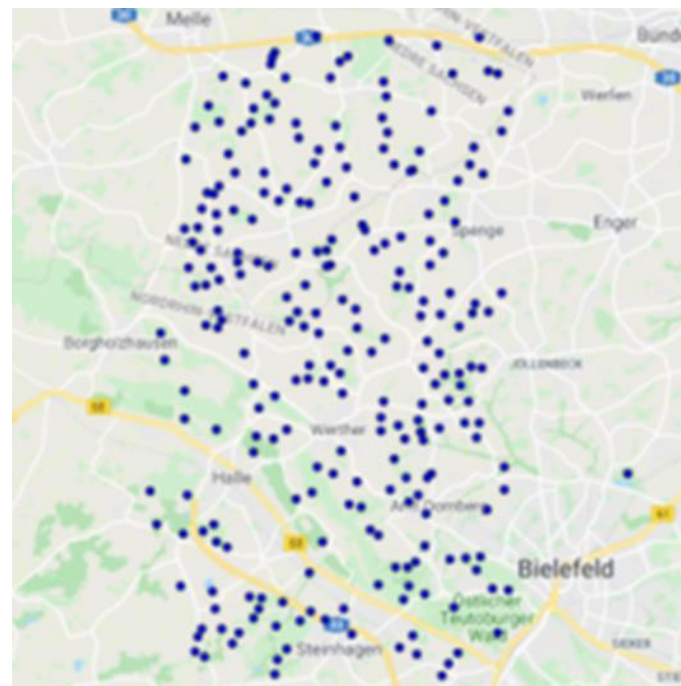
窓ガラスの雨粒 と 鳥の巣

窓ガラスに付いた雨粒



雨粒は窓ガラスに一様の確率で当たる

鳥の巣（ノスリ）の分布



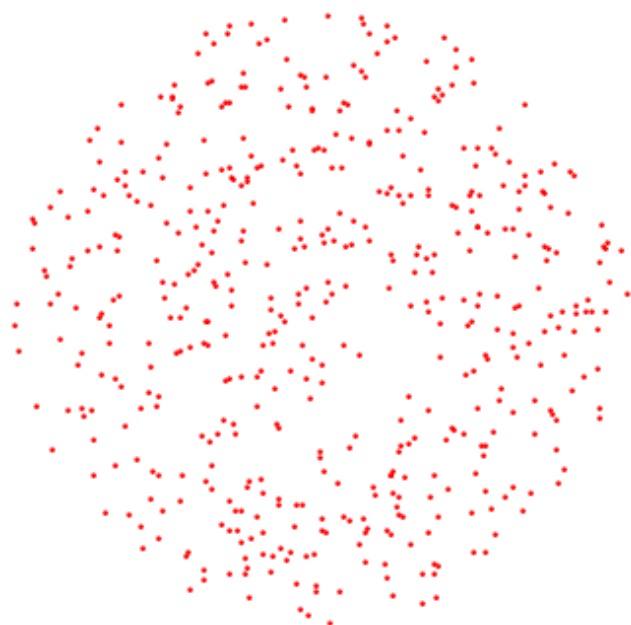
巣はお互いに距離をとる

これらは異なる点過程で表現される

右図： G. Akeman, et al., 2021, Journal of Theoretical Biology.より引用

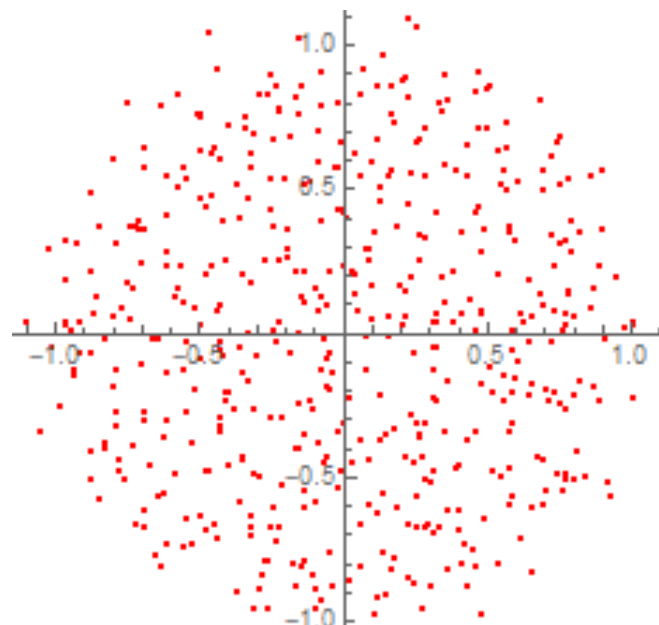
Poisson点過程 と 行列式点過程 : 窓ガラスの雨粒 と 鳥の巣

窓ガラスの雨粒
Poisson (ポアソン) 点過程



規則性がない点配置
窓ガラスの雨粒・月のクレーター
自由粒子のスナップショット

鳥の巣
行列式点過程



整然とした点配置
動物の巣・電波の基地局
相互作用を持つ粒子のスナップショット

それぞれ500個の粒子

行列式点過程の概要

➤粒子系

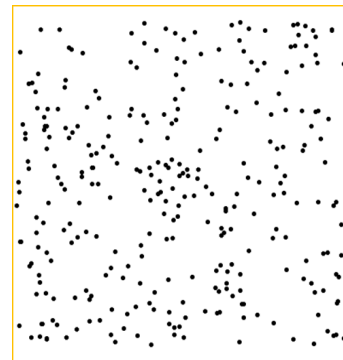
- ランダムな粒子配置のスナップショット
- 粒子配置のランダムダイナミクス

➤行列式点過程は反発力が働く粒子系を表す.

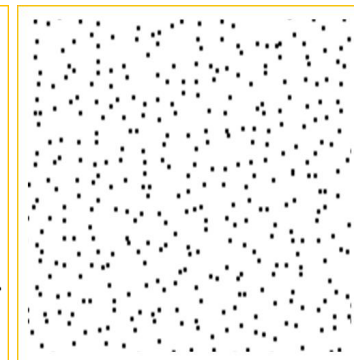
➤連続モデル, 離散モデルの様々な対象に現れる.

➤数理モデリングでは, 行列式点過程はRepulsiveなモデルのサンプリングとして注目される.

➤アイテムのピックアップ, Lebesgue測度の近似, クラスタリング



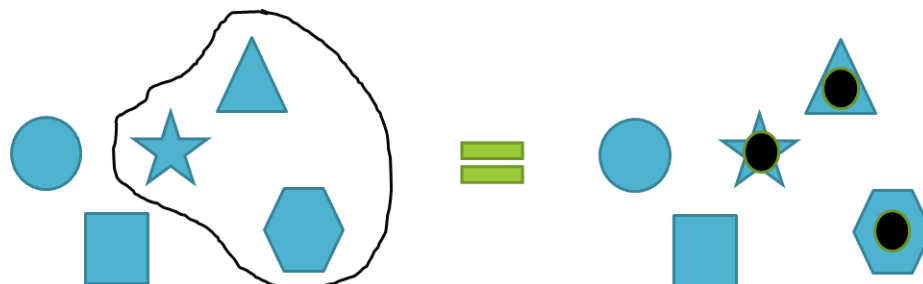
Poisson点過程
無相関



行列式点過程
負の相関 (反発力)

行列式点過程の構成定理：有限粒子 1

- $S = \{1, \dots, N\}$: 有限集合
- $L: S \times S \rightarrow \mathbb{C}$: 正定値行列.
- このとき,



$$\sum_{X \subset S} \det [L_X] = \det [I + L].$$

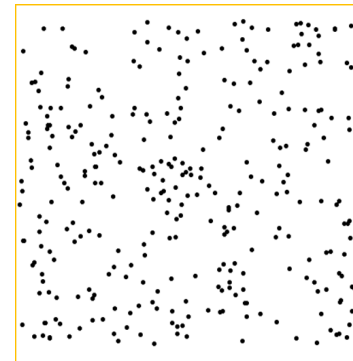
- $\{0,1\}^S$ 上の確率測度を次で定義：

$$\mathbb{P}[0^{S-X} 1^X] := \frac{\det [L_X]}{\det [I + L]}.$$

このとき, $X \subset S$ 上の粒子の存在確率は,

$$\mathbb{P}[1^X] = \det \left[\frac{L}{I + L} \Big|_X \right].$$

点過程の定義



S : Polish空間. 完備距離付け可能な可分空間

配置空間 :

$$\text{Conf}(S) = \{S \text{ 上の非負整数値Radon測度 } \xi = \sum_n \delta_{x_n} \text{ 全体} \}$$

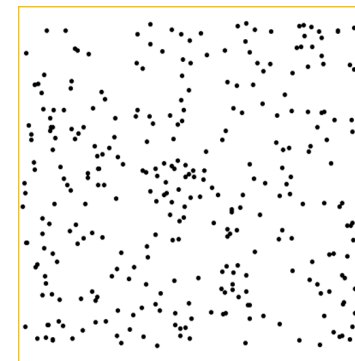
漠位相 :

$$\int_S f(x) \xi_n(dx) \rightarrow \int_S f(x) \xi(dx) \quad (\forall f \in C_c(S))$$

$\text{Conf}(S)$ -値確率変数 $\mathcal{X}$ (またはその分布 $\mathbb{P}^{\mathcal{X}}$) を点過程という.

・ S が離散位相空間かつ, 点過程 $\mathcal{X}$ が多重点を持たないとき,
配置空間 $\text{Conf}(S)$ は $\{0,1\}^S$ (または 2^S) と同一視できる.

行列式点過程の定義



- 点過程 $\mathcal{X}$ は以下の相関関数 $\{\rho^n\}$ で決定される:

$$\int_{A_1^{k_1} \times \dots \times A_l^{k_l}} \rho^n(x_1, \dots, x_n) dm^{\otimes n} = \mathbb{E} \left[\prod_{i=1}^l \frac{\mathcal{X}(A_i)!}{(\mathcal{X}(A_i) - k_i)!} \right]$$

ただし $A_1, \dots, A_l$ は disjoint, $k_1 + \dots + k_l = n$, m は S 上の Radon 測度.

- $\rho^1(x)$ は点密度を表す.
- S が離散位相空間のとき, $m = \sum_{x \in S} \delta_x$ に対して

$$\rho^n(x_1, \dots, x_n) = \mathbb{P}[1^{\{x_1, \dots, x_n\}}].$$
- 相関関数が核関数 $K: S \times S \rightarrow \mathbb{C}$ の行列式で与えられる点過程を **行列式点過程** という.

$$\rho^n(x_1, \dots, x_n) = \det[K(x_i, x_j)]_{i,j=1, \dots, n}$$

行列式点過程とPoisson点過程との違い

定常Poisson点過程

$$\rho^n(x_1, \dots, x_n) = \sum_{i=1}^n \rho^1(x_i)$$

- 粒子はお互いを気にしない

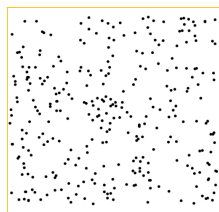
$$\rho^2(x, y) = \rho^1(x)\rho^1(y)$$

- 分散は大きい

$$\mathbb{V} \left[\sum_{x \in \mathcal{X}} 1_{B_r}(x) \right] \sim r^2 \text{ as } r \rightarrow \infty$$

- 異なる場所の粒子分布は独立

$$\sigma[\mathcal{X}_{B_r}] \perp \sigma[\mathcal{X}_{B_r^c}]$$



Poisson点過程

Ginibre点過程

$$\rho^n(x_1, \dots, x_n) = \det[K(x_i, x_j)]$$

- 粒子がお互いを避ける

$$\rho^2(x, y) \leq \rho^1(x)\rho^1(y)$$

- 分散が小さい

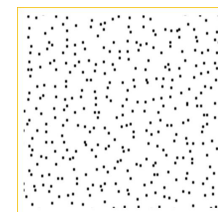
$$\mathbb{V} \left[\sum_{x \in \mathcal{X}} 1_{B_r}(x) \right] \sim r \text{ as } r \rightarrow \infty$$

- 剛性を持つ

$\mathcal{X}(B_r)$: $\sigma[\mathcal{X}_{B_r^c}]$ -可測

- 基地局の配置モデル

- ルベーク測度の近似



行列式点過程

行列式点過程の構成定理

- S : 可算基を持つ局所コンパクトHausdorff空間
- m : S 上のラドン測度
 - (S, m) の典型例は $(\mathbb{Z}^d, \sum \delta_z)$, $(\mathbb{R}^d, dx)$
- $K: S \times S \rightarrow \mathbb{C}$
- Macchi(1975)

連続な K が以下① – ③を満たすとき行列式点過程が存在する :

- ① $\overline{K(x, y)} = K(y, x)$
- ② $\mathcal{K}f := \int_S K(x, y)f(y)m(dy), f \in L^2(S, m)$ はトレイス・クラス.
- ③ $0 \leq \mathcal{K} < I$.

Macchi(1975)ではFermion processと呼んでいる.

$$K(x, y) = \sum_{n \in \mathbb{N}} \lambda_n \varphi_n(x) \varphi_n(y).$$

行列式点過程の構成定理：無限粒子

- S : 可算基を持つ局所コンパクトハウスドルフ空間
- m : S 上のラドン測度
- $K: S \times S \rightarrow \mathbb{C}$
 - (S, m) の典型例は $(\mathbb{Z}^d, \sum \delta_z)$, $(\mathbb{R}^d, dx)$
- Macchi(1975)-Shirai and Takahashi(2000, 2003)-Soshnikov(2000)
- K が以下①－③を満たすとき行列式点過程が一意に存在する.

① $\overline{K(x, y)} = K(y, x).$

② $\mathcal{K}$ は局所トレイス・クラス

i.e. $\Lambda \subset S$ コンパクトに対し, $\mathcal{K}_\Lambda f := \int_\Lambda K(x, y) f(y) m(dy)$ はトレイス・クラス.

③ $0 \leq \mathcal{K} \leq I.$

*特に, ①+②を満たす (K, m) に対し, ③は行列式点過程の存在の必要十分条件.

*①－③の条件は, $\mathcal{K}$ (または K) のpositive contraction propertyともよぶ.

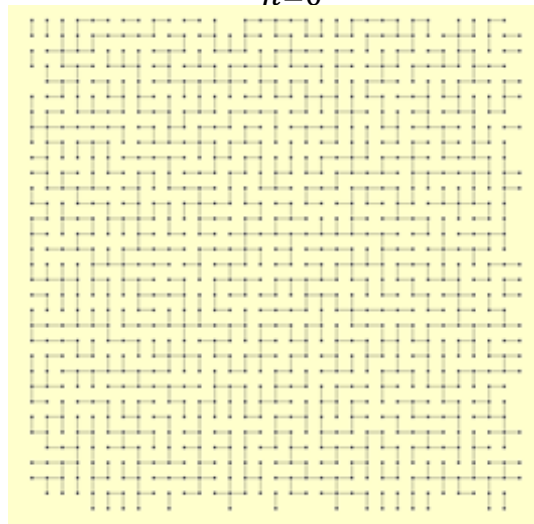
行列式点過程の構成定理：例

$$\textcircled{1} \overline{K(x, y)} = K(y, x) \quad \textcircled{2} \mathcal{K} \text{は局所トレイス・クラス.} \quad \textcircled{3} 0 \leq \mathcal{K} \leq I.$$

ランダム行列の固有値分布：

$$K^N(x, y) = \sum_{n=0}^{N-1} \varphi_n(x) \overline{\varphi_n(y)}.$$

一様全域木：



行列式点過程の例：直交多項式

Gaussian unitary ensembleの固有値分布

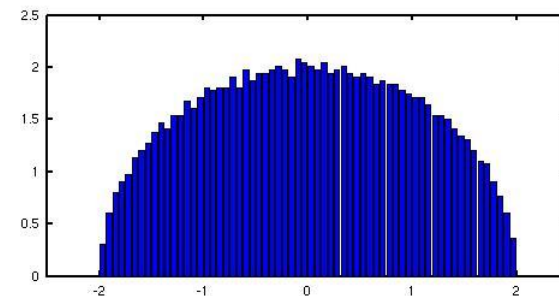
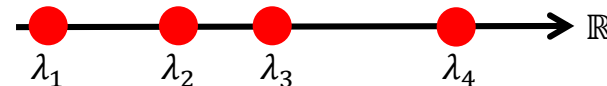
- $X_{i,j}, Y_{i,j} \sim \mathcal{N}_{\mathbb{R}}(0,1), i \leq j, i, j = 1, \dots, N$: i.i.d.
- $H^N = \{H_{i,j}\}_{i,j=1,\dots,N}$ s. t. $H_{i,j} = \begin{cases} X_{i,j} + iY_{i,j}, & i < j \\ X_{i,i}, & i = j \\ X_{i,j} - iY_{i,j}, & j < i \end{cases}$

• 固有値分布は $\mathbb{R}$ 上の行列式点過程：

$$K^N(x, y) = \sum_{n=0}^{N-1} \Psi_n(x) \Psi_n(y),$$

$$\Psi_n(x) = C_n h_n(x) \sqrt{e^{-\frac{x^2}{2}}}.$$

ここで, h_n : Hermit多項式, C_n : 正規化定数.



Monte Carlo法と行列式点過程

- Monte Carlo with determinantal point processes,

R Bardenet, A Hardy - The Annals of Applied Probability, 2020.

- 適切なサンプリング $\{x_i\}_{i=1}^N$ と重み $\{w_i\}_{i=1}^N$ を与えて f の積分値を求めたい :

$$\int f(x)\mu(dx) \approx \sum_{i=1}^N w_i f(x_i).$$

cf. ガウス求積 ($f: I = [-1,1]$ 上の多項式に対する数値積分公式)

- [BH2020]核関数がNevai-classの多変量**直交多項式**で与えられるDPP $\mathcal{X}$,

$(-1,1)^d$ 上のコンパクト台を持つ C^1 -級関数 f に対して誤差

$$\varepsilon_N(f) = \left(\sum_{i=1}^N w_i f(x_i) - \int f(x)\mu(dx) \right).$$

$$\sqrt{\mathbb{V}[\varepsilon_N(f)]} \sim \mathcal{O}\left(N^{-\left(1+\frac{1}{d}\right)/2}\right)$$

- DPPのsmall variance性と思える.

- 典型的には $\mathcal{O}\left(N^{-\frac{1}{2}}\right)$

Monte Carlo法と行列式点過程：コストと精度のトレードオフ

- MC with DPPは近似精度がいい。
- 一般にサンプリングコストは N^3 のオーダー。

- 問題意識：

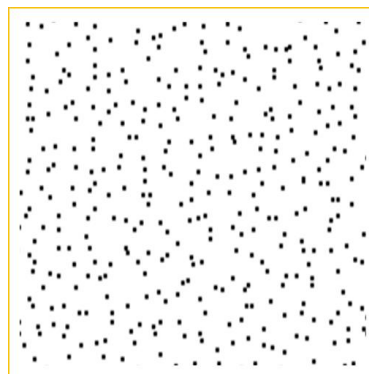
近似精度とサンプリングコストのトレードオフは可能か？

- 昨年度の提案：

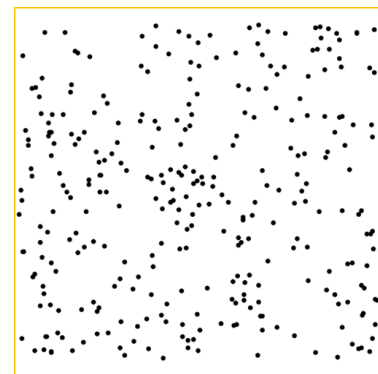
1. 行列式点過程の離散近似（離散DPPのサンプルコストを下げる研究が多々ある）
例えば, Dereziński, M., Calandriello, D., & Valko, M. (2019). Exact sampling of determinantal point processes with sublinear time preprocessing. *arXiv preprint arXiv:1905.13476*.
2. カップリングによるポアソン点過程と行列式点過程の補完
（長田&森, 2021+）

L^2 空間の間の等長変換と行列式点過程の離散近似

- L^2 空間の等長変換を与えると, 行列式点過程も変形する.
- 目的に応じて適切な行列式点過程のクラスを選択することで,
より良いサンプリングやモデリングが期待できるだろう.



行列式点過程
(反発力)

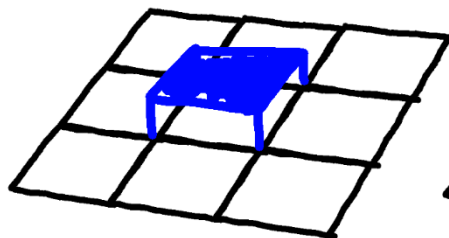


Poisson点過程
(無相関)

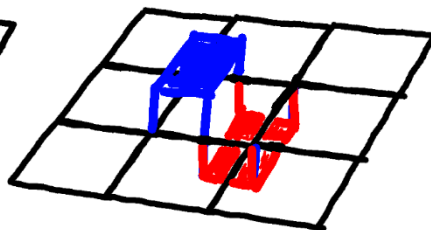
L^2 空間の間の等長変換と行列式点過程の離散近似

$\{\Delta(\ell)\}_{\ell \in \mathbb{N}}$: S の分割の単調増大列で, 各点を分離する.

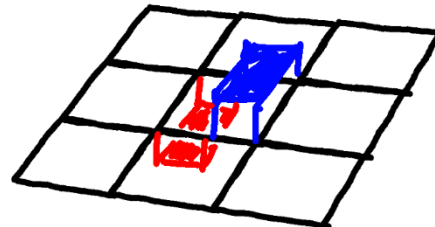
$\{\phi_{D,k}^\ell(x)\}_{(D,k) \in \Delta(\ell) \times \mathbb{N}_0}$: $\phi_{D,0}^\ell$ が $D \in \Delta(\ell)$ 上の定義関数となる $L^2(S, m)$ 上の Haar ウェーブレット.



$\phi_{D \times 0}^\ell$



$\phi_{D \times 1}^\ell$



$\phi_{D \times 2}^\ell$

...

...

L^2 空間の間の等長変換と行列式点過程の離散近似

(S, m) : 連続空間, ラドン測度 m を備える.

$\{\Delta(\ell)\}_{\ell \in \mathbb{N}}$: S の分割の単調増大列で, 各点を分離する.

$\{\phi_{D,k}^\ell(x)\}_{(D,k) \in \Delta(\ell) \times \mathbb{N}_0}$: $\phi_{D,0}^\ell$ が $D \in \Delta(\ell)$ 上の定義関数となる $L^2(S, m)$ 上の Haar ウェーブレット.

$U^\ell: \phi_{D,k}^\ell \mapsto \delta_{D,k}$ は $L^2(S, m) \rightarrow l^2(\Delta(\ell) \times \mathbb{N}_0)$ への 等長変換.

(K, m) -行列式点過程 $\mathbb{P}^K$ に対し,

$$\mathbb{K}^\ell := UKU^{-1}.$$

構成定理より $\Delta(\ell) \times \mathbb{N}_0$ 上の $\mathbb{K}^\ell$ -行列式点過程 $\mathbb{P}^{K,\ell}$ は存在する. 核関数は

$$\mathbb{K}^\ell(i, j) = \int_{S \times S} \phi_i^\ell(x) K(x, y) \phi_j^\ell(y) m(dx) m(dy), \quad i, j \in \Delta(\ell) \times \mathbb{N}_0.$$

このとき, $\mathbb{P}^K(\#_D = n) = \mathbb{P}^{K,\ell}(\#_{D \times \mathbb{N}_0} = n), D \in \Delta(\ell), n \in \mathbb{N}.$

L^2 空間の間の等長変換と行列式点過程の離散近似

$\{\Delta(\ell)\}_{\ell \in \mathbb{N}}$: S の分割の単調増大列.

$\{\phi_{D,k}^\ell(x)\}_{(D,k) \in \Delta(\ell) \times \mathbb{N}_0}$: Haar ウェーブレット. (右図)

(K, m) -行列式点過程 $\mathbb{P}^K$ に対し,

$\Delta(\ell) \times \mathbb{N}_0$ 上の $\mathbb{K}^\ell$ -行列式点過程 $\mathbb{P}^{K,\ell}$ を次で定める :

$$\mathbb{K}^\ell(i, j) = \int_{S \times S} \phi_i^\ell(x) K(x, y) \phi_j^\ell(y) m(dx) m(dy), \quad i, j \in \Delta(\ell) \times \mathbb{N}_0.$$

このとき, $\mathbb{P}^K(\#_D = n) = \mathbb{P}^{K,\ell}(\#_{D \times \mathbb{N}_0} = n), D \in \Delta(\ell), n \in \mathbb{N}.$

$\mathbb{P}^{K,\ell}$ は $\mathbb{P}^K$ の Tree representation と呼ばれる.

Tree representation は固有関数に依らない構成という意味でロバスト.

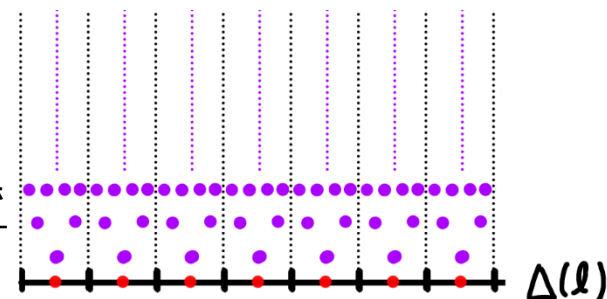
正規直行系を目的に応じて取り換えることで, 線形統計量などの DPP の持つ特性を抽出できないか?

● : $\phi_{D,k}^\ell(x) = \frac{1}{\sqrt{D}} \mathbf{1}_D(x)$



● : $\phi_{D,0}^\ell(x) = \frac{1}{\sqrt{D}} \mathbf{1}_D(x)$





ご清聴ありがとうございました