

A limit theorem for persistence diagrams of random complexes built over marked point processes ^{*1}

九州大学 マス・フォア・インダストリ研究所 須崎 清剛^{*2}

Kiyotaka Suzaki

Institute of Mathematics for Industry,

Kyushu University

概要

本稿は、2018年度確率論シンポジウムにおける筆者の発表内容とその補足をまとめたものである。マーク付き点過程に対して、その部分集合各々に発生時刻を与えることでランダムな単体複体の増大列を構成する。その増大列中のホモロジー類の発生・持続・消滅を記述するパーシステンス図を点過程とみなしたとき、その過程についてある種の大数の法則が得られることを述べる。

0 導入

パーシステントホモロジーとその表現の1つであるパーシステンス図は、様々なデータの位相的特徴を理解する際に非常に有効な道具である。その基本的な考えは、次のとおりである。まず、入力されたデータとして Euclid 空間内の有限集合が与えられたとき、各データ点を徐々に太らせていく過程を考える。すなわち、初期時刻で有限集合が与えられたとして、時刻 t で各データ点を中心とする半径 t の閉球を与える。この過程は、様々な解像度でデータを解析していると思うこともできる。次に各時刻において、「 $q+1$ 個の閉球が空でない交わりをもつとき、対応する $q+1$ 個のデータ点からなる q 単体を与える」という規則で単体複体を構成する。このような規則で構成される単体複体は、(データの被覆である閉球族の) 脈体、もしくは Čech 複体と呼ばれ、脈体定理によって閉球和と構成された単体複体はホモトピー同値になることが知られている。いま、 $t \geq 0$ をパラメータとする単体複体の増大列が得られているが、横軸を発生 (birth) 時刻、縦軸を消滅 (death) 時刻として、その増大列中の q 次ホモロジー類の発生・消滅を $\Delta = \{(x, y) \in [0, \infty]^2 : x < y\}$ 上にプロットした図は、 q 次パーシステンス図と呼ばれる。パーシステンス図は、様々なデータの新たな視点からの比較や特徴付けを行うことを可能にし、データ解析、画像処理、材料工学、センサーネットワークなど様々な分野に応

^{*1} 本研究は、JST CREST JPMJCR15D3 の支援を受けたものである。

^{*2} k-suzaki@imi.kyushu-u.ac.jp

用されている。

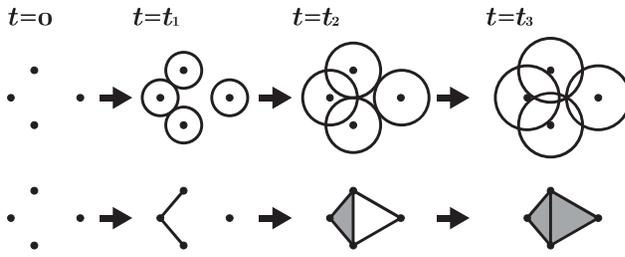


図1 下部は閉球族の脈体 (Čech 複体)

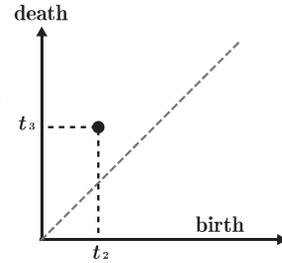


図2 1次パーシステンス図

本稿では入力データをマーク付き点過程，すなわちランダムなマーク付き集合とし，単体複体へ対応させる規則をより一般の規則としたとき，そのパーシステンス図の大数の法則について考える．ここでマークとは，各データ点に新たに付加された情報を意味する．例えば分子模型がデータとして与えられたとき，問題によっては各原子の配置だけではなくその半径も考慮すべきである．マーク付きの点集合から我々が構成する単体複体は，それぞれ異なる半径をもつ閉球族の脈体を例として含んでいる．また，様々な点過程から誘導されるランダム複体のパーシステンス図について理解を深めることは，点過程の標本配置の新たな幾何学的理解を深めることとなり，非常に重要である．本研究は，九州大学の白井朋之氏との共同研究に基づく．

1 準備

この節では主結果を述べるために必要な準備を行う．位相空間 S に対して $\mathcal{F}(S)$ を S の非空の有限部分集合全体とする． $\mathcal{F}(S)$ 上で定義された関数 f が与えられたとき，正整数 k ごとに成分に関する置換不変な S^k 上の関数 f_k が存在して， $f_k(s_1, s_2, \dots, s_k) = f(\{s_1, s_2, \dots, s_k\})$ をみたす． f が可測であるとは，すべての k に対して f_k が Borel 可測となることをいう． \mathbb{M} は Polish 空間とし，可測関数 $\kappa : \mathcal{F}(\mathbb{R}^d \times \mathbb{M}) \rightarrow [0, \infty)$ は次の条件 (K1), (K2), (K3) をみたすとする．

(K1) $A \subset B$ ならば $\kappa(A) \leq \kappa(B)$ ．

(K2) 単調非減少な $\rho : [0, \infty) \rightarrow [0, \infty)$ が存在して，すべての $(x, m), (y, n) \in \mathbb{R}^d \times \mathbb{M}$ に対して， $|x - y| \leq \rho(\kappa(\{(x, m), (y, n)\}))$ が成立する．

(K3) すべての $a \in \mathbb{R}^d$ に対して， $\kappa(T_a A) = \kappa(A)$ が成り立つ．ここで， $T_a : \mathcal{F}(\mathbb{R}^d \times \mathbb{M}) \rightarrow \mathcal{F}(\mathbb{R}^d \times \mathbb{M})$ は第1成分の a による平行移動，すなわち $T_a A = \{(x + a, m) :$

$(x, m) \in A$ である.

$\pi : \mathbb{R} \times \mathbb{M} \rightarrow \mathbb{R}$ を第 1 成分に関する射影とする. $\tilde{\Xi} \in \mathcal{F}(\mathbb{R}^d \times \mathbb{M})$ が単純マーク付き集合であるとは, $\pi|_{\tilde{\Xi}}$ が単射となるときのことをいい, このとき $\Xi = \pi(\tilde{\Xi})$ と表す. 単純な $\tilde{\Xi} \in \mathcal{F}(\mathbb{R}^d \times \mathbb{M})$ に対しては, π により自然な全単射 $\mathcal{F}(\tilde{\Xi}) \ni \tilde{\sigma} \mapsto \sigma \in \mathcal{F}(\Xi)$ が誘導され, 各 $\tilde{\sigma} = \{(x_0, m_0), (x_1, m_1), \dots, (x_q, m_q)\}$ は, \mathbb{R}^d 内の有限点配置として与えられるデータ $\sigma = \{x_0, x_1, \dots, x_q\}$ にマーク $\{m_0, m_1, \dots, m_q\}$ が付加されていると見なせる. 単純な $\tilde{\Xi} \in \mathcal{F}(\mathbb{R}^d \times \mathbb{M})$ が与えられたとき, 抽象単体複体の増大列 $\mathbb{K}(\tilde{\Xi}) = \{K(\tilde{\Xi}, t)\}_{t \geq 0}$ を

$$K(\tilde{\Xi}, t) = \{\sigma \subset \Xi : \kappa(\tilde{\sigma}) \leq t\}$$

で定める. すなわち $\kappa(\tilde{\sigma})$ は, $\mathbb{K}(\tilde{\Xi})$ における単体 σ の発生時刻である. $\mathbb{K}(\tilde{\Xi})$ を κ -フィルター付き複体, もしくは κ -フィルトレーションと呼ぶ.

例 1. $R > 0$, $\mathbb{M} = [0, R]$ とし, $\mathcal{F}(\mathbb{R}^d \times \mathbb{M})$ 上の関数 κ を

$$\kappa(\{(x_0, r_0), (x_1, r_1), \dots, (x_q, r_q)\}) = \inf_{w \in \mathbb{R}^d} \max_{0 \leq i \leq q} (\|x_i - w\| - r_i)^+$$

をみたすものとして定める. ここで $a \in \mathbb{R}$ に対して $a^+ = \max\{a, 0\}$ である. 中心 x , 半径 r の \mathbb{R}^d における閉球を $\overline{B}_r(x)$ と表すと

$$\kappa(\{(x_0, r_0), (x_1, r_1), \dots, (x_q, r_q)\}) \leq t \Leftrightarrow \bigcap_{i=0}^q \overline{B}_{t+r_i}(x_i) \neq \emptyset$$

であることが容易にわかる. したがってこの κ とマーク付き集合 $\tilde{\Xi} = \{(x_i, r_i)\} \in \mathcal{F}(\mathbb{R}^d \times \mathbb{M})$ によって構成される κ -フィルター付き複体の各時刻 t における単体複体 $K(\tilde{\Xi}, t)$ は, マークに依存して定まる $\Xi = \{x_i\}$ の被覆 $\{\overline{B}_{t+r_i}(x_i)\}$ の脈体である.

例 2. \mathbb{M} を $[0, \infty)$ 上の右連続狭義単調増加関数からなる有限集合 $\{\mathbf{r}_i(\cdot)\}_{i \in I}$ とおく. 関数 $\kappa : \mathcal{F}(\mathbb{R}^d \times \mathbb{M}) \rightarrow [0, \infty)$ を

$$\kappa(\{(x_0, \mathbf{r}_0), (x_1, \mathbf{r}_1), \dots, (x_q, \mathbf{r}_q)\}) = \inf_{w \in \mathbb{R}^d} \max_{0 \leq i \leq q} \mathbf{r}_i^{-1}(\|x_i - w\|)$$

と定める. ここで $\mathbf{r} \in \mathbb{M}$ に対して $\mathbf{r}^{-1}(t) = \inf\{s \geq 0 : \mathbf{r}(s) \geq t\}$ である. 先程の例と同様にして

$$\kappa(\{(x_0, \mathbf{r}_0), (x_1, \mathbf{r}_1), \dots, (x_q, \mathbf{r}_q)\}) \leq t \Leftrightarrow \bigcap_{i=0}^q \overline{B}_{\mathbf{r}_i(t)}(x_i) \neq \emptyset$$

であることが示せる. したがって, この κ とマーク付きデータから構成される κ -フィルター付き複体を考えることは, 各データ点にマークとして付加された関数を半径増大度として太らせていった過程に付随する脈体を考えていることに対応する.

適当な \mathbb{M} と関数 κ を定めることで, 閉球に限らず様々な集合の脈体や, Rips 型の単体複体を与えることも可能である ([5]). 以下では κ を 1 つ固定して議論する.

体 \mathbb{F} と非負整数 q に対して, $H_q(K(\tilde{\Xi}, t))$ を \mathbb{F} 上の $K(\tilde{\Xi}, t)$ の q 次ホモロジー群とし, $r \leq s$ に対して, $\iota_r^s : H_q(K(\tilde{\Xi}, r)) \rightarrow H_q(K(\tilde{\Xi}, s))$ を包含写像 $K(\tilde{\Xi}, r) \hookrightarrow K(\tilde{\Xi}, s)$ から誘導される線形写像とする. $H_q(\mathbb{K}(\tilde{\Xi})) = (\{H_q(K(\tilde{\Xi}, t))\}_{t \geq 0}, \{\iota_r^s\}_{r \leq s})$ を, q 次パーシステントホモロジー群と呼ぶ. 単項イデアル整域上の次数付き加群の構造定理により, 非負整数 n_q と $(b_i, d_i) \in \Delta$ ($i = 1, 2, \dots, n_q$) が存在して, $H_q(\mathbb{K}(\tilde{\Xi}))$ は区間分解

$$H_q(\mathbb{K}(\tilde{\Xi})) \simeq \bigoplus_{i=1}^{n_q} I(b_i, d_i)$$

をもつことが知られている ([6]). ここで $I(b_i, d_i)$ は, 線形空間と線形写像の列

$$I(b_i, d_i) : 0 \rightarrow \dots \rightarrow 0 \rightarrow \mathbb{F} \xrightarrow{\text{id}_{\mathbb{F}}} \dots \xrightarrow{\text{id}_{\mathbb{F}}} \mathbb{F} \rightarrow 0 \rightarrow \dots \rightarrow 0$$

at $t = b_i$ at $t = d_i$

である. すなわち各時刻の q 次ホモロジー群の基底を与えるのに十分な n_q 個の q 次ホモロジー類 $\{[c]_i\}_{i=1}^{n_q}$ が存在して, 各 $[c]_i$ は $t = b_i$ で発生し, $b_i \leq t < d_i$ 間で持続し, $t = d_i$ で消滅することを表している. 多重集合 $D_q(\mathbb{K}(\tilde{\Xi})) = \{(b_i, d_i) \in \Delta : i = 1, 2, \dots, n_q\}$ を q 次パーシステンス図という. 我々は $D_q(\mathbb{K}(\tilde{\Xi}))$ を数え上げ測度

$$\xi_q(\mathbb{K}(\tilde{\Xi})) = \sum_{(b,d)} m_{b,d} \delta_{(b,d)}$$

と同一視し, パーシステンス図の収束を Δ 上の Radon 測度の漠収束で考える. ここで, $m_{b,d}$ は $D_q(\mathbb{K}(\tilde{\Xi}))$ における (b, d) の多重度である.

$\tilde{\Phi}$ は $\mathbb{R}^d \times \mathbb{M}$ 上の点過程であって, \mathbb{R}^d へ射影した点過程 $\Phi(\cdot) = \tilde{\Phi}(\cdot \times \mathbb{M})$ が \mathbb{R}^d 上の単純点過程となるとき, \mathbb{M} をマーク空間とする \mathbb{R}^d 上のマーク付き点過程であるという. また, このとき Φ はマーク付き点過程 $\tilde{\Phi}$ のグラウンド過程と呼ばれる. グラウンド過程 Φ が全有限モーメントをもつとは, \mathbb{R}^d の任意の有界な Borel 可測集合 A と $p \geq 1$ に対して, $\mathbb{E}[\Phi(A)^p] < \infty$ が成り立つときをいう. $a \in \mathbb{R}^d$ と d 次直交行列 $A \in O(d)$ による $\mathbb{R}^d \times \mathbb{M}$ 上の平行移動と回転を

$$T_a : \mathbb{R}^d \times \mathbb{M} \ni (x, m) \mapsto (x + a, m) \in \mathbb{R}^d \times \mathbb{M},$$

$$R_A : \mathbb{R}^d \times \mathbb{M} \ni (x, m) \mapsto (Ax, m) \in \mathbb{R}^d \times \mathbb{M}$$

というように第 1 成分のみのそれぞれで定める. これらの作用は配置空間 $\text{Conf}(\mathbb{R}^d \times \mathbb{M})$ 上の平行移動と回転

$$\begin{aligned} (T_a)_* &: \text{Conf}(\mathbb{R}^d \times \mathbb{M}) \ni \mu \mapsto \mu \circ T_a^{-1} \in \text{Conf}(\mathbb{R}^d \times \mathbb{M}), \\ (R_A)_* &: \text{Conf}(\mathbb{R}^d \times \mathbb{M}) \ni \mu \mapsto \mu \circ R_A^{-1} \in \text{Conf}(\mathbb{R}^d \times \mathbb{M}) \end{aligned}$$

を誘導する. マーク付き点過程 $\tilde{\Phi}$ が定常であるとは, その確率分布が平行移動不変であるときをいい, エルゴード的であるとは, 定常かつその確率分布に関して平行移動不変 σ 集合体が自明となるときをいう. また, $\tilde{\Phi}$ が等方的であるとは, その確率分布が回転不変となるときをいう. マーク付き点過程の基本事項は, [1], [2] で得られる.

2 主結果

各 $L > 0$ に対して, $\Lambda_L = [-L/2, L/2]^d \times \mathbb{M}$ とおく. $A \in O(d)$ とマーク付き点過程 $\tilde{\Phi}$ から定まるランダムな κ -フィルタ付き複体 $\mathbb{K}(\tilde{\Phi}|_{R_A(\Lambda_L)}) = \{K(\tilde{\Phi}|_{R_A(\Lambda_L)}, t)\}_{t \geq 0}$ に対応する q 次パーシステンス図 $\xi_q(\mathbb{K}(\tilde{\Phi}|_{R_A(\Lambda_L)}))$ を, 簡潔に $\xi_{q,A,L}$ と表す. 主結果は次の定理である.

定理 3. $\tilde{\Phi}$ は定常マーク付き点過程で, そのグラウンド過程 Φ は全有限モーメントをもつと仮定する. このとき非負整数 q と $A \in O(d)$ に対して, Δ 上の Radon 測度 $\nu_{q,A}$ が存在して, $L \rightarrow \infty$ のとき $\mathbb{E}[\xi_{q,A,L}]/L^d \xrightarrow{v} \nu_{q,A}$ が成立する. ここで \xrightarrow{v} は漠収束を表す. さらに $\tilde{\Phi}$ がエルゴード的であれば, ある $A \in O(d)$ が存在して, ほとんど確実に $L \rightarrow \infty$ のとき

$$\frac{1}{L^d} \xi_{q,A,L} \xrightarrow{v} \nu_{q,A}$$

が成立する. とくに $\tilde{\Phi}$ が等方的であれば, 上の概収束はすべての $A \in O(d)$ に対して成立し, 極限分布 $\nu_{q,A}$ は $A \in O(d)$ に依らない.

3 主結果の証明の流れ

定理 3 の証明は, 通常の点過程のパーシステンス図の極限定理について論じている [3] で用いられている手法を, マーク付き点過程版へ拡張して行う. $r \leq s$ に対して, q 次 (r, s) -パーシステント Betti 数は

$$\beta_q^{r,s}(\mathbb{K}(\tilde{\Xi})) = \dim \frac{Z_q(\mathbb{K}(\tilde{\Xi}, r))}{Z_q(\mathbb{K}(\tilde{\Xi}, r)) \cap B_q(\mathbb{K}(\tilde{\Xi}, s))},$$

と定義される. ここで, $Z_q(K(\tilde{\Xi}, r))$ と $B_q(K(\tilde{\Xi}, r))$ はそれぞれ単体複体 $K(\tilde{\Xi}, r)$ の q 次 cycle 群と boundary 群である. この数は線形写像 $\iota_r^s : H_q(K(\tilde{\Xi}, r)) \rightarrow H_q(K(\tilde{\Xi}, s))$ の階数と等しく, 時刻 r から s までの間で持続している q 次ホモロジー類の数を表す. $\xi_q(\mathbb{K}(\tilde{\Xi}))$ の定義により,

$$\beta_q^{r,s}(\mathbb{K}(\tilde{\Xi})) = \sum_{b \leq r, s < d} m_{b,d} = \xi_q(\mathbb{K}(\tilde{\Xi}))([0, r] \times (s, \infty])$$

が成立することがわかる. すなわち, パーシステント Betti 数はパーシステンス図が定める数え上げ測度 $\xi_q(\mathbb{K}(\tilde{\Xi}))$ の分布関数の役割を果たす. Radon 測度の濃収束に関する一般論と合わせると, 定理 3 の証明は次のパーシステント Betti 数に関する大数の法則の証明へと帰着される.

定理 4. $\tilde{\Phi}$ は定常マーク付き点過程で, そのグラウンド過程 Φ は全有限モーメントをもつと仮定する. このとき $r \leq s$ と非負整数 q , そして $A \in O(d)$ に対して, 非負実数 $\hat{\beta}_{q,A}^{r,s}$ が存在して, $L \rightarrow \infty$ のとき

$$\frac{1}{L^d} \mathbb{E}[\beta_q^{r,s}(\mathbb{K}(\tilde{\Phi})_{R_A(\Lambda_L)})] \rightarrow \hat{\beta}_{q,A}^{r,s}$$

が成立する. さらに $\tilde{\Phi}$ がエルゴード的であれば, ある $A \in O(d)$ が存在して, ほとんど確実に $L \rightarrow \infty$ のとき

$$\frac{1}{L^d} \beta_q^{r,s}(\mathbb{K}(\tilde{\Phi})_{R_A(\Lambda_L)}) \rightarrow \hat{\beta}_{q,A}^{r,s}$$

が成立する.

[4] によれば, 可分な確率空間へのエルゴード的 \mathbb{R}^d 作用 $\{T'_a\}_{a \in \mathbb{R}^d}$ が与えられたとき, ある超平面の可算族 $\{P_n\}$ が存在して, 任意の $x \in \mathbb{R}^d \setminus \cup_n P_n$ に対して, T'_x はエルゴード的になることが知られている. 必要ならば回転させたマーク付き点過程を考えればよいから, 定理 4 は A が単位行列で, 概収束に関しては十分多くの $M > 0$ に対して格子の作用 $\{T_a\}_{a \in M\mathbb{Z}^d}$ がエルゴード的である場合に示せば十分であることがわかる. また, 定理 4 の証明において次の補題が重要である.

補題 5. [3, Lemma 2.11] $\mathbb{K}^1 = \{K_t^1\}_{t \geq 0}$ と $\mathbb{K}^2 = \{K_t^2\}_{t \geq 0}$ は各 $t \geq 0$ について $K_t^1 \subset K_t^2$ をみたすフィルター付き複体とする. このとき

$$|\beta_q^{r,s}(\mathbb{K}^1) - \beta_q^{r,s}(\mathbb{K}^2)| \leq \sum_{j=q, q+1} \#K_{s,j}^2 \setminus K_{s,j}^1 + \#\{\sigma \in K_{s,j}^1 \setminus K_{r,j}^1 : t_\sigma^{(2)} \leq r\},$$

が成り立つ. ここで $K_{s,j}^i$ は K_s^i に属す j 単体全体, $t_\sigma^{(i)}$ は \mathbb{K}^i における単体 σ の発生時刻である ($i = 1, 2$).

したがって、2つの異なるマーク付き集合にそれぞれ対応するパーシステント Betti 数の間の差は、2つのマーク付き集合がそれぞれ構成するフィルター付き複体の間の差で制御される。補題 5 と次の単体の数え上げに関する大数の法則により、定理 4 は証明される。

補題 6. 正整数 q と $r \geq 0$, そして $\tilde{\Xi} \in \mathcal{F}(\mathbb{R}^d \times \mathbb{M})$ に対して

$$F_q(\tilde{\Xi}, r) = \#\{\sigma \in K(\tilde{\Xi}, r) : \sigma \text{ は } q \text{ 単体}\}$$

とおく。また、 $\tilde{\Phi}$ は定常マーク付き点過程で、そのグラウンド過程 Φ は全有限モーメントをもつと仮定する。このとき $C_{q,r} \geq 0$ が存在して、任意の有界な \mathbb{R}^d の Borel 集合 A に対して

$$\mathbb{E}[F_q(\tilde{\Phi}|_{A \times \mathbb{M}}, r)] \leq C_{q,r}|A|$$

が成り立つ。ここで $|A|$ は A の Lebesgue 測度である。さらに格子作用 $\{T_a\}_{a \in \mathbb{Z}^d}$ がエルゴード的であれば、ほとんど確実に $L \rightarrow \infty$ のとき

$$\left\{ \frac{F_q(\tilde{\Phi}|_{\Lambda_L}, r)}{L^d} \right\}_{L>0}$$

は収束する。

注意 7. 定理 3 において \mathbb{R}^d の近似列として d 次元立方体 $\{[-L/2, L/2]^d\}_{L>0}$ およびそれを回転させた列を考えた。筆者は、凸集合の増大列 $\{A_L\}_{L>0}$ であって、 $L \rightarrow \infty$ のとき

$$\sup\{r \geq 0 : \text{ある } x \in \mathbb{R}^d \text{ が存在して } B_r(x) \subset A_L\} \rightarrow \infty$$

となる R^d の近似列に対して定理 3 の同様の極限定理が成立し、さらに等方的という仮定なしに極限分布は近似列のとり方に依らないと予想している。この予想の解決に関しては、[5] において述べる予定である。

参考文献

- [1] D. J. Daley, and D. Vere-Jones, An introduction to the theory of point processes. Vol. I. Elementary theory and methods. Second edition. Probability and its Applications (New York), Springer-Verlag, New York, (2003).
- [2] D. J. Daley and D. Vere-Jones, An introduction to the theory of point processes. Vol. II. General theory and structure. Second edition, Springer, New York, (2008).

- [3] Y. Hiraoka, T Shirai, and K. D. Trinh, *Limit theorems for persistence diagrams*, Ann. Appl. Probab. 28 (2018), 2740–2780.
- [4] C. Pugh and M. Shub, *Ergodic elements of ergodic actions*, Compositio Math. 23 (1971), 115–122.
- [5] T. Shirai and K. Suzuki, *A Limit theorem for persistence diagrams of random filtered complexes built over marked point processes*, In preparation.
- [6] A. Zomorodian and G. Carlsson, *Computing persistent homology*, Discrete Comput. Geom. 33 (2005), no.2, 249–274.