ランダム方体複体に対する大偏差原理

角田 謙吉

1. 序

データ解析の分野においてデータの形に着目した数学的解析手法が近年用いられている. 応用上のデータとして得られるのは高次元ユークリッド空間における点データであったり, グレースケール画像 (2 値データ) であったりそのままでは幾何学的な構造を持たない. そのようなデータに対して適切な幾何構造を与えて, データを解析するために必要な特徴量を抽出することにより, データがもつ性質を特定することが可能である. このような解析手法は近年位相的データ解析とよばれ, データ解析の分野において確立した地位を占めている. このような背景の下, データがランダムなノイズを伴う状況を考えることは自然であり, その数学的模型の解析も活発に行われている. 数学的にはランダムグラフの高次元化と考えられるランダム複体を解析することが主な話題であり, この分野はランダムトポロジーとよばれる. 多くの模型においてランダム複体に対する極限定理が得られているが, 基本的な問題ですらまだ分かっていないことも多い. 本稿では節 2 でランダム幾何複体に対する極限定理を幾つか述べた後, 節 3 で [4] の主結果であるランダム方体複体のフィルトレーションに対する大偏差原理について述べる.

2. ランダム幾何複体

$d \in \mathbb{N}$ を固定する. $\Phi = \Phi(\rho) = \{x_i\}_{i\in\mathbb{N}}$ を $\mathbb{R}^d$ 上の密度 ρ の Poisson 点過程とする. つまり, Φ は $\mathbb{R}^d$ 上のランダムな非負整数値 Radon 測度であり次の性質をみたす.

- 任意の有界 Borel 集合 $A \subset \mathbb{R}^d$ に対して $\Phi(A)$ はパラメータ $\rho|A|$ の Poisson 分布に従う. ここで $|A|$ は A の Lebesgue 測度を表す.
- 任意の互いに交わらない有界 Borel 集合 $A_1, \dots, A_n$ に対して $\Phi(A_1), \dots, \Phi(A_n)$ は独立.

$n \in \mathbb{N}$ に対して $\Lambda_n = [-n, n]^d$ とし, Φ_n を Φ の Λ_n 内への制限とする: $\Phi_n = \Phi \cap \Lambda_n$. また中心 $x \in \mathbb{R}^d$, 半径 $r > 0$ の閉球を $B(x, r)$ とかくことにする. いま Φ_n の各点を中心とする半径 r の閉球の和集合を考える:

$$\bigcup_{x\in\Phi_n} B(x, r). \tag{2.1}$$

以下では (2.1) の Betti 数に対する大数の法則について述べる. Betti 数とは (2.1) の “穴の数” を表す位相不変量であり, 正確には Čech 複体のホモロジー群の階数として定義される. ホモロジー群や Betti 数, 後で述べるパーシステントホモロジー群などについては, ランダム幾何複体に対しては [6], ランダム方体複体に対しては [4] を参照して頂きたい.

各 $q = 0, \dots, d-1$ に対して $\beta_q(\cdot)$ で q 次 Betti 数を表すものとして, $\beta_q^{(n)}(r)$ を次で定義する:

$$\beta_q^{(n)}(r) = \beta_q\Big(\bigcup_{x\in\Phi_n} B(x, r)\Big).$$

$\beta_q^{(n)}(r)$ に対する大数の法則は [10] で示された.

Theorem 2.1. ランダムでない定数 $\widehat{\beta}_q^{(d)}(\rho, r)$ が存在して確率 1 で次が成立する*:*

$$\lim_{n\to\infty}\frac{1}{|\Lambda_n|}\beta_q^{(n)}(r)=\widehat{\beta}_q^{(d)}(\rho, r).$$

次にこの結果が多様体の設定でも成立することを述べる. $\mathcal{M}\subset\mathbb{R}^d$ を滑らかな m 次元コンパクト多様体とする. $\{X_i\}_{i\in\mathbb{N}}$ を $\mathcal{M}$-値独立同分布確率変数列でその分布は密度関数 $f:\mathcal{M}\to[0,\infty)$ をもつものとし, 二項点過程を $\mathcal{X}_n$ とかくことにする: $\mathcal{X}_n=\{X_1,\dots,X_n\}$. コンパクト多様体の設定で Čech 複体の半径を r に固定してしまうと, n が大きいとき高確率で多様体を覆ってしまうため, 適切な半径の列 $r_n\downarrow 0$ をとる必要がある. そのために $r>0$ を固定して, $r_n=rn^{-1/m}$ とし, $\beta_q^{(n)}$ を次で定義する:

$$\beta_q^{(n)}=\beta_q\Big(\bigcup_{x\in\mathcal{X}_n}B(x,r_n)\Big).$$

$\beta_q^{(n)}$ に対する大数の法則は [3] で示された.

Theorem 2.2. 任意の $p\in[1,\infty)$ に対して f は p 次可積分であると仮定する. このとき任意の $q=0,\dots,m-1$ に対して次が確率 1 で成立する*:*

$$\lim_{n\to\infty}\frac{1}{n}\beta_q^{(n)}=\int_{\mathcal{M}}\widehat{\beta}_q^{(m)}(f(x),r)dx.$$

右辺の被積分関数に現れる $\widehat{\beta}_q^{(m)}(\cdot, r)$ は Theorem 2.1 で与えられるものであることを注意しておく. また Theorem 2.2 は既存の結果を大きく改善していることも注意しておく. 実際, $\mathcal{M}=\mathbb{R}^d$ の場合では f に幾つかの仮定を課した下で大数の法則が [10] で示されている. またコンパクト多様体の設定 [1] では $\beta_q^{(n)}$ の平均値がオーダー n であることのみしか示されていなかった. そのため [3] の主結果は, [10] における仮定を緩めた上で, コンパクト多様体の設定で大数の法則を示したものとなっている.

再びユークリッド空間上の Poisson 点過程から得られる Čech 複体について考える. Čech 複体の増大列 $\mathbb{C}_n=\{\mathcal{C}(\Phi_n,r)\}_{r\geq 0}$ から得られるパーシステント図を $D_q(\mathbb{C}_n)=\{(b_i,d_i): 0\leq b_i<d_i\leq\infty, i=1,\dots,m_q\}$ とかくことにする. ただし $D_q(\mathbb{C}_n)$ は多重集合としている. ここで $\Delta=\{(b,d)\in[0,\infty]^2: 0\leq b<d\leq\infty\}$ 上のランダム測度を

$$\xi_q^{(n)}=\sum_{(b,d)\in D_q(\mathbb{C}_n)}\delta_{(b,d)},$$

により定める. ただし $\delta_{(b,d)}$ は $(b,d)\in\Delta$ 上の Dirac 測度である. また Δ 上の Radon 測度全体を $\mathcal{M}(\Delta)$ とかくことにし, $\mathcal{M}(\Delta)$ には漠収束による位相を導入しておく. ここまでで $\mathcal{M}(\Delta)$-値確率変数の列 $\xi_q^{(n)}$ が構成された.

パーシステントホモロジー群はホモロジー群全てを並べたものであるため, Betti 数のレベルをあげたものとして捉えることができる. 実際, 任意の $r>0$ に対して $\xi_q^{(n)}([0,r]\times(r,\infty])=\beta_q^{(n)}(r)$ が成立する. ランダム測度 $\xi_q^{(n)}$ に対する大数の法則は [6] で示された.

Theorem 2.3. ランダムでない測度 $\widehat{\nu}_q(\rho)\in\mathcal{M}(\Delta)$ が存在して確率 1 で次が成立する*:*

$$\frac{1}{|\Lambda_n|}\xi_q^{(n)}\xrightarrow{v}\widehat{\nu}_q(\rho).$$

ただし $\xrightarrow{v}$ は漠収束についての収束を表す.

Theorem 2.3 では説明の簡単のため Poisson 点過程に付随する Čech 複体のフィルトレーションについて主張を述べたが, [6] では点過程と単体複体のフィルトレーションのどちらも一般的なものを扱っていることを注意しておく.

3. ランダム方体複体

ここからはランダム方体複体に対する極限定理を述べる. 幾つかの記法は前節と同じものを用いるが, 基本的に同じ対象を扱っているのでご容赦頂きたい. はじめにランダム方体複体を導入する. 再び $d \in \mathbb{N}$ を固定する. $\mathbb{R}^d$ の部分集合 Q は $Q = I_1 \times \cdots \times I_d$ の形にかけるとき基本方体であると言われる. ただし $I_1, \ldots, I_d$ は $a_i \in \mathbb{Z}$ を用いて $I_i = [a_i, a_i + 1]$ もしくは $I_i = \{a_i\}$ とかかれるものである. また基本方体 $Q \subset \mathbb{R}^d$ に対して Q の非退化な成分の個数を Q の次元とよび, $\dim Q$ とかくことにする:

$$\dim Q = \#\{1 \le i \le d : |I_i| = 1\}.$$

$\mathbb{R}^d$ の部分集合 X は高々可算個の基本方体の合併集合にかけるとき方体集合であるといわれる. 方体集合が与えられると, 境界作用素を導入して複体の構造が自然に導入されるので, ホモロジー群やその階数である Betti 数が定義される. 詳しい定義については再び省略するので, 詳しくは [8] を参照して頂きたい.

$q = 0, \ldots, d$ に対して, $\mathcal{K}_q^d$ を $\mathbb{R}^d$ 内の q 次元基本方体全体とし, $\mathcal{K}^d = \bigcup_{q=0}^d \mathcal{K}_q^d$ とおく. また各 $Q \in \mathcal{K}^d$ に対して確率変数 $t_Q \in [0,1]$ が与えられたとし, ランダムな方体集合 $X(t)$ を

$$X(t) = \bigcup \{Q \in \mathcal{K}^d : t_Q \le t\}, \quad t \in [0,1],$$

により定義する. 確率変数 $\{t_Q\}_{Q \in \mathcal{K}^d}$ の法則に対して次を仮定しておく:

- 各 $z \in \mathbb{Z}^d$ に対して $\{t_Q\}_{Q \in \mathcal{K}^d}$ と $\{t_{z+Q}\}_{Q \in \mathcal{K}^d}$ は同分布.
- ある $R > 0$ が存在して, $A, B \subset \mathbb{R}^d$ が $d(A,B) > R$ となるとき $\{t_Q : Q \in \mathcal{K}^d, Q \subset A\}$ と $\{t_Q : Q \in \mathcal{K}^d, Q \subset B\}$ は独立. ただし $d(A,B)$ は A と B の集合間の距離である.

確率変数 $\{t_Q\}_{Q \in \mathcal{K}^d}$ の取り方によってはボンドパーコレーションを考えることができる. 実際, $\{t_Q\}_{Q \in \mathcal{K}^d}$ は独立であるとし, $\dim Q = 0$ のとき $t_Q \equiv 0$, $\dim Q = 1$ のとき t_Q の分布は $[0,1]$ 上の一様分布, $\dim Q \ge 2$ のとき $t_Q \equiv 1$ とすると, $X(t)$ は辺の発生確率が t である $\mathbb{Z}^d$ 上のボンドパーコレーションに他ならない.

節 2 と同様にして Betti 数とパーシステントホモロジーを考えていく. ランダムな方体複体 $\mathbb{X}(t) = \{X(t)\}_{t \in [0,1]}$ に対して次を定義する:

$$X_n(t) = X(t) \cap \Lambda_n, \quad \mathbb{X}_n = \{X_n(t)\}_{t \in [0,1]}, \quad \xi_q^{(n)} = \xi_q(\mathbb{X}_n),$$
$$\beta_q^{(n)}(s,t) = \xi_q^{(n)}([0,s] \times (t, \infty]), \quad 0 \le s < t \le 1.$$

$\beta_q^{(n)}(s,t)$ はパーシステント Betti 数とよばれる. $\beta_q^{(n)}(t,t)$ は q 次 Betti 数 $\beta_q(X_n(t))$ と一致することを注意しておく.

$\beta_q(X_n(t))$ に対する大数の法則は [7] で示された. [6] の手法を援用することにより, $\beta_q^{(n)}(s,t), \xi_q^{(n)}$ に対する大数の法則を [4] において示した.

Theorem 3.1. *ランダムでない定数 $\widehat{\beta}_q(s,t)$ が存在して次が確率 1 で成立する:*

$$\lim_{n \to \infty} \frac{1}{|\Lambda_n|} \beta_q^{(n)}(s,t) = \widehat{\beta}_q(s,t).$$

またランダムでない測度 $\widehat{\nu}_q(\rho) \in \mathcal{M}(\Delta)$ が存在して次が確率 1 で成立する:

$$\frac{1}{|\Lambda_n|} \xi_q^{(n)} \xrightarrow{v} \widehat{\nu}_q(\rho).$$

[4] の主結果は Theorem 3.1 に対応する大偏差原理である. その結果を述べる前にレート関数を定義しておく. 一言で言えば極限対数モーメント母関数の Fenchel–Legendre 変

換である. $\{\beta_q^{(n)}(s,t)/|\Lambda_n|\}_{n\in\mathbb{N}}$ に対するレート関数を次で定義する: $0 \le s < t \le 1$ に対して

$$\varphi_{q,s,t}(\lambda) = \lim_{n\to\infty} |\Lambda_n|^{-1} \log \mathbb{E}\left[\exp(\lambda \beta_q^{(n)}(s,t))\right], \quad \lambda \in \mathbb{R}.$$
$$\varphi_{q,s,t}^*(x) = \sup_{\lambda\in\mathbb{R}} \{\lambda x - \varphi_{q,s,t}(\lambda)\}, \quad x \in \mathbb{R}.$$

また $\{\xi_q^{(n)}/|\Lambda_n|\}_{n\in\mathbb{N}}$ に対するレート関数を次で定義する:

$$\phi_q(f) = \lim_{n\to\infty} |\Lambda_n|^{-1} \log \mathbb{E}\left[\exp\left(\int_\Delta f\, d\xi_q^{(n)}\right)\right], \quad f \in C_c(\Delta).$$
$$\phi_q^*(\xi) = \sup_{f\in C_c(\Delta)} \left\{\int_\Delta f\, d\xi - \phi_q(f)\right\}, \quad \xi \in \mathcal{M}(\Delta).$$

次の大偏差原理が [4] の主結果である.

Theorem 3.2. $\mathbb{R}$-値確率変数列 $\{\beta_q^{(n)}(s,t)/|\Lambda_n|\}_{n\in\mathbb{N}}$ は $|\Lambda_n|$ をスピード, $\varphi_{q,s,t}^*$ をレート関数として大偏差原理をみたす. つまり, 任意の閉集合 $C \subset \mathbb{R}$ に対して

$$\limsup_{n\to\infty} \frac{1}{|\Lambda_n|} \log \mathbb{P}\left(\frac{\beta_q^{(n)}(s,t)}{|\Lambda_n|} \in C\right) \le -\inf_{x\in C} \varphi_{q,s,t}^*(x),$$

任意の開集合 $O \subset \mathbb{R}$ に対して

$$\liminf_{n\to\infty} \frac{1}{|\Lambda_n|} \log \mathbb{P}\left(\frac{\beta_q^{(n)}(s,t)}{|\Lambda_n|} \in O\right) \ge -\inf_{x\in O} \varphi_{q,s,t}^*(x),$$

が成立する. また $\mathcal{M}(\Delta)$-値確率変数列 $\{\xi_q^{(n)}/|\Lambda_n|\}_{n\in\mathbb{N}}$ は $|\Lambda_n|$ をスピード, ϕ_q^* をレート関数として大偏差原理をみたす. つまり, 任意の閉集合 $\mathcal{C} \subset \mathcal{M}(\Delta)$ に対して

$$\limsup_{n\to\infty} \frac{1}{|\Lambda_n|} \log \mathbb{P}\left(\frac{\xi_q^{(n)}}{|\Lambda_n|} \in \mathcal{C}\right) \le -\inf_{\xi\in\mathcal{C}} \phi_q^*(\xi),$$

任意の開集合 $\mathcal{O} \subset \mathcal{M}(\Delta)$ に対して

$$\liminf_{n\to\infty} \frac{1}{|\Lambda_n|} \log \mathbb{P}\left(\frac{\xi_q^{(n)}}{|\Lambda_n|} \in \mathcal{O}\right) \ge -\inf_{\xi\in\mathcal{O}} \phi_q^*(\xi),$$

が成立する.

我々の知る限り, パーシステントホモロジーに対する大偏差原理は我々の結果が初めてである. ランダムな複体から得られるパーシステント Betti 数についての大偏差原理は例えば [5] で研究されている. しかしながら技術的困難さから [5] では $d = 2$ の場合かつ単体複体も特殊なもののみを扱っており, ランダム幾何複体の場合に一般の d, q に対する Betti 数の大偏差原理は未解決である.

最後に証明の概略を簡単に述べる. 証明は幾つかの段階に分けられるが, 大きな方針として次の二つを示すことに問題は帰着される:

(1) $\mathcal{M}(\Delta)$-値確率変数列 $\{\xi_q^{(n)}/|\Lambda_n|\}_{n\in\mathbb{N}}$ はあるレート関数に対して大偏差原理をみたす.
(2) そのレート関数は ϕ_q^* である.

(1) を示すために超指数的な評価や縮約原理などの大偏差原理の理論の一般的な手法を用いる. (1) の段階ではレート関数が ϕ_q^* であることは分からないが, その各ステップで凸性が保存されることに注目すると, 実はそのレート関数が ϕ_q^* であることがわかる. これによ

り (2) が分かるので, Theorem 3.2 の証明が完了する. また Theorem 3.2 の前半の主張は (1) を示す途中で示される.

次に (1) を示すための概略を述べる. 初めに Riesz–Markov–Kakutani の定理から $\mathcal{M}(\Delta)$ は正値線形汎関数と位相を込めて同一視される. 具体的には以下の対応を考える:

$$\mathcal{M}(\Delta) \simeq \{L : L \text{ は } C_c(\Delta) \text{ 上の正値線形汎函数}\},$$
$$\xi \in \mathcal{M}(\Delta) \mapsto L_\xi : L_\xi f = \int_\Delta f\, d\xi, \quad f \in C_c(\Delta).$$

この対応があるので, 問題は $\{L_{\xi_q^{(n)}}/|\Lambda_n|\}_{n\in\mathbb{N}}$ に対して大偏差原理を示すことに帰着される.

また Dawson–Gärtner の定理 (cf. [2, Theorem 4.6.9]) から, $\{L_{\xi_q^{(n)}}/|\Lambda_n|\}_{n\in\mathbb{N}}$ に対して大偏差原理を示すことは, 有限次元の場合に示すことに帰着される. 具体的には, 任意の $f_1, \dots, f_m \in C_c(\Delta)$ に対して $\mathbb{R}^m$-値確率変数列

$$\left\{\frac{1}{|\Lambda_n|}\left(L_{\xi_q^{(n)}} f_1, \dots, L_{\xi_q^{(n)}} f_m\right) = \frac{1}{|\Lambda_n|}\left(\xi_q^{(n)} f_1, \dots, \xi_q^{(n)} f_m\right)\right\}_{n\in\mathbb{N}},$$

に対する大偏差原理を示せばよい.

一旦 Δ 内の矩形領域を $I = (s_1, s_2] \times (t_1, t_2]$ として, $L_{\xi_q^{(n)}} \mathbf{1}_I = \xi_q^{(n)}(I)$ を考えてみる. 包除原理から $L_{\xi_q^{(n)}} \mathbf{1}_I$ はパーシステント Betti 数の交代和としてかける:

$$L_{\xi_q^{(n)}} \mathbf{1}_I = \beta_q^{(n)}(s_2, t_1) - \beta_q^{(n)}(s_2, t_2) + \beta_q^{(n)}(s_1, t_2) - \beta_q^{(n)}(s_1, t_1).$$

なので $\mathbb{R}^4$-値確率変数

$$\left(\beta_q^{(n)}(s_2, t_1), \beta_q^{(n)}(s_2, t_2), \beta_q^{(n)}(s_1, t_2), \beta_q^{(n)}(s_1, t_1)\right),$$

に対して大偏差原理を示せば, 縮約原理から $L_{\xi_q^{(n)}} \mathbf{1}_I$ に対して大偏差原理が示される.

一般の $f \in C_c(\Delta)$ は Δ 内の矩形領域 I の指示関数 $\mathbf{1}_I$ の線形和で sup ノルムについて近似することが出来ることに注意して, 次の集合を導入する. 各 $l \in \mathbb{N}$ に対して $\mathcal{I}_l$ を次のどちらかの形の矩形領域全体とする:

$$I = \left[0, \frac{1}{2^{l+1}}\right] \times \left(\frac{j-1}{2^{l+1}}, \frac{j}{2^{l+1}}\right] \text{ for } j \in \mathbb{N} \text{ with } 3 \le j \le l \cdot 2^{l+1},$$
$$I = \left(\frac{i-1}{2^{l+1}}, \frac{i}{2^{l+1}}\right] \times \left(\frac{j-1}{2^{l+1}}, \frac{j}{2^{l+1}}\right] \text{ for } (i,j) \in \mathbb{N}^2 \text{ with } 2 \le i \le j \le l \cdot 2^{l+1} \text{ and } j - i \ge 2.$$

ここで $\mathbb{R}^{\mathcal{I}_l}$-値確率変数を $\mathrm{HIST}_l(\xi_q^{(n)}) = (\xi_q^{(n)}(I))_{I\in\mathcal{I}_l}$ で定義しておく. 一つ前の段落の議論と先にした注意から, $\mathrm{HIST}_l(\xi_q^{(n)})$ に対して大偏差原理を示せれば, 確率変数列

$$\left\{\frac{1}{|\Lambda_n|}\left(L_{\xi_q^{(n)}} f_1, \dots, L_{\xi_q^{(n)}} f_m\right) = \frac{1}{|\Lambda_n|}\left(\xi_q^{(n)} f_1, \dots, \xi_q^{(n)} f_m\right)\right\}_{n\in\mathbb{N}},$$

に対する大偏差原理が示される. このために超指数的評価と縮約原理を用いる.

最後に $\mathrm{HIST}_l(\xi_q^{(n)})$ に対して大偏差原理を示す必要がある. 先と同様の理由から $\mathrm{HIST}_l(\xi_q^{(n)})$ に対する大偏差原理はパーシステント Betti 数に対して大偏差原理に示すことに帰着される. 詳細は省くが, パーシステント Betti 数に対する大偏差原理を示すために Yukich–Seppäläinen の結果 [9] を我々の設定に一般化する必要がある. これらを全てまとめることで Theorem 3.2 の証明が完了する.

References

[1] O. Bobrowski and S. Mukherjee, *The topology of probability distributions on manifolds*, Probab. Theory Related Fields **161** (2015), 651–686.
[2] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Applications of Mathematics **38**, Springer.
[3] A. Goel, K.D. Trinh and K. Tsunoda, *Strong law of large numbers for Betti numbers in the thermodynamic regime*, J. Stat. Phys. **174** (2019), 865–892.
[4] Y. Hiraoka, S. Kanazawa, J. Miyanaga and K. Tsunoda, *Large deviation principle for persistence diagrams of random cubical filtrations*, arXiv:2210.12469.
[5] C. Hirsch and T. Owada, *Large deviation principle for geometric and topological functionals and associated point processes*, arXiv:2201.07276.
[6] Y. Hiraoka, T. Shirai and K.D. Trinh, *Limit theorems for persistence diagrams*, Ann. Appl. Probab. **28** (2018), 2740–2780.
[7] Y. Hiraoka and K. Tsunoda, *Limit theorems for random cubical homology*, Discrete Comput. Geom. **60** (2018), 665–687.
[8] T. Kaczynski, K. Mischaikow and M. Mrozek, *Computational homology*, Applied Mathematical Sciences **157**, Springer.
[9] T. Seppäläinen and J.E. Yukich, *Large deviation principles for Euclidean functionals and other nearly additive processes*, Probab. Theory Related Fields **120** (2001), 309–345.
[10] D. Yogeshwaran, E. Subag and R. Adler, *Random geometric complexes in the thermodynamic regime*, Probab. Theory Related Fields **167** (2017), 107–142.

(角田 謙吉) 九州大学大学院数理学研究院, 819-0395, 福岡市西区元岡 744
E-mail address: tsunoda@math.kyushu-u.ac.jp