階層的クラスタリングの高次元漸近的性質について

筑波大学・理工情報生命学術院・数理物質科学研究群　　江頭　健斗（Kento Egashira）
Degree Programs in Pure and Applied Sciences
Graduate School of Science and Technology
University of Tsukuba
筑波大学・数理物質系　　矢田　和善（Kazuyoshi Yata）
Institute of Mathematics
University of Tsukuba
筑波大学・数理物質系　　青嶋　誠（Makoto Aoshima）
Institute of Mathematics
University of Tsukuba

1　はじめに

本論文では，高次元小標本データに対する凝集型階層的クラスタリングを考える．独立な d 次元の母集団が 2 個あると考える．母集団 π_X と π_Y は平均に未知の d 次ベクトル $\boldsymbol{\mu}_X$，$\boldsymbol{\mu}_Y$，共分散行列に未知の d 次正定値対称行列 $\boldsymbol{\Sigma}_X$，$\boldsymbol{\Sigma}_Y$ をもつと仮定する．母集団 π_X から n 個のデータ $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$，母集団 π_Y から m 個のデータ $\boldsymbol{y}_1,\dots,\boldsymbol{y}_m$ を無作為に抽出する．

$$\boldsymbol{X}=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\},\quad \boldsymbol{Y}=\{\boldsymbol{y}_1,\cdots,\boldsymbol{y}_m\}$$

とおく．

高次元データに対するクラスタリングについて，Liu et al. (2008) は，“statistical significance of clustering(SigClust)” と呼ばれる 2 分割タイプのクラスタリングを提案した．Ahn et al. (2012) は，高次元における分割型の階層的クラスタリングを提案し，その漸近的性質を求めた．Huang et al. (2015) は，Liu et al. (2008) による SigClust を，ソフト閾値法によって発展させた．Yata and Aoshima (2010, 2020) は，高次元混合分布における幾何学的一致性を示し，それをクラスタリングに応用した．幾何学的一致性については，青嶋・矢田 (2019) も参照のこと．Nakayama et al. (2021) は，高次元データに対するカーネル主成分分析を用いたクラスタリングの漸近的性質を導出した．Borysov et al. (2014)

は，階層的クラスタリングの漸近的振舞いを定式化し，その性質を研究した．しかしながら，Borysov et al. (2014) では，データに正規分布を仮定し，さらに，共分散行列に単位行列のスカラー倍を仮定するなど，高次元データに対して仮定が厳しいものであった．本論文では，高次元小標本 ($d \to \infty$，n と m は固定) の枠組みにおける階層的クラスタリングの漸近的振舞いを緩い条件のもとで導出する．2 節では，単連結法による階層的クラスタリングを説明する．3 節では，階層的クラスタリングの高次元における振舞いを導入し，単連結法の高次元漸近的性質を導出する．4 節では，数値シミュレーションにより，高次元小標本データに対する単連結法の振舞いを検証する．

2　階層的クラスタリング

本節では，階層的クラスタリングを紹介する．階層的クラスタリングは，主に凝集型と分割型に分類することができる．凝集型階層的クラスタリングでは，与えられたデータをひとつひとつのクラスターとみなし，その中で，クラスター間の距離が一番近いものを結合していく．最終的に，すべてのデータは結合され，1 つのクラスターが形成される．分割型階層的クラスタリングでは，与えられた全てのデータを 1 つのクラスターとみなし，その中で，クラスター間の距離が一番離れるようにデータを 2 分割していく．最終的に，すべてのデータは分割され，データ数分のクラスターが形成される．通常，分割型階層的クラスタリングは計算量が莫大になる．本論文では，凝集型階層的クラスタリングに着目する．以降，階層的クラスタリングは，凝集型階層的クラスタリングを意味するとする．

次に階層的クラスタリングを実行するうえで必要な連結関数を導入する．連結関数は，クラスター間の類似度，または，非類似度を測る関数である．本論文では，以下のようにユークリッド距離による単連結法を考える．

単連結法:

$$D(\boldsymbol{X}, \boldsymbol{Y}) = \min_{\boldsymbol{x} \in \boldsymbol{X}, \boldsymbol{y} \in \boldsymbol{Y}} d(\boldsymbol{x}, \boldsymbol{y}).$$

ただし，$d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$ とし，$\|\cdot\|$ はユークリッドノルムとする．

単連結法では，クラスター間の中で最も小さい距離を，クラスター間の距離として定めている．

3 階層的クラスタリングの高次元漸近的性質

$Z=X,Y$ に対して，以下を仮定として考える．

$$\limsup_{d\to\infty}\frac{\|\boldsymbol{\mu}_Z\|^2}{d}<\infty,\quad \liminf_{d\to\infty}\frac{\mathrm{tr}(\boldsymbol{\Sigma}_Z)}{d}>0,\quad \limsup_{d\to\infty}\frac{\mathrm{tr}(\boldsymbol{\Sigma}_Z)}{d}<\infty.$$

また, $\mathrm{tr}(\boldsymbol{\Sigma}_{max}^2)=\max\{\mathrm{tr}(\boldsymbol{\Sigma}_X^2),\mathrm{tr}(\boldsymbol{\Sigma}_Y^2)\}$, $K_X=\mathrm{Var}[\|\boldsymbol{x}_i-\boldsymbol{\mu}_X\|^2]$, $K_Y=\mathrm{Var}[\|\boldsymbol{y}_i-\boldsymbol{\mu}_Y\|^2]$ とおき，以下を仮定する．

(A-i): $\displaystyle \liminf_{d\to\infty}\max\left\{\frac{\|\boldsymbol{\mu}_X-\boldsymbol{\mu}_Y\|^2}{d},\ \frac{|\mathrm{tr}(\boldsymbol{\Sigma}_X)-\mathrm{tr}(\boldsymbol{\Sigma}_Y)|}{d}\right\}>0;$

(A-ii): $\mathrm{tr}(\boldsymbol{\Sigma}_{max}^2)/d^2\to 0\ \text{ as } d\to\infty;$

(A-iii): $K_X=O\{\mathrm{tr}(\boldsymbol{\Sigma}_X^2)\}$ and $K_Y=O\{\mathrm{tr}(\boldsymbol{\Sigma}_Y^2)\}\ \text{ as } d\to\infty.$

母集団に正規分布を仮定したとき，$\mathrm{Var}(\|\boldsymbol{x}_i-\boldsymbol{\mu}_X\|^2)=2\mathrm{tr}(\boldsymbol{\Sigma}_X^2)$，$\mathrm{Var}(\|\boldsymbol{y}_i-\boldsymbol{\mu}_Y\|^2)=2\mathrm{tr}(\boldsymbol{\Sigma}_Y^2)$ が成立し，(A-iii) が成り立つことに注意する．

次に，高次元における階層的クラスタリングの漸近的振舞いを導入する．

(A): 一方の母集団からの全てのデータが，同じ母集団同士で結合し 1 つのクラスターを形成する．もう一方の母集団からのデータも同様に同じ母集団同士で結合し 1 つのクラスターを形成する．最後のステップで，その 2 つのクラスターが結合する．

(B): 一方の母集団からの全てのデータが，同じ母集団同士で結合し 1 つのクラスターを形成する．その後，もう一方の母集団からのデータが 1 つずつ，吸収されていくように，既に形成されているクラスターに結合していく．

振舞い（A）は，各々の母集団からの任意のデータ間の距離が，異なる母集団から取られてるデータ間よりも小さいときに起こる．振舞い（B）は，一方の母集団からの任意のデータ間の距離が十分に小さく，もう一方の母集団からの任意のデータ間の距離が，異なる母集団データ間よりも大きいときに起こる．

定理 1. *(A-i)* から *(A-iii)* を仮定する．

(1) $\displaystyle \liminf_{d\to\infty}\frac{\|\boldsymbol{\mu}_X-\boldsymbol{\mu}_Y\|^2}{|\mathrm{tr}(\boldsymbol{\Sigma}_X)-\mathrm{tr}(\boldsymbol{\Sigma}_Y)|}>1$ のとき，単連結法を用いて階層的クラスタリングを行った場合，振舞い（A）が起きる確率が $d\to\infty$ のもとで 1 に収束する．

(2) $\displaystyle \limsup_{d\to\infty}\frac{\|\boldsymbol{\mu}_X-\boldsymbol{\mu}_Y\|^2}{|\mathrm{tr}(\boldsymbol{\Sigma}_X)-\mathrm{tr}(\boldsymbol{\Sigma}_Y)|}<1$ のとき，単連結法を用いて階層的クラスタリングを

行った場合，振舞い（B）が起きる確率が $d \to \infty$ のもとで 1 に収束する．

注意 1． 単連結法を用いた階層的クラスタリングについて，*Borysov et al. (2014)* も定理 *1* と同様な結果を導出しているが，データに正規分布を仮定し，さらに，共分散行列に単位行列のスカラー倍を仮定するなど，高次元データに対して仮定が厳しいものであった．

4　数値シミュレーション

本節では，高次元小標本のもとで，単連結法を用いたときの階層的クラスタリングを数値的に検証する．母集団 π_X と π_Y について，次の分布を考える．
(i) π_X： $N_d(\mathbf{0}_d, \mathbf{\Sigma}_X)$，$\pi_Y$： $N_d(\mathbf{0}_d, \mathbf{\Sigma}_Y)$;
(ii) $\boldsymbol{x}_j - \boldsymbol{\mu}_X$ は自由度 5 で共分散行列 $\mathbf{\Sigma}_X$ の d 次元 t 分布，$\boldsymbol{y}_j - \boldsymbol{\mu}_Y$ は自由度 5 で共分散行列 $\mathbf{\Sigma}_Y$ の d 次元 t 分布にそれぞれ従う．

ここで，すべての成分が 1 である d 次元ベクトルを $\mathbf{1}_d$，最初の $\lceil d^{2/3} \rceil$ 個の成分が 1, それ以外が 0 である d 次元ベクトルを $\mathbf{1}_{2/3} = (1, ..., 1, 0, ..., 0)^T$ とする．ただし，$\lceil x \rceil$ は x 以上の最小の整数を表す．次の 4 つの場合を考える．
(I)　(i)，$\boldsymbol{\mu}_X = \mathbf{1}_d$，$\boldsymbol{\mu}_Y = \mathbf{0}_d$，$\mathbf{\Sigma}_X = \mathbf{\Phi}$, $\mathbf{\Sigma}_Y = 1.5\mathbf{\Phi}$, $(n, m) = (7, 7)$;
(II)　(i)，$\boldsymbol{\mu}_X = \mathbf{1}_{2/3}$，$\boldsymbol{\mu}_Y = \mathbf{0}_d$，$\mathbf{\Sigma}_X = \mathbf{\Phi}$, $\mathbf{\Sigma}_Y = 1.5\mathbf{\Phi}$, $(n, m) = (7, 7)$;
(III)　(ii)，$\boldsymbol{\mu}_X = \mathbf{1}_d$，$\boldsymbol{\mu}_Y = \mathbf{0}_d$，$\mathbf{\Sigma}_X = \mathbf{\Phi}$, $\mathbf{\Sigma}_Y = 1.5\mathbf{\Phi}$, $(n, m) = (7, 7)$;
(IV)　(ii)，$\boldsymbol{\mu}_X = \mathbf{1}_{2/3}$，$\boldsymbol{\mu}_Y = \mathbf{0}_d$，$\mathbf{\Sigma}_X = \mathbf{\Phi}$, $\mathbf{\Sigma}_Y = 1.5\mathbf{\Phi}$, $(n, m) = (7, 7)$.
ただし，$\mathbf{\Phi} = \boldsymbol{B}(0.3^{|i-j|^{1/3}})\boldsymbol{B}$，$\boldsymbol{B} = \mathrm{diag}(\{0.5 + 1/(d+1)\}^{1/2}, ..., \{0.5 + d/(d+1)\}^{1/2})$ とする．ここで，$\mathrm{tr}(\mathbf{\Phi}) = d$ となる．次元を $d = 2^s$，$s = 4, ..., 10$ と設定する．各設定のもとでデータを発生させ，単連結法を用いて樹形図を作成した．その樹形図が，振舞い (A), (B)，その他，の 3 つのどの状況にあるかを確認した．実験を 2000 回繰り返し，その振舞いの個数を纏めたものが表 1 である．

定理 1 より，(I) と (III) の場合では振舞い（A），(II) と (IV) の場合では振舞い（B）となるように設定されている．実際に，(a) の場合には振舞い（A）に収束し，(b) の場合には振舞い（B）に収束することを確認することができた．

表 1　:

(I)	d	単連結法		
		(A)	(B)	その他
	16	97	13	1890
	32	251	23	1726
	64	528	21	1451
	128	1127	12	861
	256	1742	1	257
	512	1988	0	12
	1024	2000	0	0

(II)	d	単連結法		
		(A)	(B)	その他
	16	3	17	1980
	32	4	75	1921
	64	3	244	1753
	128	1	650	1349
	256	0	1261	739
	512	0	1795	205
	1024	0	1977	23

(III)	d	単連結法		
		(A)	(B)	その他
	16	93	16	1891
	32	215	19	1766
	64	462	23	1515
	128	968	9	1023
	256	1532	4	464
	512	1901	0	99
	1024	1988	0	12

(IV)	d	単連結法		
		(A)	(B)	その他
	16	4	18	1978
	32	2	60	1938
	64	1	199	1800
	128	0	480	1520
	256	0	1007	993
	512	0	1596	404
	1024	0	1928	72

5　付録

一般性を失うことなく $\mathrm{tr}(\boldsymbol{\Sigma}_X) \le \mathrm{tr}(\boldsymbol{\Sigma}_Y)$ と仮定できる．ここで, $b_1 = 2\mathrm{tr}(\boldsymbol{\Sigma}_Y) + \{\|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|^2 - |\mathrm{tr}(\boldsymbol{\Sigma}_X) - \mathrm{tr}(\boldsymbol{\Sigma}_Y)|\}/2$，$b_2 = 2\mathrm{tr}(\boldsymbol{\Sigma}_X) + \{\|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|^2 + |\mathrm{tr}(\boldsymbol{\Sigma}_X) - \mathrm{tr}(\boldsymbol{\Sigma}_Y)|\}/2$，$\Delta_* = \|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|^2 + \mathrm{tr}(\boldsymbol{\Sigma}_X) + \mathrm{tr}(\boldsymbol{\Sigma}_Y)$ とおく．$\Delta_* = \|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|^2 - |\mathrm{tr}(\boldsymbol{\Sigma}_X) - \mathrm{tr}(\boldsymbol{\Sigma}_Y)| + 2\mathrm{tr}(\boldsymbol{\Sigma}_Y) = \|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|^2 + |\mathrm{tr}(\boldsymbol{\Sigma}_X) - \mathrm{tr}(\boldsymbol{\Sigma}_Y)| + 2\mathrm{tr}(\boldsymbol{\Sigma}_X)$ となることに注意し，(A-i) と

$$\liminf_{d\to\infty} \frac{\|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|^2}{|\mathrm{tr}(\boldsymbol{\Sigma}_X) - \mathrm{tr}(\boldsymbol{\Sigma}_Y)|} > 1$$

のもとで,

$$\liminf_{d\to\infty}\{b_1 - 2\mathrm{tr}(\boldsymbol{\Sigma}_Y)\}/d > 0,\ \liminf_{d\to\infty}\{\Delta_* - b_1\}/d > 0 \tag{1}$$

となり，(A-i) と

$$\limsup_{d\to\infty} \frac{\|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|^2}{|\mathrm{tr}(\boldsymbol{\Sigma}_X) - \mathrm{tr}(\boldsymbol{\Sigma}_Y)|} < 1$$

のもとで，

$$\liminf_{d\to\infty}\{b_2-2\mathrm{tr}(\boldsymbol{\Sigma}_X)\}/d>0,\quad \liminf_{d\to\infty}\{\Delta_*-b_2\}/d>0,$$
$$\liminf_{d\to\infty}\{b_1-\Delta_*\}/d>0,\quad \liminf_{d\to\infty}\{2\mathrm{tr}(\boldsymbol{\Sigma}_Y)-b_1\}/d>0 \tag{2}$$

となる．

定理 *1 (1)* の証明．(A-i) から (A-iii) と

$$\liminf_{d\to\infty}\frac{\|\boldsymbol{\mu}_X-\boldsymbol{\mu}_Y\|^2}{|\mathrm{tr}(\boldsymbol{\Sigma}_X)-\mathrm{tr}(\boldsymbol{\Sigma}_Y)|}>1$$

を仮定する．以下の事象を定義する．

$$\widetilde{B}=\{\max_{i,j}\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2<\min_{i,j}\|\boldsymbol{x}_i-\boldsymbol{y}_j\|^2\},$$
$$\widetilde{C}=\{\max_{i,j}\|\boldsymbol{y}_i-\boldsymbol{y}_j\|^2<\min_{i,j}\|\boldsymbol{x}_i-\boldsymbol{y}_j\|^2\},$$

$$\widetilde{E}_1=\{\max_{i,j}\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2<b_1\},$$
$$\widetilde{E}_2=\{\max_{i,j}\|\boldsymbol{y}_i-\boldsymbol{y}_j\|^2<b_1\},$$
$$\widetilde{E}_3=\{\min_{i,j}\|\boldsymbol{x}_i-\boldsymbol{y}_j\|^2>b_1\}.$$

このとき，共通部分 $\widetilde{B}\cap\widetilde{C}$ を考える．単連結法が用いられたとき，$\widetilde{B}\cap\widetilde{C}\subset(A)$ が成立する．また，包括関係 $\widetilde{E}_1\cap\widetilde{E}_2\cap\widetilde{E}_3\subset\widetilde{B}\cap\widetilde{C}$ が成立する．従って，以下を得ることができる．

$$P((A)^C)\le P(\widetilde{E}_1^C)+P(\widetilde{E}_2^C)+P(\widetilde{E}_3^C).$$

次に，確率 $P(\widetilde{E}_1^C)$，$P(\widetilde{E}_2^C)$，$P(\widetilde{E}_3^C)$ をそれぞれ見ていく．チェビシェフの不等式を用いることで，(1) に基づき，

$$\begin{aligned}P(\widetilde{E}_1^C)&\le\sum_{i,j=1(i\ne j)}^{n}P(\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2-2\mathrm{tr}(\boldsymbol{\Sigma}_X)>b_1-2\mathrm{tr}(\boldsymbol{\Sigma}_X))\\&\le\sum_{i,j=1(i\ne j)}^{n}P(|\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2-2\mathrm{tr}(\boldsymbol{\Sigma}_X)|>b_1-2\mathrm{tr}(\boldsymbol{\Sigma}_X))\\&\le\sum_{i,j=1(i\ne j)}^{n}\frac{\mathrm{Var}[\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2]}{(b_1-2\mathrm{tr}(\boldsymbol{\Sigma}_X))^2}=O\{(K_X+\mathrm{tr}(\boldsymbol{\Sigma}_X^2))/d^2\}\to0\ \ (d\to\infty)\end{aligned}$$

を得る．同様に，$P(\widetilde{E}_2^C) \to 0\ \ (d \to \infty)$ を得る．ここで，$\|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2 = \|\boldsymbol{x}_i - \boldsymbol{\mu}_X\|^2 + \|\boldsymbol{y}_j - \boldsymbol{\mu}_Y\|^2 + \|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|^2 - 2(\boldsymbol{x}_i - \boldsymbol{\mu}_X)^T(\boldsymbol{y}_j - \boldsymbol{\mu}_Y) + 2(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)^T\{\boldsymbol{x}_i - \boldsymbol{\mu}_X - (\boldsymbol{y}_j - \boldsymbol{\mu}_Y)\}$ となり，$E(\|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2) = \Delta_*$ となる．さらに，$\mathrm{Var}\{(\boldsymbol{x}_i - \boldsymbol{\mu}_X)^T(\boldsymbol{y}_j - \boldsymbol{\mu}_Y)\} = \mathrm{tr}(\boldsymbol{\Sigma}_X\boldsymbol{\Sigma}_Y) \le \{\mathrm{tr}(\boldsymbol{\Sigma}_X^2)\mathrm{tr}(\boldsymbol{\Sigma}_Y^2)\}^{1/2} = o(d^2), \mathrm{Var}[(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)^T\{\boldsymbol{x}_i - \boldsymbol{\mu}_X - (\boldsymbol{y}_j - \boldsymbol{\mu}_Y)\}] = (\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)^T(\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_Y)(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y) \le \|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|^2\{\mathrm{tr}(\boldsymbol{\Sigma}_X^2)^{1/2} + \mathrm{tr}(\boldsymbol{\Sigma}_Y^2)^{1/2}\} = o(d^2)$ となることに注意すれば，$\mathrm{Var}(\|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2) = o(d^2)$ を得る．よって，

$$P(\widetilde{E}_3^C) \le \sum_{i=1}^{n}\sum_{j=1}^{m} P(|\|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2 - \Delta_*| > \Delta_* - b_1) \le O\left(\frac{\mathrm{Var}(\|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2)}{(\Delta_* - b_1)^2}\right) \to 0$$

となる． 従って，定理 1 (1) を示すことができる． □

定理 1 *(2)* の証明. (A-i) から (A-iii) と

$$\limsup_{d\to\infty} \frac{\|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|^2}{|\mathrm{tr}(\boldsymbol{\Sigma}_X) - \mathrm{tr}(\boldsymbol{\Sigma}_Y)|} < 1$$

を仮定する．次に以下の事象を定義する．

$$\begin{aligned}
A &= \{\max_{i,j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 < \min_{i,j} \|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2\},\\
B &= \{\max_{i,j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 < \min_{i,j} \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2\},\\
C &= \{\max_{i,j} \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2 < \min_{i,j} \|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2\},
\end{aligned}$$

$$\begin{aligned}
E_1 &= \{\max_{i,j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 < b_2\},\\
E_2 &= \{\min_{i,j} \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2 > b_2\},\\
E_3 &= \{\max_{i,j} \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2 < b_1\},\\
E_4 &= \{\min_{i,j} \|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2 > b_1\}.
\end{aligned}$$

このとき，共通部分 $A \cap B \cap C$ を考える．単連結法が用いられたとき，$A \cap B \cap C \subset (B)$ が成立する．また，包括関係 $E_1 \cap E_2 \cap E_3 \cap E_4 \subset A \cap B \cap C$ が成立する．従って，以下を得ることができる．

$$P((B)^C) \le P(E_1^C) + P(E_2^C) + P(E_3^C) + P(E_4^C).$$

ここで，(2) に注意し，定理 1 (1) の証明と同様にして，$P(E_i^C) \to 0\ (d \to \infty),\ i = 1, ..., 4,$ を得る．従って，定理 1 (2) を示すことができる． □

謝辞
本研究は，JST 次世代研究者挑戦的研究プログラム JPMJSP2124，科学研究費補助金 基盤研究 (A) 20H00576 研究代表者: 青嶋 誠「大規模複雑データの理論と方法論の革新的展開」，学術研究助成基金助成金 挑戦的研究 (萌芽) 19K22837 研究代表者: 青嶋 誠「超高次元データによる個別化モデリングへの挑戦」，および，科学研究費補助金 基盤研究 (C) 18K03409 研究代表者: 矢田 和善「高次元データにおける高次漸近理論の開拓とその応用」から研究助成を受けています.

参考文献

[1] 青嶋 誠，矢田和善 (2019). 高次元の統計学. 共立出版.

[2] Ahn, J., Lee, M.H., Yoon, Y.J. (2012). Clustering high dimension, low sample size data using the maximal data piling distance. *Statistica Sinica* **22**. 443–464.

[3] Borysov, P., Hannig, J., Marron, J.S. (2014). Asymptotics of hierarchical clustering for growing dimension. *Journal of Multivariate Analysis* **124**. 465–479.

[4] Huang, H., Liu, Y., Yuan, M., Marron, J.S. (2015). Statistical Significance of Clustering using Soft Thresholding. *Journal of computational and graphical statistics* **24**. 975–993.

[5] Liu, Y., Hayes, D.N., Nobel, A., Marron, J.S. (2008). Statistical significance of clustering for high-dimension, low-sample size data. *Journal of the American Statistical Association* **103**. 1281–1293.

[6] Nakayama, Y., Yata, K., Aoshima, M. (2021). Clustering by principal component analysis with Gaussian kernel in high-dimension, low-sample-size settings. *Journal of Multivariate Analysis* **185**. 104779.

[7] Yata, K., Aoshima, M. (2010). Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix. *Journal of Multivariate Analysis*, **101**, 2060–2077.

[8] Yata, K., Aoshima, M. (2020). Geometric consistency of principal component scores for high-dimensional mixture models and its application. *Scandinavian Journal of Statistics* **47**. 899–921.