

377.S1
1D

博士論文

しきい値モデルの統計量の極限定理
(Limit theorems for statistics of the threshold network model)

国立大学法人 横浜国立大学大学院
工学府

井手 勇介 Yusuke Ide

2009年3月

横浜国立大学附属図書館



12256480

目次

第1章	複雑ネットワーク	5
第2章	しきい値モデルの定義と次数	13
第3章	しきい値モデルのクラスター係数	19
第4章	しきい値モデルの階層構造	27
第5章	しきい値モデルの平均頂点間距離	37
第6章	しきい値モデルの一般化	43
第7章	結論	51

第1章 複雑ネットワーク

現実の世界に見られるネットワーク（グラフ）は，従来のグラフ理論で扱われてきたグラフと比べて複雑である．しかし，これまでの多くの研究によって，現実のネットワークには複雑な中にもスモールワールド (small world) やスケールフリー (scale free) などといった特徴的な性質をもつネットワークが多いことが認識されるようになった [1, 2, 4, 5, 12, 13, 21, 23, 24, 35–37, 51, 52, 55–59]．そのような共通の性質に着目して，ネットワークを解析する研究分野（または，そのようなモデルそのもの）を複雑ネットワーク (complex networks) と総称している．

この章では，複雑ネットワークの文脈で基本的な統計量である，次数・クラスター係数・平均頂点間距離を定義し，その代表的なモデルとして，スモールワールド性を持つ WS モデル，スケールフリー性を持つ BA モデルを紹介する．

次数

頂点 v の次数 (degree) とは， v から出ている辺の数のことであり， $d(v)$ で表す． v にループがある場合は，次数を計算する際に 2 本として数える．図 1.1 のグラフ G では， $d(1) = 2$ ， $d(2) = 2$ ， $d(3) = 2$ ， $d(4) = 3$ ， $d(5) = 3$ である．本論文では，単純グラフのみを考えるので，隣接行列 $A = (a_{ij})$ により，各頂点の次数は対応する行に属する a_{ij} の和で表される．即ち， $d(i) = \sum_{j=1}^n a_{ij}$ である．

任意のグラフにおいて，各辺の両端にはちょうど 1 つずつ頂点がつながっているので，次数を計算する際に各辺は 2 回ずつ数えられることになる．従って，全ての頂点の次数を合計すると辺の総数の 2 倍となる．この事実は握手の補題 (handshaking lemma) と呼ばれることもある．握手の補題から，どのようなグラフにおいても次数が奇数となっている頂点は偶数個であることがわかる．

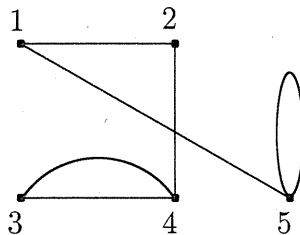


図 1.1: グラフ G

複雑ネットワークの文脈で登場するグラフは，確率的なルールによって生成されるものが多いので，その次数がある確率分布に従っている場合も考える必要がある．この次数を決め

る確率分布を次数分布 (degree distribution) と呼ぶことが多い。本論文では、 $D(i)$ を頂点 i の次数を表す確率変数とする。

クラスター係数

まず、頂点ごとのクラスター係数 (clustering coefficient) を定義する。単純グラフ G に関して、頂点 $v \in G$ のクラスター係数 $C(v)$ とは、頂点 v に隣接した $d(v)$ 個の頂点のなかで互いに隣接しているペアの数、即ち頂点 v を含む三角形の数を、頂点 v に隣接している頂点对の総数 $d(v)(d(v) - 1)/2$ で割ったものである。もし、頂点 v の次数が 0 や 1 である場合には $C(v) = 0$ とすることが多い。

また、グラフ全体のクラスター係数 C は頂点数が n 個のとき、頂点ごとのクラスター係数の相加平均

$$C = \frac{1}{n} \sum_{v \in G} C(v)$$

で定義する。

例として、図 1.2 のグラフ G の場合のクラスター係数を求める。頂点 1 に注目すると、隣接しているのは頂点 2 と頂点 4 のみである。このとき、この 3 つの頂点を使った三角形ができていないので、頂点 1 のクラスター係数は $C(1) = 1/1 = 1$ となる。次に、頂点 2 に注目すると、隣接しているのは頂点 1、頂点 3、頂点 4 の 3 つである。頂点 2 を含む三角形は 1 つしかないので、 $C(2) = 1/3$ となる。頂点 3 は次数が 1 なので三角形が作れず、 $C(3) = 0$ である。最後に、頂点 4 のクラスター係数は頂点 1 の場合と同様で、 $C(4) = 1/1 = 1$ となる。従って、グラフ全体のクラスター係数はその相加平均なので、

$$C = \frac{1}{4} \left(1 + \frac{1}{3} + 0 + 1 \right) = \frac{7}{12}$$

となる。

クラスター係数を計算するときには三角形の数を数えているが、三角形をクラスターとする根拠は特になく、例えば四角形をクラスターとして、それを数えてもよい。三角形は単純な形で他の形より数えやすいために、クラスターの定義に用いられることが多い。

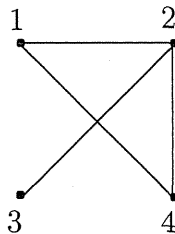


図 1.2: グラフ G

平均頂点間距離

グラフの異なる2頂点 v と w との距離 (distance) とは、 v から w へ行くために通らなくてはならない辺の数の最小値のことであり、これを $d(v, w)$ と表す。全く同じことだが、 $d(v, w)$ は v から w への道の長さの最小値であるといってもよい。但し、便宜上 $d(v, v) = 0$ と定義する。もし、 v から w への道が1つもない場合には v と w の距離は ∞ であると定義する。図 1.3 のグラフ G_1 では、頂点 1 と 2 の距離は 1、頂点 1 と 3 の距離は 3、というようになる。

頂点数が n であるとき、異なる2つの頂点对の選び方は $n(n-1)/2$ 通りある。平均頂点間距離 (average distance) L とは2頂点間の距離の $n(n-1)/2$ ペア全体にわたる相加平均のことである。つまり、

$$L = \frac{\sum_{\text{異なる頂点对 } (v,w)} d(v, w)}{n(n-1)/2} = \sum_{\text{異なる頂点对 } (v,w)} \frac{d(v, w)}{n(n-1)/2}$$

例えば、図 1.3 のグラフ G_1 では、 $n = 5$ なので $n(n-1)/2 = 10$ となる。距離 1 でつながっている頂点对が $(1, 2)$, $(1, 5)$, $(2, 4)$, $(3, 4)$ の4つ、距離 2 でつながっている頂点对が $(1, 4)$, $(2, 3)$, $(2, 5)$ の3つ、距離 3 でつながっている頂点对が $(1, 3)$ と $(4, 5)$ で2つ、距離 4 でつながっている頂点对は $(3, 5)$ のみで1つとなっているので、

$$L = 1 \times \frac{4}{10} + 2 \times \frac{3}{10} + 3 \times \frac{2}{10} + 4 \times \frac{1}{10} = 2$$

と計算される。

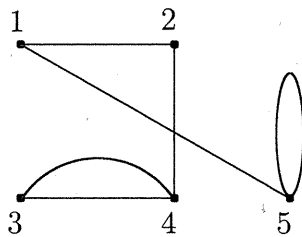


図 1.3: G_1

頂点総数が n のとき、頂点 v に注目すると、それ以外の頂点は $n-1$ 個である。そこで、頂点 v からの平均頂点間距離 $L(v)$ を

$$L(v) = \frac{\sum_{w \in G: w \neq v} d(v, w)}{n-1} = \sum_{w \in G: w \neq v} \frac{d(v, w)}{n-1}$$

で定義する。つまり、頂点 v から他の頂点への距離の相加平均を $L(v)$ とするのである。ここで、各頂点ごとにその他の頂点との距離の総和を計算して、それらを全て足すと、握手の補題と同様にして各頂点对に対してその距離をちょうど2回数えることになる。従って、

$$\frac{1}{n} \sum_{v \in G} L(v) = \frac{1}{n} \frac{\sum_{v \in G} \sum_{w \in G: w \neq v} d(v, w)}{n-1} = \frac{2 \cdot \sum_{\text{異なる頂点对 } (v,w)} d(v, w)}{n(n-1)} = L$$

となる。これは、各頂点ごとに平均頂点間距離を計算して、その相加平均を計算すれば、グラフ全体の平均頂点間距離となることを示している。

続いて、複雑ネットワークで著名な二つのモデルを紹介する。一つ目は、スモールワールド性を持つモデルとして導入されたWSモデルであり、二つ目は、スケールフリー性を持つモデルとして導入されたBAモデルである。

スモールワールド性

現実のネットワークを解析すると、正則なグラフとは異なる性質をもつネットワークが多いことがわかってきた。その性質の一つがスモールワールド性である。グラフが小さな平均頂点間距離と、大きなクラスター係数とを同時に実現しているとき、スモールワールド性をもつといわれる。スモールワールド性をもつモデルの中で最も有名な例が、次で紹介するWSモデルである。

WSモデルは、Watts and Strogatz [59]によって導入されたモデルである。このモデルは、単純なモデルでありながら、小さい平均頂点間距離、大きいクラスター係数を同時に実現できる。さらに、モデルを記述するパラメータは1つであり、そのパラメータによって、格子とランダム・グラフを極端な場合として含み、スモールワールド性がその中間の性質として自然に現れる。

以下で、WSモデルの作り方を説明する。

- (i) まず、 n 個の頂点を円環状に置く。 k を偶数として、それぞれの頂点は円環上の左右 $k/2$ 個の頂点と隣接するように辺でつなぐ。つまり、各頂点の次数は全て k であるような正則グラフである。また、 $k=2$ のときはサイクル C_n であり、一般の k に対してはサイクルの拡張であるとみなせる。本書ではこれを次数 k の拡張されたサイクルと呼ぶことにする。握手の補題により、拡張されたサイクルには全部で $kn/2$ 本の辺が存在する。
- (ii) $kn/2$ 本の辺のうち、割合 p ($0 \leq p \leq 1$)だけの辺をランダムに選ぶ。つまり、 $pkn/2$ 本の辺をランダムに選ぶ。
- (iii) 選んだそれぞれの辺に対して、片方の頂点との接続は保ったまま、もう片方の頂点と辺を切り離す。どちらの頂点との接続を保つかは確率 $1/2$ でランダムに決める。
- (iv) 各辺の新しい接続先の頂点をグラフからランダムに選んでつなぐ。この操作をつなぎかえ、新しい辺をショートカットと呼ぶことにする。但し、新しい接続先の選び方には条件がついていて、(1)自分自身、(2)そこから左右 $k/2$ 個の頂点、(3)すでにショートカットが存在する頂点は選べないこととする。即ち、ショートカットが多重辺にならないようにする (図 1.5)。

WSモデルで生成されるネットワークは、つなぎかえ確率 p の値に応じて図 1.6 のようになる。 $p=0$ の時にはつなぎかえが起こっていないので、拡張されたサイクルのままである。 $p=1$ の時には、全ての辺に対してつなぎかえが起こることになる。従って、基本的にランダム・グラフと同じである。スモールワールド性が現れる範囲を明確に決めることはできないが、小さい平均頂点間距離と大きなクラスター係数が同時に現れる p の範囲が存在することが、数値計算や近似計算によりわかっている。

次数分布に関しては、 $p=0$ のときには拡張されたサイクルであり、全ての頂点の次数は

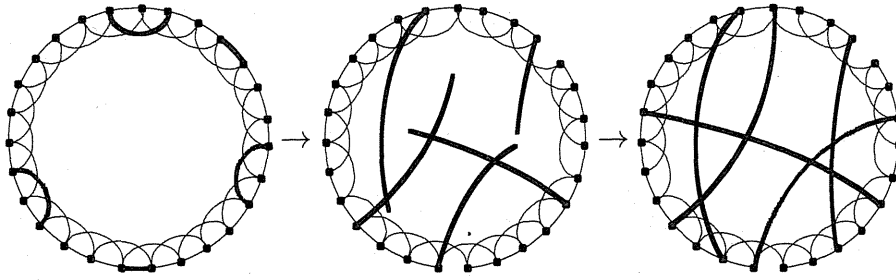


図 1.4: WSモデルの作り方 ($k=4$)

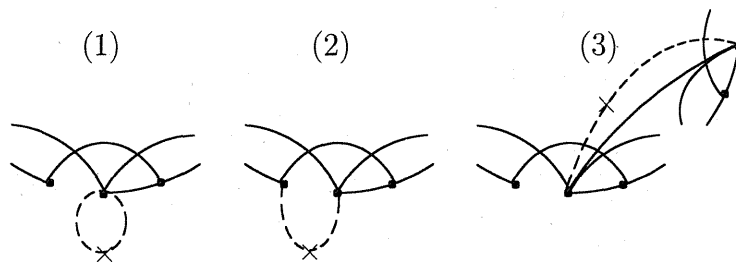


図 1.5: WSモデルで許されない辺のつなぎかえ方

k である. $p=1$ のときには, ポアソン分布で近似し得ることが知られている. $0 < p < 1$ では, これらの中間の形の次数分布が得られることがわかっている.

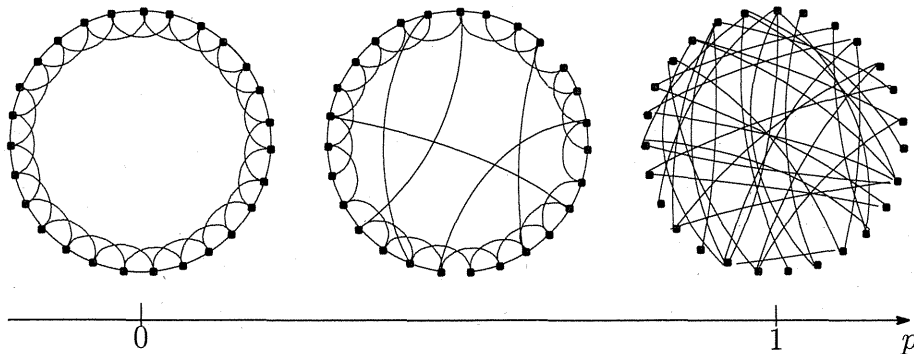


図 1.6: 様々なつなぎかえ確率 p に対する WSモデル ($k=4$)

WSモデルは単純なモデルであるので, シミュレーションや近似計算により様々な性質が調べられている (例えば増田, 今野 [36] を参照). しかし, このモデルを確率論の枠組みできちんと定義し, 解析するのは容易ではない. WSモデルの特徴は辺をつなぎ替えるルールであるが, このルールが定義や解析を難しくしているのである. そのため, WSモデルの解析を容易にするような修正版モデルが考案され, 解析されている (例えば, Bollobás and Chung [11] や Newman and Watts [44] 等を参照).

スケールフリー性

スモールワールド性と並んで、複雑ネットワークでもう1つの大きなキーワードとなるのがスケールフリー性である。現実のネットワークの次数分布を調べてみると、ベキ分布であることが多く、そのようなグラフはスケールフリー性をもつといわれる。ここで、次数分布がベキ分布であるとは、

$$p(k) \propto k^{-\gamma}, \quad (\gamma > 0)$$

を満たしていることである。但し、 α は比例していることを表す記号である。また、 γ はベキ指数と呼ばれている。

BAモデルは、Barabási and Albert [6] によって提案されたモデルである。BAモデルでは、成長と優先的選択によって、ベキ分布に従う次数分布をもつグラフを得ることができる。以下で、BAモデルの作り方を説明する。

- (i) まず、 $m_0 \geq 1$ 個の頂点を用意して、完全グラフ K_{m_0} を作る。
- (ii) 次に、新しい頂点を1つ追加する（成長）。このとき、新しい頂点は $m \leq m_0$ 本の辺を持って追加される。
- (iii) m 本の新しい辺は、既にあるグラフの各頂点に接続するが、そのとき、各頂点の次数に比例した確率で接続先が選ばれる（優先的選択）。この選択によってできるグラフが単純グラフになるように、同じ頂点に2つ以上の辺が接続されないように制限するのが一般的である。
- (iv) 上記の(ii)と(iii)を繰り返す。

ここで、優先的選択についてももう少し詳しく説明する。(iii)で、既にあるグラフの頂点が n 個であるとする。このとき、新しい辺が頂点 i ($1 \leq i \leq n$) に接続する確率は次数 $d(i)$ に比例する。つまり、確率 $d(i) / \sum_{j=1}^n d(j)$ で新しい辺の接続先として、頂点 i が選ばれる。図1.7は、 K_4 からスタートして新しい辺が毎回2本ずつ追加されるBAモデルの成長の様子を表したものである。ここで、各頂点に振ってある数字は、その頂点が新しい辺の接続先として選ばれる確率を表している。

例えば、 $t = 0$ の完全グラフ K_4 では、各頂点を選ばれる確率は全て等しくなっている。 $t = 3$ のグラフでは、辺の総数が12本なので、握手の補題により $\sum_{j=1}^n d(j) = 24$ である。このとき、各頂点の次数に偏りが出ているので、各頂点が新しい辺の接続先として選ばれる確率に差が出てきている。

このようなルールで確率的なグラフを作っていくのだが、作り方からもわかるように、最初ほどの頂点も同じ次数である。しかし、次のステップでたまたま新しい辺の接続先として選ばれた頂点は、その次のステップでは、新しい辺の接続先としてより選ばれやすくなる。このようにして、ステップを進める毎に、各頂点が獲得できる辺の数に差が出てくる。始めの数ステップで多くの辺を獲得した頂点は、大きな次数を持つ頂点に成長し易い。このような大きな次数をもつ頂点のことをハブ (hub) と呼ぶ。

このようにして作られたBAモデルの次数分布は、ベキ指数が $\gamma = 3$ のベキ分布 $p(k) \propto k^{-3}$ に従うことが、連続近似などによりわかっている。実は、近似計算やモンテカルロ・シミュレーションによってBAモデルを解析することは可能であるが、優先的選択を数学的にきち

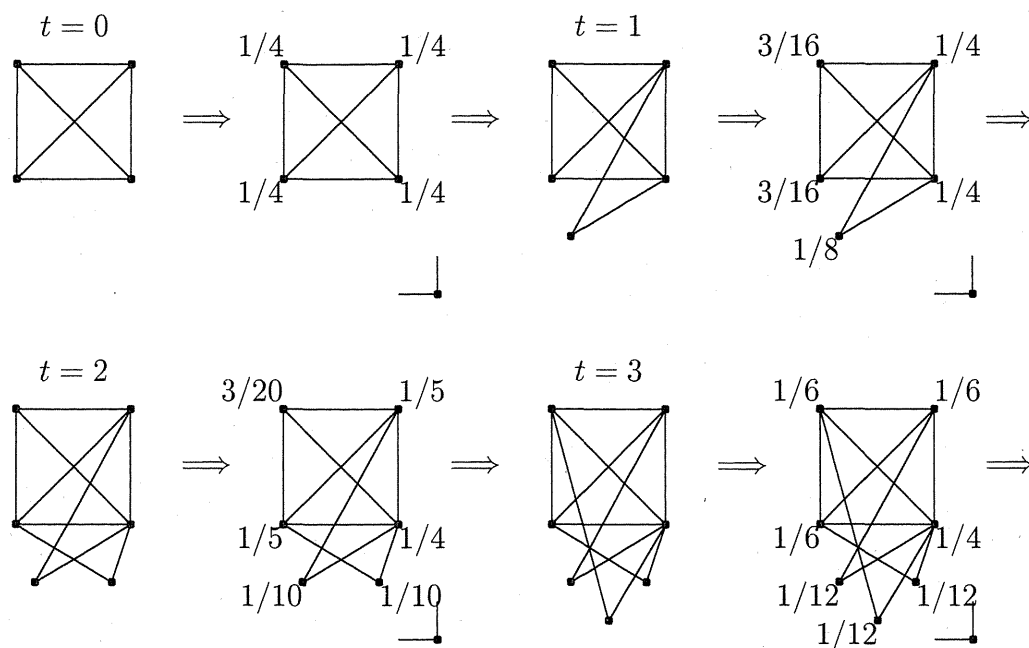


図 1.7: BAモデル

んと扱うのは一般に難しい。そこで、いくつかの数学的に厳密に扱えるモデルが導入、解析されている（例えば、Móri [41] や Chung and Lu [15] を参照）。

第2章 しきい値モデルの定義と次数

しきい値（閾値）モデルは、Caldarelli *et al.* [14] と Söderberg [50] によって独立に提案されたモデルの特別な場合であり、種々の特性量が物理的な手法 [9, 14, 25, 38–40, 49, 50], 数学的な手法 [26, 27, 32, 42] を用いて調べられている. 特に, 重みを与える確率変数の分布が指数分布やパレート分布の場合に, その次数分布がベキ則に従うこと (スケールフリー性) が示されており, 複雑ネットワークの分野でも注目されているモデルの1つである. 以下で, しきい値モデルを定義する.

しきい値グラフ $G_n = (V_n, E_n)$ は, n 個のラベル付けされた頂点の集合 $V_n = \{1, 2, \dots, n\}$ からなる有限単純グラフである. 辺集合 E_n を定義するために, n 個の実数 (重み) x_1, x_2, \dots, x_n と, しきい値 $\theta \in \mathbb{R}$ を用意する. しきい値グラフでは, 各頂点 $i \in V_n$ に重み x_i を対応させ, 異なる2頂点 $i, j \in V_n$ はそれぞれの重みの和がしきい値を超えたときに限って辺で結ばれる. 即ち, $E_n = \{(i, j) : x_i + x_j > \theta, 1 \leq i < j \leq n\}$ と定義する. しきい値グラフには何種類もの同値な定式化が存在することが知られている (例えば [34] を参照). ここで, 重みを独立同分布に従う確率変数列 X_1, X_2, \dots, X_n で与え, 従って, 辺集合を (確率的に) $E_n = \{(i, j) : X_i + X_j > \theta, 1 \leq i < j \leq n\}$ と定義した, ランダムなしきい値グラフをしきい値モデルと呼ぶことにする. つまり, しきい値モデルでは $X_i + X_j > \theta$ ($i \neq j$) のときに限り, 頂点 i と頂点 j が辺 $(i, j) \in E_n$ で結ばれる (図 2.1).

$$\begin{array}{c} X_i \\ \cdot \\ i \end{array} \text{---} \begin{array}{c} X_j \\ \cdot \\ j \end{array} \iff X_i + X_j > \theta$$

図 2.1: しきい値モデルのルール

しきい値モデルの次数を考えるために, 関数 $h_e(x, y) = I_{(\theta, \infty)}(x + y)$ を導入する. 即ち, $x + y > \theta$ のとき $h_e(x, y) = 1$ となり, $x + y \leq \theta$ のときには $h_e(x, y) = 0$ となるような関数である. しきい値モデルの定義より, 頂点 i と頂点 j は $h_e(X_i, X_j) = 1$ のとき辺 (i, j) で結ばれ, $h_e(X_i, X_j) = 0$ のときには辺で結ばれないので, 頂点 1 の次数 $D_n(1)$ は

$$D_n(1) = \sum_{2 \leq j \leq n} h_e(X_1, X_j)$$

と表すことができる. ここで, 任意の $x \in \mathbb{R}$ に対して $D_n(1; x)$ を頂点 1 の重みが x であるときの頂点 1 の次数であるとする. 即ち,

$$D_n(1; x) = \sum_{2 \leq j \leq n} h_e(x, X_j)$$

とする. このとき, 以下の定理が成り立つ.

定理 2.0.1(i) 任意の $x \in \mathbb{R}$ に対して,

$$\lim_{n \rightarrow \infty} \frac{D_n(1; x)}{n-1} = 1 - F(\theta - x), \quad \text{a.s.}$$

(ii) 特に,

$$\lim_{n \rightarrow \infty} \frac{D_n(1)}{n-1} = 1 - F(\theta - X_1), \quad \text{a.s.}$$

証明

$D_n(1; x) = \sum_{2 \leq j \leq n} h_e(x, X_j)$ は, 独立同分布な確率変数の $n-1$ 個の和である. 従って, 大数の強法則により, $D_n(1; x)/(n-1)$ は確率 1 で期待値 $\mathbb{E}[h_e(x, X_j)]$ に収束する. ここで, $h_e(x, y) = I_{(\theta, \infty)}(x+y)$ であることに気をつけると,

$$\mathbb{E}[h_e(x, X_j)] = \mathbb{P}(X_2 > \theta - x) = 1 - F(\theta - x)$$

を得る. さらに, $\{X_i\}_{1 \leq i \leq n}$ が互いに独立であることを使うと,

$$\begin{aligned} & \mathbb{P}\left(\lim_{n \rightarrow \infty} D_n(1)/(n-1) = 1 - F(\theta - X_1)\right) \\ &= \int_{\mathbb{R}} \mathbb{P}\left(\lim_{n \rightarrow \infty} D_n(1; x)/(n-1) = 1 - F(\theta - x)\right) F(dx) = \int_{\mathbb{R}} 1 \cdot F(dx) = 1 \end{aligned}$$

がいえる. つまり, $D_n(1)/(n-1)$ は確率 1 で確率変数 $1 - F(\theta - X_1)$ に収束する.

□

定理 2.0.1 の (ii) に出てくる極限は, X_1 の分布から決まる確率変数である. 従って, その分布は重みの分布によって様々に変わる可能性がある. 以下の補題は, X_1 の分布が絶対連続な場合, 即ち, 確率密度関数 f を持つ場合の極限分布の表示を与える. 以下, $\text{supp} f = \overline{\{x \in \mathbb{R} : f(x) \neq 0\}}$ を f の台とする.

補題 2.0.1(i) ある $a \in \mathbb{R}$ が存在して, $\text{supp} f = [a, \infty)$ ならば,

$$1 - F(\theta - X_1) \sim \begin{cases} \delta_1(dk) & \text{if } \theta \leq 2a, \\ I_{(1-F(\theta-a), 1)}(k) \cdot \frac{f(\theta - F^{-1}(1-k))}{f(F^{-1}(1-k))} dk \\ \quad + (1 - F(\theta - a)) \cdot \delta_1(dk) & \text{if } \theta > 2a. \end{cases}$$

(ii) ある $b \in \mathbb{R}$ が存在して, $\text{supp} f = (-\infty, b]$ ならば,

$$1 - F(\theta - X_1) \sim \begin{cases} F(\theta - b) \cdot \delta_0(dk) \\ \quad + I_{(0, 1-F(\theta-b))}(k) \cdot \frac{f(\theta - F^{-1}(1-k))}{f(F^{-1}(1-k))} dk & \text{if } \theta < 2b, \\ \delta_0(dk) & \text{if } \theta \geq 2b. \end{cases}$$

(iii) ある $a, b \in \mathbb{R}$ が存在して, $\text{supp} f = [a, b]$ ならば,

$$1 - F(\theta - X_1) \sim \begin{cases} \delta_1(dk) & \text{if } \theta \leq 2a, \\ I_{(1-F(\theta-a), 1)}(k) \cdot \frac{f(\theta - F^{-1}(1-k))}{f(F^{-1}(1-k))} dk \\ \quad + (1 - F(\theta - a)) \cdot \delta_1(dk) & \text{if } 2a < \theta < a + b, \\ I_{(0, 1)}(k) \cdot \frac{f(a+b - F^{-1}(1-k))}{f(F^{-1}(1-k))} dk & \text{if } \theta = a + b, \\ F(\theta - b) \cdot \delta_0(dk) \\ \quad + I_{(0, 1-F(\theta-b))}(k) \cdot \frac{f(\theta - F^{-1}(1-k))}{f(F^{-1}(1-k))} dk & \text{if } a + b < \theta < 2b, \\ \delta_0(dk) & \text{if } \theta \geq 2b. \end{cases}$$

さらに, f が $\text{supp} f$ 上で対称ならば,

$$1 - F(\theta - X_1) \sim I_{(0, 1)}(k) dk \quad \text{if } \theta = a + b.$$

(iv) $\text{supp} f = (-\infty, \infty)$ ならば, 全ての $\theta \in \mathbb{R}$ に対して,

$$1 - F(\theta - X_1) \sim I_{(0, 1)}(k) \cdot \frac{f(\theta - F^{-1}(1-k))}{f(F^{-1}(1-k))} dk.$$

以下で, いくつかの分布に関して極限分布の例を挙げる.

例 2.0.1 (指数分布)

X_1 が, 平均 $1/\lambda$ の指数分布に従うとき, 即ち, ある $\lambda > 0$ に対して

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (2.1)$$

であるとき,

$$1 - F(\theta - X_1) \sim \begin{cases} \delta_1(dk) & \text{if } \theta \leq 0, \\ I_{(e^{-\lambda\theta}, 1)}(k) \cdot \frac{e^{-\lambda\theta}}{k^2} dk + e^{-\lambda\theta} \cdot \delta_1(dk) & \text{if } \theta > 0. \end{cases}$$

例 2.0.2 (パレート分布)

X_1 が, パレート分布に従うとき, 即ち, ある $a, c > 0$ に対して

$$f(x) = \begin{cases} \frac{c}{a} \cdot \left(\frac{a}{x}\right)^{c+1} & \text{if } x \geq a, \\ 0 & \text{otherwise,} \end{cases}$$

であるとき,

$$1 - F(\theta - X_1) \sim \begin{cases} \delta_1(dk) & \text{if } \theta \leq 2a, \\ I_{((\frac{a}{\theta-a})^c, 1)}(k) \cdot \left(\frac{a}{\theta \cdot k^{1/c} - a}\right)^{c+1} dk + \left(\frac{a}{\theta-a}\right)^c \cdot \delta_1(dk) & \text{if } \theta > 2a. \end{cases}$$

例 2.0.3 (一様分布)

X_1 が, $[0, 1]$ 上の一様分布に従うとき, 即ち,

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

であるとき,

$$1 - F(\theta - X_1) \sim \begin{cases} \delta_1(dk) & \text{if } \theta \leq 0, \\ I_{(1-\theta, 1)}(k)dk + (1 - \theta) \cdot \delta_1(dk) & \text{if } 0 < \theta < 1, \\ I_{(0, 1)}(k)dk & \text{if } \theta = 1, \\ (\theta - 1) \cdot \delta_0(dk) + I_{(0, 2-\theta)}(k)dk & \text{if } 1 < \theta < 2, \\ \delta_0(dk) & \text{if } \theta \geq 2. \end{cases}$$

しきい値モデルの特性量を調べる際に, U 統計量 (例えば Serfling [48] を参照) は強力な道具となる. U 統計量に関して, 大数の法則, 中心極限定理等さまざまな極限定理が知られているが, ここでは特に, 独立同分布な確率変数に関する U 統計量の大数の強法則を定理の形でまとめておくことにする.

$\{X_i\}_{i=1, 2, \dots}$ を独立同分布な確率変数列であるとする. ある自然数 m に対して m 変数の実可測関数 $h(x_1, \dots, x_m)$ を用意する. ただし, $h(x_1, \dots, x_m)$ は変数の任意の入れ替えに関してその値を変えないものとする. このとき,

$$U_m(X_1, \dots, X_n) = \sum_{1 \leq i_1 < \dots < i_m \leq n} h(X_{i_1}, \dots, X_{i_m})$$

で定義される確率変数 $U_m(X_1, \dots, X_n)$ を U 統計量と呼ぶ. U 統計量に関して以下の大数の強法則が成り立つ.

定理 2.0.2

$\mathbb{E}[h(X_1, \dots, X_m)] < \infty$ を仮定する. このとき,

$$\lim_{n \rightarrow \infty} \frac{U_m(X_1, \dots, X_n)}{\binom{n}{m}} = \mathbb{E}[h(X_1, \dots, X_m)], \quad \text{a.s.}$$

が成り立つ.

定理 2.0.2 により, 直ちにグラフ全体の辺の数に関する極限定理を得る. 即ち, グラフ全体の辺の数 D_n を

$$D_n = \sum_{1 \leq i < j \leq n} h_e(X_i, X_j)$$

で定義するとき, 次の定理が成り立つ.

定理 2.0.3

$$\lim_{n \rightarrow \infty} \frac{D_n}{\binom{n}{2}} = \mathbb{E}[h_e(X_1, X_2)] = \mathbb{P}(X_1 + X_2 > \theta), \quad \text{a.s.}$$

ここで、握手の補題により

$$\frac{D_n}{\binom{n}{2}} = \frac{2 \sum_{1 \leq i < j \leq n} D_n(i, j)}{\binom{n}{2}} = \frac{1}{n} \sum_{1 \leq i \leq n} \frac{D_n(i)}{n-1}$$

なる関係が成り立っていることに注意が必要である。この関係と定理 2.0.3 により、

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{1 \leq i \leq n} \frac{D_n(i)}{n-1} = \mathbb{P}(X_1 + X_2 > \theta) = \mathbb{E}[1 - F(\theta - X_1)], \quad \text{a.s.}$$

が得られる。この主張は、次数を表す確率変数がエルゴード性を持つことを意味している。

第3章 しきい値モデルのクラスター係数

この章では、しきい値モデルのクラスター係数の極限定理を議論する。そのために、頂点 i の次数 $D_n(i)$ と頂点 i を含む三角形の個数 $T_n(i)$ を以下で定義する。

$$D_n(i) = \sum_{\substack{1 \leq j \leq n \\ j \neq i}} h_e(X_i, X_j),$$

$$T_n(i) = \sum_{\substack{1 \leq j < k \leq n, \\ j, k \neq i}} h_T(X_i, X_j, X_k).$$

ここで、 $h_e(x, y) = I_{(\theta, \infty)}(x+y)$ は辺の定義関数であり、 $h_T(x, y, z) = I_{(\theta, \infty)}(x+y) \cdot I_{(\theta, \infty)}(y+z) \cdot I_{(\theta, \infty)}(z+x)$ は三角形の定義関数である。これらを用いて、頂点 i のクラスター係数 $C_n(i)$ は

$$C_n(i) = \frac{T_n(i)}{\binom{D_n(i)}{2}} \cdot I_{\{2, \dots, n-1\}}(D_n(i)) + w \cdot I_{\{0, 1\}}(D_n(i))$$

と定義される。ここで、不定元 w はクラスター係数の右边第一項が定義できない場合の寄与を表している。もし、クラスター係数の値を $C_n(i) \in [0, 1]$ とする必要がある場合には、 w に $[0, 1]$ の実数値を代入すればよい。第1章では、 $w = 0$ とした定義を紹介したが、本章では、それ以外の場合も統一的に扱うために不定元を用いた定義を採用した。

クラスター係数の計算を容易にするために、以下で定義を変更する。頂点 i を含む「 Λ 字型」の個数を、

$$\Lambda_n(i) = \sum_{\substack{1 \leq j < k \leq n, \\ j, k \neq i}} h_\Lambda(X_i, X_j, X_k),$$

で定義する。ここで、 $h_\Lambda(x, y, z) = I_{(\theta, \infty)}(x+y) \cdot I_{(\theta, \infty)}(x+z)$ とする。この $\Lambda_n(i)$ は、頂点 i と辺で結ばれている頂点对の数を表す確率変数である。定義から、 $D_n(i) \geq 2$ と $\Lambda_n(i) \geq 1$ は同値であり、事象 $\{D_n(i) \geq 2\}$ 上で

$$\binom{D_n(i)}{2} = \Lambda_n(i)$$

が成り立つ。従って、クラスター係数の定義を以下のように変更することができる。

$$C_n(i) = \frac{T_n(i)}{\Lambda_n(i)} \cdot I_{\{1, \dots, (n-1)\}}(\Lambda_n(i)) + w \cdot I_{\{0\}}(\Lambda_n(i)).$$

前節までの議論と同様に, $T_n(i; x)$, $V_n(i; x)$, $C_n(i; x)$, は $X_i = x$ である場合の $T_n(i)$, $V_n(i)$, $C_n(i)$, の値を表すものとする. 例えば, $T_n(1; x) = \sum_{2 \leq j < k \leq n} h_T(x, X_j, X_k)$ である. このとき, 以下の定理が成り立つ.

定理 3.0.4

(i) 任意の $x \in \mathbb{R}$ に対して,

$$\lim_{n \rightarrow \infty} C_n(1; x) = C(1; x), \quad \text{a.s.}$$

(ii) 特に,

$$\lim_{n \rightarrow \infty} C_n(1; x) = C(1) \equiv C(1; X_1), \quad \text{a.s.}$$

ここで,

$$C(1; x) = \frac{\mathbb{P}(h_T(x, X_2, X_3) = 1)}{\mathbb{P}(x + X_2 > \theta)^2} \cdot I_{(0,1]}(\mathbb{P}(x + X_2 > \theta)) + w \cdot I_{\{0\}}(\mathbb{P}(x + X_2 > \theta))$$

である.

証明

まず, 任意の $x \in \mathbb{R}$ に対して,

$$\mathbb{E}[h_\Lambda(x, X_2, X_3)] = 0 \iff \mathbb{P}(\Lambda_n(1; x) = 0 \text{ for all } n \geq 1) = 1. \quad (3.1)$$

であることを示す. 実際, $\mathbb{E}[h_\Lambda(x, X_2, X_3)] = 0$ ならば,

$$\mathbb{E}[\Lambda_n(1; x)] = \mathbb{E}\left[\sum_{2 \leq j < k \leq n} h_\Lambda(x, X_j, X_k)\right] = \sum_{2 \leq j < k \leq n} \mathbb{E}[h_\Lambda(x, X_j, X_k)] = 0$$

が全ての $n \geq 1$ に対して成り立つ. 逆に, 全ての $n \geq 1$ に対して, $\mathbb{E}[\Lambda_n(1; x)] = 0$ ならば,

$$\mathbb{E}[h_\Lambda(x, X_2, X_3)] = \mathbb{E}[\Lambda_3(1; x)] = 0$$

である. 従って,

$$\mathbb{E}[h_\Lambda(x, X_2, X_3)] = 0 \iff \mathbb{E}[\Lambda_n(1; x)] = 0 \text{ for all } n \geq 1.$$

を得る. ここで, $\Lambda_n(1; x)$ は非負であることから,

$$\mathbb{E}[\Lambda_n(1; x)] = 0 \text{ for all } n \geq 1 \iff \mathbb{P}(\Lambda_n(1; x) = 0) = 1 \text{ for all } n \geq 1.$$

である. さらに, 事象 $\{\Lambda_n(1; x) = 0\}$ は n に関して単調非減少であるので,

$$\mathbb{P}(\Lambda_n(1; x) = 0) = 1 \text{ for all } n \geq 1 \iff \mathbb{P}(\Lambda_n(1; x) = 0 \text{ for all } n \geq 1) = 1$$

となり, (3.1) を得る. 従って, クラスター係数の定義より, $\mathbb{E}[h_V(x, X_2, X_3)] = 0$ は $\mathbb{P}(\lim_{n \rightarrow \infty} C_n(1; x) = w) = 1$ と同値である.

一方, 定義より $\Lambda_n(1; x)$ は $\{X_2, \dots, X_n\}$ の値の入れ替えに関して不変であるので, Hewitt-Savage の 0-1 法則より, $\mathbb{P}(\Lambda_n(1; x) = 0 \text{ for all } n \geq 1)$ は 0 か 1 かのどちらかの値しかとらない. 従って,

$$\begin{aligned} & \mathbb{E}[h_\Lambda(x, X_2, X_3)] > 0 \\ \iff & \mathbb{P}(\Lambda_n(1; x) = 0 \text{ for all } n \geq 1) = 0 \iff \mathbb{P}(\Lambda_n(1; x) \geq 1 \text{ for some } n \geq 1) = 1 \end{aligned}$$

を得る. さらに, $\Lambda_n(1; x)$ は単調非減少であるので, 事象 $\{\Lambda_n(1; x) \geq 1 \text{ for some } n \geq 1\}$ は事象

$\{\exists N \geq 1 \text{ s.t. } \Lambda_n(1; x) \geq 1 \text{ for all } n \geq N\}$. と同値である. 従って,

$$\begin{aligned} & \mathbb{E}[h_\Lambda(x, X_2, X_3)] > 0 \\ \iff & \mathbb{P}(\exists N \geq 1 \text{ s.t. } \Lambda_n(1; x) \geq 1 \text{ for all } n \geq N) = 1 \\ \iff & \mathbb{P}(\exists N \geq 1 \text{ s.t. } C_n(1; x) = T_n(1; x)/\Lambda_n(1; x) \text{ for all } n \geq N) = 1. \end{aligned} \quad (3.2)$$

となる.

ここで, $x \in \mathbb{R}$ を固定することにより, $h_T(x, x_2, x_3)$ と $h_\Lambda(x, x_2, x_3)$ は変数 x_2, x_3 に関して対称な関数であるので,

$$\begin{aligned} T_n(1; x) &= \sum_{2 \leq j < k \leq n} h_T(x, X_j, X_k), \\ \Lambda_n(1; x) &= \sum_{2 \leq j < k \leq n} h_\Lambda(x, X_j, X_k). \end{aligned}$$

は U -統計量である. 従って, 定理 2.0.2 より,

$$\lim_{n \rightarrow \infty} \frac{T_n(1; x)}{\binom{n-1}{2}} = \mathbb{E}[h_T(x, X_2, X_3)], \quad \text{a.s.} \quad (3.3)$$

$$\lim_{n \rightarrow \infty} \frac{\Lambda_n(1; x)}{\binom{n-1}{2}} = \mathbb{E}[h_\Lambda(x, X_2, X_3)], \quad \text{a.s.} \quad (3.4)$$

を得る. このとき, 二つの式 (3.3) と (3.4) により, クラスタ係数

$$C_n(1; x) = \frac{T_n(1; x)}{\Lambda_n(1; x)} = \frac{T_n(1; x)/\binom{n-1}{2}}{\Lambda_n(1; x)/\binom{n-1}{2}}$$

は確率 1 で $\mathbb{E}[h_T(x, X_2, X_3)]/\mathbb{E}[h_\Lambda(x, X_2, X_3)]$ へ収束することがわかる. よって, 式 (3.2) とあわせて,

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} C_n(1; x) = \mathbb{E}[h_T(x, X_2, X_3)]/\mathbb{E}[h_\Lambda(x, X_2, X_3)]\right) = 1. \quad (3.5)$$

を得る.

以上の議論より, X_2 と X_3 の独立性から $\mathbb{E}[h_\Lambda(x, X_2, X_3)] = \mathbb{E}[h_e(x, X_2)]^2 = \mathbb{P}(x + X_2 > \theta)^2$ が成り立っていることに注意することで (i) を得る.

(ii) は次数の場合と同様, Fubini の定理より, 以下のように得られる.

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} C_n(1) = C(1)\right) = \int_{\mathbb{R}} \mathbb{P}\left(\lim_{n \rightarrow \infty} C_n(1; x) = C(1; x)\right) F(dx) = \int_{\mathbb{R}} 1 \cdot F(dx) = 1.$$

□

ここまでに、それぞれの頂点に関するクラスター係数を議論した。次に、グラフ全体のクラスター係数

$$C_n = \frac{1}{n} \sum_{i=1}^n C_n(i).$$

に関する極限定理を考える。\$C_n\$ は、各頂点 \$i\$ に関するクラスター係数 \$C_n(i)\$ の相加平均である。ここで、\$C_n(i)\$ は同分布であるが独立ではないので、極限定理は自明ではない。しかし、\$C_n\$ は変数 \$(X_1, \dots, X_n)\$ の入れ替えに関して対称であることから、エルゴード定理を用いることにより、以下の定理を得ることができる。

定理 3.0.5

$$\lim_{n \rightarrow \infty} C_n = \mathbb{E}[C(1)], \quad \text{a.s.}$$

証明

簡単のため、\$\mathbb{E}[C(1)] = 0\$ の場合のみを考える。一般の場合は、\$C(1)\$ を \$C(1) - \mathbb{E}[C(1)]\$ とするような平行移動によって \$\mathbb{E}[C(1)] = 0\$ の場合に帰着する。\$\mathbf{x} = (x_1, x_2, \dots)\$ を重みの実現値を表す無限次元ベクトルとする。まず、任意の \$n \ge 1\$ に対して、直積測度 \$\mathbb{P}\$ に関する保測変換 \$T_n\$ を以下で定義する。

$$(T_n \mathbf{x})_k = \begin{cases} x_{k+1} & \text{if } 1 \leq k \leq n-1, \\ x_1 & \text{if } k = n, \\ x_k & \text{otherwise.} \end{cases}$$

但し、\$(T_n \mathbf{x})_k\$ はベクトル \$T_n \mathbf{x}\$ の \$k\$ 番目の成分を表すものとする。つまり、\$T_n\$ は \$\mathbf{x}\$ の最初の \$n\$ 個の成分のみを回転させて、他の成分は変化させないような変換である。ここで、\$C_n(i; \mathbf{x})\$ を \$X_1, \dots, X_n\$ の実現値が \$x_1, \dots, x_n\$ と与えられたときの \$C_n(i)\$ の実現値とすれば、重みの実現値が \$\mathbf{x}\$ であるときの \$C_n\$ の実現値を

$$C_n(\mathbf{x}) = \frac{1}{n} \sum_{i=0}^{n-1} C_n(1; T_n^i \mathbf{x}).$$

と表すことができる。

任意に固定した \$\varepsilon > 0\$ に対して、

$$C_n^\varepsilon(1; \mathbf{x}) = (C_n(1; \mathbf{x}) - \varepsilon) \cdot I_{A_\varepsilon}, \quad (3.6)$$

$$S_n^\varepsilon(\mathbf{x}) = \sum_{i=0}^{n-1} C_n^\varepsilon(1; T_n^i \mathbf{x}),$$

を定義する. 但し, $A_\varepsilon = \{\mathbf{x} : \limsup_{n \rightarrow \infty} C_n(\mathbf{x}) > \varepsilon\}$. である. 最大エルゴード定理 (例えば [8] を参照) により, 任意の $n \geq 1$ に対して,

$$\int_{M_n^\varepsilon} C_n^\varepsilon(1; \mathbf{x}) d\mathbb{P} \geq 0$$

を得る. ここで, $M_n^\varepsilon = \{\mathbf{x} : \sup_{1 \leq j \leq n} S_j^\varepsilon(\mathbf{x}) > 0\}$ である. 一方, 定義より $n \rightarrow \infty$ の極限で

$$M_n^\varepsilon \uparrow \left\{ \mathbf{x} : \sup_{k \geq 1} S_k^\varepsilon(\mathbf{x}) > 0 \right\} = \left\{ \mathbf{x} : \sup_{k \geq 1} \frac{S_k^\varepsilon(\mathbf{x})}{k} > 0 \right\} = \left\{ \mathbf{x} : \sup_{k \geq 1} C_k(\mathbf{x}) > \varepsilon \right\} \cap A_\varepsilon = A_\varepsilon$$

を得るので, 優収束定理と定理 3.0.4 により, $n \rightarrow \infty$ の極限で,

$$0 \leq \int_{M_n^\varepsilon} C_n^\varepsilon(1; \mathbf{x}) d\mathbb{P} \rightarrow \int_{A_\varepsilon} [C(1; \mathbf{x}) - \varepsilon] d\mathbb{P} \quad (3.7)$$

を得る.

\mathcal{I}_n を最初の n 個の確率変数の任意の入れ替えに関して不変な事象の族とし, $\mathcal{I} = \bigcap_{n=1}^{\infty} \mathcal{I}_n$ とする. 定義より, $A_\varepsilon \in \mathcal{I}$ であることに注意する. Hewitt-Savage の 0-1 法則 (例えば [8] を参照) によると, 任意の $A \in \mathcal{I}$ に対して, $\mathbb{P}(A)$ は 0 または 1 のどちらかの値しかとらないので, 条件付期待値 $\mathbb{E}[C(1)|\mathcal{I}]$ は確率 1 で $\mathbb{E}[C(1)]$ と等しい. よって, 式 (3.7) と $\mathbb{E}[C(1)] = 0$ より,

$$\begin{aligned} 0 \leq \int_{A_\varepsilon} [C(1; \mathbf{x}) - \varepsilon] d\mathbb{P} &= \int_{A_\varepsilon} C(1; \mathbf{x}) d\mathbb{P} - \varepsilon \mathbb{P}(A_\varepsilon) = \int_{A_\varepsilon} \mathbb{E}[C(1; \mathbf{x})|\mathcal{I}] d\mathbb{P} - \varepsilon \mathbb{P}(A_\varepsilon) \\ &= \int_{A_\varepsilon} \mathbb{E}[C(1)] d\mathbb{P} - \varepsilon \mathbb{P}(A_\varepsilon) = -\varepsilon \mathbb{P}(A_\varepsilon) \end{aligned}$$

を得る. 従って, 任意の $\varepsilon > 0$ に対して $\mathbb{P}(A_\varepsilon) = 0$ が成り立つ. これより, 確率 1 で $\limsup_{n \rightarrow \infty} C_n \leq 0$ となる. 同じ議論を $-C_n$ に対して繰り返すことで, $\liminf_{n \rightarrow \infty} C_n \geq 0$ が確率 1 で得られる.

□

ここで, 定理 3.0.5 の結果がより簡単な議論で導かれる例を以下に示す.

例 3.0.4

$\{X_i\}_{i=1,2,\dots}$ を, $\mathbb{P}(X_1 = 1) = p$, $\mathbb{P}(X_1 = 0) = 1 - p$ を満たす独立同分布な確率変数列とする. ここで, しきい値 θ を $0 \leq \theta < 1$ を満たすように固定して, しきい値モデルを考える. つまり, この場合には $X_i = X_j = 0$ の場合のみ頂点 i と頂点 j は辺で結ばれないことになる. このとき, 定義に基づいて計算すると,

$$\begin{aligned} \mathbb{E}[C(1)] &= p \cdot C(1; 1) + (1 - p) \cdot C(1; 0) \\ &= p \cdot \frac{p^2 + 2p(1 - p)}{1} + (1 - p) \cdot \frac{p^2}{p^2} \\ &= 1 - p(1 - p)^2 \end{aligned}$$

となる。これから、 C_n が $1 - p(1 - p)^2$ に確率1で収束することを見る。そのために C_n を $S_n = \sum_{i=1}^n X_i$ で場合分けして計算する。ここで、 S_n は $X_i = 1$ となっている頂点の数を表す確率変数である。

まず、 $S_n = 0$ の場合は、重みが1の頂点が1つも無いのでどの頂点も孤立点である。従って、 $C_n = w$ である。

また、 $S_n = 1$ の場合は、重みが1の頂点が1つだけあって、その他の $n - 1$ この頂点の重みは0である。このとき、重みが1の頂点と0の頂点は辺で結ばれるが、重みが0の頂点同士は辺で結ばれない(図3.1)。従って、

$$C_n = \frac{1}{n} \{0 \cdot 1 + w \cdot (n - 1)\} = \left(1 - \frac{1}{n}\right) \cdot w$$

となる。

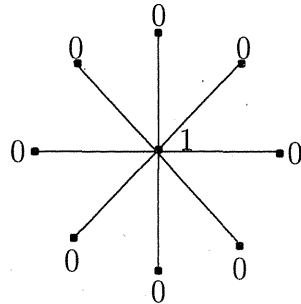


図 3.1: $S_n = 0$ の場合

次に、 $2 \leq S_n \leq n - 1$ の場合には、 $X_i = 1$ となっている S_n 個の頂点は、それぞれ自分以外の $n - 1$ 個の頂点と辺で結ばれている。一方、 $X_i = 0$ となっている $n - S_n$ 個の頂点は、 $X_i = 1$ となっている S_n 個の頂点としか辺で結ばれていない(図3.2)。従って、

$$\begin{aligned} C_n &= \frac{1}{n} \left\{ \frac{\binom{n-1}{2} - \binom{n-S_n}{2}}{\binom{n-1}{2}} \cdot S_n + \frac{\binom{S_n}{2}}{\binom{S_n}{2}} \cdot (n - S_n) \right\} \\ &= 1 - \frac{(n - S_n)(n - 1 - S_n)S_n}{n(n - 1)(n - 2)} \end{aligned}$$

となる。

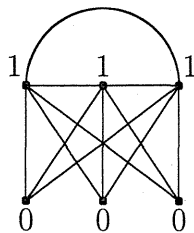


図 3.2: $2 \leq S_n \leq n - 1$ の場合

最後に, $S_n = n$ の場合には, 完全グラフとなるので $C_n = 1$ である.

以上の結果を用いて,

$$1 - \frac{(n - S_n)(n - 1 - S_n)S_n}{n(n - 1)(n - 2)} = \begin{cases} 1, & (S_n = 0), \\ 1 - \frac{1}{n}, & (S_n = 1), \\ 1 - \frac{(n - S_n)(n - 1 - S_n)S_n}{n(n - 1)(n - 2)}, & (2 \leq S_n \leq n), \end{cases}$$

であることに注意すると,

$$\begin{aligned} C_n &= \left[1 - \frac{(n - S_n)(n - 1 - S_n)S_n}{n(n - 1)(n - 2)} \right] \cdot \{I_{\{2, \dots, n\}}(S_n) + w \cdot I_{\{0, 1\}}(S_n)\} \\ &= \left[1 - \left(1 - \frac{S_n}{n}\right) \left(1 - \frac{S_n}{n - 1}\right) \left(\frac{S_n}{n - 2}\right) \right] \cdot \{1 + (w - 1) \cdot I_{\{0, 1\}}(S_n)\} \\ &\rightarrow 1 - (1 - p)^2 p = \mathbb{E}[C(1)], \quad \text{a.s. } (n \rightarrow \infty). \end{aligned}$$

となることがわかり, 定理 3.0.5 が確かめられた. ここで, 最後の収束には, 大数の強法則により, 確率 1 で $S_n/n \rightarrow p$ となることを使った.

第4章 しきい値モデルの階層構造

この章では、しきい値モデルで生成されるグラフが階層構造を持つことを示す。しきい値モデルでは、各頂点の重みが決まるごとにグラフが生成される。ここで、各頂点をその重みによって2種類に分類することを考える。つまり、与えられたしきい値 θ に対して、重みが $\theta/2$ を超える頂点からなる集合 A と、そうでない頂点からなる集合 B とに分類する。

まず、 A から任意に2つの頂点を選んで、その重みをそれぞれ x, y とする。ここで、 $x, y > \theta/2$ となっていることに注意すると、 $x + y > \theta$ となることがわかる。このことから、 A の任意の2頂点は辺で結ばれていることがわかる。つまり、重みが $\theta/2$ を超えている頂点同士は互いに辺で結ばれてクリークを形成する。同じ議論で、 B のどんな2頂点も辺で結ばれないことがわかる。つまり、 B から任意に2つの頂点を選んで、その重みを z, w とすると $z, w \leq \theta/2$ となっているので、 $z + w \leq \theta$ となってしまう、辺で結ばれないのである。

それでは、 A と B の頂点同士の結びつきはどうなっているだろうか。 B に属する頂点で、孤立点になっていないものを集めて C とする。つまり、 C の各頂点は A に属する頂点の一部とだけ辺で結ばれている。ここで、頂点 $u_* \in C$ を C の中で最も重みの小さい頂点とする。もし、そのような頂点が複数あるときには、その中で頂点の番号が一番小さいものを選ぶことにする。いま、 $v \in A$ と u_* が辺で結ばれているとする。即ち、 u_* と v の重みの和が θ を超えているとすると、任意の $w \in C$ の重みは u_* の重み以上になっていることに注意すれば、 v と w の重みの和が θ を超えることになり、 v と w は辺で結ばれる。つまり、 C に属する全ての頂点と結ばれる、 A に属する頂点が少なくとも1つ存在することがわかる(図4.1)。従って、図4.2のように繋がることはないことに注意する。

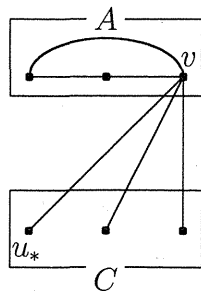


図 4.1: 許される繋がり方

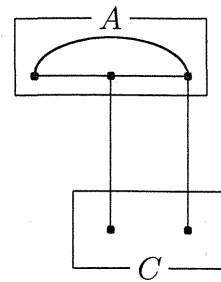


図 4.2: 許されない繋がり方

以上をまとめると、しきい値モデルには図4.3のような階層構造があることがわかる。つまり、 A はクリーク、 B は孤立点の集合と孤立点以外の頂点の集合(= C)からなり、 A と C が結ばれている。 A に属する頂点同士はクリークなので距離1でつながり、 C に属する頂点同士は5と6が3を介してつながっているように、 A に属する頂点を介して距離2でつな

がることがわかる。さらに、 A に属する頂点と C に属する頂点は3と5のように辺で直接結ばれているか、または、1と6が2を介して結ばれるように、共通の隣接する頂点を介して距離2で結ばれている。上記以外の頂点は孤立点となっている。これより、連結成分だけに注目すれば平均頂点間距離が1と2の間にあることがわかる。

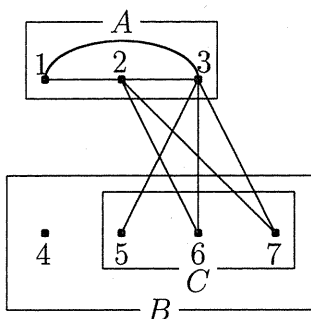


図 4.3: 階層構造

しきい値モデルには、上述のような階層構造があることがわかったが、より詳しい構造を知ることが出来る [25]. 頂点総数が n のとき、重みを与える確率変数 X_1, X_2, \dots, X_n の実現値が、それぞれ x_1, x_2, \dots, x_n で与えられているものとする。このとき、重みを小さい順に並べ替えて、 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ となるように番号を付けなおす。ここで、 $x_{(1)} + x_{(n)} > \theta$ であるとき、しきい値モデルの定義から、 $x_{(n)}$ に対応する頂点は、他の $n-1$ 個の頂点と辺で結ばれることになる。なぜなら、

$$\theta < x_{(1)} + x_{(n)} \leq x_{(2)} + x_{(n)} \leq \dots \leq x_{(n-1)} + x_{(n)}$$

が成り立っているからである。逆に、 $x_{(1)} + x_{(n)} \leq \theta$ であるときには、 $x_{(1)}$ に対応する頂点は、他の $n-1$ 個の頂点のどれとも辺で結ばれず、孤立してしまう。なぜなら、

$$x_{(1)} + x_{(2)} \leq x_{(1)} + x_{(3)} \leq \dots \leq x_{(1)} + x_{(n)} \leq \theta$$

が成り立つからである。

以上より、 $x_{(1)} + x_{(n)}$ が θ を超えれば $x_{(n)}$ に対応する頂点はそれ以外の全ての頂点とつながり、スターとなる。逆に、 $x_{(1)} + x_{(n)}$ が θ を超えない場合には $x_{(1)}$ に対応する頂点は孤立点となることがわかる。このようにして、 n 個の頂点のうちでスターになる頂点か孤立点になる頂点かのどちらか1つが決まる。スターか孤立点かのどちらかが決まったら、その頂点を除いた $n-1$ 個の頂点について考えると、同じ議論で、最大の重みと最小の重みの和がしきい値を超えるかどうかに応じて、 $n-1$ 個の頂点からなる部分グラフでスターになる頂点か孤立点になる頂点かのどちらか1つが決まる。このように、しきい値モデルで生起可能なグラフの頂点には、ある部分グラフの中でスターとなっているスター的な頂点と、ある部分グラフで孤立点となっている孤立点的な頂点の二種類が存在することがわかる。

この性質を使ってスター的な頂点には値1を、孤立点的な頂点には値0を与えて、しきい値モデルで生起可能なグラフに対してラベル付けを行う。ただし、スター的・孤立点的とい

う考え方は、他に比較する頂点がある場合に有効であるので、最後の1個に対しては値を決めることができない。そのため、ここでは最後の1つには必ず値1を与えることにする。それぞれの頂点に値を与えて、グラフ全体のラベルを決めるわけだが、その表記法も工夫が必要である。ここでは、後の議論のために n 個の $0, 1$ の組 $(g_0, g_1, \dots, g_{n-1})$ で表記することにしよう。それぞれの g_i は $0, 1$ のどちらかの値をとり、値が決まった順に右から並べることとする。但し、 $g_0 = 1$ となっていることに注意が必要である。以下では、この $(g_0, g_1, \dots, g_{n-1})$ のことを構造ベクトルと呼ぶことにする。

例として、以下で $n = 4$ の場合の構造ベクトルを実際に求める。この場合は、 $g_0 = 1$ かつ、 $g_1, g_2, g_3 \in \{0, 1\}$ であるので、8通りの構造ベクトルが考えられる。以下では、重み $x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq x_{(4)}$ が与えられているものとする。

まず、 $x_{(1)} + x_{(4)} > \theta$ の場合には、 $x_{(4)}$ に対応する頂点がスターとなっている。従って、この頂点に値1を与える。既に、 $x_{(4)}$ に対応する頂点はスターであることがわかったので、それを除いた $x_{(1)} \leq x_{(2)} \leq x_{(3)}$ で議論を続ける。

(i) $x_{(1)} + x_{(3)} > \theta$ の場合。この場合には、 $x_{(3)}$ に対応する頂点はスター的である。つまり、 $x_{(4)}$ に対応する頂点を除いた3つの頂点の中でスターになっている。従って、この頂点に値1を与える。既に、 $x_{(3)}$ に対応する頂点はスター的であることがわかったので、それを除いた $x_{(1)} \leq x_{(2)}$ で議論を続ける。

(a) $x_{(1)} + x_{(2)} > \theta$ の場合。この場合、 $x_{(2)}$ に対応する頂点に値1を与えて、最後に残った $x_{(1)}$ には値1を与える。従って、与えられる構造ベクトルは $(1, 1, 1, 1)$ である。

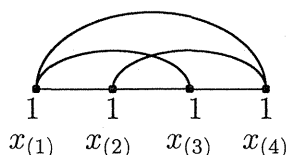


図 4.4: $(1, 1, 1, 1)$ に対応するグラフ

(b) $x_{(1)} + x_{(2)} \leq \theta$ の場合。この場合、 $x_{(1)}$ に対応する頂点に値0を与えて、最後に残った $x_{(2)}$ には値1を与える。従って、与えられる構造ベクトルは $(1, 0, 1, 1)$ である。

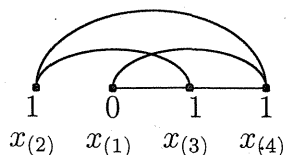
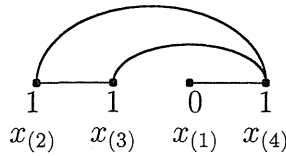


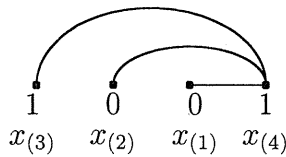
図 4.5: $(1, 0, 1, 1)$ に対応するグラフ

(ii) $x_{(1)} + x_{(3)} \leq \theta$ の場合。この場合には、 $x_{(1)}$ に対応する頂点は孤立点的である。つまり、 $x_{(4)}$ に対応する頂点を除いた3つの頂点の中で孤立している。従って、この頂点に値0を与える。既に、 $x_{(1)}$ に対応する頂点は孤立点的であることがわかったので、それを除いた $x_{(2)} \leq x_{(3)}$ で議論を続ける。

(a) $x_{(2)} + x_{(3)} > \theta$ の場合。この場合、 $x_{(3)}$ に対応する頂点に値1を与えて、最後に残った $x_{(2)}$ には値1を与える。従って、与えられる構造ベクトルは $(1, 1, 0, 1)$ である。

図 4.6: $(1, 1, 0, 1)$ に対応するグラフ

- (b) $x_{(2)} + x_{(3)} \leq \theta$ の場合. この場合, $x_{(2)}$ に対応する頂点に値 0 を与えて, 最後に残った $x_{(3)}$ には値 1 を与える. 従って, 与えられる構造ベクトルは $(1, 0, 0, 1)$ である.

図 4.7: $(1, 0, 0, 1)$ に対応するグラフ

次に, $x_{(1)} + x_{(4)} \leq \theta$ の場合は, $x_{(1)}$ に対応する頂点が孤立点となっている. 従って, この頂点に値 0 を与える. 既に, $x_{(1)}$ に対応する頂点は孤立点であることがわかったので, それを除いた $x_{(2)} \leq x_{(3)} \leq x_{(4)}$ で, $x_{(1)} + x_{(4)} > \theta$ の場合と同様の議論を続けることで, 構造ベクトルを決めることができる.

以上をまとめると, 頂点総数が $n = 4$ の場合にしきい値モデルで生成可能なグラフは図 4.8 の 8 種類であることがわかる. ここで, 図 4.9 のサイクル C_4 は生起しないことに注意しよう. しきい値モデルで生起されるグラフの平均頂点間距離は, 連結であれば 2 以下である. この C_4 は, 平均頂点間距離が 2 以下であるが, しきい値モデルで生起できないグラフの例となっている.

一般の場合には, 以下のようなアルゴリズムで, 構造ベクトルを求めることができる.

- (i) $w(n) = \{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$ とする.
- (ii) $M = \max\{j : x_{(j)} \in w(n)\}$, $m = \min\{j : x_{(j)} \in w(n)\}$ として, $x_{(m)} + x_{(M)}$ の値に応じて次のいずれかを行う.
 - (a) $x_{(m)} + x_{(M)} > \theta$ ならば, $g_{n-1} = 1$ として, $w(n)$ から $x_{(M)}$ を取り除いた集合を $w(n-1)$ とする. このとき, $x_{(M)}$ に対応する頂点に $g_{n-1} = 1$ を割り当てたことになる. この頂点は $w(n-1)$ に属する重みに対応する全ての頂点と辺で結ばれている.
 - (b) $x_{(m)} + x_{(M)} \leq \theta$ ならば, $g_{n-1} = 0$ として, $w(n)$ から $x_{(m)}$ を取り除いた集合を $w(n-1)$ とする. このとき, $x_{(m)}$ に対応する頂点に $g_{n-1} = 0$ を割り当てたことになる. この頂点は $w(n-1)$ に属する重みに対応する全ての頂点から孤立している.
- (iii) 次に, n を $n-1$ と読み替えて (ii) と同様の操作を行うことで g_{n-2} が決まり, 同時に $w(n-2)$ を得る. このように, (ii) の操作を帰納的に $w(1)$ となるまで繰り返す.
- (iv) $w(1)$ となったら, $g_0 = 1$ として $(g_0, g_1, \dots, g_{n-1})$ を得る.

上述のアルゴリズムによって, 重みの組 x_1, x_2, \dots, x_n が与えられるごとに構造ベクトル $(g_0, g_1, \dots, g_{n-1})$ を求めることができる. この構造ベクトルへの対応付けは一意的ではある

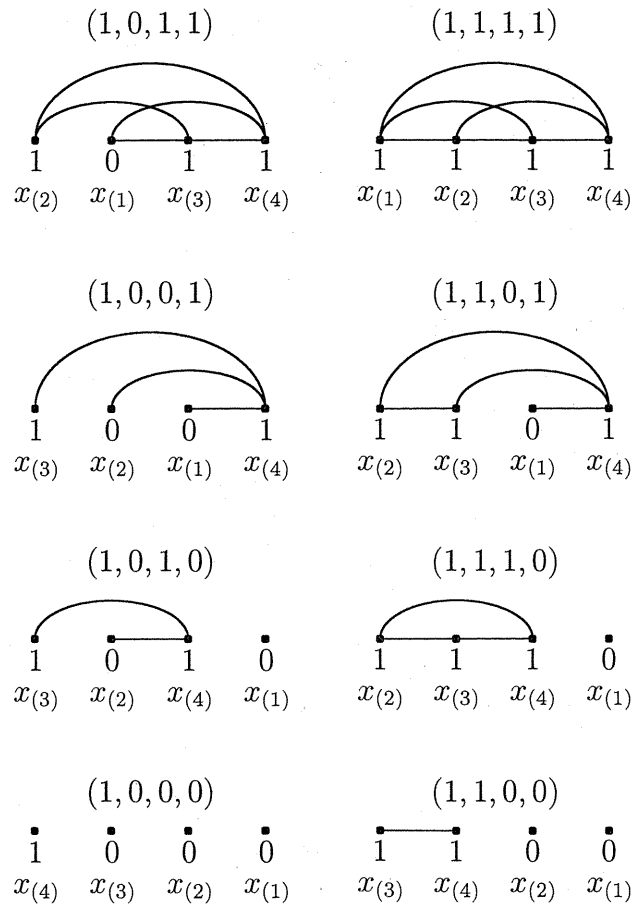


図 4.8: しきい値モデルで生起可能なグラフ ($n = 4$)

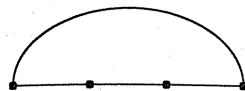


図 4.9: 生起可能でないグラフの例 ($n = 4$)

が、複数の異なる重みの組が同じ構造ベクトルに対応することもある。これは、2つの頂点間を辺で結ぶかどうかを、それぞれの重みの和がしきい値を超えるかどうかで判断していることに起因する。例えば、孤立点になっている頂点を v とすると、 v の重みをより小さい重みと取り替えてグラフを作っても、 v は孤立点である。このとき、重みの組は異なっているが、対応する構造ベクトルは全く同じものとなっている。

階層構造を見るために、しきい値モデルで生起可能なグラフとその構造ベクトルとの対応を考える。構造ベクトル $(g_0, g_1, \dots, g_{n-1})$ が与えられたとき、 n 個の各頂点には0または1が対応付けられている。このとき、上記アルゴリズムの(ii)により、 $g_i = 1$ に対応している頂点は、 g_0, g_1, \dots, g_{i-1} に対応する i 個のどの頂点とも辺で結ばれている。逆に、 $g_i = 0$ に対応する頂点は、 g_0, g_1, \dots, g_{i-1} に対応する i 個のどの頂点とも辺で結ばれない。より詳しく見ると、 $g_i = 0$ に対応する頂点は、 $g_{i+1}, g_{i+2}, \dots, g_{n-1}$ に対応する頂点の中で、1が対応付けられている頂点のみと辺で結ばれている。ここで、1が対応付けられている頂点だけに注目すると、それぞれが、自分より左にあるどの頂点とも辺で結ばれるので、クリークを形成することがわかる。しかも、これは最大クリーク（クリークの中で、頂点数が最も多くなっているもの）の1つとなっている。例えば、図4.10では、値1が割り当てられている g_0, g_1, g_2, g_3 に対応する頂点が最大クリークを形成している。また、右端から1を挟まずに連続する0は孤立点に対応している。例えば、図4.10では、 g_4, g_5 に対応する頂点は孤立点となっている。

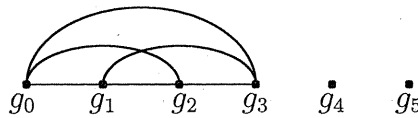


図 4.10: $(1, 1, 1, 1, 0, 0)$ に対応するグラフ

ここで、図4.11のように、 $g_1 = 0$ となっている場合は最大クリークが複数個存在することに注意が必要である。この場合、 g_0, g_3, g_4 に対応する頂点が最大クリークを形成しているが、 g_1, g_3, g_4 に対応する頂点と g_2, g_3, g_4 に対応する頂点もまた最大クリークを形成しているので、最大クリークは3個存在している。これは、構造ベクトルを決める際に最後の g_0 は0と1のどちらの値にするかをアルゴリズムでは決められず、勝手に1と決めたことによる。一般に、 $g_1 = \dots = g_k = 0$ のように、 g_1 から g_k までの値が全て0で、 $g_{k+1} = 1$ となっている場合は、 k 個の0と $g_0 = 1$ は同等の役割となり、最大クリークが $k+1$ 個存在することとなる。その理由は以下のように説明できる。まず、 g_0, g_1, \dots, g_k に対応する $k+1$ 個の頂点は、 g_{k+1}, \dots, g_n の中で、その値が1となっている頂点を作るクリーク K に属するどの頂点とも繋がっている。従って、 K に g_0, g_1, \dots, g_k に対応するどの頂点を1つ付け加えてもクリークを形成することがわかる。しかし、 g_0, g_1, \dots, g_k に対応する頂点同士は辺で結ばれていないので、このようにできた $k+1$ 個のクリークが最大クリークである。

階層性を詳しく見るために、構造ベクトルを右端から見て最初に現れる連続した0,1の列 $1, \dots, 1, 0, \dots, 0$ をまとめてレベル1と呼ぶことにする。もし、右端が1である場合はレベル1は $1, \dots, 1$ であるとする。その後、 $1, \dots, 1, 0, \dots, 0$ が出てくる度に、レベル2、レベル

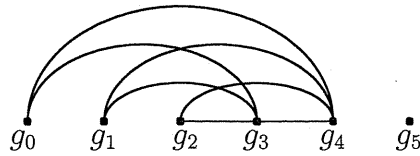


図 4.11: $(1, 0, 0, 1, 1, 0)$ に対応するグラフ

3のように名前を付けていく．つまり，下記のように名前付けを行う．

$$\overbrace{(1, \dots, 1, 0, \dots, 0, \dots)}^{\text{レベル } k}, \overbrace{(1, \dots, 1, 0, \dots, 0, \dots)}^{\text{レベル } 2}, \overbrace{(1, \dots, 1, 0, \dots, 0, \dots)}^{\text{レベル } 1},$$

または,

$$\overbrace{(1, \dots, 1, 0, \dots, 0, \dots)}^{\text{レベル } k}, \overbrace{(1, \dots, 1, 0, \dots, 0, \dots)}^{\text{レベル } 2}, \overbrace{(1, \dots, 1)}^{\text{レベル } 1}.$$

このとき，レベル1，レベル2，レベル3，のような順番で順序付けられているものとする．つまり，レベル1が最も「高い」レベルであり，レベル2がその次に高いレベル，といった順序付けを考える．例えば，図4.10の場合は，構造ベクトルが $(1, 1, 1, 1, 0, 0)$ なので，レベル1の頂点しかなく，図4.10の場合は，構造ベクトルが $(1, 0, 0, 1, 1, 0)$ なので，レベル1に g_3, g_4, g_5 に対応する3つの頂点が属し，レベル2に残りの3つの頂点が属している．

この対応関係により，0に対応する頂点は，自分より高いレベルにある1に対応する頂点のみと辺で結ばれていることが分かる．一方，1に対応する頂点は，1に対応する頂点全てと，自分より低いレベルにある0に対応する頂点と結ばれている（図4.12）．

このことから，しきい値モデルで生起可能なグラフには，2個以上の頂点からなる連結成分は高々1個しか許されないことも同時にわかる．なぜなら，レベル1の0に対応する頂点は孤立点であり，それ以外の全ての頂点は，スター的なレベル1の1に対応する頂点と結ばれているからである．

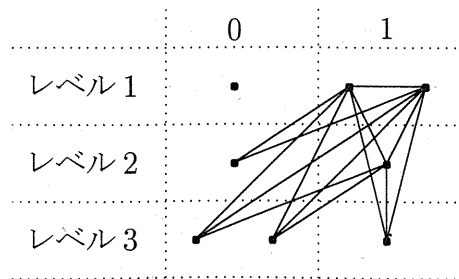


図 4.12: $(1, 0, 0, 1, 0, 1, 1, 0)$ に対応するグラフ

このように，しきい値モデルで生起可能なグラフは階層構造を持つことがわかったが，実際にしきい値モデルで生起可能なグラフはどのような分布になっているだろうか．この問いに対する一般的な答えはまだ得られていないが，重みが特定の分布に従う場合には簡単に知ることができる．以下で，例を1つ挙げる．

例 4.0.5

頂点総数が n で、重み $\{X_i\}_{i=1,2,\dots,n}$ を、 $\mathbb{P}(X_1 = 1) = p$, $\mathbb{P}(X_1 = 0) = 1 - p$ を満たす独立同分布な確率変数列とする。ここで、しきい値 θ を $0 \leq \theta < 1$ を満たすように固定して、しきい値モデルを考える。つまり、この場合には $X_i = X_j = 0$ の場合のみ頂点 i と頂点 j は辺で結ばれないことになる。 $S_n = \sum_{i=1}^n X_i$ とすると、 S_n は n 個の頂点の中で重みの値が 1 となっている頂点の数を表すことになる。さらに、定義から S_n は二項分布 $B(n, p)$ に従う確率変数である。即ち、 $k = 0, 1, \dots, n$ に対して

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

が成り立っている。以下、 $S_n = k$ として考える。

まず、 $0 \leq k \leq n-2$ の場合を考える。この場合、 k 個の頂点の重みが 1 で $n-k$ 個の頂点の重みが 0 である。従って、重みを小さい順に並べ替えると

$$\overbrace{0, \dots, 0}^{n-k}, \overbrace{1, \dots, 1}^k$$

となる。この重みの列に対して上述のアルゴリズムを適用すると、 $0+1 > \theta$ であることから、最初の k 回では値 1 が与えられて、重み 1 が取り除かれることがわかる。そうして、重み 1 がなくなると重みは 0 だけになるが、 $0+0 \leq \theta$ であることから、残りの $n-k-1$ 回では値 0 が与えられることになる。つまり、このとき与えられる構造ベクトルは

$$(1, \overbrace{0, \dots, 0}^{n-k-1}, \overbrace{1, \dots, 1}^k)$$

となる。従って、上記の構造ベクトルが与えられる確率は

$$\binom{n}{k} p^k (1-p)^{n-k}$$

となる。

次に、 $k = n-1$ の場合を考える。この場合は、 $k = n-1$ 個の頂点の重みが全て 1 で、重み 0 の頂点が 1 つだけの場合である。従って、重みを小さい順に並べ替えると

$$0, \overbrace{1, \dots, 1}^{n-1}$$

となる。このとき、上の場合と同様に $k = n-1$ 回だけ値 1 が与えられて、重み 1 が取り除かれることがわかる。すると、最後に重み 0 が 1 つだけ残されることになるが、この重みに対応する値は最後の 1 つなので 1 である。従って、この場合に与えられる構造ベクトルは

$$(1, \dots, 1)$$

となる。また、 $k = n$ の場合も全ての頂点に値 1 が与えられることが容易にわかるので、構造ベクトル $(1, \dots, 1)$ が与えられる確率は

$$\binom{n}{n-1} p^{n-1} (1-p) + \binom{n}{n} p^n = np^{n-1} (1-p) + p^n$$

となる.

以上をまとめると,

$$\mathbb{P} \left(G_n = (1, \overbrace{0, \dots, 0}^{n-k-1}, \overbrace{1, \dots, 1}^k) \right) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k = 0, 1, \dots, n-2, \\ np^{n-1}(1-p) + p^n, & k = n-1, \end{cases}$$

となることがわかる. ここで, 上記の $G_n = (\text{構造ベクトル})$ と表される事象は, 頂点総数が n のしきい値モデルがその構造ベクトルで表されるグラフを生起する事象を表している.

第5章 しきい値モデルの平均頂点間距離

前章で、しきい値モデルの階層構造を議論したが、その際にしきい値モデルには、孤立点を除くと距離1の頂点对と距離2の頂点对しか存在しないことを得た。この章では、その事実を用いて、平均頂点間距離の極限定理を議論する。そのために、以下で平均頂点間距離をもう一度定義する。

頂点総数が n のとき、 $N_n^{(0)}$ を孤立点の個数とする。平均頂点間距離を計算するには距離1となっている頂点对の個数と、距離2となっている頂点对の個数と、距離が ∞ となっている頂点对の個数が必要である。ここで、距離1となっている頂点对の個数は辺の総数と同じであるので D_n (定理 2.0.3 参照) である。また、距離が ∞ となっている頂点对の個数 $N_n^{(\infty)}$ は、孤立点から2つの頂点を選ぶか、または孤立点とそれ以外の頂点から1つずつ選べばよいので、

$$N_n^{(\infty)} = \binom{N_n^{(0)}}{2} + N_n^{(0)}(n - N_n^{(0)}) = \frac{N_n^{(0)}(2n - N_n^{(0)} - 1)}{2}$$

である。最後に、距離2の頂点对の数は距離1の頂点对と距離 ∞ の頂点对の数を加えて、可能な頂点对の総数 $n(n-1)/2$ から引くことで得られる。ここで、本来の平均頂点間距離の定義では、距離 ∞ の頂点对が存在するとその値が ∞ となってしまうことに注意する。その状況を避けるために、ここでは定義を少し変更して、距離 ∞ の頂点对が平均頂点間距離に与える寄与を不定元 w で表すことにする。この変更により、平均頂点間距離 L_n は

$$\begin{aligned} L_n &= \frac{1}{\binom{n}{2}} \left\{ 1 \times D_n + 2 \times \left(\binom{n}{2} - D_n - N_n^{(\infty)} \right) + w \times N_n^{(\infty)} \right\} \\ &= 2 - \frac{D_n}{\binom{n}{2}} - (2-w) \frac{N_n^{(\infty)}}{\binom{n}{2}} \\ &= 2 - \frac{D_n}{\binom{n}{2}} - (2-w) \frac{N_n^{(0)}(2n - N_n^{(0)} - 1)}{n(n-1)} \end{aligned} \quad (5.1)$$

で与えられる。

平均頂点間距離の極限定理を得るためには、 D_n と $N_n^{(0)}$ に関する極限定理を示せば十分であるが、定理 2.0.3 で D_n に関する極限定理は得られている。よって、ここでは $N_n^{(0)}$ に関する極限定理を示す。そのために、任意の $x \in \mathbb{R}$ に対して

$$N_n^{(0)}(i; x) = \prod_{\substack{1 \leq j \leq n \\ j \neq i}} (1 - h_e(x, X_j))$$

を定義する. ここで, $h_e(x, y) = I_{(\theta, \infty)}(x+y)$ と定めた. 定義より, $N_n^{(0)}(i; x)$ は, 頂点 i の重みの値が x であるときに, 頂点 i 以外の全ての頂点との間に辺が無い場合だけ 1 となり, それ以外の場合には 0 となる関数である. つまり, $N_n^{(0)}(i) = N_n^{(0)}(i; X_i)$ とおくと, $N_n^{(0)} = \sum_{i=1}^n N_n^{(0)}(i)$ と表せる. ここで, 以下の極限定理が成り立つことを示す.

定理 5.0.6

(i) 任意の $x \in \mathbb{R}$ に対して,

$$\lim_{n \rightarrow \infty} N_n^{(0)}(1; x) = N^{(0)}(1; x) = I_{(-\infty, \theta - x^*]}(x), \quad \text{a.s.}$$

(ii) 特に,

$$\lim_{n \rightarrow \infty} N_n^{(0)}(1) = N^{(0)}(1; X_1) = I_{(-\infty, \theta - x^*]}(X_1), \quad \text{a.s.}$$

但し, $x^* = \sup\{x \in \mathbb{R} : F(x) < 1\}$ である.

証明

任意に $x \in \mathbb{R}$ を固定する. 定義から,

$$\begin{aligned} \mathbb{P}(N_n^{(0)}(1; x) = 1) &= \mathbb{P}\left(\prod_{2 \leq j \leq n} [1 - h_e(x, X_j)] = 1\right) \\ &= \mathbb{P}\left(\bigcap_{2 \leq j \leq n} \{x + X_j \leq \theta\}\right) \end{aligned}$$

となることに注意する. ここで, X_2, X_3, \dots, X_n は独立同分布なので,

$$\begin{aligned} \mathbb{P}(N_n^{(0)}(1; x) = 1) &= \prod_{2 \leq j \leq n} \mathbb{P}(x + X_j \leq \theta) = \mathbb{P}(x + X_2 \leq \theta)^{n-1} \\ &= F(\theta - x)^{n-1} \end{aligned}$$

である. 事象列 $\{N_n^{(0)}(1; x) = 1\}$ は n に関して単調非増加であるので,

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} N_n^{(0)}(1; x) = 1\right) = \lim_{n \rightarrow \infty} \mathbb{P}(N_n^{(0)}(1; x) = 1) = \lim_{n \rightarrow \infty} F(\theta - x)^{n-1}$$

となる. 従って,

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} N_n^{(0)}(1; x) = 1\right) = \begin{cases} 1, & \text{但し, } F(\theta - x) = 1, \\ 0, & \text{但し, } F(\theta - x) < 1, \end{cases}$$

である. 最後に $N_n^{(0)}(1; x)$ は 0 か 1 しかとらないことに注意すれば,

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} N_n^{(0)}(1; x) = 0\right) = \begin{cases} 1, & \text{但し, } F(\theta - x) < 1, \\ 0, & \text{但し, } F(\theta - x) = 1, \end{cases}$$

であることも同時にわかる。つまり、確率1で

$$\lim_{n \rightarrow \infty} N_n^{(0)}(1; x) = \begin{cases} 1, & \text{但し, } F(\theta - x) = 1, \\ 0, & \text{但し, } F(\theta - x) < 1, \end{cases}$$

である。ここで、分布関数 F が $x \in \mathbb{R}$ に関して単調非減少であり、また、右連続であることから、

$$\{x \in \mathbb{R} : F(\theta - x) = 1\} = \{x \in \mathbb{R} : x^* \leq \theta - x\} = (-\infty, \theta - x^*] \quad (5.2)$$

がわかる。従って、 $N^{(0)}(1; x) = I_{(-\infty, \theta - x^*]}(x)$ と定義すれば、

$$\lim_{n \rightarrow \infty} N_n^{(0)}(1; x) = N^{(0)}(1; x) = I_{(-\infty, \theta - x^*]}(x), \quad \text{a.s.}$$

が示される。

また (ii) は、 F を X_1 の分布関数とすると、

$$\begin{aligned} & \mathbb{P}\left(\lim_{n \rightarrow \infty} N_n^{(0)}(1; X_1) = N^{(0)}(1; X_1)\right) \\ &= \int_{\mathbb{R}} \mathbb{P}\left(\lim_{n \rightarrow \infty} N_n^{(0)}(1; x) = N^{(0)}(1; x) \mid X_1 = x\right) F(dx) \\ &= \int_{\mathbb{R}} 1 \cdot F(dx) = 1 \end{aligned}$$

となる。但し、最初の等号は独立性から、二番目の等号は (i) より導かれる。

□

このことを使って、 $N_n^{(0)}$ に関しても次の極限定理を得る。但し、証明は定理 3.0.5 と同じ議論を適用すればよいので省略する。

定理 5.0.7

$x^* = \sup\{x \in \mathbb{R} : F(x) < 1\}$ とする。このとき、以下が成り立つ。

$$\lim_{n \rightarrow \infty} \frac{N_n^{(0)}}{n} = \mathbb{E}[N^{(0)}(1; X_1)] = F(\theta - x^*), \quad \text{a.s.}$$

定理 2.0.3 と定理 5.0.7 により、式 (5.1) は n を無限大にする極限で、

$$\begin{aligned} L_n &= 2 - \frac{D_n}{\binom{n}{2}} - (2 - w) \frac{N_n^{(0)}(2n - N_n^{(0)} - 1)}{n(n-1)} \\ &\rightarrow 2 - \mathbb{P}(X_1 + X_2 > \theta) - (2 - w)F(\theta - x^*)\{2 - F(\theta - x^*)\}, \quad \text{a.s.} \end{aligned}$$

となる。ここで、定理 5.0.7 より $n \rightarrow \infty$ で、

$$\begin{aligned} \frac{N_n^{(0)}(2n - N_n^{(0)} - 1)}{n(n-1)} &= \frac{N_n^{(0)}}{n} \frac{2(n-1) - N_n^{(0)} + 1}{n-1} \\ &\rightarrow F(\theta - x^*)\{2 - F(\theta - x^*)\} \end{aligned}$$

が確率1で成り立つことを使った。よって、以下の定理を得る。

定理 5.0.8

しきい値モデルの平均頂点間距離に関して次の極限定理が成立する.

$$\lim_{n \rightarrow \infty} L_n = 2 - \mathbb{P}(X_1 + X_2 > \theta) - (2 - w)F(\theta - x^*)\{2 - F(\theta - x^*)\}, \quad \text{a.s.}$$

但し, $x^* = \sup\{x \in \mathbb{R} : F(x) < 1\}$ である.

式(5.2)より, $Q = \{x \in \mathbb{R} : F(\theta - x) = 1\}$ とすると,

$$F(\theta - x^*) = \int_Q F(dx) \quad (5.3)$$

が成り立つことに注意すると, 定理 5.0.8 の結果は以下の極端な場合に, 直観と一致していることが確認できる.

(i) $\mathbb{P}(X_1 + X_2 > \theta) = 1$ の場合.

$$\begin{aligned} 1 &= \mathbb{P}(X_1 + X_2 > \theta) = 1 - \mathbb{P}(X_1 + X_2 \leq \theta) \\ &= 1 - \int_{\mathbb{R}} F(\theta - x) F(dx) \end{aligned}$$

であることから,

$$\int_{\mathbb{R}} F(\theta - x) F(dx) = 0$$

が得られる. ここで, 式(5.3)より,

$$\begin{aligned} F(\theta - x^*) &= \int_Q F(dx) = \int_Q F(\theta - x) F(dx) \\ &\leq \int_{\mathbb{R}} F(\theta - x) F(dx) = 0 \end{aligned}$$

が得られる. ここで, 2番目の等号では, Q 上で $F(\theta - x) = 1$ となることを使った. 従って, 分布関数は非負であることから, $F(\theta - x^*) = 0$ を得る. このことから,

$$\lim_{n \rightarrow \infty} L_n = 2 - 1 - (2 - w) \times 0 \times 2 = 1, \quad \text{a.s.}$$

となる. これは, $\mathbb{P}(X_1 + X_2 > \theta) = 1$ の場合に完全グラフとなることを反映した結果である.

(ii) $\mathbb{P}(X_1 + X_2 \leq \theta) = 1$ の場合.

$$1 = \mathbb{P}(X_1 + X_2 \leq \theta) = \int_{\mathbb{R}} F(\theta - x) F(dx)$$

であるので, $\int_Q F(dx) < 1$ とすると, $\mathbb{R} \setminus Q = \{x \in \mathbb{R} : F(\theta - x) < 1\}$ に気をつければ,

$$\begin{aligned} 1 &= \int_{\mathbb{R}} F(\theta - x) F(dx) \\ &= \int_Q F(\theta - x) F(dx) + \int_{\mathbb{R} \setminus Q} F(\theta - x) F(dx) \\ &< \int_Q F(dx) + \int_{\mathbb{R} \setminus Q} F(dx) = 1 \end{aligned}$$

となり矛盾. ここで, 最後の不等号では, Q 上で $F(\theta - x) = 1$ であることと, $\mathbb{R} \setminus Q$ 上で $F(\theta - x) < 1$ であることを用いた. 従って, $F(\theta - x^*) = \int_Q F(dx) = 1$ を得る. よって,

$$\lim_{n \rightarrow \infty} L_n = 2 - 0 - (2 - w) \times 1 \times (2 - 1) = 2 - (2 - w) = w, \quad \text{a.s.}$$

となる. これは, $\mathbb{P}(X_1 + X_2 \leq \theta) = 1$ の場合に空グラフとなり, 全ての頂点が孤立点となることを反映している.

第6章 しきい値モデルの一般化

この節では、しきい値モデルの一般化を考える。\$d\$次元ユークリッド空間 \$\mathbb{R}^d\$ に値をとる、独立同分布な確率変数列を \$X_1, \dots, X_n\$ として、その共通の分布関数を \$F\$ とする。ここで、\$l\$ 個のボレル可測関数 \$f_c^{l'} : (\mathbb{R}^d)^2 \to \mathbb{R}\$, \$l' \in \{1, \dots, l\}\$ で、任意の \$x, y \in \mathbb{R}^d\$ に対して \$f_c^{l'}(x, y) = f_c^{l'}(y, x)\$ を満たすものを考える。さらに、1次元ボレル集合のある族 \$\mathcal{C} = \{B_1, \dots, B_l\}\$ を用意する。このとき、異なる頂点 \$i\$ と \$j\$ は、全ての \$l' \in \{1, \dots, l\}\$ に対して、\$f_c^{l'}(X_i, X_j) \in B_{l'}\$ が成り立つときに限って辺で結ばれるというルールで、一般化されたしきい値モデルを定義する。もとのしきい値モデルは、\$d = l = 1\$ で \$f_c^1(x, y) = x + y\$, \$B_1 = (\theta, \infty)\$ とした場合に対応している。

ここではまず、一般化されたしきい値モデルで生成されるグラフに現れる部分グラフに関する極限定理を述べる。そのために必要な概念を以下で導入する。

まず、\$m \in \{2, \dots, n\}\$ 個の頂点からなるグラフ \$H\$ を考える。このとき、\$m\$ 個の頂点からなるグラフ \$H'\$ が \$H\$ と同型であることを、\$H' \sim H\$ と表すことにする。今、\$\mathcal{A}_H \sim = \{H' : H' \sim H\}\$ と定義すると、\$\mathcal{A}_H\$ は \$H\$ と同型なグラフ全体を表す集合となる。場合によっては、\$m\$ 個の頂点からなるグラフのある族 \$\{H_i\}_{i \in I}\$ に含まれるグラフに同型なグラフを全て集めたい場合もあるので、ここでは、\$\mathcal{A}_m = \bigcup_{i \in I} \mathcal{A}_{H_i}\$, と定義しておくことにする。ここで、

$$h_{\mathcal{A}_m}(x_1, \dots, x_m) = I_{\mathcal{A}_m}(G_m(x_1, \dots, x_m)), \quad (6.1)$$

で関数 \$h_{\mathcal{A}_m}\$ を定義する。但し、\$G_m(x_1, \dots, x_m)\$ は重み \$x_1, \dots, x_m\$ から生成される一般化されたしきい値モデルの実現である。このとき、\$\mathcal{A}_m\$ の定義より、\$h_{\mathcal{A}_m}\$ は変数の入れ替えに関して不変である。従って、一般化されたしきい値モデル \$G_n(X_1, \dots, X_n)\$ で生成されるグラフの部分グラフで、\$\mathcal{A}_m\$ に含まれるものの個数は

$$U_n(\mathcal{A}_m) = \sum_{1 \leq i_1 < \dots < i_m \leq n} h_{\mathcal{A}_m}(X_{i_1}, \dots, X_{i_m}),$$

で与えられ、\$U\$ 統計量となる。\$U\$ 統計量の一般論より、以下の大数の強法則、中心極限定理、重複対数の法則が得られる [48].

定理 6.0.9 (大数の強法則)

$$\lim_{n \rightarrow \infty} \frac{U_n(\mathcal{A}_m)}{\binom{n}{m}} = \mathbb{E}[h_{\mathcal{A}_m}(X_1, \dots, X_m)], \quad \text{a.s.}$$

中心極限定理，重複対数の法則では

$$\zeta(\mathcal{A}_m) = \text{Var}(\mathbb{E}[h_{\mathcal{A}_m}(X_1, \dots, X_m)|X_1]) > 0$$

を仮定する。

定理 6.0.10 (中心極限定理)

$n \rightarrow \infty$ の極限で，

$$\sqrt{\frac{n}{m^2 \zeta(\mathcal{A}_m)}} \left[\frac{U_n(\mathcal{A}_m)}{\binom{n}{m}} - \mathbb{E}[h_{\mathcal{A}_m}(X_1, \dots, X_m)] \right] \Longrightarrow \mathcal{Z},$$

但し， \Longrightarrow は弱収束を， \mathcal{Z} は標準正規分布をそれぞれ表す。

定理 6.0.11 (重複対数の法則)

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n(\log \log n)^{-1}}{2m^2 \zeta(\mathcal{A}_m)}} \left| \frac{U_n(\mathcal{A}_m)}{\binom{n}{m}} - \mathbb{E}[h_{\mathcal{A}_m}(X_1, \dots, X_m)] \right| = 1, \quad \text{a.s.}$$

また，頂点 i を含む部分グラフで \mathcal{A}_m に含まれるものの個数を，

$$U_n(\mathcal{A}_m; i) = \sum_{\substack{1 \leq i_2 < \dots < i_m \leq n \\ i_2, \dots, i_m \neq i}} h_{\mathcal{A}_m}(X_i, X_{i_2}, \dots, X_{i_m}),$$

と定義し，

$$U_n(\mathcal{A}_m; i; x) = \sum_{\substack{1 \leq i_2 < \dots < i_m \leq n \\ i_2, \dots, i_m \neq i}} h_{\mathcal{A}_m}(x, X_{i_2}, \dots, X_{i_m}),$$

とすると，定理 2.0.1 と同じ議論を適用することで，以下を得る。

定理 6.0.12

(i) 任意の $x \in \mathbb{R}$ に対して，

$$\lim_{n \rightarrow \infty} \frac{U_n(\mathcal{A}_m; 1; x)}{\binom{n-1}{m-1}} = U(\mathcal{A}_m; 1; x), \quad \text{a.s.}$$

(ii) 特に，

$$\lim_{n \rightarrow \infty} \frac{U_n(\mathcal{A}_m; 1)}{\binom{n-1}{m-1}} = U(\mathcal{A}_m; 1; X_1), \quad \text{a.s.}$$

但し，

$$U(\mathcal{A}_m; 1; x) = \int_{\mathbb{R}^d} \dots \int_{\mathbb{R}^d} h_{\mathcal{A}_m}(x, x_2, \dots, x_m) F(dx_2) \dots F(dx_m).$$

である。

ここで, $U_n(\mathcal{A}_m; i), 1 \leq i \leq n$ は同分布であることと, 以下の関係に注意する.

$$\frac{\sum_{i=1}^n U_n(\mathcal{A}_m; i) / \binom{n-1}{m-1}}{n} = \frac{m U_n(\mathcal{A}_m)}{n \binom{n-1}{m-1}} = \frac{U_n(\mathcal{A}_m)}{\binom{n}{m}}.$$

このことから, 部分グラフの個数もまた, エルゴード性を持つことがわかる.

前章までに, しきい値モデルに対する, 次数・クラスター係数・平均頂点間距離の極限定理を得た. 一般化されたしきい値モデルに関しても, 同じ議論により, 次数・クラスター係数に関する極限定理を得ることができる. しかし, 平均頂点間距離に関する極限定理はまだ得られていない.

一般化されたしきい値モデルに関する, 次数の極限分布の例を以下に挙げる. 以下の例では, 極限の確率変数を D で表すこととする. また, X_1 の分布は平均 $1/\lambda$ の指数分布に従うものとする.

例 6.0.6

$B_1 = (\theta_1, \theta_2]$, $f_c(x, y) = x + y$ とおいた場合の一般化されたしきい値モデルは, $\theta_1 < X_i + X_j \leq \theta_2$ であるときに限って, 辺 (i, j) が存在するモデルである. 但し, 二つのしきい値, $\theta_1, \theta_2 \in \mathbb{R}$ は $\theta_1 < \theta_2$ を満たすものとする. このとき, 次数の極限分布は以下で与えられる.

$$D \sim \begin{cases} \delta_0(dk) & \text{if } \theta_1 < \theta_2 \leq 0, \\ \begin{cases} e^{-\lambda\theta_2} \cdot \delta_0(dk) \\ + I_{(0, 1-e^{-\lambda\theta_2})}(k) \cdot \frac{e^{-\lambda\theta_2}}{(1-k)^2} dk \end{cases} & \text{if } \theta_1 \leq 0 < \theta_2, \\ \begin{cases} e^{-\lambda\theta_2} \cdot \delta_0(dk) \\ + I_{(0, 1-e^{-\lambda(\theta_2-\theta_1)})}(k) \cdot \frac{e^{-\lambda\theta_2}}{(1-k)^2} dk \\ + I_{(e^{-\lambda\theta_1}-e^{-\lambda\theta_2}, 1-e^{-\lambda(\theta_2-\theta_1)})}(k) \cdot \frac{e^{-\lambda\theta_1}-e^{-\lambda\theta_2}}{k^2} dk \end{cases} & \text{if } 0 < \theta_1 < \theta_2. \end{cases}$$

例 6.0.7

$N \in \{1, 2, \dots\}$ に対して, $f_c(x, y) = x + y$, $B = \bigcup_{j=1}^N (a_j, b_j]$ で, $0 \leq a_1 \leq b_1 \leq \dots \leq a_j \leq b_j \leq a_{j+1} \leq \dots \leq a_N \leq b_N$ となるもので生成されるモデルに関する次数の極限分布は, 以下のようになる.

$$D \sim \sum_{j=0}^N I_{(e^{\lambda b_j S_{j+1}}, e^{\lambda a_{j+1} S_{j+1}})} \cdot \frac{S_{j+1}}{k^2} dk + \sum_{j=1}^N I_{(e^{\lambda b_j S_{j+1}}, e^{\lambda a_j S_j})} \cdot \frac{e^{-\lambda b_j} - S_{j+1}}{(1-k)^2} dk.$$

但し, $b_0 = 0$ で, $j \in \{1, \dots, N\}$ に対して $S_j = \sum_{i=j}^N (e^{-\lambda a_i} - e^{-\lambda b_i}) = \mathbb{P}(X_1 \in \bigcup_{i=j}^N (a_i, b_i]) \in [0, 1]$ である.

例 6.0.8

$\theta > 0$ に対して, $f_c(x, y) = xy$, $B = (\theta, \infty)$ とした場合には, 極限分布は以下のようなになる.

$$D \sim I_{(0,1)}(k) \cdot \frac{\lambda^2 \theta \cdot e^{\lambda^2 \theta / \log k}}{k (\log k)^2} dk.$$

例 6.0.9

ここでは $l=2$ の例として, 固定した $\theta \in \mathbb{R}$ と $c \in [0, \infty)$ に対して, $B_1 = (\theta, \infty]$, $B_2 = (0, c]$, $f_c^1(x, y) = x + y$, $f_c^2(x, y) = |x - y|$ とおいたモデルを考える. このモデルは, Masuda and Konno [38] で提案されているものである. モデルの定義から, 辺の定義関数は

$$h(x, x_2) = \begin{cases} I_{[-c+x, c+x]}(x_2) & \text{if } \frac{\theta+c}{2} \leq x, \\ I_{(\theta-x, c+x]}(x_2) & \text{if } \frac{\theta-c}{2} \leq x \leq \frac{\theta+c}{2}, \\ 0 & \text{if } x \leq \frac{\theta-c}{2}, \end{cases}$$

のように書けるので, 次数の極限分布は以下のようになる.

$$D \sim \begin{cases} (1 - e^{-\lambda(\theta-c)/2}) \delta_0(dk) \\ + I_{(0, 2e^{-\lambda(\theta+c)/2} \sinh(\lambda c)]}(k) \cdot g(k) dk & \text{if } c \leq \theta, \\ I_{(0, e^{-\lambda\theta} - e^{-\lambda c}]}(k) \cdot \frac{1}{2 \sinh(\lambda c)} dk \\ + I_{(e^{-\lambda\theta} - e^{-\lambda c}, 1 - e^{-\lambda(\theta+c)})}(k) \cdot g(k) dk \\ + I_{(1 - e^{-\lambda(\theta+c)}, 1 - e^{-2\lambda c}]}(k) \cdot \frac{e^{2\lambda c}}{2 \sinh(\lambda c)} dk & \text{if } 0 \leq \theta \leq c, \\ I_{(0, 1 - e^{-\lambda c}]}(k) \cdot \frac{1}{2 \sinh(\lambda c)} dk \\ + I_{(1 - e^{-\lambda c}, 1 - e^{-2\lambda c}]}(k) \cdot \frac{e^{2\lambda c}}{2 \sinh(\lambda c)} dk & \text{if } -c \leq \theta \leq 0, \\ I_{(0, 1 - e^{-\lambda c}]}(k) \cdot \frac{1}{2 \sinh(\lambda c)} dk \\ + I_{(1 - e^{-\lambda c}, 1 - e^{-2\lambda c}]}(k) \cdot \frac{e^{2\lambda c}}{2 \sinh(\lambda c)} dk & \text{if } \theta \leq -c, \end{cases}$$

where

$$g(k) = \frac{4e^{-\lambda\theta}}{(k + \sqrt{k^2 + 4e^{-\lambda(\theta+c)}})^2 + 4e^{-\lambda(\theta+c)}} + \frac{1}{2 \sinh(\lambda c)}.$$

次の例は, X_1 の分布が $[0, 1]$ 区間上の一様分布である場合の, 次数の極限分布である.

例 6.0.10

$N \in \{1, 2, \dots\}$ に対して, $f_c(x, y) = x + y$, $B = \bigcup_{j=1}^N (a_j, b_j]$ で, $0 \leq a_1 \leq b_1 \leq \dots \leq a_j \leq b_j \leq a_{j+1} \leq \dots \leq a_N \leq b_N \leq 1$ となるもので生成されるモデルに関する次数の極限分布は, 以下のようになる.

$$D \sim I_{(0, S_1)}(k) + (1 - b_N) \cdot \delta_0(k) + \sum_{i=1}^N (a_i - b_{i-1}) \cdot \delta_{S_i}(k),$$

但し, $b_0 = 0$ で, $j \in \{1, \dots, N\}$ に対して $S_j = \sum_{i=j}^N (b_i - a_i) = \mathbb{P}(X_1 \in \bigcup_{i=j}^N (a_i, b_i])$. である.

例として, 以下を考える.

$$B = K_n \equiv \bigcup_{\substack{a_m=0,2 \\ 1 \leq m \leq n}} \left[\sum_{m=1}^n \frac{a_m}{3^m}, \sum_{m=1}^n \frac{a_m}{3^m} + \frac{1}{3^n} \right].$$

この集合の極限集合 $K = \bigcap_{n=1}^{\infty} K_n$ は、カントール集合である。 K_n に関する極限分布は、以下のようなになる。

$$D \sim I_{(0, 2^n/3^n)}(k)dk + \frac{1}{3^n} \cdot \sum_{i=1}^{2^{n-1}} \delta_{\frac{2i-1}{3^n}}(dk) + \sum_{i=1}^{2^{n-1}-1} \frac{1}{3 \cdot 3^{|n-1-i|}} \cdot \delta_{\frac{2i}{3^n}}(dk).$$

Vapnik-Chervonenkis のアプローチは統計的学習理論の分野でよく知られており、特に、一様な極限定理を示す際に用いられる [16, 45]。以下では、定理 6.0.9 (及び定理 2.0.3) が VC クラスと呼ばれる、ボレル集合のある族の上で一様収束であることを示す。

ある集合 M に対して、 \mathcal{D} を M の部分集合のある族であるとする。ここで、 $A \subset M$ に対して、 $\Delta^{\mathcal{D}}(A) = \#(A \cap \mathcal{D})$ とする。ここで、 $\#(A \cap \mathcal{D})$ は $A \cap \mathcal{D} = \{A \cap D : D \in \mathcal{D}\}$ に属する集合の個数を表すものとする。さらに、 $n = 0, 1, 2, \dots$ に対して、 $m_{\mathcal{D}}(n) = \max_{A \subset M} \{\Delta^{\mathcal{D}}(A) : |A| = n\}$ とする。但し、 $|M| < n$ の場合は $m_{\mathcal{D}}(n) = m_{\mathcal{D}}(|M|)$ とする。ここで、 $|A|$ は集合 A に属する元の個数を表す。

族 \mathcal{D} に対する指標 $S(\mathcal{D})$ を以下で定義する。

$$S(\mathcal{D}) = \begin{cases} \sup \{n : m_{\mathcal{D}}(n) = 2^n\} & \text{if } \mathcal{D} \text{ is non-empty,} \\ -1 & \text{if } \mathcal{D} \text{ is empty.} \end{cases}$$

この指標を使って、族 \mathcal{D} は $S(\mathcal{D}) < +\infty$ であるとき、Vapnik-Chervonenkis (VC) クラスと呼ばれる。例えば、半開区間の族 $\mathcal{D} = \{(-\infty, x] : x \in \mathbb{R}\}$ は $S(\mathcal{D}) = 1$ となり、 \mathbb{R} 上の VC クラスの例となっている。

VC クラスに関して、Dudley [16] の Chapter 4.5 の議論をもとに、以下の補題を得る。

系 6.0.1

- (i) 任意の族 $\mathcal{D}, \mathcal{D}' \subset 2^M$ に対して、 $S(\mathcal{D}) < \infty$ かつ、 $S(\mathcal{D}') < \infty$ ならば、 $S(\mathcal{D} \cup \mathcal{D}') < \infty$ かつ、 $S(\mathcal{D} \cap \mathcal{D}') < \infty$ である。
- (ii) 任意の族 $\mathcal{D} \subset 2^M$ と $\mathcal{D}' \subset 2^N$ に対して、 $S(\mathcal{D}) < \infty$ かつ、 $S(\mathcal{D}') < \infty$ ならば、 $S(\mathcal{D} \times \mathcal{D}') < \infty$ である。

ここで、上述の族はそれぞれ $\mathcal{D} \cup \mathcal{D}' = \{D \cup D' : D \in \mathcal{D}, D' \in \mathcal{D}'\}$, $\mathcal{D} \cap \mathcal{D}' = \{D \cap D' : D \in \mathcal{D}, D' \in \mathcal{D}'\}$, $\mathcal{D} \times \mathcal{D}' = \{D \times D' : D \in \mathcal{D}, D' \in \mathcal{D}'\}$ である。

与えられた関数 $h : M \rightarrow \mathbb{R}$ に対して、

$$D_h = \{(x, t) \in M \times \mathbb{R} : 0 \leq t \leq h(x) \text{ or } h(x) \leq t \leq 0\}.$$

を h の部分グラフという。関数の族 \mathcal{H} が VC-subgraph クラスであるとは、 $\mathcal{D}_{\mathcal{H}} = \{D_h : h \in \mathcal{H}\}$ が VC クラスであるときをいう。

ある自然数 m を固定する。 M 上の独立同分布な確率変数列 $\{X_i\}_{i=1,2,\dots}$ に対して、以下の M^m 上の実可測関数の族 \mathcal{H} に関する一様な大数の強法則が知られている。

補題 6.0.2 (Corollary 3.3 of [3])

\mathcal{H} が変数の入れ替えに対して不変な M^m 上の実可測関数の VC-subgraph クラスで、
 $\mathbb{E}[\sup_{h \in \mathcal{H}} |h(X_1, \dots, X_m)|] < +\infty$ を満たすならば、

$$\lim_{n \rightarrow \infty} \sup_{h \in \mathcal{H}} \left[\frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} h(X_{i_1}, \dots, X_{i_m}) - \mathbb{E}[h(X_1, \dots, X_m)] \right] = 0, \quad \text{a.s.}$$

が成り立つ。

補題 6.0.2 を使うために、以下では、式 (6.1) を書き直し、ボレル集合の VC クラスで指標付けされているそれらの関数の族は VC-subgraph クラスであることを示す。以下、簡単のため $d = l = 1$ とするが、一般の場合への拡張は自明である。

前述の整数 $m \geq 2$ と $f_c^1 : \mathbb{R}^2 \rightarrow \mathbb{R}$ に対して、グラフの定義関数 $G : \mathbb{R}^m \rightarrow \mathbb{R}^{\binom{m}{2}}$ を以下で定義する。

$$G(x_1, \dots, x_m) = (f_c^1(x_1, x_2), f_c^1(x_1, x_3), \dots, f_c^1(x_1, x_m), f_c^1(x_2, x_3), \dots, f_c^1(x_{m-1}, x_m)).$$

この関数の値の成分は、異なる頂点对を辞書式順序で並べたものと対応している。例えば、 $m = 4$ の場合は、第 1 座標が $(1, 2)$ 、第 2 座標が $(1, 3)$ 、また、第 6 座標が $(3, 4)$ に対応している。ここで、辺 $\langle i, j \rangle$ は与えられたボレル集合 B に対して、 $f_c^1(x_i, x_j) \in B$ であるときに限って存在することに注意する。

与えられた m 個の頂点からなるグラフの族 \mathcal{A}_m に対して、対応する $\mathbb{R}^{\binom{m}{2}}$ 上の集合 $\tilde{\mathcal{A}}_m$ を以下で定義する。まず、各グラフ $H \in \mathcal{A}_m$ に対して、 $\mathbb{R}^{\binom{m}{2}}$ 上の集合 \tilde{H} を対応付ける。この集合は、 H の頂点对が辺で結ばれているときには、対応する \tilde{H} の座標が B であり、そうでなければ B^c となっている集合である。例えば、 $m = 4$ で H の辺集合が $\{\langle 1, 2 \rangle, \langle 2, 3 \rangle\}$ の場合は、 $\tilde{H} = B \times B^c \times B^c \times B \times B^c \times B^c$ である。これらを用いて、 $\tilde{\mathcal{A}}_m = \bigcup_{H \in \mathcal{A}_m} \tilde{H}$ とする。定義より、 $G(x_1, \dots, x_m) \in \tilde{\mathcal{A}}_m$ は、 $G_m((x_1, \dots, x_m) \in \mathcal{A}_m$ と同値である。従って、式 (6.1) を以下のように書き換えることができる。

$$h_{\mathcal{A}_m}^B(x_1, \dots, x_m) = I_{\tilde{\mathcal{A}}_m}(G(x_1, \dots, x_m)) = I_{G^{-1}(\tilde{\mathcal{A}}_m)}(x_1, \dots, x_m), \quad (6.2)$$

但し、 $G^{-1}(\tilde{\mathcal{A}}_m)$ は $\tilde{\mathcal{A}}_m$ の逆像である。

\mathcal{D} を \mathbb{R} 上のボレル集合による、ある VC クラスとする。式 (6.2) より、 $\mathcal{H}_{\mathcal{D}} = \{h_{\mathcal{A}_m}^B(x_1, \dots, x_m) : B \in \mathcal{D}\}$ は、VC-subgraph クラスであることを示す。系 6.0.1 より、 \mathcal{D} が VC クラスであるとき、定義関数の族 $\{I_D : D \in \mathcal{D}\}$ は VC-subgraph クラスである。従って、式 (6.2) より、族 $G_{\mathcal{D}}^{-1}(\tilde{\mathcal{A}}_m) = \{G^{-1}(\tilde{\mathcal{A}}_m) : B \in \mathcal{D}\}$ が VC クラスならば、 $\mathcal{H}_{\mathcal{D}}$ は VC-subgraph クラスである。実際、 $G_{\mathcal{D}}^{-1}(\tilde{\mathcal{A}}_m)$ は [16] の Theorem 4.2.3 と上述の補題 6.0.1 によって、VC クラスであることが容易にわかる。従って、補題 6.0.2 により、以下の定理を得る。

定理 6.0.13

\mathcal{D} をボレル集合の, ある VC クラスとする. また, 一般化されたしきい値モデルを構成する f_c と \mathcal{A}_m を考える. このとき, 以下が成り立つ.

$$\limsup_{n \rightarrow \infty} \sup_{B \in \mathcal{D}} \left[\frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} h_{\mathcal{A}_m}^B(X_{i_1}, \dots, X_{i_m}) - \mathbb{E} [h_{\mathcal{A}_m}^B(X_1, \dots, X_m)] \right] = 0, \quad \text{a.s.}$$

第7章 結論

本論文では、複雑ネットワークで基本的な役割を果たす、次数・クラスター係数・平均頂点間距離に関する極限定理を中心に、しきい値モデルとその一般化モデルについての構造を議論した。

しきい値モデルの次数と平均頂点距離の概収束極限は、いずれもその期待値から予想できるものである。次数に関しては、頂点の重みを固定した場合には、通常の大数の法則からの直接の帰結として極限定理が得られた。その結果から、頂点の重みを固定しない場合の極限が独立性を用いて得られるが、この場合の極限分布は、重みを与える確率変数の分布そのものから決まるものであり、それ自身興味深い。特に、指数分布やパレート分布などの具体的な例で、ベキ分布が現れるため、簡単なルールで得られる確率的なグラフでありながら、スケールフリー性を持つこともあるモデルとして注目に値する。

平均頂点間距離の解析は、グラフ全体の情報が必要であるため困難であることが多い。しかしながら、しきい値モデルの場合には、その定義から階層構造を得ることができ、その結果としてグラフの距離が三種類に限られることが本質的な役割を果たし、平均頂点間距離の極限定理を得るに至った。その過程では、孤立点の個数に関する概収束が独立性を用いて証明されることが大きな役割を果たしている。しかしながら、一般化されたしきい値モデルに関しては、階層構造のような良い構造を利用することが一般にはできないので、平均頂点間距離の漸近的な振る舞いについては未解決のままである。

クラスター係数の極限は、次数・平均頂点間距離の極限とは異なり、統計量そのものの平均値ではなく、分母・分子のそれぞれの平均値の比となっている点が直観的な予想と異なっている。証明では、 U 統計量の一般論が適用できるように変形することで極限定理を得た。また、グラフ全体のクラスター係数の極限定理は、各頂点のクラスター係数が同分布であるが独立ではないため、その極限定理は自明ではなかった。しかしながら、モデルを構成する確率変数の入れ替えに関して対称であることから、保測変換を用いた表現に書き換えることができ、その結果として通常のエルゴード定理と同じような議論を用いて極限定理を得ることができた。

一般化されたしきい値モデルに関しても、次数・クラスター係数に関する極限定理が得られたため、今後の応用の場面で、これらの量を定量的に扱う必要性が出てきた際に利用されることが予想される。平均頂点間距離に関する更なる結果が待たれるところである。また、Vapnik-Chervonenkis のアプローチを応用して一様な極限定理を得たことで、多数のモデルを比較する必要がある場合の理論的な裏付けとして一様収束の結果が用いられることが期待される。

関連図書

- [1] Albert, R. and Barabási, A. -L. (2002). Statistical mechanics of complex networks, *Rev. Mod. Phys.*, **74**, 47–97.
- [2] 青山 秀明, 家富 洋, 池田 裕一, 相馬 亘, 藤原 義久. (2007). パレート・ファームズ – 企業の興亡とつながりの科学, 日本経済評論社.
- [3] Arcones, M. A. and Giné, E. (1993). Limit theorems for U -processes, *Ann. Probab.*, **21**, 1494–1542.
- [4] Barabási, A.-L. (2002). *Linked: The New Science of Networks*, Perseus Books Group.
- [5] アルバート＝ラズロ・バラバシ. 青木 薫 訳. (2002). 新ネットワーク思考 – 世界のしくみを読み解く, NHK 出版.
- [6] Barabási, A.-L., and Albert, R. (1999). Emergence of scaling in random networks, *Science*, **286**, 509–512.
- [7] Biggs, N. L. (1993). *Algebraic Graph Theory, 2nd ed.*, Cambridge University Press.
- [8] Billingsley, P. (1995). *Probability and Measure, 3rd ed.*, John Wiley & Sons Inc.
- [9] Boguñá, M., and Pastor-Satorras, R. (2003). Class of correlated random networks with hidden variables, *Phys. Rev. E*, **68**, 036112.
- [10] Bollobás, B. (2001). *Random Graphs, 2nd ed.*, Cambridge University Press.
- [11] Bollobás, B., and Chung, F. R. K. (1988). The diameter of a cycle plus a random matching, *SIAM J. Discrete Math.*, **1**, 328–333.
- [12] Buchanan, M. (2003). *Nexus: Small Worlds and the Groundbreaking Science of Networks*, W. W. Norton & Company, New York.
- [13] マーク・ブキャナン. 阪本 芳久 訳. (2005). 複雑な世界, 単純な法則 – ネットワーク科学の最前線, 草思社.
- [14] Caldarelli, G., Capocci, A., De Los Rios, P., and Muñoz, M. A. (2002). Scale-free networks from varying vertex intrinsic fitness, *Phys. Rev. Lett.*, **89**, 258702.

- [15] Chung, F., and Lu, L. (2006). *Complex Graphs and Networks*, The American Mathematical Society.
- [16] Dudley, R. M. (1999). *Uniform Central Limit Theorems. Cambridge Studies in Advanced Mathematics, 63*. Cambridge University Press; Cambridge.
- [17] Durrett, R. (2004). *Probability: Theory and Examples, 3rd ed.*, Duxbury Press.
- [18] Durrett, R. (2007). *Random Graph Dynamics*, Cambridge University Press.
- [19] Erdős, P., and Rényi, A. (1959). On random graphs, *Publicationes Mathematicae*, **6**, 290–297.
- [20] Franceschetti, M. and Meester, R. (2007). *Random Networks for Communication: From Statistical Physics to Information Systems*, Cambridge University Press.
- [21] L. C. フリーマン. 辻 竜平 訳. (2007). 社会ネットワーク分析の発展, NHK 出版.
- [22] Grimmett, G. R., and Stirzaker, D. R. (2001). *Probability and Random Processes, 3rd ed.*, Oxford University Press.
- [23] 林 幸雄. (2007). 噂の拡がり方 – ネットワーク科学で世界を読み解く, 化学同人.
- [24] 林 幸雄 (編著), 大久保 潤, 藤原 義久, 上林 憲行, 小野 直亮, 湯田 聡夫, 相馬 亘, 佐藤 一憲. (2007). ネットワーク科学の工具箱 – つながりに隠れた現象をひもとく, 近代科学社.
- [25] Hagberg, A., Schult, D. A., and Swart, P. J. (2006). Designing threshold networks with given structural and dynamical properties, *Phys. Rev. E*, **74**, 056116.
- [26] Ide, Y., Konno, N., and Masuda, N. (2007). Limit theorems for some statistics of a generalized threshold network model, *RIMS Kokyuroku*, no.1551, Theory of Biomathematics and its Applications III, 81–86.
- [27] Ide, Y., Konno, N., and Masuda, N. (2007). Statistical properties of a generalized threshold network model, *Methodol. Comput. Appl. Probab.*, to appear.
- [28] Janson, S., Luczak, T., and Ruciński, A. (2000). *Random Graphs*, John Wiley & Sons Inc.
- [29] 今野 紀雄. (2006). 閾値ネットワークモデルの極限定理. 数理科学, No.517, 54–59.
- [30] 今野 紀雄. (2007). 複雑ネットワーク上のコンタクト・プロセスへ向けて, 数理科学, No.524, 61–68.
- [31] 今野 紀雄, 井手 勇介. (2008). 複雑ネットワーク入門, 講談社.

- [32] Konno, N., Masuda, N., Roy, R., and Sarkar, A. (2005). Rigorous results on the threshold network model, *J. Phys. A: Math. Gen.*, **38**, 6277–6291.
- [33] 蔵本 由紀. (2007). 非線形科学, 集英社新書.
- [34] Mahadev, N. V. R. and Peled, U. N. (1995). *Threshold Graphs and Related Topics*, *Annals of Discrete Mathematics*, 56. Elsevier, New York.
- [35] 増田 直紀. (2007). 私たちはどうつながっているのか - ネットワークの科学を応用する, 中公新書.
- [36] 増田 直紀, 今野 紀雄. (2005). 複雑ネットワークの科学, 産業図書.
- [37] 増田 直紀, 今野 紀雄. (2006). 「複雑ネットワーク」とは何か, 講談社ブルーバックス.
- [38] Masuda, N., and Konno, N. (2006). VIP-club phenomenon: Emergence of elites and masterminds in social networks, *Social Networks*, **28**, 297–309.
- [39] Masuda, N., Miwa, H., and Konno, N. (2004). Analysis of scale-free networks based on a threshold graph with intrinsic vertex weights, *Phys. Rev. E*, **70**, 036124.
- [40] Masuda, N., Miwa, H., and Konno, N. (2005). Geographical threshold graphs with small-world and scale-free properties, *Phys. Rev. E*, **71**, 036108.
- [41] Móri, T. F. (2002). On random trees, *Studia Scientiarum Mathematicarum Hungarica*, **39**, 143–155.
- [42] Najim, C. A., and Russo, R. P. (2003). On the number of subgraphs of a specified form embedded in a random graph, *Methodol. Comput. Appl. Probab.*, **5**, 23–33.
- [43] Newman, M. E. J., Moore, C., and Watts, D. J. (2000). Mean-field solution of the small-world network model, *Phys. Rev. Lett.*, **84**, 3201–3204.
- [44] Newman, M. E. J., and Watts, D. J. (1999). Renormalization group analysis of the small-world network model, *Phys. Lett. A*, **263**, 341–346.
- [45] Peskir, G. (2000). *From Uniform Laws of Large Numbers to Uniform Ergodic Theorems. Lecture Notes Series (Aarhus)*, 66. University of Aarhus, Department of Mathematics; Aarhus.
- [46] Peterson, I. (1997). *The Jungles of Randomness: A Mathematical Safari*, John Wiley & Sons Inc.
- [47] アイヴァース・ピーターソン. 今野 紀雄 監訳, 高橋 佐良人 訳. (2000). カオスと偶然の数学 - ランダムネス, 確率, そして複雑性へ, 白揚社.

- [48] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons Inc.
- [49] Servedio, V. D. P., Caldarelli, G., and Buttá, P. (2004). Vertex intrinsic fitness: How to produce arbitrary scale-free networks, *Phys. Rev. E*, **70**, 056126.
- [50] Söderberg, B. (2002). General formalism for inhomogeneous random graphs, *Phys. Rev. E*, **66**, 066121.
- [51] Strogatz, S. H. (2003). *SYNC: The Emerging Science of Spontaneous Order*, Hyperion Books.
- [52] スティーヴン・ストロガッツ. 蔵本 由紀 監修, 長尾 力 訳. (2005). SYNC, 早川書房.
- [53] 竹居 正登, 今野 紀雄. (2008). 複雑ネットワーク上の確率モデル, 日本ロボット学会誌, **26**, 43–49.
- [54] Varadhan, S. R. S. (2001). *Probability Theory*, The American Mathematical Society.
- [55] Watts, D. J. (1999). *Small Worlds*, Princeton University Press.
- [56] ダンカン・ワッツ. 栗原 聡 他 訳. (2006). スモールワールド – ネットワークの構造とダイナミクス, 東京電機大学出版局.
- [57] Watts, D. J. (2003). *Six Degrees – The science of a connected age*, W. W. Norton & Company, New York.
- [58] ダンカン・ワッツ. 辻 竜平, 友知 政樹 訳. (2004). スモールワールド・ネットワーク – 世界を知るための新科学的思考法, 阪急コミュニケーションズ.
- [59] Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of ‘small world’ networks, *Nature*, **393**, 440–442.
- [60] R. J. ウィルソン. 西関 隆夫, 西関 裕子 訳. (2001). グラフ理論入門, 近代科学社.

謝辞

この場を借りて、お世話になった方々に御礼申し上げたい。

指導教官の今野紀雄教授には、博士課程に入学してからの4年間、研究指導して頂いたことに心から感謝申し上げたい。研究がなかなか思うように進まない時期にも、大きな心で見守り、激励して頂いた。多くの学会や研究集会をご紹介頂き、それらに参加・発表させて頂いた経験は掛け替えのない財産となった。また、研究以外の面でも大いに勉強させて頂いた。重ねて感謝申し上げたい。

高野清治教授には、学部・修士の3年間、指導教官としてご指導頂いた。数学を本格的に勉強し始めて間もなかった私の、拙い基本的な内容のセミナーにお付き合い頂き、懇切丁寧にご指導頂いたことに深謝申し上げたい。本論文中で扱った Vapnik-Chervonenkis の理論は学部・修士を通じてご指導頂いた内容と大いに関係している。

北田泰彦教授には、修士時代に位相の授業でお世話になった。数列や写像に関する様々な性質について、一対一のセミナー形式で証明を添削して頂き、穴だらけの証明を納得できる形に仕上げたときの感動を今でも覚えている。そのような体験をさせて頂いたことに感謝申し上げたい。

玉野研一教授には、研究室の学生さん達と共に、群論や論理のセミナーに参加させて頂き、基本的な考え方をご指導頂いた。また、様々な場面で疑問に思ったことについて議論させて頂いた。行間を読むことの大変さと楽しさを、また、大勢でのセミナーで議論しあうことの楽しさを体感させて頂いたことに感謝申し上げたい。

梶原健准教授には、普段の研究室のセミナーで有益なコメントを頂くだけでなく、授業やセミナーで、専門外の我々にもわかりやすい言葉で、ご自身の研究分野を解説して頂いた。また、授業やセミナーの進め方で悩んでいる際に、さりげなく解決の糸口を頂いたことにも深く感謝を申し上げたい。

共同研究者の増田直紀氏には、論文の内容に関してはもちろんのこと、英語の添削や査読者への返答の仕方まで、懇切丁寧にご指導頂いた。また、有益な情報を提供して頂き、研究内容に関して多くの議論をさせて頂いたことに深謝申し上げたい。

大阪電気通信大学の竹居正登氏には、研究室のセミナーに度々ご参加頂き、有益なコメントを頂いた。また、セミナーの空き時間や研究集会等でお会いした際に、様々な疑問に対して議論して頂いたことに感謝申し上げたい。

特別研究教員の斎藤革子先生、名倉真紀先生、博士後期課程の瀬川悦生さん、濱田昌寿さんをはじめとして、今野研究室のメンバーの方々には、普段の学生生活やセミナーでの議論など、数え切れないほどにお世話になった。深く御礼申し上げたい。

最後に、家族をはじめとして、私を支えて下さった全ての方々に深く御礼申し上げます。