

数理手法 VIII (2023年冬学期) 講義ノート

島田 尚

2024年1月10日

上記講義の受講生のために公開しているものです。講義期間中適宜改訂します。

1 線形代数関連

1.1 固有値・固有ベクトルと行列の対角化

1.1.1 対角化可能な場合

n 次正方行列 A の固有値に縮退が無いとき、この行列は固有 (列) ベクトルを n 本並べた行列

$$P = \begin{pmatrix} | & | & \cdots & | \\ \vec{\lambda}_1 & \vec{\lambda}_2 & \cdots & \vec{\lambda}_n \\ | & | & \cdots & | \end{pmatrix} \quad (1)$$

を用いて

$$P^{-1}AP = D \equiv \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \quad (2)$$

と対角化できる。これは行列の作用を

$$A = PDP^{-1} \quad (3)$$

とわかりやすく分解したことに対応している。すなわち、

1. P^{-1} : 入力ベクトルを A の固有ベクトルの基底での成分表示にする
2. D : 行列の作用である、各固有空間での伸び縮み (複素数の場合は回転) を行う
3. P : 元の座標での成分表示に戻す

このため、対角化を用いることでこの行列の m 乗が

$$A^m = (PDP^{-1})^m = P \begin{pmatrix} \lambda_1^m & 0 & \cdots & 0 \\ 0 & \lambda_2^m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^m \end{pmatrix} P^{-1} \quad (4)$$

またこの行列の指数関数が

$$e^{\mathbf{A}t} = \sum_{k=0}^{\infty} \left[\frac{(\mathbf{P}\mathbf{D}\mathbf{P}^{-1})^k}{k!} \right] t^k = \mathbf{P} \begin{pmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{\lambda_n t} \end{pmatrix} \mathbf{P}^{-1} \quad (5)$$

などに見通しよくコンパクトに書ける。

(重要な例) エルミート (実対称) 行列の固有値・固有ベクトルの性質

\vec{x}, \vec{y} をそれぞれエルミート行列: $\mathbf{A}^* \equiv \overline{\mathbf{A}^\top} = \mathbf{A}$ の固有値 α, β に対応する固有ベクトルとすると、

$$\overline{\vec{y}^\top} (\mathbf{A}\vec{x}) = \alpha (\overline{\vec{y}^\top} \cdot \vec{x}) = (\overline{\vec{y}^\top} \mathbf{A}) \vec{x} = \overline{(\mathbf{A}^\top \vec{y})}^\top \vec{x} = \beta (\overline{\vec{y}^\top} \cdot \vec{x}) \quad (6)$$

であるから相異なる固有値に属する固有ベクトルは互いに直交することが分かる。またこのとき、 $\mathbf{P}^{-1} = \mathbf{P}^*$ なので \mathbf{P} はユニタリ行列: $\mathbf{P}^* \mathbf{P} = \mathbf{P} \mathbf{P}^* = \mathbf{E}$ (直交行列: $\mathbf{P}^\top \mathbf{P} = \mathbf{P} \mathbf{P}^\top = \mathbf{E}$) となる。さらに、上の式で $\alpha = \beta$ ととるとただちにわかるように、エルミート行列の固有値は全て実数である。

1.1.2 対角化できない場合：一般化固有値問題

行列の固有値が s 重に縮退している場合は、普通の固有ベクトルでは次元が足りなくて対角化できない (かもしれない) が、その部分空間は高々 s 次の一般化固有値問題：

$$(\mathbf{A} - \lambda \mathbf{E})^s \vec{\lambda} = \vec{0} \quad (7)$$

の一般化固有ベクトルで張られるので¹、その部分空間について考えると、 \mathbf{A} のベキ乗は

$$\mathbf{A}^m = \left[(\mathbf{A} - \lambda \mathbf{E}) + \lambda \mathbf{E} \right]^m = \sum_{k=0}^{s-1} \left[\frac{m!}{k!(m-k)!} \right] \lambda^{m-k} (\mathbf{A} - \lambda \mathbf{E})^k \quad (8)$$

と $(s-1)$ 次までの式に、指数関数は

$$e^{\mathbf{A}t} = e^{\lambda \mathbf{E}t} e^{(\mathbf{A} - \lambda \mathbf{E})t} = \left[\sum_{k=0}^{s-1} \frac{(\mathbf{A} - \lambda \mathbf{E})^k}{k!} t^k \right] e^{\lambda t}. \quad (9)$$

と、縮退している固有値に対応する指数関数的振る舞いのプレファクターとして高々 $(s-1)$ 次の多項式がついた形になる。

¹ジョルダン標準形に相似だということと同じこと。

1.2 Perron-Frobenius の定理

1.2.1 正行列について

以下では、行列の要素が全て正のものを正行列、全て非負のものを非負行列とよぶことにする。また、同様に「行列が正である」などという。正の正方行列 \mathbf{A} について次のことが成り立つ（ペロン・フロベニウスの定理）：

1. \mathbf{A} は正の固有値を必ず一つ以上持ち、そのうちの最大のものを α とすれば α は単純特性根である（縮退していない）。また、 α に対応する固有ベクトルは正（要素が全て正）である。この α はフロベニウス根（またはペロン根）と呼ばれる。
2. α 以外の固有値の絶対値は α より小さい。
3. 転置行列 \mathbf{A}^\top のフロベニウス根は α に等しい。
4. 正の固有ベクトルはフロベニウス根に属するものに限られる。

1.2.2 非負行列について

非負行列については、 $\mathbf{A}(t) \equiv a_{ij} + t$ の $\lim_{t \rightarrow 0}$ 極限を考えることでほぼ同じことが示せる：

1. \mathbf{A} は非負の固有値を必ず持ち、そのうちの最大のものを α とすればこれに対応する非負の固有ベクトルが存在する。この α もフロベニウス根と呼ばれる。
2. α 以外の固有値の絶対値は α 以下である。
3. 転置行列 \mathbf{A}^\top のフロベニウス根は α に等しい。

1.2.3 既約な非負行列について

任意の (i, j) に対して、 $(\mathbf{A}^l)_{ij} \neq 0$ となる l がそれぞれとれるような行列を既約行列という。正の正方行列は明らかに既約行列である。 \mathbf{A} が隣接行列なら、既約性は無向グラフについての連結性、有向グラフについての強連結性に対応する。「非負の既約行列」については「正行列」についてと同じく正のフロベニウス根と対応する正の固有ベクトルが存在することが言える。「ペロン・フロベニウスの定理」と言った場合、対象が少し広がったこちらの方を指していることが多い。

1.2.4 定理の証明のしかた

$\|\vec{x}\|_\infty = 1$ なる² 適当な正のベクトル \vec{x}_0 から出発して、帰納的に

$$\vec{x}_n = \frac{\mathbf{A}\vec{x}_{n-1}}{\|\mathbf{A}\vec{x}_{n-1}\|_\infty} \quad (10)$$

² $\|\vec{x}\|_p \equiv \left(\sum_i^n |x_i|^p \right)^{\frac{1}{p}}$ はベクトルの p ノルムの記号で、 $p \rightarrow \infty$ は最大値ノルム。

と正のベクトルの列を定めて正の固有ベクトルの存在と構成法を証明してゆく。まず、上の過程が正のベクトル列を生成することなどを確かめたあと、行列を作用させた際のベクトルの各成分の伸び縮み率の最小値と最大値：

$$\alpha_n \equiv \min \left\{ \frac{y_n^i}{x_n^i} \right\}, \quad \beta_n \equiv \max \left\{ \frac{y_n^i}{x_n^i} \right\} \quad (\vec{y}_n = \mathbf{A}\vec{x}_n) \quad (11)$$

の間に以下の不等式関係が成り立つことを示す：

$$0 \leq \alpha_0 \leq \alpha_1 \leq \dots \alpha_n \leq \beta_n \dots \leq \beta_1 \leq \beta_0, \quad (12)$$

$$\alpha_n + \frac{\epsilon}{\|\mathbf{A}\|_1} \cdot \|\mathbf{A}\vec{x}_n - \alpha_n \vec{x}_n\|_\infty \leq \alpha_{n+1}. \quad (13)$$

この関係式より、このベクトル列の $n \rightarrow \infty$ 極限が \mathbf{A} の正の固有ベクトルと一致する：

$$\lim_{n \rightarrow \infty} \alpha_n = \alpha, \quad \lim_{n \rightarrow \infty} \vec{x}_n = \vec{u}, \quad \mathbf{A}\vec{u} - \alpha\vec{u} = 0 \quad (14)$$

ことが言える。より詳細については講義での解説を参照のこと。

1.2.5 応用例 1：マルコフ連鎖

時間と状態が離散なマルコフ過程：

$$\vec{p}(t+1) = \mathbf{T}_{nm} \vec{p}(t) \quad (15)$$

における遷移行列 (Transition Matrix, 遷移確率行列、推移確率行列) \mathbf{T} は非負行列で、この過程がエルゴディックであることは \mathbf{T} が既約であることと同値。また、遷移行列なので $\sum_n T_{nm} = 1$ (確率の保存) すなわち $\vec{u} = (1, 1, \dots, 1)$ を固有値 1 の左固有ベクトルとして持つ。

これらの性質により、ペロン-フロベニウスの定理からこのような一般の (エルゴディックな) マルコフ過程に対して以下のことが言える。

1. \mathbf{T} は (右) 固有値 1 をフロベニウス根として持ち、対応する唯一の正の右固有ベクトルが存在する → 平衡状態の分布
2. これ以外の固有値の絶対値は 1 より小さい → 緩和過程
 - (a) 実際、確率の保存の性質からベクトルの成分の和は保存するので、固有ベクトルについて考えると $\|\mathbf{T}\vec{\lambda}\|_1 = \sum_i \sum_j T_{ij} \lambda_j = \lambda \sum_i \lambda_i = \sum_k \lambda_k$ が成り立つ。よって、ペロン根以外の固有値に対応する固有ベクトルは $\sum_k \lambda_k = 0$ を満たす。
 - (b) ペロン根以外に対応する固有ベクトル方向の成分は時間とともに減衰していくがこれは成分の和が 0 なので確率の保存とは矛盾せず、平衡状態からのずれに対応する緩和モードである。

1.2.6 応用例 2 : Google Page Rank

ノード数 N のネットワークに対するグーグル行列 G :

$$g_{ij} = \alpha \left(\frac{a_{ij}}{k_i} \right) + (1 - \alpha) \left(\frac{1}{N} \right) \quad (a_{ij} \text{ は隣接行列の成分、} k_i \text{ はノード } i \text{ の出次数}) \quad (16)$$

の転置は既約な遷移行列なので、上記の通り唯一の平衡状態が定まる。この固有ベクトル成分を PageRank Vector と呼び、この固有ベクトルの成分の大きさに従ってつけたノードの重要性のランクを PageRank という。初期 Google 検索結果の順位づけの基本的機構。

1.2.7 関連 : ベキ乗法 (Power Method)

行列の固有値と固有ベクトルを最大固有値から順に近似的に求める簡便な方法の代表例としてベキ乗法がある。行列 A がランク落ちしておらず、また絶対値が最大の固有値 λ_1 が縮退していないとすると³、適当な初期ベクトル

$$\vec{x}_0 = \sum_i c_i \vec{\lambda}_i$$

に対して

$$\vec{x}_n \equiv A^n \vec{x}_0 = \lambda_1^n \left[c_1 \vec{\lambda}_1 + \left\{ c_2 \left(\frac{\lambda_2}{\lambda_1} \right)^n \vec{\lambda}_2 + c_3 \left(\frac{\lambda_3}{\lambda_1} \right)^n \vec{\lambda}_3 + \dots \right\} \right] \quad (17)$$

なので

$$\lim_{n \rightarrow \infty} \vec{x}_n \rightarrow c_1 \lambda_1^n \vec{\lambda}_1. \quad (18)$$

つまり、毎ステップベクトルの正規化をしながら行列のかけ算を繰り返すだけで最大固有値とそれに対応する固有ベクトルが得られる。行列 A が疎 (0 の要素が多い) な場合に最大固有値を求めたい場合 (例えば Google 行列から Page Rank ベクトルを求めたい場合など) には特に速い手法である。

実際の活用にあたっては、“固有値” : $\frac{\|\vec{x}_n\|}{\|\vec{x}_{n-1}\|}$ やベクトルの各成分が充分収束したところでそれらを最大固有値と固有ベクトルとすればよい⁴。また、その最大固有値の固有空間を除いて再びベキ乗法を行えば 2 番目以降の固有値と固有ベクトルを求めていくこともできる。

³例えば正行列についてはペロンフロベニウスの定理から保証される。

⁴「正の行列」のペロン根については、各成分の伸び率で上下から押さえ込むことでしっかりした保証が得られる。

1.3 特異値分解 (Singular Value Decomposition, SVD)

1.3.1 特異値分解の定義

$n \times m$ (n 行 m 列) の (実) 行列 A は

$$A = U \Sigma V^* \quad (U \Sigma V^T) \quad (19)$$

といつでも分解できる。ここで U と V はそれぞれ $n \times n$ と $m \times m$ のユニタリ行列 (直交行列) で、 Σ は A のランクのサイズ $r \times r$ の対角行列 D_r を含んだ次のような $n \times m$ 行列である:

$$\Sigma = \begin{pmatrix} D_r & O_{r,m-r} \\ O_{n-r,r} & O_{n-r,m-r} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & \sigma_r & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} \quad (20)$$

$$(\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0). \quad (21)$$

これを特異値分解 (singular value decomposition, SVD) という。

1.3.2 いつでも特異値分解できること

A^*A はエルミート (実対称) で半正定値⁵な行列なので、 m 個の負でない実固有値 $\{\lambda_i\}$ と、これに対応した互いに直交する右固有ベクトル $\{\vec{v}_i\}$ を持つ。以下、固有値は大きい順に番号付けし、また固有ベクトルは正規化したものを考える。

これら固有ベクトルに A をかけたもの同士も

$$(A\vec{v}_i, A\vec{v}_j) = \vec{v}_i^* A^* A \vec{v}_j = \delta_{ij} \lambda_i \quad (22)$$

と互いに直交することが分かる。従って、 $A\vec{v}_i$ をそれぞれ正規化したものを \vec{u}_i とすると、

$$\vec{u}_i^* A \vec{v}_j = \delta_{ij} \sigma_i \quad (\sigma_i = \sqrt{\lambda_i}). \quad (23)$$

よって、列ベクトル \vec{v}_i を横に m 本並べた行列を V , 列ベクトル \vec{u}_i を横に n 本並べた行列を U とそれぞれすれば⁶,

$$U^* A V = \Sigma \quad \rightarrow \quad A = U \Sigma V^* \quad (24)$$

⁵任意のベクトル \vec{x} に対して $\vec{x}^* (A^* A) \vec{x} \geq 0$ ということ。

⁶ A が縦長 ($n > m$) の場合は m 本の \vec{u}_i に加えて適当に直交基底を足し、横長 ($n < m$) の場合は A のランクは高々 n しか無いのであるから m 個の固有値のうち $m - n$ 個の 0 固有値の固有ベクトルを省けば良い。このようにして作った正方行列 V, U は当然ユニタリ行列である。

つまり、行列要素を並べて書けば以下のよう：

$$\mathbf{A} = \left(\begin{array}{c|ccc|c} \vec{u}_1 & & & \\ \cdots & & & \\ \vec{u}_i & & & \\ \cdots & & & \\ \vec{u}_n & & & \end{array} \right) \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & \sigma_r & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} -\vec{v}_1^* \\ \vdots \\ -\vec{v}_j^* \\ \vdots \\ -\vec{v}_m^* \end{pmatrix} \quad (25)$$

元の行列が $\min(n, m)$ よりランク落ちしていない場合はそれぞれ以下のような形になる。
横長の場合：

$$\mathbf{A} = \left(\begin{array}{c|ccc|c} \vec{u}_1 & & & \\ \cdots & & & \\ \vec{u}_n & & & \end{array} \right) \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & \sigma_n & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} -\vec{v}_1^* \\ \vdots \\ -\vec{v}_n^* \\ \vdots \\ -\vec{v}_m^* \end{pmatrix} \quad (26)$$

縦長の場合：

$$\mathbf{A} = \left(\begin{array}{c|ccc|c} \vec{u}_1 & & & \\ \cdots & & & \\ \vec{u}_m & & & \\ \cdots & & & \\ \vec{u}_n & & & \end{array} \right) \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_m \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \begin{pmatrix} -\vec{v}_1^* \\ \vdots \\ -\vec{v}_m^* \end{pmatrix} \quad (27)$$

thin-SVD

上述の書き方を full-SVD, 0の部分省いて、 $n < m$ のとき $(n \times n)(n \times n)(n \times m)$ の形

$$\mathbf{A} = \left(\begin{array}{c|ccc|c} \vec{u}_1 & & & \\ \cdots & & & \\ \vec{u}_n & & & \end{array} \right) \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n \end{pmatrix} \begin{pmatrix} -\vec{v}_1^* \\ \vdots \\ -\vec{v}_n^* \end{pmatrix} \quad (28)$$

同様に $n > m$ のとき $(n \times m)(m \times m)(m \times m)$ の形

$$\mathbf{A} = \left(\begin{array}{c|ccc|c} \vec{u}_1 & & & \\ \cdots & & & \\ \vec{u}_m & & & \end{array} \right) \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_m \end{pmatrix} \begin{pmatrix} -\vec{v}_1^* \\ \vdots \\ -\vec{v}_m^* \end{pmatrix} \quad (29)$$

に書く書き方を thin-SVD という。

1.3.3 擬似逆行列

Σ のゼロでない対角項の逆数を取り、さらに転置した行列を Σ^+ と書くことにすれば、

$$\Sigma\Sigma^+ = E_n^{(r)} \equiv \begin{pmatrix} E_r & O \\ O & O_{n-r} \end{pmatrix}, \quad \Sigma^+\Sigma = E_m^{(r)} \equiv \begin{pmatrix} E_r & O \\ O & O_{m-r} \end{pmatrix}. \quad (30)$$

であるから (非対角ブロックの O のサイズの添字表記は見易さのため省いた)、

$$A^+ \equiv V\Sigma^+U^* \quad (31)$$

は A の擬似逆行列になる。すなわち、

$$AA^+ = U\Sigma\Sigma^+U^* = UE_n^{(r)}U^* \quad (32)$$

$$A^+A = V\Sigma^+\Sigma V^* = VE_m^{(r)}V^* \quad (33)$$

であることから $A = U\Sigma V^*$ と $A^+ = V\Sigma^+U^*$ は互いに広義可逆元である：

$$AA^+A = A \quad \cap \quad A^+AA^+ = A^+. \quad (34)$$

上の「単位行列もどき」

$$\tilde{E}_n^{(r)} \equiv AA^+ = UE_n^{(r)}U^* \quad (35)$$

$$\tilde{E}_m^{(r)} \equiv A^+A = VE_m^{(r)}V^* \quad (36)$$

は、「行列 A のランクに対応する空間についてなにもせず、それ以外の空間はつぶす」という作用をする正方行列である。

A が対角化可能な正方行列な場合は $AA^+ = A^+A = E_n$ だから A^+ は普通の逆行列 A^{-1} に他ならない。

1.4 対角化や特異値問題とデータ解析

1.4.1 最小2乗問題

$n \times m$ 行列 \mathbf{A} と定ベクトル \vec{b} (次元 n) が与えられた時に、

$$\|\mathbf{A}\vec{x} - \vec{b}\|_2^2 = (\mathbf{A}\vec{x} - \vec{b})^\top (\mathbf{A}\vec{x} - \vec{b}) \quad (37)$$

を最小化する \vec{x} (次元 m) を求める問題を一般に最小2乗問題という⁷。

解となる最適ベクトルを \vec{x}_* と書けば、そこからの任意の方向への微小なずれ $\vec{\epsilon}$ に対して

$$\|\mathbf{A}(\vec{x}_* + \vec{\epsilon}) - \vec{b}\|_2^2 - \|\mathbf{A}\vec{x}_* - \vec{b}\|_2^2 = \vec{\epsilon}^\top (\mathbf{A}^\top \mathbf{A}) \vec{\epsilon} + 2\vec{\epsilon}^\top [\mathbf{A}^\top (\mathbf{A}\vec{x}_* - \vec{b})] \geq 0 \quad (38)$$

が成り立つのであるから、 $\mathbf{A}^\top (\mathbf{A}\vec{x}_* - \vec{b}) = \vec{0}$ でなければいけない。この条件式

$$\mathbf{A}^\top \mathbf{A} \vec{x} = \mathbf{A}^\top \vec{b} \quad (39)$$

は正規方程式と呼ばれる。正規方程式に特異値分解のかたちを代入すると、

$$(\mathbf{V}\Sigma^\top \mathbf{U}^\top \mathbf{U}\Sigma \mathbf{V}^\top) \vec{x} = (\mathbf{V}\Sigma^\top \Sigma \mathbf{V}^\top) \vec{x} = (\mathbf{V}\Sigma^\top \mathbf{U}^\top) \vec{b} \quad (40)$$

ここで \mathbf{V} は正則行列なので左から $\mathbf{V}^{-1} = \mathbf{V}^\top$ をかけて

$$(\Sigma^\top \Sigma) \mathbf{V}^\top \vec{x} = (\Sigma^\top) \mathbf{U}^\top \vec{b} \quad (41)$$

を得る。ここで行列

$$\Sigma^\top \Sigma = \begin{pmatrix} \mathbf{D}_r^2 & \mathbf{O} \\ \mathbf{O} & \mathbf{O}_{m-r} \end{pmatrix} \quad (42)$$

は \mathbf{A} のランクサイズ r まで以外は零成分ばかりなので、 $\mathbf{V}^\top \vec{x}$ の $r+1$ から m までの成分については不定性がある (= なんでもいい)。これらの、正規方程式からは不定な成分について0とした解 (最小ノルム解) は次のように擬似逆行列で書ける：

$$\vec{x}_* = \mathbf{V}\Sigma^+ \mathbf{U}^\top \vec{b} = \mathbf{A}^+ \vec{b}. \quad (43)$$

この最小ノルム解は、 \mathbf{A} がフルランク正方行列である場合についての逆行列を用いた表記：

$$\vec{x}_* = \mathbf{A}^{-1} \vec{b} \quad (44)$$

と、また \mathbf{A} が列方向にフルランク (ランク m) である場合についての、 $(m \times m)$ 正則行列 $\mathbf{A}^\top \mathbf{A}$ の逆行列を用いた表記：

$$\vec{x}_* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \vec{b} \quad (45)$$

と一致する。

⁷ $n \leq m$ で行列のランクが行数 n の場合 (行フルランク) は連立方程式 $\mathbf{A}\vec{x} - \vec{b} = \vec{0}$ の解があるので、その場合も含み「 \vec{b} にぴったり合わせる事が無理な場合についてもベストな近似を探す」問題設定とすることができる。

1.4.2 線形回帰

目的変数 y とその説明変数 p 組: $\vec{x} = (x_1, \dots, x_p)^\top$ についての n 組のデータがあるとする。 $\vec{y} = (y_1, \dots, y_n)^\top$, $\mathbf{X} = (\vec{x}_1, \dots, \vec{x}_n)^\top$ とすれば、線形回帰によるデータのモデルは

$$\vec{y} = \mathbf{X}\vec{\beta} + \vec{\xi} \quad (46)$$

となる。ただしここで $\vec{\beta}, \vec{\xi}$ はそれぞれ回帰係数ベクトル (モデルパラメタ)、誤差ベクトルである。線形モデルからのデータのずれを2乗ノルム (平均2乗誤差) で評価することになれば、

$$\min \left(\|\vec{y} - \mathbf{X}\vec{\beta}\|_2^2 \right) \quad (47)$$

は最小2乗法に他ならない。

1.4.3 スパースモデリング

同じような誤差ならより少ない説明変数の方が良い、という考え方とそのための手法の代表例を以下に挙げる。

Ridge

平均2乗誤差に、加えて、フィッティングパラメタである回帰係数の2乗ノルムを加えた量を最小化する手法は Ridge と呼ばれる:

$$\min \left(\|\vec{y} - \mathbf{X}\vec{\beta}\|_2^2 + \lambda \|\vec{\beta}\|_2^2 \right) \quad (48)$$

λ は説明パラメタをより多くとることに課すペナルティの係数で、 λ を大きくとればよりノルムの小さな β が選ばれる。

LASSO

パラメタのペナルティを絶対値にとった場合はそのまま「 L_1 ノルム正規化」もしくは LASSO (Least Absolute Shrinkage and Selection Operator) と呼ばれる:

$$\min \left(\|\vec{y} - \mathbf{X}\vec{\beta}\|_2^2 + \lambda \|\vec{\beta}\|_1 \right) \quad (49)$$

線形なペナルティ項は等高線が多面体になるため、 λ を大きくとるといくつかの β_i が0となる頂点が選択される (説明変数が明示的に減る) のが利点。

1.4.4 主成分分析

多成分の観測データが複数 ($\{\bar{x}_n\}$) あるとき、基底の取り直しによりそれらをなるべく少ない基底で良く記述しようとする手法。

各成分の次元が同じ場合はデータ間の分散共分散行列:

$$V_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle \quad (\langle \cdot \rangle \text{ は観測データにわたっての平均を表す}), \quad (50)$$

各成分の次元が違う場合 (データが { 身長、収入、年齢、... } のような場合など) はデータ間の相関係数:

$$C_{ij} = \frac{\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle}{\sigma_i \sigma_j} \quad (\langle \cdot \rangle \text{ は観測データにわたっての平均を表す}) \quad (51)$$

を計算し、この実対称行列を対角化すればよい。最大固有値の方から固有ベクトルをとっていけば、これが求める“データ間の差異を表すのに重要な順”の直交基底 (≡ **主成分**)。

行列のトレース (対角和) の巡回不変性より

$$\text{Tr} \mathbf{V} = \text{Tr} (\mathbf{P} \mathbf{D} \mathbf{P}^{-1}) = \text{Tr} (\mathbf{P}^{-1} \mathbf{P} \mathbf{D}) = \text{Tr} (\mathbf{D}) = \sum_i \lambda_i \quad (52)$$

である (トレースの相似不変性) から、全固有値の和で規格化した各固有値はデータ点の分散のうちその固有方向に起因する割合に等しいので、**寄与率** (proportion) と呼ばれる。最大固有値からの累積寄与率で主成分をどこまでとるかを勘案する。

趣旨からして全ての固有空間を求める必要はなく、固有値の大きい順にいくつか固有値・固有ベクトルを求めてゆけばよいので、一般に行列をちゃんと対角化する必要は無く、ベキ乗法なんかでもいい。固有値の大きい方から少数の基底をとる、すなわち

$$\mathbf{C} = \mathbf{P} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \mathbf{P}^\top \quad (53)$$

を

$$\tilde{\mathbf{C}} = \begin{pmatrix} \vdots & 0 & \cdots & 0 \\ \tilde{\mathbf{P}} & \vdots & & \vdots \\ \vdots & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \lambda_i & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \cdots & \tilde{\mathbf{P}}_i^\top & \cdots \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \quad (54)$$

と近似することは最小2乗問題やスパースモデリングの精神と通じる。実際、これは残った2乗誤差を最小にするように少数の基底を選んでいることになっている。また、オートエンコーダにおいて中間層を一層とし、活性化関数を恒等関数にとった特殊な場合は主成分分析による次元の圧縮に他ならない。

2 確率分布関連

2.1 離散確率分布

2.1.1 確率分布

N 個の離散事象 $\{x_i\}$ に対する確率変数 X が、各事象をとる確率の組：

$$P(X = x_i) = f(x_i) \quad (i = 1, 2, \dots, N) \quad (55)$$

のことを確率分布という。よく、 x_i の離散状態の添字を略して $f(x)$ のように書く。観測データに関わるものは極論すればいつでもこの形（ヒストグラムの値）で、後述の連続型は理論的ツールということもできる。

2.1.2 累積分布関数

$$F(x) = P(X \leq x) = \sum_{\xi \leq x} f(\xi) \quad \text{もしくは} \quad \bar{F}(x) = P(X > x) = \sum_{\xi > x} f(\xi) \quad (56)$$

を累積分布関数と呼ぶ。サンプル数が膨大とは限らない実際のデータ解析では確率分布よりこちらの分布を扱うことも多い。また、後述するように連続型の確率分布も視野に入れると累積分布の方が扱いやすい面がある。

2.1.3 ランクサイズプロット

データを大きいものもしくは小さいもの順にならべたものをランクサイズプロットという。 $F(x), \bar{F}(x)$ はそれぞれ x の小さい順、大きい順に並べた時のランクに対応するから、ランクサイズプロットはこれら累積分布の逆関数（縦軸と横軸を入れ替えたもの）。

従って、例えば $P(x)$ のテール ($x \rightarrow \infty$ での漸近形) が指数 $-\alpha$ のべき分布に従うとき、 $P(> x) \sim x^{1-\alpha} \equiv x^{-\beta}$ 。 (大きい順の) ランクサイズプロットの傾きは $x \sim r^{\frac{1}{1-\alpha}} \sim r^{-\frac{1}{\beta}}$ 。よって、狭義の Zipf 則： $s \sim r^{-1}$ は $\alpha = 2$ ($\beta = 1$) に対応する。例えば後述の Cauchy 分布は $\alpha = 2$ なので、ランクサイズ分布を取れば Zipf 則的な領域を持つ。

2.1.4 べきテールと両対数プロット

指数分布や正規分布は、 y 軸を \log スケールでとることでそれぞれ直線や放物線としてすぐ見分けがつく。べき分布については両対数プロットが便利である。両対数プロットでの直線：

$$\log y = \alpha \log x + \beta \quad (57)$$

は元の変数でのべき関係

$$y = Cx^\alpha \quad (C = e^\beta) \quad (58)$$

に対応するので、直線的分布からべき分布が伺え、さらにその傾きからべきの指数が分かる。代数的なスピードより早い減衰（指数減衰）は、両対数プロットでは上に凸な形状となる。これは良く、truncation（直訳だと「切り詰め」）と表現される。

2.1.5 各種特徴量

期待値

一般に、 $E(g(X)) \equiv \sum_x g(x)f(x)$ を X についての g の期待値という。

平均、中央値、最頻値

x の期待値

$$E(X) = \sum_x xf(x) \quad (59)$$

を平均 (average) といい、分布の中心を表す“代表値”の代表である。良く μ で表される。

他の代表値として、中央値 (median) :

$$F(x) \geq \frac{1}{2} \quad \text{を満たす最小の } x \quad (60)$$

と最頻値 (mode)

$$f(x) \quad \text{が最大である } x \quad (61)$$

がある。

分散と標準偏差

分布の広がりを示す特徴量として、平均値からのズレの2乗の期待値 :

$$V(X) = E((X - \mu)^2) = \sum_x (x - \mu)^2 f(x) \quad (62)$$

を分散といい、良く σ^2 と書く。分散の平方根、 σ は標準偏差と呼ばれる量で、元の x と同じ次元をもっているため分布の幅により対応する。

高次のモーメント

$$\mu_k = E(X^k) = \sum_x x^k f(x) \quad (63)$$

を一般に k 次のモーメントといい、

$$\mu'_k = E((X - \mu)^k) = \sum_x (x - \mu)^k f(x) \quad (64)$$

を期待値まわりの k 次モーメントという。

2.2 連続型確率分布

2.2.1 確率密度分布

確率変数 X の取りうる値が連続な場合、各点そのものを取る確率は 0 なので、各点周りの微小領域での確率密度（確率密度関数、probability distribution function: PDF）:

$$p(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x} \quad (65)$$

を用いる。誤解の恐れがないときは、普通これも確率分布と呼ぶ。

2.2.2 累積分布関数

確率分布 $P(x)$ の累積分布は

$$P(X > x) = P(> x) = \int_x^{\infty} p(\xi) d\xi, \quad P(X \leq x) = P(\leq x) = \int_{-\infty}^x p(\xi) d\xi \quad (66)$$

2.2.3 変数変換について

単調増加で微分可能な関数 ϕ を用いて $Y = \phi(X)$ と変数変換した場合、 Y の確率密度は

$$g(y(x)) \Delta y = f(x) \Delta x \rightarrow g(y) = f(x) \frac{dx}{dy} = f(x) \left(\frac{d\phi}{dx} \right)^{-1} \quad (x = \phi^{-1}(y)). \quad (67)$$

例えば線形変換: $Y = aX + b$ の場合は $g(y) = p\left(\frac{y-b}{a}\right) / a$. また、上式より直ちに分かるように、累積分布は変数変換の代入そのもの: $G(y) = F(\phi^{-1}(y))$.

2.2.4 指定の分布に従った乱数の生成（逆関数法）

下からの累積確率分布を $y = P(\leq x) = \int_{-\infty}^x p(\xi) d\xi = F(x)$ と書けば、その逆関数: $x = F^{-1}(y)$ に $(0, 1)$ 区間の一様乱数を入力したものは $p(x)$ に従う発生確率となる。

(証明) $F(x)$ は値域が $(0, 1)$ の単調増加関数だから、 $y \in (0, 1) \rightarrow x = F^{-1}(y)$ は x の全定義域に対する一対一写像。また、累積分布関数なので $\frac{dy}{dx} = p(x) \rightarrow \frac{dx}{dy} = \frac{1}{p(x)}$. 従って、 y 上の微小区間 $(y_0, y_0 + \Delta y)$ は x 上の

$$(F^{-1}(y), F^{-1}(y + \Delta y)) = \left(x_0, x_0 + \frac{dx}{dy} \Big|_{y_0} \Delta y \right) = \left(x_0, x_0 + \frac{\Delta y}{p(x_0)} \right) \quad (68)$$

に写される。よって、この変換によって得られる分布の密度 $\rho(x_0)$ は

$$\rho(x_0) = \lim_{\Delta y \rightarrow 0} \frac{P\left(x < X \leq x + \frac{\Delta y}{p(x_0)}\right)}{\left(\frac{\Delta y}{p(x_0)}\right)} = \lim_{\Delta y \rightarrow 0} \frac{\Delta y}{\left(\frac{\Delta y}{p(x_0)}\right)} = p(x_0) \quad (69)$$

2.2.5 各種特徴量

期待値

連続型の場合の期待値は

$$E(g(X)) \equiv \int_{-\infty}^{\infty} g(x)p(x)dx \quad (70)$$

平均、中央値、最頻値

連続型の場合の平均は

$$\mu = \int_{-\infty}^{\infty} xp(x)dx, \quad (71)$$

中央値 (median) は

$$P(> x) = P(\leq x) = \frac{1}{2}, \quad (72)$$

最頻値 (mode) は極値点から探せばよい：

$$\frac{d p(x)}{dx} = 0. \quad (73)$$

分散と標準偏差

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 p(x)dx \quad (74)$$

高次のモーメント、歪度、尖度

期待値周りの高次のモーメント

$$\mu'_k = \int_{-\infty}^{\infty} (x - \mu)^k p(x)dx \quad (75)$$

も分布の形の特徴づけに良く使われる。

特に、歪度 (わいど、skewness)

$$\alpha_3 = \frac{\mu'_3}{\sigma^3} \quad (76)$$

は期待値周りの分布の非対称性の指標として、

尖度 (せんど、kurtosis)

$$\alpha_4 = \frac{\mu'_4}{\sigma^4} \quad \left(\text{または} \frac{\mu'_4}{\sigma^4} - 3 \right) \quad (77)$$

は期待値周りへの集中具合 (尖り具合) の指標として良く使われる。尖度が3 (0) より大きい場合はガウス分布より裾が広く、小さい場合はより期待値周りに集中している。

2.3 特性関数

2.3.1 定義と性質

確率密度分布の Fourier 変換

$$C(\xi) \equiv \int_{-\infty}^{\infty} \exp\{i\xi x\} p(x) dx \quad (78)$$

は特性関数と呼ばれる。定義から、特性関数は以下の性質を持つ：

- 確率分布の特性関数は必ずある
- 確率分布と特性関数とは 1 対 1 対応で、対応は連続
- $C(\xi)$ は連続で、 $C(0) = 1, |C(\xi)| \leq 1, C(-\xi) = \overline{C(\xi)}$ ($P(x)$ が対称 $\rightarrow C(\xi)$ 実対称)
- $P(x)$ が k 次モーメントまで持つ場合 $C(\xi)$ は k 回微分可能で、 $\left. \frac{d^n C(\xi)}{d\xi^n} \right|_{\xi=0} = \langle (ix)^n \rangle$
- 線形変換： $C_{(aX+b)}(\xi) = e^{i\xi b} C(a\xi)$
- 重畳定理： $C_{(X_1+X_2+\dots+X_n)}(\xi) = C_{X_1}(\xi) C_{X_2}(\xi) \cdots C_{X_n}(\xi)$

$$(\because \langle e^{i\xi \sum_i X_i} \rangle = \langle \prod_i e^{i\xi X_i} \rangle = \prod_i \langle e^{i\xi X_i} \rangle)$$

累積分布関数や特性関数は元の分布の情報を全て持っているので、適宜便利に使う。

(特に便利ではないが) 分かりやすい例

区間 $(0, 1)$ の一様分布の特性関数は

$$C(\xi) = \int_0^1 e^{i\xi x} dx = \frac{e^{i\xi} - 1}{i\xi} = \frac{i\xi - \frac{\xi^2}{2} - \frac{i\xi^3}{3!} + \dots}{i\xi} = 1 + \frac{i}{2}\xi - \frac{1}{3!}\xi^2 - \frac{i}{4!}\xi^3 + \dots$$

これから、全確率、期待値、分散などがそれぞれ次のように確認できる：

$$\langle 1 \rangle = C(\xi)|_{\xi=0} = 1$$

$$\mu = \langle x \rangle = \left(\frac{1}{i} \right) \left. \frac{dC(\xi)}{d\xi} \right|_{\xi=0} = \frac{1}{2}$$

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = - \left. \frac{d^2 C(\xi)}{d\xi^2} \right|_{\xi=0} - \left(\frac{1}{2} \right)^2 = \frac{1}{12}$$

$$\mu'_3 = \langle x^3 \rangle - 3\mu\sigma^2 - \mu^3 = - \frac{1}{i} \left. \frac{d^3 C(\xi)}{d\xi^3} \right|_{\xi=0} - \left(\frac{3}{2 \cdot 12} \right) - \left(\frac{1}{2} \right)^3 = \frac{1}{4} - \frac{1}{4} = 0$$

2.4 チェビシェフの不等式

2.4.1 定義と証明

確率分布 $P(X)$ の期待値と分散をそれぞれ μ, σ^2 とすると、以下の不等式が成り立つ：

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad (79)$$

「ある分布の期待値から標準偏差の何倍もずれた事象が起こる」という稀な事象の起こる確率について**確率分布の形状の詳細によらず**その上限を与えてくれる、基本的で大事な不等式関係である。実際の各々の分布については左辺の確率はこの上限よりもっと小さいことが多い（例：正規分布なら $P(|X - \mu| \geq 2\sigma) \sim 1/20$ ）。

(証明)

$$\theta(X) = \begin{cases} 0 & (|X - \langle X \rangle| < \epsilon) \\ 1 & (|X - \langle X \rangle| \geq \epsilon) \end{cases}$$

という関数を考えると、明らかに

$$\theta(X) \leq \left(\frac{X - \langle X \rangle}{\epsilon} \right)^2$$

がいつでも成り立つ。この不等式の両辺の期待値をとれば、

$$\langle \theta(X) \rangle \leq \left(\frac{\sigma(X)}{\epsilon} \right)^2$$

左辺は $\text{Prob.}(|X - \langle X \rangle| \geq \epsilon)$ そのものである。

2.5 大数の法則

2.5.1 平均と分散の加法性

独立な同一の分布に従う確率変数 X_i (独立同一分布、independent and identically distributed: IID などと呼ぶ) N 個の和について考えると、平均について

$$M \equiv \left\langle \sum_{i=1}^N X_i \right\rangle = \sum_{i=1}^N \langle X_i \rangle = N\mu \quad (80)$$

また、分散についても

$$\Sigma \equiv \left\langle \left(\sum_{i=1}^N X_i - M \right)^2 \right\rangle = \left\langle \left(\sum_{i=1}^N X_i - N\mu \right)^2 \right\rangle = \left\langle \sum_{i,j}^N (X_i - \mu)(X_j - \mu) \right\rangle \quad (81)$$

$$= \sum_{i,j}^N \langle (X_i - \mu)(X_j - \mu) \rangle = \sum_{i,j}^N \sigma^2 \delta_{ij} \quad (\because X_i \text{ は独立}) = N\sigma^2 \quad (82)$$

と加法性が成り立つ。よって、独立同一分布に従う n 個の確率変数の平均

$$S_n = \frac{\sum_{i=1}^n X_i}{n} \quad (83)$$

の分布は平均 $\bar{\mu}_n = \mu$ 、分散 $\bar{\sigma}_n^2 = \frac{\sigma^2}{n}$ となる。

2.5.2 大数の(弱)法則

上の結果にチェビシェフの不等式を適用すると、

$$P(|S_n - \mu| \geq k\bar{\sigma}_n) \leq \frac{1}{k^2} \quad (84)$$

なので、任意の $\epsilon > 0$ に対して $k = \frac{\epsilon}{\bar{\sigma}_n} = \left(\frac{\sqrt{n}}{\sigma} \right) \epsilon$ とすれば

$$\lim_{n \rightarrow \infty} P(|S_n - \mu| \geq \epsilon) \leq \lim_{n \rightarrow \infty} \left(\frac{\sigma}{\epsilon} \right)^2 \frac{1}{n} = 0 \quad (85)$$

すなわち、 n サンプルの標本平均は $n \rightarrow \infty$ で期待値 μ に確率収束する。これは大数の弱法則と呼ばれる。

2.5.3 特性関数による証明

X_i の平均 μ が存在すれば、特性関数は $C(\xi) = 1 + i\xi\mu + o(\xi)$ ($\xi \rightarrow 0$) なので、

$$\lim_{n \rightarrow \infty} C_{S_n}(\xi) = \lim_{n \rightarrow \infty} \left(1 + \frac{i\xi\mu}{n} + \dots \right)^n = e^{i\xi\mu} \quad (86)$$

これは退化分布 $\delta(x - \mu)$ の特性関数。

2.6 中心極限定理

2.6.1 定理

平均と分散が有限（であれば各々違って良い）な互いに独立な確率変数 $\{X_k\}$ について、それら n 個の平均に \sqrt{n} をかけた量

$$Z_n \equiv \frac{1}{\sqrt{n}} \left(\sum_{k=1}^n \frac{X_k - \mu_k}{\sigma_k} \right) \quad (87)$$

の分布は標準正規分布に収束する。

2.6.2 特性関数による証明

特性関数を使うと簡単。

$$Y_k = \frac{X_k - \mu_k}{\sigma_k} \quad (88)$$

は平均 0、分散 1 に整えられた独立な確率変数だから

$$C_{Y_k}(\xi) = 1 - \frac{\xi^2}{2!} + \frac{\mu_3^k (i\xi)^3}{3!} + \dots \quad (89)$$

ここで、3 次以降の高次のモーメントは元の X_k がそれぞれ従う分布の形状によって一般にはそれぞれ違うので Y_k の 3 次のモーメントを μ_3^k と書いた。 X_k を同一な分布からとった場合 (IID) は $\mu_3^k = \mu_3$, $C_{Y_k}(\xi) = C_Y(\xi)$ と均一になる。

これより Z_n の特性関数を計算すると

$$C_{Z_n}(\xi) = \left\langle \exp \left\{ i\xi \left(\sum_k \frac{X_k - \mu_k}{\sqrt{n}\sigma_k} \right) \right\} \right\rangle = \left\langle \prod_k \left(\exp \left\{ \frac{i\xi(X_k - \mu_k)}{\sqrt{n}\sigma_k} \right\} \right) \right\rangle \quad (90)$$

$$= \prod_k \left\langle \exp \left\{ \frac{i\xi(X_k - \mu_k)}{\sqrt{n}\sigma_k} \right\} \right\rangle \quad (\because X_k \text{ は独立}) = \prod_k C_{\frac{Y_k}{\sqrt{n}}}(\xi) = \prod_k C_{Y_k} \left(\frac{\xi}{\sqrt{n}} \right) \quad (91)$$

$$= \prod_k \left[1 - \frac{\xi^2}{2n} + \frac{\mu_3^k}{6} \left(\frac{i\xi}{\sqrt{n}} \right)^3 + \dots \right] = \left[1 - \frac{\xi^2}{2n} + o\left(\frac{\xi^2}{n}\right) \right]^n \quad \left(\frac{\xi^2}{n} \rightarrow 0 \right) \quad (92)$$

従って、

$$\lim_{n \rightarrow \infty} C_{Z_n}(\xi) \sim \exp \left\{ -\frac{1}{2}\xi^2 \right\} \quad (93)$$

これは標準正規分布の特性関数。

2.6.3 安定分布

上の証明から、 $\{X_i\}$ の平均や分散が有限で無い時は中心極限定理が成り立たないことが分かる。例えば X_i が Cauchy 分布に従う場合は平均値は同一の Cauchy 分布に従う：

$$C_{\frac{X_1+X_2}{2}}(\xi) = \left(\exp \left\{ \frac{i\alpha\xi}{2} - \beta \left| \frac{\xi}{2} \right| \right\} \right)^2 = \exp \{ i\alpha\xi - \beta|\xi| \} = C_{X_1}(\xi). \quad (94)$$

このような性質を持つ分布を安定分布と呼ぶ。正規分布は安定分布の代表例である。

また、 $\{X_i\}$ が互いに独立で無い（相関がある）場合にも中心極限定理は成り立たない。

2.6.4 対数正規分布

対数正規分布：

$$\mathcal{LN}(\mu, \sigma, x) = \frac{1}{\sqrt{2\pi} \sigma x} \exp \left\{ -\frac{\ln x - \mu}{2\sigma^2} \right\} \quad (95)$$

は、 $\ln x$ が正規分布に従うような分布であり、以下に示すように乗算過程との関連もあり自然界で良く見出される。対数正規分布の特徴量は以下のように決まる

$$\begin{aligned} \text{期待値} \quad E &= e^{\mu + \frac{\sigma^2}{2}} \\ \text{分散} \quad V &= e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) = (e^{\sigma^2} - 1) E^2 \\ \text{(標準偏差)} \quad S &= (e^{\sigma^2} - 1)^{\frac{1}{2}} E = \left[\left(\frac{E}{M} \right)^{\frac{2}{3}} - 1 \right]^{\frac{1}{2}} E \\ \text{最頻値} \quad M &= e^{\mu - \sigma^2} = \left[\left(\frac{S}{E} \right)^2 + 1 \right]^{-\frac{3}{2}} E \\ \text{中央値} \quad C &= e^{\mu} = (E^2 M)^{\frac{1}{3}} \end{aligned} \quad (96)$$

ので、例えば

$$\text{最頻値の位置を一定に保つ} \rightarrow \delta\mu = \delta\sigma^2 \quad (97)$$

のようにパラメターを定める必要がある。

2.6.5 乗算ノイズ過程と対数正規分布

変数に白色ノイズが（加算的でなく）乗算的にかかる系：

$$x(t+1) = \eta(t)x(t) \quad (\eta(t) > 0) \quad (98)$$

を考えると、

$$x(t) = \left(\prod_{\tau=0}^{t-1} \eta(\tau) \right) x(0) \quad (99)$$

であるから、この式の両辺の対数をとると

$$\ln x(t) = \ln x(0) + \sum_{\tau=0}^{t-1} \ln \eta(\tau) \quad (100)$$

と、 $\ln x$ が独立な確率変数の和で書けるので、 $\ln x$ は t の増大とともに正規分布に収束する。すなわち、 x は対数正規分布に従う。

2.7 ベキ分布を生む過程の例

実データの分布では「ガウス分布やポワソン分布、指数分布的」なものに加えて「裾の幅広い分布、特にベキ分布」が良く観測される。中心極限定理から考えると不思議なこのような幅広い分布はどう理解したら良いのだろうか？

講義では「見ている量を構成している変数間の相関が無視できない」場合（臨界現象）や、「次元性と最適化の現れ」（フラクタル）と考えるのが筋が良い例を紹介した。これらは非常に重要で面白い話題であるが本「数理手法」の枠を少々超えるので、以下ではベキ分布を生む簡潔な数理過程として他に有名な2つの例について紹介する。

2.7.1 乗算ノイズ + 加算ノイズの系

変数に乗算的にかかる白色ノイズと加算的にかかる白色ノイズのある系：

$$x(t+1) = \eta(t)x(t) + \zeta(t) \quad (\eta(t) > 0, \zeta \text{の分布は対称}) \quad (101)$$

は $\langle \eta^2 \rangle > 1$, $\langle \log \eta \rangle < 0$ の時分散が発散する定常分布を持ち、その定常解の累積密度分布は $P(> x) = C/x^\beta$ とベキ分布になる。ベキ指数 β は条件 $\langle \eta^\beta \rangle = 1$ より決まる。

ノイズの強さについての条件

式 (101) の両辺の2乗の期待値をとると、

$$\langle x^2 \rangle = \langle (\eta x + \zeta)^2 \rangle = \langle \eta^2 \rangle \langle x^2 \rangle + \langle \zeta^2 \rangle \rightarrow \langle x^2 \rangle = \frac{\langle \zeta^2 \rangle}{1 - \langle \eta^2 \rangle}. \quad (102)$$

よって、 $\langle \eta^2 \rangle > 1$ が定常解の分散が発散するための条件。

このようなとき、加算的ノイズの与えるスケールのはるか上の領域を考えるとそこでは加算項が無視できるので、乗算ノイズ過程だけを考えればよく、この領域では基本的に $\ln x$ がランダムウォークするような過程である。乗算ノイズの期待値についての条件 $\langle \log \eta(t) \rangle < 0$ が満たされているとき、このランダムウォークの平均は負（下向き）になるので、このとき定常分布があることになる。

ベキ指数の決まりかた

x が分散が発散した定常分布に従うとき、この分布の特性関数 $C_X(\xi)$ を考えると式 (101) から

$$C_X(\xi) \equiv \langle e^{i\xi x} \rangle = \langle e^{i\xi \eta x} \rangle \langle e^{i\xi \zeta} \rangle = \left(\int_0^\infty p(\eta) C_X(\eta\xi) d\eta \right) (1 + o(i\xi)) \quad (\xi \rightarrow 0). \quad (103)$$

x の分布の2次のモーメントが発散しているので、特性関数の $\xi = 0$ 周りの展開は

$$C_X(\xi) = 1 - \alpha |\xi|^\beta + \dots \quad (\alpha : \text{const.}, 0 < \beta < 2) \quad (104)$$

の形になる。これを式 (103) に代入して展開の係数を両辺で比べると

$$\left(1 - \alpha \langle |\eta\xi|^\beta \rangle + \dots \right) (1 + o(i\xi)) = \left(1 - \alpha \langle \eta^\beta \rangle |\xi|^\beta + \dots \right) = 1 - \alpha |\xi|^\beta + \dots \quad (105)$$

よって、冪指数についての自己無撞着条件： $\langle \eta^\beta \rangle = 1$ が導かれる。

2.7.2 成長過程 (Yule-Simon 過程)

講義で紹介した「Wikipedia 中での各単語の出現回数の Zipf 則」のモデルとして、次のような簡単な過程を考える

1. 初期時刻 $t = 0$ で、単語数 N_0 から成る文書があるとする
2. 各ステップ (t) で 1 語づつ、次のルールで新しい単語を書き加えてゆく
 - (a) 確率 α : それまでの文書に登場したことのない新しい単語
 - (b) 確率 $1 - \alpha$: それまでの文書に既に現れている単語から、その出現頻度に比例して選ぶ

単語 i の各時点での出現回数を $n_i(t)$ 、文書の総単語数を $N(t) = \sum_i n_i(t)$ とすれば、各単語の出現後の出現回数の平均的な挙動は以下の微分方程式に従う：⁸

$$\frac{dn_i(t)}{dt} = (1 - \alpha) \frac{n_i(t)}{N(t)} = (1 - \alpha) \frac{n_i(t)}{N_0 + t} \sim (1 - \alpha) \frac{n_i(t)}{t} \quad (t \rightarrow \gg N_0). \quad (106)$$

これは変数分離形なので一般解は $n_i(t) = C_i t^{1-\alpha}$ であり、単語 i が初めて現れた時刻を τ_i とすれば $n_i(\tau_i) = 1$ より C_i も定まって

$$n_i(t) = \left(\frac{t}{\tau_i} \right)^{1-\alpha}. \quad (107)$$

$n_i(t)$ は t の単調増加関数で、また τ_i は時間的に一様に分布しているから、時刻 $t (\gg N_0)$ で n 回以上出現している単語の数は $n = \left(\frac{t}{\tau_c} \right)^{1-\alpha}$ となる時刻 τ_c に等しい。よって、単語の出現回数の累積分布は

$$P(> n) = \frac{\tau_c(n)}{t} = n^{-\frac{1}{1-\alpha}} \quad (108)$$

またこれより、単語の出現回数の分布は

$$P(n) = -\frac{d}{dn} P(> n) = \left(\frac{1}{1-\alpha} \right) n^{-(1+\frac{1}{1-\alpha})}. \quad (109)$$

これより、全く新しい単語が出現する確率 α が小さい極限でランクサイズ分布が Zipf 則になることや、中間の α では連続的に指数の違うべき分布となることが分かる。

以上でみたモデルは「系の大きさが成長し続けながら新規要素が出現する過程」として有名な Yule-Simon 過程の一例であり、モデリングの対象に応じて細かなバリエーションが色々あるがべき分布が導出される仕組みは基本的に同じである。

⁸ゆらぎもちゃんと考えた解もそんなにややこしくなく求められるが分布形状についての結論は同じになる。

3 複雑ネットワーク関係

1990年代の終わり頃から、生体内反応系の構造や人の繋がり、インターネットの構造などのデータが手に入りやすくなったことを契機として、それら現実ネットワークの構造についての研究が進んだ。以下ではネットワークの構造を特徴づけるのに用いられる特性量などを簡単に紹介する。

3.1 各種ネットワーク特性量の定義など

3.1.1 次数分布

次数の平均： $\langle k \rangle$ や次数の分布関数 $P(k)$ は基本的な特性量。格子（レギュラーグラフ）なら $P(k) = \delta(\langle k \rangle)$ （もしくは複数のデルタ関数）だし、(Erdős-Rényi) ランダムグラフなら期待値周りで指数的に減衰するのでどちらにしる次数分布は幅が狭い。

これに対し、現実の系で良く観測されるべき分布： $P(k) \sim k^{-\gamma}$ ($2 \leq \gamma \leq 3$) をもつ系は スケールフリーネットワーク と呼ばれる。

3.1.2 次数相関

Neighbor Connectivity

次数 k のノードに繋がっているノードの次数の平均：

$$\langle k_{nn} \rangle \equiv \sum_m m P(m|k) \quad (110)$$

は Neighbor Connectivity と呼ばれる。ただしここで $P(m|k)$ は次数 k のノードの隣接ノードの次数が m である条件付き確率である。 k に対して k_{nn} をプロットすれば、全体の次数-次数相関やさらには k についてのローカルな情報も読める。

Assortativity Coefficient

ネットワーク全体にわたる次数相関の指標として、リンク両端のノードの次数の Pearson 相関係数が用いられる。すなわち、リンクの両端のノードを k_i^α, k_i^β でラベルすれば

$$r \equiv \frac{\sum_i (k_i^\alpha - \langle k^\alpha \rangle) (k_i^\beta - \langle k^\beta \rangle)}{\sqrt{\sum_i (k_i^\alpha - \langle k^\alpha \rangle)^2} \sqrt{\sum_i (k_i^\beta - \langle k^\beta \rangle)^2}} \quad (111)$$

が、Assortativity Coefficient と呼ばれ良く使われる。

現実の系のネットワークでは正負 (assortative, disassortative) どちらの例も観察される (正の例：共著ネットワーク、負の例：代謝ネットワーク)。

3.1.3 クラスタリング係数

結合のクラスタリングの程度を測る指標として、(大域的) クラスタリング係数 (clustering coefficient) :

$$C \equiv 3 \times \frac{(\text{3つのノードが成す三角形の数})}{\text{連結した3つのノードの組の数}} \leq 1 \quad (112)$$

が良く使われる。これは、あるノード i に隣接しているノード間を繋ぐエッジの数 (繋がっている組の数) を m_i と書けば

$$C_i = \frac{2m_i}{k_i(k_i - 1)} \quad (113)$$

と計算される局所的なクラスタリング係数のネットワーク平均である。

3.1.4 ネットワーク上の距離

ネットワーク上でのノード間の距離は、それらをリンクに沿って結ぶ経路のうちの最短のものとして普通定義される (連結していないものについては便宜上 ∞ とする)。ノード間距離の計算法の代表例としては次の Dijkstra 法がある。

3.1.5 ネットワークの大きさ

ネットワークの大きさの指標として、ノード間距離 l_{ij} の平均 :

$$\langle l_{ij} \rangle = \left(\sum_{i < j}^N l_{ij} \right) / \frac{N(N-1)}{2} = \int_0^\infty l P(l) dl \quad (114)$$

やネットワーク直径 :

$$D = l_{max} \quad (115)$$

が使われる。平均ノード間距離は連結でないノード対があるだけで発散してしまう (そのような組を除いて定義してもいいが計算が面倒) ので、その逆数の平均 :

$$E \equiv \langle l_{ij}^{-1} \rangle = \frac{2}{N(N-1)} \sum_{i < j}^N \frac{1}{l_{ij}} \quad (116)$$

をもって代えることも多い。 E はネットワークの大域的効率と呼ばれる。

3.2 各種のノード中心性 (Centrality) 指標

ネットワーク全体の構造と共に、あるノードがどれだけそのネットワークの要になっているかということがしばしば重要となる。このようなノードの中心性を特徴付ける指標はいろいろあるが、代表的なものを以下に挙げる。

次数中心性

ノードの次数そのものも中心性の指標のひとつである。

固有ベクトル中心性

次数中心性と同じ考え方ながら、次数の大きいノードからのリンクの方が重要、... と考えた場合は既出の Page Rank ベクトルの最大固有ベクトル成分（や、講義で紹介した他のノード重要性指標）がノードの中心性の指標となる。

近接中心性

ネットワーク内の他の全てのノードとの距離が近いという指標：

$$C_{\text{近接}} = 1 / \left(\sum_j l_{ij} \right) \quad (117)$$

脆弱度

そのノードの除去によりネットワークの大域的効率がどれだけ低下するかという指標：

$$V_i = \frac{E - E(i)}{E} \quad (E(i) \text{ はノード } i \text{ を除去したネットワークの } E) \quad (118)$$

媒介中心性

ネットワーク内のノード対の最短経路がそのノードを経由する数：

$$b_i = \sum_{j < k \text{ (} j, k \neq i \text{)}} \frac{\#(j \rightarrow i \rightarrow k)}{\#(j \rightarrow k)} \quad (119)$$

は媒介中心性 (betweenness centrality) と呼ばれる。

3.3 現実のネットワークの構造の特徴

現実のネットワークに普遍に見られる特徴は以下のものである。一般に、このような特徴を指して「複雑ネットワーク」と呼ぶ。

特徴量	ランダムネットワーク ⁹	d 次元格子	現実のネットワーク
次数分布	ポアソン分布 $e^{-\langle k \rangle} \left(\frac{\langle k \rangle^k}{k!} \right)$	退化分布 $\delta(k - \langle k \rangle)$	冪テールの場合あり (Scale-Free)
次数相関	無し (= 0)	無し (デルタ関数的)	正負どちらもあり (assortative/dis-)
クラスタリング係数	$l_{rand} = p = \frac{\langle k \rangle}{N}$ ($\sim N^{-1}$)	高い ($\sim N^0$)	高い ($\sim N^0$)
ネットワーク直径	$l_{rand} \sim \frac{\ln(N)}{\ln(\langle k \rangle)}$	$l_d \sim N^{\frac{1}{d}}$	l_{rand} 的 (Small-World)

3.4 Scale-Free 性と Barabási-Albert モデル

旧来扱われてきた Erdős-Rényi ランダムグラフでは、ノードの次数は二項分布に従うことから指数関数より速く落ちるテールを持つ。また、分子のつながりのネットワークなどでも高次数のテールは幅広くなることができない。

現実のネットワークに普遍的に見られる「べき的テールを持つ次数分布」を説明するシンプルかつ代表的なモデルが以下の Barabási-Albert モデルである。

0. 連結な初期ネットワークを用意する

1. 新しいノードを一つ加える (ネットワークの成長)
2. 新しいノードと既存のノードを M 本のリンクで繋ぐ。その際、繋ぐ先は既存ノードの次数に比例して選ぶ (選択性)。1 に戻って逐次成長過程を繰り返す。

生成されたネットワークの次数分布がべきになる仕組みは既出の Yule-Simon 過程と同様に確かめることができるので、べきの指数を計算してみよう。

⁹Erdős-Rényi のタイプのランダムグラフ:「最初に N 個のノードを用意し、各ノードペアの間を確率 p で繋いで作ったネットワーク」

3.5 母関数とその活用による Small World 性の理解

3.5.1 母関数の定義

離散確率変数 ($X = 0, 1, 2, \dots$) の確率分布 $P(X)$ について、

$$G(x) = \sum_X P(X)x^X \quad (|x| \leq 1) \quad (120)$$

を (確率) 母関数という。特性関数と同様、 $G(x)$ は $P(x)$ についての情報を全て持つ。

3.5.2 母関数の一般的性質

- (i). $P(X)$ は確率分布なので上の範囲で絶対収束する (特異性無し)。
- (ii). $G(0) = \sum_X P(X)0^X = P(0)$ ($0^0 = 1$ を採用)。
- (iii). $G(1) = \sum_X P(X) = 1$.
- (iv). 確率分布の線型和の分布 $P(X) = pP_1(X) + (1-p)P_2(X)$ ($0 \leq p \leq 1$) の母関数は母関数の線型和: $G(x) = pG_1(x) + (1-p)G_2(x)$

- (v). 独立な確率変数の和: $X = \sum_{i=1}^n X_i$ の母関数は積になる: $G(x) = \prod_{i=1}^n G_i(x)$

(証明): $n = 2$ の場合について和積公式を示せば良い。 $X = X_1 + X_2$ として、

$$\begin{aligned} G_1(x)G_2(x) &= \left(\sum_{X_1=0}^{\infty} P_1(X_1)x^{X_1} \right) \left(\sum_{X_2=0}^{\infty} P_2(X_2)x^{X_2} \right) = \sum_{X_1, X_2} P_1(X_1)P_2(X_2)x^{X_1+X_2} \\ &= \sum_X P(X)x^X = G(x) \end{aligned}$$

- (vi). 畳み込み: $P(X) = \sum_{Z=0}^X P_1(Z)P_2(X-Z)$ の母関数も積: $G(x) = G_1(x)G_2(x)$

(証明)

$$G(x) = \sum_{X=0}^{\infty} \left(\sum_{Z=0}^X P_1(Z)P_2(X-Z) \right) x^X = \sum_{X=0}^{\infty} \sum_{Z=0}^X P_1(Z)x^Z P_2(X-Z)x^{X-Z}$$

$Y = X - Z$ とすると、元々の XY 平面での $(X, Z < X)$ の三角形領域に対応する点は $\lim_{X \rightarrow \infty}$ で YZ 平面の第一象限全体に写るから、

$$\text{上式} = \left(\sum_{Z=0}^{\infty} P_1(Z)x^Z \right) \left(\sum_{Y=0}^{\infty} P_2(Y)x^Y \right) = G_1(x)G_2(x)$$

- (vii). モーメント: $\langle X^n \rangle = \left[\sum_{X=0}^{\infty} X^n P(X)x^X \right]_{x=1} = \left[\left(x \frac{d}{dx} \right)^n G(x) \right]_{x=1}$

3.5.3 次数分布の母関数

ランダムに選ばれたノードの次数分布（そのネットワークの次数分布）が $P(k)$ であるとき、その母関数は

$$G_0(x) = \sum_{k=0}^{\infty} P(k)x^k \quad (121)$$

である。 $P(k) = \delta_{mk}$ の場合 $G(x) = x^m$ だから、 x はそれぞれの隣接ノードに対応し、また m 個の隣接ノードが同じノードにつながっていることがそれらの積 x^m に対応しているということができる。前述のモーメントの式より、平均次数やその揺らぎなどの母関数を用いた表式は以下のようになる。

$$\langle k \rangle = G'_0(1) \quad (\equiv \langle n_1 \rangle) \quad (122)$$

$$\langle k^2 \rangle = G'_0(1) + G''_0(1) \quad (123)$$

$$\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2 = G'_0(1) [1 - G'_0(1)] + G''_0(1) \quad (124)$$

3.5.4 隣接ノードと次近接ノード数の母関数

任意に抽出したノードの次数分布が $P(k)$ のとき、任意に抽出したエッジの片端のノードの次数分布は（エッジの数に比例した抽出確率を考慮して）

$$Q(k) = \frac{kP(k)}{\sum_{k=0}^{\infty} kP(k)} = \frac{kP(k)}{\langle k \rangle} \quad (125)$$

である。この次数分布の母関数は

$$F(x) = \sum_{k=0}^{\infty} \left(\frac{kP(k)}{\langle k \rangle} \right) x^k = x \left(\frac{G'_0(x)}{G'_0(1)} \right) \quad (126)$$

したがって、ランダムに選んだエッジの先のノードから（そのエッジの帰り道を除いて）出て行く次数の分布は

$$R(k) = \frac{(k+1)P(k+1)}{\langle k \rangle} \quad (127)$$

またその母関数は

$$G_1(x) = \sum_{k=0}^{\infty} \left(\frac{(k+1)P(k+1)}{\langle k \rangle} \right) x^k = \frac{G'_0(x)}{\langle k \rangle} = \frac{G'_0(x)}{G'_0(1)} \quad (128)$$

サイクルが無視できるとすると、次近接ノード数は上の $G_1(x)$ で扱う分布の起点ノード分布の次数分之和であるから、その分布の母関数は

$$G^{(2)}(x) = \sum_{k=0}^{\infty} P(k)[G_1(x)]^k = G_0(G_1(x)), \quad (129)$$

よって次近接ノード数の期待値は

$$\langle n_2 \rangle = \left[x \frac{d}{dx} G^{(2)}(x) \right]_{x=1} = G'_1(1)G'_0(G_1(1)) = G''_0(1) \quad (130)$$

3.5.5 l 次近接ノード数

式 (129) の導出過程をそのまま繰り返すことで、サイクルの寄与を無視した場合の l 次近接ノード数の分布の母関数は

$$G^{(l)}(x) = G_0 \left(\overbrace{G_1(G_1 \cdots G_1(G_1(x)) \cdots)}^{(l-1) \text{ 個}} \right) \quad (131)$$

となる。よって、 l 次近接ノード数の期待値は

$$\langle n_l \rangle = \left[\frac{d}{dx} G_0 \left(\overbrace{G_1(\cdots G_1(x)) \cdots}^{(l-1) \text{ 個}} \right) \right]_{x=1} \quad (132)$$

$$= \left[G'_0 \left(\overbrace{G_1(\cdots G_1(x)) \cdots}^{(l-1) \text{ 個}} \right) G'_1 \left(\overbrace{G_1(\cdots G_1(x)) \cdots}^{(l-2) \text{ 個}} \right) \cdots G'_1(G_1(x)) G'_1(x) \right]_{x=1} \quad (133)$$

$$= G'_0(1) [G'_1(1)]^{l-1} = \frac{[G''_0(1)]^{l-1}}{[G'_0(1)]^{l-2}} = \langle k \rangle \left(\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right)^{l-1} = \langle n_1 \rangle \left(\frac{\langle n_2 \rangle}{\langle n_1 \rangle} \right)^{l-1}. \quad (134)$$

これを足し上げて直ちに、あるノードから半径 L までの距離にあるノード数の平均は

$$N_L = \langle n_1 \rangle \left[\sum_{l=1}^L \left(\frac{\langle n_2 \rangle}{\langle n_1 \rangle} \right)^{l-1} \right] = \frac{\left(\frac{\langle n_2 \rangle}{\langle n_1 \rangle} \right)^L - 1}{\frac{\langle n_2 \rangle}{\langle n_1 \rangle} - 1}. \quad (135)$$

この個数の指数関数的増大：

$$N_L \sim \exp \left\{ L \ln \left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1 \right) \right\} \quad (136)$$

は、ランダムネットワークの直径がノード数の \log 的にしか増えないことと対応している

$$D_N \sim \left(\frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \right) \ln N \quad (137)$$

4 参考文献

4.1 参考図書

- 東京大学工学教程「微積分」(時弘哲治)
- 東京大学工学教程「線形代数 I」(室田一雄、杉原正顕)
- 東京大学出版会「線形代数入門」(斎藤正彦)
- 東京大学工学教程「確率・統計 I」(縄田和満)
- 東京大学出版会「データ科学のための微積分・線形代数」(藤原毅夫・藤堂眞治)
- 共立出版「複雑ネットワークとその構造」(矢久保考介)

4.2 講義中に紹介する/した話題関連の論文

1. PageRank と、他の重要性指標の試み
 - (a) “A new status index derived from sociometric analysis” Leo Katz, *Psychometrika* 18, 39-43 (1953).
<https://link.springer.com/article/10.1007/BF02289026>
 - (b) “PageRank: Standing on the Shoulders of Giants.”, M. Franceschet, *Communication of the ACM* 54, 6 (2011).
<https://dl.acm.org/doi/pdf/10.1145/1953122.1953146>
 - (c) “A dynamical systems approach to gross domestic product forecasting.”, A. Tacchella, D. Mazzilli, and L. Pietronero., *Nature Physics* 14, 861-865 (2018).
<https://doi.org/10.1038/s41567-018-0204-y>
 - (d) “Ecology of the digital world of Wikipedia”, F. Ogushi, J. Kertész, K. Kaski, and T. Shimada, *Scientific Reports* Vol. 11, 18371 (2021).
<https://doi.org/10.1038/s41598-021-97755-w>
2. スパースモデリングや次元削減関係
 - (a) “Regression Shrinkage and Selection Via the Lasso”, R. Tibshirani, *J. of Royal Statistical Society: Series B* 58, 267-288 (1996)
<https://rss.onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1996.tb02080.x>
 - (b) “Super-resolution imaging with radio interferometry using sparse modeling.”, M. Honma, K. Akiyama, M. Uemura, and S. Ikeda, *Publications of the Astronomical Society of Japan* 66 (5), 95 (1-14), (2014).
<https://doi.org/10.1093/pasj/psu070>

- (c) “Reducing the Dimensionality of Data with Neural Networks”, G. E. Hinton and R. R. Salakhutdinov, *Science* 313 (5786), 504-507 (2006).
<https://www.science.org/doi/10.1126/science.1127647>
3. 株式市場など経済データの分布則、都市のスケーリング則
- (a) “Scaling behaviour in the dynamics of an economic index.”, R. Mantegna and H. E. Stanley, *Nature* 376, 46-49 (1995).
<https://doi.org/10.1038/376046a0>
- (b) “Scaling behaviour in the growth of companies.”, M. Stanley, L. Amaral, S. Buldyrev et al. *Nature* 379, 804-806 (1996).
<https://doi.org/10.1038/379804a0>
- (c) “Identifying States of a Financial Market.”, M. Münnix, T. Shimada, R. Schäfer, et al., *Scientific Reports* 2, 644 (2012).
<https://doi.org/10.1038/srep00644>
- (d) “Growth, innovation, scaling, and the pace of life in cities.”, L. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. West, *PNAS* 104 (17), 7301-7306, (2007).
<https://doi.org/10.1073/pnas.0610172104>
4. 人や動物の移動・通信の分布関数とその役割
- (a) “Foraging success of biological Lévy flights recorded in situ.”, N. E. Humphries, H. Weimerskirch, N. Queiroz, E. J. Southall, and D. W. Sims, *PNAS* 109(19), 7169-7174, (2012).
<https://www.pnas.org/content/109/19/7169.full>
- (b) “Evidence of Lévy walk foraging in human hunter-gatherers.”, D. A. Raichlen, B. M. Wood, A. D. Gordon, A. Z. P. Mabulla, F.-W. Marlowe, and H. Pontzer, *PNAS* 111(2), 728-733, (2014).
<https://www.pnas.org/content/111/2/728.full>
- (c) “Generalized Lévy walks and the role of chemokines in migration of effector CD8+ T cells”, T. Harris, E. Banigan, D. Christian et al., *Nature* 486, 545-548, (2012).
<https://doi.org/10.1038/nature11098>
- (d) “Scaling identity connects human mobility and social interactions”, P. Deville, C. Song, N. Eagle, V. D. Blondel, A.-L. Barabási, and D. Wang, *PNAS* 113, 7047-7052 (2016)
<https://www.pnas.org/content/113/26/7047>
- (e) “Functional advantages of Lévy walks emerging near a critical point”, M. Abe, *PNAS* 117, 24336-24344 (2020)
<https://www.pnas.org/content/117/39/24336>

5. 人や自然現象の発生間隔分布（バースト）

- (a) “The origin of bursts and heavy tails in human dynamics.”, A.-L. Barabási, Nature 435, 207-211 (2005).
<https://doi.org/10.1038/nature03459>
- (b) “Darwin and Einstein correspondence patterns.”, J. Oliveira and A.-L. Barabási, Nature 437, 1251 (2005).
<https://doi.org/10.1038/4371251a>
- (c) “Correlated Dynamics in Egocentric Communication Networks.”, M. Karsai, K. Kaski, and J. Kertész, PLoS ONE 7, 7, (2012).
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0040612>

6. 対数正規分布や Zipf 則と乗算過程

- (a) “Stable Infinite Variance Fluctuations in Randomly Amplified Langevin Systems”, H. Takayasu, A. Sato, and M. Takayasu, Phys. Rev. Lett. 79, 966 (1997).
<https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.79.966>
- (b) “Zipf’s Law in Gene Expression.”, C. Furusawa and K. Kaneko, Phys. Rev. Lett. 90, 088102, (2003).
<https://doi.org/10.1103/PhysRevLett.90.088102>
- (c) “Ubiquity of log-normal distributions in intra-cellular reaction dynamics.”, C. Furusawa, T. Suzuki, A. Kashiwagi, T. Yomo, and K. Kaneko, The Biophysical Society of Japan 1, 25-31 (2005).
<https://doi.org/10.2142/biophysics.1.25>

7. 現実系のネットワークとその構造や性質関連

- (a) “Collective dynamics of ‘small-world’ networks”, D. J. Watts and S. H. Strogatz, Nature 393, 440-442 (1998).
<https://www.nature.com/articles/30918>
- (b) “Error and attack tolerance of complex networks.”, R. Albert, H. Jeong, and, A.-L. Barabási, Nature 406, 378-382, (2000).
<https://doi.org/10.1038/35019019>
- (c) “Catastrophic cascade of failures in interdependent networks”, S. Buldyrev, R. Parshani, G. Paul et al., Nature 464, 1025-1028, (2010).
<https://doi.org/10.1038/nature08932>
- (d) “Structure and tie strengths in mobile communication networks”, J. P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, PNAS 104 (18), 7332-7336, (2007).
<https://www.pnas.org/content/104/18/7332.full>