Markov Chain

李 仁 碩

article version
180705

대개 "학부 2학년 선형대수학" 강좌의 첫째 학기는 둘째 학기보다 분량이 많다. Cyclic Decomposition Theorem을 첫째 학기에 다루어야 하기 때문이다.[1] 그러나 Cyclic Decomposition Theorem은 학부 2학년에게는 좀(너무) 어려운 내용이다. 무엇보다도 그의 motivation을 설명할 길이 없기 때문이다.[2]

그래서 [I, §8.5–8.7]의 Cyclic Decomposition Theorem에 관한 내용은 적당히 '수박 겉 핥기' 정도로 끝내고, Jordan canonical form의 응용 — 예를 들어, Markov chain — 에 첫째 학기의 마지막 1주일 정도를 투자하는 것도 의미 있는 계획이라고 생각된다.

이 article에서는 우리가 흔히 **Markov chain**이라고 부르는 subject를 소개한다. Markov chain은 선형대수의 가장 유명한 — 그리고 가장 멋있는 — application이라고 할 수 있다. 이 article에서는 우선 Markov chain을 공부한 후, 이를 적용하는 다음

(가) 유전자形 빈도(genotype frequency),

(나) Google ranking system

에 관한 내용을 소개한다. 또, Markov chain의 '뿌리'라고 할 수 있는 Perron-Frobenius Theory도 (일부분이나마) 소개한다.

이 article을 읽기 위해서는 Jordan canonical form의 형태([I, 238쪽 $(\diamondsuit_j)^{\mathbf{t}}$] 와 [I, 239쪽 $\spadesuit_i$])만 알면 충분하다. 다시 말하면, Cyclic Decomposition Theorem을 몰라도 좋다는 뜻이다. 그리고, Jordan canonical form을 다루므로, 이 article에서는 (특별한 언급이 없는 한) 항상 $F = \mathbb{C}$ 라고 가정한다.

---

[1] 그래서, 예를 들어, [3]에서는 Cyclic Decomposition Theorem을 둘째 학기에 다루기도 한다. 그렇지만 "선형대수학1"에서는 1-차 연립방정식을 공부하고, "선형대수학2"에서는 2-차 연립방정식을 다룬다는 철학을 지키는 것이 바람직해 보인다.

[2] [II, 제12장] 참조.

section version

180705

제 1 절　선형대수와 극한

행렬의 극한은 componentwise 정의한다.

정의 1.1 $C_k = \big(c_{ij}^{(k)}\big) \in \mathfrak{M}_{m,n}(\mathbb{C})$ 일 때, 만약 모든 i, j 에 대하여 $\lim\limits_{k\to\infty} c_{ij}^{(k)}$ 가 존재하면, $\{C_k\}_{k=0}^{\infty}$ 가 **수렴**한다고 말하고,

$$\lim_{k\to\infty} C_k = \Big(\lim_{k\to\infty} c_{ij}^{(k)}\Big)$$

로 표기한다. 독자들은 $\mathfrak{M}_{m,n}(\mathbb{C}) = \mathbb{C}^{mn}$ 으로 identify 한다고 생각하는 쪽이 마음이 더 편할 것이다.

학부 2 학년 수준의 "선형대수학"에는 대개 다음 세 가지 상황(보기 1.2, 보기 1.3 및 보기 1.6) 에서 극한이 등장한다.

보기 1.2 (Exponential map) 함수 $\exp : \mathfrak{M}_{n,n}(\mathbb{C}) \to \mathfrak{M}_{n,n}(\mathbb{C})$ 는

$$\exp(A) = \sum_{k=0}^{\infty} \frac{1}{k\,!}\, A^k, \qquad (A \in \mathfrak{M}_{n,n}(\mathbb{C}))$$

로 정의한다. 이때 이 무한급수가 항상 수렴한다는 증명은 "해석개론" 책에서 찾을 수 있을 것이다.[3]

다음 보기는, 예를 들어, M. Artin [1, p. 119] 에서 발췌하였다.

보기 1.3 (Eigen-vector 의 motivation) (가) 우리는 행렬 $A = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} \in \mathfrak{M}_{n,n}(\mathbb{R})$ 이 (적어도 하나의) eigen-vector 를 갖는다는 사실을 다음과 같이 '직관적'으로 이해할 수 있다. 우선 $A\mathbf{e}_1 = (2,3)^{\mathbf{t}}$ 이고 $A\mathbf{e}_2 = (1,4)^{\mathbf{t}}$ 이므로, L_A 는 제 1 사분면 $S = \big\{a\mathbf{e}_1 + b\mathbf{e}_2 \mid a, b \geq 0\big\}$ 을 'cone' $AS = \big\{a(2,3)^{\mathbf{t}} + b(1,4)^{\mathbf{t}} \mid a, b \geq 0\big\}$ 으로 옮긴다. 계속해서 $A^2\mathbf{e}_1 = (7,18)^{\mathbf{t}}$ 이고 $A^2\mathbf{e}_2 = (6,19)^{\mathbf{t}}$ 이므로, L_A 는 AS 를 'cone' $A^2S = \big\{a(7,18)^{\mathbf{t}} + b(6,19)^{\mathbf{t}} \mid a, b \geq 0\big\}$ 으로 옮긴다. 이렇게 계속하면, 점점 폭이 좁아지는 'cone'들의 sequence

$$S \supseteq AS \supseteq A^2S \supseteq A^3S \supseteq A^4S \supseteq \cdots$$

을 얻게 된다.

[3] Exponential map 은 Lie group 과 Lie algebra 의 세계로 들어가는 입구이다.

(나) 결국 이 'cone'들은 half-line ℓ 로 '수렴'할 것이다.[4] 즉, $\ell = \bigcap_{m=0}^{\infty} A^m S$ 는 제 1 사분면의 (원점을 포함하는) half-line 이라는 뜻이다. 그런데,

$$A\ell = A\big(\bigcap_{m=0}^{\infty} A^m S\big) = \bigcap_{m=0}^{\infty} A^{m+1} S = \ell$$

이므로, L_A 는 half-line ℓ 을 보존할 것이다. 따라서 X 가 ℓ 의 임의의 (non-zero) vector 이면, $AX = \lambda X$ 인 A 의 eigen-value $\lambda \in \mathbb{R}$ 이 존재한다. 그리고 이때 $\lambda > 0$ 이어야 하는 것도 당연하다.

(다) 우리는 위에서 $\lambda = 5$ 이고, ℓ 은 $(1,3)^{\mathbf{t}}$ 을 포함하는 half-line 임을 잘 알고 있다([I, 보기 7.2.1 참조). (A 의 다른 eigen-value 는 1 이다.) 이 보기를 일반화한 것이 Perron-Frobenius Theory 이다(제 7 절 참조).

위 보기의 (절댓값이) 가장 큰 eigen-value $\lambda = 5$ 를 우리는 dominant eigen-value 라고 부른다.

정의 1.4 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 일 때, A 의 eigen-value(in $\mathbb{C}$) 중에서 maximum **absolute** value 를 갖는 것을 A 의 **dominant eigen-value** 라고 부른다. (물론 A 의 dominant eigen-value 는 유일하지 않을 수도 있다.) Dominant eigen-value 에 대응하는 eigen-vector 는 **dominant eigen-vector** 라고 부른다. (분야에 따라서는 **principal eigen-vector** 로 부르기도 한다.[5])

"Numerical Linear Algebra" 또는 "Matrix Computation" 의 최대 관심사는, 물론, 큰 size 의 행렬의 eigen-value(의 근삿값)와 그에 대응하는 eigen-vector(의 근삿값)를 구하는 것이다. 그러나 만약 n 이 10 억(약 2^{30}) 정도라면 characteristic polynomial 과 그의 10 억 개의 root 들을 구하는 방법은 상상하기조차 어렵다.

그런데 — 모든 eigen-vector 를 구할 수는 없지만 — dominant eigen-vector 를 구하는 방법이 알려져 있다. 사실 "Matrix Computation" 책에는, 약간 과장하면(!), 온통 dominant eigen-vector 에 관한 얘기뿐이다.[6]

---

[4] 이는 '직관적'으로는 거의 자명하지만, 엄밀한 증명을 위해서는 A 를 대각화한 후 A^m 을 구해야 할 듯하다. (그러나, 지금 목표는 eigen-vector 의 존재를 설명하는 것인데, 대각화하려면 eigen-vector 를 알아야 한다…….)

[5] 통계학의 "Principal Component Analysis(PCA)" 에서 'principal'도 비슷한 뜻일 것이다.

[6] "Matrix Computation" 분야의 참고도서로는 Golub-Van Loan[4]과 Meyer[5]가 널리 알려져 있다. 좀 두꺼운 책들이지만…….

주어진 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 의 dominant eigen-vector 를 계산하려면, 보기 1.3 을 다시 살펴볼 필요가 있다. 즉, $X \in \mathbb{C}^n$ 일 때, $\lim_{m\to\infty} A^m X$ 를 (만약 수렴한다면) 계산해 보자는 뜻이다.[7] 그러나, 예를 들어, 보기 1.3 의 경우에, m 이 커지면 A^m 이나 $A^m X$ 는 좌표들이 무한히 커질 수 있으므로, 당연히 수렴할 수 없다. 그렇다면 $\lim_{m\to\infty} \frac{A^m X}{\|A^m X\|}$ 는 어떨까?

그런데 $A, A^2, A^3, \ldots$ 을 계산하려면 각 step 마다 n^2-번의 '내적'을 계산해야 한다. 그러나 AX, $A(AX)$, $A(A^2X)$, $\ldots$ 을 계산한다면 각 step 마다 n-번의 '내적'만 계산하면 된다. 따라서 $\lim_{m\to\infty} A^m$ 이 아닌 $\lim_{m\to\infty} A^m X$ 를 생각한다. 만약 n 이 크다면, n^2 과 n 의 차이는 significant.

주의 1.5 "Numerical Linear Algebra" 는, 물론, 독자들이 지금까지 공부해 온 "Linear Algebra" 와는 분위기가 다르다.

(가) 우선 대부분의 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 는 가역행렬일 것이다. 만약 운이 정말 나빠 $\det(A) = 0$ 이었다면, A 의 좌표 하나(또는 몇 개)에 2^{-30} 쯤을 더해주면 다시 운이 좋아질 것이다. 이 새로운 행렬은 "선형대수학"의 입장에서는 본질적으로 다른 행렬이지만, "수치선형대수학"에서는 오십보백보.

(나) 마찬가지로 대부분의 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 는 diagonalizable 일 것이다. 왜냐하면 $\phi_A(t)$ 가 중근을 가질 가능성은 거의 없기 때문이다([I, 명제 7.2.18] 참조). 같은 이유로 대부분의 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 는 **unique** dominant eigen-value 를 가질 것이다.

보기 1.6 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$, $X \in \mathbb{C}^n$ 일 때, 이 보기에서는 '대충' 대부분의 경우에 $\lim_{m\to\infty} \frac{A^m X}{\|A^m X\|}$ 가 A 의 dominant eigen-vector 로 수렴함을 보일 것이다.

(가) 우선 당연히 모든 m 에 대해 $A^m X \neq 0$ 이라고 가정해야 한다(대부분의 경우 A 는 invertible 이므로, $X \neq 0$ 이면 '대충' OK). 이제 A 는 diagonalizable 이라고 가정하자. 즉, $AX_i = \lambda_i X_i$ 인 $\lambda_i \in \mathbb{C}$ 와 $X_i \in \mathbb{C}^n$ 이 존재하고,

$$U^{-1}AU = D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \qquad A^m X = UD^m U^{-1} X$$

로 쓸 수 있을 것이다. 이때 물론 $U = (X_1, \ldots, X_n)$ 이다([I, 관찰 7.2.13] 참조). 이제 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 는 **unique** dominant eigen-value λ_1 을 갖고, $\lambda_1 > 0$ 이라고 가정하자. 또 대부분의 경우 $j \neq 1$ 이면 $\lambda_1 \neq \lambda_j$ 일 것이므로, 그렇게 가정한다. 마지막으로 $U = (u_{ij})$, $U^{-1}X = (y_j)$ 로 표기하고, $y_1 \neq 0$ 이라고 가정하자.

[7] 보기 1.3 의 경우에 이를 달리 표현하면, μ 가 제1사분면의 (원점을 포함하는) half-line 일 때, $\lim_{m\to\infty} A^m \mu$ 를 계산해 보자는 뜻이다.

(나) 이제 $\lim_{m\to\infty}\frac{A^mX}{\|A^mX\|}$ 의 k-번째 좌표를 계산해 보면, $\lambda_1>0$ 은 unique dominant eigen-value 이고 $\lambda_1\neq\lambda_j$ 이므로(단, $j\neq 1$),

$$\frac{\sum_j u_{kj}\lambda_j^m y_j}{\sqrt{\sum_i|\sum_j u_{ij}\lambda_j^m y_j|^2}} = \frac{\lambda_1^m\sum_j u_{kj}\left(\frac{\lambda_j}{\lambda_1}\right)^m y_j}{\lambda_1^m\sqrt{\sum_i\left|\sum_j u_{ij}\left(\frac{\lambda_j}{\lambda_1}\right)^m y_j\right|^2}} \longrightarrow \frac{y_1}{\sqrt{\sum_i|u_{i1}y_1|^2}}u_{k1}$$

이 된다. 따라서 $\lim_{m\to\infty}\frac{A^mX}{\|A^mX\|}$ 는 $[U]^1=X_1$ 의 상수배이다($\sum_i|u_{i1}y_1|^2\neq 0$ 인 이유는?). 즉, $\lim_{m\to\infty}\frac{A^mX}{\|A^mX\|}$ 는 A 의 dominant eigen-vector.

(다) 독자들은 앞의 '대충'이라는 표현에 실망하고 있을지도 모르겠다. 위 (가)항에서 너무 많은 가정을 하고 있다고 생각할 수도 있다. 그렇지만, 주의 1.5 와 연습문제 1.7 및 아래 (마)항에 의하면, 이 가정들이 그리 대단한 건 아니다.

(라) 위 (가)항의 **essential assumption** 은 $\lambda_1>0$ 부분이다. (독자들은 위 (나)항의 증명에서 $\lambda_1>0$ 이라는 가정이 필수적임을 확인해 보기 바란다.) 사실 이 가정은 '대충'이라는 말로 적당히 넘어가기에 매우 어색하다. 임의의 복소수가 positive real number 일 **가능성은 거의 zero** 이기 때문이다.

(마) A 가 diagonalizable 이 아니더라도, A 의 Jordan canonical form 을 이용하면 '대충' 같은 결론을 얻을 수 있다. Detail 은 생략한다.[8]

(바) ("해석개론"을 수강한) 독자들은 위 (나)항에서 사용한 L^2-norm 대신 L^1-norm 을 생각해도 된다는 것을 알고 있을 것이다. 특별히 A 가 Markov matrix 이고 X 는 provability vector 이면(정의 2.2 및 연습문제 4.5(나)항 참조), 모든 m 에 대해 A^mX 의 L^1-norm 은 1 이므로, $\lim_{m\to\infty}A^mX$ 를 계산하면, A 의 dominant eigen-vector 를 구할 수 있다(따름정리 4.9 참조).[9]

연습문제 1.7 보기 1.6(가)항에서, $y_1\neq 0$ 일 필요충분조건은 $X\notin\langle X_2,\ldots,X_n\rangle$ 임을 보여라.[10]

그런데, 위와 같은 방법을 사용하면 실제로 $\lim_{m\to\infty}\frac{A^mX}{\|A^mX\|}$ 는 얼마나 빨리 수렴할까? Google Ranking System(이때 n 은 약 10억)에서는 대략 $A^{52}X$ 까지 계산하면 된다고 한다(제 6 절 참조).

[8] (한가한 시간에) 확인해 보기 바란다. 연습문제 3.2 참조.
[9] [6] 에도 L^1-norm 이 여러 번 강조되어 있다.
[10] Hint: $U\mathbf{e}_i=X_i$.

새로운 terminology 하나.

정의 1.8 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$, $X \in \mathbb{C}^n$ 일 때, AX, $A(AX)$, $A(A^2X)$, $\ldots$ 을 계산해 $\lim_{m\to\infty} \frac{A^mX}{\|A^mX\|}$ (의 근삿값)를 — 즉, A 의 dominant eigen-vector (의 근삿값)를 — 구하는 방법을 **power method** 라고 부른다.

주의 1.9 (가) 물론, power method 가 항상 성공하는 것은 아니다.
(나) 예를 들어, 만약 n 이 2^{30} 정도라면, A 와 X 가 보기 1.6 (가)항의 가정들을 만족시키는지 검증할 방법이 없다. 그래서 일단 power method 를 시행해 보고, $\lim_{m\to\infty} \frac{A^mX}{\|A^mX\|}$ 가 수렴하는지 살펴본다. 만약 수렴하지 않으면, A 와 X 의 몇몇 좌표에 2^{-30} 을 더해주는 '땜질'을 해 보고······. (이 '땜질'의 idea 가 바로 800 조 원짜리 Google Ranking System 의 핵심이다. 제 6 절 참조.)
(다) Power method 가 성공할 필요충분조건을 구하는 문제에는 별로 관심이 없다.[11]

애당초 A 의 **unique** dominant eigen-value 가 positive real number 가 아니었다면······, 이것도 별로 관심 없다. 따라서, 자연히, 우리는 unique dominant eigen-value 가 positive real number 임이 알려진 square matrix 에 관심을 갖게 된다. Positive Markov matrix (표기법 2.1 및 정의 2.2 참조)가 바로 그런 행렬이다. (더 일반적으로, non-negative matrix (표기법 2.1 참조)도 '많은 경우 대충' 비슷한 성질을 갖는다.[12] 우리는 non-negative matrix 의 dominant eigen-vector 에 관한 결과를 Perron-Frobenius Theory 라고 부른다.[13])

아차, 독자들이 (아직) $\mathbb{C}^n$-공간의 norm 과 거리의 개념에 익숙하지 않을 것 같다. 우리는 $X = (x_1, \ldots, x_n)^{\mathbf{t}} \in \mathbb{C}^n$ 일 때, X 의 norm (크기)을

$$\|X\| = \sqrt{|x_1|^2 + \cdots + |x_n|^2}$$

으로 정의한다.[14] 물론 $X, Y \in \mathbb{C}^n$ 사이의 거리는 $\|X - Y\|$.

---

[11] "관심 없다." = "자꾸 캐물으면 미워할 것이다."
[12] 제 7 절 및 Meyer [5] 참조. Non-negative case 는 제법 골치 아프다.
[13] 그러나, power method 에는 computer 의 도움이 필수적이므로, power method 를 염두에 두고 Perron-Frobenius Theory 가 만들어졌다고 하는 것은 語不成說. Perron-Frobenius Theorem 은 20세기 초에 증명되었다. 우리의 실생활 주변에 있는 행렬들은 대개 non-negative matrix 이므로, 오래 전부터 관심의 대상이 된 것은 매우 자연스러운 일이다.
[14] $\mathbb{C}^n$ 을 $\mathbb{R}^{2n}$ 과 naturally identify 했다고 생각할 수도 있다.

제 2 절 A Toy Example

section version 180102

이 節에서는 Markov matrix를 정의하고, toy example을 하나 소개한다.

표기법 2.1 $C=(c_{ij})\in\mathfrak{M}_{m,n}(\mathbb{C})$일 때, 만약 $[c_{ij}>0 \text{ for all } i,j]$이면, $C>0$으로 표기하고 C는 **positive**라고 말한다.[15] 또 만약 $[c_{ij}\geq 0 \text{ for all } i,j]$이면, $C\geq 0$으로 표기하고 C는 **non-negative**라고 말한다. 따라서, $X\in\mathbb{C}^n$일 때, $X>0$ 또는 $X\geq 0$의 표기법도 가능하다.

정의 2.2 (가) $X\in\mathbb{C}^n$일 때, 만약 $X\geq 0$이고 X의 좌표들의 합이 1이면, X를 **probability vector**라고 부른다. 또 $A=(a_{ij})\in\mathfrak{M}_{m,n}(\mathbb{C})$일 때, 만약 A의 모든 column들이 probability vector이면(즉, A의 모든 column sum이 1이면), A를 (**left**) **Markov matrix**라고 부른다(물론 이때 $A\geq 0$).[16]
(나) $A^{\mathbf{t}}$가 Markov matrix이면, A를 **right Markov matrix**라고 부른다.

Markov matrix는 여러 가지 특별한 성질을 갖는다. 우선:

관찰 2.3 (Right) Markov matrix는 항상 1을 eigen-value로 갖는다.

증명 : 먼저 $\phi_A(t)=\phi_{A^{\mathbf{t}}}(t)$이므로([I, 연습문제 7.6.24]), 형용사 'right'에는 신경 쓰지 않아도 된다. 이 관찰은 두 가지 방법으로 설명할 수 있다. 만약 A가 right Markov라면, A의 모든 row sum이 1이므로,

$$A\cdot(1,\ldots,1)^{\mathbf{t}}=1\cdot(1,\ldots,1)^{\mathbf{t}}$$

이 된다. 또, 만약 A가 Markov라면, $(A-I)$의 row들의 합이 zero vector이므로, $(A-I)$의 row들은 일차종속이고, 따라서 $\det(1\cdot A-I)=0$. □

우리가 행렬의 대각화를 공부할 때 (2×2)-행렬의 example이 결정적인 역할을 했듯이, Markov matrix의 경우에도 (2×2)-행렬부터 살펴볼 필요가 있다. 다음 toy example은 너무나 '인상적'이어서 [3]에서 (숫자도 바꾸지 않고) 그대로 복사하였다. (이 toy example의 idea는 제 5 절과 제 6 절(Random Surfer Model)에도 필요하다.)

[15] 이 용어는 positive definite matrix와 혼동할 여지가 있어 ("Linear Algebra"에서) 많이 사용하지는 않는다. "대수학 강의 홈페이지"의 [I, 참고 및 추가, 15.3] 참조.

[16] Markov matrix는 분야에 따라 stochastic matrix, probability matrix, transition matrix 등 다양한 이름으로 불리운다. Probability vector를 stochastic vector라고 부르기도 한다.

보기 2.4 (가) 어떤 도시(city)와 그 근교(suburbs)의 현재의 인구 분포 **확률벡터**(probability vector)를 $X = (c_0, s_0)^{\mathbf{t}}$ 라고 하자(따라서, 물론, $c_0, s_0 \geq 0$ 이고 $c_0 + s_0 = 1$).[17] 그리고,

$$A = \begin{pmatrix} 0.90 & 0.02 \\ 0.10 & 0.98 \end{pmatrix} = \begin{pmatrix} \text{city} \rightsquigarrow \text{city} & \text{suburbs} \rightsquigarrow \text{city} \\ \text{city} \rightsquigarrow \text{suburbs} & \text{suburbs} \rightsquigarrow \text{suburbs} \end{pmatrix}$$

라고 놓자. 이 표기법의 의미는, 예를 들어, 1년 후에는 city 인구 중 90%는 계속 city에 살고 있고, city 인구 중 10%는 suburbs로 이사 간다는 뜻이다. 그러면, 1년 후의 인구 분포 확률벡터 $(c_1, s_1)^{\mathbf{t}}$ 는

$$\begin{pmatrix} c_1 \\ s_1 \end{pmatrix} = \begin{pmatrix} 0.90\,c_0 + 0.02\,s_0 \\ 0.10\,c_0 + 0.98\,s_0 \end{pmatrix} = \begin{pmatrix} 0.90 & 0.02 \\ 0.10 & 0.98 \end{pmatrix} \begin{pmatrix} c_0 \\ s_0 \end{pmatrix} = AX$$

가 될 것이다. 마찬가지로, 2년 후의 인구 분포 확률벡터 $(c_2, s_2)^{\mathbf{t}}$ 는

$$\begin{pmatrix} c_2 \\ s_2 \end{pmatrix} = A \begin{pmatrix} c_1 \\ s_1 \end{pmatrix} = A^2 X$$

가 되고, m-년 후의 인구 분포 확률벡터는 $A^m X$ 가 된다.[18]

(나) 이제 A를 대각화하면,

$$U^{-1}AU = \begin{pmatrix} \frac{1}{6} & -\frac{1}{6} \\ \frac{5}{6} & \frac{1}{6} \end{pmatrix}^{-1} \begin{pmatrix} 0.90 & 0.02 \\ 0.10 & 0.98 \end{pmatrix} \begin{pmatrix} \frac{1}{6} & -\frac{1}{6} \\ \frac{5}{6} & \frac{1}{6} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0.88 \end{pmatrix} = D$$

가 된다. 그러므로,

$$A^m = UD^mU^{-1} = \begin{pmatrix} \frac{1+5\cdot(0.88)^m}{6} & \frac{1-(0.88)^m}{6} \\ \frac{5-5\cdot(0.88)^m}{6} & \frac{5+(0.88)^m}{6} \end{pmatrix}, \qquad \lim_{m\to\infty} A^m = \begin{pmatrix} \frac{1}{6} & \frac{1}{6} \\ \frac{5}{6} & \frac{5}{6} \end{pmatrix}$$

이고,

$$A^m X = \begin{pmatrix} \frac{1+5\cdot(0.88)^m}{6}\, c_0 + \frac{1-(0.88)^m}{6}\, s_0 \\ \frac{5-5\cdot(0.88)^m}{6}\, c_0 + \frac{5+(0.88)^m}{6}\, s_0 \end{pmatrix}, \qquad \lim_{m\to\infty} A^m X = \begin{pmatrix} \frac{1}{6} \\ \frac{5}{6} \end{pmatrix}$$

가 된다(확인해 보라).

---

[17] c for city, s for suburbs.

[18] 굳이 언급할 필요도 없겠지만……, 이 mathematical model에는 A는 항상 constant이고, 외부로(부터)의 인구 유출(유입)은 없다는 등의 (약간 비현실적인) 가정들이 필요할 것이다. 아, 그리고, 인구조사는 1년에 한 번만 한다.

(다) 행렬 A를 살펴보면, suburbs 거주자가 city 거주자보다 많아지는 것은 당연해 보인다. 그렇지만 인구 분포 확률벡터가 $\left(\frac{1}{6}, \frac{5}{6}\right)^{\mathbf{t}}$라는 equilibrium(평형점)에 도달하는 것은 직관적으로 자명해 보이지 않는다……. 오히려 city 거주자가 점점 감소해 0으로 수렴할 것으로 착각하기 쉽다. (물론 suburbs 거주자의 2%가 매년 city로 유입되고, 그 숫자는 suburbs 거주자가 증가할수록 커진다.)
(라) A는 물론 positive Markov matrix이다. 그리고 1은 A의 unique dominant eigen-value이고, 그에 대응하는 dominant eigen-vector는 U의 첫 번째 column $\left(\frac{1}{6}, \frac{5}{6}\right)^{\mathbf{t}}$이다. (Dominant eigen-vector 중에서 probability vector는 $\left(\frac{1}{6}, \frac{5}{6}\right)^{\mathbf{t}}$뿐.) 또, 위 (나)항에서 $\lim_{m\to\infty} A^m X$가 A의 dominant eigen-vector로 수렴하는 것은 보기1.6의 예상과 일치한다.
(마) 한편, $\lim_{m\to\infty} A^m$의 두 column이 모두 A의 dominant eigen-vector인 것은 무척 신기하다.
(바) 가장 예상하기 힘든 부분은 인구 분포 확률벡터의 equilibrium $\left(\frac{1}{6}, \frac{5}{6}\right)^{\mathbf{t}}$가 현재의 인구 분포 확률벡터 $X = (c_0, s_0)^{\mathbf{t}}$와는 무관하다는 점이다. 극단적으로 설령 $c_0 = 0$이거나 $s_0 = 0$이더라도 equilibrium에는 변화가 없다.
(사) 위 (나)항에서 A^m과 $A^m X$의 수렴 속도는, 지수함수 0.88^m이 0으로 수렴하는 속도와 같으므로, 매우 빠를 것이다.

이 toy example을 [I, 제7장]에 수록했어도 좋았을 것이다. 물론 이 '인상적'인 toy example의 신기한 결과들은 우연이 아니다(정리4.8과 따름정리4.9 참조).

제 3 절 Preliminary Results

section version
180102

지금부터는, 물론 $F=\mathbb{C}$ 이고, **항상** $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 라고 가정한다. 이제 A의 characteristic polynomial을

$$\phi_A(t) = (t-\lambda_1)^{e_1}(t-\lambda_2)^{e_2}\cdots(t-\lambda_k)^{e_k}$$

라고 놓자(단, $e_i \geq 1$, λ_i는 distinct). 이때 e_i를 eigen-value λ_i의 **multiplicity** 라고 부르고, $e_i = \mathrm{mult}_A(\lambda_i)$로 표기하자. 또 물론 A의 λ-eigen-space는

$$E_\lambda^A = \ker(A-\lambda I) = \{X \in \mathbb{C}^n \mid AX = \lambda X\}$$

로 정의한다([I, 정의 7.6.20] 참조).

만약 다음 연습문제에 '목숨'을 걸 수 없다면 $\cdots\cdots$, 이 article은 다음으로 미루고 "선형대수학"을 재수강해야 할 것이다!

연습문제 3.1 **Prove or disprove:** $\dim E_\lambda^A = \mathrm{mult}_A(\lambda)$.[19]

우선 Jordan canonical form부터 복습해 보자.[20] A와 similar한 Jordan canonical form J는 **λ_i-Jordan block** $J_i \in \mathfrak{M}_{e_i,e_i}(\mathbb{C})$들로 이루어진 block diagonal matrix이다. 또 각각의 J_i는 **λ_i-Jordan 小-block** $J_{ij} \in \mathfrak{M}_{r_{ij},r_{ij}}(\mathbb{C})$들로 이루어진 block diagonal matrix이다. 즉,

$$A \sim J = \mathrm{diag}(J_1, J_2, \ldots, J_k), \qquad J_i = \mathrm{diag}(J_{i1}, J_{i2}, \ldots, J_{ih_i})$$

이다.[21] 이때 λ-Jordan 小-block J_{ij}의 일반적인 형태는 다음

$$K_\lambda = (\lambda), \quad \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \quad \ldots, \quad \begin{pmatrix} \lambda & 1 & & & & \\ & \lambda & 1 & & \mathbf{0} & \\ & & \cdot & \cdot & & \\ & & & \cdot & \cdot & \\ & \mathbf{0} & & & \lambda & 1 \\ & & & & & \lambda \end{pmatrix}$$

과 같다. 따라서, 특별히 A가 diagonalizable이면, A의 Jordan canonical form은 $\mathrm{diag}(\lambda_1 I_{e_1}, \lambda_2 I_{e_2}, \ldots, \lambda_k I_{e_k})$가 된다.

[19] 만약 λ가 A의 eigen-value가 아니면, $\dim E_\lambda^A = 0 = \mathrm{mult}_A(\lambda)$.
[20] 다시 강조하면, Cyclic Decomposition Theorem은 몰라도 OK.
[21] Jordan block이라는 용어는 그 의미가 분명치 않아, Jordan 小-block과 구분하였다.

물론 Jordan canonical form 은 (upper-)triangular matrix 이고, triangular matrix 는 diagonal matrix 다음으로 예쁜 행렬이다.

우리가 Jordan canonical form 을 특히 더 예뻐하는 이유 중 하나는 다음 연습문제 때문이다. 즉, 임의의 자연수 m 에 대하여 J^m 을 묘사할 수 있기 때문이다. 그리고, J 가 A 의 Jordan canonical form 일 때, J^m 을 묘사할 수 있다면, $\{A^m\}_{m=0}^{\infty}$ 의 수렴 여부도 알 수 있기 때문이다.

연습문제 3.2 위에서 $K_\lambda = \lambda I + N \in \mathfrak{M}_{r,r}(F)$ 로 표기할 때, $N^2, N^3, \ldots$ 을 묘사하라. 또, m 이 자연수이고, $m \geq r-1$ 일 때,

$$(K_\lambda)^m = \begin{pmatrix} \lambda^m & m\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \binom{m}{3}\lambda^{m-3} & \cdots & \binom{m}{r-1}\lambda^{m-r+1} \\ & \lambda^m & m\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \cdots & \binom{m}{r-2}\lambda^{m-r+2} \\ & & \ddots & \ddots & \ddots & \vdots \\ & \mathbf{0} & & \ddots & \ddots & \vdots \\ & & & & \lambda^m & m\lambda^{m-1} \\ & & & & & \lambda^m \end{pmatrix}$$

임을 보여라.[22] (우리는, 대개, $m < i$ 일 때는 binomial coefficient 를 $\binom{m}{i} = 0$ 으로 정의한다. 그러면, 위 결과는 $\lambda \neq 0$ 이고 $m < r-1$ 일 때에도 유효하다. 물론 $\lambda = 0$ 이면, $(K_\lambda)^m = N^m$.[23])

행렬의 극한에 관한 다음 연습문제는, 솔직히, 종이가 아깝다.

연습문제 3.3 $A_k \in \mathfrak{M}_{n,n}(\mathbb{C})$, $B \in \mathfrak{M}_{m,n}(\mathbb{C})$, $C \in \mathfrak{M}_{n,r}(\mathbb{C})$ 이고 $U \in \mathfrak{M}_{n,n}(\mathbb{C})$ 는 가역일 때, 다음을 보여라.

(가) $\{A_k\}_{k=0}^{\infty}$ 가 수렴할 때, $\lim\limits_{k\to\infty} A_k = L$ 이라고 놓자. 그러면, $\{BA_k\}_{k=0}^{\infty}$ 와 $\{A_kC\}_{k=0}^{\infty}$ 도 수렴한다. 이때 $\lim\limits_{k\to\infty}(BA_k) = BL$ 이고 $\lim\limits_{k\to\infty}(A_kC) = LC$.
(나) $\{U^{-1}A_kU\}_{k=0}^{\infty}$ 가 수렴할 필요충분조건은 $\{A_k\}_{k=0}^{\infty}$ 가 수렴하는 것이다. 이때 $\lim\limits_{k\to\infty} A_k = L$ 이라고 놓으면, $\lim\limits_{k\to\infty}(U^{-1}A_kU) = U^{-1}LU$.

---

[22] Hint: λI 와 N 은 서로 commute 하므로 — 즉, $(\lambda I)N = N(\lambda I)$ 이므로 — $(\lambda I + N)^m$ 의 2-항 전개가 가능하다.
[23] 물론 N 은 nilpotent.

다음 표기법은 단지 종이를 절약하기 위한 것이다.

표기법 3.4 $\mathfrak{D} = \{\lambda \in \mathbb{C} \mid |\lambda| < 1 \text{ or } \lambda = 1\}$.[24]

위 표기법의 의미는 자명하다. 즉, $\{\lambda^m\}_{m=0}^{\infty}$ 이 수렴하기 위한 필요충분조건은 $\lambda \in \mathfrak{D}$. 이제 $\{A^m\}_{m=0}^{\infty}$ 의 수렴 여부를 판정할 수 있다.

명제 3.5 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 일 때, $\{A^m\}_{m=0}^{\infty}$ 가 수렴할 필요충분조건은

(i) $\lambda \in \mathbb{C}$ 가 A 의 eigen-value 이면, $\lambda \in \mathfrak{D}$ 이고,

(ii) 만약 1 이 A 의 eigen-value 라면, $\dim E_1^A = \mathrm{mult}_A(1)$ (즉, A 의 1-Jordan block 은 항등행렬).

증명 : J 를 A 의 Jordan canonical form 이라고 하면, 연습문제 3.3 (나) 항에 의해, $\{A^m\}_{m=0}^{\infty}$ 가 수렴할 필요충분조건은 $\{J^m\}_{m=0}^{\infty}$ 가 수렴하는 것이다. 이제 연습문제 3.2 의 $(K_\lambda)^m$ 을 살펴보면 거의 자명. (다음 연습문제도 필요.) □

독자들은 이제 앞 연습문제 3.1 을 다시 기억하면서, 다음 연습문제를 분명히 해 두기 바란다.

연습문제 3.6 λ 가 A 의 eigen-value 일 때, 다음 조건

(1) $\dim E_\lambda^A = \mathrm{mult}_A(\lambda)$.

(2) A 의 λ-Jordan block 은 λI.

은 동치임을 보여라.

"Linear Algebra" 에서는 "A 의 eigen-value 를 $\lambda_1, \ldots, \lambda_k$ 라고 놓자"고 하면 그것으로 끝이었다. 그렇지만, "Numerical Linear Algebra" 에서는 — 명제 3.5 에서 보듯이 — eigen-value 의 크기 (절댓값) 가 문제가 된다. Eigen-value 의 크기를 조사하기 위해서는 다음 정의가 필수적이다.

정의 3.7 $A = (a_{ij}) \in \mathfrak{M}_{n,n}(\mathbb{C})$ 일 때,

$$\rho_i(A) = \sum_{j=1}^n |a_{ij}|, \qquad \nu_j(A) = \sum_{i=1}^n |a_{ij}|$$

로 표기하자. 또 A 의 **row sum** 과 **column sum** 을 각각

$$\rho(A) = \max\{\rho_i(A) \mid 1 \le i \le n\}, \qquad \nu(A) = \max\{\nu_j(A) \mid 1 \le j \le n\}$$

으로 정의하자.

[24] $\mathfrak{D}$ 는 disk의 첫 글자.

다음 명제와 그 증명은 정말 simple 하지만, 그 결과는 매우 유용하다.

명제 3.8 (Gershgorin's Disk Theorem, 1931) 행렬 $A = (a_{ij}) \in \mathfrak{M}_{n,n}(\mathbb{C})$ 의 eigen-value λ 에 대응하는 eigen-vector 를 $X = (x_i) \in \mathbb{C}^n$ 이라고 표기하자. 이때 $|x_k| = \max\{|x_i| \,|\, 1 \le i \le n\}$ 이라고 놓으면,[25] 다음 부등식

$$|\lambda - a_{kk}| \le \rho_k(A) - |a_{kk}|$$

이 성립한다.[26]

증명 : $AX = \lambda X$ 이므로, $\sum_{j=1}^n a_{kj} x_j = \lambda x_k$ 이고, 따라서

$$\sum_{j \ne k} a_{kj} x_j = \lambda x_k - a_{kk} x_k$$

이다. 이제, 당연히 $x_k \ne 0$ 이므로,

$$|\lambda - a_{kk}| = \left| \frac{\sum_{j \ne k} a_{kj} x_j}{x_k} \right| \le \sum_{j \ne k} |a_{kj}| = \rho_k(A) - |a_{kk}|$$

가 된다.[27] □

이제 정의 3.7 의 row sum 과 column sum 의 역할이 드러난다.

따름정리 3.9 λ 가 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 의 eigen-value 이면, $|\lambda| \le \min\{\rho(A), \nu(A)\}$.

증명 : Gershgorin's Disk Theorem 에 의하면,

$$|\lambda| \le |\lambda - a_{kk}| + |a_{kk}| \le \big(\rho_k(A) - |a_{kk}|\big) + |a_{kk}| = \rho_k(A) \le \rho(A)$$

가 된다. 한편 λ 는 $A^{\mathbf{t}}$ 의 eigen-value 이기도 하므로 (왜 그런가?),

$$|\lambda| \le \rho(A^{\mathbf{t}}) = \nu(A)$$

이다. □

Gershgorin's Disk Theorem 과 따름정리 3.9 의 결과는 약간 미묘하게 느껴진다. 왜냐하면, 행렬 A 의 eigen-value 는 A 의 similarity class 의 invariant 이지만, $\rho(A)$, $\nu(A)$ 등은 similarity class 의 invariant 가 아닌데 ······. 아, 어쨌든, 행렬 A 의 eigen-value 는 A 의 n^2-개 성분들이 결정하는 것은 분명하다.

[25] 따라서 k 는 유일하지 않을 수도 있다.

[26] 중심이 a_{kk} 이고 반지름이 $\rho_k(A) - |a_{kk}|$ 인 disk 를 Gershgorin's disk 라고 부른다.

[27] 이렇게 단순 명료하고 초보적인 명제가 1931 년에야 알려진 것은 무척 놀라운 사실이다. (아마 A. Markov (1856–1922) 도 이 결과를 몰랐을 가능성이 크다.) 이에 관한 historical comment 는 Meyer [5, p. 497] 참조.

Positive matrix의 eigen-value에 관한 다음 명제는 'a technical lemma'라고 부르고 싶다. 증명의 motivation은 잘 모르겠고······.[28]

명제 3.10 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$일 때, $A > 0$이라고 하자. 또 A는 $|\lambda| = \rho(A)$인 eigen-value λ를 갖는다고 가정하자.[29] 그러면, 다음이 성립한다.

(가) $\lambda = \rho(A) > 0$.

(나) $\dim E_\lambda^A = 1$. 사실은 $E_\lambda^A = \big\langle (1,1,\ldots,1)^{\mathbf{t}} \big\rangle$.

증명 : 명제 3.8 (Gershgorin's Disk Theorem)의 표기법을 계속 사용하면,

$$|\lambda|\cdot|x_k| = \big|\textstyle\sum_j a_{kj}x_j\big| \le \sum_j |a_{kj}x_j| \le \sum_j |a_{kj}|\cdot|x_k| = \rho_k(A)\cdot|x_k| \le \rho(A)\cdot|x_k|$$

가 된다. 그런데, $|\lambda| = \rho(A)$라고 가정하고 있으므로, 위에서 모든 부등호($\le$)는 사실은 등호($=$)여야 한다. 즉, 다음 등식들

(i) $\big|\sum_j a_{kj}x_j\big| = \sum_j |a_{kj}x_j|$,

(ii) $\sum_j |a_{kj}x_j| = \sum_j |a_{kj}|\cdot|x_k|$,

(iii) $\rho_k(A) = \rho(A)$

이 성립한다. 우선 등식 (i)로부터, 우리는

$$a_{kj}x_j = c_j z, \quad (1 \le j \le n), \qquad |z| = 1$$

인 real number $c_1, \ldots, c_n \ge 0$과 $z \in \mathbb{C}$가 존재함을 알 수 있다(아래 연습문제 3.11 참조). 그리고, $A > 0$이므로, 등식 (ii)로부터는

$$|x_j| = |x_k|, \qquad (1 \le j \le n)$$

임을 알 수 있다. 따라서, 계속 $A > 0$임에 유의하면,

$$a_{kj}\cdot|x_k| = a_{kj}\cdot|x_j| = |c_j z| = c_j, \qquad (1 \le j \le n)$$

이다. 그러므로,

$$x_j = \tfrac{c_j}{a_{kj}}\, z = |x_k|\cdot z, \qquad (1 \le j \le n)$$

가 된다. 즉, $X = |x_k|\cdot z(1,1,\ldots,1)^{\mathbf{t}}$이다. 이는 λ에 대응하는 eigen-vector는 모두 $\mathbf{1} = (1,1,\ldots,1)^{\mathbf{t}}$의 상수배라는 뜻이다. 물론 $\mathbf{1}$도 λ에 대응하는 eigen-vector. 한편 $A\mathbf{1} > 0$임은 당연하고, $A\mathbf{1} = \lambda\mathbf{1}$이므로, $\lambda > 0$.[30] □

---

[28] 명제 3.10의 증명은 [3]를 참조하였다.

[29] 따라서, 따름정리 3.9에 의해, λ는 A의 dominant eigen-value.

[30] 이 증명에서 등식 (iii)은 이용되지 않았다.

연습문제 3.11 $\alpha_1, \ldots, \alpha_r \in \mathbb{C}$ 가 다음 조건

$$|\alpha_1 + \cdots + \alpha_r| = |\alpha_1| + \cdots + |\alpha_r|$$

을 만족시키면,

$$\alpha_i = c_i \beta, \qquad (i = 1, \ldots, r)$$

인 $\beta \in \mathbb{C}$ 와 $0 \le c_1, \ldots, c_r \in \mathbb{R}$ 이 존재함을 보여라. 물론 이때 $|\beta| = 1$ 이라고 해도 OK.[31]

연습문제 3.12 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 일 때, $A > 0$ 이라고 하자. 또 A 는 $|\lambda| = \nu(A)$ 인 eigen-value λ 를 갖는다고 가정하자. 이때 $\lambda = \nu(A) > 0$ 이고, $\dim E^A_\lambda = 1$ 임을 보여라.[32]

보기 3.13 예를 들어, $\begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix}$ 이나 $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ 을 생각하면, 명제 3.10 은 A 가 positive 가 아닐 때는 어림도 없는 얘기이다. 확인해 보라.[33]

이제 준비가 거의 끝났다. 마지막으로:

정의 3.14 $A = (a_{ij}) \in \mathfrak{M}_{n,n}(\mathbb{C})$ 일 때, A 의 **matrix norm** 을

$$\|A\| = \max\left\{ |a_{ij}| \;\middle|\; 1 \le i, j \le n \right\}$$

으로 정의한다.[34]

다음은 독자들에게 맡긴다. 거의 자명.

연습문제 3.15 $A, B \in \mathfrak{M}_{n,n}(\mathbb{C})$ 일 때, 다음을 보여라.

(가) $A \neq 0$ 이면, $\|A\| > 0$.
(나) $c \in \mathbb{C}$ 이면, $\|cA\| = |c| \cdot \|A\|$.
(다) $\|A + B\| \le \|A\| + \|B\|$.
(라) $\|AB\| \le n\|A\| \cdot \|B\|$.

---

[31] Hint: r 에 관한 귀납법. "대수학 강의 홈페이지"의 [I, 참고 및 추가, 7.5] 참조.
[32] [I, 연습문제 7.6.24(다)] 참조. (지금은 $(1, 1, \ldots, 1)^{\mathbf{t}}$ 가 eigen-vector 일 필요는 없다.)
[33] 두 번째 행렬은 Markov matrix.
[34] 물론 matrix norm 은 여러 종류가 있다. 방금 정의한 것은 흔히 max norm 이라고도 부른다. 만약 $\mathfrak{M}_{n,n}(\mathbb{C}) = \mathbb{C}^{n^2}$ 으로 identify 한다면, max norm 은 L^∞-norm 이다. Matrix norm 은 $\exp(A)$ 의 수렴성을 증명할 때도 필요하다(보기 1.2 참조).

section version

150909

제 4 절 Markov Matrix

이 節에서는 제 2 절 toy example 의 결과가 우연이 아님을 증명한다. 우선 제 3 절의 결과들을 Markov matrix 에 적용해 보자.

관찰 4.1 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 가 Markov matrix 이면,

(가) $\nu_j(A) = \nu(A) = 1$ for all $j = 1, \ldots, n$.

(나) $\lambda \in \mathbb{C}$ 가 A 의 eigen-vector 이면, $|\lambda| \le 1$.

(다) 1 is **a** dominant eigen-value of A.

증명 : (가) 항은 Markov matrix 의 정의로부터 자명. (나) 항은 따름정리 3.9 (Gershgorin's Disk Theorem 의 따름정리) 와 (가) 항의 결과. (다) 항은 관찰 2.3 과 (나) 항의 결과. □

다음에는 positive Markov matrix 를 생각한다.

관찰 4.2 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 가 **positive** Markov matrix 이면,

(가) $\lambda \in \mathbb{C}$ 가 A 의 eigen-vector 일 때, 만약 $\lambda \neq 1$ 이면, $|\lambda| < 1$. 즉, 1 은 A 의 **unique** dominant eigen-value.

(나) $\dim E_1^A = 1$.

증명 : 연습문제 3.12 (명제 3.10 의 'transpose version') 와 관찰 4.1 의 direct consequence. □

연습문제 4.3 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 가 right Markov matrix 일 때, 관찰 4.1 과 관찰 4.2 에 대응하는 명제를 쓰고 증명하라.

위 관찰 4.2 에서, A 가 positive Markov 이면, $\dim E_1^A = 1$ 임을 알게 되었지만, 아직 $\dim E_1^A = \mathrm{mult}_A(1)$ 인 것도 알 수 있는 것은 아니다. 바로 이 부분이 Markov chain 의 **essence** 이다!

다음 easy exercise 들은 독자들의 몫이다. (편의상 $\mathbf{1} = (1, 1, \ldots, 1)^{\mathbf{t}}$ 로 표기.)

연습문제 4.4 다음을 보여라.

(가) $0 \le A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 가 Markov matrix 일 필요충분조건은 $A^{\mathbf{t}} \cdot \mathbf{1} = \mathbf{1}$.

(나) $0 \le X \in \mathbb{C}^n$ 이 probability vector 일 필요충분조건은 $X^{\mathbf{t}} \cdot \mathbf{1} = (1)_{1\times 1}$.

연습문제 4.5 다음을 보여라.

(가) $A, B \in \mathfrak{M}_{n,n}(\mathbb{C})$가 Markov matrix이면, AB도 Markov matrix.

(나) $A \in \mathfrak{M}_{n,n}(\mathbb{C})$가 Markov matrix이고, $X \in \mathbb{C}^n$이 probability vector이면, AX도 probability vector.

연습문제 4.6 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$가 right Markov matrix일 때, 연습문제 4.4와 연습문제 4.5에 대응하는 명제를 쓰고 증명하라.[35]

'Final touch'는 matrix norm의 몫이다.[36]

명제 4.7 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$가 (right) Markov matrix이면, $\dim E_1^A = \mathrm{mult}_A(1)$ (즉, A의 1-Jordan block은 항등행렬).

증명 : (i) 임의의 자연수 m에 대해, A^m도 Markov이므로 (연습문제 4.4 (가) 항), $\|A^m\| \le 1$.

(ii) $U^{-1}AU = J$를 A의 Jordan canonical form이라고 놓으면,

$$\|J^m\| = \|U^{-1}A^mU\| \le n^2\|U^{-1}\|\cdot\|A^m\|\cdot\|\|U\| \le n^2\|U^{-1}\|\cdot\|U\|$$

가 된다. 즉, $\{\|J^m\| \mid m \ge 0\}$은 bounded above.

(iii) 따라서 J의 모든 1-Jordan 小-block은 (1×1)-행렬이어야 한다 (연습문제 3.2 참조). 즉, J의 1-Jordan block은 항등행렬 (연습문제 3.6 참조).

(iv) A가 right Markov인 경우는 [I, 연습문제 7.6.24] 참조. □

이제 이 article의 main theorem을 state할 수 있다.

정리 4.8 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$가 **positive** (right) Markov matrix이면,

(가) $\dim E_1^A = \mathrm{mult}_A(1) = 1$.[37] (따라서 A의 unique dominant eigen-value 1에 대응하는 dominant eigen-vector도 (up to scalar) unique.)

(나) $\lim_{m\to\infty} A^m$은 수렴.

증명 : (가) 관찰 4.1, 관찰 4.2 및 명제 4.7.

(나) 명제 3.5. □

---

[35] X가 probability vector일 때, 지금은 AX도 probability vector일 이유는 없다.

[36] 사실은, 명제 4.7을 관찰 4.1보다 먼저 다룰 수도 있으므로, 'first touch'라고 해도 OK.

[37] 만약 $\mathrm{mult}_A(1) = 1$이라면, $\dim E_1^A = 1$일 수밖에 없다. 왜 그런가?

따름정리 4.9 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$가 **positive** Markov matrix일 때, $\lim_{m\to\infty} A^m = L$이라고 표기하자. 또 A의 unique dominant eigen-value 1에 대응하는 dominant eigen-vector 중 unique probability vector를 P라고 놓으면,

(가) $AL = LA = L$.

(나) L도 Markov matrix이고, $L = (P, P, \dots, P)$.

(다) 임의의 probability vector X에 대해, $\lim_{m\to\infty} A^m X = LX = P$.[38]

(라) $P > 0$. (따라서 $L > 0$.)

증명 : (가) $AL = A \lim_{m\to\infty} A^m = \lim_{m\to\infty} A^{m+1} = L$. ($LA = L$도 마찬가지.)

(나) 우선 A^m도 Markov이므로(연습문제 4.5),

$$\mathbf{1^t} \cdot L = \mathbf{1^t} \lim_{m\to\infty} A^m = \lim_{m\to\infty} \mathbf{1^t} \cdot A^m = \lim_{m\to\infty} \mathbf{1^t} = \mathbf{1^t}$$

이므로(물론 $L \geq 0$), L은 Markov(연습문제 4.4). 따라서, L의 모든 column은 probability vector이다. 그런데, $AL = L$이므로, L의 모든 column은 eigen-value 1을 갖는 A의 eigen-vector이기도 하다. 즉, L의 모든 column은 P.

(다) $P = (p_1, \dots, p_n)^{\mathbf{t}}$, $X = (x_1, \dots, x_n)^{\mathbf{t}}$로 놓자. 그러면, $L = (P, P, \dots, P)$이므로, LX의 i-번째 좌표는

$$p_i x_1 + \cdots + p_i x_n = p_i(x_1 + \cdots + x_n) = p_i$$

가 된다. 즉, $LX = P$.

(라) 자명. $AP = P$이고, $A > 0$, $P \geq 0$이므로, $P > 0$. $\square$

따라서, 이제 positive Markov matrix A가 주어지면, '적당한' probability vector X를 골라 AX, $A(AX)$, $A(A^2X)$, ... 을 계산하면 $\lim_{m\to\infty} A^m X$의 근삿값을 — 즉, A의 dominant eigen-vector P의 근삿값을 — 구할 수 있을 것이다. (이때 $\lim_{m\to\infty} A^m = L$을 구하는 것은 시간이 훨씬 더 많이 필요하다. 왜 그런가?) 이러한 **power method**(정의 1.8)의 수렴 속도는, 당연히, X의 선택에 depend할 것이다. 즉, P를 대강이나마 추측할 수 있다면 큰 도움이 될 것이다.

따름정리 4.9의 dominant eigen-vector P는 분야에 따라 **Perron-Frobenius vector**, **stochastic vector**, **stationary vector**, **fixed probability vector** 등 다양한 이름으로 불리운다.[39]

[38] 따라서, 특별히, $LP = P$.

[39] 제 6 절에서는 PageRank vector.

다음 연습문제도 분명히 해 둘 필요가 있다.

연습문제 4.10 $A_i \in \mathfrak{M}_{n,n}(\mathbb{C})$가 positive Markov matrix 일 때, block diagonal matrix $A = \mathrm{diag}(A_1, \ldots, A_k)$를 생각하자. 이때

(가) $\lim_{m\to\infty} A^m$을 묘사하라.
(나) Eigen-space E_1^A의 basis를 묘사하라.

Right Markov matrix의 경우도 분명히 정리해 둔다. (증명은 독자들의 몫.)

따름정리 4.11 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$가 **positive** right Markov일 때, $\lim_{m\to\infty} A^m = L$이라고 표기하자. 또 $A^{\mathbf{t}}$의 unique dominant eigen-value 1에 대응하는 dominant eigen-vector 중 unique probability vector를 P라고 놓으면,[40]

(가) $AL = LA = L$.
(나) L도 right Markov matrix이고, $L = (P, P, \ldots, P)^{\mathbf{t}}$.
(다) 임의의 probability vector X에 대해, $X^{\mathbf{t}} \cdot L = P^{\mathbf{t}}$이고,[41] $LX = (P^{\mathbf{t}} \cdot X)\,\mathbf{1}$.

다음 comment(들)도 생략하면 섭섭할 것이다.

참고 4.12 (가) 언급할 필요조차 없겠지만, 만약 A가 diagonalizable인 경우만 다루었다면, 이 article의 길이가 반으로 줄었을 것이다. 뿐만 아니라 'Jordan canonical form의 응용'이라고 표현할 수도 없었을 것이다.
(나) A가 Markov matrix일 때, 만약 $A^s > 0$인 자연수 s가 존재하면, A를 **regular** Markov matrix(또는 **primitive** Markov matrix)라고 부른다. Regular Markov matrix의 경우에도 관찰 4.2는 — 따라서 정리 4.8과 따름정리 4.9 및 따름정리 4.11도 — 여전히 성립한다.[42] [증명]: 만약 λ가 A의 eigen-value이고 $|\lambda| = 1$이면, λ^s는 $A^s > 0$의 eigen-value이므로, $\lambda^s = 1$(관찰 4.2). 한편, 당연히 $A^{s+1} > 0$이므로, 같은 이유로 $\lambda^{s+1} = 1$. 따라서 $\lambda = 1$. 또 $\dim E_1^{A^s} = 1$이므로 (정리 4.8), $E_1^A = E_1^{A^s}$임은 자명.

그러고 보니, 이 article의 제목 "Markov chain"(또는 "Markov process")의 'chain'의 뜻을 아직 설명하지 않았다. 이를 설명하려면 그림을 그려야 하는데……, 아래 여백도 부족하고……. 독자들에게 맡긴다.[43]

[40] A의 dominant eigen-vector는 $\mathbf{1}$이다(관찰 2.3의 증명 참조).
[41] 'Right Markov'에서 'right'의 의미는 'right notation(왼손잡이의 표기법)'이다. 즉, L이 오른쪽에서 act한다는 뜻.
[42] 하지만, n이 매우 크면, A가 regular인지 여부를 판정하기 쉽지 않다.
[43] Wikipedia 참조. 이 기회에 stochastic process의 뜻도 조사해 보라.

section version

180102

제 5 절 Hardy-Weinberg Equilibrium

이 節에서는 "Population Genetics"의 한 부분(시작 부분)을 소개한다.[44] 이 분야는 1908–1909 년 **Hardy-Weinberg equilibrium**(principle, law)으로부터 시작되었다.[45]

§ 5.A. 일반 유전자

예를 들어, 완두콩은 둥근 것과 주름진 것 두 종류가 있다. 이를 결정하는 한 쌍의 **allele**(대립유전자)을 (T, t)로 표기하자. 이때 **genotype** TT와 Tt의 phenotype은 둥근 콩이고, genotype tt의 phenotype은 주름진 콩이다.[46]

이제, $m \geq 0$ 일 때, m-번째 세대(m-세대)의 **genotype frequency**를

$$p_m = \mathrm{Prob}_{m\text{-세대}}(TT), \qquad q_m = \mathrm{Prob}_{m\text{-세대}}(Tt), \qquad r_m = \mathrm{Prob}_{m\text{-세대}}(tt)$$

로 놓고, m-세대의 genotype frequency vector를 $P_m = (p_m, q_m, r_m)^{\mathbf{t}}$로 표기하자.[47] 그리고 현재 세대(0-세대)의 genotype frequency vector는 특별히

$$P_0 = (p_0, q_0, r_0)^{\mathbf{t}} = (p, q, r)^{\mathbf{t}}$$

로 표기하자. 또 m-세대의 **allele frequency**는

$$a_m = \mathrm{Prob}_{m\text{-세대}}(T), \qquad b_m = \mathrm{Prob}_{m\text{-세대}}(t)$$

로 놓고, m-세대의 allele frequency vector를 $Q_m = (a_m, b_m)^{\mathbf{t}}$로 표기하자. 이번에도 0-세대의 allele frequency vector는 특별히

$$Q_0 = (a_0, b_0)^{\mathbf{t}} = (a, b)^{\mathbf{t}}$$

로 표기하자. 이때, $P_m, Q_m \geq 0$ 이고,

$$p_m + q_m + r_m = 1, \qquad a_m + b_m = 1$$

인 것은 물론이다.

[44] 이때 'population'의 의미는 ……, 타자 칠 생각 없다. 독자들이 조사해 보기 바란다. 또 고등학교 생물 시간에 배운 내용도 다시 설명할 생각 없다.

[45] G. H. Hardy(1877–1947)는 Hardy's Theorem, Hardy's Inequality, Hardy-Littlewood Theorem 등으로 우리에게 친숙한 수학자(해석학자)이다. 20세기 초 neo-Darwinism(modern synthesis)의 선봉에는 Hardy가 있었던 것이다. (W. Weinberg(1862–1937)는 (산)부인과 의사.)

[46] Mendel의 유전법칙. Dominant trait과 recessive trait…….

[47] Frequency vector = probability vector.

그러면, m-세대의 allele frequency는

$$a_m = p_m + \tfrac{1}{2} q_m, \qquad b_m = \tfrac{1}{2} q_m + r_m$$

으로 나타낼 수 있다(왜 그런가?).[48] 특별히 0-세대의 경우에는

$$a = p + \tfrac{1}{2} q, \qquad b = \tfrac{1}{2} q + r$$

이 된다.

그러면, 1-세대의 genotype frequency를 조사해 보자. 먼저 아빠가 TT-type 인 경우에 자식의 genotype 각각의 확률은 다음 table(Markov matrix)

$A_{TT} =$ [아빠 TT-type]

엄마 \ 자식	TT	Tt	tt
TT	1	$\frac{1}{2}$	0
Tt	0	$\frac{1}{2}$	1
tt	0	0	0

로 주어진다. 마찬가지로 아빠가 Tt-type 인 경우와 tt-type 인 경우에도 자식의 genotype 확률은 다음 Markov matrix들

$A_{Tt} =$ [아빠 Tt-type]

엄마 \ 자식	TT	Tt	tt
TT	$\frac{1}{2}$	$\frac{1}{4}$	0
Tt	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
tt	0	$\frac{1}{4}$	$\frac{1}{2}$

$A_{tt} =$ [아빠 tt-type]

엄마 \ 자식	TT	Tt	tt
TT	0	0	0
Tt	1	$\frac{1}{2}$	0
tt	0	$\frac{1}{2}$	1

로 주어질 것이다. 이제,

$$A = p\,A_{TT} + q\,A_{Tt} + r\,A_{tt}$$

로 표기하면,

$$A = \begin{pmatrix} p + \frac{1}{2} q & \frac{1}{2} p + \frac{1}{4} q & 0 \\ \frac{1}{2} q + r & \frac{1}{2} p + \frac{1}{2} q + \frac{1}{2} r & p + \frac{1}{2} q \\ 0 & \frac{1}{4} q + \frac{1}{2} r & \frac{1}{2} q + r \end{pmatrix} = \begin{pmatrix} a & \frac{1}{2} a & 0 \\ b & \frac{1}{2} & a \\ 0 & \frac{1}{2} b & b \end{pmatrix}$$

가 된다. A는 물론 Markov matrix.

[48] 그렇지만, 거꾸로, allele frequency로부터 genotype frequency를 알아낼 수는 없다.

Markov matrix A를 제 2 절 toy example의 언어를 흉내 내어 표현하면,

$$A = \begin{pmatrix} a & \frac{1}{2}a & 0 \\ b & \frac{1}{2} & a \\ 0 & \frac{1}{2}b & b \end{pmatrix} = \begin{pmatrix} TT \rightsquigarrow TT & Tt \rightsquigarrow TT & tt \rightsquigarrow TT \\ TT \rightsquigarrow Tt & Tt \rightsquigarrow Tt & tt \rightsquigarrow Tt \\ TT \rightsquigarrow tt & Tt \rightsquigarrow tt & tt \rightsquigarrow tt \end{pmatrix}$$

로 쓸 수 있을 것이다. 이때, 예를 들어, $TT \rightsquigarrow Tt$의 표기법은 (엄마의) TT가 (자식의) Tt로 '변화'할 확률을 뜻한다. 따라서, 제 2 절의 toy example에서처럼,[49] 1-세대의 genotype frequency vector는

$$P_1 = \begin{pmatrix} p_1 \\ q_1 \\ r_1 \end{pmatrix} = AP_0 = \begin{pmatrix} a & \frac{1}{2}a & 0 \\ b & \frac{1}{2} & a \\ 0 & \frac{1}{2}b & b \end{pmatrix} \begin{pmatrix} p \\ q \\ r \end{pmatrix} = \begin{pmatrix} ap + \frac{1}{2}aq \\ bp + \frac{1}{2}q + ar \\ \frac{1}{2}bq + br \end{pmatrix} = \begin{pmatrix} a^2 \\ 2ab \\ b^2 \end{pmatrix}$$

이 된다(확인해 보라).[50] 그러므로, 1-세대의 allele frequency vector는

$$Q_1 = \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = \begin{pmatrix} a^2 + \frac{1}{2}2ab \\ \frac{1}{2}2ab + b^2 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} = Q_0$$

이다. 즉, **allele frequency는 불변**!

마찬가지로, 2-세대의 genotype frequency vector는

$$P_2 = \begin{pmatrix} p_2 \\ q_2 \\ r_2 \end{pmatrix} = \begin{pmatrix} a_1 & \frac{1}{2}a_1 & 0 \\ b_1 & \frac{1}{2} & a_1 \\ 0 & \frac{1}{2}b_1 & b_1 \end{pmatrix} \begin{pmatrix} p_1 \\ q_1 \\ r_1 \end{pmatrix} = \begin{pmatrix} a & \frac{1}{2}a & 0 \\ b & \frac{1}{2} & a \\ 0 & \frac{1}{2}b & b \end{pmatrix} \begin{pmatrix} a^2 \\ 2ab \\ b^2 \end{pmatrix}$$
$$= \begin{pmatrix} a^3 + a^2 b \\ a^2 b + ab + ab^2 \\ ab^2 + b^3 \end{pmatrix} = \begin{pmatrix} a^2 \\ 2ab \\ b^2 \end{pmatrix} = P_1$$

이고, 이번에도 allele frequency vector $Q_2 = (a,b)^{\mathbf{t}} = Q_0$는 불변. 따라서,

$$P_m = \begin{pmatrix} p_m \\ q_m \\ r_m \end{pmatrix} = A^m P_0 = \begin{pmatrix} a^2 \\ 2ab \\ b^2 \end{pmatrix} = P_1, \qquad Q_m = \begin{pmatrix} a \\ b \end{pmatrix} = Q_0, \qquad (m \geq 1)$$

임을 알 수 있다.

[49]이사 가다 = 주소가 '변화'하다.

[50]굳이 언급할 필요도 없겠지만, 이번에도, 이 mathematical model에는 random mating, sex independent genotype frequency 등의 가정들이 필요할 것이다. 또 모든 개체들은 건강하고(번식력이 있고)$\cdots\cdots$. 아, 그리고 다른 세대 간의 교배는 금지(즉, m-세대는 m-세대와만 교배).

지금까지의 논의를 종합하면,[51] allele frequency vector $(a,b)^{\mathbf{t}}$는 **constant**이고, genotype frequency vector는 — **단 한 세대만에** — equilibrium(**Hardy-Weinberg equilibrium**) $(a^2,\, 2ab,\, b^2)^{\mathbf{t}}$에 도달한다. (20세기 초 neo-Darwinism (modern synthesis)은 이 명제로부터 시작되었다.)

단 한 세대만에 equilibrium에 도달하므로, 극한은 생각할 필요도 없겠지만, 어쨌든 A는 regular Markov matrix이다.[52] 그리고, 물론 $P_1 = (a^2,\, 2ab,\, b^2)^{\mathbf{t}}$는 A의 dominant eigen-vector이다.[53] 즉, $\lim_{m\to\infty} A^m P_0 = P_1$.

§ 5.B. Sex Linked Gene

예를 들어, 인간의 (적록)색맹 유전자는 性염색체인 X-염색체에 들어 있다. 색맹에 관여하는 한 쌍의 allele을 (X,x)로 표기하자. 이때 여성의 세 genotype XX, Xx, xx 중 xx만이 색맹이고, 남성의 두 genotype XY, xY 중 xY는 색맹이다.[54]

이제, m-세대 **여성**의 genotype frequency를 각각

$$p_m = \mathrm{Prob}_{m\text{-세대}}(XX), \qquad q_m = \mathrm{Prob}_{m\text{-세대}}(Xx), \qquad r_m = \mathrm{Prob}_{m\text{-세대}}(xx)$$

로 놓고, m-세대 여성의 genotype frequency vector를 $P_m = (p_m, q_m, r_m)^{\mathbf{t}}$로 표기하자. (물론 $P_m \geq 0$이고, $p_m + q_m + r_m = 1$.) 또 m-세대 **여성**의 allele frequency를 각각

$$a_m = \mathrm{Prob}_{m\text{-세대 여성}}(X), \qquad b_m = \mathrm{Prob}_{m\text{-세대 여성}}(x)$$

로 놓으면(물론 $a_m, b_m \geq 0$이고, $a_m + b_m = 1$), 앞 § 5.A 에서처럼,

$$a_m = p_m + \tfrac{1}{2} q_m, \qquad b_m = \tfrac{1}{2} q_m + r_m, \qquad (m \geq 0)$$

인 것은 익숙하다. 그리고 m-세대 **남성**의 genotype frequency를 각각

$$c_m = \mathrm{Prob}_{m\text{-세대}}(XY) = \mathrm{Prob}_{m\text{-세대 남성}}(X),$$
$$d_m = \mathrm{Prob}_{m\text{-세대}}(xY) = \mathrm{Prob}_{m\text{-세대 남성}}(x)$$

로 놓고(당연히 남성의 allele frequency와 genotype frequency는 같다), m-세대 남성의 allele frequency vector를 $R_m = (c_m, d_m)^{\mathbf{t}}$로 표기하자.

[51] '수능의 정답'은 언제나(!) [둥근 콩 : 주름진 콩 = 3 : 1]이다. 이 비율은 어떻게 얻어진 것일까?
[52] $a, b \neq 0$이면, $A^2 > 0$. 참고 4.12(나)항 참조.
[53] A의 eigen-value는 $1, \frac{1}{2}, 0$이다. 따라서, 특별히, A는 diagonalizable이고, $\det(A) = 0$.
[54] 편의상 색맹과 색약을 모두 색맹으로 표현했다.

그러면, $(m+1)$-세대 **여성**의 genotype frequency를 알아보자. 계속 §5.A 의 표기법을 사용하면, 딸의 genotype 각각의 확률은 다음 Markov matrix들

A_{XY} = [아빠 XY-type]

딸 \ 엄마	XX	Xx	xx
XX	1	$\frac{1}{2}$	0
Xx	0	$\frac{1}{2}$	1
xx	0	0	0

A_{xY} = [아빠 xY-type]

딸 \ 엄마	XX	Xx	xx
XX	0	0	0
Xx	1	$\frac{1}{2}$	0
xx	0	$\frac{1}{2}$	1

로 주어질 것이다. 이제,

$$A_m = c_m A_{XY} + d_m A_{xY} = \begin{pmatrix} c_m & \frac{1}{2}c_m & 0 \\ d_m & \frac{1}{2} & c_m \\ 0 & \frac{1}{2}d_m & d_m \end{pmatrix}, \qquad (m \geq 0)$$

으로 표기하면(A_m은 Markov matrix), $(m+1)$-세대 여성의 genotype frequency vector는

$$P_{m+1} = \begin{pmatrix} p_{m+1} \\ q_{m+1} \\ r_{m+1} \end{pmatrix} = A_m P_m = \begin{pmatrix} c_m & \frac{1}{2}c_m & 0 \\ d_m & \frac{1}{2} & c_m \\ 0 & \frac{1}{2}d_m & d_m \end{pmatrix} \begin{pmatrix} p_m \\ q_m \\ r_m \end{pmatrix}$$

$$= \begin{pmatrix} a_m c_m \\ d_m p_m + \frac{1}{2}q_m + c_m r_m \\ b_m d_m \end{pmatrix} = \begin{pmatrix} a_m c_m \\ 1 - a_m c_m - b_m d_m \\ b_m d_m \end{pmatrix}$$

이 된다.[55]

연습문제 5.1 (가) $(m+1)$-세대 **남성**의 genotype(allele) frequency를 구하기 위해 — 위 논의를 흉내 내어 — table(Markov matrix)들 B_{XX}, B_{Xx}, B_{xx}를 묘사하라.

(나) $B_m = p_m B_{XX} + q_m B_{Xx} + r_m B_{xx}$로 놓으면,

$$R_{m+1} = \begin{pmatrix} c_{m+1} \\ d_{m+1} \end{pmatrix} = B_m R_m = \begin{pmatrix} a_m & a_m \\ b_m & b_m \end{pmatrix} \begin{pmatrix} c_m \\ d_m \end{pmatrix} = \begin{pmatrix} a_m \\ b_m \end{pmatrix}$$

임을 보여라.

[55] 이 식의 모든 vector는 probability vector이므로(연습문제 4.5), 끝 항의 두 번째 좌표는 자명.

위 연습문제 (나)항의 결과는, 사실은, 아들의 genotype 은 아빠와는 무관하므로(왜 그런가?), 직관적으로도 자명하다. 어쨌든, $d_{m+1} = b_m$. 한편, 여성의 allele frequency b_{m+1} 은

$$\begin{aligned} b_{m+1} &= \tfrac{1}{2}(1 - a_m c_m - b_m d_m) + b_m d_m \\ &= \tfrac{1}{2}\big(1 - (1-b_m)(1-d_m) - b_m d_m\big) + b_m d_m \\ &= \tfrac{1}{2}(b_m + d_m) \end{aligned}$$

이 된다. 이제 $M = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}$, $X_m = \begin{pmatrix} b_m \\ d_m \end{pmatrix}$ 으로 표기하면,

$$X_{m+1} = \begin{pmatrix} b_{m+1} \\ d_{m+1} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b_m \\ d_m \end{pmatrix} = M X_m$$

임을 알 수 있다. 즉,

$$X_1 = M X_0, \quad X_2 = M X_1 = M^2 X_0, \quad \ldots, \quad X_m = M^m X_0$$

이다.

그런데, 당연히 $M^2 > 0$ 이므로, M 은 regular **right** Markov matrix 이다. 따라서, 따름정리 4.11을 적용할 수 있을 것이다(참고 4.12(나)항 참조). 이제 $M^{\mathbf{t}}$ 의 dominant eigen-vector 는 $\big(\frac{2}{3}, \frac{1}{3}\big)^{\mathbf{t}}$ 이므로(확인해 보라),

$$\lim_{m\to\infty} \begin{pmatrix} b_m \\ d_m \end{pmatrix} = \lim_{m\to\infty} M^m \begin{pmatrix} b_0 \\ d_0 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} b_0 \\ d_0 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} b_0 + \frac{1}{3} d_0 \\ \frac{2}{3} b_0 + \frac{1}{3} d_0 \end{pmatrix}$$

가 된다. 이는, 오랜 세월이 지나면, 신기하게도 여성의 allele frequency vector 와 남성의 allele frequency vector 가 **같아진다**는 의미이다. 또, 특별히, 여성과 남성의 allele frequency vector 는 모두 **수렴**하므로,

$$b_\infty = \lim_{m\to\infty} b_m = \tfrac{2}{3} b_0 + \tfrac{1}{3} d_0, \qquad a_\infty = \lim_{m\to\infty} a_m = 1 - b_\infty$$

로 표기하면, 다음의 Hardy-Weinberg equilibrium

$$\lim_{m\to\infty} \begin{pmatrix} p_m \\ q_m \\ r_m \end{pmatrix} = \lim_{m\to\infty} \begin{pmatrix} a_{m-1} c_{m-1} \\ * \\ b_{m-1} d_{m-1} \end{pmatrix} = \lim_{m\to\infty} \begin{pmatrix} a_{m-1} a_{m-2} \\ * \\ b_{m-1} b_{m-2} \end{pmatrix} = \begin{pmatrix} (a_\infty)^2 \\ 2 a_\infty b_\infty \\ (b_\infty)^2 \end{pmatrix}$$

을 얻는다(연습문제 5.1과 그 앞 단락 참조).[56]

[56] *는 타자 치기 귀찮다는 뜻.

실제로, 통계 자료에 의하면, 남성 중 색맹의 비율은 대략 $\frac{1}{12} = 0.0833$ 쯤 되고, 여성 중 색맹의 비율은 대략 $\frac{1}{200} = 0.0050$ 쯤 된다고 한다. 한편, 우리의 결론은 남성 중 색맹의 비율은 $\lim_{m\to\infty} d_m = b_\infty$ 이고, 여성 중 색맹의 비율은 $\lim_{m\to\infty} r_m = (b_\infty)^2$ 이라는 것이다. 물론 $\left(\frac{1}{12}\right)^2 = 0.0069$ 이므로, 우리의 결과는 현실과 (거의) 일치한다.[57]

§ 5.C. Generalization

이 節의 논의는 다양한 방향으로 일반화가 가능할 것이다. 예를 들면:

(가) r-개의 유전자 $T, S, \ldots$ 가 관여하는 유전형질의 경우(이때 genotype은 모두 3^r-개),

(나) Allele이 pair (T, t)가 아니고, s-tuple $(T_1, \ldots, T_s)$인 경우(이때 genotype은 모두 $\binom{s+1}{2}$-개),

(다) 치명적인 유전병의 경우,

(라) 위 (가)–(다)항에서 남성과 여성의 genotype frequency가 다를 경우,

(마) 위 (가)–(다)항에 sex linked gene이 포함된 경우,

(바) mutation, sexual selection, ······

등 무한히 많은 일반화와 각각의 경우에 다양한 mathematical model들을 생각할 수 있다.

(미완성. 언젠가 훗날 더 보충······ (?).)

참고 5.2 이 節의 시작 부분에서 "20세기 초 neo-Darwinism(modern synthesis)의 선봉에는 G. H. Hardy가 있었던 것이다"라고 한 것을 Hardy가 들으면, 아마, 크게 화를 낼 것이다. 왜냐하면, Hardy는 "I have never done anything 'useful'. No discovery of mine has made, or is likely to make the least difference to the amenity of the world"라고 말하며, 자신이 순수수학자임을 긍지로 삼았기 때문이다. Hardy는 cricket team의 동료인 geneticist로부터 질문을 받고, (아마 다음 날) 한 쪽짜리 논문을 써 주었을 뿐이다. Hardy thus became the somewhat **unwitting** founder of a branch of applied mathematics.[58]

[57] 약간의 오차(0.019)는 아마 통계 조사의 오차일 가능성이 크다. ㅋㅋ. ($\lim_{m\to\infty} r_m = (b_\infty)^2$ 이라는 명제는 initial condition(또는 founder effect)과는 무관.)

[58] Wikipedia에서 발췌.

제 6 절 Google Ranking System

section version 180102

이 節에서는 Markov chain을 응용한 PageRank algorithm(Google ranking system)을 소개한다. Internet search engine은 본질적으로 모두 오십보백보(가능한 한 많은 web page를 수집해 저장한다). 문제의 초점은 ranking system이다: search result를 어떤 순서(rank)로 나열할 것인가?

1995년(?) Stanford University에서 Computer Science를 공부하던 박사과정생 L. Page와 S. Brin은 G. H. Golub[4] 교수의 "Matrix Computation" 강의를 듣던 중 $\cdots\cdots$ (Golub 교수는 틀림없이 한 학기 내내 dominant eigen-vector와 power method에 대해 떠들었을 것이다). Page와 Brin이 강의 내용을 잘 이해했던 것 같지는 않지만,[59] "행렬을 이용해 무언가 계산하려면 dominant eigen-vector와 Marcov chain을 생각해야만 한다"는 것만은 확실히 배운 것 같다.

PageRank의 idea는 우리 모두 알고 있는 격언에 불과했다: '좋은 벗 하나면 열 친구 안 부럽다.' 즉, 친구의 숫자보다 친구 하나하나의 질이 더 중요하다는 뜻이다.

보기 6.1 어떤 사람 i의 人格 r_i를(단, $0 \le r_i \in \mathbb{R}$) 어떻게 정의할 수 있을까? 만약 'follower'의 숫자만을 중시한다면, i를 존경하는 사람의 숫자만 세면 될 것이다. 그렇지만, 'follower'의 질까지 생각한다면, r_i를 $\sum_{j \rightsquigarrow i} r_j$로 정의하는 것을 고려할 수 있을 것이다. 이때 $j \rightsquigarrow i$의 표기법은 j가 i를 존경한다는 뜻이다. 그리고 조금만 더 생각해 보면,

$$r_i = \sum_{j \rightsquigarrow i} \frac{1}{N_j} r_j, \qquad \text{단, } N_j = \big|\{k \,|\, j \rightsquigarrow k\}\big|$$

로 정의하는 것이 합리적임을 발견하게 된다.[60]

아니, 이걸 모르는 사람도 있었던가? 이 정의가 특허의 대상이 될까? 수학자들(Golub 포함(?))은 무얼 하고 있었단 말인가? ㅠㅠ.

[59] 그들이 후에(1998년) 제출한 보고서[6]를 보면 이들은 eigen-value의 정의도 잘 모른다는 것이 드러난다. 이 보고서[6]의 보고자 중 T. Winograd는 Page의 지도교수였음에도 불구하고$\cdots\cdots$. Google의 수학자 직원의 도움 없이 보고서를 작성하고, 틀린 내용을 수정하지 않은 채 공개하고 있는 이들의 'honor system'이 오히려 돋보인다. 저자는 이 보고서[6]을 몇 번이나 읽으려고 시도했으나, 앞뒤가 들어맞지 않는 내용이 너무 많아, 번번이 실패하였다.

[60] 극단적인 예를 들면, 인격이 0인 'follower' 백만 명이 있어도 인격은 0일 뿐이다.

보기 6.1 에서처럼, rank(인격, reputation, 권위, 중요도 또는 인기도 등) r_i 는 그 값이 클수록 '높은' 것으로 해석한다. 즉, rank 는 등수가 아니다.

다음 정의가 바로 800 조 원짜리![61]

정의 6.2 어떤 web page i 의 **rank** r_i 를

$$r_i = \sum_{j \rightsquigarrow i} \frac{1}{N_j} r_j, \qquad \text{단, } N_j = \big|\{k \mid j \rightsquigarrow k\}\big|$$

로 정의한다(단, $0 \le r_i \in \mathbb{R}$).[62] 이때 $j \rightsquigarrow i$ 의 표기법은 web page j 에서 web page i 로 가는 link 가 존재한다는 의미이다.

자, 이제 r_i 를 어떻게 계산할까? r_i 는 유일하게 결정될까? 독자들은 위 정의가 r_i 들을 미지수로 갖는 1-차 연립방정식임을 알아보았는가? 즉,

$$a_{ij} = \begin{cases} \dfrac{1}{N_j} & (\text{if } j \rightsquigarrow i \text{ and } i \neq j) \\ 0 & (\text{otherwise}) \end{cases}$$

로 놓으면, 위 정의는 1-차 연립방정식

$$r_i = \textstyle\sum_j a_{ij} r_j$$

가 된다. (이때 $a_{ii} = 0$. 즉, self-link 는 무시.) 이제, 물론, $A = (a_{ij}) \in \mathfrak{M}_{n,n}(\mathbb{R})$ 로 놓고, **rank vector** 를 $R = (r_j) \in \mathbb{R}^n$ 으로 표기하면, 위 식은

$$AR = R$$

이 된다. 즉, R 은 eigen-value 1 에 대응하는 A 의 eigen-vector 이다. 그리고, 아마 A 의 matrix size n 은 대략 2^{30} 쯤 될 것이다.

연습문제 6.3 만약 A 가 zero column 을 갖지 않는다면, A 는 Markov matrix 임을 보여라. (따라서, 이때 A 는 dominant eigen-value 1 을 갖는다.)

그러나, 만약 A 에 zero column 이 있다면 — 즉, outlink 가 하나도 없는 web page 가 있다면 — A 가 항상 eigen-value 1 을 갖는 것은 아니다. (정확히 어떤 경우에 R 이 eigen-value 1 을 갖는지는 $\cdots\cdots$, 관심 없다.)

[61] 2017 년 Google의 시가 총액 약 800조 원.

[62] 모든 web page에 번호를 부여.

만약 R 이 존재한다면, R 은 (up to scalar multiple) 유일할까? 즉, $\dim E_1^A$ 는 1 일까? 또 $\mathrm{mult}_A(1)$ 은? 물론 $A \geq 0$ 이고 A 는 very very sparse matrix 이다.[63] 그리고, 0 이 많은 행렬은 대개 '문제'가 많다는 것이 우리의 경험이다.

게다가, 현실적으로 A 는 (after renumbering) block diagonal matrix 일 가능성이 크다. (Page-Brin[6] 도 "there is a small problem ······, two web pages that point to each other but to no other page" 라고 언급하고 있다. 즉, A 에 diagonal block $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ 이 나타날 수 있다는 뜻이다. 따라서, 이때는 (-1) 도 A 의 eigen-value 이므로, 1은 unique dominant eigen-value 가 아니다.) 그리고, 이때 $\mathrm{mult}_A(1)$ 은 적어도 diagonal block 의 개수보다 크거나 같을 것이다(연습문제 4.10 도 참조).[64]

"행렬을 이용해 무언가 하려면 dominant eigen-vector 와 Marcov chain 을 생각해야만 한다"는 믿음이 있다면, 다음 '땜질'은 필연적이다. 그리고 '땜질'이란, 주의 1.5 에서 언급했듯이, A 의 좌표 몇 개에 2^{-30} 쯤을 더해주는 것을 의미한다.[65]

1-차 땜질: 만약 위 행렬 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 에 zero column 이 있으면, 모든 zero column 을 $\frac{1}{n}\mathbf{1}$ 로 대체한다(단, $\mathbf{1} = (1, 1, \ldots, 1)^{\mathbf{t}}$). 이 새로운 행렬을 A' 이라고 표기하면, A' 은 **Markov matrix** 이다(연습문제 6.3 참조).

굳이 이 '땜질'에 의미를 부여하자면, outlink 가 하나도 없는 page 는 모든(자기 자신 포함) page 로 가는 outlink 가 있는 것으로 간주하자는 것이다. 즉, 아무도 존경하지 않는 것은 모두(자기 자신 포함)를 존경하는 것과 동치(?). 게다가 $\frac{1}{n}$ 은 워낙 작은 수이므로, A 와 A' 은 오십보백보. (여기까지는 제법 그럴싸하다.)

그러나, Markov matrix A' 은 아직 very very sparse. 수학자들은 — 설령 정의 6.2 를 '발견'했더라도 — 여기까지 생각해 본 후, 포기했을 것이다······.

---

[63] Sparse matrix 란 non-zero entry 가 매우 드물다는 뜻.

[64] 각각의 diagonal block 은 외부와 단절된 독립된 세계이므로, 각 독립된 세계마다 인격이 최고인 사람이 있을 것이고, 이들의 인격은 서로 비교할 수 없다.

[65] 만약 저자가 PageRank algorithm 을 공부하지 않았더라면, 주의 1.5 에서, 좌표 몇 개에 2^{-30} 쯤을 더해주면 된다고 자신 있게 말하지 못했을 것이다.

Page-Brin의 800조 원짜리 발상(신념)은 간단하고도 명료했다: "쓸 만한 theory는 positive Markov matrix의 경우에만 존재하므로, A'을 한 번 더 '땜질' 해서라도 positive Markov matrix를 만들면 된다"는 것이었다! 그렇지만, 이번에는 '땜질'이 A'의 좌표 몇 개에 2^{-30}쯤을 더해주는 것을 **의미하지 않는다**.

2-차 땜질: Google matrix $G \in \mathfrak{M}_{n,n}(\mathbb{R})$을

$$G = dA' + \frac{1-d}{n}\,\mathbf{1}_n$$

으로 정의한다(단, $\mathbf{1}_n \in \mathfrak{M}_{n,n}(\mathbb{C})$는 모든 좌표가 1인 행렬). 이때

$$d = 0.85$$

로 놓는다(d의 이름은 **damping factor**). 그러면 G는 **positive** Markov matrix가 된다(다음 단락 참조).

이제 Google matrix $G = (g_{ij})$를 구체적으로 살펴보자. 우선 $N_j = 0$일 때 (즉, j-page에는 outlink가 하나도 없고, 따라서 A의 j-th column은 0이고, A'의 j-th column은 $\frac{1}{n}\mathbf{1}$이면), G의 j-th column은 여전히 $\frac{1}{n}\mathbf{1}$이다. 즉, 만약 $N_j = 0$이면, $g_{ij} = \frac{1}{n}$ for all i. 그리고, $N_j \neq 0$일 때는

$$g_{ij} = \begin{cases} \dfrac{0.15}{n} + \dfrac{0.85}{N_j} & (\text{if } j \rightsquigarrow i \text{ and } i \neq j) \\ \dfrac{0.15}{n} & (\text{otherwise}) \end{cases}$$

임을 알 수 있다. (따라서, G는 positive Markov.)

그런데, 여기에서 $\frac{1}{n}$이나 $\frac{0.15}{n}$는 워낙 작은 수이므로 무시해 버려도 그만이겠지만, $\frac{1}{N_j}$과 $\frac{0.85}{N_j}$의 차이 $\frac{0.15}{N_j}$는 제법 의미 있는 수이다. 뿐만 아니라, 원래의 행렬 A에서 의미가 있는 non-zero component는 $\frac{1}{N_j}$ 뿐이었다. 그렇다면, damping factor를, 예를 들어, 0.5 또는 0.88이라고 하지 않고, 왜 0.85라고 못 박았을까? 오히려 damping factor를 0.9999 또는 $1 - \frac{1}{2^{30}}$로 놓는 것이 원래의 의도 — 즉, 정의 6.2 — 에 더 가깝지 않을까?[66]

사전을 찾아보면, 'damp'에는 여러 가지 뜻이 있지만, 지금은 '둔화시키다' 또는 '진동/발산을 멈추다'라는 뜻으로 사용된 듯하다. 즉, $\lim_{m\to\infty} G^m X$가 '발산을 멈추고' 수렴하게 만든다는 의미로 짐작된다(단, $X \in \mathbb{R}^n$은 probability vector).

[66] $0 < d < 1$이기만 하면, G는 positive Markov matrix.

물론 이제 G는 positive Markov matrix이므로, Perron-Frobenius vector $\lim_{m\to\infty} G^m X = P$에 관한 '멋진' theory(정리 4.8과 따름정리 4.9)가 준비되어 있다(단, $X \in \mathbb{R}^n$은 임의의 probability vector). 이때 dominant eigen-vector P를 **PageRank vector**라고 부르면 그럴듯할 것이다(P의 i-번째 좌표는 i-page의 PageRank).[67] P의 근삿값을 구하는 방법은 물론 power method. Page-Brin[6]에 따르면, 대략 $G^{52}X$까지 계산하면 충분하다고 한다.[68 69]

위 '땜질'들과 damping factor $d = 0.85$를 위한 Page-Brin[6]의 흥미로운 변명이 하나 있다.[70] 이름하여 **Random Surfer Model**.

Random Surfer Model에서는 G의 (i, j)-좌표 g_{ij}를 j-page에 있는 random surfer가 i-page로 옮겨 갈(이사 갈) 확률이라고 해석한다. 따라서, X를 web page들의 현재 '인구' 분포 확률벡터라고 놓으면,[71] ∞-번 이사 후의 '인구' 분포 확률벡터는 $\lim_{m\to\infty} G^m X$가 될 것이다. (제 2절의 toy example 참조.) 즉, 오랜 시간 후에는 backlink가 많은(중요한, 권위 있는) page에 더 많은 surfer들이 위치할 것이다. 그런데, 통계조사(?!)에 의하면, 어떤 page에 있는 surfer가 그 page의 outlink를 click해서 이사 갈 가능성이 85%이고, outlink를 무시하고 이사 갈 가능성은 15%라고 한다…….[72] (즉, $j \rightsquigarrow i$일 때, j-page에 있는 surfer가 i-page로 가는 outlink를 click하지 않고도 i-page로 이동할 확률도 $\frac{0.15}{n}$이고, 현재 사는 집(j-page)으로 다시 이사할 확률도 $\frac{0.15}{n}$.)

실제 PageRank algorithm을 현실에 implement하려면, 몇 가지 고려할 사항이 더 (많이) 있을 것이다.[73] 예를 들어, 가짜 존경(광고 등)을 어떻게 제거할 것인가, 또 방문자 수는 정말 전혀 고려하지 않을 것인가 등등…….

[67] 정의 6.2의 rank vector와는 다른 이름을 사용하였다.
[68] [2]에 따르면, "can be computed in a few hours on a medium size workstation."
[69] 예를 들어, X를 전날 계산해 놓은 PageRank vector로 놓으면, 더 빨리 수렴할 것이다.
[70] [6]에는 0.85라는 숫자는 없고, 이 숫자는 [2]에 등장한다.
[71] 이때 '인구'는 (random) surfer들의 숫자를 뜻한다.
[72] Random surfer는 85%라는 수치를 모르면 randomly 이동할 수 없다. 그러므로, 현혹적인 단어 'random'보다 'imaginary' 또는 'virtual'이라는 단어가 더 적절해 보인다.
[73] 어쩌면 0.85라는 숫자에도 우리가 짐작하기 쉽지 않은 현실적인 이유(근거)가 있을 수 있다.

section version

160102

제 7 절 Perron-Frobenius Theory

이 節에서는 우리가 흔히 Perron-Frobenius Theory 라고 부르는 결과를 소개한다. 이 節은 제 4 절보다 더 일반적인 경우를 다루므로, 논리적으로는 원래 제 4 절보다 앞서 소개하는 것이 바람직했을 것이다. 그러나, 실제 강의에서 이 節을 다루기에는 시간이 충분하지 않을 것으로 짐작되기에, 이 article 의 끝에 숨긴다. 이 節의 증명은 주로 Meyer[5] 를 참조하였다.

이 節에서는 $A>0$ 인 경우만을 다룬다. 한편 $A\geq 0$ 인 경우는 제법 골치 아프므로, 독자들에게 방학숙제로 남긴다.[74]

표기법 7.1 $C, D \in \mathfrak{M}_{m,n}(\mathbb{C})$ 일 때, 이 節에서는 다음 표기법을 사용한다.

(가) $C=(c_{ij})$ 이면, $|C| = \big(|c_{ij}|\big) \in \mathfrak{M}_{m,n}(\mathbb{C})$ 로 표기한다.[75] 따라서 $X \in \mathbb{C}^n$ 일 때, $|X|$ 의 표기법도 가능.

(나) $C>D$ if and only if $C-D>0$.

(다) $C\geq D$ if and only if $C-D\geq 0$.

우선 main theorem 을 소개하고, 증명은 잠시 후로 미룬다.

정리 7.2 (Perron-Frobenius Theorem: Positive Case) $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 이고 $A>0$ 일 때, λ 를 A 의 **dominant** eigen-value 라고 하자. 그러면:

(가) $\lambda = |\lambda| > 0$.

(나) $AX=\lambda X$ 이면 (단 $0\neq X\in\mathbb{C}^n$), $A\cdot|X| = \lambda\cdot|X|$ 이고, 따라서 $|X|>0$.

(다) $\dim E_\lambda^A = \mathrm{mult}_A(\lambda) = 1$.

따라서, 위 (가)항에 의하면, $A>0$ 은 unique dominant eigen-value $\lambda>0$ 을 갖는다. 또, 위 (나)항에 의하면, λ 에 대응하는 positive eigen-vector 가 존재한다. 뿐만 아니라, 위 (다)항에 의하면, λ 에 대응하는 positive eigen-vector 중 probability vector 를 P 라고 놓으면, P 는 유일하고 $E_\lambda^A = \langle P\rangle$ 가 된다. 우리는 P 를 A 의 **Perron-Frobenius vector** 라고 부른다.

표기법 7.3 $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 일 때, $\mathfrak{m}(A) = \max\big\{|\lambda| \,\big|\, \lambda \text{ 는 } A \text{ 의 eigen-value}\big\}$ 로 표기한다. (따라서, 만약 λ 가 A 의 dominant eigen-value 이면, $|\lambda| = \mathfrak{m}(A)$.)

[74] 엄밀하게는 $A>0$ 인 경우는 Perron Theory 라고 부르고, $A\geq 0$ 인 경우를 Perron-Frobenius Theory 라고 부르는 것 같다. Meyer[5] 참조

[75] 즉, 이 節에서는 $|C|$ 는 $\det(C)$ 와 무관.

지금부터는 **항상** $A \in \mathfrak{M}_{n,n}(\mathbb{C})$ 이고, $A > 0$ 이라고 가정한다.

연습문제 7.4 다음을 보여라.

(가) $\mathfrak{m}(A) > 0$.

(나) $\frac{1}{\mathfrak{m}(A)} A > 0$ 이고, $\mathfrak{m}\big(\frac{1}{\mathfrak{m}(A)} A\big) = 1$.

연습문제 7.5 $0 \neq c \in \mathbb{C}$ 일 때, 다음을 보여라.

(가) $E_\lambda^A = E_{c\lambda}^{cA}$.

(나) $\phi_A(t) = (t-\lambda_1)\cdots(t-\lambda_n)$ 이면, $\phi_{cA}(t) = (t-c\lambda_1)\cdots(t-c\lambda_n)$.[76] 따라서, $\mathrm{mult}_A(\lambda) = \mathrm{mult}_{cA}(c\lambda)$.

정리 7.2를 증명하려면, 당연히, 명제 3.10과 명제 4.7의 증명을 modify 해야 할 것이다. 따라서 그 증명은 'technical'······.

명제 7.6 (가) $\mathfrak{m}(A)$는 A의 eigen-value.

(나) $AX = \lambda X$ 이면(단, $|\lambda| = \mathfrak{m}(A)$, $0 \neq X \in \mathbb{C}^n$), $A \cdot |X| = \mathfrak{m}(A) \cdot |X|$ 이고, 따라서 $|X| > 0$.

증명 : (i) 편의상(notational convenience) A를 $\frac{1}{\mathfrak{m}(A)} A$ 로 'normalize'하여, $\mathfrak{m}(A) = 1$ 이라고 가정해도 좋을 것이다(확인해 보라).

(ii) $|\lambda| = \mathfrak{m}(A) = 1$ 이고, $AX = \lambda X$ 라고 하자(단, $0 \neq X \in \mathbb{C}^n$). 그러면,

$$|X| = |\lambda| \cdot |X| = |\lambda X| = |AX| \leq |A| \cdot |X| = A \cdot |X|$$

가 된다. 이제 $A \cdot |X| = |X|$ 임을 보이기 위해, $A \cdot |X| \neq |X|$ 라고 가정하면, 당연히 $A\big(A \cdot |X| - |X|\big) > 0$ 이 된다(왜 그런가?). 다음, 편의상 $Z = A \cdot |X|$ 로 표기하면, $Z > 0$ 임도 당연. 따라서 $A\big(Z - |X|\big) > \varepsilon Z$ 인 $\varepsilon > 0$ 이 존재할 것이다. 이는 $AZ - Z > \varepsilon Z$, 즉 $\frac{1}{1+\varepsilon} AZ > Z$ 라는 뜻이고, '美的감각'을 위해 $B = \frac{1}{1+\varepsilon} A$ 로 놓으면, $BZ > Z$ 가 된다. 따라서,

$$B^2 Z = B(BZ) > BZ > Z, \quad \ldots, \quad B^m Z > Z, \qquad (m \geq 1)$$

이다. 한편, $\mathfrak{m}(A) = 1$ 이므로, $\mathfrak{m}(B) = \frac{1}{1+\varepsilon} < 1$ 이고, 따라서 $\lim_{m \to \infty} B^m = 0$ 이 된다(연습문제 3.2 및 명제 3.5 참조). 그런데, 부등식 $B^m Z > Z$ 의 양변에 limit을 취하면, $0 \geq Z$ 가 된다. 모순! 따라서, $A \cdot |X| = |X| = \mathfrak{m}(A) \cdot |X|$ 이다. 한편, 당연히 $A \cdot |X| > 0$ 이므로, $A \cdot |X| = |X| > 0$. □

[76] Hint: 삼각화.

명제 7.7 λ가 A의 **dominant** eigen-value 이면(즉, $|\lambda| = \mathfrak{m}(A)$),

(가) $\lambda = |\lambda| > 0$.

(나) $\dim E_\lambda^A = \mathrm{mult}_A(\lambda) = 1$.

증명 : (i) 이번에도 편의상 A를 'normalize'하여, $\mathfrak{m}(A) = 1 = |\lambda|$ 이라고 가정해도 좋을 것이다(확인해 보라).

(ii) 이제 $0 \neq X = (x_1, \ldots, x_n)^{\mathbf{t}} \in \mathbb{C}^n$을 λ에 대응하는 A의 eigev-vector 라고 하자(즉, $AX = \lambda X$). 그러면, 앞 명제 7.6에 의해, $A \cdot |X| = |X| > 0$ 임을 알고 있다. 따라서,

$$A \cdot |X| = |X| = |\lambda| \cdot |X| = |\lambda X| = |AX|$$

이므로, i를 (아무거나) 하나 고정하고(단, $1 \leq i \leq n$), i-번째 좌표를 비교하면,

$$\textstyle\sum_j a_{ij} \cdot |x_j| = \big| \sum_j a_{ij} x_j \big|$$

가 된다. 그러므로, 이번에도 연습문제 3.11에 의해,

$$a_{ij} x_j = c_j z, \quad 즉, \quad x_j = z \tfrac{c_j}{a_{ij}} \qquad (1 \leq j \leq n)$$

인 $c_1, \ldots, c_n \geq 0$과 $z \in \mathbb{C}$가 존재할 것이다. 이는 λ에 대응하는 eigen-vector 는 모두 $Y = \big(\frac{c_1}{a_{i1}}, \ldots, \frac{c_n}{a_{in}}\big)^{\mathbf{t}}$의 상수배라는 뜻이다.[77] (그렇지만 아직 $\dim E_\lambda^A = 1$ 임을 증명한 것은 아니다. 왜냐하면 Y는 X의 선택에 depend 하기 때문이다.[78])

(iii) 앞 (ii)에 의하면 λ를 eigen-value 로 갖는 A의 eigev-vector 는 어떤 $Y \geq 0$의 상수 배이다. 따라서, 특별히, E_λ^A의 vector 는 모든 좌표가 0 이상이거나 또는 모든 좌표가 0 이하이거나 둘 중 하나일 것이다. 그러므로, $\dim E_\lambda^A = 1$ 임을 보이기 위해서는 E_λ^A에는 probability vector 가 하나뿐임을 증명하면 된다 (왜 그런가?). 이제 P, Q가 E_λ^A의 서로 다른 probability vector 라고 가정하자. 그러면 $0 \neq P - Q \in E_\lambda^A$이고, $P - Q$의 좌표들의 합은 0이므로, $P - Q$의 좌표들 중에는 양수와 음수가 모두 나타날 것이다. 모순. 따라서 $\dim E_\lambda^A = 1$.

(iv) 그런데 앞 (ii)에서, $|X| > 0$이므로, $c_1, \ldots, c_n \neq 0$이고, $Y > 0$. 이제,

$$\lambda Y = AY = |AY| = |\lambda Y| = |\lambda| \cdot Y = Y$$

이므로, $\lambda = 1 = |\lambda|$.

---

[77] 임의의 i에 대한 증명이 이채롭다.

[78] 이 gap을 지적해 준 유상우에게 감사.

(v) 이제 $\dim E_\lambda^A = \mathrm{mult}_A(\lambda)$ 인 것만 증명하면 되므로, $\dim E_\lambda^A \neq \mathrm{mult}_A(\lambda)$ 라고 가정해 보자. 이때 $J = U^{-1}AU$ 를 A 의 Jordan canonical form 이라고 놓으면, $\lim_{m\to\infty}\|J^m\| = \infty$ 가 된다(연습문제 3.2 및 정의 3.14 참조). 그리고,

$$\|J^m\| = \|U^{-1}A^mU\| \le n^2\|U^{-1}\|\cdot\|U\|\cdot\|A^m\|$$

이므로(연습문제 3.15(라)항), $\lim_{m\to\infty}\|A^m\| = \infty$ 이다. 이제 $A^m = \big(a_{ij}^{(m)}\big)$ 으로 표기하고, $\|A^m\| = a_{i_m j_m}^{(m)}$ 이라고 놓자. 또, 편의상 $Y = (y_1, \ldots, y_n)^{\mathbf{t}}$ 으로 표기하면, $AY = Y$ 이므로,

$$y_{i_m} = \textstyle\sum_j a_{i_m j}^{(m)} y_j \ge \big(\sum_j a_{i_m j}^{(m)}\big)\cdot\min_k\{y_k\} \ge \|A^m\|\cdot\min_k\{y_k\} \longrightarrow \infty$$

가 된다. 모순. □

이제 정리 7.2 의 증명(명제 7.6 과 명제 7.7) 이 끝났다. 따라서, 만약 $A > 0$ 이면, A 의 positive dominant eigen-vector 중 probability vector 를 P 라고 놓으면, P 는 유일하게 결정된다. P 의 이름은 A 의 Perron-Frobenius vector. 한편, $A^{\mathbf{t}} > 0$ 이기도 하므로, $A^{\mathbf{t}}$ 의 Perron-Frobenius vector Q 도 존재할 것이다($A^{\mathbf{t}}$ 의 dominant eigen-value 는 여전히 λ).

다음 따름정리도 대개 Perron-Frobenius Theorem 의 일부로 다루어진다.

따름정리 7.8 만약 $X \ge 0$ 이 A 의 eigen-value μ 에 대응하는 eigen-vector 이면, $\mu = \mathfrak{m}(A)$. (따라서 X 는 Perron-Frobenius vector P 의 positive scalar multiple.)

증명 : 위에서처럼 Q 를 $A^{\mathbf{t}}$ 의 Perron-Frobenius vector 라고 놓자. 그러면, 물론 $A^{\mathbf{t}}\cdot Q = \mathfrak{m}(A)Q$ 이고, $Q^{\mathbf{t}}\cdot X > 0$. 그런데, $AX = \mu X$ 이고,

$$\mu Q^{\mathbf{t}}\cdot X = Q^{\mathbf{t}}\cdot(\mu X) = (Q^{\mathbf{t}}\cdot A)X = \mathfrak{m}(A)Q^{\mathbf{t}}\cdot X$$

이므로, $\mu = \mathfrak{m}(A)$. □

연습문제 7.9 (Collatz-Wielandt Formula) $\mathcal{N} = \{X \in \mathbb{C}^n \mid X \ge 0 \text{ and } X \neq 0\}$ 으로 표기하고, 함수 $f : \mathcal{N} \to \mathbb{R}$ 을

$$f(X) = f\big((x_1, \ldots, x_n)^{\mathbf{t}}\big) = \min\Big\{\tfrac{AX\text{의 } i\text{-좌표}}{x_i} \,\Big|\, 1 \le i \le n,\ x_i \neq 0\Big\}, \qquad (X \in \mathcal{N})$$

으로 정의하면, $\mathfrak{m}(A) = \max\{f(X) \mid X \in \mathcal{N}\}$ 임을 보여라.

version

180102

참고 문헌

[1] M. Artin, *Algebra*, Prentice-Hall, 1991.

[2] S. Brin and L. Page, *The anatomy of a large-scale hypertextual Web search engine*, Computer Networks and ISDN Systems, 30, 107–117, 1998.

[3] S. H. Friedberg, A. J. Insel and L. E. Spence, *Linear Algebra*, 4th ed., Pearson, 2002.

[4] G. H. Golub and C. F. Van Loan, *Matrix Computation*, 4th ed., JHU Press, 2012.

[5] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000.

[6] L. Page, S. Brin, R. Motwani and T. Winograd, *The PageRank citation ranking: bringing order to the web*, Technical Report, Stanford InfoLab, 1999.