Чи завжди додатний степінь стохастичної матриці є також стохастичною матрицею?

Ірина Сівак[1], *Олександр Кукуш*

1. Постановка задачі та деякі позначення

Опишемо наш основний об'єкт дослідження.

Означення 1. Квадратна матриця називається стохастичною зліва, якщо усі її елементи невід'ємні і сума елементів кожного стовпчика дорівнює 1.

Надалі ми такі матриці коротко будемо називати *стохастичними*. Наприклад, стохастичною є матриця

$$C = \begin{pmatrix} 1/2 & 2/3 \\ 1/2 & 1/3 \end{pmatrix}.$$

Розглянемо дві стохастичні матриці $A = (a_{ij})_{i,j=1}^m$ та $B = (b_{ij})_{i,j=1}^m$. Їх елементи невід'ємні, тому для матриці $C = AB = (c_{il})_{i,l=1}^m$ маємо

$$c_{il} = \sum_{j=1}^m a_{ij} b_{jl} \geq 0, \ 1 \leq i, l \leq m.$$

Крім того, при кожному $1 \leq l \leq m$ виконується

$$\sum_{i=1}^m c_{il} = \sum_{i=1}^m \sum_{j=1}^m a_{ij} b_{jl} = \sum_{j=1}^m \sum_{i=1}^m a_{kj} b_{jl} =$$

$$= \sum_{j=1}^m \Big(\sum_{i=1}^m a_{ij} \Big) b_{jl} = \sum_{j=1}^m b_{jl} = 1.$$

Отже, AB є також стохастичною матрицею. Звідси випливає, що при кожному $n \in \mathbb{N}$ степінь A^n теж є стохастичною матрицею.

Нижче ми для дійсних $\lambda > 0$ та стохастичних матриць A означимо матрицю A^λ і з'ясуємо, за яких умов A^λ зберігає властивість стохастичності.

Для векторів $x = (x_1, x_2, \ldots, x_m)^T \in \mathbb{R}^m$ будемо використовувати так звану l_1-норму:

$$\|x\| = \|x\|_1 := \sum_{i=1}^m |x_i|.$$

Простір $\mathbb{R}^m$ з цією нормою позначають $\mathbb{R}_1^m$. Розглянемо матрицю A та відповідний лінійний оператор $\mathcal{A}$ в $\mathbb{R}_1^m$ такий, що $\mathcal{A}x = Ax, \ x \in \mathbb{R}^m$. Визначимо норму цієї матриці як норму оператора $\mathcal{A}$:

$$\|A\| = \|\mathcal{A}\| = \sup_{x \in \mathbb{R}^m, x \neq 0} \frac{\|\mathcal{A}x\|}{\|x\|}.$$

[1] Студентка 3 курсу механіко-математичного факультету Київського національного університету імені Тараса Шевченка.

Можна показати (див. [1], гл. 8), що

$$\|A\| = \max_{1 \le j \le m} \sum_{i=1}^{m} |a_{ij}|.$$

З властивостей операторних норм випливає, що для двох матриць розміру $m \times m$ виконується:

$$\|AB\| \le \|A\| \|B\|.$$

Через E позначатимемо одиничну матрицю розміру $m \times m$.

2. Степінь квадратної матриці

Означимо спочатку логарифм від квадратної матриці Π, близької до одиничної. Ми базуємося на розкладі, аналогічному розкладу логарифмічної функції в ряд Тейлора в околі одиниці:

$$\ln x = \ln(1 + (x-1)) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (x-1)^k, \ |x-1| < 1.$$

Означення 2. Нехай $\|\Pi - E\| < 1$. Тоді

$$\ln \Pi = \sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{k} (\Pi - E)^k, \tag{1}$$

де матричний ряд збігається поелементно.

Логарифм від матриці можна обчислювати через її жорданову нормальну форму. Нехай $\|\Pi - E\| < 1$, J_Π — жорданова нормальна форма з матрицею переходу T. Тоді $\Pi = T J_\Pi T^{-1}$, причому усі числа в жорданових клітинах додатні. Можна показати, що

$$\ln \Pi = T (\ln J_\Pi) T^{-1},$$

причому для знаходження $\ln J_\Pi$ стандартним чином використовують логарифмічну функцію від жорданових клітин, що обчислюється за допомогою похідних від логарифмічної функції.

Зауважимо, що для квадратної матриці A експонента e^A задається як сума

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k,$$

де $A^0 := E$ і матричний ряд збігається поелементно.

Означення 3. Нехай $\|\Pi - E\| < 1, \ \lambda \in \mathbb{R}$. Степенем Π^λ будемо називати матрицю

$$\Pi^\lambda = e^{\lambda \ln \Pi}.$$

Таким чином,

$$\Pi^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} (\ln \Pi)^k. \tag{2}$$

Тут і надалі усі матричні ряди збігаються поелементно. Для Π^λ можна записати розклад, аналогічний розкладу в ряд Тейлора степеневої функції:

$$x^\lambda = (1 + (x-1))^\lambda = \sum_{k=0}^{\infty} C_\lambda^k (x-1)^k, \ |x-1| < 1,$$

де $C_\lambda^k = \frac{\lambda(\lambda-1)\ldots(\lambda-k+1)}{k!}$.

Можна показати, що при $\|\Pi - E\| < 1$ виконується рівність

$$\Pi^\lambda = \sum_{k=0}^{\infty} C_\lambda^k (\Pi - E)^k.$$

Лема 1. *Нехай сума елементів кожного стовпчика матриці* $\Pi = (\pi_{ij})_{i,j=1}^m$ *дорівнює 1, причому* $\|\Pi - E\| < 1$. *Тоді при кожному* $\lambda \in \mathbb{R}$ *сума елементів кожного стовпчика* Π^λ *дорівнює 1.*

Доведення. Позначимо через $\overline{1}$ вектор-стовпчик, складений з m одиниць. Тоді

$$\Pi^T \overline{1} = \Big(\sum_{i=1}^m \pi_{ij} \Big)_{j=1}^m = \overline{1}.$$

Далі,

$$(\Pi^\lambda)^T \overline{1} = \sum_{k=0}^{\infty} C_\lambda^k ((\Pi - E)^k)^T \overline{1} = \sum_{k=0}^{\infty} C_\lambda^k (\Pi^T - E)^k \overline{1}.$$

При $k \geq 1$ маємо $(\Pi^T - E)^k \overline{1} = (\Pi^T - E)^{k-1} (\Pi^T - E) \overline{1} = \overline{0}$; $C_\lambda^0 = 1$. Тому

$$(\Pi^\lambda)^T \overline{1} = (\Pi^T - E)^0 \overline{1} = \overline{1}.$$

Звідси випливає шукане твердження. □

3. Достатні умови стохастичності Π^λ

Для стохастичної матриці Π порахуємо норму її відхилення від одиничної. Маємо

$$\|\Pi - E\| = \max_{1 \leq j \leq m} \sum_{i=1}^m |(\Pi - E)_{ij}| = \max_{1 \leq j \leq m} \Big((1 - \pi_{jj}) + \sum_{i=1, i \neq j}^m \pi_{ij} \Big) =$$
$$= \max_{1 \leq j \leq m} 2(1 - \pi_{jj}) = 2(1 - \min_{1 \leq j \leq m} \pi_{jj}) = 2(1 - \theta(\Pi)),$$

де $\theta(\Pi) := \min\limits_{1 \leq j \leq m} \pi_{jj}$.

Тому для стохастичної матриці умова $\|\Pi - E\| < 1$ рівносильна нерівності $\theta(\Pi) > \frac{1}{2}$.

При $m = 2$ цього достатньо, щоб при кожному $\lambda > 0$ матриця Π^λ також була стохастичною матрицею (див. вправу 2). Але при $m \geq 3$ невід'ємність елементів Π^λ аж ніяк не гарантована (хоча за лемою 1 суми елементів матриці Π^λ по стовпчиках дорівнюють 1).

Приклад 1. Розглянемо стохастичну матрицю

$$\Pi = \begin{pmatrix} 0,8 & 0,15 & 0 \\ 0,2 & 0,7 & 0,2 \\ 0 & 0,15 & 0,8 \end{pmatrix}.$$

Її жорданова нормальна форма є діагональною матрицею

$$J_\Pi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0,8 & 0 \\ 0 & 0 & 0,5 \end{pmatrix}$$

з матрицею переходу

$$T = \begin{pmatrix} 3 & 1 & 1 \\ 4 & 0 & -2 \\ 3 & -1 & 1 \end{pmatrix},$$

причому

$$T^{-1} = \begin{pmatrix} 0,1 & 0,1 & 0,1 \\ 0,5 & 0 & -0,5 \\ 0,2 & -0,3 & 0,2 \end{pmatrix}.$$

Маємо $\Pi = TJ_\Pi T^{-1}$, $\Pi^{1/2} = TJ_\Pi^{1/2}T^{-1}$. Остання матриця не є стохастичною, бо $(\Pi^{1/2})_{31} = 3 \cdot 0,1 - \sqrt{0,8} \cdot 0,5 + \sqrt{0,5} \cdot 0,2 \approx -0,006 < 0$.

Таким чином, при $m = 3$ не завжди додатний степінь стохастичної матриці зберігає цю властивість. Потрібні додаткові умови.

Теорема 1. *Нехай $\lambda > 0$, Π — стохастична матриця, $\theta(\Pi) > 1/2$ та $\ln \Pi$ є матрицею з невід'ємними позадіагональними елементами. Тоді Π^λ є також стохастичною матрицею.*

Доведення. З нерівності $\theta(\Pi) > 1/2$ випливає, що $\|\Pi - E\| < 1$. Тому з огляду на лему 1 достатньо довести, що усі елементи матриці Π^λ невід'ємні. Покажемо це.

Існує така стала c, що матриця

$$B = \ln \Pi + cE$$

матиме всі невід'ємні елементи. Тоді при кожному $n \in \mathbb{N},\ 1 \leq i, j \leq m$ виконується $(B^n)_{ij} \geq 0$.

Одинична матриця комутує з B, тобто $BE = EB$, тому за властивостями експоненти отримуємо

$$\Pi^\lambda = \exp(\lambda \ln \Pi) = \exp\big(\lambda(B - cE)\big) = \exp(\lambda B) \cdot \exp(-c\lambda E),$$

$$\exp(-c\lambda E) = \sum_{k=0}^{\infty} \frac{(-c\lambda E)^k}{k!} = \Big(\sum_{k=0}^{\infty} \frac{(-c\lambda)^k}{k!}\Big) E = e^{-c\lambda} E,$$

$$\Pi^{\lambda} = e^{-c\lambda} \exp(\lambda B) = e^{-c\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k B^k}{k!},$$

$$(\Pi^{\lambda})_{ij} = e^{-c\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k (B^k)_{ij}}{k!} \geq 0 \quad \text{для всіх } 1 \leq i, j \leq m.$$

Отже, Π^{λ} є стохастичною матрицею. $\square$

Зауважимо,що приклад 1 не задовольняє умови теореми 1, оскільки

$$(\ln \Pi)_{31} = (T(\ln J_{\Pi})T^{-1})_{31} = 0 \cdot 0,1 - (\ln 0,8) \cdot 0,5 + (\ln 0,5) \cdot 0,2 = \ln(0,8^{-0,5} \cdot 0,5^{0,2}) < 0,$$

бо $0,5^2 < 0,8^5$, що рівносильно правильній нерівності $5^5 = 3125 < 2^{12} = 4096$.

Вправа 1. Якщо в умовах теореми 1 вимагати, щоб матриця $\ln \Pi$ мала додатні позадіагональні елементи, то при $\lambda > 0$ всі елементи матриці Π^{λ} додатні. Довести.

Наведемо достатню умову додатності позадіагональних елементів матриці $\ln \Pi$. Позначимо

$$\mu(\Pi) = \min_{i \neq j} \pi_{ij}.$$

Нагадаємо, що π_{ij} — це елементи матриці Π.

Теорема 2. *Нехай* Π *— стохастична матриця,* $\theta(\Pi) > 1/2$,

$$\theta(\Pi) > 1 - \frac{1}{2}(\sqrt{\mu^2 + 2\mu} - \mu),$$

де $\mu = \mu(\Pi) > 0$. *Тоді* $(\ln \Pi)_{ij} > 0$ *при всіх* $i \neq j$.

Доведення. Нехай $i \neq j$. Тоді $(\Pi - E)_{ij} = \pi_{ij}$, і оскільки $\|\Pi - E\| = 2(1 - \theta(\Pi)) < 1$, то з (1) маємо

$$(\ln \Pi)_{ij} = \pi_{ij} + r, \text{ де } r := \sum_{k=2}^{\infty} \frac{(-1)^{k+1}}{k} ((\Pi - E)^k)_{ij}.$$

Далі, оскільки $\mu(\Pi) > 0$, то $\|\Pi - E\| > 0$. Маємо

$$|r| \leq \sum_{k=2}^{\infty} \frac{1}{k} |((\Pi - E)^k)_{ij}| \leq \sum_{k=2}^{\infty} \frac{1}{k} \|(\Pi - E)^k\| \leq \sum_{k=2}^{\infty} \frac{1}{k} \|(\Pi - E)\|^k <$$

$$< \frac{1}{2} \sum_{k=2}^{\infty} \|\Pi - E\|^k = \frac{1}{2} \cdot \frac{\|\Pi - E\|^2}{1 - \|\Pi - E\|}.$$

З урахуванням того, що $z = \|\Pi - E\| \in (0,1)$, нерівність

$$\frac{z^2}{2(1-z)} \le \mu = \mu(\Pi)$$

рівносильна такій:

$$z \le -\mu + \sqrt{\mu^2 + 2\mu},$$

або $\theta(\Pi) \ge 1 - \frac{1}{2}(\sqrt{\mu^2+2\mu} - \mu)$, що і дано в умові теореми 2. Тому

$$|r| < \mu \le \pi_{ij}, \ (\ln \Pi)_{ij} = \pi_{ij} + r > 0. \ \square$$

Наслідок 1. За умов теореми 2 степінь Π^λ є стохастичною матрицею при всіх $\lambda > 0$.

4. Необхідні умови стохастичності Π^λ

Теорема 3. *Нехай* Π *— стохастична матриця,* $\theta(\Pi) > 1/2$*, причому існує таке* $\delta > 0$*, що при всіх* $\lambda \in (0, \delta)$ *матриця* Π^λ *є стохастичною. Тоді* $(\ln \Pi)_{ij} \ge 0$ *для будь-яких* $i \ne j$*.*

Доведення. З розкладу (2) маємо

$$\Pi^\lambda = E + \lambda \ln \Pi + R(\lambda),$$

$\|R(\lambda)\| = O(\lambda^2), \ \lambda \to 0+$. Припустимо, що при деяких $i \ne j$ виконується $(\ln \Pi)_{ij} < 0$. Тоді

$$(\Pi^\lambda)_{ij} = \lambda (\ln \Pi)_{ij} + O(\lambda^2), \lambda \to 0+.$$

Звідси при достатньо малих $\lambda \in (0, \delta_1), \ \delta_1 \le \delta$, отримуємо $(\Pi^\lambda)_{ij} < 0$. Але тоді при таких λ матриця Π^λ не є стохастичною, що суперечить умові теореми 3.

Отже, припущення хибне, і теорему 3 доведено. $\square$

Зауважимо, що існує стохастична матриця Π, для якої $\ln \Pi$ має деякі нульові позадіагональні елементи, проте $(\Pi^\lambda)_{ij} > 0$ при всіх i, j та $\lambda > 0$.

Приклад 2. Розглянемо стохастичну матрицю:

$$\Pi = \frac{1}{4} \begin{pmatrix} 2(1+e^{-0,2}) & 2(1-e^{-0,2}) & 2(1-e^{-0,2}) \\ 1+e^{-0,4}-2e^{-0,2} & 1+e^{-0,4}+2e^{-0,2} & 1-3e^{-0,4}+2e^{-0,2} \\ 1-e^{-0,4} & 1-e^{-0,4} & 1+3e^{-0,4} \end{pmatrix}.$$

Безпосередньо перевіряється, що $\theta(\Pi) > 1/2$. Її жорданова нормальна форма є діагональною матрицею

$$J_\Pi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-0,2} & 0 \\ 0 & 0 & e^{-0,4} \end{pmatrix}$$

з матрицею переходу

$$T = \begin{pmatrix} 2 & 1 & 0 \\ 1 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix},$$

$$T^{-1} = \begin{pmatrix} 1/4 & 1/4 & 1/4 \\ 1/2 & -1/2 & -1/2 \\ 1/4 & 1/4 & -3/4 \end{pmatrix}.$$

Тоді при $\lambda > 0$

$$\Pi^\lambda = TJ_\Pi^\lambda T^{-1} = \frac{1}{4} \begin{pmatrix} 2(1+e^{-0,2\lambda}) & 2(1-e^{-0,2\lambda}) & 2(1-e^{-0,2\lambda}) \\ 1+e^{-0,4\lambda}-2e^{-0,2\lambda} & 1+e^{-0,4\lambda}+2e^{-0,2\lambda} & 1-3e^{-0,4\lambda}+2e^{-0,2\lambda} \\ 1-e^{-0,4\lambda} & 1-e^{-0,4\lambda} & 1+3e^{-0,4\lambda} \end{pmatrix}.$$

З нерівності Коші маємо $1+e^{-0,4\lambda} > 2\sqrt{e^{-0,2\lambda}} = 2e^{-0,2\lambda}$. Також

$$1-3e^{-0,4\lambda}+2e^{-0,2\lambda} = e^{-0,2\lambda}(e^{0,2\lambda}+2-3e^{-0,2\lambda}) \geq e^{-0,2\lambda}(3-3e^{-0,2\lambda}) > 0.$$

Тому $(\Pi^\lambda)_{ij} > 0$ при всіх i, j та $\lambda > 0$, а отже Π^λ є стохастичною матрицею.

Крім того,

$$\ln \Pi = T(\ln J_\Pi)T^{-1} = \begin{pmatrix} -0,1 & 0,1 & 0,1 \\ 0 & -0,2 & 0,2 \\ 0,1 & 0,1 & -0,3 \end{pmatrix},$$

тому $(\ln \Pi)_{21} = 0$.

З Теорем 1 та 3 отримуємо такий критерій стохастичності степеня:

Теорема 4. *Нехай $\delta > 0$, Π — стохастична матриця, $\theta(\Pi) > 1/2$. Степінь Π^λ є стохастичною матрицею при всіх $\lambda \in (0, \delta)$ тоді і лише тоді, коли $\ln \Pi$ має всі невід'ємні позадіагональні елементи.*

Вправа 2. Задано стохастичну матрицю Π розміру 2×2,

$$\Pi = \begin{pmatrix} \pi_{11} & 1-\pi_{22} \\ 1-\pi_{11} & \pi_{22} \end{pmatrix},$$

у якої $\pi_{11} > 1/2, \pi_{22} > 1/2$. Покажіть, що при кожному дійсному λ матриця Π^λ має вигляд

$$\Pi^\lambda = \frac{1}{1-\delta} \begin{pmatrix} 1-\pi_{22}+(1-\pi_{11})\delta^\lambda & (1-\pi_{22})(1-\delta^\lambda) \\ (1-\pi_{11})(1-\delta^\lambda) & 1-\pi_{11}+(1-\pi_{22})\delta^\lambda \end{pmatrix},$$

де $\delta = \det(\Pi) = \pi_{11}+\pi_{22}-1$. Переконайтесь, що вона є стохастичною при всіх $\lambda > 0$.

5. Стохастичні матриці в задачах з помилковою класифікацією

Втомлений читач може запитати: а кому потрібні оті степені стохастичних матриць? Коротко окреслимо їх можливе застосування.

Нехай X — випадкова величина, яка піддається помилковій класифікації. Ця величина може набувати значення $1, 2, \ldots, m$, але замість неї ми спостерігаємо дещо спотворену величину X^* з тією ж множиною значень. Процес помилкової класифікації описується за допомогою матриці $\Pi = (\pi_{ij})_{i,j=1}^m$, де $\pi_{ij} = P(X^* = i | X = j)$ задає ймовірність того, що величина кваліфікується в i-му стані, тоді як насправді вона знаходиться в j-му. Матриця Π є стохастичною, бо $\pi_{ij} \geq 0$,

$$\sum_{i=1}^{m} \pi_{ij} = P(X^* \in \{1, 2, \ldots, m\} | X = j) = 1, 1 \leq j \leq m.$$

Тут використані елементарні властивості умовних імовірностей.

У статистиці може виникнути потреба оцінити певний параметр, пов'язаний з X. Так от, у 2006 р. німецький професор Гельмут Кухенхофф (Helmut Küchenhoff) із співавторами запропонував вишуканий спосіб оцінювання, при якому матриці Π^λ, $\lambda > 0$, описують новий процес спостережень з помилковою класифікацією. Для цього, безперечно, Π^λ мусить бути стохастичною матрицею. Наприклад, виконання умов теореми 2 можуть це гарантувати.

Задачі з помилковою класифікацією природним чином виникають при діагностиці захворювань. Стан $X = 1$ може означати, що людина здорова, а $X = 2$ — що вона хвора. Тоді Π буде матрицею розміру 2×2, з якою дозволяє впоратись вправа 2.

Автори вдячні О. Шамову, студенту 1 курсу магістратури механіко-математичного факультету КНУ імені Тараса Шевченка, за плідні обговорення.

Література.

1. Ю.М. Березанский, Г.Ф. Ус, З.Г. Шефтель, *Функциональный анализ.* – К.: Выща школа, 1990. – 600 с.