Exercises in Mathematical Information Engineering 2A

Matrix scaling and its applications

Tomonari Sei*

June 7 (Wed) 2017

We learn about a version of Gordan's alternative theorem and its application to rating.

1 The theorem

The transpose of a vector $\boldsymbol{v}$ is denoted by $\boldsymbol{v}^\top$. The zero vector is denoted by $\mathbf{0}$. Define a probability simplex by

$$\Delta_p = \{\boldsymbol{w} = (w_1, \ldots, w_p)^\top \mid w_i \geq 0 \text{ and } \sum_{i=1}^{p} w_i = 1\}.$$

Theorem 1 (Gordan's alternative theorem)**.** Given a set of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p \in \mathbb{R}^n$, exactly one of the following conditions holds:

(a) there exists a vector $\boldsymbol{w} \in \Delta_p$ such that $\sum_{i=1}^{p} w_i \boldsymbol{x}_i = \mathbf{0}$.

(b) there exists a vector $\boldsymbol{b} \in \mathbb{R}^n$ such that $\boldsymbol{b}^\top \boldsymbol{x}_i > 0$ for all i.

The theorem is a corollary of the following stronger version. Let $\mathbb{R}_{>0}$ be the set of positive numbers.

Theorem 2 ([3])**.** Given a set of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p \in \mathbb{R}^n$, exactly one of the following conditions holds:

(a) there exists a vector $\boldsymbol{w} \in \Delta_p$ such that $\sum_{i=1}^{p} w_i \boldsymbol{x}_i = 0$.

(b) there exists a unique vector $\boldsymbol{w} \in \mathbb{R}_{>0}^p$ such that $(\sum_{j=1}^{p} w_j \boldsymbol{x}_j)^\top (w_i \boldsymbol{x}_i) = 1$ for all i.

Exercise. Prove Theorem 1 using Theorem 2.

The theorem is further a corollary of the following theorem. We call a symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ *strictly copositive* if $\boldsymbol{w}^\top \boldsymbol{A} \boldsymbol{w} > 0$ for all $\boldsymbol{w} \in \Delta_p$.

Theorem 3 (Marshall and Olkin [4])**.** Let $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ be a positive semi-definite and strictly copositive matrix. Then there exists a unique positive diagonal matrix $\boldsymbol{D}$ such that $\boldsymbol{DADe} = \boldsymbol{e}$, where $\boldsymbol{e} = (1, \ldots, 1)^\top \in \mathbb{R}^p$.

Exercise. Prove Theorem 2 using Theorem 3. [Hint: put $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p) \in \mathbb{R}^{n\times p}$ and $\boldsymbol{A} = \boldsymbol{X}^\top \boldsymbol{X}$.]

* 清 智也, sei@mist.i.u-tokyo.ac.jp

Proof of Theorem 3 (sketch). Define a function $\psi : \mathbb{R}^p_{>0} \to \mathbb{R}$ by

$$\psi(\boldsymbol{w}) = \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{A}\boldsymbol{w} - \sum_{i=1}^{p} \log w_i. \tag{1}$$

The stationary condition of ψ is written as $\boldsymbol{DADe} = \boldsymbol{e}$, where $\boldsymbol{D} = \mathrm{diag}(w_1, \ldots, w_p)$. Therefore it is enough to show that ψ has a unique minimum point.

The function ψ is shown to be strictly convex. Furthermore, for any given $c \in \mathbb{R}$, the set

$$L_c = \{\boldsymbol{w} \in \mathbb{R}^p_{>0} \mid \psi(\boldsymbol{w}) \le c\} \tag{2}$$

is shown to be compact. These claims are left as an exercise (→ see Assignment).

Let $\psi_0 = \inf\{\psi(\boldsymbol{w}) \mid \boldsymbol{w} \in \mathbb{R}^p_{>0}\}$. Note that ψ_0 may be $-\infty$. Take a sequence $\{\boldsymbol{w}^{(k)}\}_{k=1}^\infty$ in $\mathbb{R}^p_{>0}$ such that $\psi(\boldsymbol{w}^{(k)}) \searrow \psi_0$. Since $L = L_{\psi(\boldsymbol{w}^{(1)})}$ is compact, there exists an accumulating point $\boldsymbol{w}^* \in L$ of $\{\boldsymbol{w}^{(k)}\}_{k=1}^\infty$ due to the Bolzano-Weierstrass theorem. Then $\boldsymbol{w}^*$ is a minimum point of ψ by continuity of ψ. The uniqueness follows from the strict convexity of ψ. □

2 Interpretation of Theorem 2 using a balancing toy

Consider a balancing toy whose supporting point is the origin. Suppose that the direction of attaching p weights are given by $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p \in \mathbb{R}^3$. Then the p weights are balanced in a horizontal position if the condition (b) of Theorem 2 is satisfied.

3 Algorithm

There is a simple (but slow) algorithm for finding $\boldsymbol{D} = \mathrm{diag}(w_1, \ldots, w_p)$ in Theorem 3.

(i) Initialize $w_1, \ldots, w_p$. For example, $w_1 = \cdots = w_p = 1$.

(ii) For each $i = 1, \ldots, p$,

$w_i \leftarrow$ (the unique positive solution of the equation $\sum_{j=1}^p w_i A_{ij} w_j = 1$ with respect to w_i).

(iii) Put $\boldsymbol{D} = \mathrm{diag}(w_1, \ldots, w_p)$. If $\|\boldsymbol{DADe} - \boldsymbol{e}\|$ is small, output $\boldsymbol{D}$. Otherwise, go to Step (ii).

4 Application to rating

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p \in \mathbb{R}^n$ be scores on p academic subjects of n students. Assume that the sample mean of each $\boldsymbol{x}_i$ is zero without loss of generality. Let $\boldsymbol{w} \in \mathbb{R}^p_{>0}$ be the vector determined by Theorem 2 (b). Then the objective general index (OGI) is defined by $\boldsymbol{g} = \boldsymbol{X}\tilde{\boldsymbol{w}}$, where $\tilde{\boldsymbol{w}} = \sqrt{n}\boldsymbol{w}$. Refer to [5] for details.

5 A related theorem

For given positive square matrix $\boldsymbol{A} \in \mathbb{R}^{p\times p}_{>0}$, there exist positive diagonal matrices $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ such that $\boldsymbol{D}_1\boldsymbol{A}\boldsymbol{D}_2\boldsymbol{e} = \boldsymbol{e}$ and $\boldsymbol{e}^\top\boldsymbol{D}_1\boldsymbol{A}\boldsymbol{D}_2 = \boldsymbol{e}^\top$. This result is known as Sinkhorn's theorem [6]. See [1] for an application and [2] for further information.

References

[1] Langville, A.N. and C.D. Meyer (2012) *Who's #1?: The Science of Rating and Ranking*, Princeton University Press. (日本語訳：岩野和生・中村英史・清水咲里,「レイティング・ランキングの数理：No.1 は誰か？」, 共立出版)

[2] Idel, M. (2016). A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps, Preprint arxiv:1609.06349.

[3] Kalantari, B. (1996). A theorem of the alternative for multihomogeneous functions and its relationship to diagonal scaling of matrices, *Lin. Alg. Appl.*, **236**, 1–24.

[4] Marshall, A. W. and Olkin, L. (1968). Scaling of matrices to achieve specified row and column sums, *Numer. Math.*, **12**, 83–90.

[5] Sei, T. (2016). An objective general index for multivariate ordered data, *J. Multivariate Anal.*, **147**, 247–264.

[6] Sinkhorn, R. (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices, *Ann. Math. Statist.*, **35**, 876–879.

Assignment

Submit your report in the following manner.

To Sei's mailbox located at the 1st floor of the 6th building,
or by E-mail to `sei@mist.i.u-tokyo.ac.jp`

Due date June 21 (Wed) 2017.

Note Specify your name, student ID, and the course title. You may write the report in Japanese.

Problem Answer to at least one of the following questions.

1. Let $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ be a positive semi-definite and strictly copositive matrix. Show that
   (a) the function ψ defined by (1) is strictly convex on $\mathbb{R}^p_{>0}$, and
   (b) the set L_c defined by (2) is compact (i.e., bounded and closed).
   [Hint: For boundedness of L_c, use the strict copositivity of $\boldsymbol{A}$ to show that there exists $\rho > 0$ such that
   $$\boldsymbol{w}^\top \boldsymbol{A}\boldsymbol{w} \geq \rho \sum_i w_i^2$$
   for all $\boldsymbol{w} \in \mathbb{R}^p_{>0}$.]
2. Compute OGI for the data you collect.
3. Draw a design of the balancing toy shown in the lecture. Produce it by yourself and attach a photograph of it.