MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

An Objective General Index for Multivariate Ordered Data

Tomonari SEI

METR 2015-22 July 2015

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

An objective general index for multivariate ordered data

Tomonari SEI*

Abstract

A multivariate quantitative data is often summarized into a general index as a weighted sum when each variate has a prescribed order. Although the sum of standardized scores is a sensible choice of index, it may have negative correlation with some of the variates. In this paper, a general index that has positive correlation with all the variates is constructed. The index is applied to study the fairness of decathlon scoring. Quantification of ordered categorical data is also discussed. The limit of quantification characterizes the Gaussian distribution.

Keywords: convex minimization, correlation, general index, ordered data, quantification, ranking.

1 Introduction

Consider a data matrix of n individuals with p variates. For example, the data may be scores on p academic subjects of n students in a school, stock prices of p companies at n time points, decathlon data of n athletes about $p = 10$ events, and so on.

Our purpose is to construct a general index, that combines the p variates into a univariate index in order to rank the n individuals. The task is unsupervised in the sense that no one knows the correct index or ranking. This is a fairly classical and fundamental problem. For example, in the study of animal breeding programs, index selection is used to combine several traits without economic weights (e.g. [1], [7]). In sports data analysis, it is discussed how to score the combined events like decathlon and heptathlon (e.g. [5]). A number of university ranking systems are based on weighted average of relevant measures (e.g. [6]).

A natural index is the sum of standardized scores (Z-scores). This is a sensible choice if all the variates are uncorrelated. For correlated data, the first principal component is sometimes used as a general index since it maximizes the variance of the index under weight constraints. However, these indices can have negative correlation with some of the

*University of Tokyo, sei@mist.i.u-tokyo.ac.jp

variates. This property will be undesirable since a general index has to reflect all of the traits.

In this paper, we show that there is a general index that always has positive correlation to each variate. We will call it the objective general index. A mathematical property of positive definite matrices plays an important role in the construction. The weight is numerically obtained by a simple convex programming. As an example, we study the fairness of the scoring rule of decathlon.

Our index is extended to the case of ordered categorical variates. This is related to but distinct from the optimal scaling method for ordered categories (e.g. [2], [15]). Remarkably, it is shown that the limit of the quantified data is a Gaussian random variable.

The paper is organized as follows. In Section 2, we introduce two conditions on general indices. Then recall a key property of positive definite matrices in Section 3. The objective general index is defined in Section 4. Two examples are given in Section 5. We take into account multicollinearity and subjective importance in Section 6 as well as ordered categorical data in Section 7. Section 8 is devoted to a functional version of the index, which characterizes the Gaussian distribution. Finally, we give some discussions in Section 9.

2 Two conditions on general indices

Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p) \in \mathbb{R}^{n \times p}$ be a p-variate data matrix, where each $\boldsymbol{x}_i$ is a column vector representing scores of n individuals. Assume that each variate is centered:

$$\mathbf{1}_n' \boldsymbol{x}_i = 0,$$

where $\mathbf{1}_n$ is the vector $(1, \ldots, 1)' \in \mathbb{R}^n$ and the symbol $'$ denotes the vector/matrix transpose. We further assume that each variate has a prescribed order in that a large value means good. For example, in decathlon data, the sign of the 100m time has to be changed before analysis.

Let the covariance matrix of $\boldsymbol{X}$ be

$$\boldsymbol{S} = (S_{ij})_{i,j=1}^{p}, \quad S_{ij} = \frac{1}{n} \boldsymbol{x}_i' \boldsymbol{x}_j.$$

Suppose that $\boldsymbol{S}$ is positive definite. This assumption will be relaxed in Section 6. We often, but not always, standardize the data in advance such that $S_{ii} = 1$ for any i. For standard data, $\boldsymbol{S}$ is equal to the correlation matrix.

A general index of $\boldsymbol{X}$ is a linear combination of p variates:

$$\boldsymbol{g} = \sum_{i=1}^{p} w_i \boldsymbol{x}_i = \boldsymbol{X}\boldsymbol{w},$$

where $\boldsymbol{w} = \boldsymbol{w}(\boldsymbol{S}) = (w_1, \ldots, w_p)' \in \mathbb{R}^p$ is a weight vector depending on $\boldsymbol{S}$. The map $\boldsymbol{S} \mapsto \boldsymbol{w}(\boldsymbol{S})$ is called *a weight map.* A weight map determines a general index.

The most fundamental general index is the simple sum

$$\sum_{i=1}^{p} \boldsymbol{x}_i,$$

whose weight map is $\boldsymbol{w}(\boldsymbol{S}) = \mathbf{1}_p = (1, \ldots, 1)'$, independent of $\boldsymbol{S}$. The sum of Z-scores

$$\sum_{i=1}^{p} \frac{\boldsymbol{x}_i}{\sqrt{S_{ii}}}$$

is more sensible if the columns of $\boldsymbol{X}$ have different units.

Another example is the first principal component, where the weight map $\boldsymbol{w}$ is an eigenvector of the covariance matrix $\boldsymbol{S}$ with respect to the largest eigenvalue. This makes the variance of $\boldsymbol{X}\boldsymbol{w}$ maximum under given $\boldsymbol{w}'\boldsymbol{w}$.

Our purpose is to construct a general index as fair as possible. Consider the following two conditions. For a vector $\boldsymbol{a} = (a_i)_{i=1}^p$, denote $\boldsymbol{a} > \mathbf{0}$ if $a_i > 0$ for every i.

Definition 1. A weight map $\boldsymbol{w} = \boldsymbol{w}(\boldsymbol{S})$ is said to be *consistent* if $\boldsymbol{w} > \mathbf{0}$ for any $\boldsymbol{S}$. It is said to be *covariance consistent* if the covariance between each variate and the general index is positive, or equivalently $\boldsymbol{S}\boldsymbol{w} > \mathbf{0}$, for any $\boldsymbol{S}$.

Consistency is a natural condition for the sake of general index: if an individual A is better than B in all the variates, then the general index of A should be better than B. In contrast, the meaning of covariance consistency is not trivial, but the positive-covariance property between each variate and the general index will be an acceptable condition. The two conditions have a duality relation (see Appendix A.2).

The weight map $\boldsymbol{w} = \mathbf{1}_p$ is obviously consistent, but does not have covariance consistency if $p \geq 3$. For example, consider a positive definite matrix

$$\boldsymbol{S} = \begin{pmatrix} 1 & -7/12 & -7/12 \\ -7/12 & 1 & 0 \\ -7/12 & 0 & 1 \end{pmatrix}, \qquad (1)$$

whose eigenvalues are 1 and $1 \pm (7/12)\sqrt{2}$. Then the covariance consistency fails:

$$\boldsymbol{S}\,\mathbf{1}_3 = \begin{pmatrix} -1/6 \\ 5/12 \\ 5/12 \end{pmatrix}.$$

The first principal component has neither consistency nor covariance consistency. For example, consider

$$\boldsymbol{S} = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}.$$

Then the eigenvector $\boldsymbol{w}$ corresponding to the largest eigenvalue $\lambda_1 = 3/2$ is a scalar multiple of $(1, -1)'$, which implies both $\boldsymbol{w}$ and $\boldsymbol{S}\boldsymbol{w}(= \lambda_1 \boldsymbol{w})$ are not positive vectors. The lack of consistency is also pointed out in [1].

An example satisfying covariance consistency is $\boldsymbol{w} = \boldsymbol{S}^{-1}\mathbf{1}_p$. Indeed, we have $\boldsymbol{S}\boldsymbol{w} = \mathbf{1}_p > \mathbf{0}$. However, it is not consistent. For example, consider a positive definite matrix

$$\boldsymbol{S} = \begin{pmatrix} 1 & 7/12 & 7/12 \\ 7/12 & 1 & 0 \\ 7/12 & 0 & 1 \end{pmatrix}.$$

Then

$$\boldsymbol{w} = \boldsymbol{S}^{-1}\mathbf{1}_3 = \begin{pmatrix} -12/23 \\ 30/23 \\ 30/23 \end{pmatrix}$$

has a negative component.

The last example also shows that a weight map $\boldsymbol{w}$ is not consistent if the covariance between $\boldsymbol{x}_i$ and $\boldsymbol{g} = \boldsymbol{X}\boldsymbol{w}$ does not depend on i. Indeed, such a general index should satisfy $\boldsymbol{x}_i'\boldsymbol{g} = c$ for some $c \in \mathbb{R}$, which is equivalent to $\boldsymbol{w} = cS^{-1}\mathbf{1}_p$. In contrast, the covariance between $w_i \boldsymbol{x}_i$ and $\boldsymbol{g}$ can be made independent of i. This is the objective general index we propose in Section 4.

Table 1 summarizes the properties of the indices.

Table 1: Consistency and covariance consistency of weight maps. The symbol • indicates that the weight map satisfies the condition.

	Consistency	Covariance consistency
The sum of Z-scores	•	—
The first principal component	—	—
$\boldsymbol{w} = S^{-1}\mathbf{1}_p$	—	•
OGI (defined in Section 4)	•	•

3 The bi-unit canonical form of positive definite matrices

Let p be a positive integer and $\mathbf{1}_p = (1, \ldots, 1)' \in \mathbb{R}^p$. We define terminology.

Definition 2. Let $\boldsymbol{B}$ be a $p \times p$ matrix. Then we call $\boldsymbol{B}$ *bi-unit* if it is positive definite and satisfies $\boldsymbol{B}\mathbf{1}_p = \mathbf{1}_p$.

Note that a bi-unit matrix is not necessarily doubly stochastic since it may have negative elements. Mathematical facts on bi-unit matrices are summarized in Appendix A.1.

For example, if $p = 2$, every bi-unit matrix is represented by

$$\boldsymbol{B} = \begin{pmatrix} 1-z & z \\ z & 1-z \end{pmatrix}, \quad -\infty < z < 1/2. \tag{2}$$

The following theorem will play an essential role in the construction of our index later.

Theorem 1 ([11], Corollary 2)**.** For any positive definite matrix $\boldsymbol{S} \in \mathbb{R}^{p\times p}$, there is a unique positive diagonal matrix D such that $\boldsymbol{DSD}$ is bi-unit, that is,

$$\boldsymbol{DSD1}_p = \boldsymbol{1}_p. \tag{3}$$

We recall a sketch of the proof since some equations will be referred to later.

Proof. Denote the elements of $\boldsymbol{S}$ by S_{ij}. Put $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{w})$, where $\boldsymbol{w} = (w_1, \ldots, w_p)'$ and $w_i > 0$. Then the equation (3) is

$$\sum_{j=1}^{p} S_{ij} w_i w_j = 1, \quad i = 1, \ldots, p.$$

Dividing both sides by w_i, we obtain

$$\sum_{j=1}^{p} S_{ij} w_j - \frac{1}{w_i} = 0, \quad i = 1, \ldots, p. \tag{4}$$

The left hand side is the gradient map of a strictly convex function

$$\psi(\boldsymbol{w}) = \frac{1}{2} \sum_{i,j=1}^{p} S_{ij} w_i w_j - \sum_{i=1}^{p} \log w_i \tag{5}$$

of $\boldsymbol{w} > \boldsymbol{0}$. Therefore, if there is a solution of (4), it is unique. The existence follows from the fact that ψ diverges as $w_i \to 0$ and as $\|w\| \to \infty$, respectively (see [14], Theorem 27.2, for details). □

Note that the minimization problem of the function (5) is equivalent to a program that minimizes $\boldsymbol{w}'\boldsymbol{Sw}$ under given $\sum_{i=1}^{p} \log w_i$. Compare to the principal component analysis that *maximizes* $\boldsymbol{w}'\boldsymbol{Sw}$ under given $\sum_{i=1}^{p} w_i^2$.

Definition 3. For any positive definite matrix $\boldsymbol{S}$, the bi-unit matrix $\boldsymbol{B} = \boldsymbol{DSD}$ induced by Theorem 1 is called *the bi-unit canonical form* of $\boldsymbol{S}$.

The bi-unit canonical form of a 2×2 positive definite matrix

$$\boldsymbol{S} = \begin{pmatrix} a & b \\ b & a \end{pmatrix}, \quad a > |b|,$$

is

$$\boldsymbol{DSD} = \frac{1}{a+b}\begin{pmatrix} a & b \\ b & a \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} \frac{1}{\sqrt{a+b}} & 0 \\ 0 & \frac{1}{\sqrt{a+b}} \end{pmatrix}.$$

For another example, if $\boldsymbol{S}$ is given by the equation (1), its bi-unit canonical form is

$$\boldsymbol{DSD} = \begin{pmatrix} 8 & -3.5 & -3.5 \\ -3.5 & 4.5 & 0 \\ -3.5 & 0 & 4.5 \end{pmatrix}, \quad \boldsymbol{D} = \mathrm{diag}\left(2\sqrt{2}, \frac{3\sqrt{2}}{2}, \frac{3\sqrt{2}}{2}\right). \tag{6}$$

The equation (4) is a set of algebraic equations over the region $\boldsymbol{w} > \boldsymbol{0}$. We can numerically solve it by a coordinate descent algorithm of the convex function (5). The algorithm is described in Table 2. Note that (4) is a quadratic equation for each i given $(w_j)_{j \neq i}$.

Table 2: An algorithm computing the bi-unit canonical form.

Input A positive definite matrix $\boldsymbol{S}$, initial value $\boldsymbol{w}_0 (= \boldsymbol{1}_p)$, tolerance $\epsilon > 0$.

Output A vector $\boldsymbol{w} > \boldsymbol{0}$ such that $\boldsymbol{DSD}$ is bi-unit, where $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{w})$.

1. $\boldsymbol{w} \leftarrow \boldsymbol{w}_0$
2. For $i = 1, \ldots, p$, in order, solve the quadratic equation (4) with respect to w_i.
3. If $\|\boldsymbol{w} - \boldsymbol{w}_0\| < \epsilon$, output $\boldsymbol{w}$. Otherwise $\boldsymbol{w}_0 \leftarrow \boldsymbol{w}$ and go to step 2.

In the rest of the section, we briefly discuss relations between bi-unit matrices and correlation matrices, where the correlation matrix of a given positive definite matrix $\boldsymbol{S}$ is determined by $\boldsymbol{R} = (S_{ij}/\sqrt{S_{ii}S_{jj}})$. Both of bi-unit and correlation matrices are considered as coordinate-wise scaling of $\boldsymbol{S}$. An interesting property of the set of bi-unit matrices is closedness under powers as well as inversion, that is, if $\boldsymbol{B}$ is bi-unit, then $\boldsymbol{B}^n$ is also bi-unit for any integer n. The set of correlation matrices does not have this property.

In contrast, the set of bi-unit matrices is *not* closed under sign change of row/columns, unlike correlation matrices. More precisely, if $\boldsymbol{B}$ is bi-unit and $\boldsymbol{E} = \mathrm{diag}(e_1, \ldots, e_p)$ with $e_i \in \{-1, 1\}$, then $\boldsymbol{EBE}$ is not necessarily bi-unit. Indeed, the equation (2) for $p = 2$ is not closed under the sign change $\boldsymbol{E} = \mathrm{diag}(1, -1)$. Another distinction from correlation matrices is that a principal minor of a bi-unit matrix is not bi-unit in general.

See Appendix A.1 for other relations between correlation and bi-unit matrices.

4 An objective general index (OGI)

Recall that $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ is a centered data matrix and $\boldsymbol{S}$ is its covariance matrix.

Definition 4. Let $\boldsymbol{DSD}$ be the bi-unit canonical form of $\boldsymbol{S}$ determined by Theorem 1, and denote the diagonal components of $\boldsymbol{D}$ by $\boldsymbol{w} = \boldsymbol{w}(\boldsymbol{S}) > \boldsymbol{0}$. Then define *the objective general index* (OGI) by

$$\boldsymbol{g}_{\mathrm{OGI}} = \sum_{i=1}^{p} w_i \boldsymbol{x}_i$$

We call $\boldsymbol{w}$ *the objective weight.*

Theorem 2. The objective weight has consistency and covariance consistency.

Proof. The weight w_i is positive from the definition. The covariance between $\boldsymbol{x}_i$ and $\boldsymbol{g} = \boldsymbol{g}_{\mathrm{OGI}}$ is

$$\frac{1}{n}\boldsymbol{x}_i'\boldsymbol{g} = \sum_{j=1}^{p} S_{ij} w_j = \frac{1}{w_i} > 0, \tag{7}$$

where the last equality follows from $\boldsymbol{DSD}\boldsymbol{1}_p = \boldsymbol{1}_p$. □

For example, if the covariance matrix is given by (1), then OGI is

$$\boldsymbol{g}_{\mathrm{OGI}} = 2\sqrt{2}\boldsymbol{x}_1 + \frac{3\sqrt{2}}{2}\boldsymbol{x}_2 + \frac{3\sqrt{2}}{2}\boldsymbol{x}_3$$

since the bi-unit canonical form is (6). Examples of real data sets are given in Section 5.

The following corollary is an immediate consequence of Theorem 1.

Corollary 1. Let $\boldsymbol{w} > \boldsymbol{0}$ and $\boldsymbol{g} = \sum_{i=1}^{p} w_i \boldsymbol{x}_i$. Then the following three conditions are equivalent to each other.

(i) $\boldsymbol{w}$ is the objective weight.

(ii) For each i, the covariance between $\boldsymbol{g}$ and $w_i \boldsymbol{x}_i$ is 1.

(iii) $\boldsymbol{g}$ is orthogonal to an affine hyperplane $L = \{\sum_i a_i w_i \boldsymbol{x}_i \mid \sum_i a_i = 1\}$, and $\boldsymbol{g}'\boldsymbol{g}/n = p$.

Proof. Equivalence of (i) and (ii) is obvious from (7). Assume (iii) holds. Since both $\boldsymbol{g}/p$ and $w_i \boldsymbol{x}_i$ are in L, we have $\boldsymbol{g}'(\boldsymbol{g}/p - w_i \boldsymbol{x}_i) = 0$. By $\boldsymbol{g}'\boldsymbol{g}/n = p$, we obtain $\boldsymbol{g}'(w_i \boldsymbol{x}_i)/n = 1$. Conversely, assume (ii). Take any two vectors $\sum_i a_i w_i \boldsymbol{x}_i$ and $\sum_i b_i w_i \boldsymbol{x}_i$ in L. Then

$$\frac{1}{n}\boldsymbol{g}'\left(\sum_i a_i w_i \boldsymbol{x}_i - \sum_i b_i w_i \boldsymbol{x}_i\right) = \sum_i (a_i - b_i) = 1 - 1 = 0.$$

We also have $\boldsymbol{g}'\boldsymbol{g}/n = \sum_i \boldsymbol{g}'(w_i \boldsymbol{x}_i)/n = p$. □

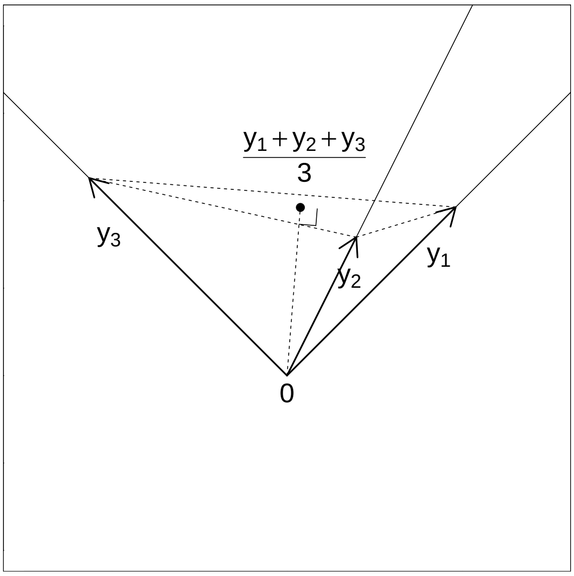


Figure 1: A geometric interpretation of OGI ($p = 3$), where $\boldsymbol{y}_i = w_i \boldsymbol{x}_i$, $1 \leq i \leq p$. In the space $\mathbb{R}^n$, the barycenter $\boldsymbol{g}/p$ of $\boldsymbol{y}_i$'s has a common inner product with each $\boldsymbol{y}_i$. The half lines out of the origin indicate the direction of $\boldsymbol{x}_i$'s.

From the condition (iii), we have a geometric interpretation of OGI as in Figure 1.

We also define *the population OGI* for a p-dimensional random vector

$$(X_1, \ldots, X_p)' : \Omega \to \mathbb{R}^p$$

on a probability space (Ω, P). The expectation is denoted by E. We assume that the variables are centered: $E[X_i] = 0$. Denote the population covariance matrix of $(X_1, \ldots, X_p)'$ by $\boldsymbol{S} = (S_{ij}) = (E[X_i X_j])$. Let $\boldsymbol{DSD}$ be the bi-unit canonical form of $\boldsymbol{S}$ (Theorem 1) and denote the diagonal components of $\boldsymbol{D}$ by $\boldsymbol{w} > \boldsymbol{0}$. Then define the population OGI by

$$G = \sum_{i=1}^{p} w_i X_i.$$

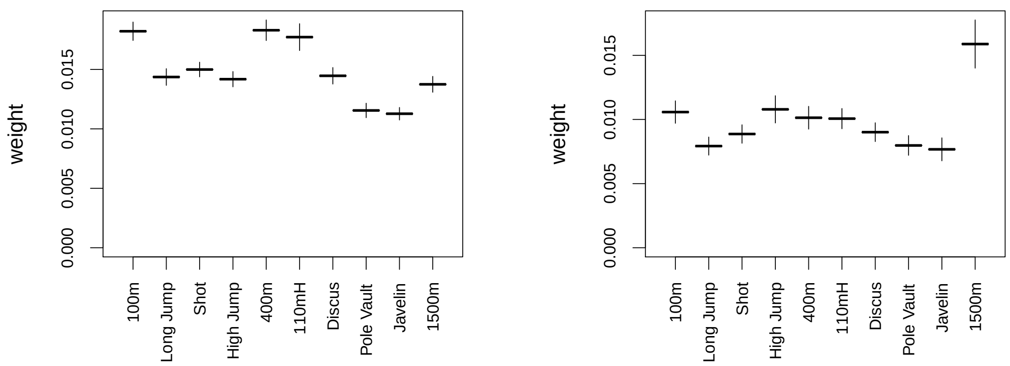
It satisfies $E[G w_i X_i] = 1$ for each i.

5 Examples

We compute the objective weights of two real datasets.

Example 1. The first example is the data for decathlon collected at the International Association of Athletics Federations (IAAF) World Athletics Championships held in 1991 to 2013. The data is available on the web site of IAAF. The data consists of $n = 235$

athletes on $p = 10$ events, after the data with missing values and an obvious outlier in 2007 were removed. We transformed the data nonlinearly according to IAAF scoring tables ([9], p.24). Figure 2 (a) shows the weight of the scaled sum, that is the reciprocal of the standard deviation. The confidence interval showing ± 1 standard error was computed by the bootstrap method. The three highest weights are 400m, 100m, and 110mH. As was reported in [5], it may be concluded that the present scoring rule could favor the athletes who are good in the field events. However, the objective weight shown in Figure 2 (b) has a different property. Only the weight of 1500m is quite larger than the other events. It suggests that the IAAF scoring method could be modified to weight 1500m more. Conservatively speaking, the fact that the correlation coefficient of the OGI and original total scores is 0.986 will support the fairness of the IAAF scoring method. Figure 3 shows the scatter plot of the two quantities.



(a) The weight of the scaled sum.

(b) The objective weight.

Figure 2: The weights of the decathlon data.

Example 2. Consider the `USJudgeRatings` data provided in `R` [13], that represents the lawyers' ratings of state judges in the US Superior Court. The data consists of $n = 43$ observations on $p = 12$ numeric variates. Figure 4 shows *the relative objective weight*, that means the objective weight of the standardized data. The standard error is computed by the bootstrap method. From the figure, the weight of the variate CONT (number of contacts of lawyer with judge) is about three times that of the other variates. This result is due to the high correlation between the variates other than CONT. In general, if $\boldsymbol{x}_1' \boldsymbol{x}_2 = 0$ and $\boldsymbol{x}_2 = \cdots = \boldsymbol{x}_p$, then the relative objective weight in an extended sense (see the following section) is given by $\boldsymbol{w} = (1, 1/\sqrt{p-1}, \ldots, 1/\sqrt{p-1})'$.

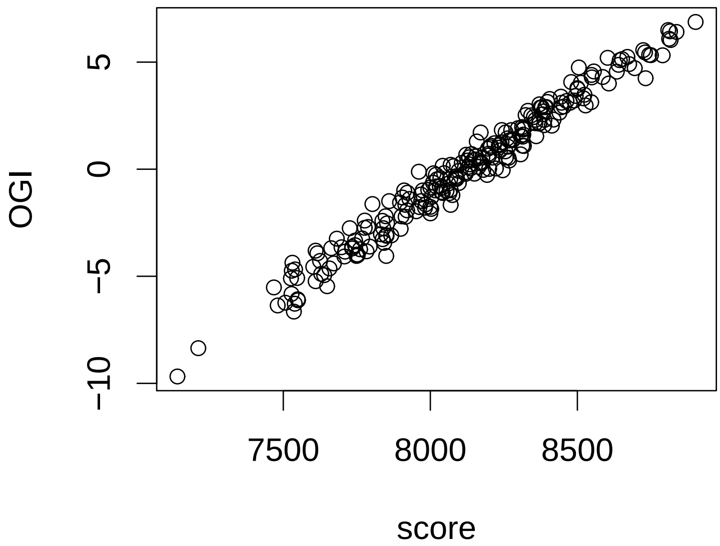


Figure 3: A scatter plot of the OGI against the total score for the decathlon data.

6 Extension of OGI

We extend the definition of OGI in two ways. First we show that even if there is multicollinearity in the data matrix $\boldsymbol{X}$, OGI is defined unless a positive combination of some columns is zero. Secondly, we incorporate *subjective importance* a priori in the definition of OGI.

6.1 Multicollinearity

The data matrix $\boldsymbol{X}$ is said to have multicollinearity if it is not column full-rank, or equivalently the covariance matrix is singular. Even for such data, a general index with the same property as OGI is constructed under a condition.

For example, consider 2-dimensional data $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$. If $\boldsymbol{x}_1 = \boldsymbol{x}_2$ and the variance of $\boldsymbol{x}_1$ is 1, then the covariance between $\boldsymbol{g} = (\boldsymbol{x}_1 + \boldsymbol{x}_2)/\sqrt{2}$ and $\boldsymbol{x}_i/\sqrt{2}$ is 1 for each i. However, if $\boldsymbol{x}_1 = -\boldsymbol{x}_2$, then there is no weight $\boldsymbol{w}$ such that the covariance between $\boldsymbol{X}\boldsymbol{w}$ and $\boldsymbol{x}_i$ is positive for each i. Difference of the two examples is described by the following condition of the covariance matrix. Denote the set of non-negative real numbers by $\mathbb{R}_{\geq 0}$.

Definition 5. Let $\boldsymbol{S}$ be a positive semi-definite matrix. Then $\boldsymbol{S}$ is called *strictly copositive* if $\boldsymbol{\lambda}'\boldsymbol{S}\boldsymbol{\lambda} > 0$ for any $\boldsymbol{\lambda} \in \mathbb{R}^p_{\geq 0} \setminus \{\boldsymbol{0}\}$.

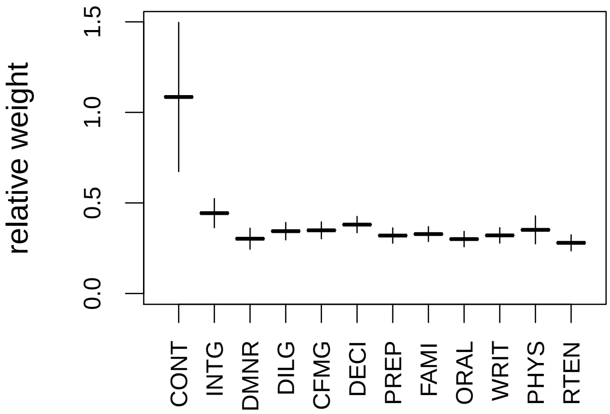


Figure 4: The relative objective weight of the lawyers' ratings data.

Theorem 1 is extended as follows. For later use, we also generalize the equation (3). The proof is similar and omitted.

Theorem 3 ([11], Theorem 1)**.** Let $\boldsymbol{S}$ be a $p \times p$ positive semi-definite matrix and $\boldsymbol{\nu}$ be a positive vector in $\mathbb{R}^p$. Then the equation

$$\boldsymbol{DSD1}_p = \boldsymbol{\nu}, \quad \boldsymbol{D} = \mathrm{diag}(\boldsymbol{w}), \quad \boldsymbol{w} > \boldsymbol{0},$$

has a unique solution if and only if $\boldsymbol{S}$ is strictly copositive.

If the covariance matrix of $\boldsymbol{X}$ satisfies the strict copositivity condition, then $\boldsymbol{g} = \boldsymbol{Xw}$ is well-defined, and called OGI when $\boldsymbol{\nu} = \boldsymbol{1}_p$.

For general square matrices, it is known that the copositivity condition is hard to confirm (see [12]). However, for positive semi-definite matrices, one can use methods of linear programming. We prepare a lemma.

Lemma 1. Let $\boldsymbol{S}$ be a positive semi-definite matrix such that $\boldsymbol{S} = \boldsymbol{X}'\boldsymbol{X}/n$ and $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p) \in \mathbb{R}^{n \times p}$. Then the following conditions are equivalent to each other.

(i) $\boldsymbol{S}$ is strictly copositive.

(ii) $(\ker \boldsymbol{S}) \cap \mathbb{R}^p_{\geq 0} = \{\boldsymbol{0}\}$.

(iii) $(\ker \boldsymbol{S}) \cap \{\boldsymbol{\lambda} \in \mathbb{R}^p_{\geq 0} \mid \boldsymbol{1}_p'\boldsymbol{\lambda} = 1\} = \emptyset$.

(iv) $(\ker \boldsymbol{X}) \cap \mathbb{R}^p_{\geq 0} = \{\mathbf{0}\}$, where $\boldsymbol{X}$ is considered as a linear map from $\mathbb{R}^p$ to $\mathbb{R}^n$.

(v) The convex cone $C = \{\sum_{i=1}^p \lambda_i \boldsymbol{x}_i \mid \lambda_1, \ldots, \lambda_p \geq 0\}$ is proper, that is, C contains no whole line.

Proof. Equivalence of (i) and (ii) follows from the identity $\{\boldsymbol{\lambda} \mid \boldsymbol{\lambda}'\boldsymbol{S}\boldsymbol{\lambda} = 0\} = \ker \boldsymbol{S}$ for positive semi-definite $\boldsymbol{S}$. Equivalence of (ii) and (iii) is obvious. Equivalence of (ii) and (iv) follows from $\ker \boldsymbol{S} = \ker \boldsymbol{X}$. To show (v) implies (iv), assume that there exists $\mathbf{0}_p \neq \boldsymbol{\lambda} \in (\ker \boldsymbol{X}) \cap \mathbb{R}^p_{\geq 0}$. There is i such that $\lambda_i > 0$. Then $\boldsymbol{x}_i = \sum_{k \neq i}(-\lambda_k/\lambda_i)\boldsymbol{x}_k$ and therefore both $\boldsymbol{x}_i$ and $-\boldsymbol{x}_i$ are contained in C. Thus C is not proper. Conversely, if C is not proper, then there exists a vector $\mathbf{0}_p \neq \boldsymbol{y} \in C$ such that $-\boldsymbol{y} \in C$. From the definition of C, there is $\boldsymbol{\lambda}, \boldsymbol{\mu} \in \mathbb{R}^p_{\geq 0} \setminus \{\mathbf{0}_p\}$ such that $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\lambda}$ and $-\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\mu}$. Then $\mathbf{0}_p = \boldsymbol{y} + (-\boldsymbol{y}) = \boldsymbol{X}(\boldsymbol{\lambda} + \boldsymbol{\mu})$ and therefore $(\ker \boldsymbol{X}) \cap \mathbb{R}^p_{\geq 0} \neq \{\mathbf{0}\}$. □

The condition (iii) is written as a feasibility problem in linear programming, and examined by standard methods (see e.g. Theorem 9.2 of [4] and its proof).

The condition (v) has geometric meaning as in Figure 1. It means that the p vectors never balance.

Example 3. Imagine a data set of baseball players with 11 attributes: the plate appearance (PA), at bat (AB), hit (H), double (DO), triple (TR), home run (HR), total base (TB), base on balls (BB), hit by a pitch (HBP), sacrifice hit (SH), and sacrifice fly (SF). This data set has the following structural multicollinearity as

$$\text{(PA)} - \text{(AB)} + \text{(BB)} + \text{(HBP)} + \text{(SH)} + \text{(SF)} = 0$$

and

$$\text{(TB)} - \text{(H)} - \text{(DO)} - 2 \cdot \text{(TR)} - 3 \cdot \text{(HR)} = 0.$$

However, the strict copositivity condition is fulfilled unless incidental multicollinearity occurs. Indeed, one can show that

$$\ker(\boldsymbol{X}) = \{(\lambda, -\lambda, -\mu, -\mu, -2\mu, -3\mu, \mu, \lambda, \lambda, \lambda, \lambda)' \mid \lambda, \mu \in \mathbb{R}\}$$

and therefore $\ker(\boldsymbol{X}) \cap \mathbb{R}^p_{\geq 0} = \{\mathbf{0}\}$.

6.2 Subjective importance

We call the parameter $\boldsymbol{\nu}$ in Theorem 3 *the subjective importance*. The OGI with the subjective importance $\boldsymbol{\nu}$ is defined by $\boldsymbol{g} = \sum_{i=1}^p w_i \boldsymbol{x}_i$, where $\boldsymbol{w}$ is determined by Theorem 3. It makes a specific variate more significant. It is also used to deal with ordered categorical data in Section 7.

7 Quantification of ordered categorical data

In this section, we define the OGI of ordered categorical variates. For ease of description, we consider random vectors instead of data matrix. Denote the probability space and expectation by (Ω, P) and E, respectively.

7.1 Univariate case

Consider a random variable X that takes values in a finite set $\{0, 1, \ldots, K\}$. The set has the usual order $0 < 1 < \cdots < K$. For each $k \in \{1, \ldots, K\}$, define a non-decreasing function h_k by

$$h_k(x) = 1_{\{x \geq k\}} - P(X \geq k), \quad x \in \{0, 1, \ldots, K\}.$$

Note that $E[h_k(X)] = 0$.

We obtain the population OGI of the K random variables $h_1(X), \ldots, h_K(X)$. Specifically, let

$$\begin{aligned} S_{kl} &= E[h_k(X)h_l(X)] \\ &= P(X \geq \max(k, l)) - P(X \geq k)P(X \geq l). \end{aligned}$$

Then the objective weight $(w_1, \ldots, w_K)$ is defined by Theorem 3, where the subjective importance is set to

$$\boldsymbol{\nu} = \left(\frac{1}{K}, \ldots, \frac{1}{K}\right)' = \frac{1}{K}\mathbf{1}_K.$$

By using the objective weight, the score of a realization x of X is defined by

$$y(x) = \sum_{k=1}^{K} w_k h_k(x), \quad x \in \{0, 1, \ldots, K\}. \tag{8}$$

The function preserves the order, that is, $y(x) < y(\tilde{x})$ if $x < \tilde{x}$. We call the procedure finding $y(x)$ *the univariate OGI quantification*. Note that $E[y(X)^2] = 1$ since

$$E[y(X)^2] = \boldsymbol{w}'\boldsymbol{S}\boldsymbol{w} = \mathbf{1}_K'\boldsymbol{\nu} = 1.$$

Our quantification method is different from the optimal scaling ([2]). In the optimal scaling, the weight $\boldsymbol{w}$ is determined in such a way that the variance of $y(X)$ is maximized under some conditions. In our method, the variance is *minimized* under given $\sum_{k=1}^{K} \log w_k$ (see the remark after Theorem 1).

Example 4. Table 3 shows an exam data ([8], Table 4.4), a result of exams on geometry and probability for 26 students. In this example, we use the set of marks $\{1-, 2, \ldots, 5\}$ instead of the set $\{0, 1, \ldots, 4\}$ for ease of explanation. The marginal distribution of the geometry exam is

$$(p_{1-}, p_2, p_3, p_4, p_5) = \left(\frac{1}{26}, \frac{2}{26}, \frac{5}{26}, \frac{14}{26}, \frac{4}{26}\right) = (0.038, 0.077, 0.192, 0.538, 0.154).$$

The objective weight is

$$(w_2, w_3, w_4, w_5) = (1.905, 1.012, 0.731, 1.162)$$

and the score (8) is

$$(y(1-), y(2), y(3), y(4), y(5)) = (-3.412, -1.506, -0.495, 0.236, 1.398).$$

Table 3: Result of examinations of geometry and probability ([8], Table 4.4).

Geometry \ Probability	5	4	3	2	1−	Total
5	2	1	1	0	0	4
4	8	3	3	0	0	14
3	0	2	1	1	1	5
2	0	0	0	1	1	2
1−	0	0	0	0	1	1
Total	10	6	5	2	3	26

7.2 Multivariate case

Consider a random vector $(X_1, \ldots, X_p)$. Each variable X_i is either ordered categorical or continuous. For ordered categorical variable X_i, assume its range is $\{0, 1, \ldots, K_i\}$ without loss of generality, and define $h_{ik}(x_i) = 1_{\{x_i \geq k\}} - P(X_i \geq k)$ for $1 \leq k \leq K_i$. For continuous variable X_i, let $K_i = 1$ and $h_{i1}(x) = x$.

The objective weight of the whole random variables

$$h_{ik}(X_i), \quad 1 \leq i \leq p, \quad 1 \leq k \leq K_i,$$

is defined by Theorem 3, where the subjective importance is set to $\nu_{ik} = 1/K_i$. Let (w_{ik}) be the objective weight. The score of a realization x_i of X_i is defined by

$$y_i(x_i) = \sum_{k=1}^{K_i} w_{ik} h_{ik}(x_i).$$

The OGI is written as

$$g(x_1, \ldots, x_p) = \sum_{i=1}^{p} y_i(x_i) = \sum_{i=1}^{p} \sum_{k=1}^{K_i} w_{ik} h_{ik}(x_i).$$

We call the procedure *the simultaneous OGI quantification.*

Again, our method is distinct from the multivariate version of the optimal scaling method proposed by [15]. In their method, the weight $\boldsymbol{w}$ is determined in such a way that the variance of the first principal component of $\{y_i(X_i)\}$ is maximized.

We also define a two-stage version of the OGI quantification. Namely, first compute the univariate OGI quantification of each categorical variable X_i as described in the last subsection. Denote it by $\tilde{g}_i(x_i)$. For continuous variables, put $\tilde{g}_i(x_i) = x_i$. Let $(v_1, \ldots, v_p)$ be the objective weight of the quantified vector $(\tilde{g}_1(X_1), \ldots, \tilde{g}_p(X_p))$. Then define the score of x_i by $\tilde{y}_i(x_i) = v_i \tilde{g}_i(x_i)$, and finally

$$\tilde{g}(x_1, \ldots, x_p) = \sum_{i=1}^{p} \tilde{y}_i(x_i) = \sum_{i=1}^{p} v_i \tilde{g}_i(x_i).$$

We call the procedure *the two-stage OGI quantification.* In general, the two-stage version is easier to compute than the simultaneous one since the former needs only covariance matrices of size $K_1, \ldots, K_p$, and p, while the latter needs a covariance matrix of size $\sum_{i=1}^{p} K_i$.

Example 5. Consider the exam data in Table 3 again. The data is bivariate. The results of OGI quantification and two-stage OGI quantification are shown in Table 4, Table 5, and Figure 5. In the figure, the diagonal axis denotes the OGI divided by 2, that is, $\sum_{i=1}^{2} y_i(x_i)/2$ for the simultaneous one and $\sum_{i=1}^{2} \tilde{y}_i(x_i)/2$ for the two-stage one, respectively. The results of the two methods are almost the same.

8 Functional OGI

The (population) general index considered so far was a linear combination $\sum_{i=1}^{p} w_i X_i$ of variables X_i. In this section, we define the functional OGI by a combination $\sum_{i=1}^{p} y_i(X_i)$ of increasing functions $y_i(X_i)$. The set of functions y_i will be determined by a nonlinear integral equation involving pairwise bivariate copulas.

Table 4: The simultaneous OGI quantification of the exam data

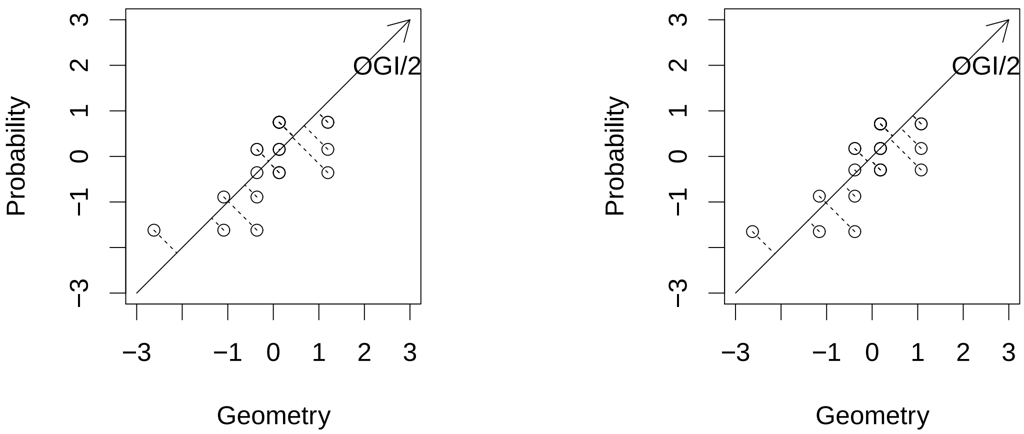
Geometry					
x	1−	2	3	4	5
weight w_{1x}		1.533	0.728	0.489	1.068
score $y_1(x)$	−2.621	−1.088	−0.360	0.129	1.197

Probability					
x	1−	2	3	4	5
weight w_{2x}		0.728	0.535	0.511	0.595
score $y_2(x)$	−1.619	−0.891	−0.357	0.155	0.750

Table 5: The two-stage OGI quantification of the exam data

Geometry ($v_1 = 0.770$)					
x	1−	2	3	4	5
marginal weight $\tilde{w}_{1x}$		1.905	1.012	0.731	1.162
marginal OGI $\tilde{g}_1(x)$	−3.412	−1.506	−0.495	0.236	1.398
score $\tilde{y}_1(x)$	−2.627	−1.160	−0.381	0.182	1.077

Probability ($v_2 = 0.770$)					
x	1−	2	3	4	5
marginal weight $\tilde{w}_{2x}$		1.011	0.747	0.612	0.701
marginal OGI $\tilde{g}_2(x)$	−2.144	−1.133	−0.386	0.226	0.927
score $\tilde{y}_2(x)$	−1.651	−0.873	−0.297	0.174	0.714



(a) The simultaneous quantification.

(b) The two-stage quantification.

Figure 5: The simultaneous and two-stage OGI quantification of exam data.

8.1 Univariate functional OGI

Let X be a random variable with the uniform distribution on $(0,1)$. For each $\xi \in (0,1)$, define a non-decreasing function $h_\xi : (0,1) \to \mathbb{R}$ by

$$\begin{aligned} h_\xi(x) &= 1_{\{x \geq \xi\}} - (1-\xi) \\ &= -(1_{\{x<\xi\}} - \xi). \end{aligned}$$

Note that $E[h_\xi(X)] = 0$. Let $S_{\xi\eta}$ be the covariance between $h_\xi(X)$ and $h_\eta(X)$:

$$S_{\xi\eta} = E[h_\xi(X)h_\eta(X)] = \min(\xi,\eta) - \xi\eta.$$

The functions $h_\xi(x)$ are considered as a basis of the set of increasing functions in the following sense. For any positive continuous function w_ξ of $\xi \in (0,1)$ with a constraint $\int_0^1 \xi(1-\xi)w_\xi d\xi < \infty$, the function

$$y(x) = \int_0^1 w_\xi h_\xi(x) d\xi \tag{9}$$

is an increasing function satisfying $y'(x) = w_x$ and $E[y(X)] = 0$ (see Appendix A.3).

The objective weight of the infinite number of the random variables $\{h_\xi(X)\}_{\xi\in(0,1)}$ is defined by the solution w_ξ of an integral equation

$$\int_0^1 S_{\xi\eta} w_\eta d\eta = \frac{1}{w_\xi}, \quad \xi \in (0,1).$$

See subsection A.4 for more details. In terms of $y(x)$ in (9), the equation is equivalent to

$$E[h_\xi(X)y(X)] = \frac{1}{y'(\xi)}, \quad \xi \in (0,1). \tag{10}$$

Then a continuous version of OGI is defined by

$$G = y(X) = \int_0^1 w_\xi h_\xi(X) d\xi. \tag{11}$$

We call G *the univariate functional OGI* for convenience even though there is no variety of the distribution of X.

We obtain the following remarkable fact. Denote the cumulative distribution function of the standard normal distribution by $\Phi(x) = \int_{-\infty}^x (2\pi)^{-1/2} e^{-u^2/2} du$.

Theorem 4. Let $y : (0,1) \to \mathbb{R}$ denote an increasing and continuously differentiable function such that $\int_0^1 x(1-x)y'(x)dx < \infty$. Then the unique solution of (10) is $y(x) = \Phi^{-1}(x)$. In particular, G has the standard normal distribution.

Proof. The equation (10) is written as

$$\int_\xi^1 y(x)dx = \frac{1}{y'(\xi)}. \tag{12}$$

The derivative of (12) is $-y(\xi) = -y''(\xi)/y'(\xi)^2$, which is equivalent to

$$\{\Phi(y(\xi))\}'' = 0.$$

By integrating twice, we obtain $\Phi(y(\xi)) = C\xi + D$, or

$$y(\xi) = \Phi^{-1}(C\xi + D).$$

By the condition (12), it is necessary to be $y'(0+) = \infty$ and $y'(1-) = \infty$. This implies $C = 1$ and $D = 0$. Therefore $y(\xi) = \Phi^{-1}(\xi)$. It certainly satisfies (12). Finally, $G = y(X)$ has the distribution Φ. □

Remark 1. Consider other weights. If we simply use $w_\xi = 1$, then the general index

$$G = \int_0^1 h_\xi(X)d\xi = X$$

is distributed according to the uniform distribution. Standardization corresponds to $w_\xi = 1/\sqrt{\xi(1-\xi)}$ since $E[h_\xi(X)^2] = \xi(1-\xi)$. Then the general index

$$G = \int_0^1 \frac{h_\xi(X)}{\sqrt{\xi(1-\xi)}}d\xi = \mathrm{Sin}^{-1}(2X-1)$$

has the distribution $P(G \le g) = (1 + \sin g)/2$ for $-\pi/2 \le g \le \pi/2$. The first principal component corresponds to the solution of

$$\int_0^1 S_{\xi\eta} w_\eta d\eta = \lambda w_\xi$$

associated with the largest eigenvalue λ. It is shown that $\lambda = 1/\pi^2$ and $w(x) = \pi\cos(\pi(x - 1/2))$. Thus

$$G = \int_0^1 w_\xi h_\xi(X)d\xi = \sin(\pi(X - 1/2))$$

has the arcsine distribution.

8.2 Multivariate functional OGI

Let $(X_1, \ldots, X_p)$ be a random vector with uniform marginals on $(0,1)$. Define the same function h_ξ as the univariate case. Let $S_{\xi i,\eta j}$ be the covariance between random variables $h_\xi(X_i)$ and $h_\eta(X_j)$:

$$S_{\xi i,\eta j} = E[h_\xi(X_i)h_\eta(X_j)] = C_{ij}(\xi,\eta) - \xi\eta,$$

where $C_{ij}(\xi,\eta) = P(X_i \le \xi, X_j \le \eta)$ is the bivariate copula of (X_i, X_j). Note that $C_{ii}(\xi,\eta) = \min(\xi,\eta)$ if $i = j$.

The objective weight $w_{\xi i}$ is the solution of

$$\sum_j \int_0^1 S_{\xi i,\eta j} w_{\eta j} d\eta = \frac{1}{w_{\xi i}}, \quad i = 1,\dots,p, \quad \xi \in (0,1).$$

In terms of $y_i(x) = \int_0^1 w_{\xi i} h_\xi(x) d\xi$, it is equivalent to

$$\sum_j E[h_\xi(X_i) y_j(X_j)] = \frac{1}{y_i'(\xi)}. \tag{13}$$

The multivariate functional OGI is defined by

$$G = \sum_{i=1}^p y_i(X_i) = \sum_{i=1}^p \int_0^1 w_{\xi i} h_\xi(X_i) d\xi.$$

The following theorem is an extension of the univariate case. Denote the multivariate normal distribution with the mean $\mathbf{0}$ and covariance matrix $\boldsymbol{S}$ by $\mathrm{N}(\mathbf{0}, \boldsymbol{S})$. Recall that Φ is the cumulative distribution function of the standard normal distribution.

Theorem 5. Let $\boldsymbol{S}$ be a $p \times p$ correlation matrix and assume that $(\Phi^{-1}(X_1), \dots, \Phi^{-1}(X_p))$ is distributed according to $\mathrm{N}(\mathbf{0}, \boldsymbol{S})$. Let $\boldsymbol{v}$ be the objective weight of $\boldsymbol{S}$. Then $y_i(x) = v_i \Phi^{-1}(x)$ solves the equation (13).

Proof. Assume $y_i(x) = v_i \Phi^{-1}(x)$ and put $Z_i = \Phi^{-1}(X_i)$. Then we obtain

$$\begin{aligned} E[h_\xi(X_i) y_j(X_j)] &= E[1_{\{X_i \ge \xi\}} y_j(X_j)] \\ &= E[1_{\{Z_i \ge \Phi^{-1}(\xi)\}} v_j Z_j] \\ &= E[1_{\{Z_i \ge \Phi^{-1}(\xi)\}} E[v_j Z_j | Z_i]] \\ &= v_j S_{ij} E[1_{\{Z_i \ge \Phi^{-1}(\xi)\}} Z_i] \\ &= v_j S_{ij} \phi(\Phi^{-1}(\xi)), \end{aligned}$$

where $\phi = \Phi'$. Then the equation (13) is written as

$$\sum_j v_j S_{ij} \phi(\Phi^{-1}(\xi)) = \frac{\phi(\Phi^{-1}(\xi))}{v_i},$$

which is satisfied if and only if $\boldsymbol{v}$ is the objective weight of $\boldsymbol{S}$. □

Remark 2. Uniqueness of the solution is obtained if one restricts the space of $w_{\xi i}$ to the domain of a convex functional. See Subsection A.4 for details.

The theorem says that (13) is explicitly solved if the bivariate copulas $C_{ij}(\xi,\eta)$ are the Gaussian copula. The author is not aware of any other copula for which (13) has an explicit solution.

For functional OGI, we can define a two-step version in the same manner as the quantification method in Section 7: compute the OGI of transformed variables $\Phi^{-1}(X_i)$. The method is, however, theoretically not interesting since it is reduced to a finite-dimensional case.

9 Discussion

9.1 High-dimensional and/or missing data

We defined the objective weight by the bi-unit canonical form of the sample covariance matrix in Section 4. The standard error of the weight can be estimated by the bootstrap method. However, if the dimension p is large, the procedure will break down. The sample covariance should be replaced with some regularized estimator.

The data usually has missing values. Our method is available as long as the covariance matrix is appropriately estimated.

9.2 OGI-based principal component analysis

Let $\boldsymbol{XD}$, $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{w})$, be the scaled data determined by the objective weight. Then we can apply the principal component analysis (PCA) to the matrix

$$\boldsymbol{Z} = \boldsymbol{XD}(\boldsymbol{I}_p - \boldsymbol{1}_p\boldsymbol{1}_p'/p).$$

Then the matrix $\boldsymbol{Z}$ is orthogonal to the OGI, $\boldsymbol{g} = \boldsymbol{XD1}_p$, since

$$\begin{aligned}\boldsymbol{g}'\boldsymbol{Z} &= \boldsymbol{1}_p'\boldsymbol{D}'\boldsymbol{X}'\boldsymbol{XD}(\boldsymbol{I}_p - \boldsymbol{1}_p\boldsymbol{1}_p'/p) \\ &= \boldsymbol{1}_p'(\boldsymbol{I}_p - \boldsymbol{1}_p\boldsymbol{1}_p'/p) \\ &= \boldsymbol{0}_p'.\end{aligned}$$

Thus $\boldsymbol{Z}$ has information other than OGI.

9.3 Relation to the textile plot

In [10], a weight vector $\boldsymbol{w}$ is used to visualize high-dimensional data $\boldsymbol{X}$ effectively, where the weight is determined in such a way that $\boldsymbol{XD}$ is aligned as *horizontal* as possible in the parallel coordinate plot. The method is called the textile plot. If $\boldsymbol{X}$ has only numeric attribute without missing values and it is standardized, then the weight $\boldsymbol{w}$ becomes the eigenvector of the sample correlation matrix associated with the largest eigenvalue

(Corollary 1 of [10]), that is, the first principal component. We call it *the textile weight* here.

We can use the objective weight to visualize the data as well. A difference from the textile plot is that the objective weight $\boldsymbol{w}$ must be always positive. Figure 6 shows the parallel coordinate plot of the USJudgeRatings data. The plot based on OGI is rather different from the textile plot.

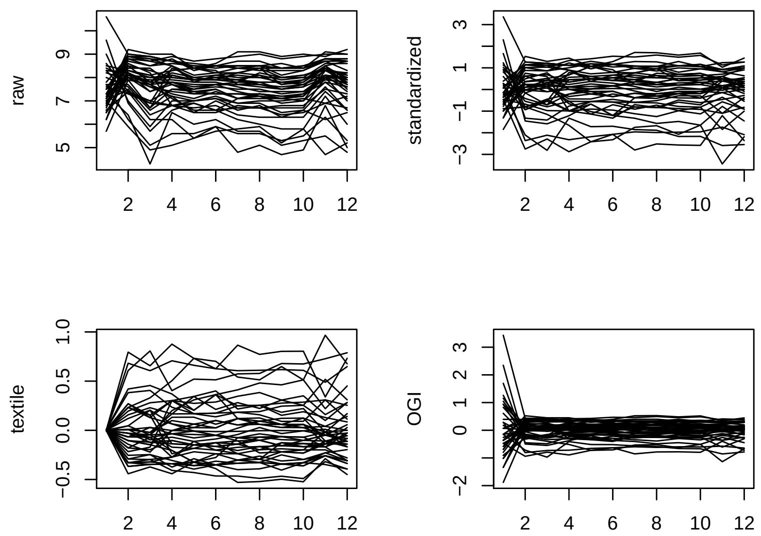


Figure 6: The parallel coordinate plot of the USJudgeRatings data based on the raw, standardized, textile-weighted, and OGI-weighted scores, respectively.

A Appendix

A.1 Mathematical properties of bi-unit matrices

We briefly summarize mathematical properties of bi-unit matrices defined in Definition 2.

Let $\mathcal{B}_p$ be the set of all bi-unit matrices in $\mathbb{R}^{p\times p}$. The set $\mathcal{B}_p$ is a convex set and closed under powers. As noted in Section 3, $\mathcal{B}_p$ is not closed under sign change of each row/columns. A principal minor of a bi-unit matrix is not bi-unit in general.

A matrix $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ is called *doubly stochastic* if every element of $\boldsymbol{A}$ is non-negative, $\boldsymbol{A}\mathbf{1} = \mathbf{1}$, and $\boldsymbol{A}'\mathbf{1} = \mathbf{1}$. Denote the set of all doubly stochastic matrices by $\mathcal{D}_p$. Then we have

$$\mathcal{B}_p \cap \mathcal{D}_p = \{\boldsymbol{B} \in \mathcal{B}_p \mid B_{ij} \geq 0 \text{ for all } i,j\} = \{\boldsymbol{A} \in \mathcal{D}_p \mid \boldsymbol{A} \text{ is positive definite}\}.$$

The set $\mathcal{B}_p \cap \mathcal{D}_p$ is studied in literature (e.g. [3]). There is no inclusive relation between $\mathcal{B}_p$ and $\mathcal{D}_p$. For example,

$$\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \in \mathcal{B}_2 \setminus \mathcal{D}_2 \quad \text{and} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathcal{D}_2 \setminus \mathcal{B}_2.$$

Denote the set of all correlation matrices by $\mathcal{C}_p$, where a correlation matrix means a positive definite matrix whose diagonal vector is $\mathbf{1}$. By Theorem 1, there is a one-to-one correspondence between $\mathcal{B}_p$ and $\mathcal{C}_p$. The correspondence is shown to be diffeomorphic.

If $p = 2$ or $p = 3$, $\mathcal{C}_p \cap \mathcal{B}_p$ consists of the identity matrix only. If $p = 4$, then $\mathcal{B}_p \cap \mathcal{C}_p$ is the set of positive definite matrices written as

$$\boldsymbol{B} = \begin{pmatrix} 1 & a & b & c \\ a & 1 & c & b \\ b & c & 1 & a \\ c & b & a & 1 \end{pmatrix}, \quad a + b + c = 0.$$

In general, $\mathcal{B}_p \cap \mathcal{C}_p$ is a $p(p-3)/2$-dimensional convex set if $p \geq 4$.

The set $\mathcal{B}_p$ is the intersection of a Lie group $\{\boldsymbol{G} \in \mathrm{GL}(p) \mid \boldsymbol{G}\mathbf{1} = \mathbf{1}\}$ and the cone of positive definite matrices. The Lie algebra is $\{\boldsymbol{A} \in \mathbb{R}^{p\times p} \mid \boldsymbol{A}\mathbf{1} = \mathbf{0}\}$. Any bi-unit matrix $\boldsymbol{B}$ is uniquely written as $\boldsymbol{B} = \exp(\boldsymbol{A}) = \sum_{k=0}^{\infty} \boldsymbol{A}^k/k!$ with a symmetric matrix $\boldsymbol{A}$ satisfying $\boldsymbol{A}\mathbf{1} = \mathbf{0}$. Any real power $\boldsymbol{B}^{\lambda}$ of a bi-unit matrix $\boldsymbol{B}$ is also bi-unit.

Choose a matrix $\boldsymbol{Q} \in \mathbb{R}^{p\times(p-1)}$ such that $(\mathbf{1}/\sqrt{p}, \boldsymbol{Q})$ is an orthogonal matrix. Then any bi-unit matrix $\boldsymbol{B}$ is uniquely written as

$$\boldsymbol{B} = \frac{1}{p}\mathbf{1}\mathbf{1}' + \boldsymbol{Q}\boldsymbol{A}\boldsymbol{Q}' \tag{14}$$

with a positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{(p-1)\times(p-1)}$. In particular, the set $\mathcal{B}_p$ is affinely isomorphic to the set of $(p-1)$-dimensional positive definite matrices.

A.2 Dual data

Denote the covariance matrix of a data matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ by $\boldsymbol{S}$. Assume that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ are linearly independent and therefore $\boldsymbol{S}$ is positive definite. Define *the dual data* by $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{S}^{-1}$. Each column of $\boldsymbol{Y}$ is written as $\boldsymbol{y}_i = \sum_j S^{ij}\boldsymbol{x}_j$, where S^{ij} is (i,j)-element of S^{-1}. The covariance matrix of $\boldsymbol{Y}$ is $\boldsymbol{S}^{-1}$.

Then the covariance between $\boldsymbol{y}_i$ and $\boldsymbol{x}_j$ is

$$\frac{1}{n}\boldsymbol{y}_i'\boldsymbol{x}_j = \frac{1}{n}\sum_k S^{ik}\boldsymbol{x}_k'\boldsymbol{x}_j = \sum_k S^{ik}S_{kj} = \delta_{ij}.$$

Therefore, for each i, $\boldsymbol{y}_i$ is orthogonal to $\{\boldsymbol{x}_j\}_{j\neq i}$. In terms of linear algebra, $\{\boldsymbol{y}_i\}$ is the dual basis of $\{\boldsymbol{x}_i\}$ with respect to the inner product. In terms of linear regression, $\boldsymbol{y}_i$ is (a scalar multiple of) the residual when $\boldsymbol{x}_i$ is explained by $\{\boldsymbol{x}_j\}_{j\neq i}$,

The following lemma shows that consistency defined in Definition 1 is equivalent to covariance consistency with respect to the dual data.

Lemma 2. A general index $\boldsymbol{g} = \boldsymbol{X}\boldsymbol{w}$ is consistent if and only if the covariance between $\boldsymbol{g}$ and $\boldsymbol{y}_i$ is positive for each i.

Proof. For any general index $\boldsymbol{g} = \sum_{i=1}^{p} w_i \boldsymbol{x}_i$, we have

$$\frac{1}{n}\boldsymbol{g}'\boldsymbol{y}_i = \sum_j w_j \frac{1}{n}\boldsymbol{x}_j'\boldsymbol{y}_i = \sum_j w_j \delta_{ij} = w_i.$$

Hence $w_i > 0$ if and only if $\boldsymbol{g}'\boldsymbol{y}_i > 0$. □

We also have an invariant property of OGI under the dual transformation.

Lemma 3. The OGI of the dual data $\boldsymbol{Y}$ is equal to that of the original data $\boldsymbol{X}$.

Proof. Denote the objective weight of $\boldsymbol{S}$ by w_i. The dual coordinate $\boldsymbol{y}_i$ has the covariance matrix $\boldsymbol{S}^{-1}$, whose objective weight is w_i^{-1}. Hence the OGI of $\boldsymbol{Y}$ is

$$\sum_i \frac{1}{w_i}\boldsymbol{y}_i = \sum_i \sum_j \frac{1}{w_i} S^{ij} \boldsymbol{x}_j = \sum_j w_j \boldsymbol{x}_j,$$

where the relation $\sum_i S^{ij}/w_i = w_j$ is used. □

A.3 The basis of increasing functions

Recall that $h_\xi(x) = 1_{\{x\geq\xi\}} - (1-\xi)$.

Lemma 4. Let w_x be a nonnegative continuous function of $x \in (0,1)$ such that $\int_0^1 w_x x(1-x)dx < \infty$. Then $y(x) = \int_0^1 w_\xi h_\xi(x) d\xi$ is a continuously differentiable non-decreasing function satisfying $y'(x) = w_x$ and $\int_0^1 y(x)dx = 0$.

Proof. Fix $x \in (0,1)$. We have $h_\xi(x) = \xi$ for $0 < \xi < x$, and $h_\xi(x) = -(1-\xi)$ for $x < \xi < 1$. Hence $y(x) = \int_0^1 w_\xi h_\xi(x)d\xi$ is finite by the assumption on w. For any $\delta \in (0,x)$, we have

$$y(x) - y(x-\delta) = \int_{x-\delta}^{x} w_\xi d\xi,$$

which implies $y(x)$ is non-decreasing and $y'(x) = w_x$. By Fubini's theorem, we obtain

$$\begin{aligned}\int_0^1 y(x)dx &= \int_0^1 \left(\int_0^1 w_\xi h_\xi(x)d\xi\right) dx \\ &= \int_0^1 w_\xi \left(\int_0^1 h_\xi(x)dx\right) d\xi = 0,\end{aligned}$$

where the condition of Fubini's theorem is checked as

$$\int_0^1 \int_0^1 w_\xi |h_\xi(x)| dx d\xi = 2\int_0^1 w_\xi \xi(1-\xi)d\xi < \infty.$$

□

A.4 Infinite-dimensional OGI

We formally define the objective general index of infinitely many random variables.

Let $\{X_\xi\}_{\xi\in\Xi}$ be a set of random variables, where Ξ is a measurable space with a finite measure μ. Assume that $E[X_\xi] = 0$ for each ξ. Then a general index of $\{X_\xi\}_{\xi\in\Xi}$ is defined by

$$G = \int_\Xi w_\xi X_\xi \mu(d\xi),$$

where w_ξ is a positive measurable function of ξ depending on the covariance process $S_{\xi\eta} = E[X_\xi X_\eta]$, $\xi, \eta \in \Xi$.

The objective weight $w = w[S]$ is defined by an integral equation

$$\int_\Xi w_\eta S_{\xi\eta}\mu(d\eta) = \frac{1}{w_\xi}, \quad \mu\text{- almost every } \xi. \tag{15}$$

This is the stationary condition of a convex functional

$$\Psi[w] = \frac{1}{2}\int_\Xi\int_\Xi w_\xi w_\eta S_{\xi\eta}\,\mu(d\xi)\mu(d\eta) - \int_\Xi (\log w_\xi)\mu(d\xi).$$

Then a uniqueness result is given in the following lemma. Existence is not discussed here.

Lemma 5. The solution of (15) over the region $\mathcal{W} = \{w \mid \Psi[w] < \infty\}$ is unique (μ-a.e.) if it exists.

Proof. Let $w \in \mathcal{W}$ and $w + \delta \in \mathcal{W}$. Then

$$\begin{aligned}&\Psi[w+\delta] - \Psi[w] \\ &= \frac{1}{2}\iint \delta_\xi\delta_\eta S_{\xi\eta}\mu(d\xi)\mu(d\eta) + \int \delta_\xi \int w_\eta S_{\xi\eta}\mu(d\eta)\mu(d\xi) - \int \log\frac{w_\xi+\delta_\xi}{w_\xi}\mu(d\xi) \\ &\geq \int \delta_\xi \int w_\eta S_{\xi\eta}\mu(d\eta)\mu(d\xi) - \int \log\frac{w_\xi+\delta_\xi}{w_\xi}\mu(d\xi) \\ &= \int \delta_\xi \left(\int w_\eta S_{\xi\eta}\mu(d\eta) - \frac{1}{w_\xi}\right)\mu(d\xi) + \int \left(\frac{\delta_\xi}{w_\xi} - \log\left(1+\frac{\delta_\xi}{w_\xi}\right)\right)\mu(d\xi) \\ &> \int_\Xi \delta_\xi \left(\int_\Xi w_\eta S_{\xi\eta}\mu(d\eta) - \frac{1}{w_\xi}\right)\mu(d\xi)\end{aligned}$$

unless δ is zero μ-almost everywhere. Hence the solution of (15) over $\mathcal{W}$ is the unique minimal point of Ψ. □

We give some examples. If $\Xi = \{1, \ldots, p\}$ and μ is the counting measure on Ξ, then (15) is equivalent to (3). The functional OGI in Section 8 uses $\Xi = [0, 1] \times \{1, \ldots, p\}$ and the product measure μ of the Lebesgue and counting measures.

Example 6 (Geostatistics)**.** Let $\Xi \subset \mathbb{R}^d$ be a bounded set with non-empty interior and define $\mu(A) = |A|/|\Xi|$, where $|A|$ is the Lebesgue measure of A. In geostatistics, the Matérn covariance function

$$S_{\xi\eta} = \frac{\phi}{2^{\nu-1}\Gamma(\nu)}(\alpha|\xi - \eta|)^{\nu} K_{\nu}(\alpha|\xi - \eta|)$$

is recommended to use (e.g. [16]), where ϕ, ν and α are positive parameters and K_ν denotes the modified Bessel function of the second kind. Figure 7 shows numerically evaluated weight functions w_ξ for several values of ν, where $\Xi = [0, 1]$ and $\alpha = \phi = 1$. The interval $\Xi = [0, 1]$ is approximated by $n = 100$ grid points. There is *an edge effect* in that the weight increases as ξ approaches to the boundary.

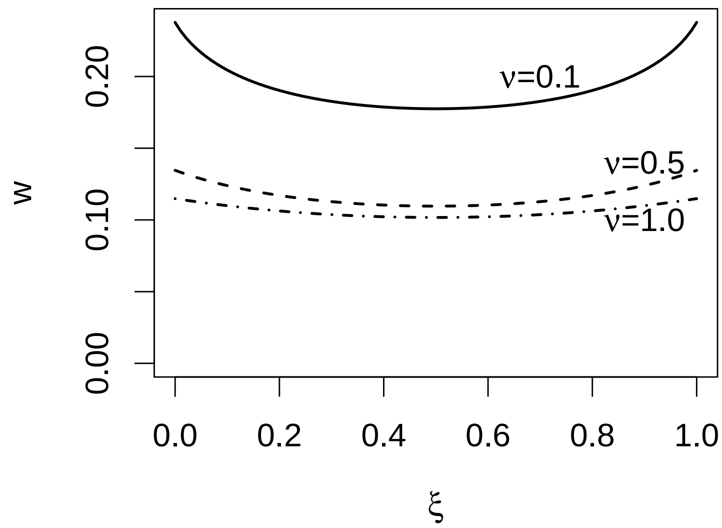


Figure 7: The objective weight function w_ξ of the Matérn class, where $\Xi = [0, 1]$ and $\alpha = \phi = 1$.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 26540013.

References

[1] R. J. Baker. Selection indexes without economic weights for animal breeding. *Canad. J. Animal Sci.*, 54(1):1–8, 1974.

[2] R. A. Bradley, S. K. Katti, and I. J. Coons. Optimal scaling for ordered categories. *Psychometrika*, 27(4):355–374, 1962.

[3] J. P. R. Christensen and P. Fischer. Positive definite doubly stochastic matrices and extreme points. *Linear Algebra Appl.*, 82:123–132, 1986.

[4] V. Chv́ital. *Linear Programming.* W. H. Freeman and Company, 1983.

[5] T. F. Cox and R. T. Dunn. An analysis of decathlon data. *J. Roy. Statist. Soc., Ser. D*, 51(2):179–187, 2002.

[6] D. D. Dill and M. Soo. Academic quality, league tables, and public policy: a cross-national analysis of university ranking systems. *Higher Education*, 49(4):495–533, 2005.

[7] R. C. Elston. A weight-free index for the purpose of ranking or selection with respect to several traits at a time. *Biometrics*, 19(1):85–97, 1963.

[8] T. Hibi, editor. *Gröbner Bases: Statistics and Software Systems.* Springer, 2013.

[9] IAAF Council, editor. *IAAF Scoring Tables for Combined Events.* IAAF, 2001.

[10] N. Kumasaka and R. Shibata. High-dimensional data visualisation: The textile plot. *Comput. Statist. Data Anal.*, 52:3616–3644, 2008.

[11] A. W. Marshall and I. Olkin. Scaling of matrices to achieve specified row and column sums. *Numer. Math.*, 12:83–90, 1968.

[12] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Math. Program.*, 39:117–129, 1987.

[13] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2014.

[14] R. T. Rockafeller. *Convex Analysis.* Princeton University Press, 1972.

[15] T. Saito and T. Otsu. A method of optimal scaling for multivariate ordinal data and its extensions. *Psychometrika*, 53(1):5–25, 1988.

[16] M. L. Stein. *Interpolation of Spatial Data – Some Theory for Kriging.* Springer, 1999.