

An objective general index for multivariate ordered data



Tomonari Sei

Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

ARTICLE INFO

Article history:

Received 19 July 2015

Available online 12 February 2016

AMS subject classifications:

15B48

62H20

62H25

Keywords:

Convex minimization

Correlation

General index

Ordered data

Quantification

Ranking

ABSTRACT

Multivariate quantitative data are often summarized into a general index as a weighted sum when the variates have 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.

© 2016 Elsevier Inc. All rights reserved.

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, or decathlon data of n athletes about $p = 10$ events.

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 in the context how to score combined events, such as decathlons and heptathlons (e.g., [5]). A number of university ranking systems are based on the 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 variates; this property is undesirable, since a general index should reflect all of the traits.

In this paper, we show that there is a general index that is always positively correlated with each of the variates. We will call this the *objective general index*. A mathematical property of positive definite matrices plays an important role in its construction. The weight is numerically obtained by simple convex programming. As an example, we study the fairness of the scoring rule for a 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., [2,18]). Remarkably, it is shown that the limit of the quantified data is a Gaussian random variable.

E-mail address: sei@mist.i.u-tokyo.ac.jp.

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

2. Two conditions on general indices

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

$$\mathbf{1}_n^\top \mathbf{x}_i = 0,$$

where $\mathbf{1}_n$ is the vector $(1, \dots, 1)^\top \in \mathbb{R}^n$, and the symbol $^\top$ denotes the vector/matrix transpose. We further assume that each variate has a prescribed order such that a larger value is associated with a better evaluation. For example, for decathlon data, the sign of the time for 100 m must be changed before analysis.

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

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

Suppose that $\mathbf{S}$ 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 standardized data, $\mathbf{S}$ is equal to the correlation matrix.

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

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

where $\mathbf{w} = \mathbf{w}(\mathbf{S}) = (w_1, \dots, w_p)^\top \in \mathbb{R}^p$ is a weight vector that depends on $\mathbf{S}$. The map $\mathbf{S} \mapsto \mathbf{w}(\mathbf{S})$ is called a *weight map*. A weight map determines a general index.

The most fundamental general index is the simple sum

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

whose weight map is $\mathbf{w}(\mathbf{S}) = \mathbf{1}_p = (1, \dots, 1)^\top$, independent of $\mathbf{S}$. If the columns of $\mathbf{X}$ have different units, then the sum of Z -scores

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

is more sensible.

Another example is the first principal component, where the weight map $\mathbf{w}$ is an eigenvector of the covariance matrix $\mathbf{S}$ with respect to the largest eigenvalue. This maximizes the variance of $\mathbf{X} \mathbf{w}$ under a given $\mathbf{w}^\top \mathbf{w}$.

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

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

Consistency is a natural requirement for a general index: if an individual A is better than B in all the variates, then the general index of A should be better than that of B. In contrast, the meaning of covariance consistency is not trivial, but at least this implies that none of the variates has negative correlation with the index. These two conditions have a duality relation (see [Appendix B](#)).

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

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

whose eigenvalues are 1 and $1 \pm (7/12)\sqrt{2}$. In this case, covariance consistency fails:

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

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

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

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	–	–
$\mathbf{w} = \mathbf{S}^{-1}\mathbf{1}_p$	–	•
OGI (defined in Section 4)	•	•

The eigenvector $\mathbf{w}$ corresponding to the largest eigenvalue $\lambda_1 = 3/2$ is a scalar multiple of $(1, -1)^\top$, which implies that neither $\mathbf{w}$ nor $\mathbf{S}\mathbf{w}(=\lambda_1\mathbf{w})$ is a positive vector. The lack of consistency was also pointed out in [1].

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

$$\mathbf{S} = \begin{pmatrix} 1 & 7/12 & 7/12 \\ 7/12 & 1 & 0 \\ 7/12 & 0 & 1 \end{pmatrix};$$

here,

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

has a negative component.

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

Table 1 summarizes the properties of these indices.

3. The bi-unit canonical form of positive definite matrices

Let p be a positive integer, and let $\mathbf{1}_p = (1, \dots, 1)^\top \in \mathbb{R}^p$. We define our terminology below.

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

Note that a bi-unit matrix is not necessarily doubly stochastic, since it may have negative elements. Some mathematical properties of bi-unit matrices are summarized in Appendix A.

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

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

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

Theorem 1 (Marshall and Olkin [13], Corollary 2). For any positive definite matrix $\mathbf{S} \in \mathbb{R}^{p \times p}$, there is a unique positive diagonal matrix $\mathbf{D}$ such that $\mathbf{DSD}$ is bi-unit, that is,

$$\mathbf{DSD}\mathbf{1}_p = \mathbf{1}_p. \quad (3)$$

We will sketch the proof, since some equations will be referred to below. Relevant information can be found in [10] and [15].

Proof. Denote the elements of $\mathbf{S}$ by S_{ij} . Let $\mathbf{D} = \text{diag}(\mathbf{w})$, where $\mathbf{w} = (w_1, \dots, w_p)^\top$ and $w_i > 0$. Then, (3) can be written as

$$\sum_{j=1}^p S_{ij}w_iw_j = 1, \quad i = 1, \dots, p.$$

Dividing both sides by w_i , we obtain

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

Table 2

Algorithm for computing the bi-unit canonical form.

Input A positive definite matrix $\mathbf{S}$, initial value $\mathbf{w}_0 (= \mathbf{1}_p)$, tolerance $\epsilon > 0$.
Output A vector $\mathbf{w} > \mathbf{0}$ such that $\mathbf{DSD}$ is a bi-unit matrix, where $\mathbf{D} = \text{diag}(\mathbf{w})$.
 1. $\mathbf{w} \leftarrow \mathbf{w}_0$
 2. For $i = 1, \dots, p$, in order, solve the quadratic Eq. (4) with respect to w_i .
 3. If $\|\mathbf{w} - \mathbf{w}_0\| < \epsilon$, output $\mathbf{w}$. Otherwise, $\mathbf{w}_0 \leftarrow \mathbf{w}$ and go to step 2.

The left-hand side is the gradient map of a strictly convex function of $\mathbf{w} > \mathbf{0}$:

$$\psi(\mathbf{w}) = \frac{1}{2} \sum_{i,j=1}^p S_{ij} w_i w_j - \sum_{i=1}^p \ln w_i. \quad (5)$$

Therefore, if there is a solution to (4), it is unique. The existence follows from the fact that ψ diverges as $w_i \rightarrow 0$ and as $\|\mathbf{w}\| \rightarrow \infty$. See [17], Theorem 27.2, for details. $\square$

Note that minimizing the function in (5) is equivalent to minimizing $\mathbf{w}^\top \mathbf{S} \mathbf{w}$ under a given $\sum_{i=1}^p \ln w_i$. Compare this to the principal component analysis that maximizes $\mathbf{w}^\top \mathbf{S} \mathbf{w}$ under a given $\sum_{i=1}^p w_i^2$.

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

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

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

is

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

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

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

Note that (4) is a set of algebraic equations over the region $\mathbf{w} > \mathbf{0}$. 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 $(w_j)_{j \neq i}$.

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

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

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

4. An objective general index (OGI)

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

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

$$\mathbf{g}_{\text{OGI}} = \sum_{i=1}^p w_i \mathbf{x}_i.$$

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

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

Proof. By definition, the weight w_i is positive. The covariance between $\mathbf{x}_i$ and $\mathbf{g} = \mathbf{g}_{\text{OGI}}$ is

$$\frac{1}{n} \mathbf{x}_i^\top \mathbf{g} = \sum_{j=1}^p s_{ij} w_j = \frac{1}{w_i} > 0, \quad (7)$$

where the last equality follows from $\mathbf{DSD}\mathbf{1}_p = \mathbf{1}_p$. $\square$

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

$$\mathbf{g}_{\text{OGI}} = 2\sqrt{2}\mathbf{x}_1 + \frac{3\sqrt{2}}{2}\mathbf{x}_2 + \frac{3\sqrt{2}}{2}\mathbf{x}_3,$$

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

The following corollary is an immediate consequence of Theorem 1.

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

- (i) $\mathbf{w}$ is the objective weight.
- (ii) For each i , the covariance between $\mathbf{g}$ and $w_i \mathbf{x}_i$ is 1.
- (iii) $\mathbf{g}$ is orthogonal to an affine hyperplane $L = \{\sum_i a_i w_i \mathbf{x}_i \mid \sum_i a_i = 1\}$, and $\mathbf{g}^\top \mathbf{g}/n = p$.

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

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

We also have $\mathbf{g}^\top \mathbf{g}/n = \sum_i \mathbf{g}^\top (w_i \mathbf{x}_i)/n = p$. $\square$

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

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

$$(X_1, \dots, X_p)^\top : \Omega \rightarrow \mathbb{R}^p$$

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

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

Note that it satisfies $E[G w_i X_i] = 1$ for each i .

5. Examples

We compute the objective weights for two real datasets.

Example 1. The first example is the decathlon data collected at the International Association of Athletics Federations (IAAF) World Athletics Championships held from 1991 to 2013. The data are available on the IAAF web site. The data are those of $n = 235$ athletes and $p = 10$ events, after the data with missing values and an obvious outlier in 2007 were removed. We transformed the data nonlinearly according to the IAAF scoring tables ([9], p. 24). Fig. 2(a) shows the weight of the scaled sum; this 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 for the 400 m, 100 m, and 110 mH events. As was reported in [5], it may be concluded that the present scoring rule could favor the athletes who perform well in the field events. However, the objective weight shown in Fig. 2(b) has a different property. The weight for the 1500 m is much larger than those for the other events. This suggests that the IAAF scoring method could be modified to give an even greater weight to the 1500 m event. The fact that the correlation coefficient of the OGI and the original total scores is 0.986 supports the fairness of the IAAF scoring method. Fig. 3 shows a scatter plot of these two quantities.

Example 2. We analyzed the USJudgeRatings data provided in R [16]; this dataset represents lawyers' ratings of state judges in the US Superior Court. The data are those of $n = 43$ observations on $p = 12$ numeric variates. Fig. 4 shows the relative objective weight, that is, 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 $\mathbf{x}_1^\top \mathbf{x}_2 = 0$ and $\mathbf{x}_2 = \dots = \mathbf{x}_p$, then the relative objective weight in an extended sense (see the following section) is given by $\mathbf{w} = (1, 1/\sqrt{p-1}, \dots, 1/\sqrt{p-1})^\top$.

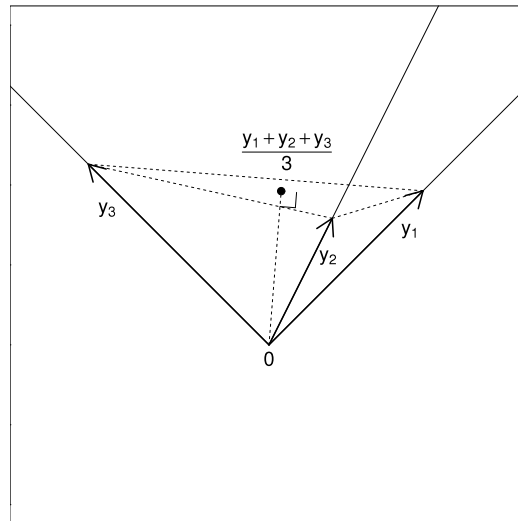


Fig. 1. Geometric interpretation of the OGI ($p = 3$), where $y_i = w_i x_i$, $1 \leq i \leq p$. In the space $\mathbb{R}^n$, the barycenter $\mathbf{g}/p$ of the y_i 's has a common inner product with each y_i . The rays from the origin indicate the direction of the x_i 's.

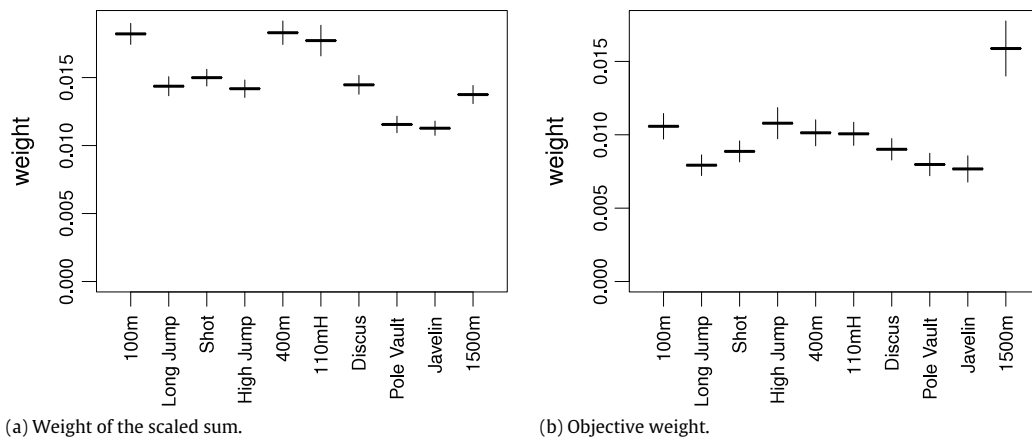


Fig. 2. Weights for the decathlon data.

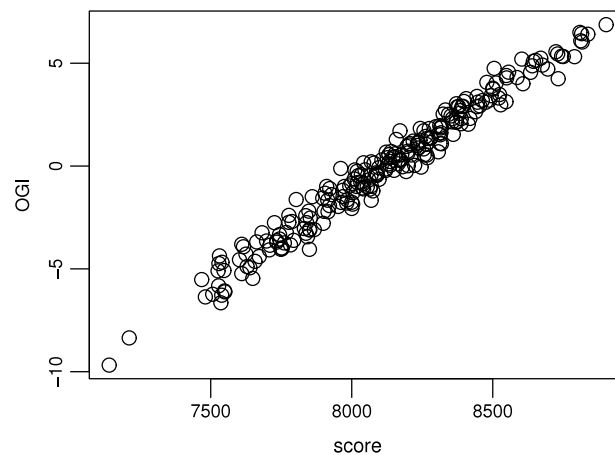


Fig. 3. Scatter plot of the OGI against the total score for the decathlon data.

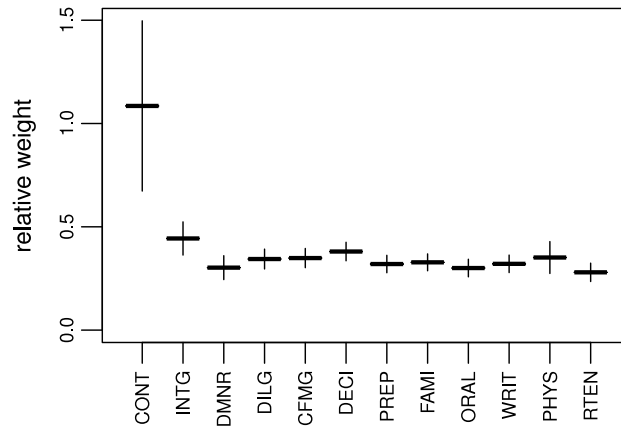


Fig. 4. Relative objective weight of the lawyers' ratings data.

6. Extension of the OGI

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

6.1. Multicollinearity

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

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

Definition 5. Let $\mathbf{S}$ be a positive semidefinite matrix. Then $\mathbf{S}$ is called *strictly copositive* if $\lambda^\top \mathbf{S} \lambda > 0$ for any $\lambda \in \mathbb{R}_{\geq 0}^p \setminus \{\mathbf{0}\}$.

Theorem 1 can be extended as follows. For later use, we also generalize the right hand side of (3). The proof is similar and is thus omitted.

Theorem 3 (Marshall and Olkin [13], Theorem 1). Let $\mathbf{S}$ be a $p \times p$ positive semidefinite matrix, and let $\mathbf{v}$ be a positive vector in $\mathbb{R}^p$. Then the equation

$$\mathbf{DSD}\mathbf{1}_p = \mathbf{v}, \quad \mathbf{D} = \text{diag}(\mathbf{w}), \quad \mathbf{w} > \mathbf{0},$$

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

If the covariance matrix of $\mathbf{X}$ is strictly copositive, then $\mathbf{g} = \mathbf{X}\mathbf{w}$ is well defined, and when $\mathbf{v} = \mathbf{1}_p$, it is the OGI in the extended sense.

For general square matrices, it is known that the copositivity condition is hard to confirm [14]. However, for positive semidefinite matrices, the methods of linear programming can be used. We prepare the following lemma.

Lemma 1. Let $\mathbf{S}$ be a positive semidefinite matrix such that $\mathbf{S} = \mathbf{X}^\top \mathbf{X}/n$ and $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p) \in \mathbb{R}^{n \times p}$. Then the following conditions are equivalent.

- (i) $\mathbf{S}$ is strictly copositive.
- (ii) $(\ker \mathbf{S}) \cap \mathbb{R}_{\geq 0}^p = \{\mathbf{0}\}$.
- (iii) $(\ker \mathbf{S}) \cap \{\lambda \in \mathbb{R}_{\geq 0}^p \mid \mathbf{1}_p^\top \lambda = 1\} = \emptyset$.
- (iv) $(\ker \mathbf{X}) \cap \mathbb{R}_{\geq 0}^p = \{\mathbf{0}\}$, where $\mathbf{X}$ is considered as a linear map from $\mathbb{R}^p$ to $\mathbb{R}^n$.
- (v) The convex cone $C = \{\sum_{i=1}^p \lambda_i \mathbf{x}_i \mid \lambda_1, \dots, \lambda_p \geq 0\}$ is proper; that is, C contains no complete line.

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

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

Condition (v) has the geometric meaning shown in Fig. 1. This means that the p vectors never balance each other.

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

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

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

$$\ker(\mathbf{X}) = \{(\lambda, -\lambda, -\mu, -\mu, -2\mu, -3\mu, \mu, \lambda, \lambda, \lambda, \lambda)^\top \mid \lambda, \mu \in \mathbb{R}\},$$

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

6.2. Subjective importance

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

7. Quantification of ordered categorical data

In this section, we define the OGI for ordered categorical variates. For ease of description, we will consider random vectors instead of a data matrix. We denote the probability space and expectation by $(\Omega, \Pr)$ and E , respectively.

7.1. Univariate case

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

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

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

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

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

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

$$\mathbf{v} = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)^\top = \frac{1}{K} \mathbf{1}_K.$$

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

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

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

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

Our quantification method is different from the optimal scaling [2]. In the optimal scaling, the weight $\mathbf{w}$ 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 a given $\sum_{k=1}^K \ln w_k$ (see the remark after Theorem 1).

Table 3
Results of examinations in 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

Example 4. Table 3 shows a dataset ([8], Table 4.4) of the results of exams in geometry and probability for 26 students. In this example, for ease of explanation, we will use the set of marks $\{1-, 2, \dots, 5\}$ instead of the set $\{0, 1, \dots, 4\}$. 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

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

7.2. Multivariate case

Consider a random vector $(X_1, \dots, X_p)$, where each variable X_i is either ordered categorical or continuous. For an ordered categorical variable X_i , without loss of generality, we can assume its range to be $\{0, 1, \dots, K_i\}$. Define $h_{ik}(x_i) = 1_{\{x_i \geq k\}} - \Pr(X_i \geq k)$ for $1 \leq k \leq K_i$. For a continuous variable X_i , let $K_i = 1$ and $h_{i1}(x) = x$.

The objective weight of the whole set of 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 $v_{ik} = 1/K_i$. 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

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

We call this procedure the *simultaneous OGI quantification*.

Again, our method is distinct from the multivariate version of the optimal scaling method proposed by Saito and Otsu [18]. In their method, the weight $\mathbf{w}$ 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 previous subsection. Denote it by $\tilde{g}_i(x_i)$. For continuous variables, put $\tilde{g}_i(x_i) = x_i$. Let $(v_1, \dots, v_p)$ be the objective weight of the quantified vector $(\tilde{g}_1(X_1), \dots, \tilde{g}_p(X_p))$. Then define the score of x_i by $\tilde{y}_i(x_i) = v_i \tilde{g}_i(x_i)$, and finally, let

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

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

Example 5. Consider again the exam data in Table 3. The data are bivariate. The results of simultaneous OGI quantification and two-stage OGI quantification are shown in Tables 4 and 5, respectively, and in Fig. 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 quantification and $\sum_{i=1}^2 \tilde{y}_i(x_i)/2$ for the two-stage one. It can be seen that the results of the two methods are almost the same.

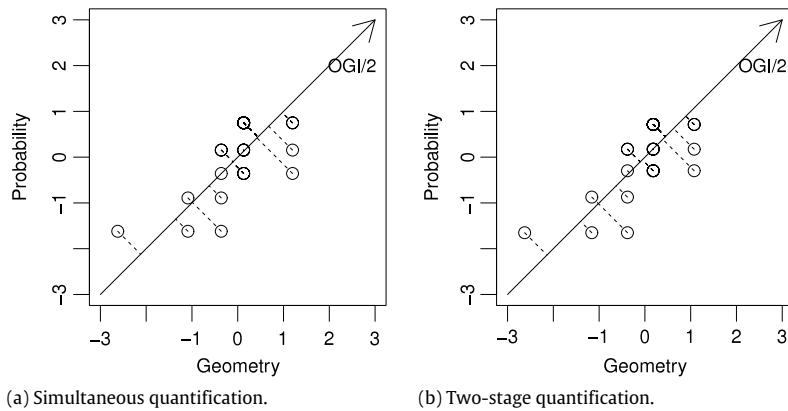


Fig. 5. Simultaneous and two-stage OGI quantifications of the exam data.

Table 4

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

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

8. Functional OGI

The (population) general index considered thus far is 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.

8.1. Univariate functional OGI

Let X be a uniformly distributed random variable on $(0, 1)$. For each $\xi \in (0, 1)$, define a nondecreasing function $h_\xi : (0, 1) \rightarrow \mathbb{R}$ 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 $h_\xi(x)$ can be 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

$$y(x) = \int_0^1 w_\xi h_\xi(x) d\xi \quad (9)$$

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

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

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

See [Appendix D](#) for more details. In terms of $y(x)$ in (9), the equation is equivalent to

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

Then, a continuous version of the OGI is defined by

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

We call G the univariate functional OGI.

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

Theorem 4. Let $y : (0, 1) \rightarrow \mathbb{R}$ 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, the univariate functional OGI has the standard normal distribution.

Proof. Eq. (10) can be written as

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

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

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

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

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

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

that is associated with the largest eigenvalue λ . One can show 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 $(X_1, \dots, X_p)$ be a random vector with uniform marginals on $(0, 1)$. Define the same function h_ξ as was defined for 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 $C_{ij}(\xi, \eta) = \Pr(X_i \leq \xi, X_j \leq \eta)$ 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 to

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

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

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

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 mean $\mathbf{0}$ and covariance matrix $\mathbf{S}$ by $\mathcal{N}(\mathbf{0}, \mathbf{S})$. Recall that Φ is the cumulative distribution function of the standard normal distribution.

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

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

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

where $\phi = \Phi'$. Eq. (13) can be 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 $\mathbf{v}$ is the objective weight of $\mathbf{S}$. $\square$

Remark 2. A unique solution is obtained if the space of $w_{\xi i}$ is restricted to the domain of a convex functional. See Appendix D 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 the functional OGI, we can define a two-step version in the same manner as the quantification method that was discussed in Section 7: compute the OGI of the transformed variables $\Phi^{-1}(X_i)$. However, this method is not theoretically interesting, since it is reduced to a finite-dimensional case.

9. Discussion

9.1. Relation to nonnegative matrix factorization

One of the reviewers suggested generalizing the OGI to a rank-two index and discussing the relation to nonnegative matrix factorization (NNMF; see e.g., [12]). The following is a possibility.

Table 6

Optimized vectors $\mathbf{t}_a$ and $\mathbf{u}_a$ for the decathlon data, where $m = 2$. The vector $\mathbf{Bt}_a$ is also shown. In the table, zero values are omitted.

	100 m	Long	Shot	High	400 m	110 mH	Discus	Pole	Javelin	1500 m
$\mathbf{t}_1$	0.53	0.64			1.00	0.88		0.85	0.10	1.00
$\mathbf{t}_2$	0.47	0.36	1.00	1.00		0.12	1.00	0.15	0.90	
$\mathbf{u}_1$	0.12	0.10			0.21	0.14		0.15		0.35
$\mathbf{u}_2$	0.08	0.10	0.23	0.23		0.06	0.23	0.05	0.20	
$\mathbf{Bt}_1$	0.55	0.50	0.11	0.12	0.80	0.62	0.09	0.62	0.22	1.36
$\mathbf{Bt}_2$	0.45	0.50	0.89	0.88	0.20	0.38	0.91	0.38	0.78	−0.36

Let $\mathbf{g} = \mathbf{XD}\mathbf{1}_p$ be the OGI of a given dataset $\mathbf{X}$. Denote the Frobenius norm of a matrix $\mathbf{A} = (A_{ij})$ by $\|\mathbf{A}\|_F = (\sum_{i,j} A_{ij}^2)^{1/2}$. Then the scaled data $\mathbf{XD}$ are approximated by a rank-one matrix $\mathbf{g}\mathbf{1}_p^\top/p$ in the sense that the unique minimizer of $\|\mathbf{XD} - \mathbf{g}\mathbf{u}^\top\|_F^2$ over $\mathbf{u} \geq \mathbf{0}_p$ is $\mathbf{u} = \mathbf{1}_p/p$. Indeed, from the definition of the OGI, one can show that

$$\frac{1}{n} \|\mathbf{XD} - \mathbf{g}\mathbf{u}^\top\|_F^2 = \text{tr}(\mathbf{B}) - 2\mathbf{u}^\top \mathbf{1}_p + p\mathbf{u}^\top \mathbf{u},$$

where $\mathbf{B} = \mathbf{DSD} = \mathbf{DX}^\top \mathbf{XD}/n$ is a bi-unit matrix.

We generalize the argument to the rank- m case, where $2 \leq m \leq p$. Let $\mathbf{D}$ and $\mathbf{B}$ be the matrices defined above. Consider the minimization problem

$$\begin{aligned} & \text{Minimize } \frac{1}{n} \left\| \mathbf{XD} - \sum_{a=1}^m (\mathbf{XD}\mathbf{t}_a)\mathbf{u}_a^\top \right\|_F^2 \\ & \text{subject to } \mathbf{t}_a \geq \mathbf{0}_p, \mathbf{u}_a \geq \mathbf{0}_p, \sum_{a=1}^m \mathbf{t}_a = \mathbf{1}_p, \quad \mathbf{t}_a^\top \mathbf{1}_p = \frac{p}{m}, \end{aligned} \quad (14)$$

where the variables are p -dimensional vectors $\mathbf{t}_a$ and $\mathbf{u}_a$, $a = 1, \dots, m$. The solution defines m new general indices $\mathbf{h}_a = \mathbf{XD}\mathbf{t}_a$. The inequality $\mathbf{t}_a \geq \mathbf{0}_p$ implies the consistency of the index $\mathbf{h}_a$ (in the nonnegative sense). The two equality conditions are assumed in order that $\sum_a \mathbf{h}_a$ be equal to $\mathbf{g}$ and have a common covariance with each $\mathbf{h}_a$ (cf. Corollary 1(ii)). Note that if $m = p$, (14) is globally minimized when $\mathbf{t}_a$ and $\mathbf{u}_a$ are the a th unit vector.

This problem is related to the NMF. The objective function (14) can be rewritten as a cost function: $\text{tr}(\mathbf{B}(\mathbf{I}_p - \mathbf{M})(\mathbf{I}_p - \mathbf{M})^\top)$ of $\mathbf{M} = \sum_{a=1}^m \mathbf{t}_a \mathbf{u}_a^\top$. As a result, the identity matrix is approximated by the rank- m nonnegative matrix $\mathbf{M}$. However, unlike the basic framework of NMF, our cost function is not written as a sum of coordinate-wise functions.

If the $\mathbf{u}_a$'s are fixed, the optimization problem is a quadratic program with respect to the $\mathbf{t}_a$'s, and vice versa. We can alternately apply a quadratic programming solver, and eventually obtain a local optimum. An investigation of the global properties of this procedure is beyond the scope of this paper.

Table 6 shows a numerical example based on the decathlon dataset (see Section 5) when $m = 2$. From the table, we can see that each of the ten events is roughly classified as either a track or a field event. The covariance consistency of the index $\mathbf{h}_2 = \mathbf{XD}\mathbf{t}_2$ fails, since $\mathbf{Bt}_2$ has a negative component. However, the corresponding component of $\mathbf{u}_2$ is zero. In other numerical experiments, we often, but not always, observed that if $(\mathbf{Bt}_a)_i < 0$, then $(\mathbf{u}_a)_i = 0$.

9.2. Characterization of consistent and covariance consistent weights

Any general index satisfying consistency and covariance consistency (Definition 1) is the OGI with a subjective importance $\mathbf{v}$ (Section 6.2), where, in general, $\mathbf{v}$ depends on $\mathbf{S}$. Indeed, if $\mathbf{w}$ satisfies the conditions $\mathbf{w} > \mathbf{0}_p$ and $\mathbf{S}\mathbf{w} > \mathbf{0}_p$, then we have the subjective importance $\mathbf{v} = \mathbf{DSD}\mathbf{1}_p$, where $\mathbf{D} = \text{diag}(\mathbf{w})$. In other words, there is a one-to-one correspondence between $\mathbf{w}$ and $\mathbf{v}$.

If $\mathbf{S}$ is fixed, then the set of consistent and covariance consistent weights $\mathbf{w}$ is a convex polyhedral cone generated by the constraints $\mathbf{w} > \mathbf{0}_p$ and $\mathbf{S}\mathbf{w} > \mathbf{0}_p$. Theorem 2 shows that the cone is not empty for every $\mathbf{S}$. It will be a challenging future problem to ascertain the properties of the cone.

9.3. High-dimensional and/or missing data

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

Datasets usually have missing values, and our method can be used, providing that the covariance matrix can be estimated appropriately.

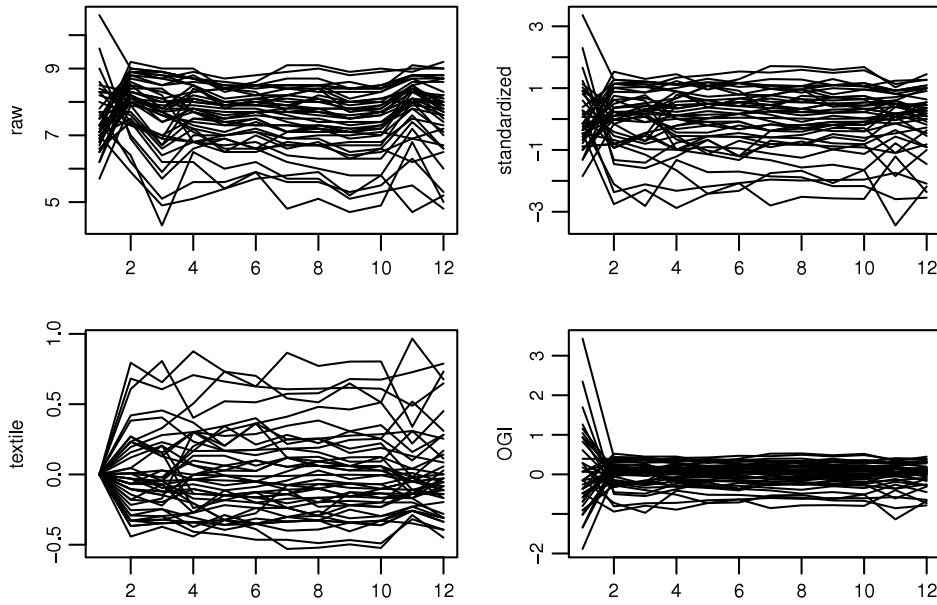


Fig. 6. Parallel coordinate plot of the USJudgeRatings data based on the raw, standardized, textile-weighted, and OGI-weighted scores.

9.4. OGI-based principal component analysis

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

$$\mathbf{Z} = \mathbf{XD}(\mathbf{I}_p - \mathbf{1}_p \mathbf{1}_p^\top / p).$$

The matrix $\mathbf{Z}$ is orthogonal to the OGI, $\mathbf{g} = \mathbf{XD}\mathbf{1}_p$, since

$$\begin{aligned} \frac{1}{n} \mathbf{g}^\top \mathbf{Z} &= \frac{1}{n} \mathbf{1}_p^\top \mathbf{D}^\top \mathbf{X}^\top \mathbf{XD}(\mathbf{I}_p - \mathbf{1}_p \mathbf{1}_p^\top / p) \\ &= \mathbf{1}_p^\top (\mathbf{I}_p - \mathbf{1}_p \mathbf{1}_p^\top / p) \\ &= \mathbf{0}_p^\top. \end{aligned}$$

Thus $\mathbf{Z}$ has information other than that contained in the OGI.

9.5. Relation to the textile plot

In [11], a weight vector $\mathbf{w}$ was used to effectively visualize high-dimensional data $\mathbf{X}$, and the weight was determined in such a way that $\mathbf{XD}$ is aligned to be as *horizontal* as possible in the parallel coordinate plot. This method is called the textile plot. If $\mathbf{X}$ has only numeric attributes without missing values and is standardized, then the weight $\mathbf{w}$ becomes the eigenvector of the sample correlation matrix associated with the largest eigenvalue ([11], Corollary 1), that is, the first principal component. We will call it the *textile weight* here.

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

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 26540013. The author is grateful to two anonymous referees for their constructive suggestions on Section 9.1, and Appendix A, which led to improvements in the presentation of this study. He also thanks Tamio Koyama and Takashi Kurokawa for their helpful comments; they pointed out the facts described in Section 9.2.

Appendix A. Mathematical properties of bi-unit matrices

Here, we briefly summarize mathematical properties of bi-unit matrices, as 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/column. In general, a principal submatrix of a bi-unit matrix is not bi-unit.

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

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

The set $\mathcal{B}_p \cap \mathcal{D}_p$ has been previously studied (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 is 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 diffeomorphic by the implicit function theorem.

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 that can be written as

$$\mathbf{B} = \begin{pmatrix} 1 & a & b & c \\ a & 1 & c & b \\ b & c & 1 & a \\ c & b & a & 1 \end{pmatrix}, \quad a + b + c = 0. \quad (\text{A.1})$$

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

One of the reviewers pointed out that the matrix [\(A.1\)](#) has the symmetry of the dihedral group D_2 , and the theory of group representations suggests the following reparameterization (e.g., [\[19\]](#), p. 50):

$$\mathbf{B} = \frac{1}{4}(\mathbf{1}\mathbf{1}^\top + \alpha \xi \xi^\top + \beta \eta \eta^\top + \gamma \zeta \zeta^\top), \quad \alpha, \beta, \gamma > 0, \quad \alpha + \beta + \gamma = 3, \quad (\text{A.2})$$

where $\xi = (1, -1, 1, -1)^\top$, $\eta = (1, 1, -1, -1)^\top$, and $\zeta = (1, -1, -1, 1)^\top$. Note that the positive definiteness of $\mathbf{B}$ is automatically satisfied. Thus, the set $\mathcal{B}_4 \cap \mathcal{C}_4$ is the interior of a two-dimensional simplex. We could not find a generalization of this notable result to $p > 4$, although we could determine that the dimension $p(p-3)/2$ of $\mathcal{B}_p \cap \mathcal{C}_p$ implies that a diagonal representation similar to [\(A.2\)](#) with fixed projections is not expected. Relevant information can be found in [\[21\]](#).

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

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

$$\mathbf{B} = \frac{1}{p}\mathbf{1}\mathbf{1}^\top + \mathbf{Q}\mathbf{A}\mathbf{Q}^\top \quad (\text{A.3})$$

with a positive definite matrix $\mathbf{A} \in \mathbb{R}^{(p-1) \times (p-1)}$. In particular, the set $\mathcal{B}_p$ is affinely isomorphic to the set of $(p-1)$ -dimensional positive definite matrices. Eq. [\(A.2\)](#) is a particular case of [\(A.3\)](#) with $\mathbf{Q} = (\xi/2, \eta/2, \zeta/2)$ and $\mathbf{A} = \text{diag}(\alpha, \beta, \gamma)$.

Appendix B. Dual data

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

The covariance between $\mathbf{y}_i$ and $\mathbf{y}_j$ is

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

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

The following lemma shows that consistency, as defined in [Definition 1](#), is equivalent to covariance consistency with respect to the dual data.

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

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

$$\frac{1}{n} \mathbf{g}^\top \mathbf{y}_i = \sum_j w_j \frac{1}{n} \mathbf{x}_j^\top \mathbf{y}_i = \sum_j w_j \delta_{ij} = w_i.$$

Hence $w_i > 0$ if and only if $\mathbf{g}^\top \mathbf{y}_i > 0$. $\square$

We also have that the OGI is invariant under the dual transformation.

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

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

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

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

Appendix C. 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 nondecreasing 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 that $y(x)$ is nondecreasing 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

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

Appendix D. Infinite-dimensional OGI

Here, we formally define the objective general index of infinitely many random variables.

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

$$G = \int_{\mathcal{E}} w_\xi X_\xi \mu(d\xi),$$

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

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

$$\int_{\mathcal{E}} w_\eta S_{\xi\eta} \mu(d\eta) = \frac{1}{w_\xi}, \quad \mu\text{-almost every } \xi. \quad (\text{D.1})$$

This is the stationary condition of a convex functional:

$$\Psi[w] = \frac{1}{2} \int_{\mathcal{E}} \int_{\mathcal{E}} w_\xi w_\eta S_{\xi\eta} \mu(d\xi) \mu(d\eta) - \int_{\mathcal{E}} (\ln w_\xi) \mu(d\xi).$$

A uniqueness result is given in the following lemma; existence will not be discussed here.

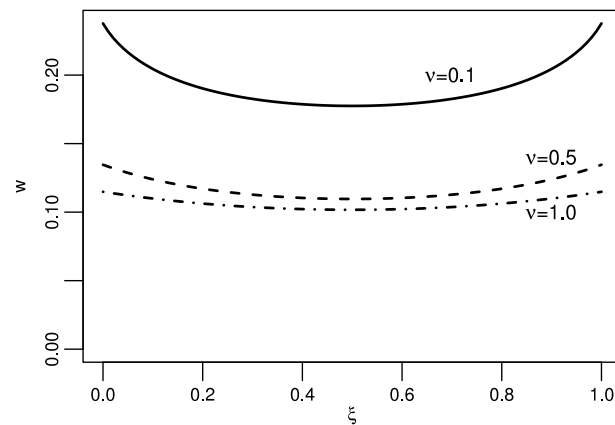


Fig. D.7. Objective weight function w_ξ of the Matérn class, where $\mathcal{E} = [0, 1]$ and $\alpha = \phi = 1$.

Lemma 5. The solution of (D.1) 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

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

unless δ is zero μ -almost everywhere. Hence, the solution of (D.1) over $\mathcal{W}$ is the unique minimal point of Ψ . $\square$

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

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

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

is recommended (e.g., [20]), where ϕ , v , and α are positive parameters, and K_v denotes the modified Bessel function of the second kind. Fig. D.7 shows numerically evaluated weight functions w_ξ for several values of v , where $\mathcal{E} = [0, 1]$ and $\alpha = \phi = 1$. The interval $\mathcal{E} = [0, 1]$ is approximated by $n = 100$ grid points. There is an *edge effect* in that the weight increases as ξ approaches the boundary.

References

- [1] R.J. Baker, Selection indexes without economic weights for animal breeding, *Can. J. Animal Sci.* 54 (1) (1974) 1–8.
- [2] R.A. Bradley, S.K. Katti, I.J. Coons, Optimal scaling for ordered categories, *Psychometrika* 27 (4) (1962) 355–374.
- [3] J.P.R. Christensen, P. Fischer, Positive definite doubly stochastic matrices and extreme points, *Linear Algebra Appl.* 82 (1986) 123–132.
- [4] V. Chvátal, *Linear Programming*, W. H. Freeman and Company, 1983.
- [5] T.F. Cox, R.T. Dunn, An analysis of decathlon data, *J. Roy. Statist. Soc. Ser. D* 51 (2) (2002) 179–187.
- [6] D.D. Dill, M. Soo, Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems, *Higher Educ.* 49 (4) (2005) 495–533.
- [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) (1963) 85–97.
- [8] T. Hibi (Ed.), *Gröbner Bases: Statistics and Software Systems*, Springer, 2013.
- [9] IAAF Council (Ed.), *IAAF Scoring Tables for Combined Events*, IAAF, 2001.
- [10] B. Kalantari, A theorem of the alternative for multihomogeneous functions and its relationship to diagonal scaling of matrices, *Linear Algebra Appl.* 236 (1996) 1–24.
- [11] N. Kumasaka, R. Shibata, High-dimensional data visualisation: The textile plot, *Comput. Statist. Data Anal.* 52 (2008) 3616–3644.

- [12] D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, in: *Advances in Neural Information Processing Systems (NIPS 2000)*, Vol. 13, 2000, pp. 556–562.
- [13] A.W. Marshall, I. Olkin, Scaling of matrices to achieve specified row and column sums, *Numer. Math.* 12 (1968) 83–90.
- [14] K.G. Murty, S.N. Kabadi, Some NP-complete problems in quadratic and nonlinear programming, *Math. Program.* 39 (1987) 117–129.
- [15] D.P. O’Leary, Scaling symmetric positive definite matrices to prescribed row sums, *Linear Algebra Appl.* 370 (2003) 185–191.
- [16] R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- [17] R.T. Rockafeller, *Convex Analysis*, Princeton University Press, 1972.
- [18] T. Saito, T. Otsu, A method of optimal scaling for multivariate ordinal data and its extensions, *Psychometrika* 53 (1) (1988) 5–25.
- [19] J.-P. Serre, *Linear Representations of Finite Groups*, Springer-Verlag, 1977.
- [20] M.L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, 1999.
- [21] M.A.G. Viana, V. Lakshminarayanan, *Dihedral Fourier Analysis*, Springer, 2013.