Article

A Quantile General Index Derived from the Maximum Entropy Principle

Tomonari Sei

Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan; sei@mist.i.u-tokyo.ac.jp

Abstract: We propose a linear separation method of multivariate quantitative data in such a way that the average of each variable in the positive group is larger than that of the negative group. Here, the coefficients of the separating hyperplane are restricted to be positive. Our method is derived from the maximum entropy principle. The composite score obtained as a result is called the quantile general index. The method is applied to the problem of determining the top 10 countries in the world based on the 17 scores of the Sustainable Development Goals (SDGs).

Keywords: check loss function; general index; SDGs; unsupervised learning

1. Introduction

Consider a data matrix, each row of which corresponds to a case, and each column represents a variable. Suppose that every variable has the meaning that a larger value indicates better. For example, ref. [1] investigated the efforts of countries to attain the SDGs (Sustainable Development Goals) and reported the 17 SDG scores for each country. The scores ranged from 0 to 100. In the report, a ranking of 163 countries on the basis of the average of the 17 scores was provided. We call such a procedure of ranking the simple sum method.

However, we sometimes find a paradoxical phenomenon in the simple sum method, in that a particular variable of a higher-score group is less than that of the remaining group. See Table 1 for illustration, where we separate the SDGs data into two groups: the 10 top countries on the basis of the simple sum method and the remaining 153 countries. The average values of each variable for the two groups are compared. On almost all the variables, the 10 top countries have larger averages than the remaining countries, as expected. However, there are reverse relations in the SDGs 12 and 13. The 10 top countries have an average value lower than the remaining countries on the two goals.

In this paper, we propose a linear weighting method that can avoid the reversal relation (in a random-decision sense). The higher-score group separated by the linear weight has average values greater than the remaining group with respect to all the variables. The idea behind the method is the objective general index (OGI; [2]), which is constructed to have a positive correlation with all the variables. The purpose of the OGI is the ranking and not the separation. The OGI is interpreted as a minimization problem of a free energy functional [3,4], which is the sum of the negative entropy and an internal energy functional. This interpretation also works in the current setting; see Section 2.

The problem of determining weights is unsupervised in the sense that no one knows the correct weights and classifications, which has been consistently discussed (e.g., [5,6]). There are many weighting methods for such purposes. Among them, the principal component analysis (PCA) is widely used. The PCA, however, does not always give positive weights; so, some modifications are necessary. It is known that a nonnegative version of the principal component analysis is a nonconvex and NP-hard optimization problem [7]. Another approach is the factor analysis, where a factor model refers to a set of multivariate distributions that have common latent factors (e.g., [8]). Although the factor analysis is

quite flexible, it needs additional assumptions such as variance–covariance structures and often does not have a unique solution. In contrast, the quantile general index we propose is reduced to a convex optimization problem and is essentially unique as we will demonstrate. The Hirsch index (or h-index) is widely used for the evaluation of scientific research reports [9], and its further application has been recently investigated by [10]. We numerically compare our method with the h-index in Section 5.

Table 1. The average values of the SDG scores for the 10 top countries (Finland, Denmark, Sweden, Norway, Austria, Germany, France, Switzerland, Ireland and Estonia) and the remaining 153 countries. The values with the reversal relations are marked by asterisks.

SDGs	Average of the 10 Top Countries	Average of the Remaining Countries
1 (no poverty)	99.6	73.8
2 (zero hunger)	68.5	58.6
3 (good health and well-being)	94.1	68.0
4 (quality education)	98.2	74.9
5 (gender equality)	84.8	60.1
6 (clean water and sanitation)	89.4	66.2
7 (affordable and clean energy)	83.8	64.8
8 (descent work and economic growth)	85.0	66.3
9 (industry, innovation and infrastructure)	91.8	43.2
10 (reduced inequalities)	92.3	59.7
11 (sustainable cities and communities)	92.8	68.8
12 (responsible consumption and production)	60.3 *	85.6 *
13 (climate action)	54.7 *	81.9 *
14 (life below water)	71.4	64.3
15 (life on land)	80.4	64.8
16 (peace, justice and strong institutions)	87.8	65.2
17 (partnerships for the goals)	71.8	58.5

Source: The Sustainable Development Report 2022 [1].

The name of the quantile general index comes from the quantile regression developed by [11]. Indeed, the objective function we use is similar to those of the quantile regression; see the explicit form in Section 3. The essential difference here is that our problem is unsupervised, whereas the regression problems are supervised.

The general indices determine an ordering of the data. The problem of well ordering multivariate data was discussed by [12], where methods of ordering were classified into four categories: marginal ordering, reduced ordering, partial ordering, and conditional ordering. Our method is considered as marginal ordering on the weighted sum.

The paper is organized as follows. In Section 2, we define the quantile general index for continuous distributions and show that it is characterized by the maximum entropy principle. In Section 3, a finite-sample counterpart of the quantile general index is derived. In Section 4, a practical method that avoids the ambiguity of data lying on the separating hyperplane is proposed. We apply the method to the SDG data in Section 5, and we conclude in Section 6.

2. Quantile General Index for Continuous Distributions

The quantile general index for continuous probability distributions is defined first. The assumption of continuity avoids the difficulty caused by the non-smoothness of the objective function. The sample counterpart of the index is constructed in the subsequent section.

Suppose that we have a random vector $x = (x_1, \ldots, x_d)^\top$ following a probability distribution P on $\mathbb{R}^d$, where $\top$ denotes the vector transpose. We assume that P has the probability density function $p(x)$ so that $P(A) = \int_A p(x)\mathrm{d}x$ for an event $A \subset \mathbb{R}^d$. For

given $h : \mathbb{R}^d \to \mathbb{R}$, we denote the expectation of a random variable $h(x)$ by $E[h(x)] = \int p(x)h(x)dx$ and the conditional expectation of $h(x)$ given an event A by

$$E[h(x)|A] = \frac{\int_A p(x)h(x)dx}{\int_A p(x)dx}.$$

We deal with a class of general indices

$$g(x) = g(x; w, c) = \sum_{i=1}^{d} w_i x_i - c \tag{1}$$

of x, where $w = (w_1, \ldots, w_d)^\top \in \mathbb{R}_+^d$ and $c \in \mathbb{R}$ are called the weight vector and the threshold, respectively. Here $\mathbb{R}_+$ denotes the set of positive numbers. The quantities w and c may depend on the underlying distribution P but do not depend on x itself.

For a given g of the form (1), the half spaces separated by the hyperplane $g(x) = 0$ are denoted by

$$H_g^+ = \{x \mid g(x) > 0\} \quad \text{and} \quad H_g^- = \{x \mid g(x) < 0\}.$$

The quantile general index is defined as follows.

Definition 1. *A general index $g(x) = \sum_i w_i x_i - c$ is called* the quantile general index *of x if it satisfies the following two equations:*

$$P(H_g^+) = \alpha \tag{2}$$

and

$$E[w_i x_i \mid H_g^+] - E[w_i x_i \mid H_g^-] = 1, \quad i = 1, \ldots, d. \tag{3}$$

The weight w is called the optimal weight.

Let us call H_g^+ and H_g^- the positive and negative group, respectively. Equation (2) means that the fraction of the positive group is α. The threshold c is the upper α-quantile of the weighted sum $w^\top x$ because $P(w^\top x > c) = \alpha$ by (2). We call α *the acceptance ratio*. Equation (3) implies that the average of each variable x_i on the positive group is greater than that on the negative group. Therefore, the reversal relation observed in Table 1 does not occur if we adopt the quantile general index.

We now state the existence and uniqueness theorem of the quantile general index. For $0 < \alpha < 1$, we define the "check" loss function $\ell_\alpha : \mathbb{R} \to \mathbb{R}$ by

$$\ell_\alpha(u) = \frac{u^-}{1-\alpha} + \frac{u^+}{\alpha}, \tag{4}$$

where $u^+ = \max(u, 0)$ and $u^- = \max(-u, 0)$ are the positive and negative parts of u, respectively. See Figure 1 for the graph of ℓ_α. The function ℓ_α is used in quantile regression [13]. The derivative of $\ell_\alpha(u)$ for $u \neq 0$ is

$$\ell_\alpha'(u) = -\frac{1}{1-\alpha}\mathbb{I}_{\{u<0\}} + \frac{1}{\alpha}\mathbb{I}_{\{u>0\}},$$

where $\mathbb{I}_{\{u>0\}}$ is 1 if $u > 0$ and 0 otherwise. The subgradient (e.g., [14]) at $u = 0$ can be also defined but is not used here.
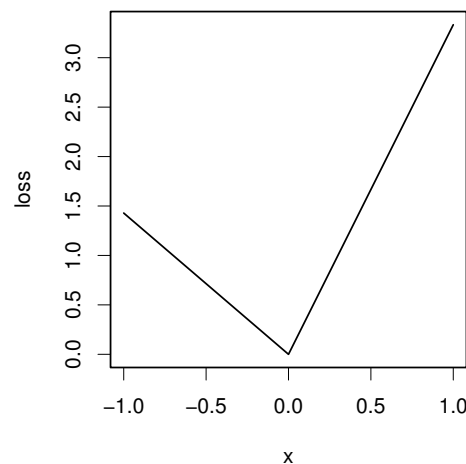
Figure 1. The check-loss function for $\alpha = 0.3$.

We define a convex function $F : \mathbb{R}_+^d \times \mathbb{R} \to \mathbb{R}$ by

$$F(\boldsymbol{w}, c) = -\sum_{i=1}^{d} \log w_i + \mathrm{E}\left[\ell_\alpha \left(\sum_{j=1}^{d} w_j x_j - c \right) \right]. \tag{5}$$

The main theorem is stated as follows.

Theorem 1. *Let $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$ be a random vector with a probability density function on $\mathbb{R}^d$ and assume that $\mathrm{E}[x_i]$ exists for each i. Let $0 < \alpha < 1$. Then, the function F in (5) admits a minimizer $(\boldsymbol{w}, c) \in \mathbb{R}_+^d \times \mathbb{R}$. The optimal $\boldsymbol{w}$ is unique, whereas c may not be unique. Furthermore, the general index $g(\boldsymbol{x}) = \sum_i w_i x_i - c$ based on the minimizer $(\boldsymbol{w}, c)$ of F satisfies the conditions (2) and (3) of the quantile general index.*

Proof. The proof of existence and uniqueness is given in Appendix A. We prove that the stationary condition of F is given by (2) and (3). The partial derivatives of F with respect to c and w_i are

$$\frac{\partial F}{\partial c} = -\mathrm{E}[\ell_\alpha'(g(\boldsymbol{x}))]$$

$$= \frac{1}{1-\alpha} \mathrm{P}(H_g^-) - \frac{1}{\alpha} \mathrm{P}(H_g^+)$$

and

$$\frac{\partial F}{\partial w_i} = -\frac{1}{w_i} + \mathrm{E}[x_i l_\alpha'(g(\boldsymbol{x}))]$$

$$= -\frac{1}{w_i} - \frac{1}{1-\alpha} \mathrm{E}[x_i \mathbb{I}_{\{g(\boldsymbol{x})<0\}}] + \frac{1}{\alpha} \mathrm{E}[x_i \mathbb{I}_{\{g(\boldsymbol{x})>0\}}].$$

Note that $\mathrm{P}(H_g^-) = 1 - \mathrm{P}(H_g^+)$, since $\mathrm{P}(g(\boldsymbol{x}) = 0) = 0$ from the assumption that $\boldsymbol{x}$ has a continuous distribution. Then, the equations $\partial F / \partial c = 0$ and $\partial F / \partial w_i = 0$ $(i = 1, \ldots, d)$ are equivalent to (2) and (3). $\square$

Example 1. *Let x_1 and x_2 be independent and identically distributed according to a continuous distribution. By the uniqueness of the optimal weight and symmetry, we have $w_1 = w_2(=w)$. We denote the upper α-quantile of $x_1 + x_2$ by y_α. Then, we have $c/w = y_\alpha$ from (2) and*

$$w = \frac{1}{\mathrm{E}[x_1 | x_1 + x_2 > y_\alpha] - \mathrm{E}[x_1 | x_1 + x_2 < y_\alpha]}$$

from (3). *For example, if* x_i *has the standard normal distribution and* $\alpha = 1/2$, *then* $c = 0$, *and* $w = \sqrt{\pi}/2$.

The quantile general index is derived from the maximum entropy principle in line with [4]. The entropy of a density function p is defined by

$$S(p) = -\int p(\boldsymbol{x}) \log p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.$$

Consider a class of transformations $T : \mathbb{R}^d \to \mathbb{R}^d$ of the form

$$T(\boldsymbol{x}) = (w_1 x_1 - c_1, \ldots, w_d x_d - c_d), \quad (w_i, c_i) \in \mathbb{R}_+ \times \mathbb{R}.$$

The push-forward density of p by T is defined by

$$\begin{aligned}(T_\sharp p)(\boldsymbol{x}) &= p(T^{-1}(\boldsymbol{x}))|(T^{-1})'(\boldsymbol{x})| \\ &= p\left(\frac{x_1 + c_1}{w_1}, \ldots, \frac{x_d + c_d}{w_d}\right)\frac{1}{w_1 \cdots w_d}.\end{aligned}$$

This is the distribution of $T(\boldsymbol{x})$ when the random variable $\boldsymbol{x}$ follows the distribution P. It is shown that the entropy of the push-forward density is

$$S(T_\sharp p) = S(p) + \sum_{i=1}^{d} \log w_i.$$

We also define an internal energy by

$$U(p) = \int p(\boldsymbol{x})\ell_\alpha(\textstyle\sum_i x_i)\mathrm{d}\boldsymbol{x},$$

where ℓ_α is the check loss function in (4). The following theorem characterizes the quantile general index in terms of entropy. The proof is straightforward.

Theorem 2. *The minimization problem of* (5) *is equivalent to*

$$\begin{aligned}&\text{Minimize } U(T_\sharp p) - S(T_\sharp p) \\ &\text{subject to } T(\boldsymbol{x}) = (w_1 x_1 - c_1, \ldots, w_d x_d - c_d), \\ &\qquad\qquad (w_i, c_i), \ldots, (w_d, c_d) \in \mathbb{R}_+ \times \mathbb{R}.\end{aligned}$$

The threshold c *in* (5) *is given by* $c = \sum_{i=1}^{d} c_i$.

3. Quantile General Index for Finite Samples

The quantile general index defined in the preceding section is valid only for continuous distributions. It is useful to define the index also for finite samples. Let $\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(n)} \in \mathbb{R}^d$ be a sample of size n. We denote the i-th coordinate of $\boldsymbol{x}_{(t)}$ by x_{ti}. We deal with a class of general indices $g_t = \sum_{i=1}^{d} w_i x_{ti} - c$, where $(\boldsymbol{w}, c) \in \mathbb{R}_+^d \times \mathbb{R}$ may depend on the whole sample $\{\boldsymbol{x}_{(t)}\}_{t=1}^{n}$ but does not depend on t.

The empirical counterpart of the objective function (5) is

$$F(\boldsymbol{w}, c) = -\sum_{i=1}^{d} \log w_i + \frac{1}{n}\sum_{t=1}^{n} \ell_\alpha\left(\sum_{j=1}^{d} w_j x_{tj} - c\right) \tag{6}$$

for $(\boldsymbol{w}, c) \in \mathbb{R}_+^d \times \mathbb{R}$.

Definition 2. *A general index* $g_t = \sum_i w_i x_{ti} - c$ *of* $\boldsymbol{x}_{(t)}$ *for* $t = 1, \ldots, n$ *is called the quantile general index if* $(\boldsymbol{w}, c)$ *minimizes the function* (6).

Remark 1. *As described in Section 1, the objective function (6) is similar to that of the quantile regression defined by*

$$\frac{1}{n}\sum_{t=1}^{n}\ell_{\alpha}\left(y_t - \sum_{j=1}^{d} w_j x_{tj}\right),$$

where y_t is a response variable and $w_1, \ldots, w_d$ are regression coefficients. See [13] for a comprehensive study of the quantile regression.

The following theorem is proved in a similar way to Theorem 1. See Appendix A.

Theorem 3. *Suppose that there is no hyperplane of $\mathbb{R}^d$ that contains all $x_{(t)}$. Then, the objective function F in (6) admits a minimizer (w, c). The weight vector w is unique. The threshold c is unique if $n\alpha$ is not an integer.*

Each case $x_{(t)}$ is classified into positive and negative groups according to $g_t > 0$ and $g_t < 0$, respectively. If the case $g_t = 0$ does not exist, then the fraction of the positive (resp. negative) group is α (resp. $1 - \alpha$), and the conditional expectation of x_{ti} on the positive group is greater than that on the negative group. This is the desired dominance relation.

However, it is not always possible to classify the data into positive and negative groups, because g_t may become 0 in some cases. Furthermore, the minimization of $F(w, c)$ is not straightforward, since the function is not differentiable. In order to avoid these issues, we modify the method in Section 4.

For illustration, we calculate the quantile general index for the following examples.

Example 2. *Consider the bivariate data*

$$x_{(1)} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad x_{(2)} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad x_{(3)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad x_{(4)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

of sample size 4. Let the acceptance ratio be $\alpha = 1/2$. In this data, any set of three points is not on a straight line. Therefore, there exists the quantile general index by Theorem 3. We show that the solution is $w_1 = 2/3$, $w_2 = 4/3$, and $c = 8/3$. We consider three disjoint subsets of $\mathbb{R}_+^2$:

$$A = \{w \mid w_2 < 2w_1\}, \quad B = \{w \mid w_2 > 2w_1\}, \quad C = \{w \mid w_2 = 2w_1\}.$$

Let $w \in A$. Then, we have

$$w^\top x_{(1)} > w^\top x_{(2)} > w^\top x_{(3)} > w^\top x_{(4)}$$

Hence, the optimal c is between $w^\top x_{(2)}$ and $w^\top x_{(3)}$, since c is the upper 1/2-quantile of $\{w^\top x_{(t)}\}$. For such c, the objective function (6) becomes

$$F(w_1, w_2, c) = -\log w_1 - \log w_2 + \frac{1}{4}\left(\frac{w^\top x_{(1)}}{1/2} + \frac{w^\top x_{(2)}}{1/2} - \frac{w^\top x_{(3)}}{1/2} - \frac{w^\top x_{(4)}}{1/2}\right)$$

$$= (-\log w_1 + 2w_1) + (-\log w_2 + w_2/2).$$

If F is minimized at some $w \in A$, then it must be $w_1 = 1/2$ and $w_2 = 2$ by the stationary condition, but this point does not belong to A. Hence, the optimal point does not exist in A.

If $w \in B$, then we have

$$w^\top x_{(1)} > w^\top x_{(3)} > w^\top x_{(2)} > w^\top x_{(4)}$$

and the objective function is

$$F(w_1, w_2, c) = (-\log w_1) + (-\log w_2 + 3w_2/2),$$

where $w^\top x_{(2)} \le c \le w^\top x_{(3)}$. It is shown again that the optimal point does not exist in B.

Therefore, the optimal point should be located in C, the boundary of A and B. The objective function is

$$F(w_1, 2w_1, c) = -\log 2 - 2\log w_1 + 3w_1,$$

where $c = w^\top x_{(2)} = w^\top x_{(3)} = 4w_1$. The optimal solution is $w_1 = 2/3$, $w_2 = 4/3$, and $c = 8/3$. The quantile general index is given by

$$\begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{pmatrix} = \begin{pmatrix} 2 & 2 & -1 \\ 2 & 1 & -1 \\ 0 & 2 & -1 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 2/3 \\ 4/3 \\ 8/3 \end{pmatrix} = \begin{pmatrix} 4/3 \\ 0 \\ 0 \\ -8/3 \end{pmatrix}.$$

The index does not provide a separation of the data because $g_2 = g_3 = 0$. In this case, however, a group $\{x_{(1)}, x_{(2)}\}$ dominates $\{x_{(3)}, x_{(4)}\}$ in the sense that the difference of averages

$$\frac{1}{2}(x_{(1)} + x_{(2)}) - \frac{1}{2}(x_{(3)} + x_{(4)}) = \begin{pmatrix} 2 \\ 1/2 \end{pmatrix}$$

is a positive vector.

If we set the acceptance ratio to $\alpha = 1/4$, then it is proved in a similar way that the optimal w is $w_1 = 3/4$ and $w_2 = 1$. In this case, c is not unique: $5/2 \le c \le 7/2$. The quantile general index is

$$\begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{pmatrix} = \begin{pmatrix} 2 & 2 & -1 \\ 2 & 1 & -1 \\ 0 & 2 & -1 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 3/4 \\ 1 \\ c \end{pmatrix} = \begin{pmatrix} 7/2 - c \\ 5/2 - c \\ 2 - c \\ -c \end{pmatrix}.$$

Therefore, $g_1 > 0$ and $g_2, g_3, g_4 < 0$ as long as $5/2 < c < 7/2$. The separation provides a dominance relation:

$$x_{(1)} - \frac{1}{3}(x_{(2)} + x_{(3)} + x_{(4)}) = \begin{pmatrix} 4/3 \\ 1 \end{pmatrix}.$$

Example 3. *Consider the bivariate data*

$$x_{(1)} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \quad x_{(2)} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \quad x_{(3)} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \quad x_{(4)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

of sample size 4. Let $\alpha = 1/2$. In a similar manner to the preceding example, the optimal parameters are shown to be $w = (1,1)^\top$ and $c = 4$. The quantile general index is

$$\begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{pmatrix} = \begin{pmatrix} 4 & 0 & -1 \\ 2 & 4 & -1 \\ 1 & 3 & -1 \\ 0 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ -2 \end{pmatrix}.$$

In this case, no separation of the sample into two groups provides a dominance relation. Indeed, all the possible combinations are

$$\frac{1}{2}(x_{(1)} + x_{(2)}) - \frac{1}{2}(x_{(3)} + x_{(4)}) = \begin{pmatrix} 5/2 \\ -1/2 \end{pmatrix},$$

$$\frac{1}{2}(x_{(1)} + x_{(3)}) - \frac{1}{2}(x_{(2)} + x_{(4)}) = \begin{pmatrix} 3/2 \\ -3/2 \end{pmatrix},$$

$$\frac{1}{2}(x_{(1)} + x_{(4)}) - \frac{1}{2}(x_{(2)} + x_{(3)}) = \begin{pmatrix} 1/2 \\ -5/2 \end{pmatrix},$$

which are not positive.

4. Practical Implementation

The quantile general index defined in the preceding section has the following two drawbacks.

- The minimization is not straightforward since F is not differentiable.
- The cases with $g_t = 0$ are not assigned to positive or negative groups.

To overcome these issues, we approximate F as

$$F_\varepsilon(\boldsymbol{w}, c) = - \sum_{i=1}^{d} \log w_i + \frac{1}{n} \sum_{t=1}^{n} \ell_{\alpha,\varepsilon}\left(\sum_i x_{ti} w_i - c \right) \tag{7}$$

where ε is a positive constant, and the function $\ell_{\alpha,\varepsilon} : \mathbb{R} \to \mathbb{R}$ is defined by
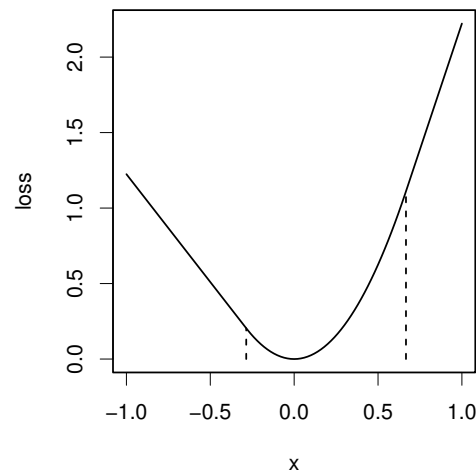
$$\begin{aligned}
\ell_{\alpha,\varepsilon}(u) &= \min_{z \in \mathbb{R}} \left\{ \ell_\alpha(z) + \frac{1}{2\varepsilon}|z - u|^2 \right\} \\
&= \begin{cases}
u/\alpha - \varepsilon/(2\alpha^2) & \text{if } u \geq \varepsilon/\alpha, \\
u^2/(2\varepsilon) & \text{if } -\varepsilon/(1-\alpha) < u < \varepsilon/\alpha, \\
-u/(1-\alpha) - \varepsilon/(2(1-\alpha)^2) & \text{if } u \leq -\varepsilon/(1-\alpha).
\end{cases}
\end{aligned} \tag{8}$$

The function is called the Moreau envelope of ℓ_α. See Figure 2 for the graph of $\ell_{\alpha,\varepsilon}$. It is shown that $l_{\alpha,\varepsilon}$ uniformly converges to ℓ_α, as $\varepsilon \to 0$.

The derivative of $\ell_{\alpha,\varepsilon}$ is piecewise linear:

$$\ell'_{\alpha,\varepsilon}(u) = \begin{cases}
1/\alpha & \text{if } u \geq \varepsilon/\alpha, \\
u/\varepsilon & \text{if } -\varepsilon/(1-\alpha) < u < \varepsilon/\alpha, \\
-1/(1-\alpha) & \text{if } u \leq -\varepsilon/(1-\alpha).
\end{cases}$$

In particular, $\ell_{\alpha,\varepsilon}$ is continuously differentiable unlike ℓ_α.

Figure 2. The Moreau envelope of the check-loss function for $\alpha = 0.3$ and $\varepsilon = 0.2$. The two vertical lines are $u = \varepsilon/\alpha$ and $u = -\varepsilon/(1-\alpha)$, respectively.

Definition 3. *A general index* $g_t = \sum_i w_i x_{ti} - c$ *is called the quantile general index within tolerance ε if $(\boldsymbol{w}, c)$ minimizes $F_\varepsilon(\boldsymbol{w}, c)$.*

The gradient of F_ε is

$$\frac{\partial F_\varepsilon}{\partial c} = -\frac{1}{n}\sum_{t=1}^{n}\left(\frac{J_t}{\alpha} - \frac{1 - J_t}{1 - \alpha}\right),$$

$$\frac{\partial F_\varepsilon}{\partial w_i} = -\frac{1}{w_i} + \frac{1}{n}\sum_{t=1}^{n}\left(\frac{J_t}{\alpha} - \frac{1 - J_t}{1 - \alpha}\right)x_{ti},$$

where

$$J_t = \begin{cases} 1 & \text{if } g_t \geq \varepsilon/\alpha, \\ \alpha(1-\alpha)(g_t/\varepsilon + 1/(1-\alpha)) & \text{if } -\varepsilon/(1-\alpha) < g_t < \varepsilon/\alpha, \\ 0 & \text{if } g_t \leq -\varepsilon/(1-\alpha). \end{cases} \tag{9}$$

These formulas prove the second part of the following theorem. See Appendix A for the proof of the first part.

Theorem 4. *Suppose that there is no hyperplane of $\mathbb{R}^d$ that contains all $\mathbf{x}_{(t)}$. Then, the objective function F_ε in (8) admits a minimizer $(\mathbf{w}, c)$, and the optimal weight vector $\mathbf{w}$ is unique. Furthermore, the optimal $(\mathbf{w}, c)$ and $J_t \in [0, 1]$ defined in (9) satisfy*

$$\frac{1}{n}\sum_{t=1}^{n}J_t = \alpha \tag{10}$$

and

$$\frac{1}{n\alpha}\sum_{t=1}^{n}w_i x_{ti} J_t - \frac{1}{n(1-\alpha)}\sum_{t=1}^{n}w_i x_{ti}(1 - J_t) = 1. \tag{11}$$

The Equations (10) and (11) correspond to (2) and (3) for continuous distributions. The quantity J_t is interpreted as the probability of assigning the case $\mathbf{x}_{(t)}$ to the positive group. We call J_t *the optimal random decision*. If the general index g_t is greater than the threshold ε/α, the case t is definitely assigned to the positive group because $J_t = 1$. Similarly, if the general index is less than $-\varepsilon/(1-\alpha)$, it is definitely assigned to the negative group.

For numerical computation, we used a general-purpose optimization solver optim in R [15] with the L-BFGS method.

Example 4 (Continuation of Example 2). *Consider four cases*

$$\mathbf{x}_{(1)} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \mathbf{x}_{(2)} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \mathbf{x}_{(3)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad \mathbf{x}_{(4)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Let $\alpha = 1/2$ and $\varepsilon = 0.001$. The optimal $\mathbf{w}$ and c are numerically obtained as $\mathbf{w} = (0.667, 1.332)^\top$ and $c = 2.666$. The quantile general index is $(g_1, g_2, g_3, g_4) = (1.333, 0.001, -0.001, -2.666)$, and the optimal random decision is $(J_1, J_2, J_3, J_4) = (1, 0.749, 0.250, 0)$, so that the optimal separation will be $\{\mathbf{x}_{(1)}, \mathbf{x}_{(2)}\}$ and $\{\mathbf{x}_{(3)}, \mathbf{x}_{(4)}\}$. This separation happens to satisfy the dominance relation as we have seen in Example 2.

Example 5 (Continuation of Example 3). *Consider four cases*

$$\mathbf{x}_{(1)} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \quad \mathbf{x}_{(2)} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \quad \mathbf{x}_{(3)} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \quad \mathbf{x}_{(4)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

Let $\alpha = 1/2$ and $\varepsilon = 0.001$. The optimal $\mathbf{w}$ and c are numerically obtained as $\mathbf{w} = (1, 1)^\top$ and $c = 4$. The quantile general index is $(g_1, g_2, g_3, g_4) = (0, 2, 0, -2)$, and the optimal random

decision is $(J_1, J_2, J_3, J_4) = (0.5, 1, 0.5, 0)$. *In this case, we cannot decide which of* $x_{(1)}$ *and* $x_{(3)}$ *has to be assigned to the positive group. This result is consistent with the discussion in Example 3.*

5. Application to the SDGs Index

We finally compute the quantile general indices of the SDGs data provided by [1], as introduced in Section 1. According to [1], countries with a fraction of missing values greater than 20% were removed from the data and then the missing values were imputed by regional averages. We applied the quantile general index with the acceptance ratio $\alpha = 10/163$ and tolerance $\varepsilon = 0.001$. The result is summarized in Table 2. The optimal weight w is shown in the second column of the table. The threshold was $c = 178.2$. The other columns of Table 2 show the average of each variable in the 10 top countries and the remaining countries, respectively. In contrast to Table 1, we do not observe the reversal relation. Table 3 shows the general index g_t and the optimal random decision J_t of the 10 top countries.

Table 2. For the SDGs data, the optimal weight w_i, the average x_i^+ of each score in the 10 top countries determined from the quantile general index (Cuba, Romania, Finland, Kyrgyz Republic, Ukraine, Chile, Poland, Georgia, Vietnam, Hungary), the average x_i^- on the remaining 153 countries, and the scaled differences $w_i(x_i^+ - x_i^-)$ are shown.

SDGs i	Weights w_i	Average of the 10 Top Countries x_i^+	Average of the Remaining Countries x_i^-	Scaled Difference $w_i(x_i^+ - x_i^-)$
1	0.049	94.9	74.1	1.02
2	0.136	65.8	58.8	0.95
3	0.079	81.3	68.8	0.99
4	0.061	91.4	75.3	0.98
5	0.122	69.2	61.1	0.99
6	0.071	81.3	66.8	1.03
7	0.091	76.0	65.3	0.97
8	0.098	77.4	66.8	1.04
9	0.082	58.4	45.4	1.07
10	0.070	75.5	60.8	1.03
11	0.088	82.0	69.5	1.10
12	0.490	85.9	83.9	0.98
13	0.412	82.4	80.1	0.95
14	0.126	71.5	64.3	0.91
15	0.138	72.6	65.3	1.01
16	0.106	75.4	66.0	1.00
17	0.079	71.2	58.5	1.00

Table 3. The 10 top countries based on the quantile general index. The last column shows the original rank based on the SDG scores.

Rank	Country	g_t	J_t	Original Rank
1	Cuba	5.09	1.00	40
2	Romania	3.82	1.00	30
3	Finland	3.33	1.00	1
4	Kyrgyz Republic	3.06	1.00	48
5	Ukraine	1.92	1.00	37
6	Chile	1.11	1.00	28
7	Poland	1.03	1.00	12
8	Georgia	0.24	1.00	51
9	Vietnam	0.01	0.68	55
10	Hungary	0.01	0.64	21

We must be careful with interpretating the result. In particular, the optimal weights had high variation: the ratio of the largest weight (SDG 12) to the smallest weight (SDG 1)

was about $0.49/0.049 = 10.0$, which means that the SDG 1 had only 10% of the impact of the SDG 12 under the quantile general index. This may discourage people or governments contributing to the SDG 1. Our main message in this paper is that there were reversal relations in the SDGs 12 and 13 under the simple sum method, as observed in Table 1, and such a phenomenon can be avoided by the proposed method. Further discussion should be needed for the use of the quantile general index.

As a reviewer suggested, we also computed the Hirsch index [9] (or h-index) of the countries based on the original SDG scores. In the current setting, the h-index is defined as the fixed point of the graph $\{(i, s_i)\}_{i=1}^{17}$, where s_i's are the 17 SDG scores in descending order (normalized into the range $[0, 17]$). The 10 top countries based on the h-index are shown in Table 4. The top three were not changed from the original SDG ranking. We also observed the reversal relations in the SDGs 12 and 13 when we adopted the h-index for separation. See [10] for a study of the scaling behavior of the h-index.

Table 4. The 10 top countries based on the h-index. The last column shows the original rank based on the SDG scores.

Rank	Country	h-Index	Original Rank
1	Finland	13.47	1
2	Denmark	13.35	2
3	Sweden	13.19	3
4	Germany	13.01	6
5	Romania	12.91	30
6	Norway	12.84	4
7	Estonia	12.77	10
8	Croatia	12.77	23
9	Ireland	12.76	9
10	Portugal	12.64	20

6. Discussion

We proposed a quantile general index that avoids reversal relations in the separated groups. The weight was defined by the solution of the convex optimization problem (6) or (7) for given data. In Section 5, we applied the proposed method to the SDG data and obtained the 10 top countries based on it. The result actually satisfies the desired properties (10) and (11). A side effect is that the obtained weights sometimes had large variation, which may be controversial.

Various applications of our method are expected. For example, one could construct a regional competitive index (e.g., [16]) based on the quantile general index if it is necessary to select a given number of top regions. The method is also applicable to admission decisions based on entrance examinations in schools or companies, where a fixed fraction of candidates are supposed to pass. Further case studies are needed to support the validity of our approach.

The quantile general index (without approximation) introduced in Section 3 was reduced to a minimization problem of a nondifferentiable objective function. It is theoretically of interest to develop an exact algorithm and also to estimate the accuracy of the practical method developed in Section 4. Another problem is to find an algorithm that decides the separability of the data into two groups without the reversal relations. In Example 3, we enumerated all possible combinations to prove that the data was not separable. However, this algorithm requires a large amount computational time when the sample size is large. Faster algorithms would be welcomed. Finally, the relation between the quantile general index and the h-index is also completely unknown.

Institutional Review Board Statement: Not applicable.

Appendix A. Proofs

We give a key lemma for the proof of Theorems 1, 3, and 4. In general, we define

$$F_\ell(\boldsymbol{w}, c) = \sum_{i=1}^{d} (-\log w_i) + \mathrm{E}\left[\ell\left(\sum_{i=1}^{d} w_i x_i - c\right)\right] \tag{A1}$$

for $(\boldsymbol{w}, c) \in \mathbb{R}_+^d \times \mathbb{R}$. Here, $\ell : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is a convex function with properties $\ell(0) = 0$ and $\ell(u) > 0$ for $u \neq 0$. The check loss function ℓ_α in Section 2 satisfies these conditions. Under the conditions, we have coercivity

$$\lim_{u \to \pm\infty} \ell(u) = \infty$$

and subadditivity

$$\ell(u + v) \leq \ell(u) + \ell(v)$$

for any real u and v.

We consider the following condition on the nondegeneracy of the distribution P of $\boldsymbol{x}$. We denote the set of nonnegative numbers by $\mathbb{R}_{\geq 0}$.

(C1) $\mathrm{P}(\sum_i w_i x_i = c) < 1$ for any $(\boldsymbol{w}, c) \in \mathbb{R}_{\geq 0}^d \times \mathbb{R}$ with $\boldsymbol{w} \neq \boldsymbol{0}$,

This condition holds if P is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, as assumed in Section 2.

Theorem 1 is immediate from the following lemma.

Lemma A1. *Suppose that $E[\ell(w_i x_i)] < \infty$ for all $i = 1, \ldots, d$ and $w_i > 0$. If the condition (C1) is satisfied, then the function F_ℓ in (A1) admits a minimizer, and the optimal $\boldsymbol{w}$ is unique. Conversely, if (C1) does not hold, then F_ℓ is not bounded from below.*

Proof. We first show that F_ℓ is finite everywhere. Indeed, by the subadditivity of ℓ, we have

$$\mathrm{E}\left[\ell\left(\sum_i w_i x_i - c\right)\right] \leq \sum_i \mathrm{E}[\ell(w_i x_i)] + \ell(-c) < \infty. \tag{A2}$$

To prove the uniqueness, let $(\boldsymbol{w}_1, c_1)$ and $(\boldsymbol{w}_2, c_2)$ be two minimizers of F_ℓ. From the strict convexity of $z \mapsto (-\log z)$ and the convexity of ℓ, we have

$$F_\ell((1 - \lambda)(\boldsymbol{w}_1, c_1) + \lambda(\boldsymbol{w}_2, c_2)) < (1 - \lambda)F_\ell(\boldsymbol{w}_1, c_1) + \lambda F_\ell(\boldsymbol{w}_2, c_2)$$

if $\boldsymbol{w}_1 \neq \boldsymbol{w}_2$. Thus, we have $\boldsymbol{w}_1 = \boldsymbol{w}_2$, and the uniqueness follows.

To prove the existence, we show that the sublevel set

$$\{(\boldsymbol{w}, c) \in \mathbb{R}_+^d \times \mathbb{R} \mid F_\ell(\boldsymbol{w}, c) \leq a\}$$

is compact for each $a \in \mathbb{R}$. We define a function $R : \mathbb{R}_{\geq 0}^d \times \mathbb{R} \to \mathbb{R}_{\geq 0}$ by

$$R(\boldsymbol{w}, c) = \mathrm{E}[\ell(\boldsymbol{w}^\top \boldsymbol{x} - c)].$$

Then, R is continuous and strictly positive unless $(\boldsymbol{w}, c) = (\boldsymbol{0}, 0)$. Indeed, the continuity of R is a consequence of Lebesgue's dominated convergence theorem, and the strict positivity follows from the condition (C1). Let

$$\gamma := \inf_{\sum_i w_i + |c| = 1} R(\boldsymbol{w}, c) > 0.$$

Since R is convex, and $R(\boldsymbol{0}, 0) = 0$, we have

$$R(\boldsymbol{w}, c) \geq \gamma \left(\sum_i w_i + |c| \right)$$

whenever $\sum_i w_i + |c| \geq 1$. For any $(\boldsymbol{w}, c) \in \mathbb{R}_+^d \times \mathbb{R}$, we have

$$\begin{aligned} F_\ell(\boldsymbol{w}, c) &= \sum_i (-\log w_i) + R(\boldsymbol{w}, c) \\ &\geq \sum_i (-\log w_i) + \gamma \left(\sum_i w_i + |c| - 1 \right) \\ &= \sum_i \{(-\log w_i) + \gamma w_i\} + \gamma |c| - \gamma. \end{aligned}$$

Since the functions $w_i \mapsto (-\log w_i) + \gamma w_i$ and $c \mapsto |c|$ have compact sublevel sets, the sublevel set of F_ℓ is also compact. $\quad\square$

In order to prove Theorems 3 and 4, it is enough to replace the distribution P by the empirical distribution $P_n = n^{-1} \sum_{t=1}^n \delta_{\boldsymbol{x}_{(t)}}$, where $\delta_{\boldsymbol{a}}$ is the Dirac measure at a point $\boldsymbol{a} \in \mathbb{R}^d$. In Theorem 3, the uniqueness of c when $n\alpha$ is not an integer follows from the observation that the optimal c for a fixed $\boldsymbol{w}$ must be $\boldsymbol{w}^\top \boldsymbol{x}_{(t)}$ for some t.

References

1. Sachs, J.; Kroll, C.; Lafortune, G.; Fuller, G.; Woelm, F. *Sustainable Development Report 2022*; Cambridge University Press: Cambridge, UK, 2022. Available online: https://dashboards.sdgindex.org (accessed on 4 June 2022).
2. Sei, T. An objective general index for multivariate ordered data, *J. Multivar. Anal.* **2016**, *147*, 247–264. [CrossRef]
3. Sei, T. Coordinate-wise transformation and Stein-type densities. In *Geometric Science of Information. GSI*; Nielsen F., Barbaresco F., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2017; Volume 10589.
4. Sei, T. Coordinate-wise transformation of probability distributions to achieve a Stein-type identity. *Inf. Geom.* **2022**, *5*, 325–354. [CrossRef]
5. Baker, R.J. Selection indexes without economic weights for animal breeding. *Can. J. Anim. Sci.* **1974**, *54*, 1–8. [CrossRef]
6. Elston, R.C. A weight-free index for the purpose of ranking or selection with respect to several traits at a time. *Biometrics* **1963**, *19*, 85–97. [CrossRef]
7. Montanari, A.; Richard, E. Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Trans. Inf. Theory* **2015**, *62*, 1458–1484. [CrossRef]
8. Bartholomew, D.; Knott, M.; Moustaki, I. *Latent Variable Models and Factor Analysis, A Unified Approach*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2011.
9. Hirsch, J.E. An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16569–16572. [CrossRef] [PubMed]
10. Ghosh, A.; Chakrabarti, B.K.; Ram, D.R.S.; Mitra, M.; Maiti, R.; Biswas, S.; Banerjee, S. Scaling behavior of the Hirsch index for failure avalanches, percolation clusters and paper citations. *arXiv* **2021**, arXiv:2109.14500.
11. Koenker, R.; Bassett, G. Regression quantiles. *Econometrica* **1978**, *46*, 33–50. [CrossRef]
12. Barnett, V. The ordering of multivariate data. *J. R. Stat. Soc. Ser. A* **1976**, *139*, 318–355. [CrossRef]
13. Koenker, R. *Quantile Regression*; Cambridge University Press: Cambridge, UK, 2005.
14. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1972.
15. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022. Available online: http://www.R-project.org (accessed on 5 October 2022).
16. Charles, V.; Sei, T. A two-stage OGI approach to compute the regional competitiveness index. *Compet. Rev.* **2019**, *29*, 78–95. [CrossRef]