An Entropy-Based Approach for Measuring Factor Contributions in Factor Analysis Models

Nobuoki Eshima [1],*, Minoru Tabata [2] and **Claudio Giovanni Borroni [3]**

[1] Center for Educational Outreach and Admissions, Kyoto University, Kyoto 606-8501, Japan
[2] Department of Mathematical Sciences, Osaka Prefecture University, Osaka 599-8532, Japan; mnrtabata@luck.ocn.ne.jp
[3] Department of Statistics and Quantitative Methods, University of Milano Bicocca, 20126 Milano, Italy; claudio.borroni@unimib.it
* Correspondence: eshima.nobuoki.2z@kyoto-u.ac.jp; Tel.: +81-75-753-2517

Abstract: In factor analysis, factor contributions of latent variables are assessed conventionally by the sums of the squared factor loadings related to the variables. First, the present paper considers issues in the conventional method. Second, an alternative entropy-based approach for measuring factor contributions is proposed. The method measures the contribution of the common factor vector to the manifest variable vector and decomposes it into contributions of factors. A numerical example is also provided to demonstrate the present approach.

Keywords: entropy coefficient of determination; factor contribution; factor loading; path analysis

1. Introduction

Factor analysis is a statistical method for extracting simple structures to explain inter-relations between manifest and latent variables. The origin dates back to the works of [1], and the single factor model was extended to the multiple factor model [2]. These days, factor analysis is widely applied in behavioral sciences [3]; hence, it is important to interpret the extracted factors and is critical to explain how such factors influence manifest variables, that is, measurement of factor contribution. Let X_i be manifest variables; ξ_j latent variables (common factors); ε_i unique factors related to X_i; and let λ_{ij} be factor loadings that are weights of factors ξ_j to explain X_i. Then, the factor analysis model is given as follows:

$$X_i = \sum_{j=1}^m \lambda_{ij}\xi_j + \varepsilon_i, \ \ i = 1, 2, \ldots, p, \tag{1}$$

where

$$\mathrm{E}(X_i) = \mathrm{E}(\xi_j) = \mathrm{E}(\varepsilon_i) = 0, \ \mathrm{var}(\xi_j) = 1, \ \mathrm{cov}(\xi_j, \varepsilon_i) = 0,$$
$$\mathrm{cov}(\varepsilon_i, \varepsilon_k) = 0 \text{ for } i \neq k \text{ and } \mathrm{var}(\varepsilon_i) = \sigma_i^2 > 0$$

For the simplicity of discussion, common factors ξ_j are assumed to be mutually independent in this section, that is, we first consider an orthogonal factor analysis model. In the conventional approach, the contribution of factor ξ_j to all manifest variables X_i, C_j, is defined as follows:

$$C_j = \sum_{i=1}^p \mathrm{cov}(X_i, \xi_j)^2 = \sum_{i=1}^p \lambda_{ij}^2 \tag{2}$$

The above definition of factor contributions is based on the following decomposition of the total of variances of the observed variables X_i [4] (p. 59):

$$\sum_{i=1}^p \mathrm{var}(X_i) = \sum_{j=1}^m \sum_{i=1}^p \lambda_{ij}^2 + \sum_{i=1}^p \sigma_i^2$$

What physical meaning does the above quantity have? Applying it to the manifest variables observed, however, such a decomposition leads to scale-variant results. For this reason, factor contribution is usually considered on the standardized versions of manifest variables X_i. What does it mean to measure factor contributions by (2)? For standardized manifest variables X_i, we have

$$\lambda_{ij} = \mathrm{cor}(X_i, \xi_j) \tag{3}$$

Then, (2) is the sum of the coefficients of determination for all standardized manifest variables X_i with respect to a single latent variable ξ_j. The squared correlation coefficients (3), that is, $\mathrm{cor}(X_i, \xi_j)^2$, are the ratios of explained variances of a manifest variable X_i, and in this sense, they can be interpreted as the contributions (effects) of factors ξ_j to the manifest variable X_i. Although, what does the sum of these with respect to all manifest variables X_i, that is, (2), mean? The conventional method may be intuitively reasonable for measuring factor contributions; however, we think it is sensible to propose a method measuring factor contributions as the effects of factors on the manifest variable vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$, which are interpretable and have a theoretical basis. There is no research on this topic as far as we have searched. The present paper provides an entropy-based solution to the problem. Entropy is a useful concept to measure the uncertainty in the systems of random variables and sample spaces [5] and it can be applied to measure multivariate dependences of random variables [6,7].

This paper proposes an entropy-based method for measuring factor contributions of ξ_j to the manifest variable vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$ concerned, which can treat not only orthogonal factors, but also oblique cases. The present paper has five sections in addition to this section. In Section 2, the conventional method for measuring factor contributions is reviewed. Section 3 considers the factor analysis model in view of entropy and makes a preliminary discussion on measurement of factor contribution. In Section 4, an entropy-based path analysis is applied as a tool to measure factor contributions. Contributions of factors ξ_j are defined by the total effects of the factors on the manifest variable vector, and the contributions are decomposed into those to manifest variables and subsets of manifest variables. Section 5 illustrates the present method using a numerical example. Finally, in Section 6, some conclusions are provided.

2. Relative Factor Contributions in the Conventional Method

In the conventional approach, for the orthogonal factor model (1), the contribution ratio of ξ_j is defined by

$$\mathrm{RC}_j = \frac{C_j}{\sum_{l=1}^{m} C_l} = \frac{\sum_{i=1}^{p} \lambda_{ij}^2}{\sum_{l=1}^{m} \sum_{k=1}^{p} \lambda_{kl}^2} \tag{4}$$

The above measure is referred to as the factor contribution ratio in the common factor space. Let R_i be the multiple correlation coefficient of latent variable vector $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_m)^T$ and manifest variable X_i. Then, for standardized manifest variable X_i, we have

$$R_i^2 = \sum_{j=1}^{m} \lambda_{ij}^2 \tag{5}$$

The above quantity can be interpreted as the effect (explanatory power) of latent variable vector $\boldsymbol{\xi} = (\xi_j)$ on manifest variable X_i; however, the denominator of (4) is the sum of those effects (5) and there is no theoretical basis to interpret it. Another contribution ratio of ξ_j is referred to as that in the whole space of $\boldsymbol{X} = (X_i)$, and is defined by

$$\widetilde{\mathrm{RC}}_j = \frac{C_j}{\sum_{i=1}^{p} \mathrm{var}(X_i)} = \frac{\sum_{k=1}^{p} \lambda_{kj}^2}{\sum_{k=1}^{p} \left(\sum_{l=1}^{m} \lambda_{kl}^2 + \sigma_k^2\right)} \tag{6}$$

If the manifest variables are standardized, we have

$$\widetilde{RC}_j = \frac{C_j}{p} = \frac{\sum_{k=1}^{p} \lambda_{kj}^2}{p}$$

Here, there is an issue similar to (4), because the denominator in (6) does not express the variation of the manifest variable vector $X = (X_i)$. Indeed, it is the sum of the variances of manifest variables and does not include covariances between them. In the next section, the factor analysis model (1) is reconsidered in the framework of generalized linear models (GLMs), and the effects (contributions) of latent variables ξ_j on the manifest variable vector $X = (X_i)$, that is, factor contributions, are discussed through entropy [8].

3. Factor Analysis Model and Entropy

It is assumed that factors ε_i and ξ_j are normally distributed, and the factor analysis model (1) is reconsidered in the GLM framework. Let $\mathbf{\Lambda} = (\lambda_{ij})$ be a $p \times m$ factor loading matrix; let $\mathbf{\Phi}$ be an $m \times m$ correlation matrix of common factor vector $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_m)^T$; and let $\mathbf{\Omega}$ be the $p \times p$ variance-covariance matrix of unique factor vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)^T$. The conditional density function of X given $\boldsymbol{\xi}$, $f(x|\boldsymbol{\xi})$, is normal with mean $\mathbf{\Lambda}\boldsymbol{\xi}$ and variance matrix $\mathbf{\Omega}$, and is given as follows:

$$f(x|\boldsymbol{\xi}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{\Omega}|^{\frac{1}{2}}} \exp\left(\frac{X^T \widetilde{\mathbf{\Omega}} \mathbf{\Lambda} \boldsymbol{\xi} - \frac{1}{2} \boldsymbol{\xi}^T \mathbf{\Lambda}^T \widetilde{\mathbf{\Omega}}^2 \mathbf{\Lambda} \boldsymbol{\xi}}{|\mathbf{\Omega}|} - \frac{\frac{1}{2} X^T \widetilde{\mathbf{\Omega}} X}{|\mathbf{\Omega}|} \right)$$

where $\widetilde{\mathbf{\Omega}}$ is the cofactor matrix of $\mathbf{\Omega}$. Let $f(x)$ and $g(\boldsymbol{\xi})$ be the marginal density functions of X and $\boldsymbol{\xi}$, respectively. Then, a basic predictive power measure for GLMs [9] is based on the Kullback–Leibler information [6], and applying it to the above model, we have

$$\text{KL}(\boldsymbol{X}, \boldsymbol{\xi}) = \iint f(x|\boldsymbol{\xi}) g(\boldsymbol{\xi}) \log \frac{f(x|\boldsymbol{\xi})}{f(x)} dx d\boldsymbol{\xi} + \iint f(x) g(\boldsymbol{\xi}) \log \frac{f(x)}{f(x|\boldsymbol{\xi})} dx d\boldsymbol{\xi} = \frac{\text{tr}\widetilde{\mathbf{\Omega}} \mathbf{\Lambda} \mathbf{\Phi} \mathbf{\Lambda}^T}{|\mathbf{\Omega}|} \quad (7)$$

The above measure was derived from a discussion on log odds ratios in GLMs [9], and is scale-invariant with respect to manifest variables X_i. The numerator of (7) is the explained entropy of X by $\boldsymbol{\xi}$, and the denominator is the dispersion of the unique factors in entropy, that is, the generalized variance of $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)^T$. Thus, (7) expresses the total effect (contribution) of factor vector $\boldsymbol{\xi} = (\xi_j)$ on manifest variable vector $X = (X_i)$ in entropy, and is denoted by $C(\boldsymbol{\xi} \to X)$ in the present paper. The entropy coefficient of determination (ECD) is calculated as follows [9]:

$$\text{ECD}(\boldsymbol{X}, \boldsymbol{\xi}) = \frac{\text{tr}\widetilde{\mathbf{\Omega}} \mathbf{\Lambda} \mathbf{\Phi} \mathbf{\Lambda}^T}{\text{tr}\widetilde{\mathbf{\Omega}} \mathbf{\Lambda} \mathbf{\Phi} \mathbf{\Lambda}^T + |\mathbf{\Omega}|} \quad (8)$$

The denominator of the above measure is interpreted as the variation of manifest variable vector $X = (X_i)$ in entropy and the numerator is the explained variation of random vector X in entropy. In this sense, ECD (8) is the factor contribution ratio of $\boldsymbol{\xi} = (\xi_j)$ for the whole entropy space of $X = (X_i)$, and it expresses the standardized total effect of $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_m)^T$ on the manifest variable vector $X = (X_1, X_2, \ldots, X_p)^T$, which is denoted by $e_T(\boldsymbol{\xi} \to X)$ [8,10]. As for (6), in the present paper, the ECD is denoted by $\widetilde{RC}(\boldsymbol{\xi} \to X)$, that is, the relative contribution of factor vector $\boldsymbol{\xi}$ for the whole space of manifest variable vector X in entropy.

Remark 1. Let $\mathbf{\Sigma}$ be the $p \times p$ variance-covariance matrix of manifest variable vector $X = (X_1, X_2, \ldots, X_p)^T$ and let $\mathbf{\Phi}$ be the $m \times m$ correlation matrix of $\boldsymbol{\xi}$. Then, we have

$$\mathbf{\Sigma} = \mathbf{\Lambda} \mathbf{\Phi} \mathbf{\Lambda}^T + \mathbf{\Omega} \quad (9)$$

For assessing the goodness-of-fit of the models, the following overall coefficient of determination (OCD) is suggested ([11], p. 60) on the basis of (9):

$$\mathrm{OCD}(\boldsymbol{X},\,\boldsymbol{\xi}) = 1 - \frac{|\boldsymbol{\Omega}|}{|\boldsymbol{\Sigma}|}\left(= \frac{|\boldsymbol{\Sigma}| - |\boldsymbol{\Omega}|}{|\boldsymbol{\Sigma}|}\right)$$

Determinant $|\boldsymbol{\Omega}|$ is the generalized variance of unique factor vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)^T$ and $|\boldsymbol{\Sigma}|$ is that of manifest variable vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^T$. Then, OCD is interpreted as the ratio of the explained generalized variance of manifest variable vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^T$ by common factor vector $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_m)^T$ in the p-dimensional Euclidian space. On the other hand, from (8), it follows that

$$\mathrm{ECD}(\boldsymbol{X},\,\boldsymbol{\xi}) = 1 - \frac{|\boldsymbol{\Omega}|}{tr\widetilde{\boldsymbol{\Omega}}\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T + |\boldsymbol{\Omega}|}$$

Hence, ECD is viewed as the ratio of the explained variation of the manifest variable vector in entropy.

Cofactor matrix $\widetilde{\boldsymbol{\Omega}}$ is diagonal and the (i, i) elements are $\prod\limits_{k \neq i} \sigma_k^2$, $i = 1, 2, \ldots, p$. If common factors are statistically independent, it follows that

$$tr\widetilde{\boldsymbol{\Omega}}\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T = \sum_{i=1}^{p} \prod_{k \neq i} \sigma_k^2 \sum_{j=1}^{m} \lambda_{ij}^2$$
$$= \sum_{i=1}^{p} \sum_{j=1}^{m} \lambda_{ij}^2 \prod_{k \neq i} \sigma_k^2$$

Thus, (7) is decomposed as

$$\mathrm{KL}(\boldsymbol{X},\,\boldsymbol{\xi}) = \sum_{i=1}^{p} \sum_{j=1}^{m} \frac{\lambda_{ij}^2}{\sigma_i^2}$$

As detailed below, in the present paper, the contribution of factor ξ_j to $\boldsymbol{X}$, $\mathrm{C}(\xi_j \to \boldsymbol{X})$, is defined by

$$\mathrm{C}(\xi_j \to \boldsymbol{X}) = \sum_{i=1}^{p} \frac{\lambda_{ij}^2}{\sigma_i^2} \tag{10}$$

Remark 2. The above contribution is different from the conventional definition of factor contribution (2); unless $\sigma_i^2 = 1$, $i = 1, 2, \ldots, p$. In this sense, we may say that the standardization of manifest variables in entropy is obtained by setting all the unique factor variances to one.

In the next section, the contributions (effects) of factors ξ_j to manifest variable vector $\boldsymbol{X}$ are discussed in a general framework through an entropy-based path analysis [8].

4. Measurement of Factor Contribution Based on Entropy

A path diagram for the factor analysis model is given in Figure 1, in which the single-headed arrows imply the directions of effects of factors and the double-headed curved arrows indicate the associations between the related variables. In this section, common factors are assumed to be correlated, that is, we consider an oblique case, and an entropy-based path analysis [8] is applied to make a general discussion in the measurement of factor contributions.
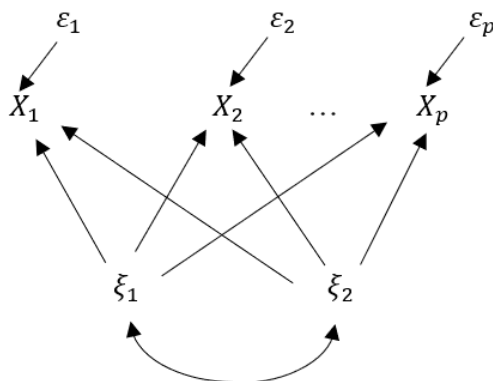
Figure 1. Path diagram for factor analysis model (1) ($m = 2$).

Theorem 1. *In the factor analysis model (1),*

$$\mathrm{KL}(\boldsymbol{X}, \boldsymbol{\xi}) = \sum_{i=1}^{p} \mathrm{KL}(X_i, \boldsymbol{\xi})$$

Proof. Let $f_i(x_i|\boldsymbol{\xi})$ be the conditional density functions of manifest variables X_i, given factor vector $\boldsymbol{\xi}$; let $f_i(x_i)$ be the marginal density functions of X_i; let $f(\boldsymbol{x})$ be the marginal density function of $\boldsymbol{X}$; and let $g(\boldsymbol{\xi})$ be the marginal density function of common factor vector $\boldsymbol{\xi}$. As the manifest variables are conditionally independent, given factor vector $\boldsymbol{\xi}$, the conditional density function of $\boldsymbol{X}$ is

$$f(\boldsymbol{x}|\boldsymbol{\xi}) = \prod_{i=1}^{p} f_i(x_i|\boldsymbol{\xi})$$

From (7), we have

$$\mathrm{KL}(\boldsymbol{X}, \boldsymbol{\xi}) = \iint \prod_{i=1}^{p} f_i(x_i|\boldsymbol{\xi}) g(\boldsymbol{\xi}) \log \frac{\prod_{k=1}^{p} f_k(x_k|\boldsymbol{\xi})}{f(\boldsymbol{x})} d\boldsymbol{x} d\boldsymbol{\xi} + \iint f(\boldsymbol{x}) g(\boldsymbol{\xi}) \log \frac{f(\boldsymbol{x})}{\prod_{k=1}^{p} f_k(x_k|\boldsymbol{\xi})} d\boldsymbol{x} d\boldsymbol{\xi}$$

$$= \iint \left(\prod_{i=1}^{p} f_i(x_i|\boldsymbol{\xi}) g(\boldsymbol{\xi}) - f(\boldsymbol{x}) g(\boldsymbol{\xi}) \right) \log \prod_{k=1}^{p} f_k(x_k|\boldsymbol{\xi}) d\boldsymbol{x} d\boldsymbol{\xi}$$

$$= \iint \prod_{i=1}^{p} f_i(x_i|\boldsymbol{\xi}) g(\boldsymbol{\xi}) \log \frac{\prod_{k=1}^{p} f_k(x_k|\boldsymbol{\xi})}{\prod_{k=1}^{p} f_k(x_k)} d\boldsymbol{x} d\boldsymbol{\xi}$$

$$+ \iint f(\boldsymbol{x}) g(\boldsymbol{\xi}) \log \frac{\prod_{k=1}^{p} f_k(x_k)}{\prod_{k=1}^{p} f_k(x_k|\boldsymbol{\xi})} d\boldsymbol{x} d\boldsymbol{\xi}$$

$$= \sum_{k=1}^{p} \iint \prod_{i=1}^{p} f_i(x_i|\boldsymbol{\xi}) g(\boldsymbol{\xi}) \log \frac{f_k(x_k|\boldsymbol{\xi})}{f_k(x_k)} d\boldsymbol{x} d\boldsymbol{\xi} + \sum_{k=1}^{p} \iint f(\boldsymbol{x}) g(\boldsymbol{\xi}) \log \frac{f_k(x_k)}{f_k(x_k|\boldsymbol{\xi})} d\boldsymbol{x} d\boldsymbol{\xi}$$

$$= \sum_{k=1}^{p} \iint f_k(x_k|\boldsymbol{\xi}) g(\boldsymbol{\xi}) \log \frac{f_k(x_k|\boldsymbol{\xi})}{f_k(x_k)} dx_k d\boldsymbol{\xi}$$

$$+ \sum_{k=1}^{p} \iint f_k(x_k) g(\boldsymbol{\xi}) \log \frac{f_k(x_k)}{f_k(x_k|\boldsymbol{\xi})} dx_k d\boldsymbol{\xi}$$

$$= \sum_{i=1}^{p} \left(\iint f_i(x_i|\boldsymbol{\xi}) g(\boldsymbol{\xi}) \log \frac{f_i(x_i|\boldsymbol{\xi})}{f_i(x_i)} dx_i d\boldsymbol{\xi} + \iint f_i(x_i) g(\boldsymbol{\xi}) \log \frac{f_i(x_i)}{f_i(x_i|\boldsymbol{\xi})} dx_i d\boldsymbol{\xi} \right) = \sum_{i=1}^{p} KL(X_i, \boldsymbol{\xi})$$

$\square$

In model (1) with correlation matrix $\Phi = (\varphi_{ij})$, we have

$$\mathrm{KL}(X_i, \boldsymbol{\xi}) = \frac{\sum_{k=1}^{m} \sum_{l=1}^{m} \lambda_{ik} \varphi_{kl} \lambda_{il}}{\sigma_i^2}$$

The above quantity is referred to as the contribution of $\boldsymbol{\xi}$ to X_i, and is denoted as $C(\boldsymbol{\xi} \to X_i)$. Let R_i be the multiple correlation coefficient of X_i and $\boldsymbol{\xi} = (\xi_j)$. Then,

$$C(\xi \to X_i) = \frac{R_i^2}{1 - R_i^2}(= \mathrm{KL}(X_i, \xi)) \tag{11}$$

From Theorem 1, we then have

$$C(\xi \to X) = \sum_{i=1}^{p} \frac{R_i^2}{1 - R_i^2}(= \mathrm{KL}(X, \xi)) \tag{12}$$

Hence, Theorem 1 gives the following decomposition of the contribution of ξ on X into those on the single manifest variables X_i (11):

$$C(\xi \to X) = \sum_{i=1}^{p} C(\xi \to X_i) \tag{13}$$

Remark 3. Notice that in the denominator of (4), the total contribution of all factors ξ_i is simply defined as the total sum assessed:

$$\sum_{l=1}^{m} C_l = \sum_{i=1}^{p} R_i^2$$

On the other hand, in the present approach, the total effect (contribution) of factor vector ξ on manifest variable vector X is decomposed into those of manifest variables X_i, (12) and (13).

Let X_{sub} be any sub-vector of manifest variable vector $X = (X_1, X_2, \ldots, X_p)^T$. Then, the contribution of factor vector ξ to X_{sub} is defined by

$$C(\xi \to X_{sub}) = \mathrm{KL}(X_{sub}, \xi)$$

From Theorem 1, we have the following corollary.

Corollary 1. *Let* $X_{(1)} = \left(X_{i_1}, X_{i_2}, \ldots, X_{i_q}\right)^T$ *and* $X_{(2)} = \left(X_{j_1}, X_{j_2}, \ldots, X_{j_{p-q}}\right)^T$ *be a decomposition of manifest variable vector* $X = (X_1, X_2, \ldots, X_p)^T$, *where* $q < p$. *Then, for factor analysis model (1), it follows that*

$$C(\xi \to X) = C\left(\xi \to X_{(1)}\right) + C\left(\xi \to X_{(2)}\right)$$
$$C\left(\xi \to X_{(1)}\right) = \sum_{k=1}^{q} C(\xi \to X_{i_k}), \quad C\left(\xi \to X_{(2)}\right) = \sum_{k=1}^{p-q} C(\xi \to X_{j_k})$$

Proof: From a similar discussion to the proof of Theorem 1, we have

$$\mathrm{KL}(X, \xi) = \mathrm{KL}\left(X_{(1)}, \xi\right) + \mathrm{KL}\left(X_{(2)}, \xi\right)$$
$$\mathrm{KL}\left(X_{(1)}, \xi\right) = \sum_{k=1}^{q} \mathrm{KL}(X_{i_k}, \xi), \quad \mathrm{KL}\left(X_{(2)}, \xi\right) = \sum_{k=1}^{p-q} \mathrm{KL}(X_{j_k}, \xi)$$

Hence, the corollary follows.

Next, the standardized total effects of single factors ξ_j on manifest variable vector X, that is, $e_T(\xi_j \to X)$, are calculated [8,10]. Let $\xi^{/j} = (\xi_1, \xi_2, \ldots, \xi_{j-1}, \xi_{j+1}, \ldots, \xi_m)^T$; $f\left(x, \xi^{/j} \mid \xi_j\right)$ be the conditional density function of X and $\xi^{/j}$ given ξ_j; $f(x \mid \xi_j)$ be the conditional density function of X given ξ_j; $g\left(\xi^{/j} \mid \xi_j\right)$ be the conditional density function of $\xi^{/j}$ given ξ_j; and $g_j(\xi_j)$ be the marginal density function of ξ_j. Then, we have

$$\mathrm{KL}\left(X, \xi^{/j} \mid \xi_j\right) = \iint f(x, \xi) \log \frac{f\left(x, \xi^{/j} \mid \xi_j\right)}{f(x \mid \xi_j) g(\xi^{/j} \mid \xi_j)} dx d\xi^{/j} d\xi_j$$
$$+ \iint f(x \mid \xi_j) g\left(\xi^{/j} \mid \xi_j\right) g_j(\xi_j) \log \frac{f(x \mid \xi_j) g(\xi^{/j} \mid \xi_j)}{f(x, \xi^{/j} \mid \xi_j)} dx d\xi^{/j} d\xi_j = \frac{\mathrm{tr}\tilde{\Omega}\Lambda \mathrm{cov}(\xi, X \mid \xi_j)}{|\Omega|}$$

where $\mathrm{cov}\left(\boldsymbol{\xi}, \boldsymbol{X}\big|\xi_j\right)$ is a $m \times p$ covariance matrix given ξ_j, of which the (k, i) elements are $\mathrm{cov}\left(\xi_k, X_i\big|\xi_j\right)$. The standardized total effect $e_T\left(\xi_j \to \boldsymbol{X}\right)$ is given by

$$e_T\left(\xi_j \to \boldsymbol{X}\right) = \frac{\mathrm{KL}(\boldsymbol{X}, \boldsymbol{\xi}) - \mathrm{KL}\left(\boldsymbol{X}, \boldsymbol{\xi}^{/j}\big|\xi_j\right)}{\mathrm{KL}(\boldsymbol{X}, \boldsymbol{\xi}) + 1} = \frac{\mathrm{tr}\widetilde{\boldsymbol{\Omega}}\boldsymbol{\Lambda}\left(\mathrm{cov}(\boldsymbol{\xi}, \boldsymbol{X}) - \mathrm{cov}\left(\boldsymbol{\xi}, \boldsymbol{X}\big|\xi_j\right)\right)}{\mathrm{tr}\widetilde{\boldsymbol{\Omega}}\boldsymbol{\Lambda}\mathrm{cov}(\boldsymbol{\xi}, \boldsymbol{X}) + |\boldsymbol{\Omega}|}$$

The standardized total effect $e_T\left(\xi_j \to \boldsymbol{X}\right)$ [8] is interpreted as the contribution ratio of factor ξ_j in the whole entropy space of $\boldsymbol{X}$, and in the present paper, it is denoted by $\widetilde{RC}\left(\xi_j \to \boldsymbol{X}\right)$. The contribution of factor ξ_j measured in entropy is defined by

$$\mathrm{C}\left(\xi_j \to \boldsymbol{X}\right) = \mathrm{KL}(\boldsymbol{X}, \boldsymbol{\xi}) - \mathrm{KL}\left(\boldsymbol{X}, \boldsymbol{\xi}^{/j}\big|\xi_j\right) = \frac{\mathrm{tr}\widetilde{\boldsymbol{\Omega}}\boldsymbol{\Lambda}\mathrm{cov}(\boldsymbol{\xi}, \boldsymbol{X})}{|\boldsymbol{\Omega}|} - \frac{\mathrm{tr}\widetilde{\boldsymbol{\Omega}}\boldsymbol{\Lambda}\mathrm{cov}\left(\boldsymbol{\xi}, \boldsymbol{X}\big|\xi_j\right)}{|\boldsymbol{\Omega}|}$$

As for (6), the relative contribution of factor ξ_j on $\boldsymbol{X}$ is given by

$$\mathrm{RC}\left(\xi_j \to \boldsymbol{X}\right) = \frac{\widetilde{RC}\left(\xi_j \to \boldsymbol{X}\right)}{\widetilde{RC}\left(\boldsymbol{\xi} \to \boldsymbol{X}\right)} = \frac{\mathrm{C}\left(\xi_j \to \boldsymbol{X}\right)}{\mathrm{C}\left(\boldsymbol{\xi} \to \boldsymbol{X}\right)}$$

Concerning factor contributions of ξ_j on the single manifest variables X_i, that is, $\mathrm{C}\left(\xi_j \to X_i\right)$, the following theorem can be stated.

Theorem 2. *In the factor analysis model (1),*

$$\mathrm{C}\left(\xi_j \to \boldsymbol{X}\right) = \sum_{i=1}^{p}\mathrm{C}\left(\xi_j \to X_i\right)$$

Proof: From Theorem 1, it follows that

$$\mathrm{KL}\left(\boldsymbol{X}, \boldsymbol{\xi}^{/j}\big|\xi_j\right) = \sum_{i=1}^{p}\mathrm{KL}\left(X_i, \boldsymbol{\xi}^{/j}\big|\xi_j\right)$$

Then, we have

$$\mathrm{C}\left(\xi_j \to X_i\right) = \mathrm{KL}(X_i, \boldsymbol{\xi}) - \mathrm{KL}\left(X_i, \boldsymbol{\xi}^{/j}\big|\xi_j\right)$$

and,

$$\begin{aligned}\mathrm{C}\left(\xi_j \to \boldsymbol{X}\right) &= \mathrm{KL}(\boldsymbol{X}, \boldsymbol{\xi}) - \mathrm{KL}\left(\boldsymbol{X}, \boldsymbol{\xi}^{/j}\big|\xi_j\right) \\ &= \sum_{i=1}^{p}\mathrm{KL}(X_i, \boldsymbol{\xi}) - \sum_{i=1}^{p}\mathrm{KL}\left(X_i, \boldsymbol{\xi}^{/j}\big|\xi_j\right) \\ &= \sum_{i=1}^{p}\left(KL(X_i, \boldsymbol{\xi}) - \mathrm{KL}\left(X_i, \boldsymbol{\xi}^{/j}\big|\xi_j\right)\right) = \sum_{i=1}^{p}\mathrm{C}\left(\xi_j \to X_i\right)\end{aligned}$$

From the above theorem, we have the following corollary.

Corollary 2. *Let* $\boldsymbol{X}_{(1)} = \left(X_{i_1}, X_{i_2}, \ldots, X_{i_q}\right)^T$ *and* $\boldsymbol{X}_{(2)} = \left(X_{j_1}, X_{j_2}, \ldots, X_{j_{p-q}}\right)^T$ *be decomposition of manifest variable vector* $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_p\right)^T$, *where* $q < p$.

$$\mathrm{C}\left(\xi_j \to \boldsymbol{X}\right) = \mathrm{C}\left(\xi_j \to \boldsymbol{X}_{(1)}\right) + \mathrm{C}\left(\xi_j \to \boldsymbol{X}_{(2)}\right)$$
$$\mathrm{C}\left(\xi_j \to \boldsymbol{X}_{(1)}\right) = \sum_{k=1}^{q}\mathrm{C}\left(\xi_j \to X_{i_k}\right), \quad \mathrm{C}\left(\xi_j \to \boldsymbol{X}_{(2)}\right) = \sum_{k=1}^{p-q}\mathrm{C}\left(\xi_j \to X_{j_k}\right)$$

Proof: From a similar discussion in the proof of Theorem 2, the corollary follows. □

Remark 4. Let $\boldsymbol{X}_{sub}$ be any sub-vector of manifest variable vector $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_p\right)^T$. By substituting $\boldsymbol{X}$ for $\boldsymbol{X}_{sub}$ in the above discussion, $\mathrm{C}\left(\boldsymbol{\xi} \to \boldsymbol{X}_{sub}\right)$, $\mathrm{C}\left(\xi_j \to \boldsymbol{X}_{sub}\right)$, $\widetilde{RC}\left(\xi_j \to \boldsymbol{X}_{sub}\right)$, and $\mathrm{RC}\left(\xi_j \to \boldsymbol{X}_{sub}\right)$ can be defined.

For orthogonal factor analysis models, the following theorem holds true.

Theorem 3. *In factor analysis model (1), if common factors ξ_j are statistically independent, then*

$$\mathrm{C}(\boldsymbol{\xi} \to \boldsymbol{X}) = \sum_{j=1}^{m} \sum_{i=1}^{p} \mathrm{C}(\xi_j \to X_i).$$

Proof: From model (1), we have

$$\mathrm{C}(\xi_j \to X_i) = \mathrm{KL}(X_i, \boldsymbol{\xi}) - \mathrm{KL}\left(X_i, \boldsymbol{\xi}^{/j} \Big| \xi_j\right) = \frac{\lambda_{ij}^2}{\sigma_i^2}$$

This completes the theorem. □

From the above discussion, if common factors ξ_j are statistically independent, (10) is derived. Moreover, we have

$$\widetilde{\mathrm{RC}}\left(\xi_j \to \boldsymbol{X}\right) = \frac{\mathrm{KL}(\boldsymbol{X}, \boldsymbol{\xi}) - \mathrm{KL}\left(\boldsymbol{X}, \boldsymbol{\xi}^{/j} \Big| \xi_j\right)}{\mathrm{KL}(\boldsymbol{X}, \boldsymbol{\xi}) + 1} = \frac{\sum_{i=1}^{p} \frac{\lambda_{ij}^2}{\sigma_j^2}}{\mathrm{KL}(\boldsymbol{X}, \boldsymbol{\xi}) + 1}$$

This measure is the relative contribution ratio of ξ_j for the variation of $\boldsymbol{X}$ in entropy. The relative contributions of ξ_j on $\boldsymbol{X}$ in entropy are calculated as follows:

$$\mathrm{RC}(\xi_j \to \boldsymbol{X}) = \frac{\mathrm{C}(\xi_j \to \boldsymbol{X})}{\mathrm{C}(\boldsymbol{\xi} \to \boldsymbol{X})} = \frac{\sum_{i=1}^{p} \frac{\lambda_{ij}^2}{\sigma_i^2}}{\sum_{j=1}^{m} \sum_{i=1}^{p} \frac{\lambda_{ij}^2}{\sigma_i^2}}$$

Remark 5. It is difficult to use OCD for assessing factor contributions, because $|\Sigma|$ cannot be decomposed as in the above discussion.

5. Numerical Example

In order to illustrate the present method, we use the data shown in Table 1 [12]. In this table, manifest variables X_1, X_2, and X_3 are subjects in liberal arts and variables X_4 and X_5 are those in sciences. First, orthogonal factor analysis (varimax method by S-PLUS ver. 8.2) is applied to the data and the results are illustrated in Table 2. From the estimated factor loadings, the first factor is interpreted as an ability relating to liberal arts, and the second factor as that for sciences. According to the factor contributions $\mathrm{C}(\xi_j \to \boldsymbol{X})$ shown in Table 3, the contribution of factor ξ_2 is about twice as big than that of factor ξ_1 from a view point of entropy, and from the relative contributions $\widetilde{\mathrm{RC}}(\xi_j \to \boldsymbol{X})$, about 30% of variation of manifest variable vector $\boldsymbol{X}$ in entropy is explained by factor ξ_1 and about 60% by factor ξ_2. The relative contribution $\widetilde{\mathrm{RC}}(\boldsymbol{\xi} \to \boldsymbol{X})$ in Table 3 implies about 90% of the entropy of manifest variable vector $\boldsymbol{X}$ is explained by the two factors. On the other hand, in the conventional method, the measured factor contributions of ξ_1 and ξ_2, that is, C_j, are almost equal (Table 4). As discussed in the present paper, the conventional method is intuitive and does not have any logical foundation for multidimensionally measuring contributions of factors to manifest variable vectors. Table 5 decomposes "the contribution of $\boldsymbol{\xi}$ to $\boldsymbol{X}$" into components $\mathrm{C}(\xi_j \to X_i)$. The contribution of ξ_2 to X_5 is prominent compared with the other contributions.

Table 1. Data for illustrating factor analysis.

Subject	Japanese X_1	English X_2	Social X_3	Mathematics X_4	Science X_5
1	64	65	83	69	70
2	54	56	53	40	32
3	80	68	75	74	84
4	71	65	40	41	68
5	63	61	60	56	80
6	47	62	33	57	87
7	42	53	50	38	23
8	54	17	46	58	58
9	57	48	59	26	17
10	54	72	58	55	30
11	67	82	52	50	44
12	71	82	54	67	28
13	53	67	74	75	53
14	90	96	63	87	100
15	71	69	74	76	42
16	61	100	92	53	58
17	61	69	48	63	71
18	87	84	64	65	53
19	77	75	78	37	44
20	57	27	41	54	30

Table 2. Factor loadings of orthogonal factor analysis ($\chi^2 = 0.55$, $df = 1$, $P = 0.45$).

	X_1	X_2	X_3	X_4	X_5
ξ_1	0.60	0.75	0.65	0.32	0.00
ξ_2	0.39	0.24	0.00	0.59	0.92
uniqueness	0.50	0.38	0.58	0.55	0.16

Table 3. Factor contributions based on entropy (orthogonal case).

	ξ_1	ξ_2	Total
$C\left(\xi_j \to X\right)$	3.11	6.23	$9.34 = C(\xi \to X)$
$\widetilde{RC}\left(\xi_j \to X\right)$	0.30	0.60	$0.90 = \widetilde{RC}(\xi \to X)$
$RC\left(\xi_j \to X\right)$	0.33	0.67	1

Table 4. Factor contributions with the conventional method.

	ξ_1	ξ_2	Total
C_j	1.44	1.39	2.83
$\widetilde{RC}_j$	0.29	0.28	0.57
RC_j	0.51	0.49	1

Table 5. Decomposition of factor contribution $C(\xi \to X)$ into $C\left(\xi_j \to X_i\right)$.

	X_1	X_2	X_3	X_4	X_5	Total $= C\left(\xi_j \to X\right)$
ξ_1	0.72	1.49	0.72	0.19	0.00	3.11
ξ_2	0.30	0.15	0	0.63	5.14	6.23
total $= C(\xi \to X_i)$	1.01	1.64	0.72	0.82	5.14	9.34

From the discussion in the previous section, the contributions of factors are flexibly calculated. For example, it is reasonable to divide the manifest variable vector into $X_{(1)} = (X_1, X_2, X_3)$ and

$X_{(2)} = (X_4, X_5)$, because the first sub-vector is related to the liberal arts and the second one to the sciences. First, the contributions of ξ_1 and ξ_2 to $X_{(1)}$ are calculated according to the present method, and the details are given as follows:

$$C\left(\xi_1 \to X_{(1)}\right) = C(\xi_1 \to X_1) + C(\xi_1 \to X_2) + C(\xi_1 \to X_3) = 0.72 + 1.49 + 0.72 = 2.93$$
$$C\left(\xi_2 \to X_{(1)}\right) = 0.30 + 0.15 + 0.00 = 0.45$$

$$C\left(\xi \to X_{(1)}\right) = C\left(\xi_1 \to X_{(1)}\right) + C\left(\xi_2 \to X_{(1)}\right) = 2.93 + 0.45 = 3.38$$
$$\widetilde{RC}\left(\xi \to X_{(1)}\right) = \frac{C\left(\xi \to X_{(1)}\right)}{C\left(\xi \to X_{(1)}\right) + 1} = \frac{3.38}{3.38 + 1} = 0.77 \tag{14}$$

$$\widetilde{RC}\left(\xi_1 \to X_{(1)}\right) = \frac{C\left(\xi_1 \to X_{(1)}\right)}{C\left(\xi \to X_{(1)}\right) + 1} = \frac{2.93}{3.38 + 1} = 0.67 \tag{15}$$

$$\widetilde{RC}\left(\xi_2 \to X_{(1)}\right) = \frac{0.45}{3.38 + 1} = 0.10 \tag{16}$$

$$RC\left(\xi_1 \to X_{(1)}\right) = \frac{C\left(\xi_1 \to X_{(1)}\right)}{C\left(\xi \to X_{(1)}\right)} = \frac{2.93}{2.93 + 0.45} = 0.87 \tag{17}$$

$$RC\left(\xi_2 \to X_{(1)}\right) = \frac{0.45}{2.93 + 0.45} = 0.13 \tag{18}$$

From (14), 77% of the entropy of manifest variable sub-vector $X_{(1)}$ are explained by the two factors, in which 67% of that are explained by factor ξ_1 (15) and 10% by factor ξ_2 (16). From the relative contributions (17) and (18), 87% of the total contribution of the two factors are made by factor ξ_1 and 13% by factor ξ_2.

On the other hand, the contributions of ξ_1 and ξ_2 on $X_{(2)} = (X_4, X_5)$ are calculated as follows:

$$C\left(\xi_1 \to X_{(2)}\right) = C(\xi_2 \to X_4) + C(\xi_2 \to X_5) = 0.19 + 0.00 = 0.19$$
$$C\left(\xi_2 \to X_{(2)}\right) = 0.63 + 5.14 = 5.77$$

$$C\left(\xi \to X_{(2)}\right) = C\left(\xi_1 \to X_{(2)}\right) + C\left(\xi_2 \to X_{(2)}\right) = 0.19 + 5.77 = 5.96$$
$$\widetilde{RC}\left(\xi \to X_{(2)}\right) = \frac{C\left(\xi \to X_{(2)}\right)}{C\left(\xi \to X_{(2)}\right) + 1} = \frac{5.96}{5.96 + 1} = 0.86 \tag{19}$$

$$\widetilde{RC}\left(\xi_1 \to X_{(2)}\right) = \frac{C\left(\xi_1 \to X_{(2)}\right)}{C\left(\xi \to X_{(2)}\right) + 1} = \frac{0.19}{5.96 + 1} = 0.03 \tag{20}$$

$$\widetilde{RC}\left(\xi_2 \to X_{(2)}\right) = \frac{5.77}{5.96 + 1} = 0.83 \tag{21}$$

$$RC\left(\xi_1 \to X_{(2)}\right) = \frac{C\left(\xi_1 \to X_{(2)}\right)}{C\left(\xi \to X_{(2)}\right)} = \frac{0.19}{5.96} = 0.03 \tag{22}$$

$$RC\left(\xi_2 \to X_{(2)}\right) = \frac{5.77}{5.96} = 0.97 \tag{23}$$

From (19), 86% of entropy of manifest variable sub-vector $X_{(2)}$ is explained by the two factors, in which 3% of the entropy are explained by factor ξ_1 (20) and 83% by factor ξ_2 (21). The contribution ratios of the factors to sub-vector $X_{(2)}$ are calculated in (22) and (23). Ninety-seven percent of the entropy was made by factor ξ_2.

Second, factor contributions in an oblique case are calculated. The estimated factor loadings and the correlation matrix of factors based on the covarimin method are shown in Tables 6 and 7, respectively. Based on factor loadings in Table 6, factor ξ_1 is interpreted as an ability for subjects in the liberal arts and factor ξ_2 as an ability for subjects in sciences. The results are similar to those in the orthogonal case mentioned above, because the correlation between the factors is not strong. Table 8 shows the decomposition of $C(\xi \to X)$ based on Theorems 1 and 2. In this case, it is noted that $C(\xi \to X) \neq C(\xi_1 \to X) + C(\xi_2 \to X)$; however, $C(\xi \to X) = \sum_{i=1}^{5} C(\xi \to X_i)$. According to the table, the contributions of ξ_1 and ξ_2 to sub-vectors of manifest variable vector X can also be calculated as in the above orthogonal factor analysis. Table 9 illustrates the contributions of factors on manifest variable vector X. Factor ξ_1 explains 42% of the entropy of X and factor ξ_2 explains 71%.

Table 6. Factor loadings of oblique factor analysis ($\chi^2 = 0.55$, $df = 1$, $P = 0.45$).

	X_1	X_2	X_3	X_4	X_5
ξ_1	0.59	0.77	0.68	0.29	0
ξ_2	0.24	0.00	−0.12	0.52	0.92
uniqueness	0.50	0.41	0.58	0.55	0.16

Table 7. Correlation matrix of factors.

	ξ_1	ξ_2
ξ_1	1	0.315
ξ_2	0.315	1

Table 8. Decomposition of factor contribution $C(\xi \to X)$ into $C\left(\xi_j \to X_i\right)$ (oblique case).

	X_1	X_2	X_3	X_4	X_5	Total = $C\left(\xi_j \to X\right)$
ξ_1	0.90	1.44	0.70	0.37	0.54	3.95
ξ_2	0.37	0.14	0.01	0.68	5.43	6.65
$C(\xi \to X_i)$	1.01	1.44	0.73	0.82	5.43	$C(\xi \to X) = 9.43$

Table 9. Factor contributions based on entropy (oblique case).

	ξ_1	ξ_2	Effect of ξ on X
$C\left(\xi_j \to X\right)$	3.95	6.65	$C(\xi \to X) = 9.43$
$\widetilde{RC}\left(\xi_j \to X\right)$	0.38	0.64	$\widetilde{RC}(\xi \to X) = 0.90$
$RC\left(\xi_j \to X\right)$	0.42	0.71	

6. Discussion

For orthogonal factor analysis models, the conventional method measures factor contributions (effects) by the sums (totals) of squared factor loadings related to the factors (2); however, there is no logical foundation for how they can be interpreted. It is reasonable to measure factor contributions as the effects of factors on the manifest variable vector concerned. The present paper has proposed a method of measuring factor contributions through entropy, that is, applying an entropy-based path analysis approach. The method measures the contribution of factor vector ξ to manifest variable vector X and decomposes it into those of factors ξ_j to manifest variables X_i and/or those to sub-vectors of X. Comparing (2) and (10), for standardization of unique factor variances $\sigma_i^2 = 1$, the present method equals to the conventional method. As discussed in this paper, the present method can be employed in oblique factor analysis as well, and it has been illustrated in a numerical example. The present method has a theoretical basis for measuring factor contributions in a framework of entropy, and it is a

novel approach for factor analysis. The present paper confines itself to the usual factor analysis model. A more complicated model with a mixture of normal factor analysis models [13] is excluded, and a further study is needed to apply the entropy-based method to the model.

Author Contributions: N.E. conceived the study; N.E., M.T., and C.G.B. discussed the idea for measuring factor contribution; N.E. and M.T. proven the theorems in the paper; N.E. and C.G.B. computed the numerical example; N.E. wrote the paper, and the coauthors reviewed it.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Spearman, S. "General-intelligence", objectively determined and measured. *Am. J. Psychol.* **1904**, *15*, 201–293. [CrossRef]
2. Thurstone, L.L. *Vector of Mind: Multiple Factor Analysis for the Isolation of Primary Traits*; The University of Chicago Press: Chicago, IL, USA, 1935.
3. Young, A.G.; Pearce, S. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Quant. Methods Psychol.* **2013**, *9*, 79–94. [CrossRef]
4. Bartholomew, D.J. *Latent Variable Models and Factor Analysis*; Oxford University Press: New York, NY, USA, 1987.
5. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1946**, *27*, 379–423. [CrossRef]
6. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
7. Harry, J. Relative entropy measures of multivariate dependence. *J. Am. Stat. Assoc.* **1989**, *84*, 157–164.
8. Eshima, N.; Tabata, M.; Borroni, G.C.; Kano, Y. An entropy-based approach to path analysis of structural generalized linear models: A basic idea. *Entropy* **2015**, *17*, 5117–5132. [CrossRef]
9. Eshima, N.; Tabata, M. Entropy coefficient of determination for generalized linear models. *Comput. Stat. Data Anal.* **2010**, *54*, 1381–1389. [CrossRef]
10. Eshima, N.; Borroni, C.G.; Tabata, M. Relative importance assessment of explanatory variables in generalized linear models: An entropy-based approach. *Stat. Appl.* **2016**, *16*, 107–122.
11. Everitt, B.S. *An Introduction to Latent Variable Models*; Chapman and Hall: London, UK, 1984.
12. Adachi, K.; Trendafilov, N.T. Some mathematical properties of the matrix decomposition solution in factor analysis. *Psychometrika* **2018**, *83*, 407–424. [CrossRef] [PubMed]
13. Attias, H. Independent factor analysis. *Neural Comput.* **1999**, *11*, 803–851. [CrossRef] [PubMed]