



An Information-Theoretic Alternative to the Cronbach's Alpha Coefficient of Item Reliability

E. Fokoué^{1*} and N. Gündüz²

¹*School of Mathematical Sciences, Rochester Institute of Technology, 98 Lomb Memorial Drive, Rochester, NY 14623, USA.*

²*Department of Statistics, Faculty of Science, Gazi University, Ankara, Turkey.*

Authors' contributions

This work was carried out in collaboration between all authors. From the genesis of the idea through its development all the way to its completion, all the authors worked in synergistic collaboration to create the work featured in this paper. All the authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJMCS/2016/23944

Editor(s):

(1) Nikolaos Dimitriou Bagis, Department of Informatics and Mathematics, Aristotelian University of Thessaloniki, Greece.

Reviewers:

(1) S. Lakshmi, Jeppiaar Engineering College, Chennai, India.

(2) Jen-Yu Shieh, National Formosa University, Taiwan.

Complete Peer review History: <http://sciencedomain.org/review-history/13502>

Received: 29th December 2015

Accepted: 5th February 2016

Published: 2nd March 2016

Original Research Article

Abstract

We propose an information-theoretic alternative to the popular Cronbach alpha coefficient of item reliability. Particularly suitable for contexts in which instruments are scored on a strictly nonnumeric and non ordinal scale, our proposed index is based on functions of the entropy of the empirical distributions defined on the sample space of responses. Our proposed reliability index tracks the Cronbach alpha coefficient uniformly while offering several other advantages discussed in great details in this paper.

Keywords: Cronbach's alpha coefficient; entropy; Likert scale; item reliability; distance; uncertainty; variation of information; probability; internal data consistency.

2010 Mathematics Subject Classification: 62H30, 62H25.

**Corresponding author: E-mail: epfeqa@rit.edu;*

1 Introduction

Suppose that we are given a dataset represented by an $n \times p$ matrix \mathbf{X} whose i th row $\mathbf{x}_i^\top \equiv (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ip})$ denotes the p -tuple of characteristics, with each $\mathbf{x}_{ij} \in \{1, 2, 3, 4, 5\}$ representing the Likert-type level (order) of preference of respondent i on item j . This Likert-type score is obtained by translating/mapping the response levels {Strong Disagree, Disagree, Neutral, Agree, Strongly Agree} into pseudo-numbers $\{1, 2, 3, 4, 5\}$.

Strong Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○
1	2	3	4	5

A usually crucial part in the analysis of questionnaire data is the calculation of Cronbach's alpha coefficient which measures the internal consistency or reliability/quality of the data. Let $X = (X_1, X_2, \dots, X_p)^\top$ be a p -tuple representing the p items of a questionnaire. Initially proposed by [1] and later used and re-explained extensively by thousands of researchers and practitioners like [2] Cronbach's alpha coefficient is a function of the ratio of the sum of the idiosyncratic item variances over the variance of the sum of the items, and is given by

$$\alpha = \left(\frac{p}{p-1} \right) \left[1 - \frac{\sum_{j=1}^p \mathbb{V}(X_j)}{\mathbb{V}(\sum_{\ell=1}^p X_\ell)} \right].$$

The coefficient of Cronbach α will be 1 if the items are all the same and 0 if none is related to another. Because it is depend on the variance of the sum of a group of independent variables and the sum of their variances. If the variables are positively correlated, the variance of the sum will be increased. If the items making up the score are all identical and so perfectly correlated, all the $\mathbb{V}(X_j)$ will be equal and $\mathbb{V}(\sum_{\ell=1}^p X_\ell) = p^2\mathbb{V}(X_j)$, so that $\frac{\sum_{j=1}^p \mathbb{V}(X_j)}{\mathbb{V}(\sum_{\ell=1}^p X_\ell)} = \frac{1}{p}$ and $\alpha = 1$.

The empirical version of Cronbach's alpha coefficient of internal consistency is given by

$$\hat{\alpha} = \left(\frac{p}{p-1} \right) \left[1 - \frac{\sum_{j=1}^p \sum_{i=1}^n \left(\mathbf{x}_{ij} - \frac{1}{n} \sum_{i=1}^n \mathbf{x}_{ij} \right)^2}{\sum_{i=1}^n \left(\sum_{j=1}^p \mathbf{x}_{ij} - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p \mathbf{x}_{ij} \right)^2} \right].$$

Definition 1.1. Let $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ be a dataset with $\mathbf{x}_i^\top = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ip})$. An observation vector \mathbf{x}_i will be called a *zero variation* vector if $\mathbf{x}_{ij} = \mathbf{constant}$, $j = 1, \dots, p$. Respondents with *zero variation* response vectors will be referred to as **single minded** respondents/evaluators.

In fact, *zero variation* responses essentially reduce a p items survey to a single item survey.

Theorem 1.1. Let $X = (X_1, X_2, \dots, X_p)^\top$ be a p -tuple representing the p items of a questionnaire. If X is zero variation, then the Cronbach's alpha coefficient will be equal to 1.

Proof. If $X = (X_1, X_2, \dots, X_p)^\top$ is zero variation, then $X_j = W$ for $j = 1, \dots, p$, and $\sum_{j=1}^p X_j = pW$. As a result, $\sum_{j=1}^p \mathbb{V}(X_j) = p\mathbb{V}(W)$ and $\mathbb{V}(\sum_{j=1}^p X_j) = \mathbb{V}(pW) = p^2\mathbb{V}(W)$. Therefore,

$$\alpha = \left(\frac{p}{p-1} \right) \left[1 - \frac{p\mathbb{V}(W)}{p^2\mathbb{V}(W)} \right] = \left(\frac{p}{p-1} \right) \left[1 - \frac{1}{p} \right] = 1$$

□

We use a straightforward adaptation of the Cronbach's alpha coefficient to measure *respondent reliability*.

Definition 1.2. Let $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ be a dataset with $\mathbf{x}_i^\top = (x_{i1}, x_{i2}, \dots, x_{ip})$. Let the estimated variance of the i th respondent be $\tilde{S}_i^2 = \sum_{j=1}^p (x_{ij} - \bar{x}_i)^2 / (p - 1)$. Let $Z_j = \sum_{i=1}^n x_{ij}$ represent the sum of the scores given by all the n respondents to item j . Our respondent reliability is estimated by

$$\hat{\alpha} = \left(\frac{n}{n-1} \right) \left[1 - \frac{\sum_{i=1}^n \sum_{j=1}^p \left(x_{ij} - \frac{1}{p} \sum_{j=1}^p x_{ij} \right)^2}{\sum_{j=1}^p \left(\sum_{i=1}^n x_{ij} - \frac{1}{p} \sum_{j=1}^p \sum_{i=1}^n x_{ij} \right)^2} \right]$$

Given a data matrix \mathbf{X} , respondent reliability can be computed in practice by simply taking the Cronbach's alpha coefficient of \mathbf{X}^\top , the transpose of the data matrix \mathbf{X} . Let m be the number of *nonzero variation*. If $m \ll p$ and m/n is very small, then respondent reliability will be very poor.

Despite its widespread use of Likert-type data since its creation, Cronbach's alpha coefficient is rigorously speaking not suitable for categorical data for the simple reason that averages on ordinal measurements are often difficult to interpret at best and misleading at worst. For many years researchers working on the clustering of Likert-type inappropriately resorted to average-driven methods like kMeans clustering. Fortunately, there has been a surge of contributions to the clustering of categorical data whereby appropriate methods have been used. At the heart of the clustering of categorical data is the need to define appropriate measure of similarity. Recognizing the possibility to preprocess Likert-type questionnaire data into a collection of estimate probability distributions over the sample spaces of responses, many authors have developed powerful, highly scalable techniques for clustering categorical data, most of them based on information-theoretic concepts like entropy, mutual information, variation of information [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], along with many other distances and measures on probability distributions like the Bhattacharya distance, the Kullback-Leibler divergence and the Hellinger distance just to name a few [14], [15], [16], [17], [18], [19]. In this paper, we use information-theoretic tools and concepts to create several measures of internal consistency of questionnaire data.

2 Information-theoretic Measures of Consistency

Let X_j represent one of the questions on the questionnaire, and consider the n responses represented by $\{x_{1j}, \dots, x_{ij}, \dots, x_{nj}\}$ provided by the n evaluators. Let $\mathbf{v}_j = (v_{j1}, \dots, v_{jk}, \dots, v_{jK})^\top$ denote the vector containing the relative frequencies of each Likert level for question j . With a total of n questionnaires collected, we have

$$\hat{v}_{jk} = \frac{1}{n} \sum_{i=1}^n I(x_{ij} = k), \quad k = 1, 2, \dots, K \quad \text{and} \quad j = 1, 2, \dots, p. \quad (2.1)$$

Using (2.1), one can then form probabilistic vectors $\hat{\mathbf{v}}_j^\top = (\hat{v}_{j1}, \dots, \hat{v}_{jk}, \dots, \hat{v}_{jK})$, for $j = 1, 2, \dots, p$. Each vector $\hat{\mathbf{v}}_j$ essentially represents an approximate probability distribution on the sample space made up of the K response levels. Using this probabilistic representation of each question j , we can compare the variability of each item of the questionnaire using the entropy, specifically

$$H(\hat{\mathbf{v}}_j) = - \sum_{k=1}^K \hat{v}_{jk} \log_2(\hat{v}_{jk}) \quad (2.2)$$

We can imagine a transformation of the $n \times p$ data matrix \mathbf{X} into a probabilistic $p \times K$ counterpart \mathbf{V} where each row represent the approximate probability distribution of the corresponding question (item). The entropy of each question indicates the variability of the answers given by students on that question. For a given course and a given instructor, a small value of this entropy would indicate a greater degree of agreement of his/her student on that item, and therefore suggest a more careful examination of the scores on that item. As far as the relationship between items is concerned, information theory also provides a wealth of measures. The symmetrized Kullback-Leibler divergence given by

$$\text{KL}_2(\mathbf{v}_i, \mathbf{v}_j) = \frac{1}{2} \left\{ \text{KL}(\mathbf{v}_i, \mathbf{v}_j) + \text{KL}(\mathbf{v}_j, \mathbf{v}_i) \right\} = \frac{1}{2} \sum_{k=1}^K \left\{ v_{ik} \log \left(\frac{v_{ik}}{v_{jk}} \right) + v_{jk} \log \left(\frac{v_{jk}}{v_{ik}} \right) \right\},$$

where

$$\text{KL}(\mathbf{v}_i, \mathbf{v}_j) = \sum_{k=1}^K v_{ik} \log \left(\frac{v_{ik}}{v_{jk}} \right) \quad \text{and} \quad \text{KL}(\mathbf{v}_j, \mathbf{v}_i) = \sum_{k=1}^K v_{jk} \log \left(\frac{v_{jk}}{v_{ik}} \right),$$

is usually the default measure used by most authors. The *Kullback-Leibler* divergence is closely related the *mutual information*

$$I(\mathbf{v}_i, \mathbf{v}_j) = \sum_{k=1}^K \left\{ \sum_{l=1}^K \left\{ v_{ik,jl} \log_2 \left(\frac{v_{ik,jl}}{v_{ik} v_{jl}} \right) \right\} \right\},$$

which has been used extensively in machine learning to define a distance known as the *Variation of Information*, and defined by

$$\text{VI}(\mathbf{v}_i, \mathbf{v}_j) = H(\mathbf{v}_i) + H(\mathbf{v}_j) - 2I(\mathbf{v}_i, \mathbf{v}_j).$$

Many other non-information-theoretic similarity and variation measures operating on probabilistic vectors can be used to further investigate several aspects of the categorical data at hand. One that have been extensively used in the machine learning and data mining community is the Bhattacharya distance [14] is given by

$$\text{BC}(\mathbf{v}_i, \mathbf{v}_j) = -\log F(\mathbf{v}_i, \mathbf{v}_j),$$

where

$$F(\mathbf{v}_i, \mathbf{v}_j) = \sum_{k=1}^K \sqrt{v_{ik} v_{jk}},$$

is known as the Bhattacharya coefficient or Fidelity coefficient. The Bhattacharya distance $\text{BC}(\mathbf{v}_i, \mathbf{v}_j)$ measures the overlap between \mathbf{v}_i and \mathbf{v}_j . The Bhattacharya distance has been immensely used in various data mining and machine learning applications [15], [16], [17], [18]. It is interesting to note that the Bhattacharya distance is related to *total variation* measure defined by

$$\Delta(\mathbf{v}_i, \mathbf{v}_j) = \frac{1}{2} \sum_{k=1}^K |v_{ik} - v_{jk}| = \frac{1}{2} \|\mathbf{v}_i - \mathbf{v}_j\|_1$$

where $\|\cdot\|_1$ is the ℓ_1 norm. Another very commonly used distance is the Hellinger distance between \mathbf{v}_i and \mathbf{v}_j is given by

$$\text{Hellinger}(\mathbf{v}_i, \mathbf{v}_j) = \frac{1}{\sqrt{2}} \sqrt{\sum_{k=1}^K (\sqrt{v_{ik}} - \sqrt{v_{jk}})^2} = \frac{1}{\sqrt{2}} \|\sqrt{\mathbf{v}_i} - \sqrt{\mathbf{v}_j}\|_2,$$

where $\|\cdot\|_2$ is the Euclidean norm or ℓ_2 norm, $\sqrt{\mathbf{v}_i} = (\sqrt{v_{i1}}, \dots, \sqrt{v_{iK}})$ and $\sqrt{\mathbf{v}_j} = (\sqrt{v_{j1}}, \dots, \sqrt{v_{jK}})$.

Definition 2.1. Let Q denote an instrument (questionnaire) for which the realized matrix of obtained responses is given by \mathbf{X} with entries $\mathbf{x}_{ij} \in \{1, 2, \dots, K\}$. We propose an information-theoretic measure of the reliability of Q , referred to as the information consistency ratio of Q and given by

$$\varphi = 1 - \frac{\min_{i=1, \dots, n} \{H(\widehat{\mathbf{z}}_i)\}}{\max_{\mathbf{z}} \{H(\mathbf{z})\}} = 1 - \frac{\min_{i=1, \dots, n} \{H(\widehat{\mathbf{z}}_i)\}}{H\left(\frac{1}{K}, \dots, \frac{1}{K}\right)} = 1 - \frac{\min_{i=1, \dots, n} \{H(\widehat{\mathbf{z}}_i)\}}{\log_2(K)}, \quad (2.3)$$

where each $\widehat{\mathbf{z}}_i = \{\widehat{z}_{ik}, k = 1, 2, \dots, K\}$ defines an approximate probability distribution on the sample space of possible responses, and $H(\cdot)$ is the entropy function, with

$$\widehat{z}_{ik} = \frac{1}{p} \sum_{j=1}^p I(\mathbf{x}_{ij} = k) \quad \text{and} \quad H(\widehat{\mathbf{z}}_i) = - \sum_{k=1}^K \widehat{z}_{ik} \log_2(\widehat{z}_{ik}). \quad (2.4)$$

Lemma 2.1. Let \mathbf{z} denote any probability measure defined on some K -dimensional sample space, with each $z_k = \Pr\{E_k\}$, $k = 1, 2, \dots, K$. Let $H(\cdot)$ denote the entropy function, such that for every \mathbf{z} , we have $H(\mathbf{z}) = - \sum_{k=1}^K z_k \log_2(z_k)$. Then

$$\max_{\mathbf{z}} \{H(\mathbf{z})\} = \log_2(K).$$

Proof. Since entropy essentially measures uncertainty (disturbance), the probability measure for which the uncertainty is the largest is the probability measure \mathbf{z}^* in which all the events are equally likely, i.e., $z_k^* = \Pr\{E_k\} = \frac{1}{K}$, $k = 1, 2, \dots, K$.

$$\max_{\mathbf{z}} \{H(\mathbf{z})\} = H(\mathbf{z}^*) = H\left(\frac{1}{K}, \dots, \frac{1}{K}\right) = - \sum_{k=1}^K \frac{1}{K} \log_2\left(\frac{1}{K}\right) = \log_2(K).$$

□

Theorem 2.2. Let Q_0 denote a special questionnaire whose items are all mutually independent (unrelated). Then the corresponding information consistency ratio φ_0 of Q_0 , is such that

$$\lim_{p \rightarrow \infty} \varphi_0 = 0.$$

Proof. With Q_0 denoting a questionnaire whose items that are all mutually independent (unrelated), the matrix of realized responses has entries \mathbf{x}_{ij} that a realization of the discrete uniform distribution on $\{1, 2, \dots, K\}$, or specifically, $\mathbf{x}_{ij} \sim \text{uniform}(1, 2, \dots, K)$. It follows that for each $i = 1, 2, \dots, n$, we must have

$$\lim_{p \rightarrow \infty} \widehat{z}_{ik} = \lim_{p \rightarrow \infty} \left\{ \frac{1}{p} \sum_{j=1}^p I(\mathbf{x}_{ij} = k) \right\} = \frac{1}{K}, \quad k = 1, 2, \dots, K.$$

In other words, given enough questions (items), the empirical proportion of answers will converge to its theoretical counterpart by the law of large number. We therefore have the uniform generation of answers, the limiting distribution

$$\lim_{p \rightarrow \infty} \widehat{\mathbf{z}}_i = \mathbf{z}^* = \left(\frac{1}{K}, \dots, \frac{1}{K}\right).$$

Finally, since all the response distributions will tend to converge to the same maximal measure \mathbf{z}^* , i.e. $\widehat{\mathbf{z}}_i \xrightarrow{\mathcal{D}} \mathbf{z}^*$, for $i = 1, 2, \dots, n$, we must have

$$\min_{i=1, \dots, n} \{H(\widehat{\mathbf{z}}_i)\} \xrightarrow{P} H(\mathbf{z}^*) = \max_{\mathbf{z}} \{H(\mathbf{z})\},$$

and therefore

$$\lim_{p \rightarrow \infty} \varphi_0 = 1 - \frac{\min_{i=1, \dots, n} \{H(\widehat{\mathbf{z}}_i)\}}{\max_{\mathbf{z}} \{H(\mathbf{z})\}} = 1 - \frac{H(\mathbf{z}^*)}{H(\mathbf{z}^*)} = 1 - 1 = 0.$$

□

Theorem 2.3. Let Q_+ denote a special questionnaire whose items are all identical. Then the corresponding information consistency ratio φ_+ of Q_+ , is such that $\lim \varphi_+ = 1$.

Proof. With Q_+ denoting a questionnaire whose items that are all identical, the matrix of realized responses has entries $\mathbf{x}_{ij} = c$, for some constant $c \in \{1, 2, \dots, K\}$. Then for each $i = 1, 2, \dots, n$, there exists $k_+ \in \{1, 2, \dots, K\}$ such that

$$\widehat{\mathbf{z}}_{ik} = \begin{cases} 1 & k = k_+ \\ 0 & k \neq k_+ \end{cases}$$

In other words, with Q_+ , the approximate distributions $\widehat{\mathbf{z}}_i$ of the answers of each respondent are of the form $(1, 0, \dots, 0)$, or $(0, 1, \dots, 0)$ or $(0, 0, \dots, 1)$. Therefore, for Q_+ , we must have $H(\widehat{\mathbf{z}}_i) = 0$, $i = 1, \dots, n$, with the result being $\min_{i=1, \dots, n} \{H(\widehat{\mathbf{z}}_i)\} = 0$, and therefore

$$\varphi_+ = 1 - \frac{\min_{i=1, \dots, n} \{H(\widehat{\mathbf{z}}_i)\}}{\max_{\mathbf{z}} \{H(\mathbf{z})\}} = 1 - \frac{0}{H(\mathbf{z}^*)} = 1 - 0 = 1.$$

□

Definition 2.2. Let Y_i represent the most frequently occurring answer in respondent i 's vector of p answers. It is easy to see that Y_i has the same sample space as each question/item, namely the same Likert scale in our case. Using the random variables Y_i , we can then define $\widehat{\mathbf{w}} = (\widehat{w}_1, \dots, \widehat{w}_k, \dots, \widehat{w}_K)^\top$ in the same manner that we define \mathbf{v}_j earlier. More specifically, we have

$$Y_i = \operatorname{argmax}_{k=1, \dots, K} \left\{ \frac{1}{p} \sum_{j=1}^p I(X_{ij} = k) \right\} \quad \text{and} \quad \widehat{w}_k = \frac{1}{n} \sum_{i=1}^n I(Y_i = k). \quad (2.5)$$

The entropy of $\widehat{\mathbf{w}}$ is given by

$$H(\widehat{\mathbf{w}}) = - \sum_{k=1}^K \widehat{w}_k \log_2(\widehat{w}_k). \quad (2.6)$$

The random variable Y_i is maximal in a set-theoretic sense, and can be thought of as the categorical analogue of the sum of numeric X_j 's. Using $\widehat{\mathbf{w}}$, an alternative definition of the information consistency ratio φ is

$$\varphi = 1 - \frac{\min_{i=1, \dots, n} \{H(\widehat{\mathbf{z}}_i)\}}{H(\widehat{\mathbf{w}})}. \quad (2.7)$$

An even more stringent measure of the information consistency ratio is given by

$$\varphi = 1 - \frac{\max_{i=1, \dots, n} \{H(\widehat{\mathbf{z}}_i)\}}{H(\widehat{\mathbf{w}})}. \quad (2.8)$$

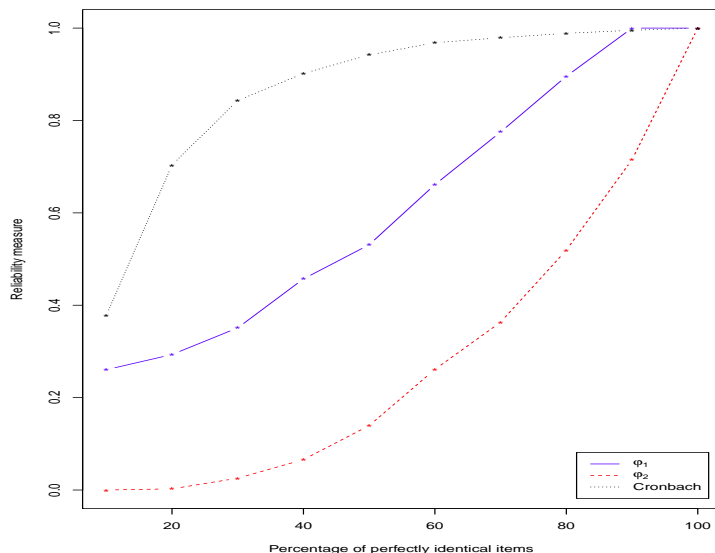


Fig. 1. Comparative curves of φ and Cronbach alpha as measures of internal consistency

3 Demonstration of Properties of φ

We use a simple simulation setup to empirically compare the different measures presented in this paper. We set $p = 50$ and $n = 1000$ and we vary the ratio of perfectly reliable components from 10% to 100% by 10%. For $i = 1, \dots, n$ and $j = 1, \dots, n$, draw the x_{ij} 's uniformly with replacement from $\{1, 2, \dots, K\}$, that is,

$$\text{Draw } x_{ij} \sim \text{uniform}(1, 2, \dots, K).$$

Randomly replace $100c\%$ of the columns of \mathbf{X} with the same column of constant values, where $c \in \{0.1, 0.2, \dots, 0.9, 1\}$. Table 1 shows the simulated values of the information consistency ratio and Cronbach's alpha coefficient for different fractions of reliable components in the instrument. Fig. 1 is a direct pictorial representation of the numbers from Table 1, and we can see that the Cronbach alpha coefficient is less strict than the information consistency ratio.

Table 1. Simulated values of the information consistency ratio and Cronbach's alpha coefficient for different fractions of reliable components in the instrument

Fraction	10	20	30	40	50	60	70	80	90	100
φ_1	0.230	0.270	0.330	0.440	0.520	0.630	0.740	0.900	1.000	1.000
φ_2	0.000	0.000	0.020	0.080	0.140	0.240	0.360	0.520	0.720	1.000
φ_3	0.230	0.270	0.330	0.440	0.520	0.630	0.740	0.900	1.000	1.000
φ_4	0.000	0.000	0.020	0.080	0.140	0.240	0.360	0.520	0.720	1.000
Cronbach	0.380	0.700	0.820	0.910	0.940	0.960	0.980	0.990	1.000	1.000

4 Discussion and Conclusion

We have proposed and developed an information-theoretic measure of internal data consistency and demonstrated via straightforward simulation that it does indeed capture the amount of information potentially contained in the data for the purposes of extracting all kinds of pattern from the data. We have also provided several other measures of similarity over probabilistic vectors that we intend to use to further refine our proposed information consistency ratio . We intend to conduct a larger simulation study to establish our proposed measure on a much stronger footing. We also plan to compare the predictive power of ICR to Cronbachs alpha coefficient on various real and simulated data.

Acknowledgement

Ernest Fokoué wishes to express his heartfelt gratitude and infinite thanks to Our Lady of Perpetual Help for Her ever-present support and guidance, especially for the uninterrupted flow of inspiration received through Her most powerful intercession.

Competing Interests

The authors declare that no competing interests exist.

References

- [1] Cronbach IJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297-333.
- [2] Bland JM, Altman DG. Cronbach's alpha. *BMJ*. 1997;314:572.
- [3] Cover T, Thomas J. *Elements of Information Theory*. Wiley-Interscience; 1991.
- [4] Barbará D, Li Y, Couto J. Coolcat: An entropy-based algorithm for categorical clustering. In *CIKM 02: Proceedings of the Eleventh International Conference on Information and Knowledge Management*, New York, NY, USA. 2002;582-589.
- [5] Huang Z. Extensions to k-means algorithm for clustering large data sets with categorical values. *Data Mining Knowledge Discovery*. 1998;2(3):283-304.
- [6] Guha S, Rastogi R, Shim K. Rock: A robust clustering algorithm for categorical attributes. *Information Systems*. 2000;25(5):345-366.
- [7] San OM, Huynh VN, Nakamori Y. An alternative extension of the k-means algorithm for clustering categorical data. *International Journal of Applied Mathematics and Computer Science*. 2004;14(2):241-247.
- [8] Li T, Ma S, Ogihara M. Entropy-based criterion in categorical clustering. In *ICML 04: Proceedings of the twenty-first international Conference on Machine Learning*, New York, NY, USA, ACM, ACM. 2004;68.
- [9] Chen K, Liu L. The best k for entropy-based categorical data clustering. In *Proceedings of the 17th International Conference on Scientific and Statistical Database Management (SSDBM2005)*, Berkeley, CA, USA. 2005;253-262.
- [10] Li T. A unified view on clustering binary data. *Machine Learning*. 2006;62(3):199-215.
- [11] Meila M. Comparing clusteringsan information based distance. *Journal of Multivariate Analysis*. 2007;98(5):873-895.

- [12] Cai Z, Wang D, Jiang L. K-distributions: A new algorithm for clustering categorical data. In Third International Conference on Intelligent Computing (ICIC 2007), Qingdao, China. 2007;436-443.
- [13] Meila M. Learning theory and kernel machines. Volume 2777 of Lecture Notes in Computer Science, Chapter Comparing Clusterings by the Variation of Information, Springer. 2003;173-187.
- [14] Bhattacharya A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.* 1943;35:99-109.
- [15] Mak B. Phone clustering using the Bhattacharyya distance. *Proceedings of the Fourth International Conference on Spoken Language.* 1996;4:2005-2008.
- [16] Choi E, Lee C. Feature extraction based on the Bhattacharyya distance. *Pattern Recognition.* 2003;36(8):1703-1709.
- [17] Goudail F, Réfrégier P, Delyon G. Bhattacharyya distance as a contrast parameter for statistical processing of noisy optical images. *JOSA A.* 2004;21(7):1231-1240.
- [18] You CH. An SVM Kernel with GMM-supervector based on the Bhattacharyya distance for speaker recognition. *Signal Processing Letters, IEEE.* 2009;16(1):49-52.
- [19] Reyes-Aldasoro CC, Bhalerao A. The Bhattacharyya space for feature selection and its application to texture segmentation. *Pattern Recognition.* 2006;39(5):812-826.

©2016 Fokoué and Gündüz; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

<http://sciencedomain.org/review-history/13502>