Food Quality and Preference 32 (2014) 93–97

Contents lists available at SciVerse ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Significance test of the adjusted Rand index. Application to the free sorting task

El Mostafa Qannari *, Philippe Courcoux, Pauline Faye

LUNAM University, ONIRIS, USC "Sensometrics and Chemometrics Laboratory", Nantes F-44322, France
INRA, Nantes F-44316, France

ARTICLE INFO

Article history:
Received 1 October 2012
Received in revised form 5 March 2013
Accepted 14 May 2013
Available online 30 May 2013

Keywords:
Adjusted rand index
Permutation test
Free sorting task

ABSTRACT

The adjusted Rand index is widely used in connection with the free sorting task to assess the degree of association between two partitions of a set of stimuli. A hypothesis testing framework regarding the significance of this index is set up. It consists in a permutation test which involves the simulation of a large number of partitions from the original two partitions being compared. However, since this strategy of analysis may be time consuming, an alternative method is also proposed. It is based on statistical properties regarding the distribution of the values generated by the procedure of permutation. The two strategies of analysis are illustrated and compared on the basis of free sorting data and a simulation study.

1. Introduction

Numerous studies, particularly in sensory and consumer science, indicated that free sorting task is a simple and efficient way of assessing similarities among a set of stimuli (Faye et al., 2004; King, Cliff, & Wall, 1998; Lawless, Sheng, & Knoops, 1995). This task consists in a categorization procedure which relies on the natural human tendency to grouping things into categories as a cognition process for learning and apprehending the complexity of the world. More precisely, the subjects are presented with a set of stimuli and instructed to sort them in as many groups as they believe it necessary, considering that stimuli in the same group are perceived as similar.

For the statistical treatment of the data collected in the course of fee sorting task, several strategies of analysis have been proposed ranging from MDS techniques (Faye et al., 2006) to methods of analysis pertaining to multiple correspondence analysis (Cadoret, Lê, & Pagès, 2009; Qannari, Cariou, Teillet, & Schlich, 2010; Takane, 1982; Van der Kloot & Van Herk, 1991). Another method of analysis which is worth mentioning is DISTATIS (Abdi, Valentin, Chollet, & Chrea, 2007). All these methods seek a representation space of low dimension in order to depict the similarities among the stimuli. Courcoux, Faye, and Qannari (2012) discuss alternative methods of analysis of sorting data. This consists in segmenting the subjects who take part in the experiment and, for each segment, determining a compromise partition of the stimuli that stands as a sort of average partition for the subjects under consideration. In such studies, the question of assessing the similarity between the partitions given by two subjects or the partitions associated with two groups of subjects arises. Several similarity measures between two partitions have been proposed in the literature (Albatineh, Niewiadomska-Bugaj, & Mihalko, 2006; Fowlkes & Mallows, 1983). Among these similarity measures, the Adjusted Rand index (ARI) (Hubert & Arabie, 1985) is widely used (Faye et al., 2004; Soufflet et al., 2004; Faye et al., 2006; Parizet, Guyader, & Nosulenko, 2008; Parizet & Koehl, 2012).

We aim at setting up a hypothesis testing framework to assess whether a given ARI is significantly larger than 0. For this purpose, we recourse to a permutation test based on randomly simulating a large number of pairs of partitions and, for each pair, we compute the ARI. Thereafter, the decision as to whether the actual (observed) ARI is significantly larger than zero is taken on the basis of the distribution of these simulated values. However, since this procedure involves intensive computations it can be time consuming, particularly when the number of products is large and when there are several pair-wise comparisons to perform. As an alternative, we propose a straightforward test which consists in approaching the distribution of the simulated values by a normal distribution where the mean and standard deviation are derived from theoretical developments by Lancaster (1969). The two approaches of hypothesis testing (*i.e.* simulation and approximation by a normal distribution) are illustrated and compared on the basis of data from a free sorting experiment and from a simulation study.

* Corresponding author. Fax: +33 02 51 78 54 38.
E-mail address: elmostafa.qannari@oniris-nantes.fr (E.M. Qannari).

0950-3293/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodqual.2013.05.005

2. Material and methods

2.1. The Rand and adjusted Rand indices

Let us consider two partitions of the same n stimuli. For instance, these could be the partitions of the same stimuli operated by two subjects in the course of a free sorting task. The Rand index (Rand, 1971) is defined by:

$$\text{Rand} = \frac{a+d}{T}$$

where a is the number of pairs of stimuli that are placed in the same group in both partitions, d is the number of pairs that are not placed in the same group in both partitions and, T is the total number of pairs of stimuli (*i.e.* $T = \frac{n(n-1)}{2}$). Obviously, The Rand index lies between 0 and 1. In particular, it is equal to 1 in case of perfect agreement (*i.e.* identity of the two partitions).

Notwithstanding its intuitive appeal, The Rand index suffers from several pitfalls. In particular, it highly depends on the number of groups in the two partitions and the number of stimuli at hand. So much so that, in some situations, it can take high values even for two random partitions (Youness & Saporta, 2004; Courcoux et al., 2012; Santos & Embrechts, 2009). In order to cope with this problem, ARI was proposed as a form of the Rand index that is corrected for the grouping of the stimuli by chance. The general form of ARI is:

$$\text{ARI} = \frac{\text{Rand} - \text{Expected (Rand)}}{1 - \text{Expected (Rand)}}$$

where Expected (Rand) is the mean value of Rand under the hypothesis that the two partitions are independent, subject to the assumption that the number of stimuli in the groups are fixed and that a generalized hypergeometric distribution is considered as a model of randomness (Albatineh, 2010; Fowlkes & Mallows, 1983; Hubert & Arabie, 1985). More details about ARI are given in the appendix.

From the expression of ARI, it follows that this index ranges between −1 and +1. It takes the value 1 in case of perfect agreement between the two partitions and takes the value 0 when the two partitions are independent.

2.2. Permutation test

Once ARI between two given partitions A and B is computed, the question that often emerges is whether this index is significantly larger than 0. We propose to recourse to a simulation study to address this issue. In the following, we shall denote by ARI_0 the actual (observed) value for ARI between the two partitions A and B.

Denoting by H_0 the hypothesis that stipulates the absence of association between the two partitions at hand, the simulation study consists in setting up an (empirical) distribution under H_0 of a statistic tightly related to ARI.

The simulation study runs as follows:

(i) Randomly simulate a large number (1000, say) of pairs of partitions with the following constraints: for each pair, the first (respectively, second) partition should have as many groups as A (respectively, B) and, moreover, the groups in this partition should have as many elements as those of A (respectively, B). This entails that the original structure of the partitions remains unchanged (*i.e.* the same number of groups and the same number of elements in the various groups). As a matter of fact, this procedure of simulation amounts to a permutation test since in each partition, we randomly shuffle the group labels (*i.e.* group names) of the stimuli.

(ii) For each pair of partitions thus generated, we compute ARI.

(iii) We compute the average (m) and the variance (v) of all the values thus obtained.

(iv) All the ARI values are transformed into $\text{NARI} = \frac{\text{ARI}-m}{\sqrt{v}}$.

(v) The actual value ARI_0 is also transformed into $\text{NARI}_0 = \frac{\text{ARI}_0-m}{\sqrt{v}}$.

(vi) On the basis of the values of NARI thus obtained, we draw a frequency histogram which stands as a probability distribution for the statistic NARI under the null hypothesis, H_0. More precisely, if we choose a significance level equal to α (*e.g.* $\alpha = 5\%$), we can identify a threshold value, γ_0, which is the value such that only a proportion equal to α of NARI values are larger than γ_0. Thereafter, the association between the two partitions A and B will be considered as significant if the observed value NARI_0 is larger than γ_0. Alternatively, the simulated values from step (iv) can be ranked in an increasing order and a p-value is computed as the proportion of values larger than the observed value NARI_0. The decision as to whether ARI between the two partitions A and B is significant (*i.e.* rejection of H_0) or not is taken by assessing whether the p-value thus obtained is smaller than α or not.

2.3. Alternative test

The drawback of the permutation test outlined above is that since it is based on a simulation study, it is likely to be time consuming especially when there are several pair-wise comparisons to perform as in the case of a free sorting task performed by a panel of subjects. Indeed, since the distribution of the simulated NARI depends on the configurations of the two partitions being compared (*i.e.* number of groups and number of stimuli in the groups), one should run as many simulations studies as there are different configurations that is, practically, as many pairs of partitions to be compared. To remedy this problem, we propose an alternative approach based on theoretical results derived from properties demonstrated by Lancaster (1969). These results concern the expected value and the variance of the distribution of ARI under the hypothesis that the two partitions are independent, subject to the assumption that the number of stimuli in the groups are fixed and that a generalized hypergeometric distribution is considered as a model of randomness (Fowlkes & Mallows, 1983; Hubert & Arabie, 1985). By construction of ARI, the theoretical value of the mean is equal to 0 since, as we stated above, ARI is a form of the Rand index corrected to account for agreement by chance. The expression of the variance which we denote by v is given in the appendix. This value depends on the total number of stimuli, the number of groups in the two partitions at hand and the number of stimuli in the various groups.

From these (theoretical) results, we propose a simplified test of hypotheses concerning the significance of the association between two partitions. The rationale behind this test is to approach the distribution of the statistic $\text{NARI} = \frac{\text{ARI}}{\sqrt{v}}$ (Normalized ARI) by a normal distribution with mean 0 and unit standard deviation. More precisely, let us denote by ARI_0 the observed value for ARI between two partitions A and B. The observed value for the statistic NARI is given by $\text{NARI}_0 = \frac{\text{ARI}_0}{\sqrt{v}}$. Thereafter, a p-value can be found from the normal distribution. If the calculated p-value is below the chosen significance level α, then the null hypothesis (*i.e.* independence of the two partitions) is rejected in favor of the alternative hypothesis.

3. Illustration

3.1. Testing the association between two partitions

The data are extracted from a case study involving the assessment of 16 wine aromas by a panel of consumers using a free sort-