Springer Proceedings in Mathematics & Statistics

Roger E. Millsap
L. Andries van der Ark
Daniel M. Bolt
Carol M. Woods *Editors*

New Developments in Quantitative Psychology

Presentations from the 77th Annual Psychometric Society Meeting

Springer

Springer Proceedings in Mathematics & Statistics

Volume 66

For further volumes:
http://www.springer.com/series/10533

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of select contributions from workshops and conferences in all areas of current research in mathematics and statistics, including OR and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

Roger E. Millsap • L. Andries van der Ark
Daniel M. Bolt • Carol M. Woods
Editors

New Developments in Quantitative Psychology

Presentations from the 77th Annual Psychometric Society Meeting

Springer

Editors
Roger E. Millsap
Department of Psychology
Arizona State University
Tempe, AZ, USA

L. Andries van der Ark
Department of Methodology and Statistics
Tilburg University
Tilburg, The Netherlands

Daniel M. Bolt
Department of Educational Psychology
University of Wisconsin
Madison, WI, USA

Carol M. Woods
Department of Psychology
University of Kansas
Lawrence, KS, USA

ISSN 2194-1009 ISSN 2194-1017 (electronic)
ISBN 978-1-4614-9347-1 ISBN 978-1-4614-9348-8 (eBook)
DOI 10.1007/978-1-4614-9348-8
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2014930294

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume represents presentations given at the 77th annual meeting of the Psychometric Society, held at the Cornhusker Hotel in Lincoln, Nebraska, during July 9–12, 2012. The annual meeting of the Psychometric Society typically attracts participants from around the world, and the 2012 conference was no exception. Attendees came from more than 15 different countries, with 149 papers being presented, along with 50 poster presentations, three workshops, two keynote speakers, six state-of-the-art speakers, five invited presentations, and seven invited symposia. A full list of the conference presentation titles can be found in the January 2013 issue of *Psychometrika*, pp. 188–201. We thank the local organizer Ralph de Ayala, along with his staff and students, for hosting a successful conference.

The idea for the present volume began with the recognition that many of the useful ideas presented at the conference do not become available to a wider audience unless the authors decide to seek publication in one of the quantitative journals. This volume provides an opportunity for the presenters to make their ideas available to the wider research community more quickly, while still being thoroughly reviewed. The 31 chapters published here address a diverse set of topics, including item response theory, reliability, test design, test validation, response styles, factor analysis, structural equation modeling, categorical data analysis, longitudinal data analysis, test equating, and latent score estimation. For the published chapters, we asked the authors to include the ideas presented in their conference papers, and we also gave them the opportunity to expand on these ideas in the published chapters. Psychological measurement is playing a larger role internationally than ever before, not only in educational applications but also in medicine and neuroscience. It is important that this expanding role be supported by rigorous and thoughtful research. We thank all of the chapter authors for their fine contributions to this volume. We hope that the contents of this volume will stimulate wider interest in psychometric research, both theoretical and applied.

Tempe, AZ, USA — Roger E. Millsap
Madison, WI, USA — Daniel M. Bolt
Tilburg, The Netherlands — L. Andries van der Ark
Lawrence, KS, USA — Carol M. Woods

Contents

A Nonparametric Ability Measure

Nan L. Kong

1 Introduction

Before we define an ability measure, we need to make clear about the concept of measure. In this section, we look into several well-defined measures from which we try to find the property in common across these measures. We believe that the ability measure, which is the topic of this paper, should also be defined on the basis of this common property.

It is well known that the area of a rectangle is measured by the product of its length and width. For example, for a rectangle with length of 2 and width of 1, the area can be directly measured with $2 = 2 \times 1$. Actually, this rectangle can also be measured indirectly: (i) split this rectangle into two unit squares with both length and width equal to 1; (ii) the areas of these two unit squares are measured with $1 = 1 \times 1$; (iii) make summation of these two area measures in (ii) with $2 = 1 + 1$. The summation in (iii) is the "indirect" measure of the area of the rectangle with length of 2 and width of 1. As we can see, both "direct" and "indirect" area measures on this rectangle produce the same value which is 2 in this example. The relation between "direct" and "indirect" area measures is mathematically expressed by $2 \times 1 = 1 \times 1 + 1 \times 1$. The left-hand side of this equation corresponds to "direct" measure while the right-hand side corresponds to "indirect" measure. Generally, for the same area, both "direct" and "indirect" measures must produce the same value—this is called *additivity* according to the measure theory (Halmos 1974). In the same example, if we measure the area of the rectangle by summation of length and width, instead of product of its length and width, with the steps in (i)–(iii), we will receive two different values for the "direct" measure, which is $3 = 1 + 2$, and the "indirect" measures which is $4 = (1 + 1) + (1 + 1)$. Obviously, with summation of length and

N.L. Kong (✉)
Educational Testing Service, 270 Hampshire Dr., Plainsboro, NJ 08536, USA
e-mail: nankg@yahoo.com

R.E. Millsap et al. (eds.), *New Developments in Quantitative Psychology*,
Springer Proceedings in Mathematics & Statistics 66, DOI 10.1007/978-1-4614-9348-8_1,

width, the area of the rectangle is measured in a wrong way—the way that has no additivity. Any measure without additivity is similar to measuring area of rectangle by summation of its length and width.

In measure theory (Halmos 1974), a set function is a function whose domain of definition is a class of sets. An extended real-valued set function $\mu(.)$ defined on a class S of sets is additive if, whenever $E \in S$, $F \in S$, $E \cup F \in S$, and $E \cap F = \emptyset$, then $\mu(E \cup F) = \mu(E) + \mu(F)$. For the measure of the rectangle area, the class S contains all rectangles (each rectangle is a set of points) and $\mu(.)$ is defined by the product of its length and width.

The next well-defined measure is called probability which measures randomness (Hays 1970). If two events A and B are exclusive, we have

$$Prob(A \cup B) = Prob(A) + Prob(B). \tag{1}$$

Equation (1) is called *additivity*.

In information theory, the entropy (Shannon 1948; Wiener 1948) is defined to measure the uncertainty in the random variables. One of the entropy fundamental properties is the following equation:

$$H(X,Y) = H(X) + H(Y) - I(X,Y), \tag{2}$$

where X and Y are two categorical random variables; $H(X)$ and $H(Y)$ are the entropies for X and Y, respectively; $H(X,Y)$ is the entropy of X and Y; $I(X,Y)$ is the mutual information among X and Y.

If X and Y are independent from each other, which implies $I(X,Y) = 0$, Eq. (2) becomes

$$H(X,Y) = H(X) + H(Y). \tag{3}$$

Equation (3) is called *additivity*.

Unlike Shannon's entropy, Fisher information (Fisher 1922 and 1925) is defined to measure the parameter(s)' information given random variable(s). If random variables X and Y are independent, we have

$$I_{X,Y}(\theta) = I_X(\theta) + I_Y(\theta), \tag{4}$$

where $I_{X,Y}(\theta)$ is the Fisher information given X and Y; $I_X(\theta)$ and $I_Y(\theta)$ are the Fisher information given X and Y, respectively. θ is the parameter(s).

Equation (4) is called *additivity*.

So far, we have looked into the theoretical structures for several well-defined measures. All of these structures reveal the same property—*additivity* as shown in (1), (3) and (4). We believe that the *additivity* is the general property for a measure. The purpose of this paper is to study a new ability measure and, therefore, it is requested that this ability measure be of the property of the additivity. In the next section, an ability measure is defined and studied according to the additivity.

2 A Nonparametric Ability Measure

In testing and psychometrics, the term ability means the knowledge, skills, or other characteristics of a test taker measured by the test. A test question, with any stimulus material provided with the test question, and the response choice or the scoring rules, is called an item. Items that are scored in two categories - right (R) or wrong (W) - are referred to as dichotomous items. In this section, the test taker's ability will be measured on the basis of a test consisting of a set of dichotomous items. For a test consisting of I items, let X_i be the item-score variable for the item i ($i = 1,\ldots,I$), with realization $X_i \in \{W,R\}$. Also, we suppose that a respondent answers $L(0 \le L \le I)$ items correctly, then these correctly answered items are indicated by $i_1,\ldots,i_l,\ldots,i_L$. For example, suppose an item-response vector of RRWWWR, then $I = 6, L = 3, i_1 = 1, i_2 = 2$, and $i_3 = 6$. The probability of right response for i_1 is denoted by $P(X_{i_1} = R)$ and, the probability of right responses for both i_1 and i_2 is denoted by $P(X_{i_l} = R, X_{i_2} = R)$, etc.

Definition 1. The ability with right (R) response(s) for items i_l ($l = 1,\ldots,L; L \ge 1$) is defined as

$$\theta(i_1,\ldots,i_l,\ldots,i_L) = -ln(P(X_{i_1} = R,\ldots,X_{i_l} = R,\ldots,X_{i_L} = R)).(L \ge 1) \qquad (5)$$

In (5), $\theta(i_1,\ldots,i_l,\ldots,i_L)$ is called the measure of the ability with right (R) response(s) for the items $i_l(l = 1,\ldots,L)$. We also request that the examinee's ability be measured as zero if this examinee does not respond to any item correctly, i.e. $L = 0$ in (5).

In Definition 1, only the probabilities on correctly responded items are used for measuring abilities, some probabilities such as those for incorrectly responded items are not shown in (5). Because the probabilities on any combinations of the correctly responded items and the incorrectly responded items can be fully expressed by the probabilities on those correctly responded items, the probabilities on correctly responded items have fully represented all of the information associated with the joint probabilities. Therefore, the ability measure in Definition 1 has lost nothing in terms of the information associated with the joint probabilities.

If items $i_1,\ldots,i_L$ are (jointly) independent, the following equation can be obtained directly from Definition 1 and shows that the ability measure in Definition 1 is additive

$$\theta(i_1,\cdots,i_L) = \theta(i_1) + \cdots + \theta(i_L). \qquad (6)$$

As we can see in Eq. (6) that, if the items are jointly independent, the measure of examinee's total ability with right responses on all these items is the summation of the measures of the examinee's abilities with right responses on each of these items. The additivity in Eq. (6) implies that the summation of the ability measures on subscales can be the total ability measure if and only if these subscales are jointly independent. For the case that the items are not jointly independent, not only the ability measure on each subscale but also the interactions among the items play the roles in total ability measure. In Sect. 4, the total ability measure will be studied in more detail.

Corollary 1.

$$0 \leq \theta(i_1, \cdots, i_L) \leq +\infty. \tag{7}$$

Proof. This is obvious from Definition 1.

Corollary 2.

$$\theta(i_1, \cdots, i_L) = 0 \iff P(X_{i_l} = R, \cdots, X_{i_L} = R) = 1 \tag{8}$$

Proof. This is obvious from Definition 1.

Corollary 3.

$$\theta(i_1, \cdots, i_L) = +\infty \iff P(X_{i_l} = R, \cdots, X_{i_L} = R) = 0 \tag{9}$$

Proof. This is obvious from Definition 1.

As shown in Corollary 1, the ability measure defined in (5) is nonnegative which implies the total ability measure is always greater than or equal to the ability measure on each subscale according to the additivity. Because the minus sign has no meaning in the ability measure, the additivity requests that the ability measure be nonnegative (generally, the measure theory always requests that a measure be nonnegative).

Now, assume that $0 < M \leq L$, there is

$$\begin{aligned}
&\theta(i_1, \cdots, i_M) = -ln(P(X_{i_l} = R, \cdots, X_{i_M} = R)) \\
&\leq -ln(P(X_{i_l} = R, \cdots, X_{i_M} = R) \\
&\quad \times P(X_{i_{M+1}} = R, \cdots, X_{i_L} = R | X_{i_l} = R, \cdots, X_{i_M} = R)) \\
&= -ln(P(X_{i_l} = R, \cdots, X_{i_L} = R)) = \theta(i_1, \cdots, i_L)
\end{aligned}$$

Therefore, the following theorem is obtained:

Theorem 1. *For* $0 < M \leq L$,

$$\theta(i_1, \cdots, i_M) \leq \theta(i_1, \cdots i_L) \tag{10}$$

Theorem 1 is another fundamental property of the ability measure: the measure of the ability associated with subset of all correctly responded items is no greater than the measure of the ability associated with all correctly responded items, i.e. the measure of the ability associated with subscale can not be greater than the measure of its total ability.

In summary, the ability measure defined in (5) has the following properties: (a) Additivity (if the items are independent) as shown in Eq. (6). (b) The ability measure is nonnegative. Therefore, the total ability measure is greater than or equal to the ability measure on each subscale. (c) The ability measures with the same response patterns are the same (this is obvious by Definition 1). (d) The ability measure on a response pattern is greater than or equal to the ability measure on the subset of its response pattern (Theorem 1). (e) The ability measure is determined by the difficulties of the items and the interactions among those items. The more difficult and more jointly independent items cause higher ability measure. (f) The ability measure in Definition 1 has no specific parametric structure. Therefore, the ability measure in Definition 1 has no those assumptions or limitations associated with the specific parametric structure. (g) The ability measure is defined with the joint probability of the items in a given test and all of the response vectors out of these items are utilized for measuring ability, therefore, the ability is measured with full information for given joint probabilities.

In the next two sections, the following properties of the ability measure defined in (5) will be studied: (h) With the additivity, it is possible to measure the shared ability and unique ability. Generally speaking, an examinee's ability consists of two parts: the unique part that belongs to the examinee and the part shared with others. (i) The total ability measure and the ability measures on subscales are related to the additivity. Therefore, the interactive structures of the total ability and those abilities associated with the subscales can be mathematically expressed.

3 Shared Ability Measure and Conditional Ability Measure

Because the ability measure in Definition 1 has the property of additivity, it is possible to measure the shared ability among the correctly responded items and unique ability of each correctly responded item.

Definition 2. The shared ability among correctly responded items i_1 and i_2 is measured with

$$\theta(i_1 * i_2) = \theta(i_1) + \theta(i_2) - \theta(i_1, i_2), \tag{11}$$

where $\theta(i_1)$, $\theta(i_2)$, and $\theta(i_1, i_2)$ are defined in Definition 1.

According to Definitions 1 and 2, the following equation can be obtained:

$$\theta(i_1 * i_2) = -ln \frac{P(X_{i_1} = R)P(X_{i_2} = R)}{P(X_{i_1} = R, X_{i_2} = R)} \tag{12}$$

By (12), it is obvious that $\theta(i_1 * i_2) = \theta(i_2 * i_1)$.

The following theorem offers a sufficient and necessary condition for no shared ability between two items i_1 and i_2.

Theorem 2.

$$\theta(i_1 * i_2) = 0 \iff i_1 \; and \; i_2 \; are \; independent.$$

Proof. Let X_{i_1} and X_{i_2} be the item-score variables of the items i_1 and i_2. By Definition 1,

$$\theta(i_1) = -ln(P(X_{i_1} = R), \tag{13}$$

$$\theta(i_2) = -ln(P(X_{i_2} = R), \tag{14}$$

$$\theta(i_1, i_2) = -ln(P(X_{i_1} = R, X_{i_2} = R)). \tag{15}$$

Therefore, X_{i_1} and X_{i_2} are independent if and only if

$$\theta(i_1, i_2) = \theta(i_1) + \theta(i_2)$$

By Eq. (11), we have

$$\theta(i_1 * i_2) = 0$$

This is the proof of Theorem 2.

In concept, the shared ability is closer to the concept of interaction between those items associated with different respondents or subscales. The stronger association between those items implies that the more abilities are shared. For example, if two items are identical, the shared ability is the same as the ability associated with each of those items. Another extreme case is that, if two items are independent, the shared ability is zero. The shared ability is also related to the redundant or overlapped information among the items, i.e. the items could be heavily similar to each other in which the scope for those items to cover for testing could be limited. Therefore, the shared ability among the different items should not be too big.

Unlike the ability measure in Definition 1 which is nonnegative, the shared ability measure in Definition 2 can be negative. If an examinee with correct response on one item tends to correctly respond to another item, this examinee has positive shared ability among these two items. If an examinee with correct response on one item tends to wrongly respond to another item, this examinee has negative shared ability among these two items. In practice, for most of cases, the shared ability is positive. The negative shared ability only happens for two items associated with the exclusive abilities.

Definition 3. The unique or conditional ability with i_1 given i_2 is measured with

$$\theta(i_1|i_2) = -lnP(X_{i_1} = R|X_{i_2} = R). \tag{16}$$

Corollary 4.

$$\theta(i_1, i_2) = \theta(i_2) + \theta(i_1|i_2) \tag{17}$$

Proof. The proof is obvious from Definitions 1 and 3 with noting that:

$$\theta(i_1|i_2) = -ln(P(X_{i_1} = R|X_{i_2} = R)) = -ln(P(X_{i_1} = R, X_{i_2} = R)) + ln(P(X_{i_2} = R))$$

Corollary 5.

$$\theta(i_1 * i_2) = \theta(i_1) - \theta(i_1|i_2) \tag{18}$$

Proof. The proof is obvious from Definition 2 and Corollary 4.

The unique or conditional ability $\theta(i_1|i_2)$ measures the part of the ability with i_1, but exclusive of i_2, that is, $\theta(i_1|i_2)$ measures the unique ability associated with i_1 out of the ability associated with i_1 and i_2. The following equation, which can be proved with Corollaries 4 and 5, describes the relation among total ability, shared ability, and unique ability:

$$\theta(i_1,i_2) = \theta(i_1 * i_2) + \theta(i_1|i_2) + \theta(i_2|i_1). \tag{19}$$

In (19), the $\theta(i_1,i_2)$ is decomposed into three parts—the shared ability associated with i_1 and i_2, the unique ability associated with i_1 with exclusive of the ability associated with i_2, and the unique ability associated with i_2 with exclusive of the ability associated with i_1. Equation (19) is also available in probability and entropy:

$$P(A \cup B) = P(A \cap B) + P(A \cap B^c) + P(B \cap A^c),$$
$$H(X,Y) = I(X,Y) + H(X|Y) + H(Y|X),$$

where A and B are events; A^c and B^c are the events "not A" and "not B". X and Y are two random variables; $H(X,Y)$ is the entropy of X and Y; $H(X)$ and $H(Y)$ are the entropies for X and Y, respectively; $H(X|Y)$ is the conditional entropy of X given Y; $I(X,Y)$ is the mutual information among X and Y.

Theorem 3.

$$\theta(i_1 * i_2) \leq \theta(i_1) \tag{20}$$

Proof.

$$P(X_{X_{i_2}} = R) \geq P(X_{i_1} = R, X_{i_2} = R) \iff ln\frac{P(X_{i_2} = R)}{P(X_{i_1} = R, X_{i_2} = R)} \geq 0$$
$$\iff -ln\frac{P(X_{i_1} = R, X_{i_2} = R)}{P(X_{i_1} = R)P(X_{i_2} = R)} \leq -lnP(X_{i_1} = R)$$
$$\iff \theta(i_1 * i_2) \leq \theta(i_1).$$

This is the proof of Theorem 3.

The measure of the shared ability associated with i_1 and i_2 in Definition 2 can be extended into the measure of the shared ability associated with i_1, i_2, $\cdots$, i_L which is denoted by $\theta(i_1 * \cdots * i_L)$. Without loss of generality, $\theta(i * i_2 * i_3)$ can be defined by:

$$\begin{aligned}\theta(i_1 * i_2 * i_3) = {} & \theta(i_1) + \theta(i_2) + \theta(i_3) - \theta(i_1, i_2) \\ & - \theta(i_1, i_3) - \theta(i_2, i_3) + \theta(i_1, i_2, i_3).\end{aligned} \tag{21}$$

Obviously, according to (21), (joint) independence among i_1, i_2, and i_3 implies that $\theta(i_1 * i_2 * i_3) = 0$. Similar to $\theta(i_1 * i_2)$, $\theta(i_1 * i_2 * i_3)$ can be negative, but the interpretation for this is more complicated. Roughly speaking, $\theta(i_1 * i_2 * i_3)$ is the interactive ability contribution by i_1, i_2, and i_3 to the total ability $\theta(i_1, i_2, i_3)$.

4 Total Ability and Abilities Associated with Subscales

Given the item responses $i_1 \ldots i_L$ answered correctly by a respondent, the examinees' abilities can be measured according to (5). The ability measured by (5) is called the overall or total ability because it is measured by all correctly answered items. In case that those correctly answered item responses $i_1 \ldots i_L$ contain several subscales in which each subscale is associated with a subset of $\{i_1 \ldots i_L\}$, we need to measure the examinees' abilities on the basis of each subscale. First, let us look into the case of two subscales: S_1 and S_2 which S_1 is associated with the subset $\{i_{j_1}, \ldots, i_{j_M}\}$ and S_2 is associated with the subset $\{i_{k_1}, \ldots, i_{k_N}\}$ where $M \le L$ and $N \le L$. Here the intersection of $\{i_{j_1}, \ldots, i_{j_M}\}$ and $\{i_{k_1}, \ldots, i_{k_N}\}$ may not be empty set $\emptyset$, that is, some items may be associated with both S_1 and S_2. We also assume that $\{i_{j_1}, \ldots, i_{j_M}\} \cup \{i_{k_1}, \ldots, i_{k_N}\} = \{i_1 \ldots i_L\}$.

Without loss of generality, the total ability and the abilities associated with the subscales S_1 and S_2 are measured with

$$\theta(Total) = -ln(P(X_{i_1} = R, \cdots, X_{i_L} = R)), \tag{22}$$

$$\theta(S_1) = -ln(P(X_{i_{j_1}} = R, \cdots, X_{i_{j_M}} = R)), \tag{23}$$

$$\theta(S_2) = -ln(P(X_{i_{k_1}} = R, \cdots, X_{i_{k_N}} = R)). \tag{24}$$

Here X_i is the item-score variable for the item i. Because $\theta(S_1)$ and $\theta(S_2)$ in (23) and (24) are defined with the subsets $\{i_{j_1}, \ldots, i_{j_M}\}$ and $\{i_{k_1}, \ldots, i_{k_N}\}$ out of total correctly answered items $\{i_1 \ldots i_L\}$, the $\theta(S_1)$ and $\theta(S_2)$ are also called marginal measures of the abilities associated with S_1 and S_2.

Similar to Definition 2, we can define the measure for the shared ability associated with S_1 and S_2.

Definition 4. The shared ability associated with S_1 and S_2 is measured with

$$\theta(S_1 * S_2) = \theta(S_1) + \theta(S_2) - \theta(S_1, S_2), \tag{25}$$

where

$$\theta(S_1,S_2) = \theta(Total) = -ln(P(X_{i_1} = R,\cdots,X_{i_L} = R)). \tag{26}$$

Equivalently, by Definition 4

$$\theta(Total) = \theta(S_1) + \theta(S_2) - \theta(S_1 * S_2). \tag{27}$$

Equation (27) expresses the relation among the measures of the total ability and the abilities associated with the S_1 and S_2. From Definition 4, it is obvious that, if S_1 and S_2 are independent, the measure of the total ability is the summation of the measures of the abilities associated with S_1 and S_2, i.e. $\theta(Total) = \theta(S_1) + \theta(S_2)$. Also, similar to (12), $\theta(S_1 * S_2)$ can be negative in case that the abilities associated with S_1 and S_2 are exclusive from each other.

In Eq. (22), some items may be shared by both S_1 and S_2. Obviously, these shared items contribute the relation between S_1 and S_2 (the items which are not shared by S_1 and S_2 also contribute the relation between S_1 and S_2 because those not-shared items may be related across the different subscales) and relation between S_1 and S_2 determines $\theta(S_1 * S_2)$ in Eq. (27). Therefore, the total ability measure is affected by the shared items through their interactive ability measure $\theta(S_1 * S_2)$.

Definition 5. The conditional ability associated with S_1 given the ability associated with S_2 is measured with

$$\theta(S_1|S_2) = \theta(Total) - \theta(S_2), \tag{28}$$

where $\theta(Total) = \theta(S_1,S_2)$ which is defined in (22).

$\theta(S_1|S_2)$ in (28) measures the ability associated with S_1 with exclusion of S_2. If S_1 and S_2 are independent, $\theta(S_1|S_2)$ is equal to $\theta(S_1)$, i.e. $\theta(S_1|S_2) = \theta(S_1)$.

Similar to Eq. (19), the following theorem shows the same decomposition of the total ability in terms of the subscales.

Theorem 4.

$$\theta(Total) = \theta(S_1|S_2) + \theta(S_2|S_1) + \theta(S_1 * S_2) \tag{29}$$

Proof. By Definition 5, there is

$$\theta(S_1|S_2) = \theta(Total) - \theta(S_2), \tag{30}$$

$$\theta(S_2|S_1) = \theta(Total) - \theta(S_1). \tag{31}$$

By (30) + (31) and (27),

$$\theta(S_1|S_2) + \theta(S_2|S_1) = 2\theta(Total) - \theta(S_1) - \theta(S_2)$$

$$\Longleftrightarrow$$

$$\theta(S_1|S_2) + \theta(S_2|S_1) = \theta(Total) - \theta(S_1 * S_2)$$

$$\Longleftrightarrow$$

$$\theta(Total) = \theta(S_1|S_2) + \theta(S_2|S_1) + \theta(S_1 * S_2).$$

This is the proof of Theorem 4.

In Theorem 4, the measure of the total ability is the summation of the measure of the ability associated with S_1 with exclusion of S_2 and the measure of the ability associated with S_2 with exclusion of S_1 and the measure of the shared ability among S_1 and S_2. Obviously, if S_1 and S_2 are independent, the measure of the total ability is the summation of the measures of the ability associated with S_1 and the ability associated with S_2, i.e. $\theta(Total) = \theta(S_1) + \theta(S_2)$.

So far, we have discussed the measures on the abilities associated with two subscales. In case of multiple subscales, the measures can be defined in the similar way. Without loss of generality, let us look into the case of three subscales S_1, S_2, and S_3 which their items are those items in $\mathscr{S}_1$, $\mathscr{S}_2$ and $\mathscr{S}_3$, the subsets of all correctly responded items, which is $\{i_1, \ldots, i_L\}$, respectively.

$$S_1 \sim \mathscr{S}_1 \subseteq \{i_1, \ldots, i_L\}$$
$$S_2 \sim \mathscr{S}_2 \subseteq \{i_1, \ldots, i_L\}$$
$$S_3 \sim \mathscr{S}_3 \subseteq \{i_1, \ldots, i_L\}$$
$$Total \sim \{i_1, \ldots, i_L\},$$

where "$S_1 \sim \mathscr{S}_1$" means the items that belong to subscale S_1 are those in the set $\mathscr{S}_1$, which is a subset of all correctly responded items $\{i_1, \ldots, i_L\}$. Also, we assume $\mathscr{S}_1 \cup \mathscr{S}_2 \cup \mathscr{S}_3 = \{i_1, \ldots, i_L\}$.

Definition 6. The measure of the shared abilities associated with S_1, S_2, and S_3 is defined by

$$\begin{aligned} \theta(S_1 * S_2 * S_3) = {} & \theta(S_1) + \theta(S_2) + \theta(S_3) \\ & - \theta(S_1, S_2) - \theta(S_1, S_3) - \theta(S_2, S_3) + \theta(S_1, S_2, S_3), \end{aligned} \qquad (32)$$

where

$$\theta(S_1, S_2, S_3) = \theta(Total) = -ln(P(X_{i_1} = R, \ldots, X_{i_L} = R)), \qquad (33)$$

$$\theta(S_j, S_k) = -ln(P(X_{i_1} = R, \ldots, X_{i_{M_{j,k}}} = R)). \qquad (34)$$

In Eq. (34), the $M_{j,k}$ correctly responded items $i_1, \ldots, i_{M_{j,k}}$ are exactly those in $\mathscr{S}_j \cup \mathscr{S}_k$, i.e. $\{i_1, \ldots, i_{M_{j,k}}\} = \mathscr{S}_j \cup \mathscr{S}_k$ for $j, k = 1, 2, 3$.

It is interesting to compare the similar structure between Eqs. (21) and (32) and, in fact, Eq. (21) is nothing but a special case of Eq. (32) if each subscale only contains a single item. Similar to $\theta(S_1 * S_2)$, $\theta(S_1 * S_2 * S_3)$ can be negative, but its interpretation is more complicated. Although $\theta(S_1 * S_2 * S_3)$ is called shared ability here, this concept is closer to the interaction among the abilities associated with S_1, S_2, and S_3.

Corollary 6. *If S_1, S_2 and S_3 are (jointly) independent, then*

$$\theta(Total) = \theta(S_1) + \theta(S_2) + \theta(S_3). \tag{35}$$

Proof. The proof is obvious by the definitions:

$\theta(S_i) = -ln(P(X_{i_1} = R, \ldots, X_{i_{M_i}} = R))$ where the M_i correctly responded items $i_1, \ldots, i_{M_i}$ are exactly those in $\mathscr{S}_i$, i.e. $\{i_1, \ldots, i_{M_i}\} = \mathscr{S}_i$ for $i = 1,2,3$.

$\theta(Total) = -ln(P(X_{i_1} = R, \ldots, X_{i_L} = R))$ where the L correctly responded items $i_1, \ldots, i_L$ are exactly those in $\mathscr{S}_1 \cup \mathscr{S}_2 \cup \mathscr{S}_3$, i.e. $\{i_1, \ldots, i_L\} = \mathscr{S}_1 \cup \mathscr{S}_2 \cup \mathscr{S}_3$.

Equation (35) in Corollary 6 is another example of additivity in terms of their subscales. Equation (6) can be thought as a special case of Eq. (35) for each subscale to associate with a single item. Although there are three subscales in Corollary 6, the property of additivity is also true for the case of multiple subscales.

Corollary 7. *If S_1, S_2 and S_3 are (jointly) independent, then*

$$\theta(S_1 * S_2 * S_3) = 0. \tag{36}$$

Proof. The proof is similar to that in Corollary 6.

Theorem 5.

$$\begin{aligned}\theta(Total) = \; & \theta(S_1) + \theta(S_2) + \theta(S_3) - \theta(S_1 * S_2) - \theta(S_1 * S_3) \\ & -\theta(S_2 * S_3) + \theta(S_1 * S_2 * S_3).\end{aligned} \tag{37}$$

Proof. First, similar to (25), there are

$$\theta(S_j, S_k) = \theta(S_j) + \theta(S_k) - \theta(S_j * S_k) \qquad for \;\; j,k = 1,2,3 \tag{38}$$

By Definition 6 and (38), there is

$$\begin{aligned}&\theta(S_1 * S_2 * S_3) \\ &\quad = \theta(S_1) + \theta(S_2) + \theta(S_3) - \theta(S_1,S_2) - \theta(S_1,S_3) - \theta(S_2,S_3) + \theta(S_1,S_2,S_3) \\ &\quad = \theta(S_1) + \theta(S_2) + \theta(S_3) - \theta(S_1) - \theta(S_2) + \theta(S_1 * S_2) - \theta(S_1) - \theta(S_3) \\ &\qquad + \theta(S_1 * S_3) - \theta(S_2) - \theta(S_3) + \theta(S_2 * S_3) + \theta(S_1,S_2,S_3) \\ &\quad = -\theta(S_1) - \theta(S_2) - \theta(S_3) + \theta(S_1,S_2) + \theta(S_1,S_3) \\ &\qquad + \theta(S_2,S_3) + \theta(S_1,S_2,S_3)\end{aligned}$$

Therefore,

$$\begin{aligned}\theta(S_1,S_2,S_3) = {} & \theta(S_1)+\theta(S_2)+\theta(S_3)-\theta(S_1*S_2)-\theta(S_1*S_3)\\ & -\theta(S_2*S_3)+\theta(S_1*S_2*S_3).\end{aligned}$$

This is the proof of Theorem 5.

Theorem 5 shows that the measure of the total ability can be linearly expressed with the measures of the shared abilities. In fact, according to (32) and (37), $\theta(S_1,S_2,S_3)$ and $\theta(S_1*S_2*S_3)$ are two conjugate concepts.

Theorem 6.

$$\theta(Total) = \theta(S_1|S_2)+\theta(S_2|S_3)+\theta(S_3|S_1)+\theta(S_1*S_2*S_3). \tag{39}$$

Proof. First, by (38), there is

$$\theta(S_1*S_2) = \theta(S_1)+\theta(S_2)-\theta(S_1,S_2) \tag{40}$$

By Theorem 5 and (40), there is

$$\begin{aligned}\theta(Total) = {} & \theta(S_1)+\theta(S_2)+\theta(S_3)-\theta(S_1*S_2)-\theta(S_1*S_3)\\ & -\theta(S_2*S_3)+\theta(S_1*S_2*S_3)\\ = {} & \theta(S_1,S_2)+\theta(S_3)-\theta(S_1*S_3)-\theta(S_2*S_3)+\theta(S_1*S_2*S_3)\end{aligned}$$

Equivalently, Eq. (28) can be rewritten as

$$\theta(S_1,S_2) = \theta(S_1|S_2)+\theta(S_2). \tag{41}$$

By applying (41), we have

$$\theta(Total) = \theta(S_1|S_2)+\theta(S_2)+\theta(S_3)-\theta(S_1*S_3)-\theta(S_2*S_3)+\theta(S_1*S_2*S_3)$$

In the same way, by applying the following equations,

$$\begin{aligned}\theta(S_1*S_3) &= \theta(S_1)+\theta(S_3)-\theta(S_1,S_3),\\ \theta(S_2*S_3) &= \theta(S_2)+\theta(S_3)-\theta(S_2,S_3),\\ \theta(S_1,S_3) &= \theta(S_1|S_3)+\theta(S_3),\\ \theta(S_2,S_3) &= \theta(S_2|S_3)+\theta(S_3).\end{aligned}$$

We finally have

$$\begin{aligned}\theta(Total) &= \theta(S_1|S_2) + \theta(S_2,S_3) - \theta(S_1 * S_3) + \theta(S_1 * S_2 * S_3)\\ &= \theta(S_1|S_2) + \theta(S_2|S_3) + \theta(S_3) - \theta(S_1 * S_3) + \theta(S_1 * S_2 * S_3)\\ &= \theta(S_1|S_2) + \theta(S_2|S_3) + \theta(S_3|S_1) + \theta(S_1 * S_2 * S_3).\end{aligned}$$

This is the proof of Theorem 6.

It is obvious that, if S_1, S_2, and S_3 are jointly independent, Eq. (39) becomes (6) and therefore, Eq. (39) in Theorem 6 can be thought as a general form of additivity. In Theorem 6, the total ability is decomposed into four parts which are $\theta(S_1|S_2)$, $\theta(S_1|S_3)$, $\theta(S_2|S_3)$ and $\theta(S_1 * S_2 * S_3)$. The decomposition in Theorem 6 is not unique. In similar way, the total ability can also be decomposed as follows:

$$\theta(Total) = \theta(S_1|S_3) + \theta(S_3|S_2) + \theta(S_2|S_1) + \theta(S_1 * S_2 * S_3). \tag{42}$$

Although the total ability is decomposed into four components in Theorem 6, each of these four decomposed components can still be further decomposed. In the remaining part of this section, a unique and complete decomposition for the total ability will be derived. First, the following concepts are introduced:

$$\theta(S_1,S_2,S_3) = \theta(Total) = -ln(P(X_{i_1} = R,\ldots,X_{i_L} = R)), \tag{43}$$

$$\theta(S_j,S_k) = -ln(P(X_{i_1} = R,\ldots,X_{i_{M_{j,k}}} = R)). \tag{44}$$

In Eq. (43), the L correctly responded items $i_1,\ldots,i_L$ are exactly those in $\mathscr{S}_1 \cup \mathscr{S}_2 \cup \mathscr{S}_3$, i.e. $\{i_1,\ldots,i_L\} = \mathscr{S}_1 \cup \mathscr{S}_2 \cup \mathscr{S}_3$. In Eq. (44), the $M_{j,k}$ correctly responded items $i_1,\ldots,i_{M_{j,k}}$ are exactly those in $\mathscr{S}_j \cup \mathscr{S}_k$, i.e. $\{i_1,\ldots,i_{M_{j,k}}\} = \mathscr{S}_j \cup \mathscr{S}_k$ for $j,k = 1,2,3$.

With $\theta(S_1,S_2,S_3)$ and $\theta(S_j,S_k)$ in (43) and (44), we can define the following ability measures conditioned on the subscale(s):

Definition 7.

$$\theta(S_1|S_2,S_3) = \theta(S_1,S_2,S_3) - \theta(S_2,S_3), \tag{45}$$

where $\theta(S_j,S_k)$ for $j,k = 1,2,3$ and $\theta(S_1,S_2,S_3)$ are defined in (44) and (43).

Definition 8.

$$\theta(S_1,S_2|S_3) = \theta(S_1,S_2,S_3) - \theta(S_3), \tag{46}$$

where $\theta(S_1,S_2,S_3)$ is defined in (44).

By Definition 7, there is

$$\begin{aligned}\theta(S_1,S_2,S_3) &= \theta(S_3|S_1,S_2)+\theta(S_1,S_2)\\ &= \theta(S_3|S_1,S_2)+\theta(S_2|S_1)+\theta(S_1).\end{aligned} \tag{47}$$

Equation (47) is also called additivity.

Definition 9.

$$\theta(S_1 * S_2|S_3) = \theta(S_1|S_3)+\theta(S_2|S_3)-\theta(S_1,S_2|S_3), \tag{48}$$

where $\theta(S_i|S_3)$ for $i=1,2$ is defined in (28).

Theorem 7.

$$\begin{aligned}\theta(Total) &= \theta(S_1|S_2,S_3)+\theta(S_2|S_1,S_3)+\theta(S_3|S_1,S_2)+\theta(S_1 * S_3|S_2)\\ &\quad +\theta(S_1 * S_2|S_3)+\theta(S_2 * S_3|S_1)+\theta(S_1 * S_2 * S_3).\end{aligned} \tag{49}$$

Proof. First, by Definitions 7 and 9,

$$\begin{aligned}\theta(S_1|S_2,S_3)+\theta(S_1 * S_2|S_3) &= \theta(S_1,S_2,S_3)-\theta(S_2,S_3)+\theta(S_1|S_3)\\ &\quad +\theta(S_2|S_3)-\theta(S_1,S_2|S_3).\end{aligned} \tag{50}$$

Second, by Definition 8 and Eq. (41),

$$\theta(S_1,S_2|S_3) = \theta(S_1,S_2,S_3)-\theta(S_1,S_2), \tag{51}$$

$$\theta(S_1|S_3) = \theta(S_1,S_3)-\theta(S_3), \tag{52}$$

$$\theta(S_2|S_3) = \theta(S_2,S_3)-\theta(S_3). \tag{53}$$

By substituting (51), (52), and (53) into (50) and rearranging the terms, we have

$$\begin{aligned}\theta(S_1|S_2,S_3)+\theta(S_1 * S_2|S_3) &= \theta(S_1,S_2,S_3)-\theta(S_2,S_3)\\ &\quad +\theta(S_1,S_3)-\theta(S_3)+\theta(S_2,S_3)\\ &\quad -\theta(S_3)-\theta(S_1,S_2,S_3)+\theta(S_3)\\ &= \theta(S_1,S_3)-\theta(S_3)=\theta(S_1|S_3).\end{aligned} \tag{54}$$

By (54) and in the same way as (54), we have

$$\theta(S_1|S_3)=\theta(S_1|S_2,S_3)+\theta(S_1 * S_2|S_3), \tag{55}$$

$$\theta(S_3|S_2)=\theta(S_3|S_1,S_2)+\theta(S_1 * S_3|S_2), \tag{56}$$

$$\theta(S_2|S_1)=\theta(S_2|S_1,S_3)+\theta(S_2 * S_3|S_1). \tag{57}$$

Finally, by substituting (55), (56), and (57) into (42), we have

$$\begin{aligned}\theta(Total) &= \theta(S_1|S_3)+\theta(S_3|S_2)+\theta(S_2|S_1)+\theta(S_1*S_2*S_3)\\ &= \theta(S_1|S_2,S_3)+\theta(S_2|S_1,S_3)+\theta(S_3|S_1,S_2)+\theta(S_1*S_3|S_2)\\ &\quad +\theta(S_1*S_2|S_3)+\theta(S_2*S_3|S_1)+\theta(S_1*S_2*S_3).\end{aligned}$$

This is the proof of Theorem 7.

In Theorem 7, the total ability of three subscales is decomposed into seven basic components. The interpretation of each component is different from one to another. With the decomposition in Theorem 7, we can look into the details of subscale structure of the total ability.

Although we have discussed the decomposition (49) for the case of three subscales in Theorem 7, the decomposition for the case of arbitrary number of subscales can also be derived in the similar way. Readers are encouraged to derive the decomposition for the cases of four subscales or more.

5 Discussion

In this paper, the measure of the ability defined in (5) shows (1) additivity; (2) nonnegativity; (3) the measure of the ability with incorrect responses for all items is equal to zero. Therefore, the definition in (5) conceptually can be called the measure of the ability according to Measure Theory (Halmos 1974). Here, we place emphasis on the concept of measure because, without additivity, an "ability measure" can cause unexpected results. For example, without additivity, the directly measured value and indirectly measured value for the same total ability are not the same for most of cases. This is similar to measuring the area of a rectangle by summation of its length and width (see *Introduction* of this paper).

In Sect. 3, the measure of the shared abilities is defined. We point out that the measure of the shared abilities does not make sense without additivity. Unlike the ability measure in Definition 1 which is nonnegative, measure of the shared abilities can be negative. The negative value of the measure of the shared abilities is interpreted as the conflicted or exclusive interaction among these two abilities. For two exclusive abilities, the higher for one ability, the lower will be for another ability. The positive value of the measure of the shared abilities implies that these two abilities are not conflicted which means that, the higher for one ability, the higher will be also for another ability. In practice, it is very rare for the measure of the shared ability to be negative although it is possible.

The marginal measure of the ability associated with the subscale is defined in Sect. 4. We also look into the relation between the measure of the total ability and the measures of those abilities associated with the subscales by decomposing the measure of the total ability in terms of the measures of those abilities associated with the subscales. Like the measure of the shared ability, without additivity, it is impossible to decompose the measure of the total ability in terms of the measures of those abilities associated with the subscales.

Although, throughout this paper, we assume all items are dichotomous, the definition in (5) can be expanded to include partial credits, i.e. the items can have more than two categories of right (R) and wrong (W). Under the case of partial credits, the property of additivity is still reserved, i.e. the ability measure with the partial credits is on the basis of measure theory. The nonparametric ability measure with partial credits currently is under organization and will meet with readers in the near future.

Finally, in this paper, most conclusions can be extended to more general form in the same way. Also, the ability measures defined in this paper may be parameterized with some reasonable constraints such as the log-linear model. In practice, the parameterized measures is possible to handle the datasets of small size. How to parameterize the ability measures defined in this paper could be the topic for the future work.

Acknowledgments Author would like to express his thanks to Prof. Andries van der Ark for his valuable comments with which this paper can be significantly improved for its readability. Author also thanks Gwen Exner for his assistances and helps.

References

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A*, *222*, 309–368.

Fisher, R. A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, *22*, 700–725.

Halmos, P. R. (1974). *Measure theory*. New York: Springer.

Hays, W. L. (1970). *Statistics (Volume 1) - Probability, Inference, and Decision*. New York: Holt, Rinehart and Winston.

Shannon, C. E. (1948). A mathematical theory of communications. *The Bell System Technical Journal*, *27*, 379–423.

Wiener, N. (1948). Cybernetics or Control and Communication in the Animal and the Machine. Cambridge, Mass.: The MIT Press.

An Alternative to Cronbach's Alpha: An *L*-Moment-Based Measure of Internal-Consistency Reliability

Todd Christopher Headrick and Yanyan Sheng

1 Introduction

Coefficient alpha (Cronbach 1951; Guttman 1945) is a commonly used index for measuring internal-consistency reliability. Consider alpha (α) in terms of a model that decomposes an observed score into the sum of two independent components: a true unobservable score t_i and a random error component e_{ij}. The model can be summarized as

$$X_{ij} = t_i + e_{ij} \tag{1}$$

where X_{ij} is the observed score associated with the i-th examinee on the j-th test item, and where $i = 1, \ldots, n$; $j = 1, \ldots, k$; and the error terms (e_{ij}) are independent with a mean of zero. Inspection of (1) indicates that this particular model restricts the true score t_i to be the same across all k test items. The reliability measure associated with the test items in (1) is a function of the true score variance and cannot be computed directly. Thus, estimates of reliability such as coefficient α have been derived and will be defined herein as (e.g., Christman and Van Aelst 2006)

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_j \sigma_j^2}{\sum_j \sigma_j^2 + \sum\sum_{j \neq j'} \sigma_{jj'}}\right). \tag{2}$$

A conventional estimate of α can be obtained by substituting the usual OLS sample estimates associated with σ_j^2 and $\sigma_{jj'}$ into (2) as

T.C. Headrick (✉) • Y. Sheng
Section of Statistics and Measurement, Department of EPSE,
Southern Illinois University Carbondale, Carbondale, IL 62901, USA
e-mail: headrick@siu.edu; ysheng@siu.edu

R.E. Millsap et al. (eds.), *New Developments in Quantitative Psychology*,
Springer Proceedings in Mathematics & Statistics 66, DOI 10.1007/978-1-4614-9348-8_2,

$$\hat{\alpha}_C = \frac{k}{k-1}\left(1 - \frac{\sum_j s_j^2}{\sum_j s_j^2 + \sum\sum_{j \neq j'} s_{jj'}}\right) \quad (3)$$

where s_j^2 and $s_{jj'}$ are the diagonal and off-diagonal elements from the variance–covariance matrix, respectively.

Although coefficient α is often used as an index for reliability, it is also well known that its use is limited when data are non-normal, in particular leptokurtic, or when sample sizes are small (e.g., Bay 1973; Christman and Van Aelst 2006; Sheng and Sheng 2012; Wilcox 1992). These limitations are of concern because data sets in the social and behavioral sciences can often possess heavy tails or consist of small sample sizes (e.g., Micceri 1989; Yuan et al. 2004). Specifically, it has been demonstrated that $\hat{\alpha}_C$ can substantially underestimate α when heavy-tailed distributions are encountered. For example, Sheng and Sheng (2012, Table 1) sampled from a symmetric leptokurtic distribution and found the empirical estimate of α to be approximately $\hat{\alpha}_C = 0.70$ when the true population parameter was $\alpha = 0.80$. Further, it is not uncommon that data sets consist of small sample sizes, e.g., $n = 10$ or 20. More specifically, small sample sizes are commonly encountered in the contexts of rehabilitation (e.g., alcohol treatment programs, group therapy, etc.) and special education as student–teacher ratios are often small. Furthermore, Monte Carlo evidence has demonstrated that $\hat{\alpha}_C$ can underestimate α—even when small samples are drawn from a normal distribution (see Sheng and Sheng 2012, Table 1).

L-moment estimators (e.g., Hosking 1990; Hosking and Wallis 1997) have demonstrated to be superior to the conventional product-moment estimators in terms of bias, efficiency, and their resistance to outliers (e.g., Headrick 2011; Hodis et al. 2012; Hosking 1992; Vogel and Fennessy 1993). Further, L-comoment estimators (Serfling and Xiao 2007) such as the L-correlation have demonstrated to be an attractive alternative to the conventional Pearson correlation in terms of relative bias when heavy-tailed distributions are of concern (Headrick and Pant 2012a,b,c,d,e).

In view of the above, the present aim here is to propose an L-comoment-based coefficient L-α, and its estimator denoted as $\hat{\alpha}_L$, as an alternative to conventional alpha $\hat{\alpha}_C$ in (3). Empirical results associated with the simulation study herein indicate that $\hat{\alpha}_L$ can be substantially superior to $\hat{\alpha}_C$ in terms of relative bias and relative standard error (RSE) when distributions are heavy-tailed and sample sizes are small.

The rest of the paper is organized as follows. In Sect. 2, summaries of univariate L-moments and L-comoments are first provided. Coefficient L-α ($\hat{\alpha}_L$) is then introduced and numerical examples are provided to illustrate the computation and sampling distribution associated with $\hat{\alpha}_L$. In Sect. 3, a Monte Carlo study is carried out to evaluate the performance of $\hat{\alpha}_C$ and $\hat{\alpha}_L$. The results of the study are discussed in Sect. 4.

2 *L*-Moments, *L*-Comoments, and Coefficient *L*-α

The system of univariate L-moments (Hosking 1990, 1992; Hosking and Wallis 1997) can be considered in terms of the expectations of linear combinations of order statistics associated with a random variable Y. Specifically, the first four L-moments are expressed as

$$\lambda_1 = E[Y_{1:1}]$$
$$\lambda_2 = \frac{1}{2}E[Y_{2:2} - Y_{1:2}]$$
$$\lambda_3 = \frac{1}{3}E[Y_{3:3} - 2Y_{2:3} + Y_{1:3}]$$
$$\lambda_4 = \frac{1}{4}E[Y_{4:4} - 3Y_{3:4} + 3Y_{2:4} - Y_{1:4}]$$

where $Y_{\ell:m}$ denotes the ℓth smallest observation from a sample of size m. As such, $Y_{1:m} \leq Y_{2:m} \leq \ldots \leq Y_{m:m}$ are referred to as order statistics drawn from the random variable Y. The values of λ_1 and λ_2 are measures of location and scale and are the arithmetic mean and one-half of the coefficient of mean difference (or Gini's index of spread), respectively. Higher order L-moments are transformed to dimensionless quantities referred to as L-moment ratios defined as $\tau_r = \lambda_r/\lambda_2$ for $r \geq 3$, where τ_3 and τ_4 are the analogs to the conventional measures of skew and kurtosis. In general, L-moment ratios are bounded in the interval $-1 < \tau_r < 1$ as is the index of L-skew (τ_3) where a symmetric distribution implies that all L-moment ratios with odd subscripts are zero. Other smaller boundaries can be found for more specific cases. For example, the index of L-kurtosis (τ_4) has the boundary condition for continuous distributions of $(5\tau_3^2 - 1)/4 < \tau_4 < 1$.

L-comoments (Olkin and Yitzhuki 1992; Serfling and Xiao 2007) are introduced by considering two random variables Y_j and Y_k with distribution functions $F(Y_j)$ and $F(Y_k)$. The second L-moments associated with Y_j and Y_k can alternatively be expressed as

$$\lambda_2(Y_j) = 2\mathrm{Cov}(Y_j, F(Y_j))$$
$$\lambda_2(Y_k) = 2\mathrm{Cov}(Y_k, F(Y_k)). \tag{4}$$

The second L-comoments of Y_j toward Y_k and Y_k toward Y_j are

$$\lambda_2(Y_j, Y_k) = 2\mathrm{Cov}(Y_j, F(Y_k))$$
$$\lambda_2(Y_k, Y_j) = 2\mathrm{Cov}(Y_k, F(Y_j)). \tag{5}$$

The ratio $\eta_{jk} = \lambda_2(Y_j, Y_k)/\lambda_2(Y_j)$ is defined as the L-correlation of Y_j with respect to Y_k, which measures the monotonic relationship (not just linear) between two variables (Headrick and Pant 2012c). Note that in general, $\eta_{jk} \neq \eta_{kj}$. The estimators of (4) and (5) are U-statistics (Serfling 1980; Serfling and Xiao 2007) and their sampling distributions converge to a normal distribution when the sample size is sufficiently large.

In terms of coefficient L-α, an approach that can be taken to equate the conventional and L-moment (comoment) definitions of α is to express (2) as

Table 1 Data (Items) for computing the second L-moment–comoment matrix in Table 2

X_{i1}	X_{i2}	X_{i3}	$\hat{F}(X_{i1})$	$\hat{F}(X_{i2})$	$\hat{F}(X_{i3})$
2	4	3	0.15	0.45	0.15
5	7	7	0.75	0.95	1.00
3	5	5	0.35	0.65	0.40
6	6	6	0.90	0.80	0.75
7	7	6	1.00	0.95	0.75
5	2	6	0.75	0.10	0.75
2	3	3	0.15	0.25	0.15
4	3	6	0.55	0.25	0.75
3	5	5	0.35	0.65	0.40
4	4	5	0.55	0.45	0.40

The data are part of the "Satisfaction With Life Data" from McDonald (1999, p. 47)

Table 2 Second L-moment–comoment matrix for coefficient $\hat{\alpha}_L$ in Eq. (9)

Item	1	2	3
1	$\ell_{2(1)} = 0.989$	$\ell_{2(12)} = 0.500$	$\ell_{2(13)} = 0.789$
2	$\ell_{2(21)} = 0.500$	$\ell_{2(2)} = 1.022$	$\ell_{2(23)} = 0.411$
3	$\ell_{2(31)} = 0.667$	$\ell_{2(32)} = 0.333$	$\ell_{2(3)} = 0.733$

$$\alpha = \frac{1}{1+(R-1)/k} = \frac{k}{k-1}\left(1 - \frac{\sum_j \sigma_j^2}{\sum_j \sigma_j^2 + \sum\sum_{j \neq j'} \sigma_{jj'}}\right) \tag{6}$$

where $R > 1$ is the common ratio between the main and off-diagonal elements of the variance–covariance matrix, i.e. $R = \sigma_j^2 / \sigma_{jj'}$. (See the appendix for the derivation of Eq. (6)). As such, given a fixed value of R in (6) will allow for α to be defined in terms of the second L-moments and second L-comoments as

$$\alpha = \frac{1}{1+(R-1)/k} = \frac{k}{k-1}\left(1 - \frac{\sum_j \lambda_{2(j)}}{\sum_j \lambda_{2(j)} + \sum\sum_{j \neq j'} \lambda_{2(jj')}}\right) \tag{7}$$

where $R = \lambda_{2(j)}/\lambda_{2(jj')}$. Thus, the estimator of L-α is expressed as

$$\hat{\alpha}_L = \frac{k}{k-1}\left(1 - \frac{\sum_j \ell_{2(j)}}{\sum_j \ell_{2(j)} + \sum\sum_{j \neq j'} \ell_{2(jj')}}\right) \tag{8}$$

where $\ell_{2(j)}$ ($\ell_{2(jj')}$) denotes the sample estimate of the second L-moments (second L-comoment) in (4) and (5). An example demonstrating the computation of $\hat{\alpha}_L$ is provided below in Eq. (9). The computed estimate of $\hat{\alpha}_L = 0.807$ in (9) is based on the data in Table 1 and the second L-moment–comoment matrix in Table 2. The corresponding conventional estimate for the data in Table 1 is $\hat{\alpha}_C = 0.798$.

$$\begin{aligned}\hat{\alpha}_L = 0.807 = {} & (3/2)(1 - (\ell_{2(1)} + \ell_{2(2)} + \ell_{2(3)})/(\ell_{2(1)} + \ell_{2(2)} + \ell_{2(3)} \\ & + \ell_{2(21)} + \ell_{2(31)} + \ell_{2(32)} + \ell_{2(12)} + \ell_{2(13)} + \ell_{2(23)})).\end{aligned} \tag{9}$$

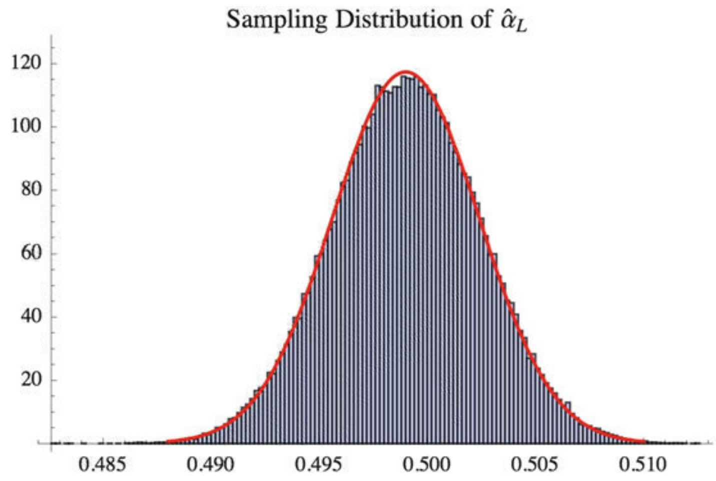

Fig. 1 Approximate normal sampling distribution of $\hat{\alpha}_L$ with $\alpha = 0.50$. The distribution consists of 25,000 statistics based on samples of size $n = 100{,}000$ and the heavy-tailed distribution (kurtosis of 25) in Fig. 2

The estimator $\hat{\alpha}_L$ in (8) and (9) is a ratio of the sums of U-statistics and thus a consistent estimator of α in (7) with a sampling distribution that converges, for large samples, to the normal distribution (e.g., Olkin and Yitzhuki 1992; Schechtman and Yitzhaki 1987; Serfling and Xiao 2007). For convenience to the reader, provided in Fig. 1 is the sampling distribution of $\hat{\alpha}_L$ that is approximately normal and based on $\alpha = 0.50$, $n = 100{,}000$, and a symmetric heavy-tailed distribution (kurtosis of 25, see Fig. 2) that would be associated with t_i in (1).

3 Monte Carlo Simulation

An algorithm was written in MATLAB (Mathworks 2010) to generate 25,000 independent sample estimates of conventional and L-comoment α. The estimators $\hat{\alpha}_C$ and $\hat{\alpha}_L$ were based on the parameters (α, k, R) given in Tables 3 and 4 and the distributions in Figs. 2–4. The parameters of α were selected because they represent commonly used references of various degrees of reliability, i.e. 0.50 (poor); 5/7 = 0.714 (acceptable); 0.80 (good); and 0.90 (excellent). Further, for each set of parameters in Tables 3 and 4, the empirical estimators $\hat{\alpha}_C$ and $\hat{\alpha}_L$ were generated based on sample sizes of $n = 10$, 20, 1,000. For all cases in the simulation, the error term e_{ij} in (1) was normally distributed with zero mean and with the variance parameters (σ_e^2) listed in Tables 3 and 4.

The three distributions depicted in Figs. 2–4 are associated with the true scores t_i in Eq. (1). These distributions are referred to as: Distribution 1 is symmetric and leptokurtic (skew = 0, kurtosis = 25; L-skew = 0, L-kurtosis = 0.4225);