Ability Measure

Nan L. Kong*

Abstract

This paper defines a new measure of examinees abilities using additivity, one of the fundamental properties of a measure. By employing mathematical proofs, other fundamental properties of the new measure are demonstrated. This paper also shows that shared ability and unique ability can be measured with additivity. Finally, the paper looks at ability measure associated with subscales and ability measure with partial credits of items is also discussed.

Key Words: Additivity, Ability measure, Item response, Shared ability, Unique ability, Measures with Subscales and Partial Credit.

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), (ii) and (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 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.

*nkong.usa@gmail.com

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

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

The 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 $I(X,Y) = 0$, the equation 2 becomes

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

The equation 3 is called *additivity*.

Unlike Shannon's entropy, Fisher information (Fisher, 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).

The equation 4 is called *additivity*.

In 2009, Kong and Lewis (Kong and Lewis, 2009) mathematically proved the following equation for the *K*-dependence coefficient (Kong and Lewis, 2009).

$$K(X:Y,Z) = K(X:Y) + K(X:Z) - K(X:Y \wedge Z). \tag{5}$$

where X, Y and Z are three categorical random variables; K(X:Y,Z) is the *K*-Dependence coefficient of X dependence on X and Y; K(X:Y) and K(X:Z) are the *K*-Dependence coefficients of X dependence on X and Y respectively; K(X:Y $\wedge$ Z) is the *K*-Dependence coefficients of X dependence on the interaction among X, Y and Z.

If there is no interaction among X, Y and Z, which implies K(X:Y $\wedge$ Z) = 0, the equation (5) becomes:

$$K(X:Y,Z) = K(X:Y) + K(X:Z). \tag{6}$$

The equation (6) is called additivity.

So far, we have looked into the theoretical structures for several well defined measures. All of these structures reveals the same property – *additivity* as shown in (1), (3), (4) and (6). 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 next section, an ability measure is defined and studied according to the additivity.

2. Ability Measure Defined with Dichotomous Items

In this section, the examinee's ability will be measured on the basis of a set of item responses given a test. Here, the items are the questions in the given test that have right (R) or wrong (W) responses. 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 $L(0 \leq L \leq I)$ 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 ($l = 1, ..., L; L \geq 1$) is defined as

$$\theta(i_1, \ldots, i_l, \ldots, i_L) = -ln(P(X_{i_1} = R, \ldots, X_{i_l} = R, \ldots, X_{i_L} = R)). \tag{7}$$

In (7), $\theta(i_1, \ldots, i_l, \ldots, i_L)$ is called the measure of the ability with right (R) response(s) for the items $i_l(l = 1, ..., L)$. We also request that the examinee's ability be measured as zero if this examinee does not respond any item correctly, i.e $L = 0$ in (7).

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 up in (7). 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

$$\theta(i_1, \cdots, i_L) = \theta(i_1) + \ldots + \theta(i_L). \tag{8}$$

As we can see in equation 25 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 equation 25 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.

Corollary 1.

$$0 \leq \theta(i_1, \cdots, i_L) \leq +\infty. \tag{9}$$

Proof: This is obvious from the Definition 1.

Corollary 2.

$$\theta(i_1, \cdots, i_L) = 0 \iff P(X_{i_1} = R, \cdots, X_{i_L} = R) = 1 \tag{10}$$

Proof: This is obvious from the Definition 1.

Corollary 3.

$$\theta(i_1, \cdots, i_L) = +\infty \iff P(X_{i_1} = R, \cdots, X_{i_L} = R) = 0 \tag{11}$$

Proof: This is obvious from the Definition 1.

As shown in Corollary 1, the ability measure defined in (7) 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

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

Therefore, the following theorem is obtained:

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

$$\theta(i_1, \cdots, i_M) \leq \theta(i_1, \cdots i_L) \tag{12}$$

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 abilty associated with subscale can not be greater than the measure of its total ability.

In summary, the ability measure defined in (7) has the following properties: (a) Additivity (if the items are independent) as shown in equation 25. (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 the 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 utlized for measureing abiliity, 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 (7) 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.

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

$$\theta(i_1 * i_2) = \theta(i_1) + \theta(i_2) - \theta(i_1, i_2). \tag{13}$$

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 :

$$\theta(i_1 * i_2) = -ln\frac{P(X_{i_1} = R)P(X_{i_2} = R)}{P(X_{i_1} = R, X_{i_2} = R)} \tag{14}$$

By (14), 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,

$$\theta(i_1) = -ln(P(X_{i_1} = R), \tag{15}$$
$$\theta(i_2) = -ln(P(X_{i_2} = R), \tag{16}$$
$$\theta(i_1, i_2) = -ln(P(X_{i_1} = R, X_{i_2} = R)). \tag{17}$$

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 equation 13, 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 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 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 another item, this examinee has negative shared ability among these two items. In practice, for most of cases, the shared ability are 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

$$\theta(i_1|i_2) = -lnP(X_{i_1} = R|X_{i_2} = R). \tag{18}$$

Corollary 4.

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

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.

$$\theta(i_1 * i_2) = \theta(i_1) - \theta(i_1|i_2) \tag{20}$$

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 7 and 8, describes the relation among total ability, shared ability and unique ability:

$$\theta(i_1, i_2) = \theta(i_1 * i_2) + \theta(i_1|i_2) + \theta(i_2|i_1). \tag{21}$$

In (21), 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. The equation in (21) 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.

$$\theta(i_1 * i_2) \leq \theta(i_1) \tag{22}$$

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:

$$\begin{aligned}\theta(i_1 * i_2 * i_3) = \theta(i_1) + \theta(i_2) + \theta(i_3) - \theta(i_1, i_2) \\ -\theta(i_1, i_3) - \theta(i_2, i_3) + \theta(i_1, i_2, i_3).\end{aligned} \tag{23}$$

Obviously, according to (23), (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. Ability Measure with Partial Credits

In this section, the examinee's ability will be measured on the basis of a set of item responses from a given test. Here, the items are the questions in the given test that have scores $v_1 \prec v_2 \prec \ldots \prec v_m$ where the ordinal symbol $\prec$ means *less than* and v_1 is (fully) wrong response while v_m is (fully) correct response. Those scores between v_1 and v_m are called partial credits. The number of the scores, m, could be different from item to item. Let i_l (*l=1,...,L*) be the item response with possible scores $v_1 \prec v_2 \prec \ldots \prec v_{m_l}$. Similar to the number of the scores m_l, those possible scores also could be different from item to item. In this paper, without loss of generality, the possible scores $v_1 \prec v_2 \prec \ldots \prec v_{m_l}$ are assumed to be the same for all items (therefore, the lowest score v_1 is assumed to be the same for all items). The probability of the event $\{i_l = s_l\}$ is denoted by P($i_l = s_l$) where $s_l \in \{v_1, \ldots, v_{m_l}\}$ and, the probability of the event $\{i_{l_1} = s_{l_1}\} \bigcap \{i_{l_2} = s_{l_2}\}$ is denoted by P($i_{l_1} = s_{l_1}, i_{l_2} = s_{l_2}$) where $s_{l_1} \in \{v_1, \ldots, v_{m_{l_1}}\}$ and $s_{l_2} \in \{v_1, \ldots, v_{m_{l_2}}\}$ etc.

Definition 4.

The ability for the items i_l with scoring $i_l = s_l$ ($l = 1, ..., L; L \geq 1$) is measured with

$$\theta(i_1 = s_1, \cdots, i_L = s_L) = -ln(P(i_1 \geq s_1, \cdots, i_L \geq s_L)). \tag{24}$$

In (24), $\theta(i_1 = s_1, \cdots, i_L = s_L)$ is called the measure of the ability associated with the items i_l with scoring $i_l = s_l$ (*l=1,...,L*). Obviously $P(i_l \geq v_1, \cdots, i_L \geq v_1) = 1$, where v_1 is the lowest score for i_l (*l=1,...,L*), implies $\theta(i_1 = v_1, \cdots, i_L = v_1) = 0$ showing the measure of the ability associated with all items of the lowest scores is equal to zero.

The following theorem shows that the ability measure defined in (24) is additive.

Theorem 4 (Additivity).

If items i_1,..., i_L are (jointly) independent, then

$$\theta(i_1 = s_1, \cdots, i_L = s_L) = \theta(i_1 = s_1) + \ldots + \theta(i_L = s_L) \tag{25}$$

Proof: This is obvious from Definition 4.

In Theorem 4, the measure on left hand side of the equation 25 is based on $i_1 = s_1, \cdots, i_L = s_L$ which are all the items in a given test. Therefore, this measure on left hand side of the equation 25 is thought as the measure of examinee's total ability. If each item defines a subscale, the measure based on an individual item can be thought as the measure of examinee's ability associated with the subscale defined by that individual item. Therefore, those measures on the right hand side of the equation 25 are the measures of examinees' abilities associated with the subscales defined by the items $i_1, \cdots, i_L$ respectively. The equation 25 in Theorem 4 shows that summation of the measures of the abilities associated with all the subscales is equal to the total ability measure if and only if these subscales are jointly independent. Therefore, the equation 25 is called additivity. For the case that the subscales are not jointly independent, not only the ability measure on each subscale but also the interactions among the subscales play the roles in the total ability measure. The general theory to measure the total ability associated with several subscales will be studied in the section *The Total Ability and the Abilities Associated with Subscales*.

Theorem 5.

$$0 \leq \theta(i_1 = s_1, \cdots, i_L = s_L) \leq +\infty. \tag{26}$$

Proof: This is obvious from Definition 4.

Theorem 6.

$$\theta(i_1 = s_1, \cdots, i_L = s_L) = 0 \qquad \Longleftrightarrow \qquad P(i_l = s_1, \cdots, i_L = s_L) = 1 \tag{27}$$

$$\theta(i_1 = s_1, \cdots, i_L = s_L) = +\infty \qquad \Longleftrightarrow \qquad P(i_l = s_1, \cdots, i_L = s_L) = 0 \tag{28}$$

Proof: These are obvious from Definition 4.

First, Theorem 5 shows that the ability measure defined in (24) is nonnegative. This conclusion is also implied by the following facts: the measure of the ability for all the items with the lowest scores (fully wrong) is zero and this zero should be lower than or equal to the measure of ability for any other response pattern. More generally, the non-negativity is a fundamental property required by measure theory. Theorem 5 also shows that the ability measure can be any values greater than or equal to zero, but no one could reach plus infinite because probability of that situation is zero (Theorem 5).

Similarly, the (partially-credited) shared ability measure and the (partially-credited) conditional ability measure can also be defined in the same way as the (dichotomously-credited) ability measure discussed in the section 3. It should be pointed out that Definition 4 is an extension of Definition 1, i.e., in case that the number of the scores, which is *m*, is equal to 2 for all of the items, Definition 4 is reduced to Definition 1.

The properties of the ability measure with the partial credits are the same as those discussed in the section 2. Readers are encouraged to verify the properties (a) - (h) with the ability measure in Definition 4.

5. Discussion

In this paper, the measure of the ability defined in (7) and (24) shows (i) additivity; (ii) non-negativity; (iii) the measure of the ability with incorrect responses for all items is equal to zero. Therefore, the definition in (7) and (24) 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 the section 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 7 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 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.

In the section 4, the ability measure for the items of partial credits is defined in (24). The shared ability and conditional ability can also be defined based on the definition in (24) in the same way as those defined in (13) and (18). Under the case of partial credits, the additivity and non-negativity still hold that makes it possible to decompose the total ability measure into the shared ability measures and the unique ability measures.

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 etc. 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.

REFERENCES

Fisher, R.A. (1921), “On the mathematical foundations of theoretical statistics,” Ser. A, Vol. 222, pp. 309-368. London: Philos. Trans. Roy. Soc.

Fisher, R.A. (1925), “Theory of statistical estimation,” (Vol. 22, pp. 700-725). Proc. Cambridge Philos. Soc.

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

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

Kong, L. N. & Lewis, C. (2009), “Multivariate Information and Interaction,” *JSM Proceedings (2009)*, Nonparametric statistics section [CD-ROM], Alexandria, VA: American Statistical Association.

Shannon, C. E. (1948), “A mathematical theory of communications.,” *Bell System Technical Journal*, 27.