

Correlation Functions over Sets with Involutions

Ildar Z. Batyrshin*

Instituto Politécnico Nacional,
Centro de Investigación en Computación, Mexico City,
Mexico

batyr1@gmail.com

Abstract. In recent years, it has been shown that many correlation and association coefficients used in statistics can be viewed as functions defined on a set with an involution operation. In this case, the involution can be thought of as a mapping of elements of a set into “opposite” elements, with the correlation between mutually opposite elements being -1 . The methods constructing such correlation functions using similarity and dissimilarity functions defined over a set with involution have been proposed, and many known in statistics correlation and association coefficients have been constructed in such a way. It was shown that these correlation functions can be obtained by rescaling bipolar similarity functions; for this reason, they are referred to as similarity correlation functions. The paper reconsiders and summarizes some basic results on methods of constructing correlation functions on sets with involution called here involutive sets. The considered methods can be used for constructing new correlation functions on sets with involution if a suitable similarity or dissimilarity function is defined.

Keywords. Correlation, association, similarity, distance, involution.

1 Introduction

Correlation and association coefficients proposed in statistics play important roles in data analysis in biology, medicine, business, etc. [1-4]. Recently [5-8], a general approach to the analysis of relationships between data has been proposed based on works on fuzzy relations [9,10], aggregation functions and pseudo-difference operations [11], measures of similarity and interestingness [12-14], etc. In the works [5-8], a functional-algebraic approach was applied to the analysis of similarity measures, correlation, and association coefficients, which were considered as

functions of two arguments defined over a universal set with involution operation and satisfying several properties, such as symmetry, reflexivity, irreflexivity, inverse relationship etc.

With this approach, similarity and dissimilarity functions can also be viewed as fuzzy relations [9, 10]. The general methods forming new similarity and correlation functions with given properties on almost any domain are considered in [7]. Particular attention in [5, 6] is given to functions defined over sets with an involution operation, where the correlation and association coefficients are considered as correlation functions (association measures) satisfying the inverse relationship property.

The functional-algebraic approach, which considers similarity measures, correlation and association coefficients as functions defined over a set with involution operation, made it possible to establish a connection between these functions and to propose methods for constructing new correlation and association coefficients from similarity measures and distances [5-8, 15-20]. In the paper, we introduce the concept of an involutive set, with some changes in definitions. We reconsider and give proof of all statements.

The paper has the following structure. Section 2 introduces the basic properties of sets with involution used further for defining similarity, dissimilarity, and correlation functions over these sets. Section 3 considers the properties of co-symmetric functions and defines correlation function (association measure) over set with involution.

Section 4 considers co-symmetric, consistent, and bipolar similarity and dissimilarity functions and the methods of constructing correlation

functions from them. Section 5 contains the conclusion and directions for future research.

2 Involutive Sets

Definition 1. Let Ω be a non-empty set. A function $N: \Omega \rightarrow \Omega$ satisfying for all $x \in \Omega$ the involutivity property: $N(N(x)) = x$, is called an *involution* on Ω .

The involution is also referred to as a reflection or negation. The pair $\langle \Omega, N \rangle$ will be called an *involution algebra*. This is the simplest algebra with involution.

For example, on the power set $\mathcal{P}(X)$ of a non-empty set X , the complement \bar{A} of subsets A of X is an involution $N(A) = \bar{A}$ because $N(N(A)) = \overline{\bar{A}} = A$ for any subset A . The algebra $\langle \mathcal{P}(X), \bar{\cdot} \rangle$ is the involutive algebra associated with the Boolean algebra of sets $\langle \mathcal{P}(X), \cap, \cup, \bar{\cdot} \rangle$.

Definition 2. Let Ω be a non-empty set with an involution $N: \Omega \rightarrow \Omega$. Such a set will be referred to as an *involution set*. A non-empty subset V of Ω will be called an *involution subset* of Ω if V is closed under N , i.e., $N(x)$ belongs to V for all x in V .

An element $x \in \Omega$ satisfying the property $N(x) = x$ is called a *fixed point* of N . A set of fixed points of N in the set Ω will be denoted by $FP(\Omega, N)$ or FP . Depending on a set Ω and the definition of involution N on Ω , it may have no fixed points, one fixed point, or a finite or infinite number of fixed points.

For example, the complement \bar{A} of subsets $A \subseteq X$ in the power set $\mathcal{P}(X)$ has no fixed points because $\bar{A} \neq A$ for any subset A . The negation $-x$ of real numbers is an involution on the set of real numbers \mathbb{R} . It has one fixed point 0 because the equation $-x = x$ satisfied only for $x = 0$. On the set of real-valued n -tuples \mathbb{R}^n with elements $x = (x_1, \dots, x_n)$ the operation $N(x) = (-x_1, \dots, -x_n)$ is an involution with n -tuple $(0, \dots, 0)$ as a fixed point.

Theorem 1. Let Ω be a set with an involution N , and the set of non-fixed points $\Omega \setminus FP(\Omega, N)$ be non-empty. Then $\Omega \setminus FP(\Omega, N)$ is an involutive subset of Ω .

Proof. If x belong to $\Omega \setminus FP(\Omega, N)$ then x is not a fixed point: $N(x) \neq x$, and from the involutivity of N it follows: $N(N(x)) = x \neq N(x)$. Hence, $N(x)$ is not a fixed point and belongs to $\Omega \setminus FP(\Omega, N)$.

Therefore, the set $\Omega \setminus FP(\Omega, N)$ is closed under N , and it is an involutive subset of Ω ■

From theorem 1, it follows that the restriction $N|_{\Omega \setminus FP}$ of the involution N to $\Omega \setminus FP$ will be an involution on $\Omega \setminus FP$, and $\Omega \setminus FP$ will be an involutive set with involution $N|_{\Omega \setminus FP}$. Further, we define correlation functions on involutive subset without fixed points. Without the loss of generality, we will suppose that Ω has no fixed points, $FP(\Omega, N) = \emptyset$, $\Omega \setminus FP = \Omega$ and $N|_{\Omega \setminus FP} = N$.

For example, the correlation of real numbers with involution $N(x) = -x$ will be defined on the set $\Omega = \mathbb{R} \setminus \{0\}$ of real numbers without fixed point $x = 0$, see [15] and example in Section 4.

Below, we consider the main definitions and properties of (dis)similarity and correlation functions defined over involutive sets [5-8]. Some changes in definitions are based on the concept of involutive set introduced in this section.

3 Correlation Functions

Definition 3. Let Ω be a set with an involution N . The real-valued function $R: \Omega \times \Omega \rightarrow \mathbb{R}$ is called a *co-symmetric function* if, for all x, y in Ω , it satisfies the *co-symmetry* property:

$$R(N(x), N(y)) = R(x, y). \quad (1)$$

Theorem 2. Let Ω be a set with an involution N . The function $R: \Omega \times \Omega \rightarrow \mathbb{R}$ satisfies co-symmetry property (1) if and only if for all x, y in Ω it satisfies the property:

$$R(x, N(y)) = R(N(x), y). \quad (2)$$

Proof. Replacing x by $N(x)$ in (1) from involutivity of N obtain (2):

$$R(N(x), y) = R(N(N(x)), N(y)) = R(x, N(y)).$$

Similarly, from (2) and involutivity of N replacing y by $N(y)$ obtain (1):

$$R(N(x), N(y)) = R(x, N(N(y))) = R(x, y) \blacksquare.$$

Further, we will consider co-symmetric functions taking values in the intervals $[-1, 1]$ or $[0, 1]$. Due to the equivalence of two forms (1) and (2) of the co-symmetry property, we usually consider only one of them.

Definition 4. Let Ω be a set with involution N without fixed points. The function $A: \Omega \times \Omega \rightarrow$

$[-1,1]$ is called an *association measure* (*correlation function*) on Ω if for all x, y in Ω it satisfies the following properties:

$$A1. A(x, y) = A(y, x), \quad (\text{symmetry})$$

$$A2. A(x, x) = 1, \quad (\text{reflexivity})$$

$$A3. A(x, N(y)) = -A(x, y), \quad (\text{inverse relationship})$$

$$A4. A(x, N(x)) = -1. \quad (\text{opposite elements})$$

Note that A4 follows from A3 and A2.

Comments 1. Definition 4 is based on [5]. Further, the correlation functions will be obtained from similarity functions. They will also be referred to as *similarity correlation functions* to emphasize this property of correlation functions.

Comment 2. In some papers (for example, in fuzzy set theory), the term correlation coefficient denotes the functions $A(x, y)$, taking values in the interval $[-1,1]$ and satisfying the properties A1 and A2. Here, such functions will be referred to as *weak correlation functions* [6], and the correlation functions satisfying A1-A4 over involutive sets will be referred to as *strong correlation functions* or *strong similarity correlation functions*.

Proposition 1. The strong correlation function is co-symmetric.

Proof. Let us show that the strong correlation function satisfies for all x, y in Ω the co-symmetry property:

$$A(x, N(y)) = A(N(x), y). \quad (3)$$

From A3, A1, A3 and A1 obtain (3): $A(x, N(y)) = -A(x, y) = -A(y, x) = A(y, N(x)) = A(N(x), y)$ ■.

4 Similarity and Dissimilarity Functions

Consider similarity and dissimilarity functions used for constructing strong correlation functions.

Definition 5. Let Ω be a set with involution N without fixed points. The function $S: \Omega \times \Omega \rightarrow [0,1]$ is called a similarity function if for all x, y in Ω it satisfies the following properties:

$$S(x, y) = S(y, x), \quad (\text{symmetry}), \quad (4)$$

$$S(x, x) = 1. \quad (\text{reflexivity}). \quad (5)$$

The function $D: \Omega \times \Omega \rightarrow [0,1]$ is called a *dissimilarity function* if, for all x, y in Ω it satisfies the following properties:

$$D(x, y) = D(y, x), \quad (\text{symmetry}), \quad (6)$$

$$D(x, x) = 0. \quad (\text{irreflexivity}). \quad (7)$$

Definition 6. Similarity and dissimilarity functions are called *complementary* if for all x, y in Ω it is fulfilled:

$$S(x, y) + D(x, y) = 1. \quad (8)$$

These complementary functions can be obtained one from the other as follows:

$$S(x, y) = 1 - D(y, x), \quad D(x, y) = 1 - S(y, x). \quad (9)$$

Definition 7. Let Ω be a set with involution N without fixed points. Similarity and dissimilarity functions $S, D: \Omega \times \Omega \rightarrow [0,1]$ are called *consistent* if for all x, y in Ω they satisfy the following *consistency* properties, respectively:

$$S(x, N(x)) = 0, \quad (10)$$

$$D(x, N(x)) = 1. \quad (11)$$

When elements x and $N(x)$ are considered as "opposite" elements, the consistency of the similarity and dissimilarity functions means the minimal similarity and maximal dissimilarity between the "opposite" elements, respectively.

Definition 8. Let Ω be a set with involution N without fixed points. Similarity and dissimilarity functions $S, D: \Omega \times \Omega \rightarrow [0,1]$ are called *co-symmetric* if for all x, y in Ω they satisfy the following co-symmetry properties, respectively:

$$S(N(x), N(y)) = S(x, y), \quad (12)$$

$$S(x, N(y)) = S(N(x), y), \quad (13)$$

$$D(N(x), N(y)) = D(x, y), \quad (14)$$

$$D(x, N(y)) = D(N(x), y). \quad (15)$$

Theorem 3. Let Ω be a set with involution N without fixed points, and $S: \Omega \times \Omega \rightarrow [0,1]$ be a co-symmetric and consistent similarity function, then the function defined for all x, y in Ω by:

$$A(x, y) = S(x, y) - S(x, N(y)), \quad (16)$$

will be a strong correlation function (association measure) satisfying A1-A4.

Proof. From (16), symmetry (4), co-symmetry (13), and (4) obtain A1: $A(x, y) = S(x, y) - S(x, N(y)) = S(y, x) - S(x, N(y)) = S(y, x) - S(N(x), y) = S(y, x) - S(y, N(x)) = A(y, x)$.

From (16), reflexivity (5) and consistency (10) obtain A2: $A(x, x) = S(x, x) - S(x, N(x)) = 1 - 0 = 1$.

From (16) and involutivity of N obtain A3: $A(x, N(y)) = S(x, N(y)) - S(x, N(N(y))) = S(x, N(y)) - S(x, y) = -A(x, y)$.

A4 follows from A3 and A2: $A(x, N(x)) = -A(x, x) = -1$ ■.

Dually to (16), we obtain from (16) and (9) a correlation function from the complementary co-symmetric and consistent dissimilarity function:

$$A(x, y) = D(x, N(y)) - D(x, y). \quad (17)$$

Definition 9. Let Ω be a set with involution N without fixed points, and $S, D: \Omega \times \Omega \rightarrow [0, 1]$ be similarity and dissimilarity functions, respectively. These functions are called *bipolar* if for all x, y in Ω they satisfy the following conditions, respectively:

$$S(x, y) + S(x, N(y)) = 1, \quad (18)$$

$$D(x, y) + D(x, N(y)) = 1. \quad (19)$$

Theorem 4. Bipolar similarity S and dissimilarity D functions are co-symmetric and consistent.

Proof. From bipolarity (18) and reflexivity (5) of S obtain consistency (10) of S : $S(x, N(x)) = 1 - S(x, x) = 1 - 1 = 0$.

From bipolarity (18), symmetry (4), (18) and (4) obtain co-symmetry (13) of S : $S(x, N(y)) = 1 - S(x, y) = 1 - S(y, x) = S(y, N(x)) = S(N(x), y)$.

Similar results we obtain for bipolar dissimilarity function D ■.

From Theorems 3, 4, and (18), we obtain the following result.

Theorem 5. Let Ω be a set with involution N without fixed points, and $S: \Omega \times \Omega \rightarrow [0, 1]$ be a bipolar similarity function, then the function, defined for all x, y in Ω , by:

$$A(x, y) = 2S(x, y) - 1, \quad (20)$$

is a correlation function (association measure) satisfying A1-A4.

Dually, a bipolar dissimilarity function D defines a correlation function by:

$$A(x, y) = 1 - 2D(x, y). \quad (21)$$

From (20), it follows that the correlation function (association measure) is a rescaled bipolar similarity function.

Corollary 1. Complementary bipolar similarity and dissimilarity functions define the correlation function:

$$A(x, y) = S(x, y) - D(x, y). \quad (22)$$

Using (20) and (21), one can obtain bipolar similarity and dissimilarity functions from the strong correlation function:

$$S(x, y) = \frac{1}{2}(A(x, y) + 1), \quad (23)$$

$$D(x, y) = \frac{1}{2}(1 - A(x, y)). \quad (24)$$

From (8), (9), (18), and (19), it follows for bipolar complementary similarity and dissimilarity functions:

$$D(x, y) = S(x, N(y)), \quad (25)$$

$$S(x, y) = D(x, N(y)). \quad (26)$$

Formulas (8), (9), (18) - (26) define a *bipolar complementary correlation triplet* $\langle S, D, A \rangle$ that gives possibility to obtain from one function of this triplet other two functions.

Example 1. In [15], it was considered the problem of constructing a correlation function on the set of real numbers with involution defined by the negation of numbers: $N(x) = -x$. This involution has a fixed point $x = 0$. On the involutive set of real numbers without fixed points $\mathbb{R} \setminus \{0\}$, the bipolar similarity

$$S(x, y) = \frac{(x+y)^2}{2(x^2+y^2)}$$

and dissimilarity

$$D(x, y) = \frac{(x-y)^2}{2(x^2+y^2)}$$

functions have been introduced. These complementary functions define by (20) and (21) the following correlation function:

$$A(x, y) = \frac{2xy}{x^2+y^2}, \quad (27)$$

satisfying the properties A1-A4 of correlation functions.

5 Conclusion

In statistics, the prevailing view is that correlation coefficients are functions based on the standard deviations of measurements. For example, it was shown that the Spearman correlation can be obtained from the Pearson correlation coefficient. The considered here methods for constructing correlation functions from similarity and dissimilarity functions can be viewed as a generalization of Spearman's distance-based look on correlation coefficient defined by him by the distance between rankings similar to (21) [6, 8].

The essential point of the described here approach to the definition of correlation functions is considering them as functions defined over involutive sets that is not used in statistics. The new approach to constructing correlation functions and association measures shows [5-8] that many classical correlation and association coefficients can be introduced as functions defined on suitable involutive sets and satisfying the properties A3 and A4.

They can be obtained by rescaling bipolar similarity functions as in (20). Moreover, there exists a one-to-one correspondence between bipolar similarity and dissimilarity functions and corresponding correlation function that compose a bipolar complementary correlation triplet $\langle S, D, A \rangle$.

This new approach makes it possible to introduce correlation functions over almost any involutive sets using suitable similarity and dissimilarity functions. These new functions describe relationships between domain data related to the symmetry corresponding to the involution operation defined on data.

Based on the proposed methods, new correlation functions (association measures) for new data types have been introduced. For example, the paper [15] introduced correlation functions on the set of real numbers, see example (27) above. The paper [16] introduced the involutive negation on the set of finite probability distributions used in constructing correlation function on the set of finite probability distributions

and relative frequency distributions defined on the set of categorical data [17, 18].

Also, the involutive negation of probability distributions was used to introduce new, co-symmetric dissimilarity measures on the set of probability distributions [19]. The paper [20], based on the Jaccard similarity measure, introduced new strong similarity correlation functions on involutive sets of sets and binary n -tuples.

We plan to use the considered methods for constructing correlation functions over new types of involutive sets. Also, we plan to use more sophisticated methods of constructing correlation functions (association measures).

Acknowledgments

This work has been funded by projects SIP IPN 20240936 and Año Sabatico 2023-2024 IPN.

References

1. **Kendall, M. G. (1970).** Rank correlation methods. 4th ed., Griffin, London.
2. **Liebetrau, A. M. (1983).** Measures of Association. Sage Publications, Iowa.
3. **Chen, P. Y., Popovich, P. M. (2002).** Correlation: Parametric and nonparametric measures. Sage, No. 139.
4. **Gibbons, J. D., Chakraborti, S. (2003).** Nonparametric statistical inference. 4th ed. CRC Press is an imprint of the Taylor & Francis Group.
5. **Batyrrshin, I. Z. (2015).** On definition and construction of association measures. Journal of Intelligent & Fuzzy Systems, Vol. 29, No. 6, pp. 2319-2326. DOI: 10.3233/IFS-151930.
6. **Batyrrshin, I. Z. (2019).** Constructing correlation coefficients from similarity and dissimilarity functions. Acta Polytechnica Hungarica, Vol. 16, No. 10, pp. 191-204.
7. **Batyrrshin, I. (2019).** Towards a general theory of similarity and association measures: similarity, dissimilarity and correlation functions. Journal of Intelligent and Fuzzy Systems, Vol. 36, No. 4, pp. 2977-3004. DOI: 10.3233/JIFS-181503.

8. **Batyrshin, I. Z. (2019).** Data science: similarity, dissimilarity and correlation functions. In: Osipov, G., Panov, A., Yakovlev, K. (eds) Artificial Intelligence, Lecture Notes in Computer Science, Springer, Cham., Vol. 11866. DOI: 10.1007/978-3-030-33274-7_2.
9. **Zadeh, L. A. (1971).** Similarity relations and fuzzy orderings. Information Sciences, Vol. 3, No. 2, pp. 177–200. DOI: 10.1016/S0020-0255(71)80005-1.
10. **Fodor, J. C., Roubens, M. R. (2013).** Fuzzy preference modelling and multicriteria decision support. Springer Science & Business Media. Vol. 14.
11. **Grabisch, M., Marichal, J. L., Mesiar, R., Pap, E. (2009).** Aggregation functions. Cambridge University Press, Vol. 127.
12. **Clifford, H. T., Stephenson, W. (1975).** An introduction to numerical classification. Academic Press, New York.
13. **Batagelj, V., Bren, M. (1995).** Comparing resemblance measures. Journal of Classification, Vol. 12, pp. 73–90. DOI: 10.1007/BF01202268.
14. **Tan, P. N., Kumar, V., Srivastava, J. (2002).** Selecting the right interestingness measure for association patterns. 8th. Proceeding of. Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 32–41. DOI: /10.1145/775047.7750.
15. **Batyrshin, I. Z., Tóth-Laufer, E. (2022).** Bipolar dissimilarity and similarity correlations of numbers. Mathematics, Vol. 10, No. 5, p. 797. DOI: 10.3390/math10050797.
16. **Batyrshin, I. Z. (2021).** Contracting and involutive negations of probability distributions. Mathematics, Vol. 9, No. 19, p. 2389. DOI: 10.3390/math9192389.
17. **Rudas, I. J., Batyrshin, I. Z. (2023).** Explainable correlation of categorical data and bar charts. In: Shahbazova, S.N., Abbasov, A.M., Kreinovich, V., Kacprzyk, J., Batyrshin, I.Z. (eds) Recent Developments and the New Directions of Research, Foundations, and Applications, Studies in Fuzziness and Soft Computing, Springer, Cham. Vol. 422. DOI: 10.1007/978-3-031-20153-0_7.
18. **Batyrshin, I. Z., Rudas, I. J., Kubysheva, N., Akhtyamova, S. (2022).** Similarity correlation of frequency distributions of categorical data in analysis of cognitive decline severity in asthmatics. Computación y Sistemas, Vol. 26, No. 4, pp. 1603–1609. DOI: 10.13053/cys-26-4-4439.
19. **Ensastegui-Ortega, M. E., Batyrshin, I., Gelbukh, A., Kubysheva, N. (2024).** Analysis of relationships between co-symmetric dissimilarity measures of probability distributions with involutive negations. Computación y Sistemas, Vol. 28, No. 2. DOI: 10.13053/CyS-28-2-5049.
20. **Batyrshin, I., Rudas, I. (2024).** New similarity correlation functions for sets and binary data based on Jaccard similarity measure. 2024 IEEE 18th International Symposium on Applied Computational Intelligence and Informatics, pp. 000145–000150. DOI: 10.1109/SACI60582.2024.10619901.

Article received on 01/09/2024; accepted on 23/10/2024.
*Corresponding author is Ildar Z. Batyrshin.