




Statistical dependence and shape of Young tableau

Jesús E. García , Verónica Andrea González-López , and Maria Magdalena Kcala Alvaro* 

Department of Statistics, University of Campinas, Sergio Buarque de Holanda, 651, Campinas, SP CEP: 13083-859, Brazil

Received 16 September 2022, Accepted 27 March 2023

Abstract – Given two continuous random variables, X and Y , we study the relationship between their statistical dependence and the Young tableau of the permutation defined from the graph of a bivariate sample coming from (X, Y) . From a sample of size n of (X, Y) , we identify the Young tableau of the permutation which maps the ranks of the X observations on the ranks of the Y observations. Procedures to detect statistical dependence between pairs of random variables, based on statistics calculated on the permutation defined by the graph of a bivariate sample have been developed, see García and González-López (2020) [Symmetry 12, 9, 1415. <https://doi.org/10.3390/sym12091415>] and García and González-López (2014) [J Multivar Anal 127, 126–146. <https://doi.org/10.1016/j.jmva.2014.02.010>]. In those papers, the information used is the length of the longest increasing (decreasing) subsequence, identified as the first line (the first column) of the Young tableau of the permutation. In this paper, we expose the information captured by the shape of the Young tableau of the permutation.

Keywords: Permutation, Young tableau, Statistical dependence

Introduction

Detecting dependence between random variables is one of the main problems in statistics and applied sciences in general. There are diverse methodologies for the problem of detecting dependence between continuous random variables. Some of them are based on dependence coefficients, such as Pearson's rho, Spearman's rho, and Kendall's tau. There are also hypothesis tests procedures based on those coefficients as well as other tests not based on correlation type coefficients as, for example, Hoeffding's test (see Hoeffding [1]) and the independence test from Genest and Rémillard [2].

Recently, has been proposed a new class of independence tests based on the length of the longest increasing subsequence (LISS) and the length of the longest decreasing subsequence (LDSS) of the permutation defined by a sample of a bivariate random vector (see e.g., García and González-López [3, 4]). In this paper, we propose the use of the Young tableau of the permutation to study the possible dependence between two random variables. The Young tableau of the permutation contains much more information about the dependence in the sample than the LISS and LDSS, which are obtained from the first line and the first column of the Young tableau, respectively. We show, through examples, that the shape of the empirical Young tableau depends on the kind of dependence. It is important to note that the shape of the Young tableau only

depends on the copula of the joint distribution of the two random variables and not on the marginal distributions.

The paper is organized as follows. Firstly, we explain how to find the permutation defined from a sample. Secondly, the computation of Young tableau of a permutation is given. Thirdly, we study how the profile of the Young tableau changes with diverse types of dependence. The last section deals with our conclusions.

Permutation defined from a sample

Definition 2.1

Let $D = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ be n independent realizations of the continuous random vector (X, Y) , with joint cumulative distribution function H . The random permutation $\pi_D: \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ corresponding to D is defined by,

$$\pi_D(\text{rank}(X_i)) = \text{rank}(Y_i), \quad i = 1, \dots, n.$$

Remark 2.2

Note that because of the continuity of the distribution, we do not have the presence of ties, being the function detailed in Definition 2.1 a one-to-one function.

Remark 2.3

We observe that the pairs $(\text{rank}(X_i), \text{rank}(Y_i))$, $i = 1, \dots, n$ define the empirical copula of the sample. For a

*Corresponding author: m229256@dac.unicamp.br

continuous random vector with joint distribution H , the distribution of the empirical copula does not depend on the marginal distributions of X or Y . Its distribution depends only on the copula corresponding to the joint distribution H .

Example 2.4

Consider the following sample of a random vector (X, Y) ,

$$D = \{(6.2, 4.1), (7.3, 5.0), (3.2, 3.2), (9.6, 2.0), (6.8, 6.1), (4.9, 3.3), (4.7, 6.3), (9.1, 4.4), (8.1, 3.7)\}.$$

Table 1 shows the sample sorted on the X coordinate and the corresponding marginal ranks.

The resulting permutation is $\pi_D = \{2, 9, 3, 5, 8, 7, 4, 6, 1\}$. Figure 1a, shows the scatter plot of the sample. Figure 1b, shows the scatter plot for the corresponding ranks of the sample. Note that the permutation can be defined directly from the sample X versus Y graph.

The definition presented below allows us to catalog the subsequences that could be formed with the elements of $\pi_D = \{2, 9, 3, 5, 8, 7, 4, 6, 1\}$, and what is observed in this classification is whether the subsequence increases or decreases.

Definition 2.5

Given the size n set $Q = \{q_1, \dots, q_n\} \subset \mathbb{R}$,

- (i) the subsequence $\{q_{i_1}, \dots, q_{i_k}\}$ of Q is an increasing subsequence of Q if for $i_1 < \dots < i_k, q_{i_1} < q_{i_2} < \dots < q_{i_k}$,
- (ii) the subsequence $\{q_{i_1}, \dots, q_{i_k}\}$ of Q is a decreasing subsequence of Q if for $i_1 < \dots < i_k, q_{i_1} > q_{i_2} > \dots > q_{i_k}$.

For instance, for the permutation $\pi_D = \{2, 9, 3, 5, 8, 7, 4, 6, 1\}$ in the example, we can identify increasing subsequences like $\{2, 3, 6\}$ or $\{2, 3, 5, 6\}$ and decreasing subsequences like $\{9, 8, 7, 6\}$ or $\{9, 8, 7, 4, 1\}$. This identification allows us to verify trends between the values of X and Y , say, the more positively correlated, the longer the increasing subsequences will be, in the same way, that the more negatively correlated, the longer the decreasing subsequences.

With a suitable procedure we can identify all the increasing subsequences and all the decreasing subsequences that can be built with the elements of $\pi_D = \{2, 9, 3, 5, 8, 7, 4, 6, 1\}$. In this example, the length of the longest increasing subsequence is 4 and is obtained from the subsequences $\{2, 3, 5, 6\}$, $\{2, 3, 4, 6\}$, $\{2, 3, 5, 7\}$ and $\{2, 3, 5, 8\}$. The length of the longest decreasing subsequence for π_D is 5 and is reached by the subsequences $\{9, 8, 7, 4, 1\}$ and $\{9, 8, 7, 6, 1\}$.

In the following section is presented a strategy (Schensted [5]) to extract from the permutation π_D the increasing subsequences and the decreasing subsequences that can be formed with the permutation π_D . We show how the strategy works in a concrete case, furthermore, we formally define a Young tableau and its shape.

Table 1. Artificial data set, and its marginal ranks.

x_i	y_i	Rank (x_i)	Rank (y_i)
3.2	3.2	1	2
4.7	6.3	2	9
4.9	3.3	3	3
6.2	4.1	4	5
6.8	6.1	5	8
7.3	5.0	6	7
8.1	3.7	7	4
9.1	4.4	8	6
9.6	2.0	9	1

Young tableau of a permutation

In this section we define the Young tableau and its shape, and we show how to obtain the Young tableau given a permutation.

Definition 3.1 (Schensted [5])

A standard Young tableau of order n is a disposition of n natural numbers, in left-justified rows of strictly increasing numbers and columns of strictly increasing numbers, without gaps between numbers, so that the lengths of the rows are non-increasing, and each column has an element in the first row.

Example 3.2

Example of a possible Young tableau for $n = 9$.

1	3	4	6
2	7		
5			
8			
9			

Definition 3.3

The shape of a standard Young tableau is the arrangement of squares resulting of replacing each number by one square in the standard Young tableau.

The shape of Definition 3.3 can be represented by a vector, in which we do not consider the values from 1 to n entered in each square of the Young tableau. That vector counts line by line (from top to bottom) the total number of squares. We will represent the shape of a standard Young tableau T of order n with K rows by the vector $S(T) = (s_1, \dots, s_n)$, where for $1 \leq i \leq K, s_i$ is the length of the row i of the Young tableau, and for $i: K < i \leq n, s_i = 0$.

In the following example, we take the case given in Example 3.2 and rescue only its shape (scheme without numbers) and then we construct S .

Example 3.4

Young tableau from Example 3.2 and its corresponding shape (see Definition 3.3)

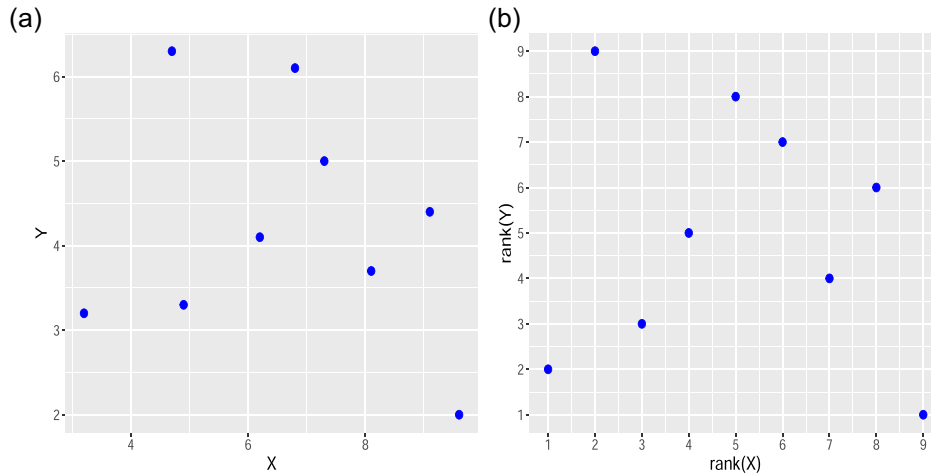


Figure 1. Scatter plots for the sample in Example 1. (a) X vs. Y . (b) $\text{ranks}(X)$ vs. $\text{ranks}(Y)$.



For this Young tableau, $S = (4, 2, 1, 1, 1, 0, 0, 0, 0)$.

To simplify the exposition, from now on, we will use a short notation for the vector representing the shape of a Young tableau which consists on only showing the non zero coordinates of S . In this way the shape of the Young tableau from Example 3.4, will be represented in the short notation by $S = (4, 2, 1, 1, 1)$.

We describe now an algorithm, introduced in Schensted [5], used to build a Young tableau.

Schensted insertion algorithm

Given a Young tableau T , and a number x , the operation

$$T \leftarrow x \text{ Schensted insertion}$$

is defined by

1. x will be inserted on the first row of T in the following way:
 - (a) if x is larger than all the numbers in the row, then x is added at the end of the row,
 - (b) otherwise, x will replace the smallest number in the row that is larger than x .
2. If x replaced a number from the first row, then this number will be inserted in the second row following the same rules.
3. Repeat this process row by row until some number is added at the end of a row.

In the following section, we show the notions that are defined by applying the algorithm for identifying the Young

tableau, on a permutation generated from a sample, as is the case of π_D reported in Example 2.4.

Standard Young tableau of a permutation

Consider now a sequence of numbers x_1, x_2, \dots, x_n (without repetitions). The standard Young tableau of the sequence is defined as

$$(\dots(x_1 \leftarrow x_2) \leftarrow x_3) \dots) \leftarrow x_n \text{ Schensted insertion}$$

Definition 3.5

The standard Young tableau of a permutation π_D , defined on $\{1, 2, \dots, n\}$ is the Young tableau resulting of applying the Schensted insertion algorithm to the sequence $\{\pi_D(1), \pi_D(2), \dots, \pi_D(n)\}$.

Remark 3.6

The length of the longest increasing subsequence for the permutation, is the length of the first row of the Young tableau shape, while the length of the longest decreasing subsequence for the permutation, is the length of the first column of the Young tableau shape. For more information about the properties of the longest increasing subsequence (see Romik [6]).

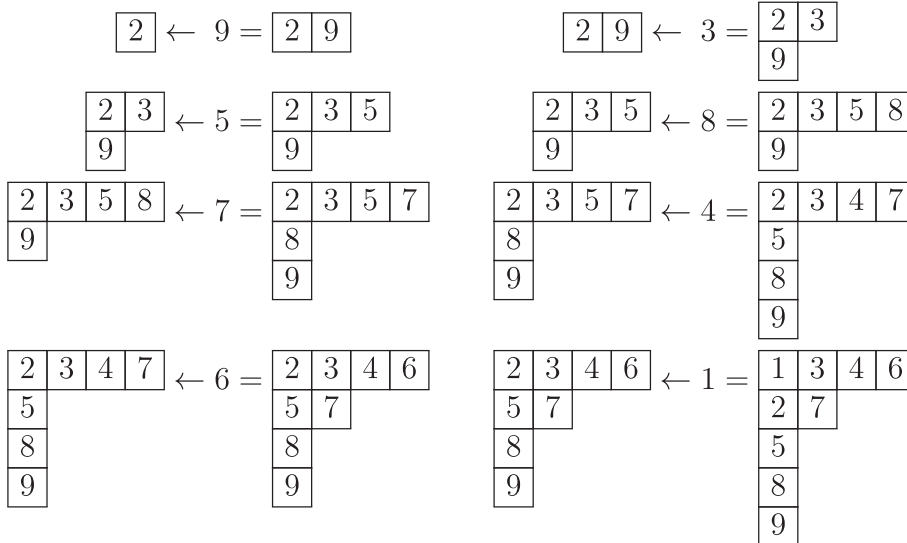
The example to follow considers the permutation of Example 2.4 and builds step by step the Young tableau, its shape, and the vector S .

Example 3.7

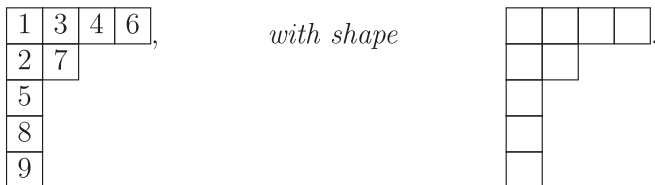
Consider the permutation coming from Example 2.4, where

$$(\pi_D(1), \pi_D(2), \dots, \pi_D(n)) = (2, 9, 3, 5, 8, 7, 4, 6, 1).$$

The step by step construction of the corresponding Young tableau (from left to right and from top to bottom) is given by



The Young tableau corresponding to the permutation π_D is



As we can see this is the Young tableau of Example 3.4 and with the short notation it can be represented by $S = (4, 2, 1, 1, 1)$.

The next section investigates the forms (shapes) taking by the Young tableau under certain distributions. The idea is to extract these shapes to represent the dependence types. We observe that this goes beyond finding the Young tableau to identify the length of the longest increasing subsequence or the length of the longest decreasing subsequence (see Remark 3.6) which was the proposal of García and González-López (2020) [3] and García and González-López (2014) [4], in this paper, we investigate the shape of the Young tableau comparing it with the type of dependence existing in the data.

Young tableau and dependence

This section shows how the Young tableau changes with the dependence type. We will consider diverse types of dependence, including (i) independence, (ii) positive Pearson correlation, (iii) negative Pearson correlation, and (iv) two cases of dependence with zero Pearson correlation. In all the examples, composed by artificial data, the size of the sample is the same, $n = 200$, and because of this all the shapes have the same area (200), here area is the number of squares in the Young tableau.

The first case is the independence between two random variables. Figure 2a shows a size 200 sample of the random

vector (X, Y) , where X and Y are independent with Uniform distribution on the interval $(0, 1)$. Figure 2b shows the shape of the corresponding Young tableau.

Using the short notation, the shape can also be described by the vector

$$S = (24, 22, 21, 17, 15, 14, 11, 10, 10, 9, 7, 7, 6, 5, 5, 3, 3, 3, 2, 2, 1, 1, 1, 1).$$

The second case corresponds to positive Pearson coefficient, $\rho = 0.7$ (positive dependence). Figure 3a shows a size 200 sample of the random vector (X, Y) , with bivariate Normal distribution and correlation 0.7, X and Y with expected value equal to zero and variance equal to 1. Figure 3b shows the shape of the corresponding Young tableau.

In this example, using the short notation,

$$S = (43, 31, 21, 20, 19, 16, 13, 11, 8, 7, 5, 3, 2, 1).$$

Note that the shape of the Young tableau (reported in Fig. 3b) shows much larger values for the length of the initial rows, compared to the independence case, Figure 2b.

The third case corresponds to negative Pearson coefficient, $\rho = -0.7$ (negative dependence). In Figure 4a, it is shown a size 200 sample of the random vector (X, Y) , with bivariate Normal distribution and correlation -0.7 , X and Y with expected value equal to zero and variance equal to 1. Figure 4b shows the shape of the corresponding Young tableau.

Using the short notation, the shape of the young tableau for this example, corresponds to

$$S = (14, 13, 12, 12, 10, 10, 9, 9, 9, 9, 8, 7, 6, 5, 5, 5, 5, 5, 5, 4, 4, 4, 3, 3, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1).$$

In the case reported in Figure 4, the shape of the Young tableau shows much larger values for the length of the initial columns, compared to the independence case,

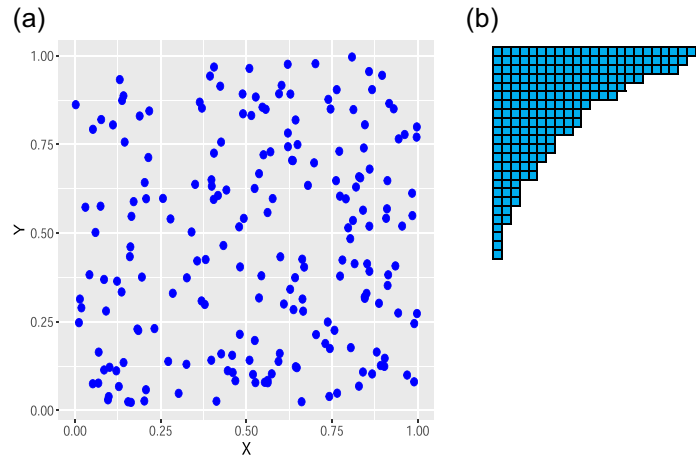


Figure 2. (a) Scatter plot X vs. Y . (b) Young tableau shape. For an independent sample of size $n = 200$ of (X, Y) with margins Uniform in $(0,1)$.

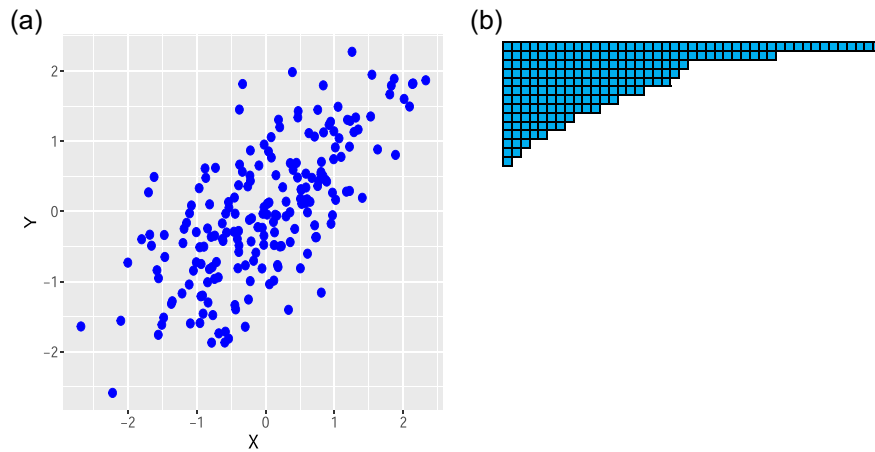


Figure 3. (a) Scatter plot X vs. Y . (b) Young tableau shape. For a sample of size $n = 200$ of (X, Y) bivariate Normal distribution with correlation 0.7.

Figure 2b. The shape in Figure 4b is a reflection of the shape given in Figure 3b.

If we compare case (i) with (ii) or (iii), we can see that the vector S captures what is happening in the shapes of each Young tableau, allowing us to speculate that S could represent each phenomenon. Despite this, research on it is beyond the purpose of this article.

The following two distributions (case (iv)), are cases of zero Pearson correlation, which are usually tricky for any method of dependence detection (see García and González-López (2014) [4]).

The fourth case corresponds to a mixture of 50% bivariate Normal distribution with Pearson correlation 0.7 and 50% bivariate Normal distribution with Pearson correlation -0.7 . The Pearson correlation for this distribution is zero. Figure 5 shows a size 200 sample of the random vector (X, Y) , with this mixture. Figure 5b shows the shape of the corresponding Young tableau.

Using the short notation, the shape of the Young tableau for this example is

$$S = (39, 26, 24, 17, 11, 9, 8, 7, 6, 5, 5, 3, 3, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1).$$

For this mixture (Fig. 5a), the shape of the Young tableau (Fig. 5b) shows larger values for the length of the initials rows and also for the length of the initials columns when compared to the independence case (Fig. 2b).

The last setting corresponds to Uniform distribution on the disk of radius 1. This distribution also has zero Pearson correlation and is hard to detect using traditional methods (see García and González-López (2020) [3]; García and González-López (2014) [4]). In Figure 6a, it is shown a size 200 sample. Figure 6b shows the shape of the corresponding Young tableau.

Using the short notation, we obtain the following vector S related to the shape of the Young tableau,

$$S = (23, 19, 17, 16, 16, 13, 13, 11, 11, 9, 9, 8, 7, 6, 6, 4, 3, 2, 1, 1, 1, 1, 1, 1, 1).$$

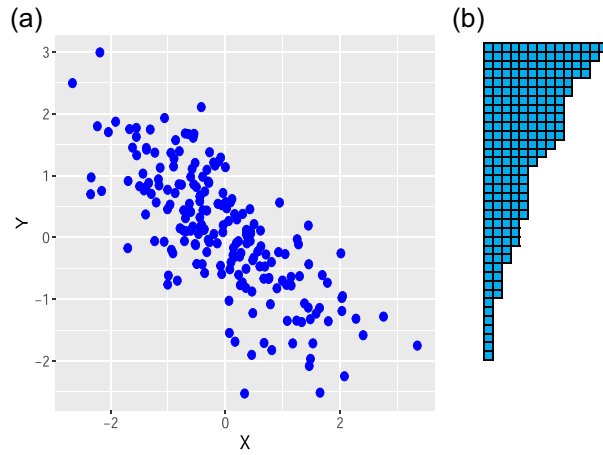


Figure 4. (a) Scatter plot X vs. Y . (b) Young tableau shape. For a sample of size $n = 200$ of (X, Y) bivariate Normal distribution with correlation -0.7 .

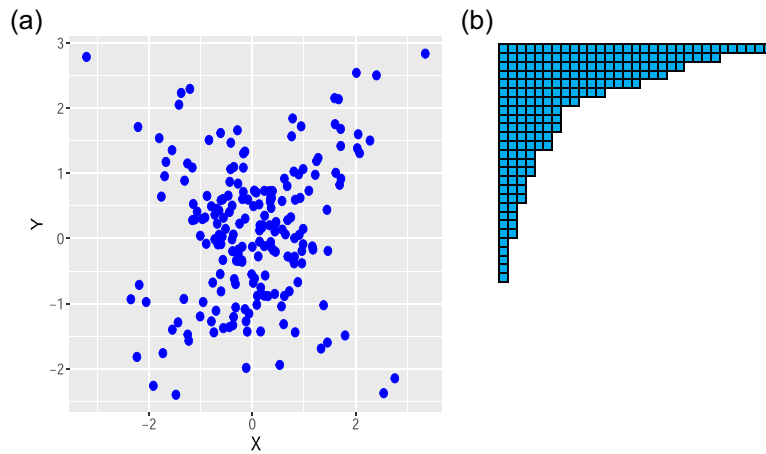


Figure 5. (a) Scatter plot X vs. Y . (b) Young tableau shape. For a sample of a mixture of 50% bivariate Normal distribution with Pearson correlation 0.7 and 50% bivariate Normal distribution with Pearson correlation -0.7 . Sample of size $n = 200$.

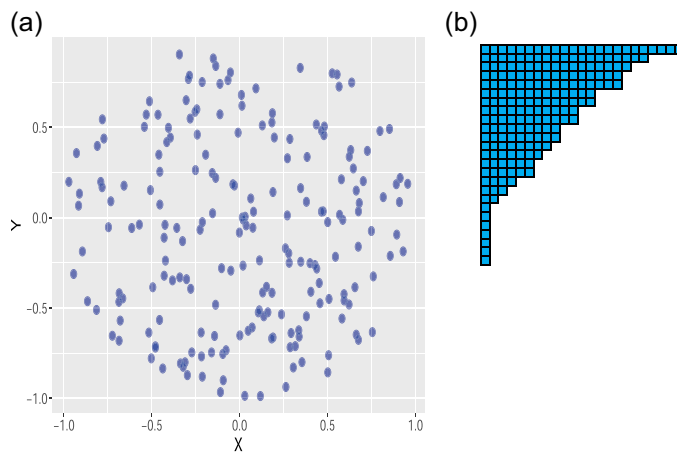


Figure 6. (a) Scatter plot X vs. Y . (b) Young tableau shape. For a size 200 sample of bivariate Uniform distribution on the disk of radius 1.

Note that the initial rows and columns are shorter than in the independence case (i), compare Figure 6b with Figure 2b. For the other rows and columns, the shape of the Young tableau shows an almost linear behavior in the way that the length of the consecutive rows changes.

Conclusions

This paper introduces the concept of the Young tableau shape as a tool for describing types of dependence that could exist in bivariate continuous random vectors. The Young tableau is built from the algorithm introduced in Schensted (1961) [5], and it is applied to the permutation that associates the ranks of the observations of X with the ranks of the observations of Y . From the Young tableau, several indicators can be obtained, as is the case of the shape of the Young tableau, allowing a representation of the type of dependence in a sample. In previous research (García and González-López [3]; García and González-López [4]), the length of the longest increasing subsequence and the length of the longest decreasing subsequence are used to detect independence. Whereas in this article we aim to use not only the size of such subsequences but to use the size of all possible increasing/decreasing subsequences.

We show a simulation study computing the shape of the Young tableau for several settings (i) independence, (ii)–(iii) Normal distributions and (iv) cases with zero Pearson correlation. For each situation, by extracting the Young tableau shape, we obtain a profile for each dependence type. We show how the shape of the Young tableau is altered for diverse types of dependence. For positive dependence, we observe that the length of the initial row is larger than the length of the initial column. For negative dependence, we observe the opposite, the length of the initial row is smaller than the length of the initial column. For the two cases of dependence with zero Pearson correlation,

we observe that the length of the initial row is similar to the length of the initial column, and those lengths are different (compare Fig. 6 with Fig. 5). Our findings provide evidences that the shape of the Young tableau can be appropriated to develop procedures of dependence detection.

Acknowledgments

Maria Magdalena Kcala Alvaro gratefully acknowledge the financial support provided by CAPES with a fellowship from the PhD Program in Statistics – University of Campinas. The authors wish to thank the referees and editors for their many helpful comments and suggestions on an earlier draft of this paper.

References

1. Hoeffding W (1948), A non-parametric test of independence. *Ann Math Statist* 19, 4, 546–557. <http://dml.mathdoc.fr/item/1177730150>.
2. Genest C, Rémillard B (2004), Test of independence and randomness based on the empirical copula process. *Test* 13, 335–369. <https://doi.org/10.1007/BF02595777>.
3. García JE, González-López VA (2020), Random permutations, non-decreasing subsequences and statistical independence. *Symmetry* 12, 9, 1415. <https://doi.org/10.3390/sym12091415>.
4. García JE, González-López VA (2014), Independence tests for continuous random variables based on the longest increasing subsequence. *J Multivar Anal* 127, 126–146. <https://doi.org/10.1016/j.jmva.2014.02.010>.
5. Schensted C (1961), Longest increasing and decreasing subsequences. *Can J Math* 13, 179–191. <https://doi.org/10.4153/CJM-1961-015-3>.
6. Romik D (2015), *The surprising mathematics of longest increasing subsequences*, Cambridge University Press, New York. <https://doi.org/10.1017/CBO9781139872003>.

Cite this article as: García JE, González-López VA & Kcala Alvaro MM 2023. Statistical dependence and shape of Young tableau. *4open*, 6, 4.