Psychological Reports, 2000, 86, 457-465.

NOTE ON MATRICES OF RANDOM NUMBERS[1]

HORACIO J. A. RIMOLDI

CONICET–U B A, Universidad del Salvador

E. BEI DE LIBONATTI, A. SALINAS, AND F. LAGOUTTE

CIIPME: CONICET, Universidad de Buenos Aires
Universidad del Salvador

Summary.—The statistical and structural characteristics of 13 matrices of random numbers in which both the cells and the entries were randomly chosen are discussed. Each matrix was explored considering row means, standard deviations, and correlations as well as column means, standard deviations, and correlations. A study concerning the sequential arrangement of digits was performed by finding out in tables of random numbers how many times the values 0 to 9 are followed by any other digit. Analyses indicate clear factor structures when factor analyzing correlations of rows and of columns and when examining sequential arrangements, concluding that for a given set of digits it is possible to assert both randomness and nonrandomness depending on how the data are examined.

The literature related to the concepts of chaos, random, order, and so forth is so vast that it would require a lengthy presentation to discuss the mathematical, statistical, and practical connotations involved, including the application of the concepts to interpret educational, biological, economic, and other kinds of problems. Only a few studies that relate to our problem are examined, running the risk of missing many significant contributions to the problem.

As it is well known the words random and randomness do not seem to have a clearly defined meaning. As far back as 1961 Feller stated that "The word random is not well defined" so that, when speaking of random samples, the adjective random implies "that all possible samples" of the same size, "have the same probability, namely n^{-r} in sampling with replacement and $1/(n)_r$ in sampling without replacement" (p. 29).

Recently, Mandelbrot (1996) discussed the kind of randomness needed to explain discontinuous functions and strong fluctuations. The development of fractal geometry led him to conclude that the idea of a single random model, including the computation of probabilities, may be misleading. This suggestion deserves a careful consideration from both the theoretical and

[1]Address enquiries to H. J. A. Rimoldi, Centro Interdisciplinario de Investigaciones en Psicología Matemática y Experimental, Tte. Gral. Perón 2158, 1040 Buenos Aires, Argentina or e-mail (ciipme@ssdnet.com.ar).

practical connotations of probability theory as usually understood. Other authors have discussed the idea of randomness and chaos in interpreting biological, physical, sociological, educational, and other kinds of data.

In 1971 Iverson, Longcor, Mosteller, Gilbert, and Youtz asked the question: "How well the laws of chance actually work?" (p. 1), examining what happens when a die is thrown and its outcome recorded considering imperfections in the die, in the throwing, in the perception of the outcome, and in recording the obtained results. After several million trials were completed, they concluded that "the dice and the experimenter seem to have behaved nearly like an unbiased, independent randomized device for the variables that we have measured" (1971, p. 14).

The concept of randomness was examined by Ayton, Hunt, and Wright (1991) and by Falk (1991). Both authors agreed on the lack of a clear definition of the concept and, it might be added, on the danger implied in some statements based on the concept of randomness and its relationship to the idea of chaos.

Crutchfield, Farmer, Packard, and Shaw (19986) stated that randomness limits prediction and that unsuspected relationships may be found in chaos, adding that certain structures may be found in what is considered to be random. A similar opinion has been held by Solé, Bascompte, Delgado, Luque, and Manrubia (1996) when stating that complex systems do appear between order and disorder, as can be found in ant-colonies, tropical forests, and other physical and biological phenomena. Also Goldberger, Rigney, and West (1990, p. 36) stated that "physiological systems may behave most erratically when they are young and healthy. Counterintuitively, increasingly regular behavior sometimes accompanies aging and disease." They further state that "Under some circumstances deterministic nonlinear systems . . . behave erratically, a state called chaos."

Examining the performance of 32 subjects asked to generate random digits, Ginsburg and Karpiuk (1994, p. 1065) found three factors that correspond to cycling seriation and repetition, so that the "ability to generate random sequences turns out to be a nonunitary trait."

The main purpose of this article is centered on the statistical and factorial analysis of collections of random numbers selected following different procedures. The study was mostly concerned with (1) the descriptive and structural characteristics that seem to be common to different matrices of random numbers and (2) the possible existence of sequential arrangements in matrices of random numbers.

Description of Matrices of Random Numbers Used in This Article

A total of 13 matrices were prepared following different procedures. The five matrices included in the first group were built by selecting randomly the cells in which the entries would be recorded. The matrices of the second group contain numbers randomly selected from Arkin and Colton (1962), from tables of random numbers of Meredith (1971), and from 1,500 extractions with reposition from a basket containing 10 balls numbered from 0 to 9. Besides 3,000 digits obtained using Microsoft Visual Basic® V© 1997 were randomly assigned to Matrices C_1, C_2, and C_3. All these matrices are listed in Table 1.

In the third group the frequency with which sequences of single digits and pairs of digits as they occur in tables of random numbers as those of Arkin and Colton (1962), Meredith (1971), the 1,500 randomly selected numbers used to build Matrix B, as well as those obtained using Microsoft Visual Basic® V© 1997 were recorded in Tables S_1 to S_5 that showed the frequency with which every number from 0 to 9 was followed by itself or by any other digit.

The first column of Table 1 gives the name and size of the matrices. For instance, Matrix M has 25 rows and nine columns, while A_1 is a matrix with 250 rows and 25 columns. Under "Analysis by Columns," Column A indicates the difference between the means of the columns, Column B gives the number of correlations between columns, and under C is shown the significance of the differences between the column correlations. Under the title "Analyses by Rows" are given for rows the same kind of information described under the title "Analysis by Columns." For example, Matrix A_2 has 320 rows and 25 columns; the differences among the means of all the columns are significant at the 1% level. There are 300 correlations between the columns of which 7% are significant at the 5% level and 1% at the 1% level, and 51,040 correlations between the rows, of which 6% are significant at the 5% level and 2% at the 1% level.

In building the described matrices different procedures were followed to assign values to the cells. There is no question that other procedures might have been used, but the approach, as briefly described in the Appendix (p. 465), ensured a reasonable diversity that may help to interpret the obtained results.

Method

(1) Analyses of variance were performed to find the significance of the differences between the means of the rows and of the columns in all the matrices. The homogeneity of variance for the rows and columns was calculated using both Cochran's and Hartley's tests. (2) The correlations and inner product of the rows and column vectors of all the matrices were calculated and factor analyzed. For the principal axes, varimax and promax solutions were computed. (3) Chi-squared tests were used to assess the significance of differences between the expected and the observed values in all the cells of

the five matrices S_1 to S_5. Where the frequency with which every digit was followed by any other, the digit was recorded. The chi squared tests were performed using for the expected values (a) the assumption of homogeneity, that is, equal frequencies in all the cells and (b) the expected value resulting from the product of the row and column totals for the corresponding entry divided by the total. These values were computed for the whole table as well as for each row and each column of the table showing in all cases significance at the 1% level.

Results

Except for Matrices M and N where the difference between row means was significant at the 1% level, and Matrix A_1 wherein the differences between column means were significant at the 1% level, in all the other cases no difference between rows or between columns reached the 5% level as shown in Table 1.

TABLE 1
ANALYSES BY COLUMNS AND BY ROWS OF THE MATRICES

Matrix		No. Cells	Analysis by Columns				Analysis by Rows			
			A	B	C		D	E	F	
					5%	1%			5%	1%
M	25 × 9	225	no	36	6	0	sig	300	18	3
N	25 × 9	225	no	36	6	3	sig	300	5	1
M_1	9 × 10	90	no	45	4	2	no	36	3	0
N_1	9 × 10	90	no	45	11	7	no	36	5	3
L(M+N)	50 × 9	450	no	36	3	0	sig	1225	7	0
A_1	250 × 25	6250	no	300	8	1	no	31125	5	1
A_2	320 × 25	8000	sig	300	7	1	no	51040	6	2
T_1	30 × 9	270	no	36	8	3	no	435	11	6
T	100 × 10	1000	no	45	2	2	no	4950	5	1
B	50 × 30	1500	no	435	0	0	no	1225	6	2
C_1	100 × 30	3000	no	435	5	1	no	4950	7	2
C_2	200 × 15	3000	no	105	8	1	no	19900		
C_3	300 × 10	3000	no	45	9		no	44850		
Total				1899				160372		

Note.—A: Difference between column means at 1% level, B: Number correlations by columns, C: Proportion of significant column correlations at the 5% and 1% levels, D: Difference between row Means at 1% level, E: Number of row correlations, and F: Proportion of significant row correlations at the 5% and 1% levels.

Chi squared tests were performed to test the assumption that in terms of randomness all digits should appear with approximately the same frequency. All were statistically significant only at above the 25% level suggesting nonsignificant differences at usually accepted levels in the frequency with which each digit occurs. The various tests performed to evaluate the homogeneity of variance also indicated nonsignificance ($p > .05$) for either rows or columns. In short, means, variances, and frequencies did not show significant differences except for Matrices M, N, and A_1 as previously stated.

The inner product between the row vectors and the column vectors of all the matrices as well as their correlations were calculated considering the 5% and 1% levels. At each level and for both rows and columns the proportion of significant correlations exceeded the given level, suggesting some kind of relationship between the rows and or between the columns in the matrices here examined. These proportions are shown in Table 1 for the 5% and 1% levels.

The varimax solution for the correlations between the columns of Matrices C_1, C_2, and C_3 show that (a) in the (100 × 30) C_1 matrix, 13 factors explain 90% of the variance due to the 435 correlations among the 30 columns, (b) in the (200 × 15) C_2 matrix, eight factors explain 94% of the variance due to the 105 correlations among the 15 columns, and (c) in the (300 × 10) matrix C_3, five factors explain 94% of the variance due to the correlations among the 10 columns of the matrix.

The association between values was further explored by factor analyzing both the correlations and the inner-products of the nontransformed values of the rows and columns of the matrices here reported. It is not possible to show all the obtained results but the overall findings indicate that in all cases the number of factors explaining between 90% and 95% of the variance is less than the number of variables involved. In our case the number of variables are either the number of rows or the number of columns of the corresponding matrices.

TABLE 2
VARIMAX SOLUTION FOR CORRELATIONS BETWEEN COLUMNS OF MATRIX N

Column	Factor			
	A	B	C	D
1	–.00	–.41	.61	.08
2	–.02	–.60	–.09	–.01
3	.63	–.06	.15	.33
4	–.30	.10	–.13	.00
5	.11	–.19	.05	.54
6	.28	.25	–.63	.16
7	–.09	.55	–.17	–.03
8	–.06	.06	–.05	.61
9	.73	–.06	–.14	.02

These results indicate associations between the rows and between the columns of matrices randomly prepared. One example of the results obtained when performing a factor analysis (varimax solution) of the intercorrelations among the nine columns of Matrix N is shown in Table 2. Taking

.40 as an arbitrary limit, we find that Columns 3 and 9 converge in Factor A. In bipolar Factor B Columns 1 and 2 appear on the negative side and Column 7 on the positive side. In Factor C Columns 1 and 6 with different signs converge, while Columns 5 and 8 define Factor D. Similar results were obtained when factorizing the correlations between the rows or between the columns of all the matrices. As an example, the varimax solution for the correlations among the nine columns of Matrix N shows that 95% of the variance was explained by four factors. Fig. 1 shows that Factor A is essentially defined by Columns 3 and 9 of Matrix N, while in bipolar Factor B, Columns 1 and 2 define the negative pole and Column 7 the positive end.

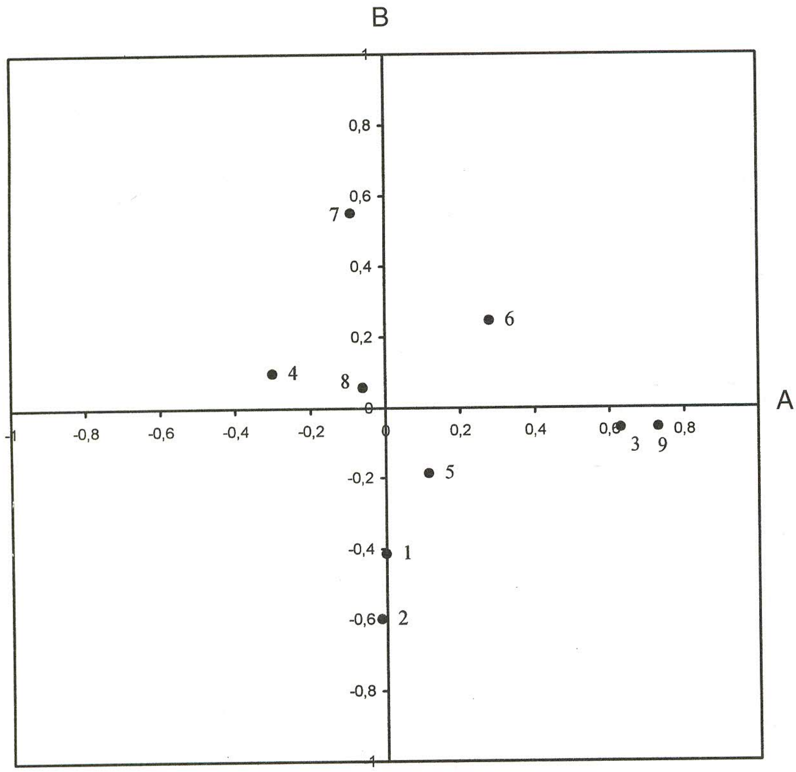

FIG. 1. Varimax solution–Matrix N by columns (Factors A and B)

The distribution of factor loadings indicates in all cases the existence of a latent structure that, as expected, varies in terms of the random matrix under consideration.

In terms of the results just described it was thought appropriate to explore the sequential arrangement of the values given in the various matrices. By sequential arrangement we mean the orderly distribution of values as shown by the relative frequency with which each number is followed by any other number in Matrices S_1 to S_5. In our case with the numbers 0 to 9 the number of possible sequences of two numbers is 100.

It was assumed that the differences between the frequencies in the entries would not be significant, hypothesizing that in matrices of random numbers all possible sequences will occur with approximately the same frequency. The chi squared values for all the entries in Matrices S_1 to S_5 show significant values beyond the 1% level. These results suggest that sequential arrangements in tables of random numbers occur with frequencies that are significantly different in terms of an assumed homogeneity.

Discussion

As expected and in terms of frequency, dispersion, and central tendency, the data here examined show nonsignificant differences between the rows and between the columns of all the matrices here examined, with the exceptions corresponding to the means of the rows in Matrices M and N and columns of Matrix A_1. Additionally, the chi squared tests indicate that from a statistical point of view the assumption that all the random values occur with similar frequencies should be accepted.

A different picture emerges when considering the possible associations that may exist between random numbers. The existence of a good proportion of significant correlations between the column and row vectors of the examined matrices suggest a relationship between groups of random numbers. This adds weight to the hypothesis that in matrices of random numbers may exist an underlying latent structure. This possibility is reinforced by the results obtained by factor analyzing correlations and cross-products. Using a principal component solution, followed by a varimax and a promax rotation, clear structures appeared.

The possibility of the existence of latent structures between random values—which may imply some order—is further strengthened by the fact that, when examining sequences in tables of random numbers, some pairs appear with a greater frequency than that expected in terms of a hypothesis of homogeneity or independence.

One reasonable interpretation of the described results seems to be that randomness does not necessarily imply lack of order, as factor structures and significant sequences of values occur among different collections of assumed random numbers so that the same collection of numbers may give very different pictures depending on what one is looking for and on how the values are analyzed. In short, the fact that a collection of values results from a random selection does not mean that the results of such selection will necessarily lack of some kind of structure.

The ideas of random and chaos are related. The place of nonlinear systems to explain chaotic events that occur in so-called random systems is of crucial interest and the existence of relations between groups of random numbers seems to require a better definition of the properties that characterize what is usually called random.

Recently, to cite but one, Quinodoz (1997, p. 1) examined how changes in psychic structures may be explained in terms of deterministic chaos theory, so that the "disorder of a system may be merely apparent and that a chaotic order governs systems previously thought to be completely random" so that "determinate causes can give rise to impredictable events."

To our mind the definition of random as "at haphazard, without aim or purpose or principle, heedlessly," as given in The Concise Oxford Dictionary of Current English (1939), requires further exploration to demonstrate that in the so-called random events exist nonrandom, well structured components that can be identified by looking at random samples using different approaches. In our case measures of central tendency, frequency, and variance satisfied the idea of randomness but, as soon as we began to examine relationships between random numbers and sequences, the problem changed.

In short, nonrandomness may be found in random samples, which brings to mind Alexander Pope's statement, "A mighty maze but not without a plan."

REFERENCES

Arkin, H., & Colton, R. (1962) *Tables for statisticians.* New York: Barnes & Noble.

Ayton, P., Hunt, A., & Wright, G. (1991) Randomness and reality. *Journal of Behavioral Decision Making*, 4, 222-226.

Clarendon Press. (1938) *The concise Oxford dictionary of current English.* Oxford, UK: Clarendon Press.

Crutchfield, J. P., Farmer, J. D., Packard, N. H., & Shaw, R. S. (1986) Chaos. *Scientific American*, 255, 38-49.

Falk, R. (1991) Randomness—an ill-defined but much needed concept. *Journal of Behavioral Decision Making*, 4, 215-218.

Feller, W. (1961) *An introduction to probability theory and its applications.* New York: Wiley.

Ginsburg, N., & Karpiuk, P. (1994) Random generation: analysis of the responses. *Perceptual and Motor Skills*, 79, 1059-1067.

Goldberger, A. L., Rigney, D. R., & West, B. J. (1990) Chaos and fractals in human physiology. *Scientific American*, 262, 36-41.

Iverson, G. R., Longcor, W. H., Mosteller, F., Gilbert, S. P., & Youtz, C. (1971) Bias in runs and dice throwing and recording: a few million throws. *Psychometrika*, 36, 1-19.

Mandelbrot, B. (1996) Del azar benigno al azar salvaje. *Investigación y Ciencia*, Diciembre, 14-20.

Meredith, W. M. (1971) *Manual de tablas estadísticas con aplicación a las ciencias de la conducta.* México City: Trillas.

Quinodoz, J. N. (1997) Transitions in psychic structures in the light of deterministic chaos theory. *International Psycho-Analysis*, 78, 699-718.

Solé, R. V., Bascompte, J., Delgado, J., Luque, B., & Manrubia, S. C. (1996) Complejidad en la frontera del caos. *Investigación y Ciencia*, Mayo, 14-21.

Accepted February 24, 2000.

APPENDIX

The cells in Matrices M and N were identified by randomly chosen digits from the Arkin and Colton tables (1962). Once the cell had been selected one entry was recorded in Matrix M while in Matrix N the value to be recorded resulted from blindly picking up from a basket one out of 10 balls numbered from 0 to 9. Matrix L was prepared by joining Matrices M and N. Matrices M_1 and N_1 were obtained from Matrices M and N, respectively. For instance, if in Column 3 of Matrix M number 1 appeared five times, in Matrix M_1 5 was entered into the cell at the intersection of Row 3 and Column 1.

The pairs of values between 00 and 99 given in Manual de Tablas Estadísticas (Meredith, 1971) were recorded in Table A_1, while the 8,000 numbers appearing in the Arkin and Colton tables (1962) were incorporated in Table A_2. The T matrix resulted from picking up at random 1,000 entries in the Meredith tables, of which 270 were used to build Matrix T_1. From a basket containing 10 balls numbered from 0 to 9 the results of 1,500 extractions were recorded in Matrix B.

The (300 × 10) C_1 matrix, the (200 × 15) C_2 matrix, and the (100 × 30) C_3 matrix were built after obtaining 3,000 random numbers using the Microsoft Visual Basic® V© 1997.