# Multivariate analyses for the study of behavior: an integrated approach

Maurizio Casarrubea
*Università degli Studi di Palermo, Facoltà di Medicina e Chirurgia, Dipartimento di Medicina Sperimentale, Sezione di Fisiologia Umana, Palermo, Italy, m.casarrubea@unipa.it*

## Introduction

Simple quantitative evaluations of isolate behavioral elements (i.e. frequencies, durations, per cent distributions) are not representative of the whole behavioral structure [1]. As suggested in a landmark paper from Spruijt and Gispen [2], it is only through the evaluation of the inter-relations among behavioral elements that it is possible to explore behavior from very different points of view, greatly beyond what the human eye can intuitively interpret. In the present paper a brief outline of different multivariate techniques for behavioral analyses will be presented in the attempt to underline the feasibility of their integration.

## Approaching transition matrices

Different levels of multivariate analyses (MVA) can be described on the basis of computational requirements and complexity of the approach. The first step is the construction of an ethogram that is a formal list containing descriptions of behavioral elements. After that, using specific software coders such as The Observer (Noldus Information Technology), behavioral elements have to be coded from the collected video-files and transitions from an element to another one traced in a transition matrix (TM). In brief, a TM is a table representing shifts among the behavioral elements, according to the selected ethogram [3][4][5][6]. A first and relatively simple approach to the analysis of a TM can be represented by the so called stochastic analysis [4][5]. A stochastic analysis requires transition matrix to be transformed into a matrix containing relative frequencies of transition from a given behavioral element to the others. Matrices containing relative frequencies can be graphically expressed through pathway diagrams. Figure 1 illustrates an example of pathway diagram representing *probabilistic relations* among five different behavioral elements identified in a group of 42 rats observed in open field. Three probability ranges were selected: between 0.10 and 0.24 (thin dotted arrows), between 0.25 and 0.49 (medium arrows), and between 0.50 and 1 (large arrows).

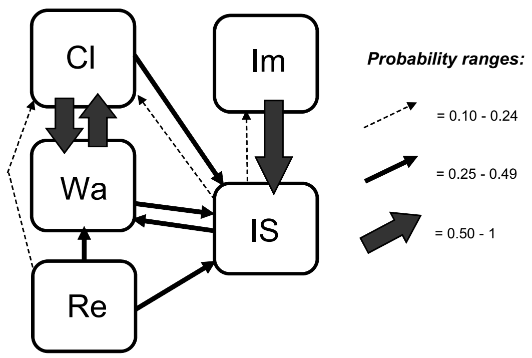


***Figure 1**. Transitional probabilities among 5 different behavioral elements observed in 42 rats in open field. Different probabilities of transition are depicted by arrows of different thickness. Cl = Climbing, Im = Immobility, Wa = Walking, IS = Immobile-Sniffing, Re = Rearing.*

A higher level of approaching matrices is represented by the cluster analysis which transforms transition matrices into similarity tables through an aggregative procedure. Cluster analysis allows the identification of behavioral clusters. The result of such procedure is a matrix containing only absolute affinity values, i.e. an half matrix where each cell indicates the "vicinity" between two given elements. Cluster analysis could be considered in some extent less intuitive than the stochastic one because of the underlying aggregative algorithm that converts transitions into similarity values (i.e. the direction of the behavioural flow is not expressed). Main outcome of cluster analysis is a dendrogram that is a tree diagram. Dendrogram in figure 2 represents *"vicinity" relations* among the same five behavioral elements presented in figure 1. However, even if both stochastic and tree diagrams present behavior in a graphically intuitive way, underlying matrices may need further statistical analyses.

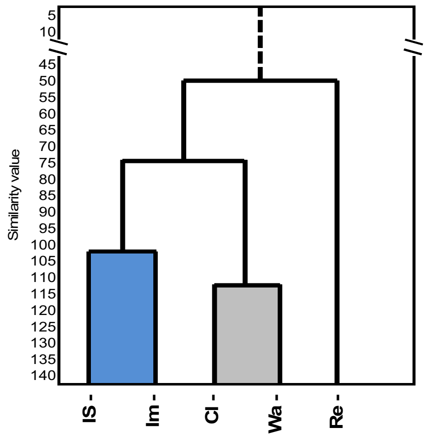


***Figure 2**. Dendrogram representing similarities among 5 different behavioral elements observed in 42 rats in open field. IS-Im and Cl-Wa are closely linked each other molding two different clusters. For abbreviations see figure 1.*

## Transitional matrices statistics: adjusted residuals

To compare transitional matrices, an useful method has been suggested by Tavaré et al. which proposed a correction factor – the so called gamma correction – to multiply by the chi-square value [7]. However, an elegant method to assess significance of cells within matrices is the one used by B.S. Everitt [8] and B.M. Spruijt [2] and, after them, by different Authors [5] [9]. Through this method a transition matrix is transformed into a matrix containing adjusted residuals. In summary, for each given transition, adjusted residual represents the difference between the observed cell value and an expected one, the latter calculated on the basis of a random distribution of transitions. Positive residuals indicate transitions occurring more often than expected, negative residuals represent transitions occurring less often than expected. Adjusted residuals can be evaluated through MatMan 1.1 (Noldus Information Technology), i.e. a specific software for matrix manipulation and analysis. The consistent advantage of adjusted residuals is that they can be expressed according to a Z-distribution so that P-values can be easily found in a common Z-table. Advantages provided by dendrograms and stochastic pathway diagrams is that they represent patterning among behavioral elements in a simple and intuitive way. However, both stochastic and cluster analyses are extremely sensitive to noise (i.e. disturbing and/or uncommon behavioral elements need to be removed before running the analysis), moreover they are close to “snapshots” of the whole observational period.

## Events along time: T-patterns

T-pattern detection doesn’t require an a priori “noise reduction” and, more important, it represents events *along* the session time. T-pattern analysis is carried out through the specific software Theme (Patternvision and Noldus Information Technology). This program performs a recursively test, checking the distribution of every combination of events *along* a specific time window [1].

Figure 3 represents an hypothetical T-pattern of four elements. If the time lag of an events sequence is not randomly distributed a simple T-pattern is detected (elements “e” and “d” in figure 3). In following steps, above simple t-pattern is processed again and if there is a temporal relation with other events, they are combined into higher order T-patterns (e-d-l-n, figure 3), and so on, repetitively, following a “bottom-up” process [1]. Since the graphical representation of T-patterns and the results of cluster analysis are both visualized through tree diagrams (figure 2 and figure 3), it is important to mention that cluster analysis is based on the similarities between events (figure 3). On the other hand, the tree structure provided by Theme does not represent such similarities, but the existence of significant relationships along time. Figure 4 shows an highly recurring T-pattern found in ten subjects randomly taken from the main group of animals represented in figure 1 and 2.

## Discussion

An even swift comparison among figures 1, 2 and 4 makes clear how these three, rather different, multivariate approaches strengthen each other in representing animal behavior. What’s more, each representation perfectly fits with the remaining ones: pathway diagrams and dendrograms show patternings among behavioral elements from stochastic and “aggregative” points of view respectively. On the other hand, T-patterns represent the behavioral structure along time. The results presented here show that these different MV approaches can be successfully used together for a better and more realistic description of behavioral patternings (i.e. to identify the presence of possible relationships between the elements of a behavioral sequence).

It is my contention that even hundreds of purely descriptive parameters make available only a partial and/or incomplete view of a given behavior. In other words, descriptive analyses reduce behavior to simple numbers. This is a reductionistic conception similar to the so called Cartesian Reductionism [10]. This approach does not work for behavioral analyses because a behavior is characterized by emergent phenomena arising from the inter-relations among events. Thus, quantitative evaluations such as durations, per cent distributions or frequencies of disjointed behavioral elements are not able to offer answers to *the* crucial question in all behavioral studies: what about the relationships between the observed elements? It is exactly here that descriptive approaches to behavioral study should give way to the multivariate analyses.


## Acknowledgements

The author would like to thank Prof. G. Crescimanno, Prof. F. Sorbera (Università degli Studi di Palermo, Italy) and Prof. F. A. Lopes (Universidade Federal do Rio Grande do Norte, Brazil) for their kind and constructive suggestions provided during the preparation of the present paper.


### References


1. Magnusson, M.S. (2004). Repeated Patterns in Behavior and Other Biological Phenomena. Eds: Oller, D.K, Griebel, U. *Evolution of Communication Systems: A Comparative Approach*. Cambridge: MIT Press 111-128.
2. Spruijt, B.M., Gispen, W.H. (1984). Behavioral sequences as an easily quantifiable parameter in experimental studies. *Physiology & Behavior* **32**, 707-710.
3. Hemerik, L, Bukovinszky, T., Gols, R., van Lenteren, JC., Vet, L.E.M. (2006). Enter the matrix: How to analyze the structure of behavior. *Behavioral Research Methods* **38**, 357-363
4. Casarrubea, M., Sorbera, F., Crescimanno, G. (2006). Effects of 7-OH-DPAT and U 99194 on the behavioral response to hot plate test, in rats. *Physiology & Behavior* **89**, 552-562
5. Casarrubea, M., Sorbera, F., Crescimanno, G. (2007). Multivariate analysis of the modifications induced by an environmental acoustic cue on rat exploratory behavior, *Physiology & Behavior* doi: 10.1016/*j.physbeh*.2007.11.010
6. Espejo, E.F., Mir D. (1993) Structure of the rat's behaviour in the hot plate test. *Behavioral Brain Research* **56**, 171-176.
7. Tavaré, S., Altham, P.M.E. (1983). Serial dependence of observations leading to contingency tables, with corrections to chi-square statisticss. *Biometrika* **70**, 139-144
8. Everitt, B.S. (1977). *The analysis of contingency tables*. New York: John Wiley and Sons
9. van Lier, H., Coenen, A.M., Drinkenburg, W.H. (2003). Behavioral transitions modulate hippocampal electroencephalogram correlates of open field behavior in the rat: support for a sensorimotor function of hippocampal rhythmical synchronous activity. *The Journal of Neuroscience* **23**, 2459-2465
10. Mikulecky, D.C. (2001) The emergence of complexity: science coming of age or science growing old? Computers and Chemistry **25**, 341-348.


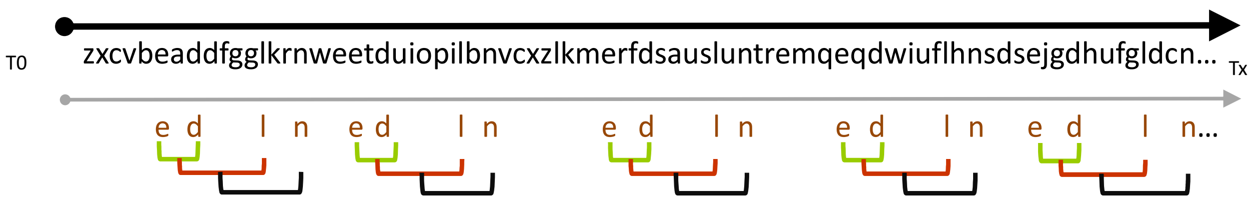


***Figure 3***. *Black upper arrow: observational period ($T_0 – T_x$) consisting of 26 hypothetical behavioral events (letters). The T-pattern represented in the bottom grey arrow (e-d-l-n) becomes evident when all the others behavioral occurrences are left out from the whole observational time line (black upper arrow).*

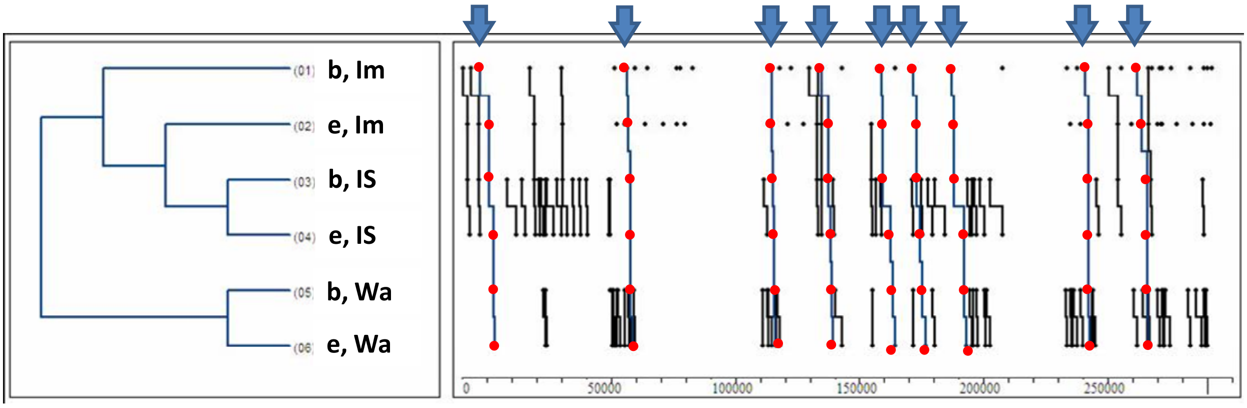


***Figure 4***. *T-Pattern encompassing Immobility, Immobile-Sniffing and Walking. T-pattern structure is shown in the left panel while the occurrences of the responses (dots) along x-axis (from frame 0 to frame 30000) are shown in the right panel. Larger dots = events encompassed in the T-pattern. b = beginning, e = ending. For other abbreviations see figure 1.*