The role of the membership function to model university students' flow

Il ruolo della funzione di appartenenza nello studio dei percorsi universitari degli studenti.

Franca Crippa, Marcella Mazzoleni and Mariangela Zenga

Abstract Students' pathways have long been the focus of educational research. Recent efforts have been directed to model levels of a progression embodied in a programme or in a theory, providing evidence about an individual level on that progression (Choi, 2012). In order to meet different schooling systems, the recognition of educational pathways as Markov processes (Trivellato, 1980, Ghellini et al., 2009) has unfolded into extensions such as fuzzy Markov chains (Symeonaki and Kalamatianou, 2011; Crippa, Mazzoleni and Zenga, 2013). The latter models, together with fuzzy set theory contributions to learning evaluation (Hameed and Sorensen, 2010), have enhanced the purpose of the membership function, whose role is investigated in this paper with respect to the issue of Italian university students' retention.
Abstract *I percorsi scolastici sono da tempo un tema cruciale della ricerca. Recentemente, la modellazione si e' rivolta ai livelli di programmi o teorie, sulla base dell'evidenza empirica (Chio, 2011). Nell'affrontare diversi sistemi di istruzione, il riconoscimento dei percorsi scolastici come processi Markoviani (Trivellato, 1980, Ghellini, 2009) si esteso grazie alle catene fuzzy di Markov (Symeonaki and Kalamatianou, 2011; Crippa, Mazzoleni and Zenga, 2013). Quest'ultimo avanzamento, insieme ai contributi della teoria degli insiemi fuzzy alla valutazione (Hameed and Sorensen, 2010), ha sollevato il tema della funzione di appartenenza, del cui ruolo nell'istruzione terziaria in Italia si discute in queste pagine.*

Key words: Membership Function, Fuzzy States, Markov Chain, University, Students' Flow

Franca Crippa
Department of Psychology, Piazza Ateneo Nuovo, 8, 20126 Milano (Italy)e-mail: franca.crippa@unimib.it

Marcella Mazzoleni and Mariangela Zenga
Department of Statistics and Quantitative Methods e-mail: m.mazzoleni8@campus.unimib; mariangela.zenga@unimib.it

1 Introduction

In their way through Higher education (HE), students navigate different trajectories that have long been the focus of educational research, in the perspective both of the institutional framework of the students' administrative flow and of learning achievements in specific disciplines. Statistical modeling of succeeding levels of these progressions, embodied either in an educational programme or in a learning theory, has developed from evidence in the educational field (Choi, 2012). In this respect, several questions, that are still open and at the centre of the educational debate, have recently been benefitted by the contribution provided by fuzzy set theory (Hameed and Sorensen, 2010).

Among the major issues in HE, a crucial role is played by students' retention, i.e. to students remaining within one higher education institution and completing their programme of study within a specific timeframe. The acknowledgement of educational pathways as Markov processes in the classical model (Trivellato, 1980, Ghellini, 2009) has undergone further advancements, due to the profound differences in HE systems. As a matter of fact, in several western countries HE students are entitled to register to the next curricular year only in case they have attained all the credits in the current yearly syllabus, whereas in some other countries, like Italy and Greece, this attainment is not mandatory. With respect to the latter system, a sound contribution has been given by the application of the theory of non homogeneous Markov systems with fuzzy states (Symeonaki and Kalamatianou, 2011; Crippa, Mazzoleni and Zenga, 2013), as a nominal academic registration to a specific curricular year does not identify a unique, homogeneous credits accumulation condition, instead it is likely to correspond to a wide spectrum of actual learning progressions. The levels of students' career progression, in this methodological approach, are related to the academic years using a stochastic model that assumes the formal levels of progress in learning, the registration year, as fuzzy states, related to a student's actual condition in terms of credits attainment, named crisp state, by means of a membership function.

Henceforth, within the frame of university students' retention, analytical modifications of the original formulation of the membership function are discussed and specified, in order to meet the specific characteristic of the Italian academic system, in particular of the local system under scrutiny. Our proposal of a membership function allows us to sketch some consideration on its key elements, in light of further insight.

2 Levels of progression as fuzzy states: the membership function

Learning achievement evaluation indicates how much a learner achieves by a set standard, as a composite and hierarchical process (Weon and Kim, 2001) that determines an order of individual students' performance levels in relation to specific educational learning objectives. It has previously been underlined how the Italian

HE systems host a wide spectrum of situations in terms of the amount and the composition of formative credits. In this scenario, the membership function allows to relate each fuzzy state to the formal requirements in the syllabus for a specific administrative year. Formally, the fuzzy space state is

$$F = \{F_1, F_2, ..F_N\} \tag{1}$$

where the number of elements N is lower than the one in the state space (S, k), where the state space (S) represents the set of values that the stochastic process can assume and with $k < N$. The relationship between the observable k space state elements and the fuzzy state space N elements is specified in terms of the function $\mu_{F_r}(j)$ of the element of the fuzzy set F_r, r=*1,,2,..,N*:

$$F_r() : S[0;1] \tag{2}$$

where the expression (2) is named membership function $\mu_{F_r}(j)$. The value of the membership function expresses to what extent an observable state belong to a given fuzzy state. If $\mu_{F_r}(j) = 0$ then the j does not belong to F_r, while if $\mu_{F_r}(j) = 1$ then j completely belongs to F_r. If $0 < \mu_{F_r}(j) < 1$ then j partially belongs to F_r and its membership function in F_r increases according to the value of $\mu_{F_r}(j)$. Besides, $F = \{F_1, F_2, ..F_N\}$ is assumed to define a fuzzy partition of S so that:

$$\sum_{r=1} \mu_{F_r}() = 1 \tag{3}$$

In HE systems like the Italian one, each actual combination of gained credits and passed exams can be linked to the correspondent fuzzy state, *i.e.* to the condition where all the requirements in the present and past syllabuses are satisfied. The record of l passed examinations of *sth* student are taken into account. In its original version (Symeonaky and Kalamatianou, 2011), the index md_i is computed for all students s, $i = 1, 2, .., l$, for each i passed exam, $i = 1, 2, .., l$, in order to evaluate its relative difficulty in comparison with the average performance in all the other examinations, according to the expression:

$$md_i = \frac{\sum_{s=1}^{n} [y_{si} - \frac{\sum_{k=1, k \neq i}^{l} y_{sk}}{l}]}{n} \tag{4}$$

being y_{si} the *s.th* student's grade in the *i.th* examination, l the number of exams that the students were able to pass in a given year of registration, n the number of students that successfully passed these *lth* exams in the same year. Therefore, the index in equation (4) relies on the individual difference between the specified mark and the average mark in all other passed exams in an academic year and it estimates the relative difficulty of a single examination. Then, from this index, the membership function $\mu_{F_r}(.)$ for the fuzzy sets F_r can be estimated:

$$\mu_{F_r} = \frac{1 + \sum md_i}{m_1} \tag{5}$$

where $\sum md_i$ refers to all exams not yet passed in a year, m_1 to the total number of examinations included in the syllabus for that year and it quantifies the relative compositional quota of each observable state with respect to the fuzzy state under scrutiny. Unfortunately, the adjustment of the index in (5) for the Italian situation as follows, for the the most part of the Italian courses this formulation leads to fairly unstable results (Crippa, Mazzoleni and Zenga, 2013). Therefore, as an innovative part of the methodology, we introduce a new definition of the membership function, related to the proportion of the successfully passed exams over the total for each academic year. Let X be a vector that contains the total formative credits contemplated for the t academic years, whose elements can be named 'levels of progress':

$$X = \begin{cases} X(1) \\ \dots \\ X(h) \\ \dots \\ X(t) \end{cases} \tag{6}$$

where $X(1)$ is the total value of the formative credits of the '1.st level of progress' , $X(2)$ is the total value of the formative credits of the '2.nd level of progress' and so on. Moreover, let X_s^h be the vector expressing the number of formative credits attained by the sth student at the end of her/his hth academic year for each 'level of progress':

$$X_s^h = \begin{cases} X(1)_s \\ \dots \\ X(h)_s \end{cases} \tag{7}$$

where $X(1)_s$ is the total value of the formative credits of the '1.st level of progress' gained progressively by the sth student during the hth academic years, $X(2)_s$ is the total value of the formative credits of the '2.nd level of progress' gained progressively by the sth student during the hth academic years Lastly, we have the relation with the non-crisp an absorbing states 'Dropout' and 'Graduation', the latter only for students enrolled at third or later academic year.

3 An example

With reference to an undergraduate course at the University of Milano-Bicocca and in line with the aforementioned model, we stratify students, at the end of each *h.th* year of administrative registration, into nonfuzzy states consisting of the amount of attained credits with respect to the credits required in the current and the previous yearly syllabuses. The situation of fully adherence to the curricular requirements within the academic year identifies the fuzzy states in (1), named 'level of progress' as designed in the HE programme. According to our proposal, the membership function (3) of the sth student can be expressed as:

$$\mu_h^s = \begin{cases} \mu_{F_1}^s = \frac{X(1)-X(1)_s}{X(1)+X(2)+...+X(h)} \\ \mu_{F_2}^s = \frac{X(2)-X(2)_s}{X(1)+X(2)+...+X(h)} \\ ... \\ \mu_{F_h}^s = \frac{X(h)-X(h)_s}{X(1)+X(2)+...+X(h)} \\ \mu_{F_{h+1}}^s = 1 - \sum_{i=1}^{h} \mu_{F_h}^s \end{cases} \tag{8}$$

The values in (8)are first estimated for all students for each registration year. A specific student we considered, for instance, at the end of his second administrative year, shows the values of the membership function to the first and to the second level of progress $\mu_{F_1}^s = 0.1250, \mu_{F_2}^s = 0.4472$, meaning that she/he is still due a tenth and slightly less than half of the credits of the first and the second year respectively. Therefore, she/he has accomplished approximately a third of the requirements for accessing the third level of progress, expressed by the value $\mu_{F_3}^s = 0.3778$. This estimation leads the probabilities of meeting syllabus requirements in relation to the academic year of registration (Symeonaki and Kalamatianou, 2011), henceforth illustrated for a subpopulation of 183 male undergraduate students at the University of Milano-Bicocca we in Table 1. Undergraduate male students show a pattern of

	Level of progression			Absorbing states	
Academic Year	1st	2nd	3rd	Graduation	Dropout
First	0.392	0.368	0.000	0.000	0.240
Second	0.294	0.323	0.257	0.000	0.126
Third	0.217	0.221	0.082	0.371	0.109
First out of programme	0.291	0.314	0.132	0.193	0.068
Second out of programme	0.111	0.182	0.099	0.245	0.364

Table 1 Probabilities of meeting syllabus requirements in relation to the academic year of registration, undergraduate course in Milano-Bicocca University, male population.

relative adherence to the curricular requirements, with some constraints in the range of choices within the pertaining academic year, giving place to a relevant dropout rate. Relative results in Table 1 highlight how the probability of failing to adhere to the first year syllabus regards 39.2 out of 100 enrolled students, at the end of their first curricular year (first raw in the table, label First). Instead, successfully acquiring all the due 60 credits has lower chances, as it involves 36.8% enrolled students, owed to a higher dropout probability (24%). At the end of second curricular year, the corresponding fuzzy state scores a higher probability, showing how a large amount of students failed to adhere to the first and second year syllabus altogether. Clearly, no graduations take place. In fact, at the end of the second curricular year, the correspondent fuzzy state scores a higher probability, showing how a large amount of students failed to adhere to the first and second year syllabus altogether: the probability of failing to adhere to the second year syllabus regards 61.7 out of 100 enrolled to the second year students (29.4% of the students are yet more related to the first year and 32.3% to the second year), while the probability of successfully

acquiring all the 120 credits is lower, as it involves 25.7% enrolled students and the dropout probability rate is equal to 12.6%. At the end of the third year, the probability of successfully acquiring all the 180 credits is higher than the last year and it involves 37.1% of the 'survived' students that can graduate, even if more than 44% failed to adhere to the first, second and third year syllabus altogether, and so on.

4 Conclusive remarks.

The introduction of fuzzy states in the theoretical frame of Markov chains can be welcomed as an advancement in the dynamic analysis of university paths, that adds relevant information with respect to other statistical techniques. Indeed, in this framework, a composite situation in terms of individual trajectories in tertiary education can be modeled, even in presence of a very scattered spectrum of university paths. To this respect, the membership function has proved to play a key role in the understanding of students' curricular choices, in line with a long-lasting discussion on its properties and characteristics in learning evaluation, on its aptitude to convey the organizational aspects of the educational system under scrutiny, such as the extent of free choice afforded to the students in their curricula, the grading system, the time span allowed to pass examinations successfully. Therefore, it opens new perspectives in the comprehension and improvement of a circular fine tuning between undergraduates strategies and the academic offer, leading to a more favourable HE outcome.

References

1. Choi, Y. *Dynamic Bayesian Inference Networks and Hidden Markov Models for Modeling Learning Progressions Over Multiple Time Points*, http://hdl.handle.net /1903/12739 (2012)
2. Crippa, F., Mazzoleni, M. and Zenga, Ma.*Never leave a job undone. Fuzzy Markov chains for Italian university students retention*, Proceedings of the SIS 2013 Statistical Conference - Advances in Latent Variables - Methods, Models and Applications http://meetings.sis-statistica.org/index.php/sis2013/ALV/paper/view/2628/435 (2013)
3. Ghellini, G., Maccari, F., Cordone, R.*Un approccio Markoviano per l'analisi lungitudinale dei percorsi universitari*, Working paper n.7 (2009)
4. Hameed, I. A. C., Sorensen, G. *Fuzzy Systems in Education: A More Reliable System for Student Evaluation*, Fuzzy Systems, Ahmad Taher Azar (Ed.), ISBN: 978-953-7619-92-3, InTech, Available from: http://www.intechopen.com/books/fuzzy-systems/fuzzy-systems-in-education-a-more-reliable-system-forstudent- evaluation (2010)
5. Symeonaki, M., Kalamatianou, A.*Markov system with fuzzy states for describing students' educational progress in Greek universities*, World Statistics Congress (ISI), Dublin, (2011)
6. Symeonaki, M., Stamoub, G.B.*Theory of Markov systems with fuzzy states* Fuzzy Sets and Systems, Volume 143, **3**, 427-445, (2004)
7. Trivellato, U. *Un modello markoviano del processo scolastico: specificazione e procedimento di stima*, Rivista di Statistica Applicata, **13**(1),3-20, (1980).
8. Weon, S., Kim, J. *Learning achievement evaluation strategy using fuzzy membership function*, Proceedings of the 31st ASEE/IEEE Frontiers in Education Conference, Reno: NV.(2001)