

Modelling Students' Length of Stay at University Using Coxian Phase-Type Distributions

Adele H. Marshall¹, Mariangela Zenga² & Sabrina Giordano³

¹ Centre for Statistical Science and Operational Research (CenSSOR), Queen's University Belfast, Belfast, Northern Ireland, UK

² Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milano, Italy

³ Department of Economics, Statistics and Finance, University of Calabria, Arcavacata di Rende, Cosenza, Italy

Correspondence: Mariangela Zenga, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via B. degli Arcimboldi, Milano 20126, Italy. Tel: 39-2-6448-3158. E-mail: mariangela.zenga@unimib.it

Received: December 17, 2012 Accepted: January 11, 2013 Online Published: January 22, 2013

doi:10.5539/ijsp.v2n1p73 URL: <http://dx.doi.org/10.5539/ijsp.v2n1p73>

Abstract

The time that Italian students spend at university is remarkably longer than in other European universities. For this reason, the government has recently introduced new rules for academic courses, in order to reduce the issue of long term students. In addition to this, universities need to address the growing problem of students prematurely leaving university before completing their courses. This paper considers the analysis of the length of stay of groups of students at Milano-Bicocca University using Coxian phase-type distributions classified according to the student individual characteristics.

Keywords: Markov chain, classification and regression tree, hazard function, survival function, students progression

1. Introduction and Motivations

Economists and sociologists regard university education as an investment where the costs of the education is balanced against the future benefits of having a better educated population and employable workforce. When a student leaves the university without completing their degree, it is at a cost to the student's family, the university as an institution and society. The costs to society are due to the economic output loss: the graduated is more productive than the non-graduated and society does not make a profit from the taxes of the missed graduate. Following this line of thought, the reform of Italian universities, through Law 509/1999, aims to prevent university drop outs and shorten the time taken to obtain a degree.

Moreover, graduation and drop out rates are adopted as criteria for evaluating the performance of universities. Therefore, the challenge for university managers is to make better informed policy decisions that can streamline the degree completion process, reduce the length of time it takes students to complete a degree and develop effective programmes to prevent drop outs.

The focus of this paper is on the time-to-event data where the time is the number of days elapsed from when a student first enrolls until he/she experiences the event, that is, he/she graduates or drops out from that university course.

The analysis of university education has been developing systematically for a long time and a wide literature exists on this subject. Event-History analysis (DesJardin et al., 1999, 2006; Ishitani, 2003; Kalamatianou & McClean, 2003) focuses on discrete events (the student drop out or graduation) occurring over time to establish risk factors. This technique is particularly interesting for analysing the departure process because the assessment of the transition from one state to another, that is, from enrolled to not enrolled, and the identification of the factors (e.g. personal, academic, socio-economic status of family) which influence the students' decision of leaving, are attainable (Triventi et al., 2009).

Another piece of work makes use of Markov chain models to analyze the progression of students at university. In these models, every student occupies a state at time t and transits from state to state at time $t + 1$ (the first and the last state represent enrolment and graduation/drop out respectively, while other states represent educational progress).

Gani (1963) originally used a Markov chain for estimating the probability of Australian students completing their degree course. Shah and Burke (1999) provided estimates for the mean time a student takes to complete the course, and mean time students spend in the higher education system in Australia. Harden and Tchong (1971) built the transition matrices from available historical data. Other examples are given in Song and Chissom (1994), Sah and Degtriarev (2005). Recently, Symeonaki and Kalamatianou (2011) proposed the theory of non-homogeneous Markov systems with fuzzy states for describing student educational progress in Greek universities.

The aim of this paper is to analyse student progression in the Italian reformed degree system and to estimate the influence of various factors on the probability that students, with certain characteristics, will progress successfully towards their degree or drop out. We propose to use the Coxian phase-type distribution for modelling the length of stay (in days) of the students enrolled at University of Milano-Bicocca. Student status has been observed for six academic years, during which time the student can graduate (as of the third year), drop out, change course or university, or can still be enrolled at the end of the observation (considered right-censored in the analysis).

There has been a wealth of literature devoted to investigating the determinants of the propensity of students to drop out of the academic career or to complete their degree programme. Personal characteristics such as gender and age; or individual abilities; income, education and socio-economic status of the family, academic-specific factors (services, quality in teaching, etc), and time-varying variables such as number of passed exams or credits are found to affect student choices at university (Arulampalam et al., 2004; Arulampalam et al., 2005; Checchi & Flabbi, 2006; Johnes, 1990; Light & Strayer, 2000; Robst et al., 1998; Smith & Naylor, 2001; among others).

A recent technical report by a consortium of Italian universities (Almalaurea, 2012) showed that one out of two students, enrolled at university, makes a wrong decision about their own education thus resulting in a high rate of drop out and a low level of satisfaction for graduated students. Motivated by this reason, we consider characteristics of students, known on enrolment, such as individual information (age, gender) and pre-college qualification (high school, mark). In this paper, we investigate the potential of using the Coxian phase-type distributions to give insights to the risk of drop out and graduation, and on the probability to “survive” at university for groups of students sharing common characteristics. It is hoped that by making students aware that, on the basis of their education to date, their background and/or their personal characteristics, they are more likely to have a particular outcome; either to drop out, complete their degree in time, or take up to six years or more to graduate, and that this will help steer students towards an appropriate course choice which meets their expectations.

A classification tree is introduced to identify the different student profiles and, for each profile, we model student’s time at university (length of stay, LoS) using different Coxian phase-type distributions. A new student upon identifying to which of the considered groups they belong, can gain valuable perspective on his/her probability of finishing the studies or dropping out, and on how long it should take him/her to complete or give up. Within the fitted Coxian phase-type distributions, each phase could represent a specific stage in academic career or behaviour. These issues are also of interest in assessing efficiency at the system and institutional level.

This enhances the use of the Coxian models for offering university leaders possible insights into the actual needs of change in management and lead universities to develop effective retention programmes and initiatives aimed at reducing drop outs and the time taken to complete the degree.

This paper is organized as follows: Section 2 introduces the Coxian phase-type distribution, Section 3 and 4 report the analysis of length of stay at Milano-Bicocca University using Coxian phase-type distributions for groups of students classified according to their individual characteristics. Section 5 concludes the work and reports on possible future developments.

2. Coxian Phase-Type Distribution

Coxian phase-type distributions (Neuts, 1989) are used to describe the time to absorption of a finite Markov chain in continuous time, where there is a single absorbing state ($n + 1$) and n ordered transient states or phases. The process starts in the first phase, then moves through sequential phases with the choice of entering the absorbing state at any time. For example, the student career at university can be thought of as a series of transitions through latent phases until an event of leaving university occurs due to graduation, drop out or transfer. Absorption from the first phases would represent the drop out of academic programmes, while absorption from the latest phases would indicate the conclusion for those students who complete a degree.

Let $\{X(t); t \geq 0\}$ be a (latent) Markov chain in continuous time with states $\{1, 2, \dots, n, n + 1\}$, where $\{1, 2, \dots, n\}$ are latent (transient) states of the process and state $(n + 1)$ is the (absorbing) state, and $X(0) = 1$.

For $i = 1, 2, \dots, n - 1$ the probability that a unit moves from one phase to the next one in the time interval δt is

$$\text{Prob}\{X(t + \delta t) = i + 1 | X(t) = i\} = \lambda_i \delta t + o(\delta t) \tag{1}$$

and for $i = 1, 2, \dots, n$ the probability that a unit leaves the system by entering the absorbing state is

$$\text{Prob}\{X(t + \delta t) = n + 1 | X(t) = i\} = \mu_i \delta t + o(\delta t) \tag{2}$$

The parameters of the Coxian phase-type distribution, λ_i and μ_i , describe the transition rates through the ordered transient states (from state i to state $i + 1$) and the transition rates from the transient states to the absorbing state (from state i to the absorbing state $n + 1$), respectively, see Figure 1.

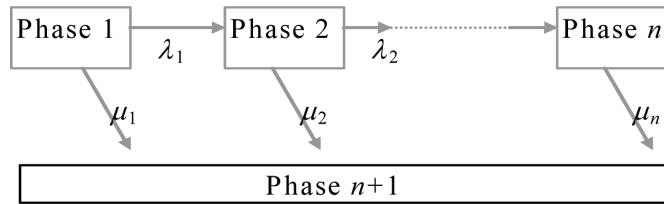


Figure 1. An illustration of the Coxian phase-type distribution

The density and survival functions of the variable T , the time until absorption, are given by:

$$f(t) = \mathbf{p} \exp\{\mathbf{Q}t\} \mathbf{q} \text{ and } S(t) = \mathbf{p} \exp\{\mathbf{Q}t\} \mathbf{1}$$

and the hazard function is $h(t) = f(t)/S(t)$, where $p = (1, 0, 0, \dots, 0)$ is the $1 \times n$ vector of probabilities defining the initial transient phases, $q = -Q\mathbf{1} = (\mu_1, \mu_2, \dots, \mu_n)'$ is the $n \times 1$ vector of transition rates from transient phases to the absorbing phase, and Q is the matrix of transition rates restricted to the transient phases

$$Q = \begin{pmatrix} -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \dots & 0 & 0 \\ 0 & -(\lambda_2 + \mu_2) & \lambda_2 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & 0 & \dots & -(\lambda_{n-1} + \mu_{n-1}) & \lambda_{n-1} \\ 0 & 0 & 0 & \dots & 0 & -\mu_n \end{pmatrix} \tag{3}$$

The probability that the individual leaves the system at phase i , say π_i is determined as a function of the estimated parameters μ_i and λ_i for $i = 1, \dots, n$, as follows:

$$\begin{aligned} \pi_1 &= \int_0^\infty \mu_1 e^{-(\lambda_1 + \mu_1)t} dt = \frac{\mu_1}{\lambda_1 + \mu_1}; \\ \pi_2 &= \int_0^\infty \mu_2 e^{-(\lambda_2 + \mu_2)t} dt \int_0^\infty \lambda_1 e^{-(\lambda_1 + \mu_1)t} dt = \left(\frac{\lambda_1}{\lambda_1 + \mu_1}\right) \left(\frac{\mu_2}{\lambda_2 + \mu_2}\right); \\ &\vdots \\ \pi_i &= \prod_{l=1}^{i-1} \left(\frac{\lambda_l}{\lambda_l + \mu_l}\right) \left(\frac{\mu_i}{\lambda_i + \mu_i}\right); \\ &\vdots \\ \pi_n &= 1 - \sum_{l=1}^{n-1} \pi_l. \end{aligned}$$

It is a usual procedure to aggregate phases sharing common characteristics to form stages, the interpretation of the stages is, often, more intuitive and meaningful.

Time (length of stay) may then be divided into intervals. In general, the k^{th} length of stay interval (at the k^{th} stage for example) can be determined by $S_k = \{t^{(j)} : m \sum_{i=1}^{k-1} \pi_i < j < m \sum_{i=1}^k \pi_i\}$ where $t^{(1)}, t^{(2)}, \dots, t^{(m)}$ represent the

ordered lengths of stay data for each individual and m represents the number of observations (Marshall & McClean, 2003).

Parameters μ_i and λ_i , $i = 1, \dots, n$, are estimated by fitting Coxian phase-type distributions via the EM algorithm (Asmussen et al., 1996) appropriately modified to take censored data into account. Likelihood ratio tests are performed to determine the most suitable number of phases. The likelihood function with censored observations is:

$$L = \prod_{j=1}^m f(t_j)^{\alpha_j} S(t_j)^{1-\alpha_j}$$

where α_j is an indicator variable which equals 1, if t_j is a complete time for the j^{th} unit and $\alpha_j = 0$, if t_j is a censored for the j^{th} unit (that is, the event does not occur before the end of the observational period).

Previous research has successfully used Coxian phase-type distributions to represent survival times as the length of time until a certain event occurs, where the phases are considered to be stages in the survival and the absorbing, final stage, the event that occurs causing the individual or element to leave the system completely. For instance, this event could be a patient recovering from an illness, a patient having a relapse, an individual leaving a certain type of employment, a piece of equipment failing, or a patient dying. Faddy (1994) illustrates how useful the Coxian phase-type distributions are in representing survival times for various applications such as the length of treatment spell of control patients in a suicide study, the time prisoners spend on remand and the lifetime of rats used as controls in a study of ageing.

In particular, Faddy and McClean (1999) used the Coxian phase-type distribution to find a suitable distribution for modelling the duration of stay of a group of male geriatric patients in hospital. They found that the phase-type distributions were ideal for measuring the lengths of stay of patients in hospital and showed how it was also possible to consider other variables that may influence the duration. More recently, Marshall and McClean (2003) have demonstrated how the Coxian phase-type distribution can, unlike alternative approaches, adequately model the survival of various groups of elderly patients in hospital uniquely capturing the typical skewed nature of such survival data in the form of a conditional phase-type model (C-Ph) which incorporates a Bayesian network of inter-related variables.

3. Length of Stay at Milano-Bicocca University

3.1 Data

The empirical analysis is conducted on administrative data of students at the Milano-Bicocca University (hereafter MIB). This university was established on June 10, 1998, to serve students from the northern Italy, to relieve some of the pressure on the over-crowded university of Milan, but first of all to offer the opportunity to take a degree at a much more affordable public university than the two renowned private universities of Milan.

The analyzed data refers to just over twenty thousand (20,069) students enrolled in the academic years 2000/01, 2001/02, 2002/3, 2003/04 at MIB in one of the 8 three year degree programmes: Economics (Ec), Educational Science (ED), Law, Mathematics-Physics-Natural Sciences (MPN), Medicine (Med), Psychology (Psy), Sociology (Soc) and Statistics (Stat). Conditions required for admission to the programmes at Milano-Bicocca university vary according to the Faculty. Students who apply for Psychology, Educational Science, Sociology, or to some of the Mathematics-Physics-Natural Sciences degree programmes are selected through an entry test while students of Economics only need to pass a mathematical and Italian language test. In addition, the undergraduate technical courses in Medicine such as Biomedical Laboratory Techniques, Dental Hygiene, Midwifery, Nursing and so on, require students to pass an aptitude test to get enrolled.

For each surveyed student, the time-to-exit from university is measured as the number of days elapsed from when a student first enrolls until she/he graduates or drops out or changes institution. Drop out students never finish their degree and are academically dismissed if they provide formal renunciation at any time during the year, if they do not pay taxes or if they do not take exams for at least one year. Students who are still studying but transfer to another institution exit in the data analysis. We follow the performances of the students for six academic years; if a student is still enrolled at the end of the observation time, his/her length of stay at university is considered as right censored in the analysis.

The event of interest is whether the student leaves during their study, gets a degree, transfers to another university, or takes six or more years to finish. The life table of all the students during the six years under study is represented graphically in Figure 2. In the first 3 years, 6167 students (31%) drop out and leave their academic career, 832

(4%) change university and 4675 (23%) take the degree in the regular time. Surprisingly, out of those completing within three years, 34% did so at the beginning of the third year. It is interesting to observe that 4817 students complete the degree programme after the legal duration of the courses. They are known as *fuoricorso* students representing 50.7% of graduated students, which is consistent with the other Italian universities (Miur, 2011). At the end of the observation, 1256 students are still enrolled (6%) taking at least six years to complete their study.

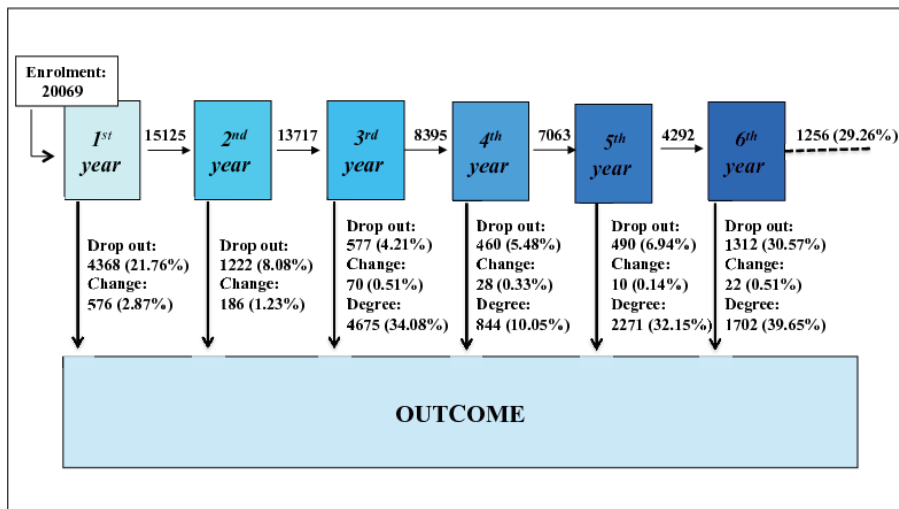


Figure 2. The progression of students during the six years of observations. The percentage is calculated with respect to students enrolled at the beginning of each academic year

As well as the length of stay, various attributes of each student are collected at the time of enrolment: gender, residence (Milan, out of Milan), high school (liceo, technical-vocational-training schools), high school mark (between 0.6 and 1), age at enrolment (at most 19, over 19 years old), enrolment (immediately after the high school, one or more years after the high school), cohort for enrolment (2000/01, 2001/02, 2002/03, 2003/04).

Table 1. Distribution of students by faculties and information at enrolment

		Faculty							
		Econ.	Law	Med.	Psych.	Educ. Science	Math. Physics, Nat. sc.	Stat.	Soc.
Gender	Female	46.17%	59.62%	78.28%	74.22%	82.88%	36.55%	50.71%	70.89%
	Male	53.83%	40.38%	21.72%	25.78%	17.12%	63.45%	49.29%	29.11%
Residence	Milan	66.95%	66.17%	31.54%	53.19%	54.46%	59.89%	58.40%	52.56%
	Out/Milan	33.05%	33.83%	68.46%	46.81%	45.54%	40.11%	41.60%	47.44%
High Sch.	Liceo	34.16%	37.14%	31.30%	57.98%	33.29%	54.04%	51.85%	41.44%
	Other	65.84%	62.86%	68.70%	42.02%	66.71%	45.96%	48.15%	58.56%
Mark	Average	0.75	0.74	0.74	0.81	0.75	0.77	0.79	0.79
	Stand. Dev.	0.12	0.11	0.11	0.12	0.11	0.13	0.12	0.12
Age	<= 19	57.26%	55.23%	49.88%	59.84%	49.49%	68.73%	70.09%	49.15%
	> 19	42.74%	44.77%	50.12%	40.16%	50.51%	31.27%	29.91%	50.85%
Enrolment	Immediat.	75.25%	74.88%	61.46%	70.38%	63.59%	85.14%	81.48%	63.62%
	Later	24.75%	25.12%	38.54%	29.62%	36.41%	14.86%	18.52%	36.38%
Cohort	2000/01	24.63%	20.17%	24.78%	25.03%	26.51%	26.89%	20.57%	26.43%
	2001/02	21.82%	24.26%	25.31%	23.67%	25.09%	22.72%	21.51%	25.23%
	2002/03	23.94%	26.05%	24.90%	25.50%	23.82%	23.11%	25.06%	22.21%
	2003/04	29.60%	29.53%	25.01%	25.80%	24.58%	27.28%	32.86%	26.14%
Enrolled Students		4953	1481	1243	2211	3241	4772	351	1817

Table 1 shows the distribution of students by personal information and Faculty. The composition of the surveyed students is very heterogeneous among the Faculties. In particular, the courses of Mathematics and Physics attract much more male students, while Medicine, Psychology, Sociology and Educational Science have typically a female setting. Most students of Statistics and Mathematics-Physics-Natural Sciences enrol immediately after the high school graduation. In general, more than half students live in Milan, but two out of three students enrolled on Medicine programmes come from other cities, probably because the offered courses are not commonly supplied in other Italian universities.

The preferences of students coming from different types of high school are in one case unexpected. Two-thirds of the students enrolled in Economics, Educational Sciences, Law and Medicine attended professional high schools while half the group of students who prefer to apply for the other programmes are qualified at liceo. It is unexpected that 63% and 68% of enrolled students into Law and Medicine courses respectively are graduated at a technical high school. More or less the same percentage of students enrolled in the four cohorts.

The final mark at the high school (Note: As the type of scale of mark at the high school has changed during the last 10 years, we considered a homogeneous version with minimum value 0.6 and maximum 1) seems not to be a discriminating factor. The average mark is quite similar for all the Faculties, except for the Psychology and Statistics programmes which are preferred by students who graduate with a medium-high level. The smallest average high school mark is for students on Medicine courses.

Table 2 shows the percentage of students of every Faculty by the cause of exit from MIB. Overall, almost half of students (47%) succeed to earn a degree, but a high percentage of enrolled students (42%) do not complete the degree programmes. At the end of the observation time (6 years), 6% of the students are still enrolled. As regards the Faculties, students attending courses of Law and Economics perform the worst. The rate of students who start studying Law (Economics) and never finish is dramatically high, 60% (53%). The Faculty of Law has also the highest percentage (10%) of students still “in progress” toward the degree after six years.

On the other hand, the percentage of students who graduated in Medicine programmes is substantially higher: nearly eighty percent of enrolled students complete their degree, while only 17% of students decide to drop out and less than 1% take over five years to finish. Note that these students have the smallest average mark at high school, but they do better in attaining a degree than their colleagues.

Students enrolled on Educational Sciences, Psychology, Sociology and Statistics courses perform similarly, with approximately 55% of these students completing their degree and nearly 35% dropping out.

Table 2. Distribution of students by Faculties and exit from MIB

		Faculty							
		Econ.	Law	Med.	Psych.	Educ. Science	Math. Physics, Nat. sc.	Stat.	Soc.
Exit	Drop out	52.80%	59.89%	16.49%	35.19%	35.24%	41.60%	35.33%	38.14%
	Degree	34.79%	26.00%	79.40%	55.99%	55.57%	45.66%	55.27%	54.21%
	Transfer	5.69%	3.78%	3.38%	3.98%	1.97%	6.29%	3.42%	2.64%
	Still enroled	6.72%	10.33%	0.72%	4.84%	7.22%	6.45%	5.98%	5.01%
Enroled Students		4953	1481	1243	2211	3241	4772	351	1817

3.2 A Classification Tree

A CART (Classification and Regression Tree) is a binary decision tree that is constructed by splitting a node into two child nodes repeatedly, beginning with the root node that contains the whole learning sample. Specifically, it is a non-parametric tree-structured recursive partitioning method, introduced by Breiman et al. (1984), to predict a response variable on the basis of certain predictors observed on a learning sample. The algorithm consists of two main stages: growing and pruning. In growing, the tree is recursively partitioned into subsets (nodes); each partition is obtained by examining all the possible binary splits along the observed data of each predictor variable and selecting the split that most reduces some measure of node impurity. The result is a sequence of nested trees, with increasing numbers of leaves (terminal nodes), until no more splits are possible and the fully grown tree is

reached. The pruning operation on the fully grown tree aims then to select the best subtree and consists by declaring an internal node as terminal and deleting all its descendants; this makes the tree more general and prevents any over-fitting on the training set. The aim of the classification tree is to predict the level of the response on the basis of the vector of the explanatory variables. In this paper the results for CART are obtained using the R package *rpart* (Therneau, 2012).

We created the classification tree reported in Figure 3 to classify the students according to their propensity for completing their degree based on individual characteristics collected at enrolment. Actually, the categorical response variable of interest is the event that determines the exit from MIB with four categories: degree, drop out, change, still enrolled. In the nodes of the tree, students with common attitude towards study who behave in a similar way are joined.

The Gini index is used to evaluate the node impurity and the misclassification rate at the final stage is 0.387. The mode of the four categories is shown on the final nodes in Figure 3. Faculty (1), age at enrolment (1), mark at high school (0.7), enrolment time (0.6) and high school type (0.2) give the most relevant contribution in the classification. The normalized measure of the importance of each predictor variable in relation to the final tree is reported in brackets.

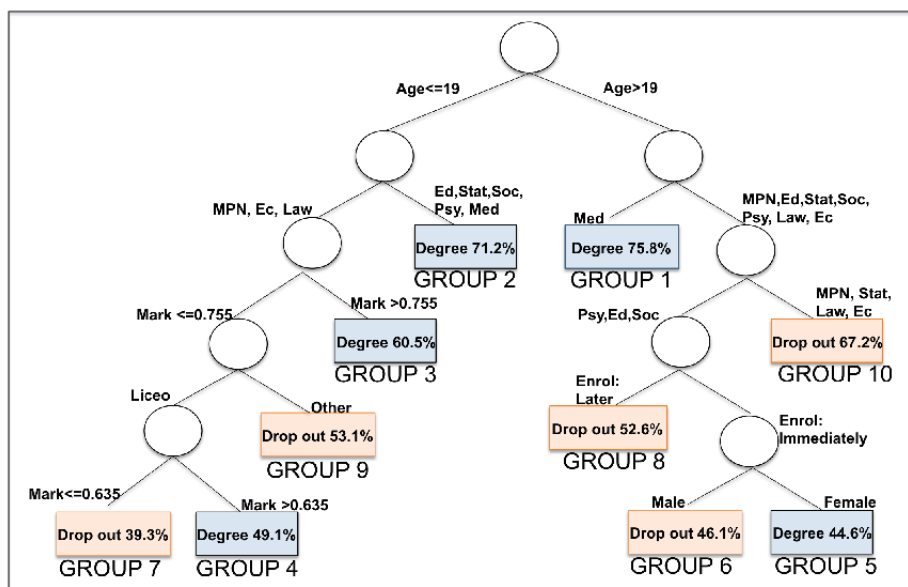


Figure 3. The classification tree for the MIB students by the risk of graduation or drop out

At the end of the pruning procedure, the terminal nodes of the tree identified 10 groups. Table 3 reports the description of the 10 groups ranked according to the likelihood of graduation, the graduation risk. For every group identified, the most frequent cause of exit is reported (column 6, Table 3). The last three columns of the table contain the mean, median and coefficient of variation of the length of stay (in days) at MIB of all the students belonging to each group.

The first group, for example, associated to a high risk of graduation, consists of over 19 years old students in Medicine. These students are probably the most motivated in studying: as they have been forced to pass an admission test and are mostly living outside of Milan, deciding to come to Milan probably only to attend courses. The study programmes give professional training where students interested in such courses make job-oriented choices without wasting time. The students tend to be older than their colleagues probably as most of them have failed the test previously. Their length of stay at MIB on average is 1031 days, so they are likely to complete their degree in a timely manner. At the end of the six years only 2 students are censored.

The last node (hereafter, the tenth), corresponding to the highest risk of drop out (67%) combined with the smallest percentage of graduated students (14%), and refers to the group of students who enrol in Economics, Mathematics-Physics-Natural Sciences, Statistics and Law at least one year later than graduation at school when they are older than 19 and drop out of the academic programmes during the first 2 years (average LoS is 775 days). They

resemble the category of students who enter university without any real motivation, probably under the pressure of their family who believe in the usefulness of the degree to getting a job, but who give up along the way.

The ninth group consists of students qualified at high schools different from liceo with at most a medium mark (≤ 0.775), enrolled immediately after school into Mathematics-Physics-Natural Sciences, Economics and Law. They could be those students who enter college with a weak background and do not perform at the level required to meet the Faculty standards and decide to leave, in fact half of them drop out.

Another group, the third, associated to a high risk of graduation (60%) consists of students enrolled immediately after school with the highest marks at Faculties Mathematics-Physics-Natural Sciences, Economics and Law. Their colleagues of the same Faculties and same age, but graduated at liceo with low-medium marks, compose the fourth group which rises above the others for the longest stay at MIB, 50% of them do not complete degree programmes within normal time and takes more than 3.64 years. Another node (the sixth), corresponding to a medium risk of drop out (46%) combined with a slightly lower percentage of graduated students (36%), refers to the group of male students over 19 years old, enrolled immediately after school into Educational Science, Sociology and Psychology Faculties. They could be those students who prefer to reconcile work with study, the programmes of these Faculties are in fact suitable also to part-time students.

Table 3. Description of the groups identified by the classification tree

Group	Description	N. students	Chance of Graduation	Drop out Risk	Outcome	Mean LoS	Median LoS	Variation Coefficient LoS
1	Age>19; Med	623	75.80%	19.40%	DEGREE	1031.53	1163.00	0.44
2	Age<=19; Fac: Ed, Stat, Soc, Psy, Med	4686	71.10%	17.40%	DEGREE	1230.65	1231.50	0.45
3	Age<=19 Fac: MPN, Ec, Law Mark >0.755	3457	60.50%	21.40%	DEGREE	1195.67	1189.00	0.51
4	Age<=19 0.635<Mark<=0.755 Fac: MPN, Ec, Law High School: Liceo	1223	49.10%	26.10%	DEGREE	1245.10	1358.00	0.56
5	Age>19 Fac: Ed, Soc, Psy Enrolment: Immediately Gender: Female	668	44.60%	32.30%	DEGREE	1229.69	1320.00	0.57
6	Age>19 Fac: Ed, Soc, Psy Enrolment: Immediately Gender: Male	360	36.40%	46.10%	DROP OUT	1068.93	1183.00	0.67
7	Age<=19 Mark<=0.635 Fac: MPN, Ec, Law High School: Liceo	496	29.20%	39.30%	DROP OUT	1177.41	1311.00	0.68
8	Age>19 Fac: Ed, Soc, Psy Enrolment: Later	2421	28.70%	52.60%	DROP OUT	1019.69	980.00	0.76
9	Age<=19 Mark<=0.755 Fac: MPN, Ec, Law High School: Prof/Tech	1758	23.90%	53.10%	DROP OUT	991.84	763.00	0.80
10	Age>19 Fac: MPN, Ec, Law, Stat	4377	14.70%	67.2%	DROP OUT	775.16	362.00	0.96

Table 4 reports the percentage of students in each group who enter university and go on to complete degree programmes within normal time (*in corso*) or take at least 4 or more years to finish or dropping out (*fuoricorso*).

Table 4. Rates of *in corso* and *fuoricorso* students by groups

		Groups										Total
		1	2	3	4	5	6	7	8	9	10	
Graduation	<i>In corso</i>	86.2%	74.5%	64.8%	42.8%	59.1%	51.1%	23.4%	53.3%	31.4%	36.8%	49.3%
	<i>Fuori corso</i>	13.8%	25.5%	35.2%	57.2%	40.9%	48.9%	76.6%	46.7%	68.6%	63.2%	50.7%
	Total	472	3333	2093	601	298	131	145	696	420	642	8831

Students belonging to the first group are considerably more likely to graduate within three regular years. Most of students in the second and third groups complete their degree in a timely manner. The percentage of students who graduate *in corso* is higher for group five than for group six even if the graduation rate is in the opposite order (see Table 3). The other groups can be characterized by high drop out rates and low graduation rates, having also a high rate of *fuoricorso* students.

3.3 The Survival and Hazard Functions for the 10 Groups of Students

To complete the explorative study of the length of stay at MIB, the empirical survival and hazard functions for the ten groups of students can be determined.

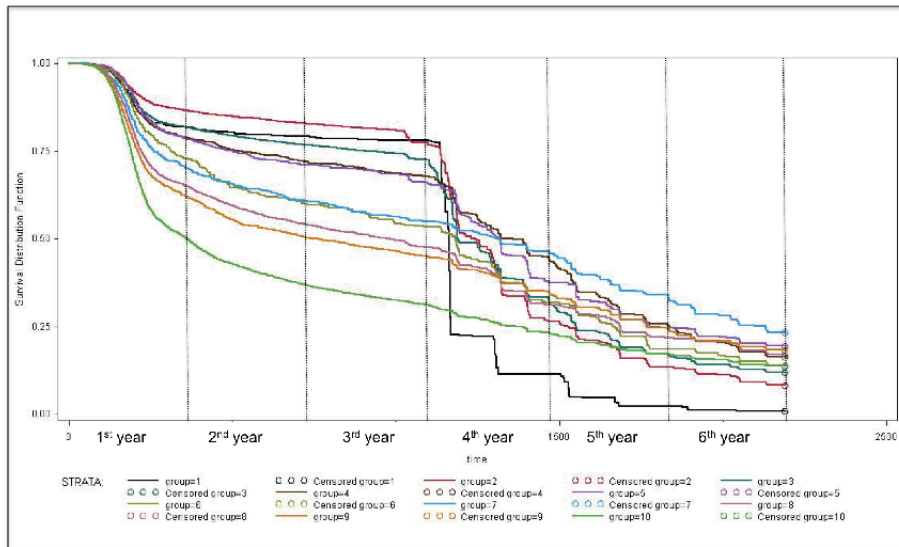


Figure 4. Empirical survival curves for the ten groups of students

Figure 4 displays the empirical survival curves for each group according to the product limit estimator. The shape of the curves seems quite similar: in the first academic year, a steep decline appears overall due to the high rate of drop outs, followed by a stationary trend in the next two years where a reduced number of drop outs and transfers usually registered. Starting from the end of the third year, there is a gradual decrease of the survival probability of students completing their degree.

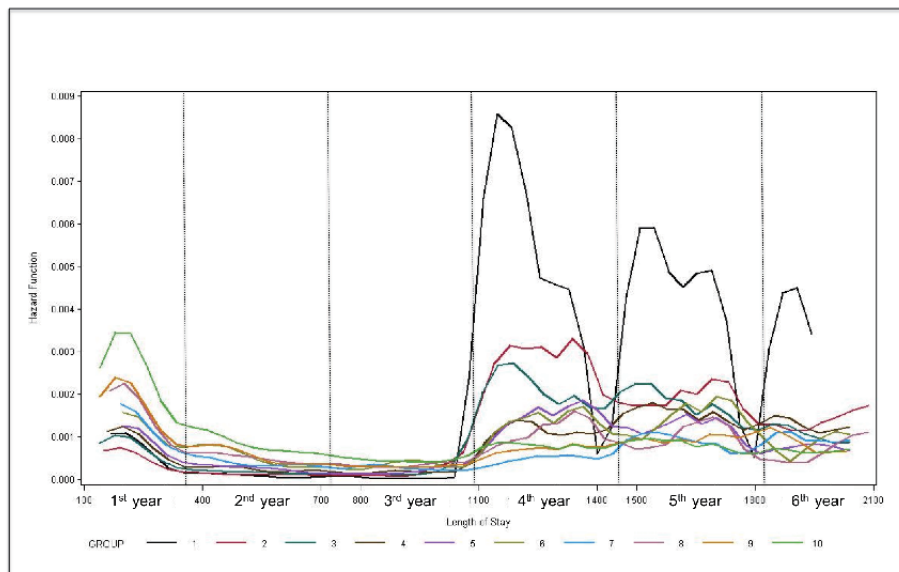


Figure 5. Empirical hazard functions for the ten groups of students

Although the general trend of the survival curves is common to all the groups, the plotted functions do not overlap and the log-rank test confirms that a significant difference exists among them (Chisq=538, df=9, p-value=0.0001).

In particular, the empirical survival curve of the first group stands out from all the other groups. It shows a sudden decline around the end of the third academic year, where most students take the degree. The curve of the second group, instead, dominates the other curves until the graduation time occurs, as the drop out/transfer rate registered for this group of students in the first three academic years is the smallest.

The performance of students decreases as the group number rises.

An overall inspection of Figure 5 shows that the higher the number of group the higher the risk of drop out at the beginning of the academic career, moreover, the first groups are characterized by a greater risk of completing the degree programme than the latest groups. This seems reasonable to expect given that groups one and two have the highest percentage of students completing within three years, therefore the risk of completing should be greater. Likewise, it is in the earlier period of study that you would expect students to be most indecisive of their course choice and most likely to drop out.

4. Fit of the Coxian Phase-Type Distribution

Tables 5a and 5b report the fit of the Coxian phase type distribution parameters using the EM-algorithm (Asmussen et al., 1996) adjusted for censored data. From inspection of the results, it is apparent that a 19 phase Coxian phase-type distribution is the most suitable for all the data (Note: We use a Chi-square test for nested model where the Loglikelihood for 19 phases was -136603.4351 and the Loglikelihood for 20 phases was -136602.9433, p-value=0.3884) together. However, it is important to note that some of the parameters μ_i associated with phase i in Table 5a are equal to zero. This would suggest that no one is observed leaving this phase which can therefore be aggregated with the neighbouring phase. Hereafter, the term stage will indicate a set of sequential phases with estimates of μ_i approaching to zero aggregated together with the closest phase associated to a strictly positive μ_i . In effect, the phases with small values of μ_i parameters are redundant and only the most dominant phases with the largest μ_i values are meaningful. This will also prevent an over-fitted model as reported in earlier literature (Marshall et al., 2012). In each stage we calculated the probability of leaving university due to degree, or drop out, or transfer. Actually, in each of the estimated phases where $\hat{\mu}_i = 0$ it is possible to leave the process of study, but the probability is so small that it is more likely for students to stay in the process than leave.

Table 5a. Results of fitting Coxian phase-type distribution

Phase i	Stage	$\hat{\lambda}_i$	$\hat{\mu}_i$	π_i
1	Explorative	0.0179	0	0.0003
2		0.0177	0	0.0000
3		0.0177	0	0.0000
4		0.0121	0.0056	0.3183
5		0.0143	0	0.0000
6	Intermediate	0.0143	0	0.0000
7		0.0143	0	0.0000
8		0.0143	0	0.0000
9		0.0132	0.0010	0.0488
10	Outcome	0.0116	0	0.0000
11		0.0116	0	0.0000
12		0.0116	0	0.0000
13		0.0116	0	0.0000
14		0.0116	0	0.0000
15		0.0116	0	0.0000
16		0.0116	0	0.0000
17		0.0116	0	0.0000
18		0.0030	0.0086	0.4682
19	Tardive	-	0.0003	0.1644

Table 5b. Average and interval of length of stay at each stage

<i>Stage k</i>	<i>Number of students (leaving at the stage)</i>	<i>Lower bound LoS</i>	<i>Upper bound LoS</i>	<i>Average of LoS</i>
1	6394	0	464	217.53
2	980	465	732	584.98
3	9366	733	1916	1326.26
4	3299	1917	2190	2161.23

The length of stay of all the surveyed students at MIB is analysed and found to be most suitably represented by a 4 phase Coxian distribution. The estimated parameters indicate four positive absorption probabilities (Note: Actually the first value of π_i is positive but strongly close to zero and it is ignored). So, the career of students seems to go through four sequential stages which we name explorative, intermediate, outcome and tardive. The four stages appear to represent student behaviour appropriately. At the beginning of a university course (the first explorative stage), we imagine an explorative stage in which the students face a new form of study; some of which realize that they cannot perform at the level required to meet the faculty standards and are discouraged to continue. So, the impact of the new environment results in a peak of students leaving (the drop out students). The second intermediate stage relates to students who have previously been unsure of their career path and after an initial attempt to go on, decide not to pursue their studies further where they either drop out or transfer, other students rest "in progress" towards a degree, proceeding step by step. The third stage, the outcome stage, comprises of the motivated students who complete their degree, in a timely manner however there will also be some students who are 'resting in progress'. The final stage (the fourth, tardive stage) are those students, with the longest length of stay at university (*fuoricorso* or censored data) taking six or more years to complete their degrees.

In particular,

- The explorative stage has length of stay between 17 and 464 days. The mean time to departure from MIB is 218 days, that is, less than three quarters of the first academic year. Among students leaving university at this stage, 88% gives up studying completely, while the remaining 12% decide to transfer to another university.
- The intermediate stage has length of stay interval between 465 and 732 days. Here we see the group of students who have been in doubt with a mean stage of 585 days (about one and half years) on whether they keep up the pace of study. Unfortunately, 92% of students who leave university during this stage are drop out students while 8% of them choose to transfer. Only 2 students take the degree in this stage (formally they graduated in the first degree session, on June).
- The outcome stage has length of stay between 733 and 1916 days. This comprises strongly motivated students who complete their degree. The average time it takes students to earn the degree in this phase is 1236 days corresponding to about three academic years. In particular, 87 out of 100 students will exit their academic career at this stage and graduate within the regular time, the other students drop out.
- The tardive stage has length of stay varying between 1917 and 2190 days. This final stage is regarded as the tail period where *fuoricorso* students graduate or remain still enrolled after six years (censored LoS). Eighty percent of students complete their degree while 20% percent remain enroled up to the end of the period. Students leaving at this stage belong to the group of those who wish to graduate but they succeed only after a very long stay at university. The average time to degree is equal to 2161 days, about six years, so twice the regular duration of a degree programme. Understanding the factors that are influential to such a long degree completion time is one of the most crucial issues for the university managers.

Table 5b reports the students leaving the university, the interval and the average LoS for each stage. So, for example, 6394 students leave during the explorative stage within 464 days and their career lasts on average 217.53 days.

In describing the stages, we focused on students who left the system, but of course there are all the motivated students who rest in progress toward a degree and proceed through these sequential stages increasing their abilities (and their human capital) by attending courses and passing exams.

Figure 6 displays how the estimated Coxian phase-type distribution fits the empirical distribution of the lengths-of-stay at MIB university. The fitted density seems to meet the characteristic shape with two peaks, the former due to the high drop out rate and the latter related to students graduating with a degree.

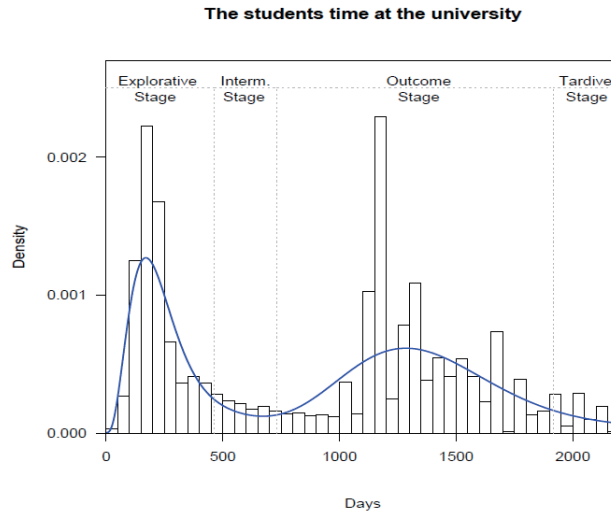


Figure 6. Empirical and estimated distribution of the length of stay at MIB university (on the complete sample of students)

At this point, we investigate the fit of a Coxian phase-type distribution to model the length of stay at MIB for each group of student. Table 7 reports the results of the fitting procedure. The estimated phases (first row) are then aggregated to form stages considering the positive values of μ_i . The likelihood statistic and p-values show that there is no significant improvement in fit by adding one more phase to the distribution. The final rows indicate the bounds (in days) of the intervals of the length of stay of students leaving at every stage of the distribution. So, for example, the LoS of students in the 10th group are divided into 3 intervals (delimiting the 3 stages): 0-412, 413-745, 746-2190 (days).

At first glance, the results make it clear that the number of stages is quite different within the different groups. Only the distribution for the second group has the 4 stages detected in the distribution fitted on the complete sample of students. Moreover, for students in groups 3 to 6, the intermediate stage seems not to be relevant.

A case which deserves attention is the first group, in such a case a remarkable high number of phases are registered and the algorithm for parameter estimation appears to have difficulty in reaching convergence. It is possible that a more suitable mixed model (continuous-discrete) for the student performances of this group can avoid the convergence difficulty. This problem needs further investigation.

A comparison between the empirical and estimated distributions for each group is shown in Figure 7. The Coxian phase-type distributions differ according to the different student groups but almost all the fitted densities seem to suitably represent the empirical trends, even if for some groups the fitting is substantially better than for other groups.

Table 7. Results of fitted Coxian distribution to the length of stay in the 10 groups of students

	Groups									
	1	2	3	4	5	6	7	8	9	10
# Phases	32	22	14	12	11	11	13	12	10	12
Log likelihood	-4338.0	-32799.4	-23787.2	-8225.2	-4356.1	-2461.9	-3085.9	-15865.1	-11295.5	-28277.6
Test Statistics	4.57	1.22	1.21	3.82	0.84	0.92	2.45	4.88	3.13	3.90
p-value	0.05	0.27	0.27	0.07	0.33	0.31	0.15	0.04	0.10	0.07
# $\mu_i > 0$ (n. of stages)	3*	4	2*	2	2	2	3	3	3*	3*
up. bound 1	382	464	585	663	568	606	403	460	417	412
up. bound 2	732	653	2190	2190	2190	2190	866	706	732	745
up. bound 3	2190	1889	-	-	-	-	2190	2190	2190	2190
up. bound 4	-	2190	-	-	-	-	-	-	-	-

Description: the* indicates that actually one more value of μ_i is positive but strongly close to zero and it is ignored. The upper bound (up. bound in the table) of the intervals of LoS at each stage is indicated, the first lower bound is zero.

Recall that the groups are ordered according to the chance of graduation, so the later groups involve students whose performance is poor and are more likely to drop out than their colleagues of the first groups with a marked propensity towards study. In Figure 7, the higher the number of group, the better the fit of the Coxian phase-type model, thus the use of the Coxian phase-type distribution seems to be more appropriate for modeling the MIB of students who perform worse and have very long lengths of stay. This agrees with previous research where Coxian phase-type distributions are used to represent survival or length of stay of elderly patients in hospital. There is more heterogeneity in the earlier stages of survival which is to be expected as there is a bigger case mix of individuals present in the first phases. In fact Faddy and McClean (1999) and Marshall et al. (2003) both highlight that the first phase includes elderly patients who either leave the system quite quickly due to having minor problems and thus return home or who have critical health problems and die within a short period of time in hospital. The approach is good at representing the very long stay patients who are consuming large amounts of hospital resources by staying in hospital for a long period of time. Likewise, the research presented in this paper is primarily concerned with those students who have very long stays at University and do not complete their course within six years of stay.

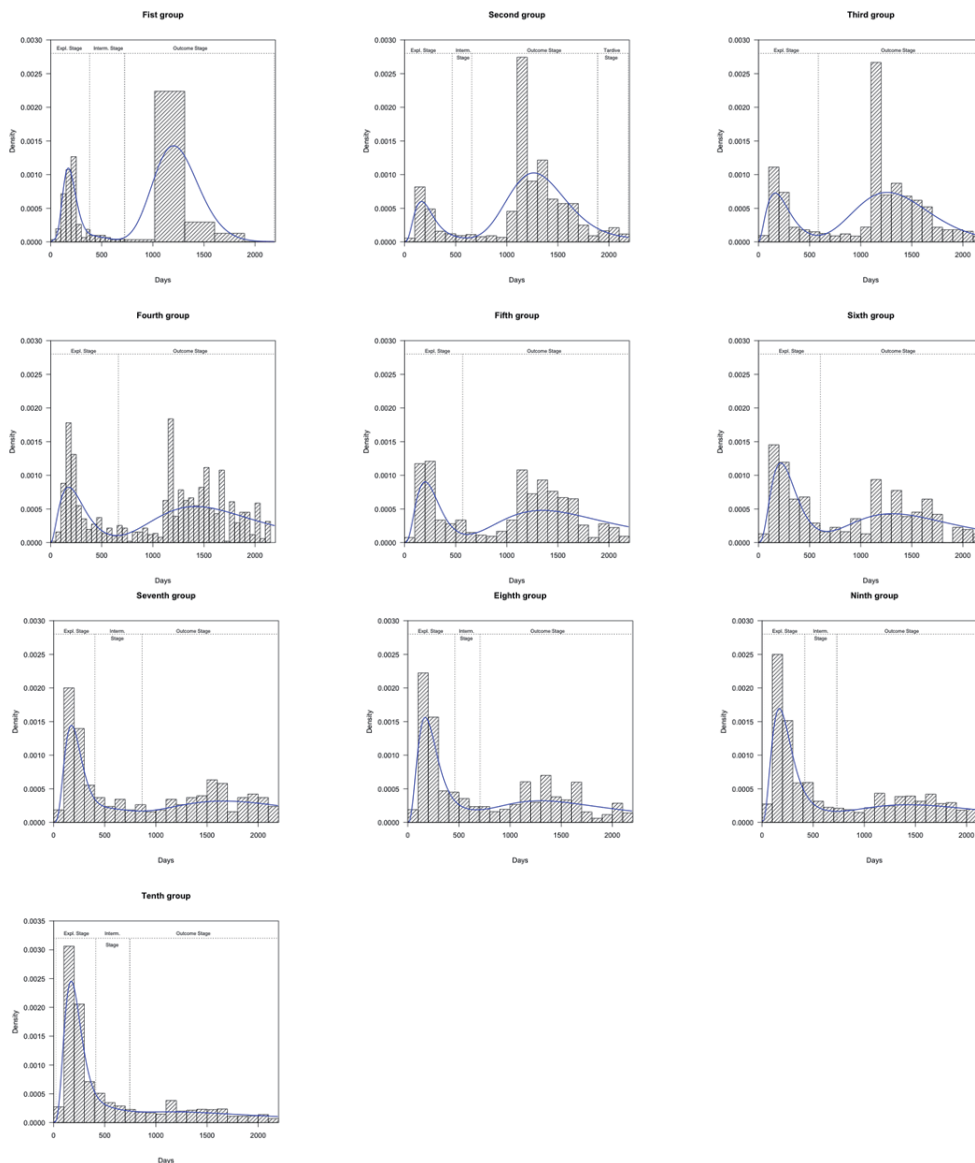


Figure 7. The empirical and estimated density distributions for the ten groups

As expected, at first sight, comparing the Coxian phase-type distributions in succession from the first to the latest ones, it appears clear that the second peak in the distributions, relates to lengths of stay of graduated students,

gradually tailing off as the number of group increases, while the peak in the initial stage caused by the dropping out students rises. Thus, the distributions are initially (for the first groups) bimodal and then tend to become highly skewed with only one large peak at the explorative stage (for the last groups).

In the first 4 plots, students are likely to enter the absorbing state either for dropping out and for graduation. Plots relative to all the groups from the 5th to the 10th instead show that most of the students reach the absorbing state in the first explorative stage, a high rate of enrolled students never finish their degree. This represents the very challenge of university leaders who need to ensure policies and practices to prevent this academic failure.

The percentage of students who graduated is substantially higher for the first three groups of students, the corresponding plots exhibit the largest peak of the density at the outcome stage. In the third group, for example, students with a good background at school who decide to enroll straight into the academic programmes of Mathematics-Physics-Natural Sciences, Economics and Law are more likely to complete their degree. The Coxian phase-type distribution captures this performance. On the other hand, the over 19 year old students in the eight group, enrolled in Education, Sociology and Psychology Faculties at least one year after the high school, do not overcome the initial difficulties and most give up in the first two years (approximately 460 days). The Coxian phase-type distribution is able to represent this empirical mode in the explorative stage (see Figure 9).

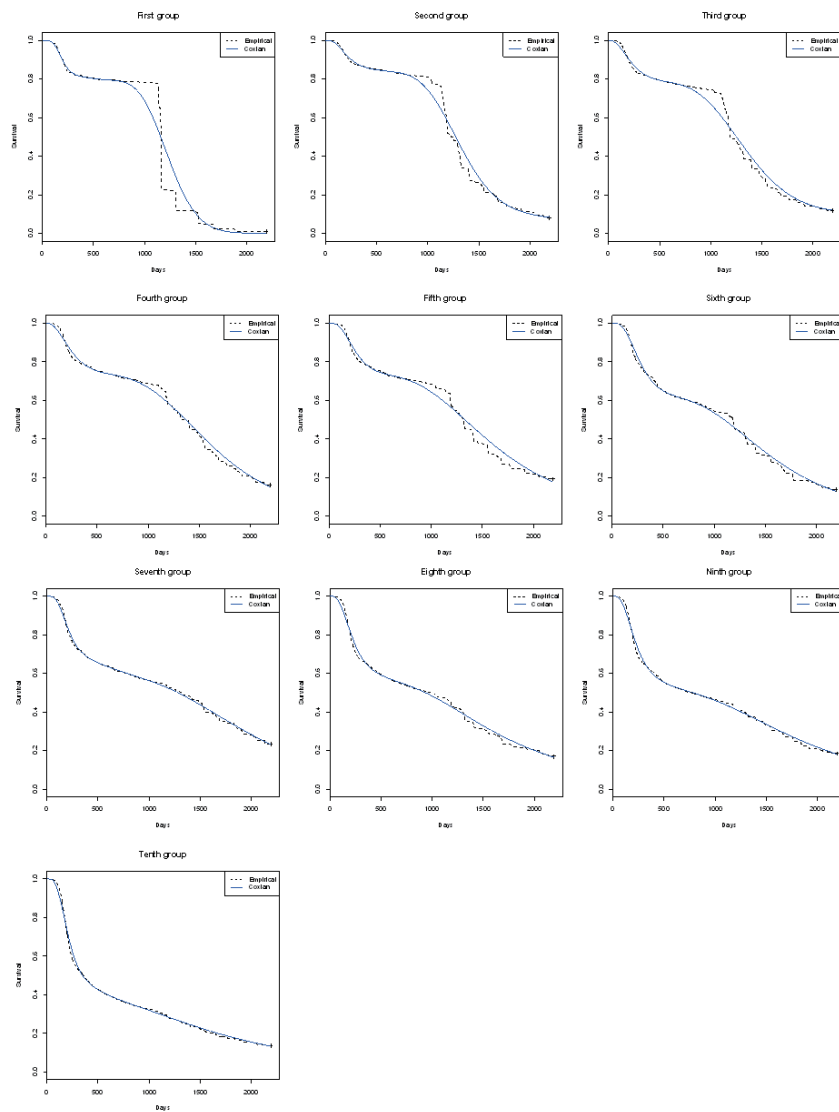


Figure 8. The empirical and estimated survivals for the ten groups

For the first group the fitted distribution does not capture the extremely high second peak of the empirical distri-

bution due to the fixed graduation dates where graduation falls on fixed days in the academic year. An alternative representation is to consider a mixture of Coxian phase-type distributions. However, doing so does not improve the fit any further than what is presented in this paper. This aspect will undergo further investigation and we restrict the focus of this paper on the extremely long stay students.

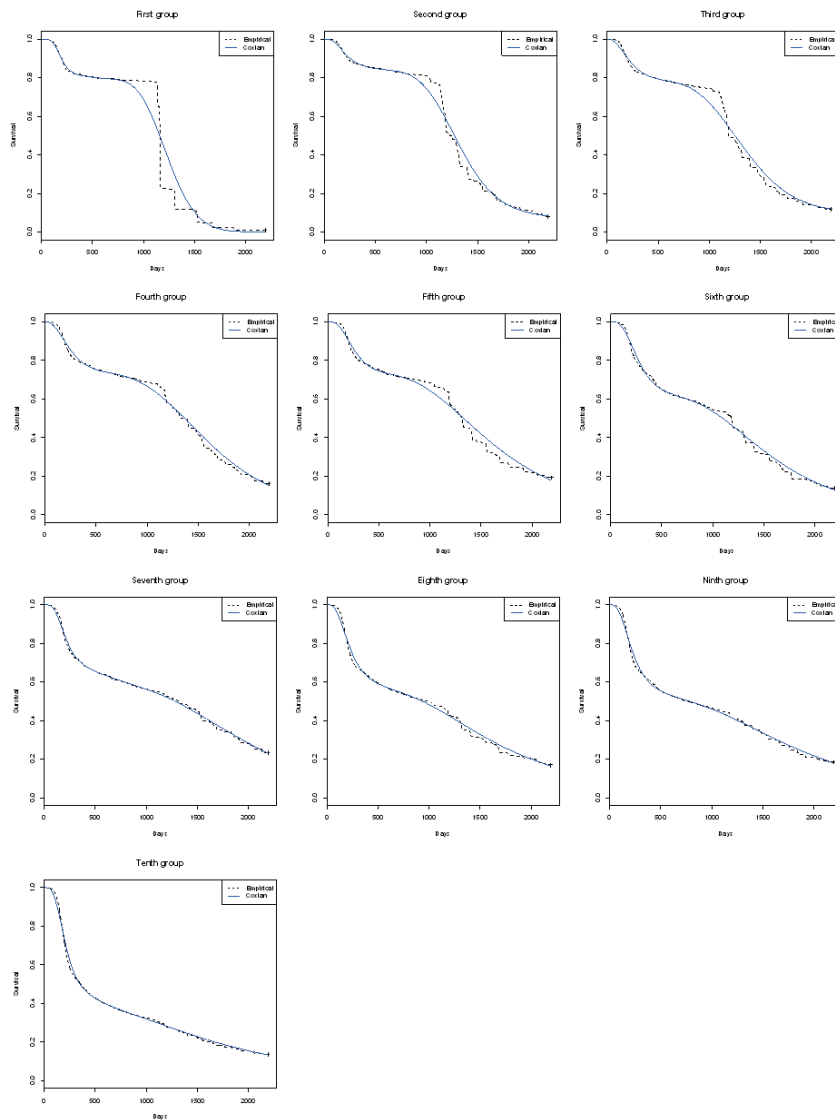


Figure 9. The empirical and estimated hazard for the ten groups

Fitting the Coxian phase-type models enables us to offer possible insights on estimating the risk of leaving university due to drop out or graduation reasons according to the group to which the student belongs. The Coxian phase-type distribution also provides the estimates of the probability to “survive” in the university system towards the degree and distinguish between different groups of students depending on the survival probability they have.

Survival and hazard functions are estimated using the Coxian phase-type densities of each group, and are compared with the empirical curves determined by the non-parametric Kaplan-Meier procedure in Figures 8 and 9, respectively. The estimated functions approximate the empirical curves well, where the fitted results for the densities represent the highest ranked groups most appropriately (Figure 7). In particular, the estimated survival functions overlap the empirical ones for all of the later groups (particularly for groups 7-10).

5. Conclusion

The work presented in this paper introduces an innovative application of the Coxian phase-type distribution to the University student progression and drop out phenomena. There are two models considered. The first, introduces a

classification tree to divide the students into different profiles of stay according to their characteristics known on enrolment of their course. This produced ten groups of student with differing characteristics across the groups and time at university represented using different Coxian phase-type distributions. A new student, upon identifying the group to which he/she belongs, can gain valuable perspective on his/her probability of finishing the studies or dropping out, and on how long it should take him/her to complete or give up. Within the fitted Coxian phase-type distributions, each phase represents a specific stage in academic career or behaviour. These issues are also of interest in assessing efficiency at the system and institutional level. This enhances the use of the Coxian models that would offer university leaders possible insights into the actual needs of change in management and lead universities to develop effective retention programmes and initiatives aimed at reducing drop outs and reducing the times taken to complete a degree. Upon developing a model for the ten different student groups, the Coxian phase-type distribution was fitted again separately for each group of student. This provides further refinement of the student length of stay by modeling each student group as a sequence of phases in a separate Coxian phase-type distribution. In doing so, an improvement in survival predictions can be made to the student stay.

This second model follows a similar format to that by Harper et al. (2012) and Marshall et al. (2012) who introduce the Discrete Conditional Phase-type distribution using a classification tree to model patient characteristics on admission to hospital as the first component in the model which is conditioned on the second component, the patient length of stay in hospital represented by a Coxian phase-type distribution. Such an approach is very applicable to student time at University and consistent with previous research. This paper extends that work to another application area and in doing so is able to use the fitted Coxian phase-type distribution to define four stages of student behaviour in University. Linked with these stages different student characteristics and associated likely result for that student in terms of graduating on time or dropping out. Student progression at an University is a concern for many countries particularly the costs incurred and the stress to the student. As further work, it is planned that the models presented in this paper will be applied to student data for other countries. One particular example that will be considered is the application of the model for University students in Greece. Another possible extension to this work is to incorporate the costs into the model.

References

- Almalaurea. (2012). XIV Indagine sulla Condizione occupazionale dei laureati. Retrieve from <http://www.almalaurea.it/universita/occupazione/occupazione10/>
- Arulampalam, W., Naylor, R. A., & Smith, J. P. (2004). A hazard model of the probability of medical school drop out in the UK. *Journal of Royal Statistical Society Series A*, 167, 157-178. <http://dx.doi.org/10.1046/j.0964-1998.2003.00717.x>
- Arulampalam, W., Naylor, R. A., & Smith, J. P. (2005). Effects of in-class variation and student rank on the probability of withdrawal: cross-section and time series analysis of UK universities students. *Economics of Education Review*, 24, 251-262. <http://dx.doi.org/10.1016/j.econedurev.2004.05.007>
- Asmussen, S., Nerman, O., & Olsson, M. (1996). Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23, 419-441.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Tree*. Chapman & Hall.
- Checchi, D., & Flabbi, L. (2006). Intergenerational Mobility and Schooling Decisions in Italy and Germany, *Mimeo*, University of Milan.
- Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes*. London: Chapman.
- DesJardinis, S. L., Ahlburg, D. A., & McCall, B. P. (1999). An event history model of student departure. *Economics of Education Review*, 18(3), 375-390. [http://dx.doi.org/10.1016/S0272-7757\(98\)00049-1](http://dx.doi.org/10.1016/S0272-7757(98)00049-1)
- DesJardinis, S. L., Ahlburg, D. A., & McCall, B. P. (2006). The effects of interrupted enrolment on graduation from college: Racial, income and ability differences. *Economics of Education Review*, 25(6), 575-590. <http://dx.doi.org/10.1016/j.econedurev.2005.06.002>
- Faddy, M. (1994). Examples of fitting structured phase-type distribution. *Applied Stochastic models and Data Analysis*, 10, 247-255. <http://dx.doi.org/10.1002/asm.3150100403>
- Faddy, M., & McClean, S. I. (1999). Analysing data on lengths of stay of hospital patients using phase-type distribution. *Applied Stochastic Models in Business and Industry*, 311-317.

- Gani, A. (1963). Formulae for projecting Enrollments and Degrees awarded in Universities. *Journal of the Royal Statistical Society*, 126(3), 400-409. <http://dx.doi.org/10.2307/2982224>
- Harper, P. R., Knight, V. A., & Marshall, A. H. (2012). Discrete Conditional Phase-type models utilising classification trees: Application to modelling health service capacities. *European Journal of Operational Research*, 291, 522-530. <http://dx.doi.org/10.1016/j.ejor.2011.10.035>
- Ishitani, T. T. (2003). A longitudinal approach to assessing attrition behavior among first-generation students: Time-Varying Effects of Pre-College Characteristics. *Research in Higher Education*, 44(4), 433-449. <http://dx.doi.org/10.1023/A:1024284932709>
- Johnes, G. (1990). Determinants of student wastage in higher education. *Studies in Higher Education*, 15, 87-99. <http://dx.doi.org/0.1080/03075079012331377611>
- Kalamatianou, A. G., & McClean, S. (2003). The Perpetual Student: Modeling Duration of Undergraduate Studies Based on Lifetime-Type Education Data. *Lifetime Data Analysis*, 9, 311-330. <http://dx.doi.org/10.1023/B:LIDA.0000012419.98989.d4>
- Light, A., & Strayer, W. (2000). Determinants of college completion: school quality or student ability? *Journal of Human Resources*, 35, 299-332. <http://dx.doi.org/10.2307/146327>
- Marshall, A. H., & McClean, S. I. (2003). Conditional phase-type distributions for modelling patient length of stay in hospital. *International Transactions in Operational Research*, 10, 565-576. <http://dx.doi.org/10.1111/1475-3995.00428>
- Marshall, A. H., Payne, K., & Cairns K. J. (2012). Modelling the development of late onset sepsis and length of stay using discrete conditional survival models with a classification tree component. *Proceedings of IEEE Computer Based Medical Systems 2012*. <http://dx.doi.org/10.1109/CBMS.2012.6266407>
- Marshall, A. H., & Zenga, M. (2012). Experimenting with the Coxian Phase-Type distribution to Uncover Suitable Fits. *Methodology and computing in Applied probability*, 14, 71-86. <http://dx.doi.org/10.1007/s11009-010-9174-y>
- Miur. (2011). Undicesimo Rapporto sullo Stato del Sistema Universitario. Retrieved from http://www.cnvsu.it/publidoc/datistat/default.asp?id_documento_padre=11777
- Neuts, M. (1989). Structured Stochastic Matrices of M/G/1 Type and Their Application (Marcel Dekker, NY).
- Robst, J., Keil, J., & Russo, D. (1998). The effect of gender composition of faculty on student retention. *Economics of Education Review*, 17, 429-438. [http://dx.doi.org/10.1016/S0272-7757\(97\)00049-6](http://dx.doi.org/10.1016/S0272-7757(97)00049-6)
- Sah, M., & Degtiarev, K. (2005). Forecasting enrollment model based on First-Order Fuzzy time series, World Academy of Science, Engineering and Technology I, 132-135.
- Schultz, T. W. (1963). *The economic value of education*. New York: Columbia University Press.
- Shah, C., & Burke, G. (1999). An Undergraduate Student Flow Model: Australian Higher Education. *Higher Education*, 37, 359-375. <http://dx.doi.org/10.1023/A:1003765222250>
- Smith, J. P., & Robin, A. N. (2001). Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students. *Journal of the Royal Statistical Society. Series A*, 164(Part 2), 389-405. <http://dx.doi.org/10.1111/1467-985X.00209>
- Song, Q. & Chissom, B. S. (1994). Forecasting enrollments with fuzzy time series. *Fuzzy Sets Syst.*, 62, 1-8. [http://dx.doi.org/10.1016/0165-0114\(94\)90067-1](http://dx.doi.org/10.1016/0165-0114(94)90067-1)
- Symeonaki, M., & Kalamatianou, A. (2011). Markov Systems with Fuzzy States for Describing Students' Educational Progress in Greek Universities, ISI 2011. Retrieved from <http://isi2011.congressplanner.eu/showabstract.php?congress=ISI2011&id=1896>
- Therneau, T., & Atkinson, B. (2012). Package 'rpart'. Retrieve from <http://cran.rproject.org/web/packages/rpart/rpart.pdf>
- Triventi, M., & Trivellato, P. (2009). Participation, Performance and Inequality in Italian Higher Education in 20th Century. *Higher Education*, 57(6), 681-702. <http://dx.doi.org/10.1007/s10734-008-9170-0>