

Un approccio Markoviano per l'analisi
longitudinale dei percorsi degli universitari

Giulio Ghellini, Francesco Maccari, Riccardo Cordone

Working Paper n. 77, Marzo 2009

Un approccio Markoviano per l'analisi longitudinale dei percorsi degli universitari

Giulio Ghellini, Francesco Maccari, Riccardo Cordone

Abstract

Il presente lavoro si inserisce nel filone di ricerca riguardante l'utilizzo di modelli Markoviani per la rappresentazione e l'analisi dei sistemi educativi. In particolare sono presentate alcune proposte di adattamento del modello per coorti scolastiche, recentemente formalizzato da Ghellini e Maccari (2007) relativamente allo studio delle carriere delle scuole medie superiori, all'analisi delle carriere universitarie. Particolarmente rilevante appare l'individuazione del livello di CFU raggiunto dagli studenti in determinati istanti temporali come elemento fondamentale per la misurazione del progresso nella formazione dello studente. Mediante opportune aggregazioni in classi dei CFU, si definiscono pertanto gli stati in cui si possono venire a trovare gli studenti nel corso del tempo, permettendo una definizione delle transizioni e delle uscite in analogia con quanto presente nel sistema scolastico tradizionale. Tale definizione rende quindi possibile la formulazione, anche nel contesto universitario, del concetto di traiettoria e la conseguenziale formalizzazione di alcuni indicatori sintetici di *performance*, assimilabili, con i dovuti distinguo, a quelli già esistenti per il sistema scolastico tradizionale. Il caso pratico di studio è rappresentato da alcune coorti di studenti immatricolati presso la Facoltà di Economia "R. M. Goodwin" dell'Università di Siena negli anni immediatamente successivi alla riforma universitaria introdotta dal DM 509/99, per le quali vengono presentati alcuni primi indicatori di *performance* ed effettuati confronti sia temporali che tra sottogruppi di studenti (per genere, provenienza geografica, ecc.)

1. Premessa

Il sistema Universitario del nostro Paese è stato, come noto, interessato in questi ultimi anni da ripetuti interventi legislativi, volti a profonde modifiche della sua offerta formativa. Basti ricordare la riforma dei cicli di studio delineata dal DM 509/99, che ha introdotto il cosiddetto “3+2”, e il DM 270/2004, con il relativo travagliato iter per la definizione delle Classi di Laurea del nuovo assetto (sintetizzabile nella formula “3 e 2”), per comprendere l’importanza di disporre di strumenti conoscitivi idonei a rendere conto in modo coerente degli effetti che tali interventi hanno avuto, e stanno avendo, nella definizione delle carriere formative delle generazioni coinvolte. Modifiche così rilevanti degli assetti, non solo organizzativi ma anche di contenuto, dell’offerta formativa dovrebbero infatti, a nostro avviso, essere sempre accompagnate dalla contemporanea messa a punto di coerenti sistemi informativi in grado di supportare la produzione di indicatori di monitoraggio, utili sia per formulare valutazioni sugli effetti dell’intervento, sia per permettere esercizi simulativi. Inoltre la disponibilità di indicatori sul processo in atto dovrebbero fornire anche utili elementi alle strutture direttamente responsabili dell’offerta formativa (si pensi in particolare alle Facoltà e ai Corsi di Studio), per programmare interventi di aggiustamento in corso d’opera.

Il presente lavoro vuole quindi rappresentare un primo modesto contributo in tale direzione, proponendo la formulazione di un compatto modello di funzionamento del processo formativo, in analogia con quanto è possibile fare per i certamente più strutturati processi formativi del sistema scolastico pre-universitario. Tale proposta si basa sulla disponibilità presso i vari Atenei di coerenti informazioni sulle effettive carriere degli studenti universitari, disponibilità che non di rado si presenta non del tutto agevole e comunque contrassegnata da rilevanti problemi di coerenza e qualità nelle informazioni raccolte. Troppo spesso, infatti, il sistema informativo sulle carriere risulta disegnato per rispondere a meri compiti amministrativi, con una marginale attenzione agli aspetti più propriamente statistico-informativi dei dati raccolti, così come evidenzia anche questo primo tentativo di applicazione del modello proposto (vedi sezione 5).

Dopo aver brevemente ricordato l’ambito metodologico all’interno del quale si pone tale modello, i processi stocastici markoviani, e richiamate le principali assunzioni per la loro applicazione nello studio delle carriere formative (sezione 2), vengono introdotte la formalizzazione del modello per coorti - basato sull’innovativo concetto di “traiettoria” - la conseguente proposta di una definizione degli stati sperimentabili dagli studenti, basata sui CFU conseguiti, e la conseguente definizione di matrici di probabilità di transizione nel e di uscita dal sistema universitario (sezione 3). Sono quindi introdotti alcuni primi indicatori di sicuro interesse per l’analisi delle carriere degli universitari, direttamente desumibili dal modello (sezione 4) e una prima applicazione sui dati relativi a due coorti di studenti della facoltà di Economia dell’università di Siena (sezione 5).

2 Modelli Markoviani per l’analisi delle carriere degli studenti: filone di ricerca e stato dell’arte

Il lavoro si inserisce nel filone di ricerca riguardante l’utilizzo di modelli Markoviani per la rappresentazione e l’analisi dei sistemi educativi. In tal senso, il sistema educativo viene visto come un sistema sociale, caratterizzato dalle transizioni che gli individui, nel corso del tempo, effettuano tra diversi stati possibili. Pertanto, un sistema sociale è rappresentato come una catena di Markov, coerentemente con un insieme di assunzioni prefissate.

Il filone di ricerca in questione trova le sue origini nei lavori di Quesnay (1758) e Quetelet (1849), che furono tra i primi ad intuire la possibilità di introdurre l’uso di modelli matematici nello

studio dei fenomeni sociali, andando oltre ad un approccio puramente descrittivo. Nella seconda metà del secolo scorso, tale ambito di ricerca ha goduto di un considerevole sviluppo. Riferendosi nello specifico alla rappresentazione del sistema scolastico tramite modelli Markoviani, i contributi più rilevanti alla base di questo lavoro sono quelli di Stone (1965 e 1972), Thonstad (1969) e, specialmente, Bartholomew (1982). In ambito italiano, il modello proposto nel seguito trova i suoi fondamenti nel modello SFINGE (Trivellato, 1980; Bernardi e Trivellato, 1980; Bernardi et al., 1986), particolarmente per quanto attiene alla produzione di indicatori di *performance* scolastica. Una rilevante caratteristica comune a tutti i modelli citati risiede nel fatto che essi utilizzano dati di stock, che, com'è noto, costituiscono un serio limite alla misurazione delle reali transizioni degli studenti all'interno del sistema analizzato.

La recente disponibilità di un set di dati di natura individuale, particolarmente ampio e accurato, sulle carriere scolastiche degli studenti delle scuole superiori della provincia di Pisa ha fornito il primo impulso a ricercare affinamenti dei modelli presenti in letteratura per la rappresentazione e la misurazione delle transizioni degli studenti. In particolare, dal momento che tali dati sono stati raccolti in un ampio arco temporale (10 anni scolastici, dal 1992/93 al 2001/02), tale da coprire per intero le carriere di alcune generazioni di studenti, si è reso possibile uno sviluppo metodologico dei modelli basati su dati di stock, utilizzando basi di dati di natura longitudinale, certamente più coerenti con l'approccio Markoviano. La disponibilità di tali informazioni ha implicato una precisa definizione dell'aggregato *coorte scolastica*, definita come l'insieme degli studenti che si iscrivono per la prima volta alla classe iniziale della scuola superiore nello stesso anno scolastico. Alla base di questo sviluppo risiede la convinzione che gli esiti scolastici degli studenti siano fortemente dipendenti dalla coorte di appartenenza. In tale ottica, il principale lavoro di riferimento è quello di Légaré (Légaré, 1972), che ben formalizza la dimensione temporale del sistema scolastico.

Questo primo sviluppo ha portato alla proposta di un modello di rappresentazione delle carriere scolastiche basato sul concetto fondamentale di *traiettoria scolastica* (Maccari, 2006; Ghellini e Maccari, 2007). Il concetto di traiettoria è definito come l'insieme di tutte le possibili successioni di transizioni (promozioni e ripetenze) che portano da un dato stato iniziale ad uno finale in un determinato numero di anni scolastici. Particolare importanza riveste il fatto che la probabilità osservata di ciascuna traiettoria è ottenibile semplicemente sulla base delle matrici di transizione, ossia utilizzando le probabilità osservate delle sole promozioni e ripetenze occorse nel periodo in questione. I principali indicatori di *performance* formalizzati (indicatori di permanenza, di uscita e di aspettativa di vita scolastica) sono calcolati utilizzando le probabilità osservate delle traiettorie interne e di uscita dal sistema scolastico considerato.

Il modello di rappresentazione delle transizioni e delle traiettorie presenta il pregio della compattezza e della velocità di calcolo delle matrici per mezzo di apposite procedure informatiche. Gli indicatori proposti sono particolarmente utili ai fini di un confronto di *performance* tra coorti diverse o tra gruppi distinti (per genere, età, scuola frequentata, ecc.) di studenti appartenenti ad una stessa coorte.

I promettenti risultati ottenuti con l'approccio testé ricordato hanno stimolato l'interesse a verificare se e in che misura fosse possibile proporre un'estensione di tale modello per lo studio delle carriere universitarie. E' apparsa subito chiara, a tal fine, la necessità di apportare al modello affinamenti in grado di rendere conto della diversa, e certamente più complessa, evoluzione delle carriere che caratterizzano la vita universitaria. Questa maggiore complessità rende certamente più difficile ottenere misure idonee e ben interpretabili dell'*outcome* del sistema universitario, in particolare dei tempi di permanenza, degli abbandoni, delle uscite con titolo, ma anche di eventi

intermedi, quali il livello dei crediti formativi universitari (CFU) conseguiti dagli studenti nel corso della loro carriera a determinati istanti di tempo. Nel caso del sistema universitario, i livelli formativi intermedi possono essere in qualche misura rappresentati dal numero di CFU ottenuti rilevati in determinati momenti della carriera dello studente; questi costituiscono infatti la misura ufficiale del “progresso” dello studente in termini di formazione universitaria.

Prima di passare alla formalizzazione del modello oggetto di questo lavoro, sembra utile richiamare le assunzioni che stanno alla base dei modelli Markoviani classici per lo studio dei sistemi educativi e della versione per coorti richiamata sopra.

Assunzioni dei modelli Markoviani classici:

- a) l'unità elementare del processo scolastico è costituita dal singolo studente;
- b) il sistema scolastico è rappresentato da un numero finito di stati, che possono essere transienti (n classi) o assorbenti (m livelli di istruzione);
- c) ciascuno studente può trovarsi in un solo stato ad un dato istante di tempo;
- d) uno studente che si trova in uno stato transiente può passare ad un altro stato o uscire dal sistema soltanto alla fine del corrente anno scolastico;
- e) la probabilità che uno studente si trovi in un certo stato (transiente o assorbente) al tempo $t+1$ dipende soltanto dallo stato in cui si trovava al tempo t (in altre parole si ipotizza l'indipendenza delle transizioni dalla storia pregressa);
- f) le probabilità di transizione tra due qualsivoglia stati sono costanti nel tempo (di conseguenza, non sono posti vincoli al numero di ripetenze possibili).

Assunzioni del modello scolastico per coorti (Ghellini-Maccari):

- a) uno studente appartiene alla coorte t solo se entra nella classe iniziale del sistema scolastico considerato all'inizio dell'anno scolastico t ;
- b) al termine di ciascun anno scolastico, uno studente può i) passare alla classe successiva, ii) ripetere la classe corrente, o iii) uscire dal sistema;
- c) gli eventi si realizzano in un arco temporale di tipo discreto, ossia gli studenti possono cambiare stato solo al termine di ciascun anno scolastico;

3. Il modello proposto per lo studio delle carriere universitarie di coorte

In questa sezione viene formulata la proposta di un modello per l'analisi delle carriere universitarie. Ci si soffermerà in particolar modo sulle caratteristiche formali del modello, sulla definizione dello spazio degli stati e del concetto di traiettoria.

Come nel caso del modello applicato al sistema scolastico superiore, è opportuno innanzi tutto definire le assunzioni fondamentali, che qui di seguito presentiamo.

- a) uno studente appartiene alla coorte t solo se si immatricola per la prima volta al primo anno di corso del sistema universitario considerato (Ateneo, Facoltà, Corso di Studio, ecc.) all'inizio dell'anno accademico t (ed entra, di conseguenza nella classe di crediti zero);
- b) in corrispondenza di ciascun istante di osservazione, uno studente può: i) restare nel sistema passando ad un livello di CFU uguale o maggiore del precedente; ii) uscire dal

sistema in corrispondenza di un livello di CFU uguale o maggiore del precedente; iii) uscire dal sistema con titolo finale;

- c) gli eventi si realizzano in un arco temporale di tipo discreto, ossia gli studenti possono cambiare stato solo in corrispondenza di ciascun istante di osservazione;
- d) non sono posti vincoli al tempo in cui lo studente può permanere all'interno del sistema (e quindi, al numero di osservazioni per studente).

Si può subito notare come le assunzioni siano coerenti con quelle alla base dei modelli Markoviani classici, richiamate nel paragrafo precedente; in particolare, si può sin da ora affermare che il sistema universitario, al pari di quello scolastico, sotto tali assunzioni, sia rappresentabile come una catena di Markov finita assorbente. Si riscontrano tuttavia alcune differenze, che dipendono principalmente da due aspetti, rispetto ai quali i due sistemi educativi si distinguono abbastanza nettamente.

- i) Innanzi tutto, nel caso del sistema universitario non è possibile considerare l'anno di corso come indicazione del progresso dello studente. La nostra proposta è quella di sostituire le classi scolastiche con i CFU nel misurare il livello di avanzamento degli studenti all'interno del sistema. Per ovvie ragioni di praticità e facilità di interpretazione, i CFU devono essere raggruppati in classi, al fine di ottenere un numero ragionevole di stati che uno studente può visitare nell'arco della propria carriera. Si noti che tale procedura di aggregazione in classi rappresenta una fase alquanto delicata dell'analisi, in quanto la scelta del numero e dell'ampiezza delle classi di CFU può influenzare notevolmente i risultati del modello in termini di transizioni. Ad ogni modo, come si desume anche dall'assunzione b), le carriere universitarie presentano una struttura ben più complessa rispetto a quelle della scuola secondaria superiore; infatti, mentre nel caso scolastico le transizioni interne al sistema sono caratterizzate da una dicotomia (promozione o ripetenza), nel caso universitario lo studente, in ciascun istante in cui è osservato, può, almeno in teoria, compiere una molteplicità di transizioni (verso tutte le classi di crediti maggiori o uguali alla precedente). Ciò si riflette, come vedremo, anche in una maggiore complessità nella definizione di traiettoria.
- ii) Il secondo aspetto di discontinuità con il modello scolastico è costituito dai tempi di rilevazione, o, in altre parole, dai momenti in cui possono avvenire le transizioni e le uscite dal sistema. Nel caso della scuola superiore, si può assumere che le transizioni si realizzino unicamente alla fine di ciascun anno scolastico; in virtù di tale assunzione, la disponibilità di rilevazioni annuali non comporta alcuna perdita di informazione; in altre parole, ogni evento d'interesse della carriera degli studenti può essere rilevato esattamente. Nel caso del sistema universitario, il quadro è molto più articolato, in quanto ogni studente può teoricamente conseguire CFU aggiuntivi in qualsiasi momento della propria carriera. Ciò evidenzia il problema della scelta dei tempi di rilevazione dei CFU accumulati dagli studenti della coorte considerata: all'aumentare della frequenza delle rilevazioni si riduce la perdita d'informazione sulle effettive transizioni effettuate dagli studenti nei periodi intercorrenti tra osservazioni successive, ma, chiaramente, tale perdita non può mai essere del tutto evitata. Inoltre, è ovvio che l'aumento della frequenza delle osservazioni comporta anche un aumento dei costi di rilevazione. Pertanto, la scelta deve essere mirata ad un ragionevole equilibrio tra parsimonia (contenere la frequenza delle rilevazioni) e rappresentatività (contenere la perdita d'informazione).

Un discorso a parte merita la definizione dello spazio degli stati, che, nel caso del modello scolastico, è dato dall'insieme di tutti i possibili stati ottenibili considerando congiuntamente la coppia (*Classe di iscrizione, Numero di ripetenze passate*). Il concetto di ripetenza, nel sistema universitario attuale, non è però applicato, pertanto non è ovviamente possibile considerarlo nella specificazione del modello. La misura del progresso dello studente è invece rappresentata dal livello di CFU acquisiti. Di conseguenza, lo spazio degli stati è qui costituito dall'insieme delle possibili combinazioni tra le classi di CFU accumulati e gli istanti di rilevazione considerati.

In analogia formale con il modello scolastico, si definisce dapprima un insieme di matrici di transizioni interne $\mathbf{T}^{(y)}$ ($y = 1, \dots, Y$), rappresentabili come segue:

$$\mathbf{T}^{(y)} = \begin{bmatrix} n_{1,1}^{(y)} & n_{1,2}^{(y)} & \dots & n_{1,g_2}^{(y)} & \dots & n_{1,K}^{(y)} \\ 0 & n_{2,2}^{(y)} & \dots & n_{2,g_2}^{(y)} & \dots & n_{2,K}^{(y)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & n_{g_1,g_2}^{(y)} & \dots & n_{g_1,K}^{(y)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 0 & n_{K,K}^{(y)} \end{bmatrix},$$

dove Y è dato dal numero totale delle osservazioni successive all'evento origine della coorte (con $y=0$ si indica proprio l'evento origine della coorte), mentre K è pari al numero delle classi in cui sono raggruppati i CFU.

Il generico elemento $n_{g_1,g_2}^{(y)}$ indica il numero di studenti che, nell'intervallo di tempo compreso tra l'istante $y-1$ e l'istante y , sono rimasti all'interno del sistema considerato, passando dalla classe di crediti g_1 alla classe di crediti g_2 (eventi di transizione interna).

Si definisce poi un secondo insieme di matrici $\mathbf{U}^{(y)}$ ($y = 1, \dots, Y$), riguardanti invece gli eventi di uscita. Occorre precisare che anche nel caso delle uscite viene rilevato il livello (finale) di CFU raggiunto dallo studente, raggruppato secondo la medesima classificazione utilizzata in precedenza; di conseguenza, il modello prevede $K+1$ stati assorbenti, di cui i primi K sono riferiti alle uscite senza titolo (ognuno in corrispondenza di una diversa classe di CFU), mentre l'ultimo è riferito alle uscite avvenute in seguito al conseguimento del titolo finale (laurea). Si ha pertanto:

$$\mathbf{U}^{(y)} = \begin{bmatrix} u_{1,1}^{(y)} & u_{1,2}^{(y)} & \dots & u_{1,k}^{(y)} & \dots & u_{1,K}^{(y)} & u_{1,K+1}^{(y)} \\ 0 & u_{2,2}^{(y)} & \dots & u_{2,k}^{(y)} & \dots & u_{2,K}^{(y)} & u_{2,K+1}^{(y)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & u_{g,k}^{(y)} & \dots & u_{g,K}^{(y)} & u_{g,K+1}^{(y)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 0 & u_{K,K}^{(y)} & u_{K,K+1}^{(y)} \end{bmatrix}$$

Il generico elemento $u_{g,k}^{(y)}$ indica il numero di studenti che, nell'intervallo di tempo compreso tra l'istante $y-1$ e l'istante y , hanno compiuto una transizione in uscita dalla classe di crediti g allo stato assorbente k .

In conclusione, l'insieme delle matrici $\mathbf{T}^{(y)}$ e $\mathbf{U}^{(y)}$ contiene le frequenze osservate di tutti gli eventi che si sono verificati nell'intero periodo di osservazione della coorte.

Sia poi $N_{g_1}^{(y)}$ il numero totale di studenti che, trovandosi nello stato g_1 all'istante $y-1$, hanno effettuato una transizione verso un qualsiasi altro stato, transiente o assorbente, ossia:

$$N_{g_1}^{(y)} = \sum_{g_2=1}^K n_{g_1, g_2}^{(y)} + \sum_{k=1}^{K+1} u_{g_1, k}^{(y)}.$$

Si possono costruire facilmente le matrici $\mathbf{P}^{(y)}$, contenente le probabilità osservate delle transizioni interne al sistema universitario nel periodo che va da $y-1$ a y , e $\mathbf{L}^{(y)}$, contenente le probabilità osservate delle uscite dal sistema, in corrispondenza dei possibili stati (livello finale di CFU), nel medesimo periodo di osservazione.

Gli elementi di $\mathbf{P}^{(y)}$ e $\mathbf{L}^{(y)}$ sono calcolati nel modo seguente:

$$p_{g_1, g_2}^{(y)} = \frac{t_{g_1, g_2}^{(y)}}{N_{g_1}^{(y)}} \quad (g_1, g_2 = 1, \dots, K) \quad [1]$$

$$l_{g_1, k}^{(y)} = \frac{u_{g_1, k}^{(y)}}{N_{g_1}^{(y)}} \quad (g_1 = 1, \dots, K; k = 1, \dots, K+1) \quad [2]$$

Fissati y e g_1 , risulta immediatamente verificata la seguente proprietà:

$$\sum_{g_2=1}^K p_{g_1, g_2}^{(y)} + \sum_{k=1}^{K+1} l_{g_1, k}^{(y)} = 1$$

Pertanto, le matrici $\mathbf{P}^{(y)}$ e $\mathbf{L}^{(y)}$ sono sufficienti a descrivere interamente tutti gli eventi che possono caratterizzare la storia della coorte analizzata, limitatamente all'arco temporale di osservazione. Le due matrici sono rappresentabili graficamente come segue.

$$\mathbf{P}^{(y)} = \begin{bmatrix} p_{1,1}^{(y)} & p_{1,2}^{(y)} & \dots & p_{1,g_2}^{(y)} & \dots & p_{1,K}^{(y)} \\ 0 & p_{2,2}^{(y)} & \dots & p_{2,g_2}^{(y)} & \dots & p_{2,K}^{(y)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & p_{g_1, g_2}^{(y)} & \dots & p_{g_1, K}^{(y)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 0 & p_{K, K}^{(y)} \end{bmatrix}$$

$$\mathbf{L}^{(y)} = \begin{bmatrix} l_{1,1}^{(y)} & l_{1,2}^{(y)} & \dots & l_{1,k}^{(y)} & \dots & l_{1,K}^{(y)} & l_{1, K+1}^{(y)} \\ 0 & l_{2,2}^{(y)} & \dots & l_{2,k}^{(y)} & \dots & l_{2,K}^{(y)} & l_{2, K+1}^{(y)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & l_{g_1, k}^{(y)} & \dots & l_{g_1, K}^{(y)} & l_{g_1, K+1}^{(y)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 0 & l_{K, K}^{(y)} & l_{K, K+1}^{(y)} \end{bmatrix}$$

In analogia con il modello scolastico, l'informazione contenuta nelle matrici $\mathbf{P}^{(y)}$ è utilizzata per pervenire alla *matrice delle traiettorie*, che, come nel caso del modello scolastico, servirà da base per la formulazione di *indicatori di performance* universitaria.

Nello studio di una coorte universitaria, una traiettoria è un possibile percorso interno al sistema, caratterizzato dai seguenti parametri:

- g_b : la classe iniziale di crediti della traiettoria ($1 \leq g_b \leq K$);
- y_b : l'istante a cui si riferisce l'osservazione iniziale della traiettoria ($0 \leq y_b \leq Y-1$);
- g_e : la classe finale di crediti della traiettoria ($g_b \leq g_e \leq K$);
- d : la durata della traiettoria, espressa come numero di osservazioni intercorse tra l'inizio e la fine della traiettoria ($1 \leq d \leq Y - y_b$).

La probabilità osservata di una qualunque traiettoria, indicata con $w_{g_b, y_b, g_e, d}$, ancora in analogia col modello scolastico, è ottenibile dalla seguente espressione:

$$w_{g_b, y_b, g_e, d} = \begin{cases} p_{g_b, g_e}^{(y_b+1)} & \text{se } d = 1 \\ \sum_{j=g_b}^{g_e} p_{g_b, j}^{(y_b+1)} \cdot w_{j, y_b+1, g_e, d-1} & \text{se } d > 1 \end{cases} \quad [3]$$

Come si vede, la struttura della formula è ricorsiva, in quanto il calcolo della probabilità di una traiettoria di durata superiore ad un periodo è funzione delle probabilità delle sotto-traiettorie di durata inferiore. In definitiva, la probabilità di qualsiasi traiettoria può essere scomposta nel prodotto di singoli elementi delle matrici $\mathbf{P}^{(y)}$.

4. Indicatori di *performance*

Vengono qui formalizzati alcuni primi indicatori di *performance*, in merito ai quali si potrà notare la forte continuità con quelli relativi al modello scolastico. Gli indicatori proposti si concentrano sui due eventi fondamentali della carriera degli studenti universitari: il conseguimento della laurea e l'abbandono (ossia, l'uscita dal sistema universitario senza titolo di studio).

Indicatori di conseguimento della laurea

La prima classe di indicatori è riferita al conseguimento del titolo finale di studio (laurea). In coerenza con la formalizzazione introdotta nel precedente paragrafo, il seguente indicatore misura la probabilità osservata di conseguire la laurea (ossia di raggiungere la classe di uscita $K+1$) dopo y periodi di osservazione dall'evento origine, condizionata al raggiungimento, dopo y_b periodi di osservazione, della classe di CFU g_b .

$$IL_{g_b, y_b}^{(y)} = \begin{cases} l_{g_b, K+1}^{(y)} & \text{se } y - y_b = 1 \\ \sum_{g_e=g_b}^K w_{g_b, y_b, g_e, y-y_b-1} \cdot l_{g_e, K+1}^{(y)} & \text{se } y - y_b > 1 \end{cases} \quad [4]$$

$$(1 \leq g_b \leq K; \quad 1 \leq y \leq Y; \quad 0 \leq y_b < y)$$

Si noti che, per $y_b = 0$, l'indicatore è non condizionato, mentre, per $0 < y_b < y$, risulta condizionato ad uno stato intermedio, rappresentato dalla classe g_b di CFU conseguiti in corrispondenza di una data osservazione (y_b). Occorre inoltre precisare che, dato y , il relativo indicatore $IL^{(y)}$ indica la probabilità di conseguire la laurea esattamente in corrispondenza dell' y -esimo tempo di rilevazione. Scegliendo per y il valore equivalente al conseguimento del titolo in regola, si ottiene l'*Indicatore di laurea in regola (ILR)*.

Componendo gli opportuni valori di $IL^{(y)}$ si ottiene inoltre l'indicatore IL^* , che misura la probabilità di laurea *entro* un dato numero di periodi osservati, nel modo che segue.

$$IL_{g_b, y_b}^{*(y)} = \sum_{h=1}^y IL_{g_b, y_b}^{(h)} \quad (1 \leq g_b \leq K; \quad 1 \leq y \leq Y; \quad 0 \leq y_b < y) \quad [5]$$

E' infine possibile ottenere, tramite una semplice media ponderata, una generalizzazione degli indicatori [4] e [5], condizionandosi non già ad una singola classe di crediti g_b , ma all'unione di più classi (per semplicità si considerano qui classi di crediti contigue). Supponiamo infatti di voler calcolare la probabilità del conseguimento della laurea in y periodi di tempo, condizionata al raggiungimento di un livello di crediti compreso nell'intervallo di classi $[g_1, g_2]$ al tempo y_b . L'indicatore di conseguimento del titolo diviene:

$$IL_{g_1, g_2, y_b}^{(y)} = \frac{\sum_{g=g_1}^{g_2} IL_{g, y_b}^{(y)} \cdot N_{g, \cdot}^{(y_b+1)}}{\sum_{g=g_1}^{g_2} N_{g, \cdot}^{(y_b+1)}} \quad (1 \leq g_1 < g_2 \leq K; \quad 1 \leq y \leq Y; \quad 0 \leq y_b < y) \quad [6]$$

Si può procedere in modo analogo anche per l'indicatore IL^* .

Indicatori di uscita senza titolo

L'indicatore di uscita senza titolo si ottiene in modo del tutto simile a quello di conseguimento del titolo. Occorre però tener presente che l'uscita senza titolo può verificarsi in corrispondenza di K stati diversi (corrispondenti alle possibili classi di CFU in uscita). Pertanto, la sua formulazione, come si vede nel seguito, richiede l'introduzione di un ulteriore parametro.

L'indicatore di uscita senza titolo misura la probabilità osservata di uscita dal sistema universitario in corrispondenza della g_e -esima classe di CFU dopo y periodi di osservazione dall'evento origine, condizionatamente al raggiungimento, dopo y_b periodi di osservazione, della classe di CFU g_b .

$$IU_{g_b, g_e, y_b}^{(y)} = \begin{cases} l_{g_b, g_e}^{(y)} & \text{se } y - y_b = 1 \\ \sum_{g=g_b}^{g_e} w_{g_b, y_b, g, y - y_b - 1} \cdot l_{g, g_e}^{(y)} & \text{se } y - y_b > 1 \end{cases} \quad [7]$$

$$(1 \leq g_b \leq g_e \leq K; \quad 1 \leq y \leq Y; \quad 0 \leq y_b < y)$$

Le considerazioni fatte in merito all'indicatore IL valgono anche per IU . Anche in questo caso, inoltre, è possibile calcolare l'indicatore di uscita *entro* un dato numero di anni, nel modo che segue:

$$IU_{g_b, g_e, y_b}^{*(y)} = \sum_{h=1}^y IU_{g_b, g_e, y_b}^{(h)} \quad (1 \leq i \leq j \leq K; \quad 1 \leq y \leq Y; \quad 0 \leq y_b < y) \quad [8]$$

Ancora in analogia con quanto visto per l'indicatore di conseguimento della laurea, si può ottenere una forma più generale degli indicatori [7] e [8]: si noti che, in questo caso, la generalizzazione riguarda non solo il livello di CFU di partenza a cui ci si condiziona, ma anche quello di uscita.

Siano pertanto $[g_{b1}, g_{b2}]$ l'intervallo di CFU iniziale e $[g_{e1}, g_{e2}]$ quello di uscita, l'indicatore IU è dato dalla seguente espressione:

$$IU_{g_{b1}, g_{b2}, g_{e1}, g_{e2}, y_b}^{(y)} = \frac{\sum_{i=g_{b1}}^{g_{b2}} \sum_{j=g_{e1}}^{g_{e2}} IU_{i,j,y_b}^{(y)} \cdot N_{i,\cdot}^{(y_b+1)}}{\sum_{i=g_{b1}}^{g_{b2}} N_{i,\cdot}^{(y_b+1)}} \quad [9]$$

$$(1 \leq g_{b1} < g_{b2} \leq K; \quad g_{b1} \leq g_{e1} < g_{e2} \leq K; \quad 1 \leq y \leq Y; \quad 0 \leq y_b < y)$$

5. Il caso di studio

Il modello sinora descritto è in corso di applicazione per l'analisi delle coorti di studenti immatricolati presso la Facoltà di Economia "R. M. Goodwin" dell'Università degli Studi di Siena negli A.A. 2001-02 e 2002-03. I dati utilizzati nell'analisi sono stati forniti dagli archivi amministrativi di Ateneo.

Per ciascuna coorte, sono stati selezionati, all'interno del set di informazioni disponibili, i valori di tre variabili qualitative di specifico interesse per l'analisi, volte a descrivere le caratteristiche "in ingresso" degli studenti, (genere, area geografica di provenienza e tipo di diploma di maturità di scuola secondaria superiore). Inoltre, al fine di generare i dati di input del modello per coorti, sono state utilizzate le rilevazioni dello status d'iscrizione e del numero di CFU conseguiti a vari istanti temporali successivi lungo l'arco di vita della coorte. In particolare, sono state selezionate le rilevazioni effettuate al 30 aprile di ogni anno di vita delle coorti, dato che tale data costituisce il termine entro il quale ogni studente ha la possibilità di terminare gli esami previsti nell'anno accademico precedente. Pertanto, per la coorte 2001 sono disponibili 4 osservazioni (30 aprile 2003, 2004, 2005 e 2006), mentre per la coorte seguente le osservazioni sono soltanto 3 (30 aprile 2004, 2005 e 2006).

Purtroppo, a causa di problemi di alcune carenze informative presenti nelle basi di dati utilizzate, non è stato possibile ricostruire la serie delle osservazioni degli studenti che hanno effettuato un cambiamento di corso di studi all'interno della Facoltà. Tali trasferimenti interni, pertanto, sono stati considerati alla stregua di uscite; ciò comporta, nell'applicazione che segue, una sovrastima degli indicatori di uscita senza titolo e, allo stesso tempo, una sottostima di quelli di laurea.

La tabella 1 mostra la composizione delle due coorti analizzate per opportune aggregazioni delle variabili "di ingresso".

Tabella 1 – Distribuzione percentuale delle coorti per variabile di ingresso

<i>Variabile</i>	<i>Specificazione</i>	<i>Coorte 2001</i>	<i>Coorte 2002</i>	<i>Δ coorti (2002 - 2001)</i>
Genere	Maschio	54,2	52,5	-1,7
	Femmina	45,8	47,5	+1,7
Area geogr. di provenienza	Toscana	61,8	57,0	-4,8
	Sud e Isole	28,6	31,2	+2,6
	Altra prov.	9,6	11,8	+2,2
Tipo di diploma	Liceo	40,4	35,6	-4,8
	Altro dipl.	59,6	64,4	+4,8
(Numerosità coorte)		(793)	(651)	(-142)

Osservando i dati in tabella 1, si può notare, in primo luogo, una diminuzione abbastanza rilevante della numerosità totale della coorte 2002 rispetto alla coorte 2001 (142 studenti in meno, pari ad una diminuzione percentuale di quasi il 18%). Tale evidenza può essere in parte legata alla particolarità della coorte 2001, che è stata la prima dopo l'introduzione della riforma.

Per quanto attiene alla distribuzione delle variabili rilevate, nella coorte 2002 si rafforza leggermente la componente femminile (+1,7%) e, in modo più deciso, la provenienza extra-Toscana (+4,8%); si riduce invece del 4,8% la presenza di studenti con maturità liceale.

Passando ora all'applicazione del modello per l'analisi delle carriere, si fa riferimento agli indicatori formalizzati nella sezione precedente, relativi al conseguimento del titolo finale (in regola e in ritardo) e a quelli di uscita ai diversi istanti temporali.

La tabella 2 mostra il valore dell'indicatore di laurea in regola (*ILR*) calcolato per le due coorti complete per gruppi distinti in base alle variabili "di ingresso".

Tabella 2 – *ILR per specificazioni delle variabili "di ingresso" e per coorte*

<i>Variabile</i>	<i>Specificazione</i>	<i>Coorte 2001</i>	<i>Coorte 2002</i>	<i>Δ ILR (C2002-C2001)</i>
Genere	Maschio	0,112	0,120	+0,008
	Femmina	0,207	0,230	+0,023
Area geogr. di provenienza	Toscana	0,118	0,178	+0,060
	Sud e Isole	0,242	0,167	-0,075
	Altra prov.	0,132	0,156	+0,024
Tipo di diploma	Liceo	0,203	0,233	+0,030
	Altro dipl.	0,123	0,138	+0,015
Totale coorte		0,155	0,172	+0,017

Per quanto riguarda il confronto tra le coorti complete, si nota un leggero incremento dell'*ILR* (pari a circa l'1,7%), pur nell'ambito di valori ben poco confortanti (solo il 17,2% degli studenti appartenenti alla coorte 2002 si è laureato in regola con i tempi)¹.

¹ Si tenga sempre presente, nell'interpretazione dei risultati del modello, l'esistenza di significative distorsioni dovute all'impossibilità di rilevare correttamente i trasferimenti interni alla Facoltà, come già spiegato in precedenza.

Più interessanti i confronti tra gruppi, in merito ai quali si nota, per entrambe le coorti, una migliore *performance* da parte delle femmine rispetto ai maschi, e degli ex-liceali rispetto ai possessori di altri tipi di diploma. Riguardo all'analisi per provenienza, invece, le due coorti si differenziano in modo rilevante: nella coorte 2001 l'*ILR* assume un valore ben più elevato per gli studenti provenienti da fuori Toscana, mentre nella coorte seguente la situazione si livella su valori molto simili per tutte le provenienze.

La tabella 3 riporta, per la sola coorte 2001 e per le relative disaggregazioni, i valori degli indicatori $IL^{(4)}$ (laurea in 4 anni) e $IL^{(4)*}$ (laurea entro 4 anni).

Innanzitutto, è opportuno rimarcare come l'indicatore di laurea in 4 anni, per la coorte complessiva, sia quasi equivalente a quello di laurea in regola (0,153 contro 0,155). Osservando poi i valori dell'indicatore per specificazione delle diverse variabili di ingresso, si nota subito come al quarto anno si annulli quasi del tutto la differenza di performance per genere (che era invece assai rilevante per quanto riguarda la laurea in regola), mentre restano marcate quelle relative alle diverse aree di provenienza e ai differenti tipi di diploma.

Tabella 3 – IL4 e IL4 per la coorte 2001 e per specificazioni delle variabili “di ingresso”*

<i>Variabile</i>	<i>Specificazione</i>	<i>IL4 C2001</i>	<i>IL4* C2001</i>
Sesso	Maschio	0,151	0,263
	Femmina	0,154	0,361
Area geogr. di provenienza	Toscana	0,139	0,257
	Sud e Isole	0,177	0,419
	Altra prov.	0,171	0,303
Tipo di diploma	Liceo	0,191	0,394
	Altro dipl.	0,126	0,249
Totale coorte		0,153	0,308

Ovviamente, date le variabili a disposizione, vi è la possibilità di definire gli indicatori sin qui visti anche per specifici sottogruppi di interesse nell'ambito delle coorti analizzate².

Per quanto riguarda l'analisi delle uscite, la tabella 4 riporta il valore dell'indicatore di uscita senza titolo, ad un qualsivoglia livello di CFU, entro i possibili istanti temporali di rilevazione (da 1 a 4 anni per la coorte 2001, da 1 a 3 anni per la coorte 2002) per genere. I valori in ogni colonna rappresentano in sostanza la “progressione” nel tempo delle uscite senza titolo.

² A titolo di esempio, si osservi che, in riferimento alla coorte 2001, il sottogruppo caratterizzato dal livello più elevato dell'*ILR* è costituito dalla popolazione di genere femminile, proveniente dal Sud+Isole e in possesso di un diploma di maturità liceale ($ILR = 0,347$), mentre per il sottogruppo costituito dai maschi provenienti dalla Toscana e in possesso di un diploma di maturità diversa da quella liceale, il valore di tale indicatore è pari a 0,032.

Tabella 4 – $IU^{(y)}$ * delle coorti 2001 e 2002 per genere e per i possibili valori di y.

Indicatore di uscita senza titolo	Coorte 2001			Coorte 2002		
	Maschio	Femmina	Totale	Maschio	Femmina	Totale
$IU^{(1)}$ *	0,151	0,134	0,144	0,143	0,133	0,138
$IU^{(2)}$ *	0,391	0,336	0,366	0,345	0,262	0,306
$IU^{(3)}$ *	0,465	0,375	0,424	0,398	0,301	0,352
$IU^{(4)}$ *	0,509	0,386	0,453	N.D.	N.D.	N.D.

I risultati presenti in tabella 4 mostrano, coerentemente con quanto visto per gli indicatori di laurea, un considerevole miglioramento complessivo della *performance* nella coorte 2002 rispetto alla coorte 2001 (ovviamente il confronto è possibile solo per i primi tre anni di vita delle due coorti). Tuttavia, i valori ottenuti sono piuttosto elevati (anche se occorre sempre tener presente il problema della sovrastima delle uscite – vedi ancora nota 1) e richiamano l’attenzione sull’entità del fenomeno degli abbandoni (si pensi al fatto che il valore dell’indicatore di uscita senza titolo in 3 anni, per entrambe le coorti, è pari a oltre il doppio di quello di laurea in regola). Osservando i valori dell’indice per i vari periodi, si può notare come le uscite senza titolo avvengano prevalentemente nel periodo che intercorre tra il primo e il secondo istante di osservazione, che nel nostro caso corrisponde pressapoco al secondo anno accademico di vita della coorte, e poi tendano a ridursi progressivamente. Analizzando infine i risultati per genere, si nota che il fenomeno dell’uscita senza titolo, misurato in qualsiasi intervallo di tempo, è sensibilmente più diffuso tra i maschi che tra le femmine.

La tabella 5 mostra il medesimo indicatore calcolato sulle due coorti per le diverse aree di provenienza.

Tabella 5 - $IU^{(y)}$ * delle coorti 2001 e 2002 per provenienza e per i possibili valori di y.

Indicatore di uscita senza titolo	Coorte 2001			Coorte 2002		
	Toscana	Sud+Isole	Altro	Toscana	Sud+Isole	Altro
$IU^{(1)}$ *	0,141	0,132	0,197	0,129	0,143	0,169
$IU^{(2)}$ *	0,420	0,247	0,368	0,356	0,232	0,260
$IU^{(3)}$ *	0,486	0,295	0,408	0,404	0,261	0,338
$IU^{(4)}$ *	0,514	0,330	0,421	N.D.	N.D.	N.D.

E’ abbastanza evidente come il fenomeno dell’uscita senza titolo tenda progressivamente a diversificarsi per area di provenienza degli studenti al prolungarsi della vita della coorte; in particolare, si può notare che i valori riguardanti la Toscana e il Sud+Isole, abbastanza omogenei al termine del primo periodo, si sbilanciano in misura sempre maggiore, per entrambe le coorti, a scapito degli studenti toscani.

La tabella 6 mostra infine i risultati dell’analisi delle uscite senza titolo per tipo di diploma di maturità conseguito presso la scuola secondaria.

Tabella 6 – $IU^{(y)*}$ delle coorti 2001 e 2002 per tipo di diploma e per i possibili valori di y .

Indicatore di uscita senza titolo	Coorte 2001		Coorte 2002	
	Liceo	Non Liceo	Liceo	Non Liceo
$IU^{(1)*}$	0,175	0,123	0,108	0,155
$IU^{(2)*}$	0,319	0,397	0,237	0,344
$IU^{(3)*}$	0,362	0,465	0,263	0,401
$IU^{(4)*}$	0,384	0,499	N.D.	N.D.

Nel complesso, com'era logico attendersi, tra gli studenti ex-liceali l'uscita senza titolo è molto meno diffusa che tra i possessori di altri tipi di diploma di maturità. L'unico dato in controtendenza è rappresentato dall'indicatore di uscita al termine del primo periodo per la coorte 2001 (0,175 per gli studenti provenienti da un liceo, 0,123 per gli altri).

I risultati fin qui visti riguardano il calcolo di indicatori in forma piuttosto semplificata. In particolare, in questa applicazione si è sinora tralasciato la possibilità di condizionarsi ad eventi intermedi della carriera degli studenti, prevista, come evidenziato nella presentazione formale del modello, sia dagli indicatori di laurea, sia da quelli di uscita. E' tuttavia proprio in questa caratteristica che risiede gran parte della capacità informativa del modello proposto; il semplice calcolo degli indicatori "lordi" di laurea e di uscita, infatti, serve prevalentemente a scopi descrittivi e valutativi, in quanto fornisce misure complessive di *performance* di certe coorti d'interesse. Al contrario, il calcolo di indicatori condizionati apre prospettive ben più ampie di applicazione del modello (si pensi, a titolo di esempio, alla possibilità di misurare la probabilità di laurea o di uscita dato un certo livello di CFU conseguiti al termine di un dato periodo della carriera), che vanno dal controllo "in itinere" delle *performance* di coorte fino a forme di supporto individuale rivolto ai singoli studenti.

Riferimenti bibliografici

- Bartholomew D. (1982), *Stochastic models for social processes*, Wiley, Chichester
- Bernardi L., G. Lovison, U. Trivellato (1986), *Markovianità ed allargamento dello spazio degli stati in un modello stocastico del processo scolastico*, Atti della XXXIII Riunione Scientifica S. I. S., Bari, I, 341-349
- Bernardi L., U. Trivellato (1980), *Un modello markoviano del processo scolastico: (II) suo impiego per l'analisi della selettività del sistema preuniversitario italiano*, Rivista di Statistica Applicata, 13 (2), 55-90
- Ghellini G., F. Maccari (2007), *Inseguire gli studenti: un modello compatto per l'analisi delle carriere scolastiche di coorte*, in: Demografia e statistica: un ricordo di Enzo Lombardo tra scienza e cultura, Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale, Università degli Studi "La Sapienza" di Roma, in corso di pubblicazione
- Légaré J. (1972), *Methods for measuring school performance through cohort analysis*, Demography, 4, 617-624
- Maccari F. (2006), *Following students: a compact cohort-based model with longitudinal data for student careers and educational system analyses*, Tesi di Dottorato di Ricerca, Dipartimento di Statistica "G. Parenti", Firenze
- Quesnay F. (1758), *Tableau Économique des Physiocrates*, Calmann-Lévy, Paris, 1969 edition
- Quetelet A. (1849), *Letters on the theory of probabilities as applied to the moral and political sciences*, C. & Layton, London
- Stone R. (1965), *A model of educational system*, Minerva, 3, 177-186
- Stone R. (1972), *A markovian education model and other examples linking social behavior to the economy*, Journal of the Royal Statistical Society, series A, 135, 511-543
- Thonstad T. (1969), *Education and Manpower: theoretical models and empirical applications*, Oliver & Boyd, Ltd., Edinburgh
- Trivellato U. (1980), *Un modello markoviano del processo scolastico: (I) specificazione e procedimento di stima*, Rivista di Statistica Applicata, 1, 3-20