Università degli Studi di Pisa
Dipartimento di Statistica e Matematica
Applicata all'Economia

Report n. 231

I tempi di laurea presso l'Università di Pisa: un'applicazione dei modelli di durata in tempo discreto

Andrea Mercatanti

Pisa, Ottobre 2002

- Stampato in Proprio –

Via Cosimo Ridolfi, 10 - 56124 PISA - Tel. Segr. Amm. 050 945231 Segr. Stud. 050 945317 Fax 050 945375
Cod. Fisc. 80003670504 - P. IVA 00286820501 - Web http://statmat.ec.unipi.it/ - E-mail: dipstat@ec.unipi.it

I tempi di laurea presso l'Università di Pisa: un'applicazione dei modelli di durata in tempo discreto

Andrea Mercatanti
Dipartimento di Statistica e Matematica Applicata all'Economia
Università di Pisa

1 Introduzione

L'oggetto del presente elaborato è un'analisi statistica dei tempi necessari al conseguimento del diploma di laurea riguardante un'insieme di studenti dell'Università di Pisa.

La questione dei tempi di laurea risulta particolarmente importante non soltanto a livello locale ma anche se si considera il sistema universitario Italiano nel suo complesso. La lentezza delle carriere degli studenti costituisce infatti uno dei principali aspetti caratterizzanti in negativo il sistema universitario del nostro paese (Capano, 2000; Luzzato, 2000). L'estrema importanza della questione è evidente, basti solo pensare allo spreco di risorse umane provocato da una tardiva entrata di individui giovani nel mondo produttivo. Ovviamente le responsabilità non vanno indirizzate totalmente a carico degli individui, anzi un'utile obiettivo può essere proprio quello di scindere le responsabilità di carattere personale da quelle dovute ad inefficienze del sistema formativo o ad altre cause di carattere economico e sociale. A questi fini emerge l'importanza di studi di tipo quantitativo in quanto costituiscono un mezzo efficace per riuscire sia ad identificare i fattori che contribuiscono ad un allungamento del tempo di laurea che a quantificarne i contributi.

Il presente lavoro risulta essenzialmente finalizzato alla identificazione delle principali variabili che contribuiscono significativamente alla spiegazione dei tempi di laurea presso l'Università di Pisa. Il tempo necessario al conseguimento della laurea può infatti essere considerato come una variabile risultato correlata ad una serie di variabili esplicative di varia natura (ad esempio: il sesso, la Facoltà di appartenenza, o il voto conseguito alla maturità). Lo scopo è di quantificare in senso probabilistico il contributo apportato da ogni variabile esplicativa al tempo di laurea. Si intende cioè arrivare a valutare di quanto varia la probabilità di laurearsi in un certo numero di anni al variare della modalità assunta da una certa variabile esplicativa, fermo restando il valore assunto da tutte le altre variabili. In tal modo risulta possibile, ad esempio, la risposta ad una domanda di questo tipo: "per uno studente iscritto alla Facoltà di Lettere e Filosofia, residente a Pisa, e proveniente dal Liceo Scientifico, di quanto varia la probabilità di laurearsi in sette anni a seconda del sesso?".

Il lavoro si presenta come uno studio di tipo longitudinale. Si tratta infatti di prendere in considerazione un insieme di individui, e di studiare un'esito comune (il conseguimento del diploma di laurea) nella sua dinamica temporale. Il lavoro si caratterizza ulteriormente per rientrare nella categoria degli "studi di sopravvivenza", essendo interessati all'analisi di un solo esito non ricorrente. La metodologia statistica utilizzata a questi fini è un modello per dati di durata in tempo discreto. Si tratta di una metodologia basata su di una modellizzazione lineare logistica della funzione di rischio (Cox, 1972) le cui caratteristiche verranno illustrate nella prossima sezione.

La popolazione oggetto della presente analisi è costituita dagli studenti immatricolati nell'anno accademico 1990/91 presso una qualsiasi delle Facoltà appartenenti all'Università di Pisa, e che si sono laureati entro 10 anni dall'immatricolazione (cioè entro il 31/12/2000). La numerosità risulta essere di 2367 unità statistiche. La stessa popolazione è già stata oggetto di un precedente lavoro di tipo descrittivo (Mercatanti, 2002) al quale si può far riferimento per una descrizione esauriente delle variabili. Il dataset dal quale è stata estratta la popolazione di riferimento è stato fornito dall'Ufficio Statistico dell'Università di Pisa. Interessanti studi di tipo longitudinale sullo stesso argomento e sempre a livello locale toscano, sono rintracciabili in: Rampichini, 1995; AA.VV., 2002.

La prossima sezione illustra la metodologia utilizzata, segue l'applicazione e l'esposizione dei risultati nella sezione 3, per poi riportare le conclusioni nell'ultima sezione.

2 Un modello di durata in tempo discreto

In questa sezione viene sinteticamente illustrato il modello statistico che sarà poi impiegato per l'analisi empirica oggetto del lavoro.

I dati derivanti dall'osservazione dei tempi nei quali si verificano determinati eventi prendono generalmente il nome di *dati di durata*; nel nostro caso l'evento che si intende studiare è rappresentato dall'ottenimento del diploma di laurea. Gli eventi oggetto di uno studio di durata si classificano poi a seconda del numero di volte che possono essere sperimentati. Si parla infatti di eventi *ricorrenti* quando possono essere sperimentati più volte, e di eventi *non ricorrenti* quando possono essere sperimentati una volta al massimo. Se l'analisi di durata è caratterizzata da un solo tipo di evento non ricorrente viene detta *analisi di sopravvivenza*. Questo è il caso dello studio oggetto del paper, si studiano infatti individui per i quali la laurea rappresenta un evento unico e non ricorrente. Il seguente schema (Grilli, 1996) sintetizza la precedente classificazione:

$$\text{Dati di durata} \begin{cases} \text{un tipo di evento} \begin{cases} \text{non ricorrente } (\textit{dati di sopravvivenza}) \\ \text{ricorrente} \end{cases} \\ \text{più tipi di evento} \end{cases}$$

Nelle analisi di sopravvivenza il tempo nel quale un evento si verifica è rappresentato da una variabile aleatoria non negativa T, che viene analizzata per mezzo di una funzione basilare: la funzione di *rischio*. Nel caso di una variabile aleatoria T discreta, la funzione di rischio è così definita:

$$h(t) = P(T = t | T \geq t);$$

indica quindi la probabilità di sperimentare l'evento al tempo t, dato che non è stato sperimentato in nessuno dei tempi precedenti a t. Dalla funzione di rischio è possibile ottenere tutte le altre probabilità di sperimentare l'evento. Ad esempio la probabilità di sperimentare l'evento in un tempo successivo a t è chiamata *funzione di sopravvivenza* ed è così definita:

$$S(t) = P(T > t) = \prod_{s=1}^{t} [1 - h(s)].$$

Oppure la probabilità che l'unità statistica sperimenti l'evento in un certo tempo t, è calcolabile come:

$$P(T = t) = h(t)\, S(t-1) = h(t) \prod_{s=1}^{t-1} [1 - h(s)].$$

La $P(T = t)$ differisce da $h(t)$ poichè è una probabilità non condizionata alla sopravvivenza al tempo precedente a t. Una volta introdotta la funzione di rischio, è possibile definire la verosimiglianza di un campione di numerosità N. Nel caso le osservazioni siano i.i.d., la verosimiglianza risulta:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \left\{ [P(T = t_i)]^{\delta_i} \, [P(T > t_i)]^{1-\delta_i} \right\} =$$

$$= \prod_{i=1}^{N} \left\{ \left[h(t_i) \prod_{s=1}^{t_i - 1} [1 - h(s)] \right]^{\delta_i} \left[\prod_{s=1}^{t} [1 - h(s)] \right]^{1-\delta_i} \right\},$$

dove

$$\delta_i = \begin{cases} 1 \text{ se l'unità } i \text{ sperimenta l'evento al tempo } t_i \\ 0 \text{ se l'unità } i \text{ è } \textit{censurata} \text{ al tempo } t_i. \end{cases}$$

e dove abbiamo indicato con t_i l'ultimo periodo osservato per l'i-esima unità. L'i-esima unità si dice *censurata* al tempo t_i, se non ha ancora sperimentato l'evento a quel tempo.

Una modellizzazione statistica che ben si adatta alla funzione di rischio è stata proposta da Cox (1972). Si tratta sostanzialmente di un modello lineare logistico per la probabilità $h_i(t)$:

$$\log \left\{ \frac{h_i(t)}{1 - h_i(t)} \right\} = \mathbf{d}_t \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta}, \tag{1}$$

dove:

- $\mathbf{d}_t$ è un vettore riga $(d_{t1}, d_{t2}, ..., d_{tT})$ di variabili indicatrici; indicando, come in precedenza, con t_i l'ultimo periodo osservato per l'i-esima unità, allora dalla successione $\{t_1, .., t_i, .., t_N\}$ si definisce $T = \sup \{t_i\}$. Le variabili indicatrici sono allora definite come:

  $$d_{tl} = \begin{cases} 1 \text{ se } l = t \\ 0 \text{ se } l \neq t \end{cases}$$

  per $l = 1, 2, ..., T$;

- $\mathbf{x}_{it}$ è un vettore riga di p variabili esplicative per l'i-esima unità al tempo t;
- $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$ sono i vettori colonna dei coefficienti del modello, di dimensioni rispettivamente $(T \times 1)$ e $(p \times 1)$. In particolare, i T elementi del vettore $\boldsymbol{\alpha}$ rappresentano i valori del rischio di sperimentare l'evento a ciascun tempo per l'*individuo tipo*[1] (cosiddetti rischi di base), mentre i p elementi del vettore $\boldsymbol{\beta}$ rappresentano i contributi di ogni covariata al logit della funzione di rischio.

Data questa definizione per il vettore $\boldsymbol{\alpha}$, la precedente funzione (1) si può riscrivere in maniera più semplice:

$$\log\left\{\frac{h_i(t)}{1-h_i(t)}\right\} = \alpha_t + \mathbf{x}_{it}\boldsymbol{\beta}.$$

In presenza di covariate tutte indipendenti dal tempo, il modello si caratterizza per soddisfare l'ipotesi di proporzionalità tra odds. Ciò significa che presi due individui con valori delle variabili esplicative $\mathbf{x}_i$ e $\mathbf{x}_{i+1}$, allora il rapporto tra odds ad ogni tempo t risulta essere indipendente dal tempo, in quanto:

$$\Psi = \frac{h_i(t)/([1-h_i(t)])}{h_{i+1}(t)/\left[1-h_{i+1}(t)\right]} = \exp\left[(\mathbf{x}_i - \mathbf{x}_{i+1})\boldsymbol{\beta}\right].$$

Oltre alle ipotesi di linearità ed eventualmente di proporzionalità tra odds, il modello si caratterizza inoltre per l'assenza di un termine di errore, il che presuppone l'assenza di eterogeneità non osservabile. Ciò significa che l'eterogeneità tra le unità statistiche è spiegata soltanto da variazioni nei valori delle covariate.

Come già espresso i coefficienti presenti nel vettore $\boldsymbol{\beta}$ quantificano i contributi di ogni covariata al logit della funzione di rischio. Questi risultano però essere di non facile interpretabilità. Meglio far riferimento ai più intuibili e facilmente interpretabili effetti marginali delle covariate sul rischio, cioè ai contributi delle covariate alla probabilità (condizionata) di sperimentare l'evento. Tali effetti marginali possono essere così calcolati:

[1]Per *individuo tipo* si intende l'unità statistica caratterizzata dal fatto di presentare valore nullo in ogni covariata.

$$\frac{\partial\, h(t)}{\partial\, x_j} = \beta_j\, h(t)\; [1 - h(t)]$$

nel caso di una covariata x_j continua, oppure

$$h(t|x_j = 1) - h(t|x_j = 0)$$

nel caso di una covariata x_j binaria.

Per le stime dei coefficienti presenti nei due vettori parametrici del modello, $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$, è consigliabile l'utilizzazione del metodo della massima verosimiglianza. Infatti per il modello in tempo discreto proposto in questa sezione, le stime di massima verosimiglianza sono calcolabili utilizzando le routine di calcolo per modelli logit, usualmente presenti nei pacchetti statistici, a condizione di apportare opportune modifiche al dataset (Allison, 1982; Singer e Willet, 1993; Yamaguchi, 1991).

3 Il caso di studio: l'analisi dei tempi di laurea presso l'Ateneo Pisano

Questa sezione riporta i risultati dell'applicazione del modello di durata in tempo discreto (proposto in sezione 2) alla popolazione oggetto di studio. Tale popolazione è costituita dai 2367 studenti immatricolati nell'anno accademico 1990/91 al primo anno di un qualsiasi corso di laurea presso l'Università di Pisa, e che si sono laureati entro 10 anni dall'immatricolazione risultando stabili sia per corso di laurea che per sede universitaria. Ulteriore condizione è data dal fatto che al momento dell'immatricolazione gli studenti siano stati in possesso al massimo di un diploma di scuola media superiore. Per un'analisi delle principali caratteristiche descrittive di questa popolazione si può far riferimento ad un precedente lavoro (Mercatanti, 2002).

Il modello adottato è quello proposto nella sezione precedente che viene qui brevemente richiamato. Ricordiamo trattasi di una modellizzazione logistica lineare della funzione di rischio:

$$\log\left\{\frac{h_i(t)}{1 - h_i(t)}\right\} = \mathbf{d}_t\boldsymbol{\alpha} + \mathbf{x}_{it}\boldsymbol{\beta}. \tag{2}$$

In questa applicazione la variabile aleatoria tempo, T, che viene definita come "numero di anni accademici a partire dall'anno di immatricolazione", assume valori: $t = 4, 5, 6, 7, 8, 9, 10$. Di conseguenza il vettore delle variabili indicatrici del tempo, $\mathbf{d}_t$, è di dimensione (6×1), essendo a rischio di laurea soltanto gli anni dal quarto al nono dal momento dell'immatricolazione. Si rammenta a proposito che tutti gli studenti presenti nel dataset si sono laureati entro dieci anni e quindi il decimo anno non si può considerare a rischio di laurea. Nell'analisi si fa inoltre riferimento al seguente individuo tipo:

- immatricolato nel medesimo anno nel quale ha conseguito il Diploma di maturità;
- con voto di maturità posto uguale al voto di maturità medio per la popolazione in esame;
- di sesso maschile;
- residente a Pisa o provincia;
- proveniente dal Liceo Scientifico.

Il vettore delle covariate $\mathbf{x}_{it}$ è di dimensione (10×1) ed è costituito dalle variabili:

- anni di ritardo nell'immatricolazione rispetto all'anno nel quale è stato conseguito il Diploma di maturità;
- voto di maturità;
- una dummy per il sesso: femmina;
- due dummy per la zona geografica di residenza: province toscane (esclusa la provincia di Pisa), altre province italiane;
- cinque dummy per la Scuola Media Superiore di provenienza: Liceo Classico, Ist. Magistrale, Ist. Tecnico, Ist. Professionale, Altre scuole.

L'utilizzazione del modello di durata proposto è essenzialmente giustificata dalla natura discreta della variabile aleatoria tempo.

Una prima analisi riguarda la verifica dell'ipotesi di proporzionalità tra odds. Rispetto a quello proposto nella sezione precedente, un modello nel quale si rilasci l'ipotesi di proporzionalità prevede la presenza di termini di interazione tra le covariate e le variabili indicatrici del tempo. Un modello del genere produrrà sicuramente una migliore bontà di adattamento ai dati. L'obiettivo è allora verificare quali siano i termini di interazione significativi e quindi da introdurre nell'analisi ai fini di un migliore bontà di adattamento del modello. Un primo passo è costituito dalla verifica dell'ipotesi nulla di proporzionalità tra odds, per confronto tra le massime verosimiglianze ottenibili sui modelli senza e con i termini di interazione. L'usuale statistica Λ del test del rapporto di verosimiglianza assume il valore: 118.15, con 48 gradi di libertà. L'ipotesi nulla è quindi ampiamente rifiutata. Non tutti i singoli termini di interazione assumono però valore significativo; si sono considerati allora nell'analisi soltanto quelli significativi avendo avuto cura di sottoporre a test l'ipotesi che i restanti termini siano nulli.

I principali risultati dell'analisi vengono riportati nelle tabelle 1-5. Le stime dei coefficienti del modello sono state calcolate con il metodo della massima verosimiglianza; le deviazioni standard degli effetti marginali con il metodo Delta (Greene, 1993). Più in dettaglio vengono illustrati nelle tabelle: i rischi di sperimentare l'evento di interesse, laurea, nei vari tempi per l'intero Ateneo e per alcune Facoltà (Tab. 1); le probabilità di laurea nei vari tempi per l'intero Ateneo e per alcune Facoltà (in Tab. 2, e riportate graficamente nelle Fig. 1-3); gli effetti marginali delle covariate sul rischio di laurea al sesto, settimo ed ottavo anno per l'intero Ateneo (Tab. 3); gli effetti marginali delle covariate sul rischio di laurea al sesto anno per alcune Facoltà (Tab. 4 e 5).

Le principali considerazioni emergenti dall'analisi dei risultati possono essere così sintetizzate:

- le tabelle 1-2 e le figure 2-3 mettono in evidenza un diverso andamento dei rischi e delle probabilità di laurea per le diverse Facoltà; in particolare si notano tempi mediamente più lunghi per Ingegneria e Giurisprudenza, al contrario tempi mediamente più brevi per Scienze Politiche e Lettere.

- il valore modale della probabilità di laurea sia per quanto riguarda l'intero Ateneo che per le singole Facoltà è di sette anni (Tab. 2). Ben superiore quindi alle durate legali che per l'anno accademico di immatricolazione 1990/91 erano di quattro, cinque, oppure (soltanto per Medicina) sei anni;

- si osserva un contributo sempre significativo e positivo del voto alla Maturità sui rischi di laurea sia per l'intero Ateneo che per le singole Facoltà. In ogni caso infatti un buon voto alla Maturità contribuisce alla diminuzione del tempo di laurea;
- si osserva un contributo significativo e generalmente negativo della provenienza da un comune non-toscano sul rischio di laurea (quindi una tendenza verso un aumento del tempo di laurea);
- si osservano contributi significativi (anche se di direzioni opposte) delle scuole di provenienza sui rischi di laurea. In particolare si osserva come rispetto alla scuola di riferimento (Liceo Scientifico) la provenienza dal Liceo Classico contribuisca in generale ed in particolare per la Facoltà di Lettere ad abbreviare la durata degli studi; una considerazione opposta vale invece per le provenienze dagli Istituti Tecnici e Professionali per le quali in generale si osserva una maggiore durata.

E' utile osservare che nelle tabelle 1 e 2 i rischi e le probabilità di laurea si riferiscono all'individuo tipo. Il modello consente però anche il calcolo di rischi e probabilità per tutte le tipologie di studenti. A titolo di esempio nella figura 4 vengono riportate le probabilità di laurea relative a due studenti molto diversi. Si confrontano infatti uno studente con voto di maturità di 58/60, di sesso femminile, proveniente dal Liceo Classico, (individuo A), con uno studente con voto di maturità di 38/60, di sesso maschile, proveniente da un Istituto Professionale, residente fuori dalla Toscana, e immatricolato con due anni di ritardo rispetto al conseguimento della maturità (individuo B). Si osserva come le due curve abbiano andamenti molto diversi. L'individuo A assume caratteristiche che contribuiscono ad un'accelerazione del tempo di laurea, il contrario per l'ind. B. E infatti la probabilità di laurearsi in 6 anni risulta molto più alta per l'individuo A: il 31.85% contro il 6.27% dell'individuo B. Col proseguire del tempo le due curve si intersecano e a 10 anni dall'immatricolazione si ha una probabilità di laurea praticamente nulla per l'individuo A (0.62%) contro una probabilità del 35.49% per l'individuo B.

4 Conclusioni

L'applicazione di un modello di durata in tempo discreto ha consentito un'adeguata analisi dei tempi di laurea presso l'Università di Pisa. Ricordiamo che la popolazione di riferimento per lo studio è costituita dall'insieme degli studenti immatricolati nell'anno accademico 1990/91 presso una qualsiasi delle facoltà appartenenti all'Università di Pisa, e che si sono laureati entro 10 anni dall'immatricolazione.

E' stato utilizzato un modello per dati di durata in tempo discreto basato su di una modellizzazione lineare logistica della funzione di rischio. Al fine dell'ottenimento di un'adeguata bontà di adattamento sono stati tenuti in considerazione i termini di interazione tra le covariate e le variabili indicatrici dei tempi di laurea.

Anche per l'Università di Pisa i risultati dell'analisi quantitativa confermano alcune caratteristiche del sistema universitario a livello nazionale. In particolare si sono osservati tempi di laurea in media molto lunghi; tempi mediamente minori per gli studenti provenienti dai Licei e/o con buoni voti di maturità e/o iscritti a Scienze Politiche o Lettere; tempi mediamente più lunghi per gli studenti iscritti a Ingegneria o Giurisprudenza.

Il modello utilizzato è adatto anche per l'analisi di altre interessanti problematiche, ad esempio gli abbandoni ai primi anni di Università. Data l'importanza sociale di queste studi, è auspicabile quindi un'approfondimento dell'analisi delle carriere presso l'Università di Pisa basato su di una maggiore copertura informativa sia quantitava che qualitativa.

References

[1] AA.VV. (2002); *Il sistema universitario, l'istruzione post-diploma in Toscana;* IRPET e Regione Toscana; Giunti Editore.

[2] Allison P. (1982); *Event history analysis : regression for longitudinal event data;* Sage Publ.

[3] Capano G. (2000); *L'Università in Italia;* Il Mulino, Bologna.

[4] Cox D. (1972); *Regression models and life tables*; Journal of the Royal Statistical Society, Series B, 34, 187-220.

[5] Greene W.H. (1993); *Econometric analisys*; Macmillan Publishing Company, New York.

[6] Grilli L. (1996); L'approccio dei processi di conteggio ai modelli di regressione per dati di durata; Working paper n. 69, Dip. di Statistica "G.Parenti", Università di Firenze.

[7] Luzzato G. (2000); *L'Università del 2000;* La Nuova Italia.

[8] Mercatanti A. (2002); *Un'analisi descrittiva dei laureati dell'Università di Pisa*; Report n.221 del Dipartimento di Statistica e Matematica applicata all'Economia, Università di Pisa.

[9] Rampichini C. (1995); *Modelli di durata a tempo discreto e rischi concorrenti. Un'applicazione all'analisi delle carriere degli studenti universitari*; Tesi di Dottorato, Dip. di Statistica "G.Parenti", Università di Firenze.

[10] Rampichini C. (1997); *Problemi e metodi di analisi della durata degli studi universitari;* in: "La popolazione studentesca e le Università italiane: indagini, modelli e risultati"; CLEUP.

[11] Singer J., J. Willet (1993); *It's about time: using discrete time survival analysis to study duration and the timing of events*; The journal of educational statistics.

[12] Windmeijer F. (1995); *Goodness of fit measures in binary choice models;* Econometric Reviews, 14, 101-116.

[13] Yamaguchi K. (1991); *Event history analysis*; Applied social research methods series, 28, Newbury Park, California, Sage Pubblications inc.

Tab. 1: rischi di laurea, P(T=t| T>=t) per l'intero Ateneo e per alcune Facoltà, individuo tipo

t	Ateneo	Economia	Ingegn.	Giurisp.	Lettere	Sc. M.F.N.	Sc. Polit.
4	0.0193	0.0414	0.0000	0.0082	0.0487	0.0210	0.0454
5	0.0917	0.1407	0.0080	0.0649	0.1694	0.1126	0.1828
6	0.2344	0.2892	0.1619	0.1529	0.3933	0.2808	0.3885
7	0.4444	0.5091	0.3478	0.4424	0.7074	0.4670	0.7819
8	0.4412	0.5092	0.4817	0.4159	0.5720	0.4608	0.6513
9	0.7901	0.8440	0.7536	0.7947	0.9259	0.8215	1.0000

Tab. 2: probabilità di laurea, P(T=t) per l'intero Ateneo e per alcune Facoltà, individuo tipo

t	Ateneo	Economia	Ingegn.	Giurisp.	Lettere	Sc. M.F.N.	Sc. Polit.
4	0.0193	0.0414	0.0000	0.0082	0.0487	0.0210	0.0454
5	0.0899	0.1348	0.0080	0.0643	0.1611	0.1102	0.1745
6	0.2087	0.2382	0.1606	0.1418	0.3107	0.2439	0.3030
7	0.3030	0.2980	0.2891	0.3475	0.3391	0.2917	0.3729
8	0.1671	0.1463	0.2611	0.1821	0.0802	0.1534	0.0677
9	0.1672	0.1190	0.2117	0.2033	0.0555	0.1475	0.0362
10	0.0444	0.0220	0.0692	0.0525	0.0044	0.0320	0.0000

Tab. 3: effetti marginali delle covariate sul rischio di laurea al sesto, settimo e ottavo anno

anni	sesto	settimo	ottavo
Anni di ritardo nell'immatric.	0.0089 (0.008)	0.0123 (0.011)	0.0123 (0.011)
Voto di maturità	0.0089 (0.000)	0.0123 (0.001)	0.0123 (0.011)
Femmina	0.0233 (0.011)	0.0322 (0.016)	0.0322 (0.016)
Extra-Toscana	-0.0289 (0.011)	-0.0400 (0.015)	-0.0399 (0.015)
Toscana	0.0056 (0.006)	0.0078 (0.008)	0.0078 (0.008)
Liceo Classico	0.0322 (0.015)	0.0446 (0.022)	0.0445 (0.022)
Ist. Magistrale	-0.0387 (0.026)	-0.0535 (0.036)	-0.0534 (0.036)
Ist. Tecnico	-0.0444 (0.013)	-0.0614 (0.017)	-0.0613 (0.017)
Ist. Professionale	-0.1658 (0.053)	-0.2291 (0.072)	-0.2288 (0.072)
Altre scuole	-0.0376 (0.022)	-0.0520 (0.030)	-0.0519 (0.030)

Tab. 4: effetti marginali delle covariate sul rischio di laurea al sesto anno, per alcune Facoltà

	Economia	Ingegneria	Legge
Anni di ritardo nell'immatric.	-0.0164 (0.024)	0.0162 (0.024)	0.0067 (0.007)
Voto di maturità	0.0111 (0.002)	0.0062 (0.002)	0.0063 (0.001)
Femmina	-0.0383 (0.026)	-0.0301 (0.031)	0.0283 (0.019)
Extra-Toscana	-0.0281 (0.035)	-0.0628 (0.031)	0.0005 (0.014)
Toscana	0.0440 (0.028)	-0.0147 (0.029)	-0.0064 (0.020)
Liceo Classico	0.0942 (0.062)	0.0113 (0.049)	0.0089 (0.022)
Ist. Magistrale	-0.0313 (0.126)	-	-0.1455 (0.047)
Ist. Tecnico	-0.0761 (0.030)	-0.0164 (0.026)	-0.0773 (0.032)
Ist. Professionale	-0.1670 (0.113)	-	-0.2553 (0.101)
Altre scuole	-0.2295 (0.109)	-0.0818 (0.063)	-0.0998 (0.060)

Tab. 5: effetti marginali delle covariate sul rischio di laurea al sesto anno, per alcune Facoltà

	Lettere	Scienze	Sc. Politiche
Anni di ritardo nell'immatric.	0.0080 (0.019)	-0.0212 (0.020)	0.0368 (0.030)
Voto di maturità	0.0158 (0.002)	0.0115 (0.002)	0.0119 (0.005)
Femmina	-0.0808 (0.052)	-0.0032 (0.020)	-0.0033 (0.037)
Extra-Toscana	-0.0201 (0.048)	0.0113 (0.035)	-0.1138 (0.119)
Toscana	0.0702 (0.040)	0.0157 (0.033)	-0.0050 (0.088)
Liceo Classico	0.0842 (0.046)	-0.1300 (0.057)	-0.1109 (0.112)
Ist. Magistrale	-0.0830 (0.060)	0.0936 (0.101)	-0.2725 (0.162)
Ist. Tecnico	-0.0930 (0.069)	-0.099 (0.033)	-0.1943 (0.101)
Ist. Professionale	-0.4352 (0.162)	-0.0133 (0.173)	-0.1954 (0.246)
Altre scuole	-0.1194 (0.052)	0.0268 (0.108)	-0.2431 (0.115)

Fig.1

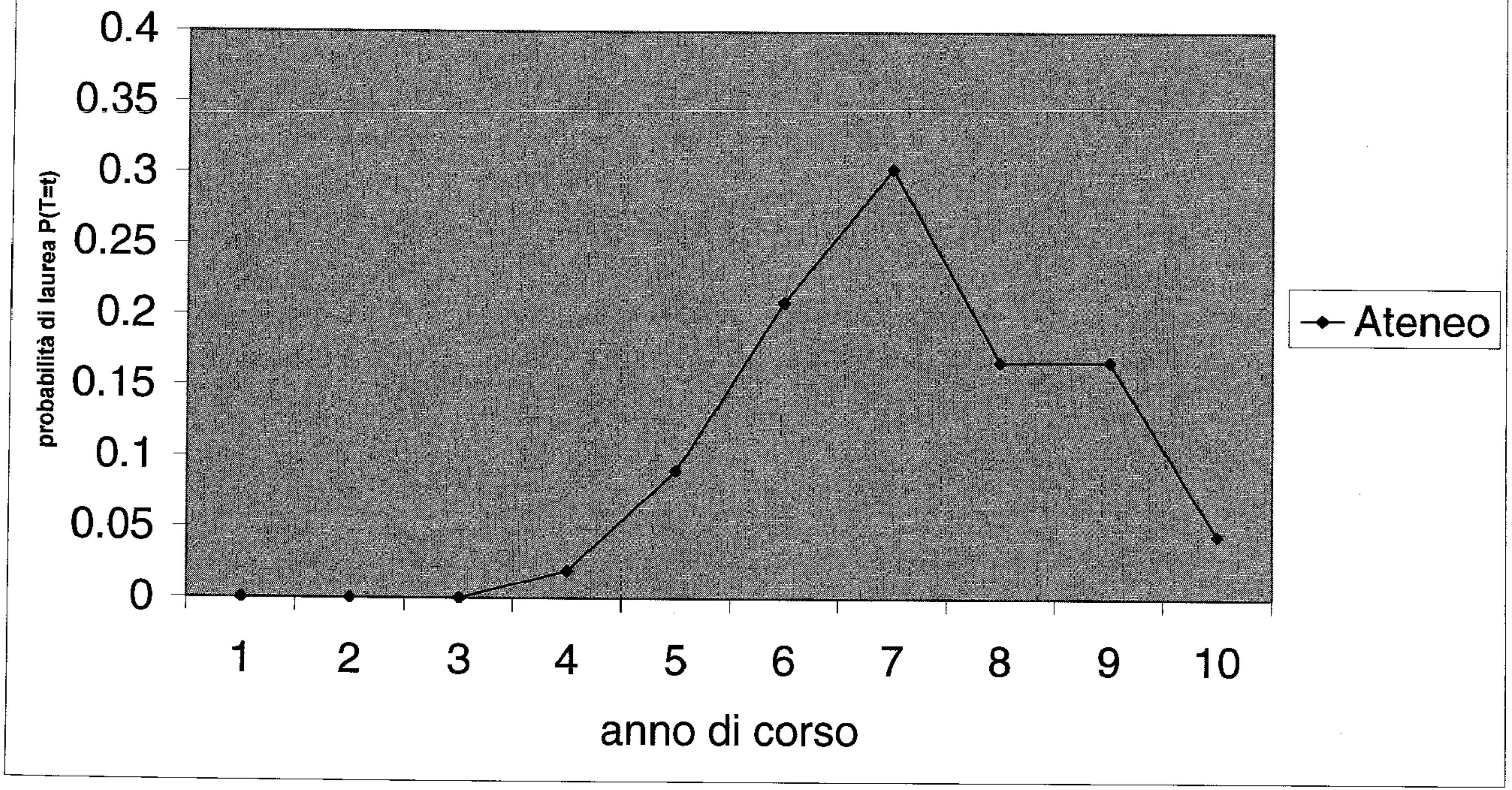

0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
probabilità di laurea P(T=t)
1
2
3
4
5
6
7
8
9
10
anno di corso
Ateneo

Fig.2

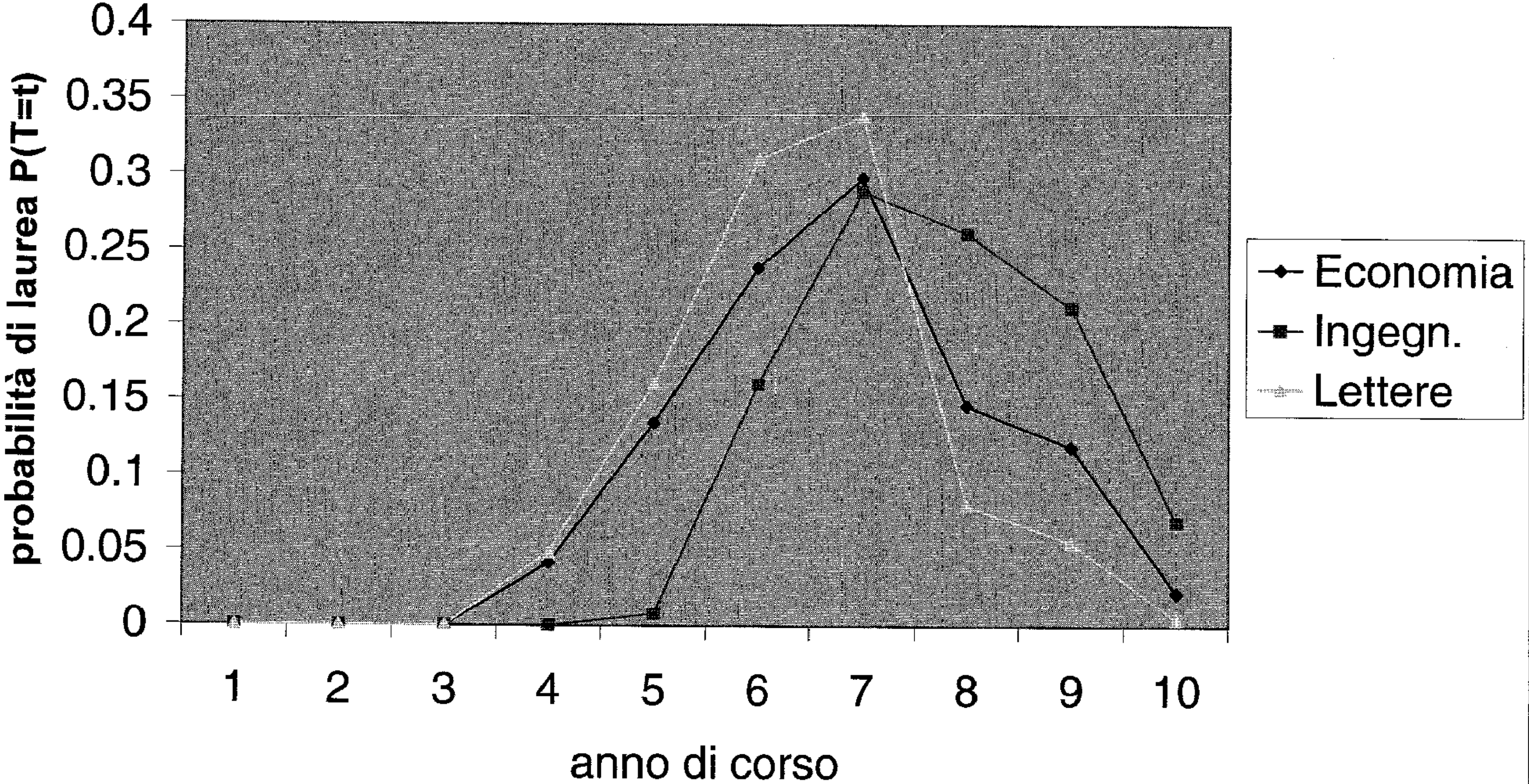

probabilità di laurea P(T=t)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10
anno di corso
Economia
Ingegn.
Lettere

Fig.3

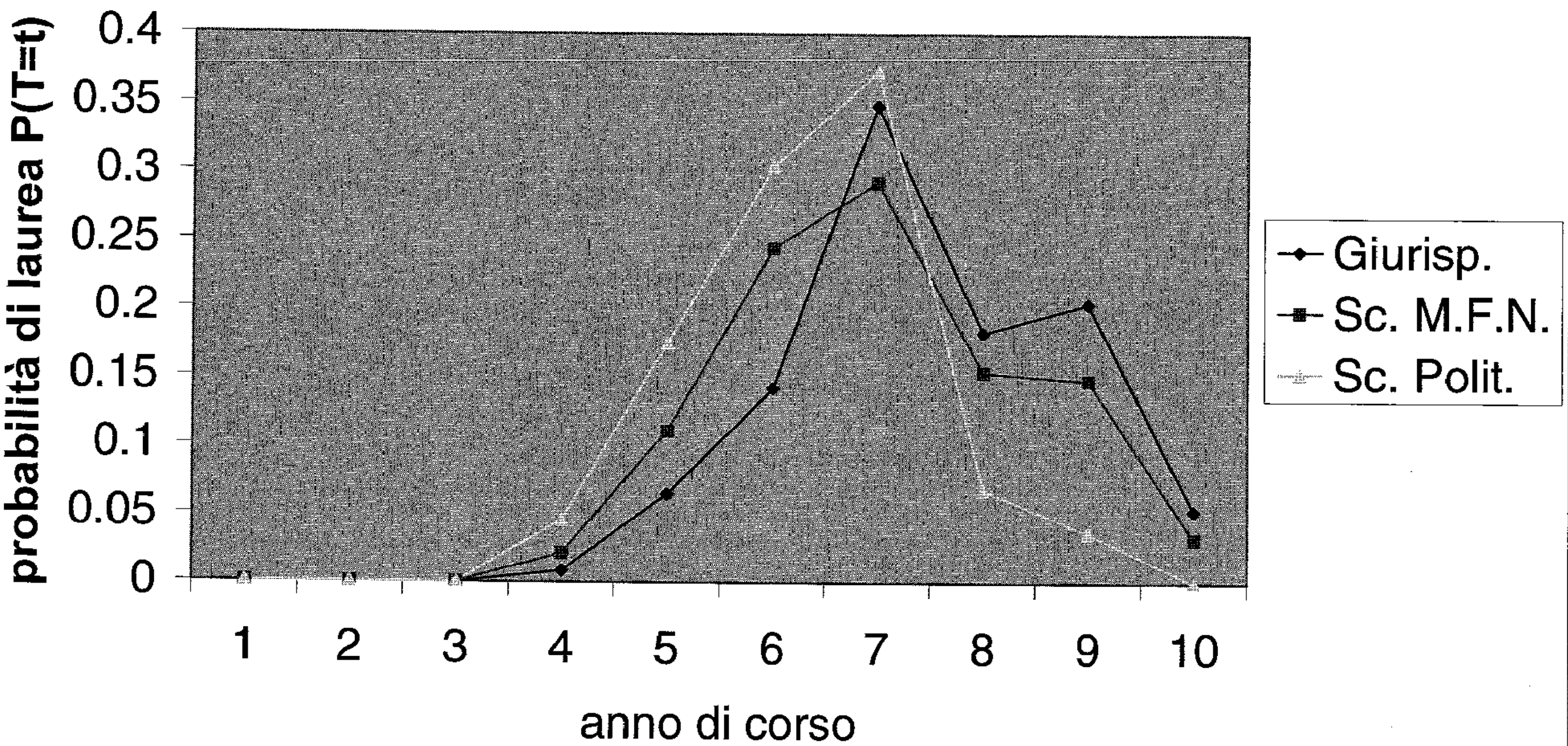
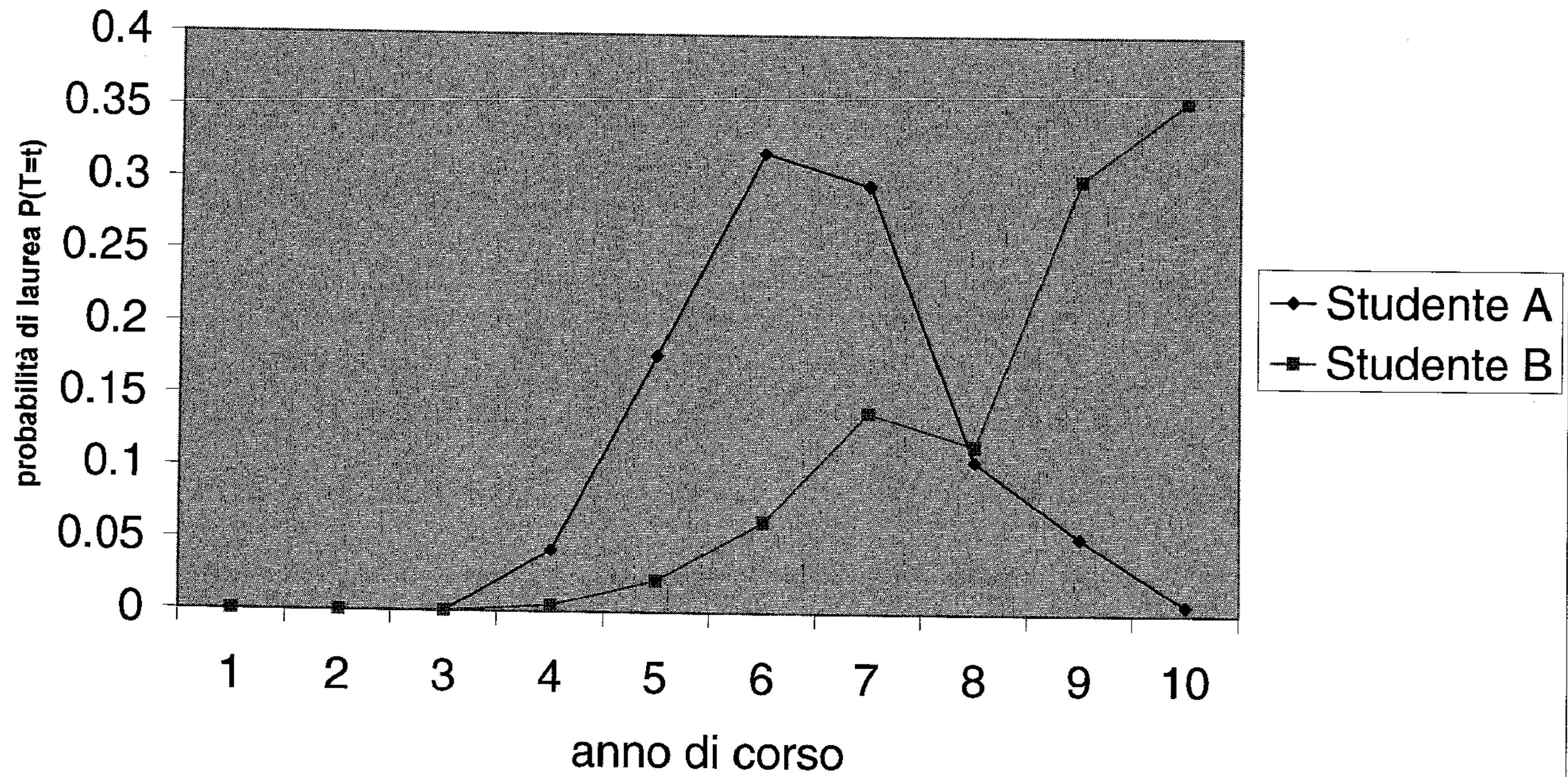

probabilità di laurea P(T=t)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10
anno di corso
Giurisp.
Sc. M.F.N.
Sc. Polit.

Fig.4


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
probabilità di laurea P(T=t)
1
2
3
4
5
6
7
8
9
10
anno di corso
Studente A
Studente B