CALCOLO DELLE PROBABILITÀ E STATISTICA

percorsi di insegnamento nel triennio delle superiori

*di Marco Bramanti**

Questo contributo nasce dall'esperienza di un corso di formazione di calcolo delle probabilità e statistica, rivolto a insegnanti di scuole superiori, che l'autore ha tenuto negli anni 2007 e 2008 presso il Laboratorio Didattico "effediesse" del Dipartimento di Matematica del Politecnico di Milano.

Dal momento che il calcolo delle probabilità e la statistica sono discipline che non fanno parte del curriculum di studi di molti insegnanti laureati in matematica, né dei tradizionali programmi di matematica delle scuole, l'indicazione, contenuta ormai da anni in diversi programmi scolastici, di inserire queste materie nel proprio insegnamento può porre evidentemente vari problemi agli insegnanti.

Al di là dell'ovvia necessità degli insegnanti di colmare eventuali lacune nella propria formazione di base, studiando alcuni elementi di queste discipline, esiste un problema non banale di scelta di che cosa insegnare, esattamente: quali argomenti, con quale taglio, in quale successione, in quali classi, con quali obiettivi e con quali collegamenti alle altre materie.

Queste domande si pongono perché, realisticamente, contenuti e obiettivi indicati dai programmi ufficiali appaiono spesso sovradimensionati rispetto alle concrete possibilità degli insegnanti, in termini sia di tempo a disposizione che di recettività della classe: occorre dunque scegliere. Inoltre, a differenza di altre materie matematiche tradizionali, come l'analisi matematica, calcolo delle probabilità e statistica si sviluppano per loro natura con un percorso meno lineare, che offre fin dall'inizio varie biforcazioni del discorso e possibili alternative didattiche.

Non c'è quindi una definizione univoca di quali siano «le prime cose» da insegnare, quando non si può sviluppare un percorso ampio. Si tratta perciò di scelte non facili. Infine, ci sono alcuni nodi concettuali insiti in queste discipline che pongono specifici problemi didattici, per l'insegnamento nelle scuole.

Mi riferisco in particolare all'utilizzo del calcolo integrale per la trattazione della probabilità nel continuo, e alla delicata «partenza» del discorso sulla

*Professore di Analisi Matematica presso il Politecnico di Milano.

probabilità, con i sottili problemi legati alla sua definizione.
Nel seguito si descrivono sinteticamente alcuni «moduli», sottolineando per ciascuno di essi diversi aspetti significativi e indicando quindi un percorso didattico essenziale.
Concretamente, si suggerisce, esplicitamente o implicitamente, una gerarchia di importanza dei vari aspetti di un argomento, in vista di certi obiettivi, che aiuti anche a scegliere «cosa fare e cosa saltare».
Quindi si presentano alcuni schemi, tra loro alternativi, di percorsi complessivi con cui presentare, nell'arco degli ultimi tre anni di scuola, con vari gradi di approfondimento, parte di queste unità didattiche. Ciascun percorso ha le proprie motivazioni, le proprie sottolineature (e necessariamente le proprie omissioni). La scelta di usare solo gli ultimi tre anni sembra in molti casi ragionevole, ma può essere, a maggior ragione, adattata a una distribuzione su tutto l'arco della scuola superiore.
Un'avvertenza generale.
La mentalità della probabilità e della statistica è particolare, e molto diversa da quella del resto della matematica che lo studente incontra a scuola. Una volta scelti i contenuti da insegnare in un dato anno, mi sembra opportuno presentarli in modo abbastanza concentrato nel tempo, non troppo annacquato: diversamente, si rischia di veder dimenticare tutto nel passare da un argomento all'altro. Per lo stesso motivo in linea generale mi sembra preferibile distribuire la materia sull'arco di tre anni e non di più.

Modulo 1
Statistica descrittiva

Contenuti
Statistica univariata: variabili statistiche, tabelle di distribuzione (frequenze assolute, relative, cumulative), indici (media, mediana, varianza, quantili, eccetera) e grafici (istogrammi, distribuzioni cumulative); statistica bivariata: diagramma di dispersione, correlazione, retta di regressione, metodo dei minimi quadrati.

Obiettivi
Per cominciare, si tratta di alfabetizzare all'uso di *termini* (come media, mediana, quantili, percentili, varianza), *concetti* (correlazione diretta, inversa), e anche alla lettura di *grafici* (istogrammi, diagrammi di dispersione) che oggi sono sempre più onnipresenti: infatti, fanno parte sia della cultura generale scientifica (in particolare, nelle discipline sperimentali), sia delle scienze sociali e dell'economia, sia del linguaggio comune dei mass media, che sempre più spesso riportano dati statistici in forma grafica o tabellare. Questo primo obiettivo, per un verso abbastanza «umile»,

è pur sempre un modo di attrezzare i ragazzi alla comprensione di un linguaggio diffuso nel mondo in cui vivono, e quindi è importante. In secondo luogo, l'introduzione dei primi elementi di statistica descrittiva consente di insegnare un modo preciso e quantitativo di esplorare dati numerici ed effettuare raffronti tra gruppi di dati, in qualunque ambito, andando oltre ciò che potrebbe suggerire il puro buon senso.
Più specificamente, lo studio della statistica bivariata (diagramma di dispersione, correlazione, regressione) si collega alla ricerca scientifica di leggi fenomenologiche o correlazioni empiriche tra due grandezze: può essere quindi utilmente affiancata alla fisica sperimentale. Infine, non è difficile trovare esempi, didatticamente significativi, in cui un uso ingenuo di questi strumenti porti ad affermazioni prive di buon senso (come l'esistenza di una correlazione - in senso statistico - tra fenomeni che evidentemente non hanno tra loro una relazione di nesso causale), e questo può offrire lo spunto per stimolare negli studenti una riflessione sull'uso critico di questi strumenti, questione sempre cruciale nell'applicazione della matematica alla realtà.

Prerequisiti matematici
Sostanzialmente nessuno, (equazione della retta per la regressione lineare!); per *giustificare* la scelta dei coefficienti nell'equazione della retta di regressione (metodo dei minimi quadrati) si può usare un argomento unidimensionale, basato sul vertice della parabola.

Modulo 2
Introduzione alla probabilità nel discreto

Contenuti
Introduzione alla probabilità (nel discreto), probabilità di eventi e operazioni con essa, probabilità condizionata, indipendenza; problemi nel discreto, affrontati utilizzando l'algebra della probabilità, e *non necessariamente* o *unicamente* il calcolo combinatorio per problemi di probabilità classica.

Obiettivi
Sono molto più «concettuali» che nel modulo precedente. Si tratta qui di introdurre la mentalità probabilistica, ossia l'idea di dipendenza delle opinioni, e quindi delle decisioni da prendere, dalle *informazioni incomplete* in nostro possesso, che possono cambiare nel tempo e da persona a persona. La mentalità probabilistica è a mio avviso un elemento importante della mentalità scientifica: insegnarla significa anche introdurre una sensibilità critica nei confronti delle «certezze» che la mentalità comune pretende dalla scienza, segnalando la presenza di un margine di incertezza

in molte nostre affermazioni, ma sdrammatizzando questa incertezza in presenza di probabilità molto grandi o molto piccole (si pensi, per esempio, al dibattito su *rischi* e *precauzioni* in campo costruttivo, ingegneristico o medico: quando i margini di sicurezza sono «ragionevoli»?).
Altro obiettivo è offrire spunti di riflessione sulle *teorie assiomatiche*: anche una presentazione elementare della probabilità, infatti, a mio avviso deve comprendere (fin dall'inizio) l'enunciazione degli *assiomi* della probabilità, e quindi presentare il calcolo delle probabilità come uno schema di discorso che può essere applicato in molti modi.
Occorre evitare, a mio avviso, di appiattire il discorso della probabilità sul concetto di probabilità classica. Questa sottolineatura porta con sé un altro elemento importante: la possibilità di far apprezzare il ruolo distinto che, nella *applicazione alla realtà* di una teoria assiomatica, giocano i giudizi iniziali (in base all'osservazione della realtà o alle opinioni) e la deduzione formale conseguente.
Detto più esplicitamente: a certi eventi elementari possiamo *assegnare* certe probabilità sulla base delle simmetrie oggettive (casi finiti equiprobabili, schema della probabilità classica), o sulla base delle frequenze relative osservate in passato (punto di vista frequentista); a questo punto, fissate queste premesse, il calcolo delle probabilità con le sue regole consente di determinare (cioè *calcolare*) la probabilità di eventi più complessi.
Tutti questi concetti si possono fare emergere anche senza sviluppare esempi complicati di combinatoria e probabilità classica. Certamente questo discorso diventa significativo se approfondito, ma la combinatoria non è l'unica possibile direzione di approfondimento.

Prerequisiti matematici
Linguaggio insiemistico, simbolo di sommatoria, eventualmente il concetto di serie (introdotta intuitivamente); questo modulo è concettualmente indipendente dal Modulo 1.

Modulo 2.A
Calcolo combinatorio e probabilità classica

Contenuti
Introduzione della principali situazioni combinatorie utilizzando due «schemi di ragionamento»: lo schema delle scelte successive (diagramma ad albero - principio del prodotto delle possibilità); lo schema delle scelte simultanee (sottoinsiemi di k elementi estratti da un insieme di n elementi - coefficienti binomiali).
Questo approccio consente di introdurre sinteticamente permutazioni

semplici, disposizioni semplici, disposizioni con ripetizione, combinazioni, e affrontare anche problemi più complessi in cui si combinano insieme più idee di questo tipo.
Applicazione del calcolo combinatorio a problemi e situazioni di probabilità classica.

Obiettivi
Questi argomenti sono certamente una buona palestra per educare il *ragionamento analitico*: analizzare un problema in sottoproblemi, calcolare a ogni passo il numero di opzioni possibili, il tutto senza perdere il filo del discorso.
Nella misura in cui lo studente è in grado di padroneggiare questo ragionamento analitico, può essere istruttivo (e di soddisfazione) mostrare come poche regole e idee di base possano combinarsi tra loro in modo creativo, per affrontare situazioni apparentemente molto diverse tra loro, e di complessità crescente.
Altro aspetto che il problema combinatorio può servire ad educare è la *capacità di astrazione*, necessaria per applicare uno schema generale alle situazioni concrete.
Per esempio: chi sono gli «oggetti» e chi sono i «posti», in questo problema concreto?
Nell'utilizzare il calcolo combinatorio per i problemi di probabilità classica, non ci si dimentichi delle regole di calcolo della probabilità già apprese (vedi modulo 2): per esempio, non sempre la strada più semplice consiste nel calcolare casi favorevoli e casi possibili per l'evento di cui interessa la probabilità: potrebbe essere più semplice calcolare la probabilità dell'evento complementare, o di altri eventi da cui poi può essere dedotta quella che interessa. In altre parole: se il calcolo combinatorio viene insegnato in funzione della probabilità, facciamo interagire il più possibile le due discipline, anziché muoverci con i paraocchi.

Qualche osservazione generale
Il calcolo combinatorio, nel curriculum scolastico, a mio parere val la pena sia insegnato solo se si utilizza in questo contesto probabilistico: il suo valore intrinseco è limitato, rispetto agli altri argomenti che si toccano nei programmi di scuola.
L'affronto di problemi combinatori, non appena escano dalla prevedibile ripetitività (come quando si assegnano esercizi esclusivamente sull'ultima formula spiegata) richiede intelligenza vivace, capace di astrazione creativa; non mi sembra utile cercare «a tutti i costi» di renderlo alla portata di tutti riducendolo a schemi ripetitivi mnemonici. Se la classe non lo consente, si faccia altro: si possono dire tante cose interessanti sulla probabilità anche senza usare problemi combinatori.

Modulo 3
Variabili aleatorie, modelli discreti

Contenuti
Introduzione dei primi concetti generali sulle variabili aleatorie discrete (legge, valore atteso, varianza, indipendenza, covarianza, eccetera), per introdurre successivamente il modello di Bernoulli, con le sue varie leggi (Bernoulliana, Binomiale, Geometrica), e le situazioni che permette di descrivere. L'enfasi è posta sul modello di Bernoulli.

Obiettivi
Anzitutto, si mostra uno sviluppo significativo dei primi concetti di probabilità (vedi modulo 2), evidenziando nuovamente, in particolare, come con poche ipotesi semplici e ragionevoli su un certo fenomeno, il calcolo delle probabilità consenta di trarre deduzioni potenti. Il modello di Bernoulli, pur nella sua semplicità, è estremamente significativo a questo riguardo, e perciò didatticamente prezioso.
Mediante il concetto di modello probabilistico, si può dare concretezza all'idea scientifica generale di *modello*, che significa in sostanza quanto detto qui sopra (da poche ipotesi ragionevoli si traggono conclusioni potenti grazie alla teoria). Inoltre, l'applicazione del modello alla realtà richiede e quindi favorisce l'*educazione della capacità analitica e di astrazione*. Pensiamo per esempio al percorso logico che lo studente deve compiere nell'affronto di un problema di questo tipo: definiamo con precisione una variabile aleatoria nel problema in esame (per esempio: «sia X la variabile che conta il numero di volte che lanciando un dado ho ottenuto un sei, su 10 lanci»); ne riconosciamo la legge (per esempio: «X ha legge binomiale di parametri n = 10, p = 1/6»); ci chiediamo che cosa ci interessa calcolare - la probabilità che essa assuma certi valori (per esempio: «voglio calcolare la probabilità che X sia maggiore di 8»), oppure il suo valore atteso-; impostiamo il calcolo; eseguiamo il calcolo.
Questo richiede, come già detto, capacità di analisi e di astrazione, che poggiano su e si esprimono con un *linguaggio* verbale, scritto e simbolico *estremamente preciso* che, a mio avviso, fa parte della maturità e consapevolezza a cui si deve e si può educare attraverso la matematica e le scienze.

Prerequisiti matematici
I contenuti del modulo 2; coefficienti binomiali e loro significato combinatorio (ma solo questo, senza grandi allenamenti all'uso di situazioni diverse di conteggio).
In teoria, occorrerebbe la nozione di serie, ma se ne può fare a meno, utilizzando la somma (finita) geometrica e qualche discorso intuitivo.

Relazioni con altri argomenti
Se si è studiata in precedenza la statistica descrittiva, si apre l'opportunità di un confronto tra i concetti di *variabile statistica* e *variabile aleatoria*, con gli analoghi, ma diversi, concetti di media, varianza, covarianza; anzi, questo confronto è doveroso, se si vuole mettere ordine nelle idee; ma se non è stata studiata la statistica descrittiva non c'è alcun problema: il concetto di variabile aleatoria ha vita autonoma.

Modulo 3.A
Modello di Poisson
e introduzione «soft» alle variabili aleatorie continue

Contenuti
Proseguendo il discorso sui modelli discreti, si introduce il modello di Poisson, che viene affiancato a quello di Bernoulli. Dalla legge di Poisson (come modello di variabile che conta il numero di arrivi casuali in un intervallo di tempo $[0,t]$) si introduce poi la legge esponenziale (primo esempio di legge di variabile aleatoria continua) mediante la sua funzione di ripartizione, ossia calcolando la probabilità che il primo arrivo avvenga prima dell'istante t.

Obiettivi
Sono gli stessi elencati nel modulo 3, con qualche elemento specifico in più. Anzitutto, avere a disposizione due modelli (Bernoulli e Poisson) permette di affrontare «problemi reali» (sia pur ipersemplificati) in cui si chiede di *riconoscere qual è il modello adeguato*, facendo fare allo studente esperienza concreta di cosa significa scegliere il modello migliore per descrivere un fenomeno. Questo ovviamente non si può fare fintanto che si conosce un modello solo!
Dal punto di vista matematico, il modello di Poisson offre un'interessante applicazione del limite notevole che definisce il numero e.
Infine, il modello di Poisson dà lo spunto naturale per introdurre la *legge esponenziale*, che è la legge di una variabile *continua*, non più *discreta*, mediante la sua funzione di ripartizione, quindi *senza usare integrali*. Può essere un modo per dare una prima idea delle variabili aleatorie continue, più presto di quanto non consentirebbe la trattazione degli integrali nel programma di analisi dell'ultimo anno.
Ecco perché nel titolo di questo modulo ho parlato di un'introduzione *soft* alle variabili continue.

Prerequisiti matematici
I contenuti dei moduli 2 e 3; limiti, definizione di e, serie.

Modulo 3.B
Variabili aleatorie continue

Contenuti
Il concetto di legge di una variabile aleatoria continua (mediante la sua densità e funzione di ripartizione, senza l'introduzione esplicita della variabile come funzione definita su uno spazio di probabilità astratto) e le sue prime proprietà (in analogia a quelle discrete). Quindi si introduce la legge normale per discuterne le varie applicazioni, arrivando al teorema centrale del limite.

Obiettivi
Le proprietà delle variabili aleatorie continue *in parte* ricalcano quelle delle variabili aleatorie discrete (e quindi, in un certo senso, non insegnano molto di nuovo); in parte invece queste proprietà, o le loro dimostrazioni, vi si discostano, e in tal caso diventano troppo difficili per essere trattate rigorosamente a scuola.
Di conseguenza, l'obiettivo con cui si svolge questa parte non è tanto quello di presentare i *concetti generali* riguardanti le variabili continue, quanto quello di introdurre una *specifica* classe di variabili continue, e cioè le variabili di *legge normale*.
La legge normale ha anzitutto una sua utilità intrinseca per le numerose applicazioni modellistiche (dalla curva degli errori alla distribuzione di numerose grandezze fenomenologiche); in secondo luogo, è il punto di partenza per poter discutere il *teorema del limite centrale* e l'*approssimazione normale*, argomenti fondamentali nel calcolo delle probabilità. Questi infatti, da una parte costituiscono strumenti potenti per il calcolo approssimato delle probabilità di certi eventi che sarebbero altrimenti ben difficili da valutare; in secondo luogo costituiscono la giustificazione teorica del fatto che il modello normale sia così «onnipresente». In particolare, se si discute l'*approssimazione normale della binomiale*, si dà lo spunto per mostrare interessanti nessi tra il mondo discreto e quello continuo, apparentemente lontani.
Infine, ricordiamo che la trattazione della legge normale e del teorema del limite centrale è una base teorica necessaria se si vuole successivamente discutere la *stima per intervalli* (che è sostanzialmente l'unico argomento di *statistica inferenziale* presente nei programmi scolastici).

Prerequisiti matematici
Integrali; i contenuti dei moduli 2 e 3. Per dar corpo alla comprensione del concetto di media campionaria, occorrerebbero anche i contenuti del modulo 4, e quindi del modulo 1; questi però non tassativamente.

Nota didattica
Non è necessario pensare di aver effettivamente svolto gli integrali. Ci si può accontentare di usarne il simbolo, col suo significato geometrico intuitivo (come si fa spesso all'inizio dell'insegnamento della fisica), e poi lavorare soprattutto con la funzione di ripartizione, facendo leva sull'analogia tra le proprietà formali delle variabili aleatorie discrete e continue. Operativamente, per calcolare gli integrali della densità normale si usano le *tavole*, non la primitiva, quindi non è richiesta la familiarità con il *calcolo integrale* nell'accezione consueta. Si dovranno, almeno in un primo momento, dare per buoni i risultati su media e varianza della normale standard, e il teorema che dà il legame tra la normale qualunque e la normale standard. Eventualmente, dopo aver introdotto gli integrali, si può ritornare su questi risultati e dimostrarli.

Modulo 4
Relazione tra variabili statistiche e variabili aleatorie

Contenuti
Prime idee sul campionamento e la statistica inferenziale; concetto di campione casuale; media campionaria. Disuguaglianza di Cebicev e legge dei grandi numeri. Stima puntuale. Il modello di Bernoulli è sufficiente per esemplificare queste idee.

Obiettivi
Si mostra la relazione tra le *variabili statistiche*, studiate nella *statistica descrittiva*, e i *modelli probabilistici*, descritti dal concetto di variabile aleatoria, per evidenziare il nesso tra probabilità, statistica descrittiva, statistica inferenziale. Tecnicamente il concetto chiave che fa da ponte tra questi due mondi è quello di *campione casuale*. La sua introduzione e discussione è uno degli obiettivi di questa parte e dà lo spunto per qualche chiarimento concettuale della problematica della statistica applicata al mondo reale: sondaggi e altri concetti comuni nei media. Si tratta in realtà di una problematica molto ampia, che qui si potrà appena accennare, ma può essere comunque istruttivo farlo. Il concetto di campione casuale consente anche di dare una formulazione precisa della *legge dei grandi numeri*, che è senz'altro una delle «grandi idee» della probabilità. Studiarla dà lo spunto per rivisitare la *concezione frequentista della probabilità* e, quindi, per riflettere sulla natura «sfuggente» del concetto di probabilità che non si lascia ridurre in alcun modo a ciò che è deterministico e, quindi, in un certo senso, non si lascia «definire».

Prerequisiti matematici
Se ci si appoggia al modello di Bernoulli per dare concretezza al discorso, i prerequisiti di analisi sono pochi; più lungo invece è il cammino che occorre aver fatto in statistica descrittiva e probabilità: i contenuti dei moduli 1, 2, 3.

Modulo 5
Statistica inferenziale: stima per intervalli

Contenuti
Stima per intervalli di una proporzione (ossia del parametro di una Bernoulliana) per grandi campioni; stima della media di una popolazione normale di varianza nota.
Ampiezza del campione, livello di confidenza, precisione della stima.

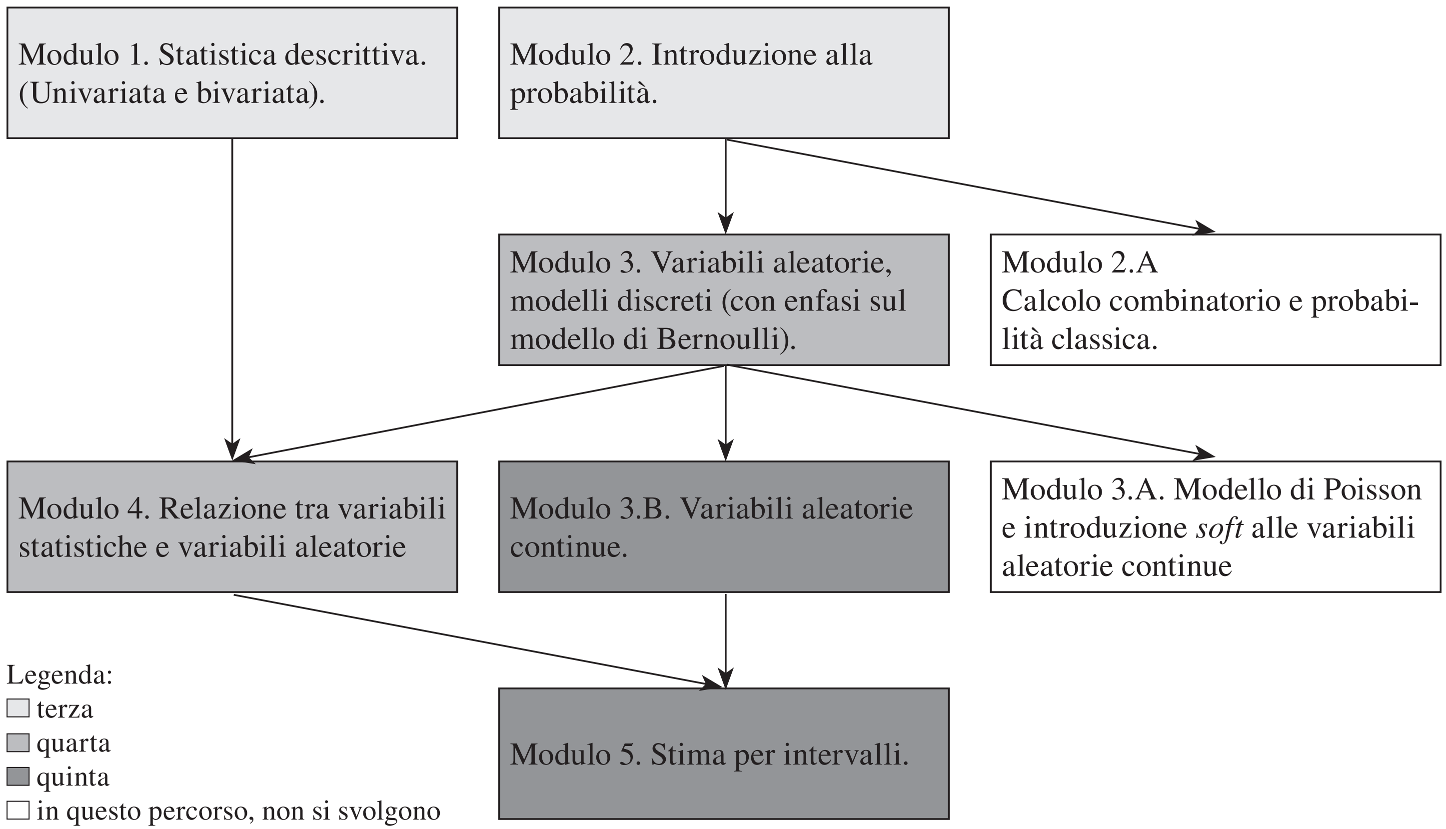
Obiettivi
Se si arriva fino a questo punto (cosa non facile per il lungo cammino che questo presuppone) si ha la possibilità di toccare con mano la potenza quantitativa dello strumento statistico inferenziale, applicato per esempio a un tema «caldo» com'è quello dei sondaggi d'opinione (elettorali o d'altro tipo) o, più in generale, di indagini statistiche campionarie. Dal punto di vista culturale, questo argomento mostra poi l'unità di statistica descrittiva, statistica inferenziale, calcolo delle probabilità e modelli probabilistici.

Prerequisiti matematici
Sono molti, naturalmente: buona parte dei contenuti descritti fin qui (moduli 1, 2, 3, 4)

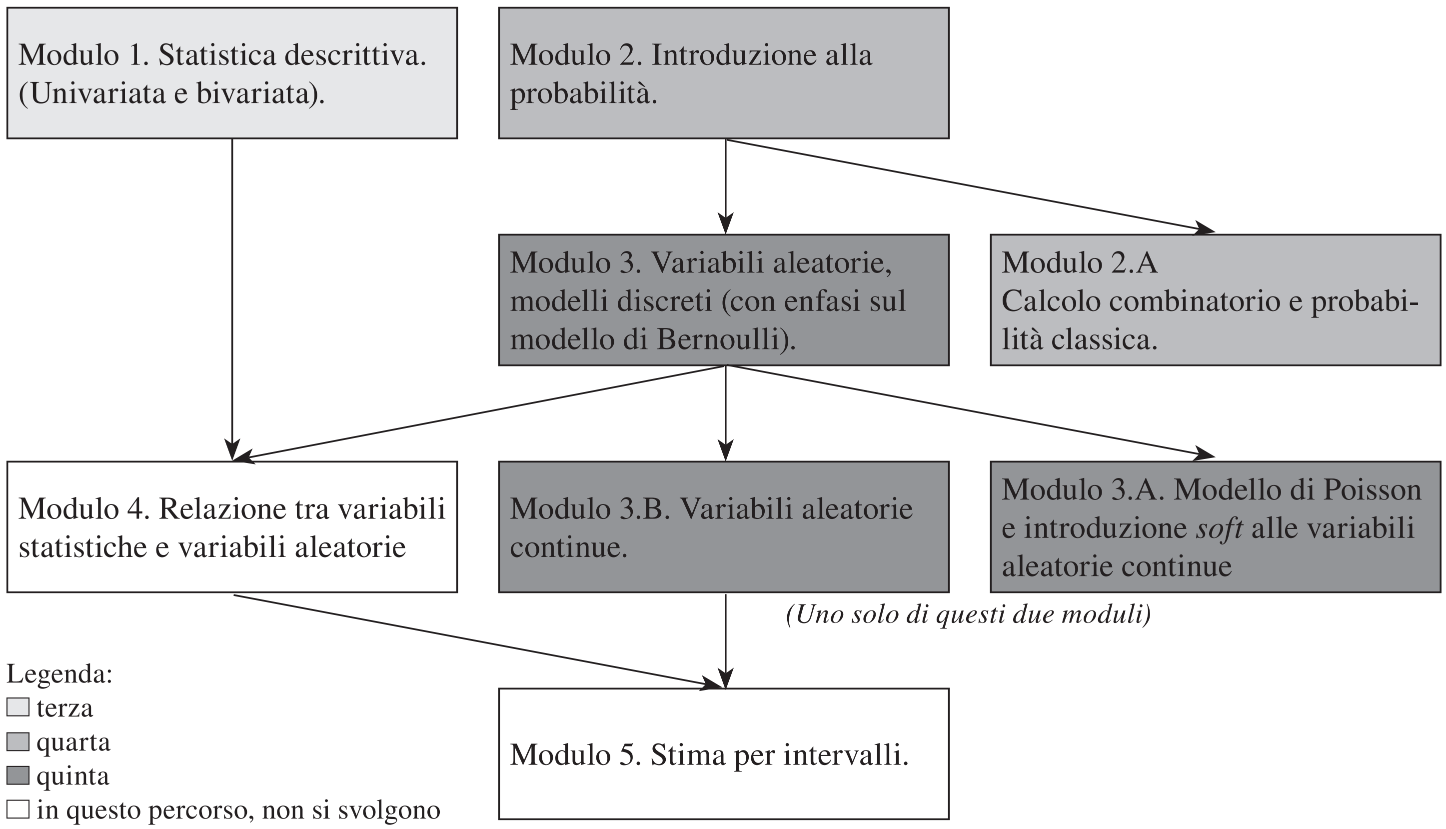
INDICAZIONI BIBLIOGRAFICHE
Un riferimento bibliografico sintetico, per l'insegnante, che contiene tutto il materiale a cui si è accennato, è: M. Bramanti, *Calcolo delle Probabilità e Statistica*, Ed. Progetto Leonardo, Esculapio, Bologna 1997.
Il relativo eserciziario è: D. Bertacchi, M. Bramanti, G. Guerra, *Esercizi di Calcolo delle Probabilità e Statistica*, Ed. Progetto Leonardo, Esculapio, Bologna 2003.
Un ciclo di libri di testo per la scuola che copre questi argomenti, è invece: M. Andreini, R. Manara, F. Prestipino, *Matematica Controluce – per i programmi sperimentali*, ETAS, Milano, 2007.
Per motivi di spazio in questo articolo non ho discusso del possibile uso del computer nell'insegnamento della probabilità e statistica, ma naturalmente questo è senz'altro opportuno. Riguardo all'uso didattico di *Excel* per queste materie, si vedano in particolare i seguenti testi:
E. Battistini, *Probabilità e Statistica: un approccio interattivo con Excel*, McGraw-Hill, Milano 2004, (con cd-rom).
A. M. Paganoni, L. Pontiggia, *Laboratorio di Statistica con Excel*, Pearson Educational, 2007 (con files scaricabili dal sito).
Entrambi i testi sono pensati per l'università, ma possono essere senz'altro utili per il docente di scuola; in particolare, entrambi insegnano a usare *Excel* per questi scopi anche a chi non ne abbia già conoscenze, e contengono molto materiale.

Per eventuali contatti con l'autore: marco.bramanti@polimi.it - www1.mate.polimi.it/~bramanti/

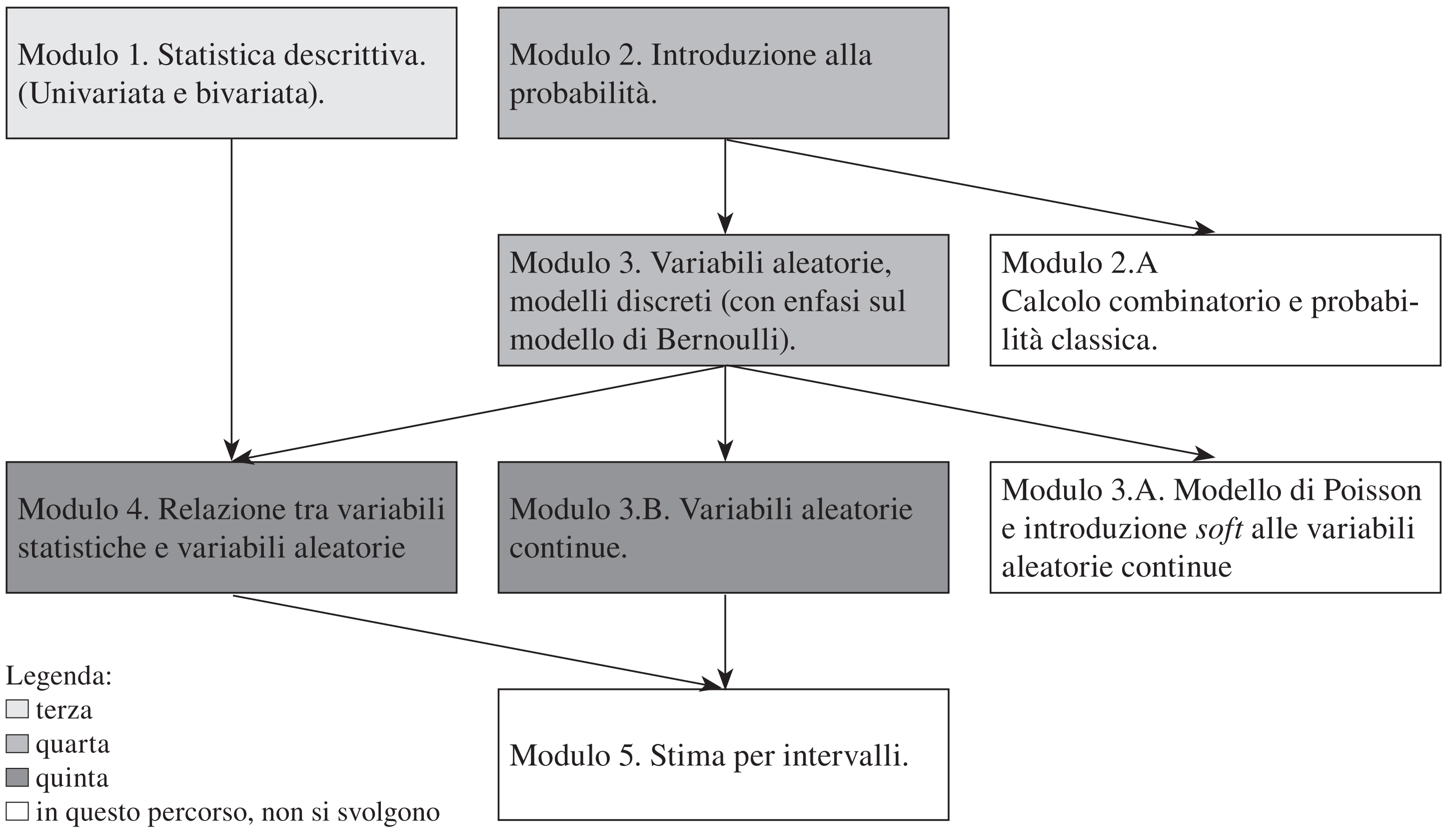
Percorso finalizzato a mostrare le tre discipline: statistica descrittiva, probabilità, statistica inferenziale



Percorso statistica descrittiva e probabilità (senza collegamento tra le due, e senza statistica inferenziale).
Enfasi posta sui modelli probabilistici; la statistica descrittiva qui è uno strumento separato



Percorso variabili statistiche e aleatorie (con collegamento tra le due, ma senza la stima per intervalli)



Percorso di calcolo delle probabilità e modelli probabilistici, senza statistica

Si dà più spazio a variabili aleatorie e modelli probabilistici