

Prefazione per il docente

Questo testo è stato scritto per un corso di calcolo delle probabilità e statistica (di una semiannualità) rivolto a studenti del primo anno del Diploma in Ingegneria, che hanno già seguito un corso di matematica generale. L'esigenza da cui è nato il testo è quella di fornire un'introduzione elementare e molto sintetica (il corso si articola in 30 ore di lezione e 20 di esercitazione) a queste discipline, toccando argomenti di calcolo delle probabilità, statistica descrittiva e statistica inferenziale. Trattandosi di un corso di diploma, si privilegia l'aspetto operativo all'approfondimento teorico. La sintesi necessaria per trattare in così poche ore gli argomenti citati, tuttavia, non può essere raggiunta semplicemente "omettendo le dimostrazioni": si tratta di ridisegnare opportunamente il susseguirsi degli argomenti, rispetto all'ordine consueto con cui sono presentati in un corso rivolto agli studenti di un corso di laurea. Le principali scelte didattiche compiute sono le seguenti.

La statistica descrittiva è trattata come primo argomento. Lo scopo è introdurre il concetto di *variabile*, che sarà sviluppato più tardi mediante quello di *variabile aleatoria*, cominciare a familiarizzare con i concetti di *media* e *varianza*, e introdurre la nozione di *correlazione tra variabili* e *retta di regressione*: questi ultimi sono argomenti che in un corso così breve sarebbero altrimenti del tutto sacrificati. Parlarne da un punto di vista descrittivo è quindi un compromesso tra il parlarne dal punto di vista della statistica inferenziale (che richiederebbe alcune ore in più) e non parlarne del tutto.

La combinatoria è un argomento che in un corso di diploma non può essere considerato "prerequisito", in quanto difficilmente trova spazio nel corso di matematica generale, già molto nutrito di argomenti. Perciò i concetti elementari del calcolo combinatorio sono stati introdotti, non come argomento a sé stante, ma in relazione ai problemi della probabilità classica.

La probabilità è introdotta in spazi discreti, e tutte le proprietà "algebriche" della probabilità di eventi sono sviluppate in questo contesto: questo consente di non citare neppure il concetto di "sigma algebra degli eventi", senza perdere di rigore. Si introducono quindi le variabili aleatorie discrete come funzioni (qualsiasi!) definite su uno spazio campionario discreto. Solo dopo aver sviluppato una certa confidenza con le variabili aleatorie si introduce la nozione di *legge di una variabile aleatoria continua*, definita mediante la densità continua.

Una certa enfasi, nella parte sulle variabili aleatorie, è posta sull'aspetto modellistico e in particolare su tre modelli fondamentali: il processo di Bernoulli, il processo di Poisson, il modello normale. Le distribuzioni notevoli sono introdotte con riferimento a questi modelli, anziché assegnandone la legge; questo ha alcuni vantaggi: ad esempio il fatto che una binomiale sia somma di bernoulliane indipendenti diventa un'immediata conseguenza della definizione, e quindi permette di calcolare facilmente media e varianza di questa legge, sfruttando le proprietà del valore atteso e della varianza (enunciate e non dimostrate); il fatto che la densità discreta di

questa legge abbia una certa espressione analitica è invece dimostrato, come applicazione dei ragionamenti elementari di combinatoria e probabilità visti in precedenza.

I concetti di *modello statistico*, *campione casuale*, *stimatore* sono introdotti subito dopo le prime nozioni sulle variabili aleatorie discrete: anche se queste nozioni saranno utilizzate esplicitamente solo nell'ultima parte del corso (statistica inferenziale), permettono di chiarire il punto di vista con cui si useranno i modelli introdotti; inoltre, il concetto di campione casuale getta un ponte tra la statistica descrittiva e lo studio delle variabili aleatorie, ed è quindi utile a mostrare l'unità tra gli argomenti trattati nel corso.

Anche se nessun argomento ingegneristico è stato sviluppato in dettaglio, ho cercato di dare un'idea di alcune tipiche applicazioni della teoria e dei modelli presentati: affidabilità di un sistema, controllo di qualità, tempo di vita di un apparecchio, code di attesa, indagini comparative \hat{a}

Naturalmente, la sottolineatura posta sui modelli "concreti" ha un prezzo, nell'economia del corso: sacrificare molti argomenti più teorici riguardanti le variabili aleatorie. In particolare, non si trattano qui: densità congiunte; densità condizionate e valore atteso condizionato; metodi di calcolo della legge di una funzione di una o più variabili aleatorie.

In sostanza, questi "tagli" corrispondono alla scelta di trattare solo variabili aleatorie singole o **8**-uple di v.a. indipendenti.

Coerentemente a queste scelte, nella parte di statistica inferenziale non si è data quasi nessuna nozione generale sugli stimatori (massima verosimiglianza, stimatori UMVUE, lemma di Neyman-Pearson \hat{a}), concentrandosi invece sulle nozioni di *intervallo di confidenza* e *test di ipotesi*, sviluppate nei casi dei modelli studiati. Nel far questo, si è cercato di fornire metodi che arrivino fino alla "ricetta operativa", dando risposte pragmatiche (perciò certamente opinabili) ad alcune tipiche domande che sorgono nell'applicazione concreta delle procedure statistiche: quando posso applicare questa approssimazione? Quando posso applicare questo test? Non si pretende certo di aver dato gli strumenti per affrontare qualsiasi problema, ma almeno qualche strumento per affrontare *effettivamente* qualche problema.

Il testo, come il corso, è pensato come un unico fluire di lezioni ed esercitazioni. Di conseguenza la successione degli argomenti è stata studiata in modo che non ci siano mai troppi paragrafi "teorici" consecutivi, e gli esempi necessari per illustrare la teoria sono stati sviluppati fin nel dettaglio di come eseguire certi calcoli o costruire certe tabelle: il fatto che certe operazioni siano un po' noiose non significa necessariamente che lo studente sia in grado di svolgerle correttamente senza alcun suggerimento, e non è quindi a mio avviso un buon motivo per ometterle dall'insegnamento.

Il corso è affiancato da alcune ore di esercitazione in aula informatizzata. L'utilizzo di un software statistico (nel nostro caso StataQuest 4 per Windows) può essere didatticamente molto utile a illustrare le idee e le procedure statistiche, e anche in questo testo ci sono frequenti cenni a ciò che si può fare con un computer. Fin tanto che i nostri studenti non disporranno di un computer su cui lavorare quotidianamente, però, ritengo sia inevitabile

che un corso del genere insegni anzitutto ad eseguire i calcoli con carta, penna e calcolatrice.

Desidero ringraziare particolarmente le persone che, tenendo le esercitazioni del mio corso nell'A.A. '96/'97, hanno letto questi capitoli mano a mano che venivano prodotti, dando utili suggerimenti e segnalandomi errori: Giacomo Aletti, Anna Cagnacci e Franco Tornaghi. La responsabilità di eventuali altri errori rimane naturalmente mia.

Questo testo è stato composto dall'autore, utilizzando il seguente software, in ambiente Windows:

• 4.0 (word processor);

StataQuest 4.0 (grafici statistici, calcoli statistici);

Excel 5.0 (diagrammi a barre, calcoli per le tabelle, tavole a fine testo);

Mathematica 2.0 (grafici di funzioni).

Le immagini prodotte con gli ultimi 3 programmi sono semplicemente copiate nei files • .

M. B.

Milano, novembre '97