

Politecnico di Milano
Appunti di calcolo delle probabilità per il corso di
Fondamenti di Statistica e Segnali Biomedici [Mod 1]¹

Ilenia Epifani

¹Il contenuto di queste dispense è protetto dalle leggi sul copyright e dalle disposizioni dei trattati internazionali. Il materiale qui contenuto può essere copiato (o comunque riprodotto) ed utilizzato liberamente dagli studenti, dagli istituti di ricerca, scolastici ed universitari afferenti ai Ministeri della Pubblica Istruzione e dell'Università e della Ricerca Scientifica e Tecnologica per scopi istituzionali, non a fine di lucro. Ogni altro utilizzo o riproduzione (ivi incluse, ma non limitatamente a, le riproduzioni a mezzo stampa, su supporti magnetici o su reti di calcolatori) in toto o in parte è vietata, se non esplicitamente autorizzata per iscritto, a priori, da parte degli autori. L'informazione contenuta in queste pagine è ritenuta essere accurata alla data della pubblicazione. Essa è fornita per scopi meramente didattici. L'informazione contenuta in queste pagine è soggetta a cambiamenti senza preavviso. Gli autori non si assumono alcuna responsabilità per il contenuto di queste pagine (ivi incluse, ma non limitatamente a, la correttezza, completezza, applicabilità ed aggiornamento dell'informazione). In ogni caso non può essere dichiarata conformità all'informazione contenuta in queste pagine. In ogni caso questa nota di copyright non deve mai essere rimossa e deve essere riportata anche in utilizzi parziali. Copyright 2010 Ilenia Epifani.

Indice

1	Introduzione	1
1.1	Introduzione	1
1.2	Organizzazione degli appunti	2
1.3	Riferimenti bibliografici	3
2	Probabilità	1
2.1	Esperimento aleatorio, spazio campionario, eventi, probabilità	1
2.1.1	Proprietà della probabilità	4
2.2	Probabilità in spazi campionari finiti o numerabili	6
2.2.1	Esperimenti con esiti equiprobabili	6
2.3	Regole di conteggio	7
2.3.1	Campionamento da urne	8
2.4	Probabilità condizionata ed indipendenza	10
2.4.1	Alcune formule importanti	11
2.4.2	Indipendenza	13
2.4.3	Modelli probabilistici basati sull'indipendenza	14
3	Variabili aleatorie discrete	17
3.1	Variabili aleatorie	17
3.2	Variabili aleatorie discrete	18
3.3	Esempi di densità discrete notevoli	20
3.3.1	Densità binomiale e bernoulliana	20
3.3.2	Densità Geometrica	23
3.3.3	Densità di Poisson	24
3.4	Trasformazioni di variabili aleatorie discrete	26
3.5	Valore atteso di variabili aleatorie discrete	27
3.5.1	Proprietà del valore atteso	28
3.6	Varianza di variabile aleatoria discreta	29
3.6.1	Proprietà della varianza	30
3.6.2	Disuguaglianza di Chebychev	31
3.7	Standardizzazione di una variabile aleatoria	31
4	Variabili aleatorie continue	33
4.1	Introduzione	33
4.2	Variabili aleatorie assolutamente continue	34
4.2.1	Trasformazioni di una variabile aleatoria continua	35

4.2.2	Indici di variabili aleatorie assolutamente continue	37
4.3	Esempi di densità continue notevoli	38
4.3.1	Densità uniforme continua	38
4.3.2	Densità esponenziale	40
4.3.3	Densità gaussiana	42
4.3.4	Densità lognormale	44
4.4	Approssimazione gaussiana della funzione di ripartizione binomiale	45
4.5	Accenno al caso di variabili aleatorie miste	46
5	Variabili aleatorie multivariate	49
5.1	Variabili aleatorie indipendenti	50
5.2	Coppie di variabili aleatorie	51
5.2.1	Coppie di variabili aleatorie discrete	51
5.3	Variabili aleatorie congiuntamente continue	54
5.3.1	Densità congiunta uniforme	55
5.4	Trasformazioni di variabili aleatorie	57
5.5	Valore atteso di trasformazioni di variabili aleatorie	57
5.5.1	Somme di variabili aleatorie	58
5.6	Covarianza, coefficiente di correlazione	60
5.6.1	Riproducibilità	62
5.6.2	Densità di Erlang	64
5.7	Densità bivariata gaussiana	64
6	Media campionaria e teoremi limite	65
6.1	Campione casuale e media campionaria	65
6.2	Legge dei grandi numeri	66
6.3	Teorema centrale del limite	68
7	Test diagnostici, valutazione di riconoscitori e curve ROC	71
7.1	Test diagnostici	71
7.1.1	Stima della sensibilità e specificità	72

Capitolo 1

Introduzione

Premessa

I primi 5 crediti del corso di Fondamenti di Statistica e Segnali Biomedici sono volti a familiarizzare lo studente con alcune delle tecniche statistiche di base più usate nelle applicazioni dell'ingegneria biomedica. Cercherò di presentare in forma chiara ed accessibile i concetti base di statistica e teoria della probabilità. Il corso è introduttivo, strumentale, di servizio. Quindi vedremo solo qualche semplice applicazione. Non risolveremo mai un problema ingegneristico biomedico “vero”. Proviamo ad introdurci al problema.

1.1 Introduzione¹

Ross 2005 definisce la statistica *l'arte di imparare dai dati*. Per risolvere un problema complicato scientifico o pratico che sia, occorre 1) raccogliere i dati, 2) descriverli e 3) analizzarli per trarne delle conclusioni generali. La branca della statistica che si occupa di raccogliere e descrivere i dati si chiama statistica descrittiva. Le quantità statistiche che sintetizzano un insieme di dati si chiamano *statistiche*. I valori delle statistiche sono determinati dai dati. La statistica descrittiva sarà trattata nella parte di laboratorio. La branca della statistica che analizza i dati per trarne conclusioni si chiama statistica inferenziale.

Tipicamente, quando si fa statistica non si censisce una intera popolazione, ma si procede a “campione”. Inoltre, alcuni dati sono di per se variabili, soggetti al caso, cioè “puramente casuali”. Per esempio, se lanciamo 10 volte una moneta e 8 volte otteniamo testa, possiamo concludere che la moneta è truccata? Oppure, la moneta è regolare ed è un fatto puramente casuale avere osservato testa così tante volte?

Non solo: la moneta è un corpo rigido e il moto di un corpo rigido nello spazio è ben descritto dalle equazioni della meccanica newtoniana, quindi in linea di principio, se ad ogni lancio della moneta riusciamo a tenere conto della velocità iniziale con la quale viene lanciata la moneta, dell'attrito effettuato dall'aria e degli urti anelastici che la moneta subisce quando ricade a terra, potremmo calcolare se alla fine la moneta esibirà sulla faccia superiore testa o croce. Tuttavia un conto reale di questo genere risulta infattibile, sia perché non è possibile

¹Si veda Capitolo 1 di Ross 2008, *Introduzione alla Statistica*, Apogeo, e Capitolo 1 in Epifani, I., Ladelli, L.M. e Posta, G. (2006) *Appunti per il corso di Calcolo delle Probabilità*, AA 2005/2006 <http://www1.mate.polimi.it/~ileepi/dispense/0506CP/>

in generale misurare sperimentalmente le grandezze fisiche coinvolte, sia perché il sistema in esame esibisce una dipendenza sensibile dalle condizioni iniziali: una piccola (infinitesima) variazione delle condizioni iniziali (ad esempio la forza applicata nel lancio o posizione dalla quale si lancia) porta ad un effetto macroscopico notevole (ad esempio esce testa piuttosto che croce). Risulta invece chiaro che se la moneta è sufficientemente simmetrica ci attendiamo che la “possibilità” che dopo un lancio si presenti testa sia la stessa che si presenti croce. Da qui l’esigenza di modellare questo fenomeno attraverso una teoria diversa dalla teoria deterministica meccanica newtoniana. Ergo, per trarre conclusioni generali dai dati, non possiamo prescindere dall’influenza del caso. Per trarre conclusioni “sensate” sui dati, in modo cioè da porterne valutare l’affidabilità, bisogna fare assunzioni sulle “probabilità” con cui i dati misurati assumono ciascun valore osservato. Cioè bisogna far di conto con il caso, cioè non possiamo prescindere dal calcolo delle probabilità. L’insieme delle assunzioni su queste probabilità prende il nome di *modello probabilistico per i dati*.

Sintetizzando: Il fondamento dell’inferenza statistica consiste nel formulare un modello probabilistico atto a descrivere i dati. Da una parte, i dati raccolti migliorano la conoscenza del modello probabilistico. Dall’altra, l’assunzione di un modello probabilistico permette di misurare l’affidabilità delle conclusioni cui si arriva con l’inferenza statistica.

1.2 Organizzazione degli appunti

Gli appunti sono organizzati nel modo seguente.

Nel secondo capitolo vengono introdotte le nozioni di base della teoria delle probabilità, quali: spazio campionario, eventi, funzione di probabilità e sue proprietà, esperimenti con esiti equiprobabili, probabilità condizionata e alcune formule notevoli di calcolo delle probabilità quali la regola del prodotto, la formula delle probabilità totali e di Bayes. Viene poi sviluppato il concetto di indipendenza e le sue applicazioni all’affidabilità di sistemi in serie e in parallelo e allo schema delle prove di Bernoulli.

Nel terzo capitolo vengono introdotte le variabili aleatorie univariate discrete e le alcune caratteristiche deterministiche ad esse associate quali: valore atteso, varianza, deviazione standard e coefficiente di variazione. Vengono poi descritte alcune distribuzioni notevoli discrete quali: uniforme discreta, binomiale, di Bernoulli, di Poisson e geometrica con le loro principali proprietà.

Nel quarto capitolo vengono trattate le variabili aleatorie univariate assolutamente continue, con le caratteristiche deterministiche ad esse associate. Sono poi introdotte alcune distribuzioni notevoli continue quali: distribuzione uniforme continua, esponenziale, Erlang, gaussiana e lognormale con le loro principali proprietà.

Nel quinto capitolo vengono descritte le variabili aleatorie multivariate discrete e continue, con alcune caratteristiche deterministiche ad esse associate, quali covarianza e coefficiente di correlazione lineare. In particolare è studiata anche la somma di variabili aleatorie. Vengono poi forniti alcuni modelli notevoli di variabili multivariate quali la densità multinomiale, gaussiana bivariata e uniforme sul cerchio.

Nel sesto capitolo sono introdotti campione casuale e media campionaria e viene discussa qualche legge limite del calcolo delle probabilità, quale la legge debole dei grandi numeri e il teorema centrale del limite.

Infine, il Capitolo 7 è dedicato a test diagnostici, valutazione di riconoscitori, sensibilità, specificità e curve ROC.

1.3 Riferimenti bibliografici

In quanto “appunti”, queste note possono contenere ripetizioni, bachi e mancanze.

Un’ampia parte di questi appunti si basano sul seguente materiale:

1. Bland, M. (2000) *Statistica medica*, Apogeo. Titolo originale: *An introduction to medical statistics, 3rd edition*, Oxford University Press.
2. Epifani, I. (2009) *Esercizi di Calcolo delle Probabilità*, AA 2008/2009
<http://www1.mate.polimi.it/~ileepi/esercizi/0809CP/>
3. Epifani, I., Ladelli, L.M. e Posta, G. (2006) *Appunti per il corso di Calcolo delle Probabilità*, AA 2005/2006 <http://www1.mate.polimi.it/~ileepi/dispense/0506CP/>
4. Fawcett, T. (2004) ROC Graphs: Notes and Practical Considerations for Researchers, *Machine Learning*, **31**, 1–38
5. Ross, Sheldon M. (2004) *Probabilità e Statistica per l'ingegneria e le scienze*, Apogeo Milano.
6. Ross, Sheldon M. (2008) *Introduzione alla Statistica*, Apogeo Milano.
7. Simon, S. (2000-2008) Roc in <http://www.childrensmemory.org/stats/ask/roc.asp>
8. Verri, M. (AA 09/10) *Appunti di lezione*; insegnamento di Calcolo delle Probabilità e Statistica Matematica, scaricabile dalla pagina della didattica del prof. Verri.

Ma se inesattezze in questi appunti vi sono, la responsabilità è tutta mia.

Capitolo 2

Probabilità

2.1 Esperimento aleatorio, spazio campionario, eventi, probabilità¹

Il calcolo delle probabilità si occupa dello studio e della formalizzazione matematica di fenomeni “casuali”, (o *esperimenti aleatori*) cioè di fenomeni per i quali non possiamo predire a priori l’esito, e il termine “probabilità” è usato per quantificare la possibilità che un certo evento, inteso come *esito* di un esperimento, si verifichi. I motivi per i quali non è possibile dare una descrizione deterministica di un certo fenomeno sono molteplici: può accadere che le informazioni riguardanti il fenomeno sul quale vogliamo fare previsioni siano incomplete, può accadere che non esista una teoria che permetta di arrivare a dedurre delle conseguenze per il fenomeno in osservazione, oppure che la teoria esista ma risulti di difficile applicazione. Infine, come abbiamo già visto nell’Introduzione può accadere semplicemente che il fenomeno sia “casuale”.

Per capire più in dettaglio il problema che ci accingiamo a descrivere, partiamo da un esempio.

Esempio 2.1.1 Un sistema di comunicazione consiste di 4 antenne “apparentemente” identiche e allineate di cui 2 sono difettose. Il sistema funziona, cioè è in grado di ricevere tutti i segnali che arrivano, se non si susseguono (consecutivamente) 2 antenne difettose. Se le 4 antenne sono allineate *in modo casuale*,

1. qual è la probabilità che il sistema funzioni?
2. Qual è la probabilità che il sistema NON funzioni?

Cosa significa “allineate in modo casuale”? Cosa significano le domande 1. e 2.? Abbiamo bisogno di una terminologia, abbiamo bisogno di modellizzare.

Nell’Esempio 2.1.1 l’esperimento aleatorio consiste nell’allineare le 4 antenne. Con la convenzione che 1 indica un’antenna funzionante e 0 un’antenna difettosa, tutte le possibili configurazioni sono descritte dal seguente insieme:

$$S = \{(1, 1, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1), (0, 1, 1, 0), (0, 1, 0, 1), (0, 0, 1, 1)\}$$

¹Si veda Capitolo 4 di Ross 2008, *Introduzione alla Statistica*, Apogeo, e Capitolo 1 in Epifani, I., Ladelli, L.M. e Posta, G. (2006) *Appunti per il corso di Calcolo delle Probabilità*, AA 2005/2006 <http://www1.mate.polimi.it/~ileepi/dispense/0506CP/>

Ciascun elemento di S , che rappresenta ciascuna possibile configurazione, si chiama *evento elementare* e l'insieme S viene detto *spazio campionario* o *spazio degli esiti* o *spazio degli eventi elementari* relativo all'esperimento aleatorio. Indichiamo ogni evento elementare con la lettera greca minuscola ω .

Procediamo ora a vedere come rappresentare eventi del tipo “il sistema funziona”. Osserviamo l'evento “il sistema funziona” se e solo se si verifica una delle possibili configurazioni $(1, 0, 1, 0)$ o $(0, 1, 1, 0)$ o $(0, 1, 0, 1)$. Allora, rappresentiamo l'evento “il sistema funziona” con il sottoinsieme di S dato da $E = \{(1, 0, 1, 0), (0, 1, 1, 0), (0, 1, 0, 1)\}$.

In logica, l'evento è una proposizione della logica matematica suscettibile di assumere solo due valori: vero o falso. La scelta che facciamo in calcolo delle probabilità è di rappresentare un evento relativo a un esperimento come un sottoinsieme dello spazio campionario. Cioè: un *evento* A è un sottoinsieme dello spazio campionario ($A \subseteq S$) e si verifica se l'esito dell'esperimento aleatorio appartiene ad A .

Da qui la corrispondenza fra operatori logici e operazioni insiemistiche. In particolare, abbiamo il seguente parallelo:

operatore logico	operazione insiemistica
“o”	\cup (unione)
“e”	\cap (intersezione)
“non”	c (complementare)

Essendo gli eventi sottoinsiemi di S , essi costituiscono una *famiglia* o *collezione* di sottoinsiemi di Ω , che –se mai ci servirà– indicheremo con la lettera calligrafica \mathcal{F} (questo significa che se $E \in \mathcal{F}$ allora $E \subseteq \Omega$).

Se abbiamo due o più eventi e operiamo con \cup, \cap o complementazione (o con operazioni composte da queste elementari) otterremo dei nuovi eventi. Abbiamo così:

A sottoevento di B : se $A \subseteq B$;

evento unione: $A \cup B$ è l'evento che si verifica se si verifica A o B o entrambi;

evento intersezione: $A \cap B$ è l'evento che si verifica se si verificano entrambi A e B ;

evento complementare (o contrario): A^c è l'evento che si verifica se e solo se A non si verifica;

evento certo: è l'evento che si verifica sempre, quindi l'evento certo è S ;

l'evento impossibile (o nullo) non si verifica mai, quindi \emptyset = evento nullo;

A, B sono **eventi incompatibili (o mutuamente escludentesi)** se non si verificano mai contemporaneamente, cioè se $A \cap B = \emptyset$;

evento differenza: $A \setminus B = A \cap B^c$ è l'evento che si verifica se si verifica A ma non B .

Inoltre, gli eventi unione, intersezione e contrario godono delle proprietà elencate in Tabella 2.1.

Notate che le operazioni di unione, intersezioni così come le leggi di de Morgan possono essere estese a un numero qualunque di eventi (anche non finito).

Proprietà	unione	intersezione
commutativa	$A \cup B = B \cup A$	$A \cap B = B \cap A$
associativa	$A \cup (B \cup C) = (A \cup B) \cup C$	$A \cap (B \cap C) = (A \cap B) \cap C$
distributiva	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
di inclusione	$A \subseteq B$ se e solo se $A \cup B = B$	$A \subseteq B$ se e solo se $A \cap B = A$
	$A \cup \Omega = \Omega$	$A \cap \Omega = A$
	$A \cup \emptyset = A$	$A \cap \emptyset = \emptyset$
	$A \cup A = A$	$A \cap A = A$
	$A \cup A^c = \Omega$	$A \cap A^c = \emptyset$
Leggi di De Morgan	$(A \cup B)^c = A^c \cap B^c$	$(A \cap B)^c = A^c \cup B^c$

Tabella 2.1: Alcune proprietà di unione, intersezione e contrario

Infine, osserviamo che gli spazi campionari possono essere finiti, numerabili, continui, come esemplificato negli esempi che seguono.

Esempio 2.1.2 Si consideri l'esperimento aleatorio: "Giuseppe lancia due dadi, uno rosso l'altro blu, e osserva i numeri che compaiono sulle facce superiori". In questo caso i risultati possibili sono tutte le coppie ordinate di numeri interi tra uno e sei. Uno spazio degli eventi elementari è

$$S := \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), (2, 2), \dots, (2, 6), \dots, (6, 1), (6, 2), \dots, (6, 6)\} = \\ = \{(r, b) : r = 1, 2, \dots, 6; b = 1, 2, \dots, 6\}$$

dove il generico evento elementare (r, b) in S rappresenta il risultato "è uscito il numero r sul dado rosso e b sul dado blu".

Gli Esempi 2.1.1 e 2.1.2 sono esempi di *spazio campionario finito*.

Esempio 2.1.3 Si consideri l'esperimento aleatorio consistente nel lanciare una moneta equilibrata fino a quando non si presenta testa e si annotino quanti lanci sono necessari. Il risultato dell'esperimento casuale può essere un qualunque numero naturale $1, 2, \dots$; quindi per spazio campionario si può scegliere $S = \{1, 2, \dots, \infty\}$. Questo è un esempio di *spazio campionario discreto, numerabile*.

Esempio 2.1.4 Si consideri l'esperimento aleatorio che consiste nell'osservare il tempo in secondi che intercorre tra l'inizio del funzionamento di un componente di un circuito ed il suo primo guasto (*tempo di vita del componente*). Il risultato dell'esperimento casuale può essere un qualsiasi numero reale non negativo; pertanto, per spazio campionario possiamo scegliere $S := [0, +\infty) = \mathbb{R}^+$. Questo è un esempio di *spazio campionario continuo*.

Ora che sappiamo cos'è un evento, costruiamo la probabilità.

Definizione 2.1.5 Una *probabilità* su S è una funzione che a ogni evento A associa un numero $P(A)$ che rispetta i seguenti assiomi:

1. $P(A) \geq 0$;
2. $P(S) = 1$;

3. se A, B sono eventi disgiunti, cioè $A \cap B = \emptyset$, allora $P(A \cup B) = P(A) + P(B)$.

Gli assiomi che definiscono la probabilità sono assolutamente naturali. L'assioma 1. ci dice che la probabilità associa ad ogni evento un numero non negativo che interpretiamo come la sua probabilità di accadere. Scopriremo che l'assioma 2. ci dice semplicemente che attribuiamo all'evento certo S il valore massimo che può assumere la probabilità. Infine, l'assioma 3. esprime il fatto che se due eventi A, B non possono verificarsi simultaneamente, allora la probabilità dell'evento "almeno uno di due si verifica" è dato dalla somma delle singole probabilità di A e B . L'assioma 3. vale anche quando anziché due eventi incompatibili si consideri un numero n qualunque di eventi incompatibili. Anzi, deve valere anche quando addirittura si consideri un'infinità numerabile di eventi incompatibili. L'assioma 3. prende il nome di *additività*.

Esempio 2.1.6 (Esempio 2.1.1. Continuazione) Riprendiamo l'Esempio 2.1.1 e probabilizziamo $S = \{(1, 1, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1), (0, 1, 1, 0), (0, 1, 0, 1), (0, 0, 1, 1)\}$ con la probabilità

$$P(A) = \frac{\text{n° di casi elementari in } A}{6}$$

Effettivamente, $P(A) \geq 0$, $P(S) = 6/6 = 1$ e, quando A, B non hanno elementi in comune, si ha:

$$\begin{aligned} P(A \cup B) &= \frac{\text{n° di casi elementari in } A \cup B}{6} = \\ &= \frac{\text{n° di casi elementari in } A}{6} + \frac{\text{n° di casi elementari in } B}{6} = P(A) + P(B). \end{aligned}$$

Con questa scelta di P , che scopriremo fra breve non è strampalata, otteniamo:

1. P ("il sistema funziona") = $P(\{(1, 0, 1, 0), (0, 1, 1, 0), (0, 1, 0, 1)\}) = 3/6 = 0.5$;
2. P ("il sistema NON funziona") = $P(\{(1, 1, 0, 0), (1, 0, 0, 1), (0, 0, 1, 1)\}) = 3/6 = 0.5$.

2.1.1 Proprietà della probabilità

Un'immediata conseguenza delle proprietà che una probabilità deve rispettare sono le seguenti regole di calcolo delle probabilità.

Proposizione 2.1.7 (a) $P(A^c) = 1 - P(A)$ (probabilità del complementare);

(b) $P(\emptyset) = 0$ (probabilità dell'evento impossibile);

(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, per ogni evento A, B (regola dell'addizione della probabilità dell'unione);

(d) se A è un sottoevento di B , cioè $A \subseteq B$ allora $P(B \setminus A) = P(B) - P(A)$ e quindi $P(A) \leq P(B)$ (regola delle differenze e monotonia);

(e) $P(A) \leq 1$, per ogni evento A .

Dimostrazione

(a) Notiamo che $S = A \cup A^c$ e $A \cap A^c = \emptyset$ e quindi per gli assiomi 1. e 3.: $1 = P(S) = P(A) + P(A^c)$ che implica $P(A^c) = 1 - P(A)$.

(b) Scegliamo nel punto precedente $A = S$; quindi $P(\emptyset) = P(S^c) = 1 - P(S) = 1 - 1 = 0$.

(c) Vale che $A \cup B = A \cup (B \setminus A)$ e gli eventi A e $B \setminus A$ sono disgiunti. Applicando ripetutamente la proprietà 3. della probabilità per eventi disgiunti otteniamo

$$\begin{aligned} P(A \cup B) &= P(A) + P(B \setminus A) \\ &= P(A) + \boxed{P(B \setminus A) + P(A \cap B)} - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \quad [\text{in quanto } (B \setminus A) \cup (A \cap B) = B]. \end{aligned}$$

(d) se $A \subseteq B$ allora $B = (B \setminus A) \cup A$ e gli eventi $B \setminus A$ e A sono disgiunti. Quindi per la proprietà 3. $P(B) = P(B \setminus A) + P(A)$ da cui $P(B \setminus A) = P(B) - P(A)$. Inoltre, siccome tutte le probabilità sono non negative, allora $P(B \setminus A) \geq 0$ e quindi $P(B) - P(A) \geq 0$.

(e) Poiché $A \subseteq S$, allora, per il punto precedente abbiamo $P(A) \leq P(S) = 1$. ■

Esempio 2.1.8 Relativamente alla prima sessione d'esame del primo anno del corso di laurea XXX è noto che la probabilità che uno studente superi:

- l'esame A è 0.4,
- l'esame B è 0.5,
- l'esame C è 0.3,
- l'esame A e l'esame B è 0.35,
- l'esame A e l'esame C è 0.2,
- l'esame B e l'esame C è 0.25,
- tutti e tre esami è 0.15,

Determinare la probabilità che nella prima sessione uno studente scelto a caso

1. non superi l'esame A;
2. superi A ma non superi B;
3. superi almeno un esame;
4. non superi alcun esame.

Soluzione Indichiamo con A l'evento "lo studente supera l'esame A", con B l'evento "lo studente supera l'esame B" e con C l'evento "lo studente supera l'esame C". Allora le probabilità richieste sono:

1. $P(A^c) = 1 - P(A) = 0.6$;
2. $P(A \cap B^c) = P(A \setminus (A \cap B)) = P(A) - P(A \cap B) = 0.4 - 0.35 = 0.05$;
3. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - [P(A \cap B) + P(A \cap C) + P(B \cap C)] + P(A \cap B \cap C) = 0.4 + 0.5 + 0.3 - 0.35 - 0.2 - 0.25 + 0.15 = 0.55$;
4. $P(A^c \cap B^c \cap C^c) = P((A \cup B \cup C)^c) = 1 - 0.55 = 0.45$. ■

2.2 Probabilità in spazi campionari finiti o numerabili

In questa sezione vediamo come si costruiscono modelli probabilistici per esperimenti aleatori che hanno un numero finito o al più numerabile di esiti possibili, cioè, vediamo come probabilizzare uno spazio campionario finito o numerabile.

Supponiamo $S = \{\omega_1, \omega_2, \dots, \omega_N\}$, con N che può essere un numero finito o $N = \infty$. Assegniamo ora dei pesi p_1, p_2, \dots non negativi a somma 1, cioè tali che

$$p_k \geq 0 \text{ per ogni } k = 1, 2, \dots$$

e

$$\sum_{k=1}^N p_k = 1$$

e interpretiamo questi pesi come le probabilità degli eventi elementari: $p_1 = P(\{\omega_1\})$, $p_2 = P(\{\omega_2\})$ e così via. Allora, necessariamente, la probabilità di ogni evento $A \subseteq S$ risulta automaticamente data da

$$P(A) = \sum_{k: \omega_k \in A} P(\{\omega_k\}) = \sum_{k: \omega_k \in A} p_k \quad (2.2.1)$$

È immediato verificare che la funzione di probabilità P definita in (2.2.1) è una probabilità su S .

Esempio 2.2.1 Lancio un dado a 10 facce truccato con una faccia numerata 1, due facce numerate 2, tre facce numerate 3 e quattro facce numerate 4. Allora $S = \{1, 2, 3, 4\}$ e, per ragioni di simmetria, è sensato scegliere i seguenti pesi: $p_1 = 1/10$, $p_2 = 2/10$, $p_3 = 3/10$, $p_4 = 4/10$. La probabilità dell'evento $A = \text{"esce un numero dispari"}$ è data da: $P(A) = P(\{1, 3\}) = 1/10 + 3/10 = 4/10 = 0.4$.

2.2.1 Esperimenti con esiti equiprobabili

Consideriamo ora un esperimento aleatorio che ammette solo un numero finito N di risultati possibili con spazio campionario $S = \{\omega_1, \omega_2, \dots, \omega_N\}$, $N < \infty$. Supponiamo che la natura dell'esperimento aleatorio ci suggerisca di ritenere tutti i possibili esiti equiprobabili, cioè $p_1 = p_2 = \dots = p_N = p$. In questo caso si parla di *modello uniforme* e si dice che l'esito dell'esperimento è *scelto casualmente*. Poiché $P(S) = 1$ e

$$P(S) = p_1 + \dots + p_N = N \times p,$$

allora necessariamente ogni peso deve essere preso pari a $1/N$, cioè al reciproco della numerosità dello spazio campionario. Inoltre, la probabilità di ogni evento A è data da

$$P(A) = \sum_{k: \omega_k \in A} p_k = \sum_{k: \omega_k \in A} \frac{1}{N} = \frac{n^o \text{ degli elementi in } A}{N}$$

Solitamente scriviamo

$$P(A) = \frac{n^o \text{ casi favorevoli a } E}{n^o \text{ casi possibili}} \quad (2.2.2)$$

La probabilità P definita in (2.2.2) è la definizione classica/frequentista di probabilità.

Esempio 2.2.2 (segue Esempio 2.1.2) Consideriamo ancora l'esempio del lancio di due dadi. In questo caso lo spazio degli eventi elementari è $S = \{(i, j) : i, j = 1, 2, \dots, 6\}$. Per quanto riguarda l'assegnazione di una probabilità P su S osserviamo che se assumiamo che i due dadi non siano truccati e vogliamo che il nostro modello di probabilità modellizzi questo fatto fisico, dobbiamo ammettere che tutti gli eventi elementari di S abbiano la stessa probabilità $p = 1/36$. Sia E_k l'evento "la somma dei due dadi è k " per $k = 2, 3, \dots, 12$. Allora,

$$\begin{aligned} E_2 &= \{(1, 1)\} \\ E_3 &= \{(1, 2), (2, 1)\} \\ E_4 &= \{(1, 3), (2, 2), (3, 1)\} \\ E_5 &= \{(1, 4), (2, 3), (3, 2), (4, 1)\} \\ E_6 &= \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\} \\ E_7 &= \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} \\ E_8 &= \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} \\ E_9 &= \{(3, 6), (4, 5), (5, 4), (6, 3)\} \\ E_{10} &= \{(4, 6), (5, 5), (6, 4)\} \\ E_{11} &= \{(5, 6), (6, 5)\} \\ E_{12} &= \{(6, 6)\} \end{aligned}$$

Applicando la formula $P(E) = (n^\circ \text{ casi favorevoli ad } E)/36$ otteniamo: $P(E_2) = P(E_{12}) = 1/36$, $P(E_3) = P(E_{11}) = 1/18$, $P(E_4) = P(E_{10}) = 1/12$, $P(E_5) = P(E_9) = 1/9$, $P(E_6) = P(E_8) = 5/36$, $P(E_7) = 1/6$.

Esempio 2.2.3 Estraggo a caso una carta da un mazzo di 52 carte (13 valori per ognuno di 4 semi). Allora i casi possibili sono 52, e se il gioco non è truccato, è naturale assegnare a ogni possibile carta pescata probabilità $1/52$. Segue che

$$\begin{aligned} P(\text{"un asso"}) &= \frac{4}{52} = \frac{1}{13} \\ P(\text{"una carta diversa da asso"}) &= 1 - P(\text{"un asso"}) = \frac{12}{13} \\ P(\text{"una carta di picche"}) &= \frac{13}{52} = \frac{1}{4} \\ P(\text{"un asso di picche"}) &= \frac{1}{52} \end{aligned}$$

2.3 Regole di conteggio

In un modello uniforme, per calcolare la probabilità di un evento A non serve descrivere esplicitamente gli eventi elementari che compongono A e quelli che compongono S , ma basta contare quanti sono. Da qui la necessità di regole di conteggio generali, quindi di calcolo combinatorio. Noi dobbiamo saper contare negli esempi classici di modelli uniformi della Sezione 2.3.1, perché i problemi che affronteremo si riducono agli schemi di campionamento da urne lì descritti.

2.3.1 Campionamento da urne

Consideriamo l'esperimento aleatorio consistente nel

campionamento di n palline da un'urna che ne contiene M numerate da 1 a M

Per il resto le palline sono indistinguibili.

Possono darsi tre seguenti schemi di estrazione che danno luogo a tre differenti spazi campionari.

1. **Campionamento SENZA ripetizione e interessa l'ordine**, cioè estraiamo una dopo l'altra $n \leq M$ palline dall'urna eliminando di volta in volta la pallina estratta. Allora, l'evento elementare è una stringa ordinata di n palline da M cioè (a_1, \dots, a_n) con tutte le componenti diverse e lo spazio campionario è l'insieme di queste stringhe, altrimenti dette *disposizioni semplici di n elementi estratti fra M* . Queste stringhe sono nel numero di

$$M(M-1) \cdots (M-n+1) = \frac{M!}{(M-n)!}$$

Per provare questo risultato, generalizzate il ragionamento seguito nella soluzione del punto 1. dell'Esempio 2.3.1.

Se $n = M$:

$$\text{le permutazioni di } \{1, \dots, M\} \text{ sono } \frac{M!}{(M-M)!} = M!$$

Esempio 2.3.1 10 velocisti partecipano a una gara. Apparentemente tutti hanno la stessa preparazione e la stessa chance di salire sul podio. Vincono i primi tre classificati.

1. In quanti diversi modi può concludersi la gara?
2. Qual è la probabilità che i velocisti che indossano le maglie 7 e 10 salgano sul podio?
3. Qual è la probabilità che il velocista con la maglia 10 salga sul podio?

Soluzione

1. La prima posizione sul podio può essere ricoperta da uno dei 10 velocisti, per ogni possibile velocista che arriverà primo possono essercene 9 diversi che arrivano secondi e per ognuna delle possibili 10×9 coppie di vincitori classificati al primo e secondo posto possono esserci 8 diversi velocisti classificati al terzo. Quindi la gara può concludersi in $10 \times 9 \times 8 = 720$ modi.

$$2. P(\text{"i velocisti con la maglia 7 e 10 salgano sul podio"}) = \frac{3 \times 2 \times 8}{720} = \frac{1}{15}.$$

$$3. P(\text{"il velocista con la maglia 10 sale sul podio"}) = \frac{3 \times 9 \times 8}{720} = 0.3. \blacksquare$$

2. **Campionamento SENZA ripetizione e NON interessa l'ordine**. Ogni esito di questo esperimento si può descrivere come un sottoinsieme di n palline estratte da un insieme di M . Quindi tutti i possibili casi elementari sono tanti quanti tutti i sottoinsiemi di $\{1, \dots, M\}$ numerosi n , cioè tanti quante le *combinazioni di classe n* di $\{1, \dots, M\}$. Questo numero è dato da $\binom{M}{n}$. Infatti, se permuti gli n elementi di un generico sottoinsieme $C \subseteq \{1, \dots, M\}$, ottengo $n!$ stringhe ordinate di elementi di C . Quindi il numero, chiamiamolo y , che sto cercando soddisfa $n! \times y = \frac{M!}{(M-n)!}$ da cui $y = \binom{M}{n}$.

Esempio 2.3.2 Se una persona gioca a poker con un mazzo di 32 carte, così ripartite: quattro semi \heartsuit , \diamondsuit , \clubsuit e \spadesuit , per ognuno dei quali si hanno le 8 carte distinte: $A, K, Q, J, 10, 9, 8, 7$,

1. in quanti modi può essere servito?

Ogni mano è un insieme di 5 carte scelte dal mazzo. Allora il numero di mani possibili è $\binom{32}{5} = 201376$.

2. Qual è la probabilità che il giocatore abbia un tris “servito”?

Sia A l'evento: “il giocatore ha un tris servito (e non un gioco migliore)”. Allora $P(A) = (n^0 \text{ di casi favorevoli ad } A) / \binom{32}{5}$. Per calcolare la numerosità di A , scegliamo il valore del tris (Es. tris di K) tra gli 8 disponibili, per ciascuna scelta abbiamo $\binom{4}{3}$ modi di scegliere i semi delle carte che compongono il tris (Es. \heartsuit , \diamondsuit e \clubsuit): in totale abbiamo $8 \times \binom{4}{3}$ modi di scegliere il tris. Ora dobbiamo prendere le rimanenti 2 carte. I valori di queste carte devono necessariamente essere differenti tra di loro (altrimenti avremmo un “full”) e differenti dal valore precedentemente scelto per il tris (altrimenti avremmo un “poker”), abbiamo quindi $\binom{7}{2}$ modi di scegliere i valori delle rimanenti 2 carte². Rimangono da decidere i semi delle 2 carte: per ciascuna carta abbiamo 4 semi possibili. In definitiva $|A| = 8 \times \binom{4}{3} \times \binom{7}{2} \times 4 \times 4$ e la probabilità del tris servito è

$$\frac{8 \times \binom{4}{3} \times \binom{7}{2} \times 4 \times 4}{\binom{32}{5}} = \frac{48}{899} \simeq 0.0534 \simeq 5.3\%$$

3. Campionamento CON ripetizione e interessa l'ordine. L'esperimento questa volta consiste nell'estrarre una pallina dalla solita urna, registrarne il numero prima di procedere alla successiva estrazione rimettere la pallina nell'urna. Quindi ripetiamo n volte le estrazioni secondo questo schema. In questo caso n può essere un numero naturale qualunque. Ogni caso elementare si può rappresentare come una stringa ordinata di lunghezza n le cui componenti possono ripetersi. Altrimenti detto, lo spazio campionario è l'insieme di tutte le disposizioni con ripetizione di M elementi di ordine n . Segue che questo spazio campionario contiene M^n elementi.

Esempio 2.3.3 Scegliendo a caso 5 lettere dall'alfabeto italiano (costituito da 5 vocali e 16 consonanti),

- Qual è la probabilità di comporre una parola che contiene una sola lettera “a”?
- Qual è la probabilità di comporre una parola di 5 vocali?
- Qual è la probabilità di comporre la parola “esame”?

Si risponda alle precedenti domande nelle diverse due ipotesi:

- le lettere possono essere ripetute,
- ogni lettera può essere usata una sola volta.

N.B. Vengono contate anche le parole di senso non compiuto!

Soluzione Sotto (i): a) $\frac{5 \times 20^4}{21^5}$; b) $(5/21)^5$; c) 21^{-5} ; Sotto (ii): a) $\frac{5 \times 20! / (20 - 4)!}{(21)_5} = 5/21$;

b) $1 / \binom{21}{5}$; c) 0] ■

²7 sono i valori disponibili e ne scegliamo 2 senza ripetizione e senza tenere conto dell'ordine

2.4 Probabilità condizionata ed indipendenza

Qualche volta interessa valutare la probabilità di un evento A avendo qualche informazione parziale sull'esito dell'esperimento, per esempio sapendo che si è verificato B . Per far ciò usiamo la nozione di probabilità condizionata di A dato B . Partiamo da un esempio.

Esempio 2.4.1 Supponiamo vengano lanciati un dado rosso e uno blu e supponiamo che ci venga chiesto di calcolare la probabilità che la somma dei due dadi sia 12. È facile verificare che questa probabilità vale $1/36$. Rispondiamo ora alla stessa domanda ma sapendo che sul dado rosso è uscito 6. Questa ulteriore informazione cambia radicalmente le nostre valutazioni. Infatti, se sappiamo che sul dado rosso è uscito un 6, la probabilità che la somma dei due dadi dia 12 è uguale alla probabilità che sia uscito 6 anche sul dado blu, cioè $1/6$.

Come calcolare in generale una probabilità condizionata quando il ragionamento euristico non aiuta? Ragioniamo: siamo arrivati a $1/6$ lavorando nel seguente modo: l'ulteriore informazione "si è verificato 6 sul dado rosso" ha ristretto lo spazio campionario di tutte le 36 coppie alle coppie che hanno prima componente uguale a 6. Queste 6 coppie sono ora da considerare come i soli casi possibili. I casi favorevoli sono invece le coppie tali che la somma è 12 e la componente rossa è 6, cioè $(6, 6)$. Quindi:

$$P(A|B) = \frac{1}{6} = \frac{1/36}{6/36} = \frac{P(A \cap B)}{P(B)}$$

Definizione 2.4.2 (Probabilità condizionata) Se $P(B) > 0$, si chiama *probabilità condizionata di A dato B* il numero

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

Esercizio 2.4.3 Un lotto è costituito da 25 transistor accettabili, 10 parzialmente difettosi (cioè che si rompono dopo qualche ora d'uso) e 5 difettosi (cioè che si guastano immediatamente). Un transistor viene preso a caso dal lotto. Se non si rompe subito qual è la probabilità che sia accettabile?

Soluzione In questo caso abbiamo tre eventi A "il transistor è accettabile", B "il transistor è parzialmente difettoso", C "il transistor è difettoso". Ci viene chiesto di calcolare $P(A|C^c)$. Abbiamo che:

$$P(A|C^c) = \frac{P(A \cap C^c)}{P(C^c)} = \frac{P(A)}{1 - P(C)} = \frac{25/40}{35/40} = \frac{5}{7} \blacksquare$$

Esempio 2.4.4 Fate vedere che $P(A|B) = 1$ se $B \subseteq A$ e che $P(A|B) = 0$ se $A \cap B = \emptyset$.

Osservazione 2.4.5 Se pensiamo l'evento condizionante "a destra" B fissato, e leggiamo $P(A|B)$ come funzione dell'evento condizionato "a sinistra" A , allora $A \mapsto P(A|B)$ è una funzione di probabilità cioè

1. $P(A|B) \geq 0$;
2. $P(S|B) = 1$;

3. Se $A_1 \cap A_2 = \emptyset$ allora $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$.

Osservazione 2.4.6 L'evento " $A|B$ " non esiste. Ma, se $P(B) > 0$, esiste la probabilità dell'evento A valutata nell'ipotesi che B si è verificato, cioè $P(A|B)$.

Osservazione 2.4.7 Attenzione a distinguere fra $P(A|B)$ e $P(A \cap B)$: sono probabilità diverse! Infatti nell'esempio del lancio di due dadi dove A = "somma=12" e B = "dado rosso=6", abbiamo ottenuto: $P(A \cap B) = 1/36$ e $P(A|B) = 1/6$.

Esempio 2.4.8 Un'inchiesta sulla popolazione della città xxx ha fornito i seguenti dati: il 10% della popolazione è ricco (R), il 5% è famoso (F) e il 3% è ricco e famoso. Per un cittadino di xxx scelto a caso,

- (a) qual è la probabilità che sia ricco ma non famoso?
- (b) Per un cittadino NON famoso, qual è la probabilità di essere ricco?
- (c) Per un cittadino famoso, qual è la probabilità di essere ricco?

Soluzione 2.4.8 Poiché $P(R) = 0.1$, $P(F) = 0.05$ e $P(R \cap F) = 0.03$, allora

- (a) $P(R \cap F^c) = P(R) - P(R \cap F) = 0.1 - 0.03 = 0.07 = 7\%$;
- (b) $P(R|F^c) = \frac{P(R \cap F^c)}{P(F^c)} = \frac{0.07}{1-0.05} = \frac{7}{95} \approx 0.07368$;
- (c) $P(R|F) = \frac{P(R \cap F)}{P(F)} = \frac{0.03}{0.05} = 60\%$. ■

2.4.1 Alcune formule importanti

Riuniamo in questa sezione alcune formule utili nelle applicazioni e che coinvolgono il concetto di probabilità condizionata.

Regola del prodotto

La probabilità condizionata può essere usata strumentalmente per calcolare probabilità non condizionate di eventi.

Come conseguenza immediata della definizione di probabilità condizionata $P(A|B) = P(A \cap B)/P(B)$ – abbiamo la *regola del prodotto*:

$$P(A \cap B) = P(A|B)P(B) \quad (2.4.1)$$

Effettivamente, a volte siamo interessati a calcolare $P(A \cap B)$ ma è più semplice determinare $P(B)$ e $P(A|B)$ direttamente. Applicando la regola del prodotto otteniamo $P(A \cap B)$.

La regola del prodotto può essere generalizzata a un numero arbitrario di eventi, applicando la Formula (2.4.1) ripetutamente. Per esempio, se abbiamo tre eventi A_1, A_2, A_3 allora³

$$P(A_1 \cap A_2 \cap A_3) = P(A_3|A_1 \cap A_2) \times P(A_2|A_1)P(A_1) .$$

Se, invece, ne abbiamo un generico numero $n \geq 2$ la formula del prodotto⁴ diventa:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1) \cdot \dots \cdot P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1})$$

³Nella formula stiamo considerando eventi tali che $P(A_1 \cap A_2) > 0$

⁴Nella formula stiamo considerando eventi tali che $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$

Esempio 2.4.9 Estraiamo in sequenza e senza ripetizione delle biglie da un'urna che inizialmente ne contiene 5 rosse e 7 bianche.

1. Qual è la probabilità che la prima biglia estratta sia rossa e la seconda bianca?

Possiamo procedere come segue. Siano B_k l'evento "la k -esima biglia estratta è bianca" ed R_k l'evento "la k -esima biglia estratta è rossa". La probabilità richiesta è

$$P(R_1 \cap B_2) = P(B_2|R_1)P(R_1) = \frac{7}{5+7-1} \times \frac{5}{5+7}$$

2. Qual è la probabilità che la prima biglia estratta sia rossa, la seconda bianca, la terza rossa e la quarta ancora bianca?

La soluzione è data da:

$$\begin{aligned} P(R_1 \cap B_2 \cap R_3 \cap B_4) &= P(R_1)P(B_2|R_1)P(R_3|R_1 \cap B_2)P(B_4|R_1 \cap B_2 \cap R_3) = \\ &= \frac{5}{5+7} \times \frac{7}{5+7-1} \times \frac{5-1}{5+7-2} \times \frac{7-1}{5+7-3} \end{aligned}$$

Formula di Bayes

Ci poniamo il seguente problema: Conosciamo $P(B)$ e $P(B|H)$ e sappiamo che $P(B) > 0$. Siamo in grado di determinare $P(H|B)$? La risposta è positiva e la otteniamo ragionando come segue:

$$P(H|B) = \frac{P(H \cap B)}{P(B)} = \frac{P(B \cap H)}{P(B)} = \frac{P(B|H)P(H)}{P(B)}$$

Abbiamo così provato la *Formula di Bayes*:

Se un evento B è tale che $P(B) > 0$ allora si ha $P(H|B) = \frac{P(B|H)P(H)}{P(B)}$

Formula delle probabilità totali

Spesso calcolare direttamente una probabilità $P(B)$ è operazione complessa, ma sono facili da calcolare la probabilità condizionata di B dato il verificarsi o meno di un evento H . Combinando le probabilità condizionate $P(B|H), P(B|H^c)$ con la probabilità $P(H)$ mediante la *formula delle probabilità totali* si riesce ad ottenere $P(B)$:

se $P(H) > 0$ allora $P(B) = P(B|H)P(H) + P(B|H^c)(1 - P(H))$ per ogni B

Infatti

$$B = (B \cap H) \cup (B \cap H^c)$$

e quindi

$$\begin{aligned} P(B) &= P(B \cap H) + P(B \cap H^c) = P(B|H)P(H) + P(B|H^c)P(H^c) \\ &= P(B|H)P(H) + P(B|H^c)(1 - P(H)) . \end{aligned}$$

Esempio 2.4.10 Ci sono due urne A e B . L'urna A contiene 1000 biglie bianche e 1 nera mentre la B ne contiene 2 nere. Si lancia una moneta equa e, se viene testa, si pesca una biglia dall'urna A mentre se viene croce si pesca una biglia dall'urna B .

1. Qual è la probabilità che la biglia pescata sia nera?

Poniamo H “esce testa” e B “viene pescata una biglia nera”. Allora H^c = “esce croce”, $P(H) = 1/2$, $P(B|H) = 1/1001$ e $P(B|H^c) = 1$. Dalla formula delle probabilità totali deriva

$$P(B) = P(B|H)P(H) + P(B|H^c)(1 - P(H)) = \frac{1}{1001} \times \frac{1}{2} + 1 \times \frac{1}{2} \simeq 0.5$$

Esercizio 2.4.11 (Test di collaudo) [Tratto da Ross, *Probabilità e statistica per l'ingegneria, Apogeo, 2004*] Un'impresa industriale ha installato un sistema automatico per il controllo di qualità, che garantisce che, se un pezzo è difettoso, esso viene eliminato con probabilità 0.995. Tuttavia, c'è una probabilità (piccola) pari a 0.001 che un pezzo non difettoso sia eliminato. Inoltre, si sa anche che la probabilità che un pezzo sia difettoso è 0.2. Si calcoli la probabilità che un pezzo non eliminato dopo il controllo di qualità sia difettoso.

Formula delle probabilità totali; caso generale. Una formula delle probabilità totali vale anche per il caso in cui anziché partizionare l'evento certo S in $H \cup H^c$, abbiamo una partizione finita⁵ di S negli eventi H_1, \dots, H_N , tali che $P(H_k) > 0$ per ogni $k = 1, \dots, N$. Allora per ogni evento B si ha

$$P(B) = \sum_{k=1}^n P(B|H_k)P(H_k)$$

Esercizio 2.4.12 Dimostrare la formula delle probabilità totali per una partizione finita di eventi H_1, H_2, \dots, H_N e per una numerabile F_1, F_2, \dots .

Test diagnostici Una rilevante applicazione della formula delle probabilità totali e di Bayes è data dai test diagnostici. Descriveremo test diagnostici, valutazioni di riconoscitori e curve di Roc nel Capitolo 7.

2.4.2 Indipendenza

In generale, la probabilità condizionata $P(A|B)$ non è uguale alla probabilità non condizionata $P(A)$. Per esempio, nel lancio di due dadi regolari dove A = “somma=12” e B = “dado rosso=6”, abbiamo ottenuto $P(A|B) = 1/6$ e $P(A) = 1/36$. Cioè, alla luce della nuova informazione contenuta in B la probabilità dell'evento A è aumentata. Se succede che $P(A|B) = P(A)$ allora si dice che A è *indipendente da* B . D'altro canto, se A è indipendente da B , allora, per la regola del prodotto, abbiamo

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B)$$

da cui otteniamo

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$$

⁵ H_1, H_2, \dots, H_N è una partizione finita di S se $H_k \neq \emptyset \forall k$, $\bigcup_{k=1}^N H_k = S$ e $H_h \cap H_k = \emptyset$ se $h \neq k$

Cioè, se A è indipendente da B allora B è indipendente da A e quindi la nozione di indipendenza è simmetrica negli eventi A, B .

Diamo la seguente definizione di indipendenza:

Definizione 2.4.13 Due eventi A, B si dicono *indipendenti* se la probabilità dell'intersezione $A \cap B$ fattorizza nelle singole probabilità di A e B : $P(A \cap B) = P(A)P(B)$.

Osservazione 2.4.14 Se A, B sono indipendenti, allora lo sono anche A e B^c , A^c e B , A^c e B^c . Perché? Quale è il significato euristico di questa proprietà?

Osservazione 2.4.15 Come definiamo l'indipendenza fra più di 2 eventi? Analogamente al caso di due: un numero finito e qualunque di eventi sono indipendenti se il realizzarsi di un numero finito di essi non influenza il verificarsi dei rimanenti. Traduciamo in formule, nel caso di tre eventi. Tre eventi A, B, C si dicono indipendenti se valgono *tutte* e quattro le seguenti relazioni:

$$\begin{aligned}P(A \cap B) &= P(A)P(B) \\P(A \cap C) &= P(A)P(C) \\P(B \cap C) &= P(B)P(C) \\P(A \cap B \cap C) &= P(A)P(B)P(C).\end{aligned}$$

Attenzione nessuna delle 4 relazioni implica le altre.

Osservazione 2.4.16 INCOMPATIBILITÀ e INDIPENDENZA sono due proprietà diverse: L'incompatibilità è una proprietà fisica degli eventi, significa: $A \cap B = \emptyset$. Invece, l'indipendenza è una proprietà della probabilità P degli eventi.

Se $A \cap B = \emptyset$ e $P(A) > 0$ e $P(B) > 0$ allora A, B NON possono essere indipendenti.

Esempio 2.4.17 (Esempio 2.2.3. Continuazione) Abbiamo visto che se estraiamo a caso una carta da un mazzo di 52 carte abbiamo

$$\begin{aligned}P(\text{"un asso"}) &= \frac{1}{13} \\P(\text{"una carta di picche"}) &= \frac{1}{4} \\P(\text{"un asso di picche"}) &= \frac{1}{52}\end{aligned}$$

In particolare, "un asso di picche" è l'evento intersezione degli eventi "estraggo un asso" ed "estraggo picche" e $P(\text{"un asso di picche"}) = P(\text{"un asso"})P(\text{"una carta di picche"})$. Concludiamo che gli eventi "estraggo un asso" ed "estraggo picche" sono indipendenti. Non ve l'aspettavate, vero?

2.4.3 Modelli probabilistici basati sull'indipendenza

Affidabilità di sistemi Siano A_1, \dots, A_n i componenti di un sistema S . L'*affidabilità del componente* A_j è la probabilità che il componente funzioni (nel senso che fornisca certe prestazioni in limiti di tempo e condizioni prefissate) e l'*affidabilità di* S è la probabilità che S funzioni.

Se i componenti sono supposti tra loro indipendenti e sono connessi in serie (cioè il sistema funziona se e solo se tutti i componenti funzionano) allora l'affidabilità del sistema è:

$$P(S) = P(A_1) \cdots P(A_n) \quad (\text{Sistema in serie})$$

Se i componenti sono supposti tra loro indipendenti e sono connessi in parallelo (cioè il sistema funziona se e solo se almeno un componente funziona) allora l'affidabilità del sistema è:

$$P(S) = 1 - (1 - P(A_1)) \cdots (1 - P(A_n)) \quad (\text{Sistema in parallelo})$$

Esercizio 2.4.18 ⁶ Calcolate l'affidabilità del sistema S_4 in figura 2.1, costituito da una copia del componente R_1 , una del componente R_2 , tre del componente R_3 , due del componente R_4 e una del componente R_5 , sapendo che i componenti R_1, R_2, R_3, R_4, R_5 sono indipendenti e hanno affidabilità, rispettivamente, 0.95, 0.99, 0.7, 0.75, 0.9.

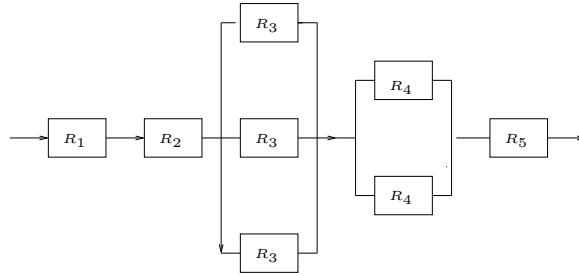


Figura 2.1: Sistema S_4

Prove di Bernoulli Alcuni fenomeni aleatori consistono in prove ripetute in condizioni identiche di uno stesso esperimento che può dare luogo solo a due risultati: successo o fallimento. L'esempio tipico è il lancio di una moneta. Allora è sensato assumere che *i*) tutte le prove siano caratterizzate dalla stessa probabilità di successo e *ii*) ciascuna prova non influenzi nessuna delle altre, cioè i risultati delle prove siano indipendenti. Il modello probabilistico che traduce un esperimento aleatorio di questa natura prende il nome di modello delle *prove di Bernoulli*. Riprenderemo le prove di Bernoulli nella Sezione 3.3.1, trattando delle variabili aleatorie binomiali e geometriche.

⁶Esempio 1.11 in Trivedi, K S. (2002) *Probability and statistics with reliability, queuing, and computer science applications*, 2. ed. Wiley New York.

Capitolo 3

Variabili aleatorie discrete

3.1 Variabili aleatorie

Spesso quando effettuiamo un esperimento aleatorio non ci interessano in dettaglio gli esiti possibili dell'esperimento, quanto dei valori numerici ad essi associati. Per esempio interessano: il tempo di vita T di un componente elettronico, oppure il numero di teste X che si presentano se lanciamo 3 volte una moneta equilibrata da un euro. Queste quantità di interesse sono determinate dal risultato dell'esperimento e si chiamano *variabili aleatorie*.

Cioè, le variabili aleatorie sono *funzioni sullo spazio campionario* S . Per meglio capire questo concetto vediamo un esempio.

Esempio 3.1.1 Lanciamo tre volte una moneta non truccata e sia X il numero di teste che si presentano. Chiaramente X è un numero casuale che può assumere i valori 0, 1, 2, 3 e lo spazio campionario è

$$S = \{(T, T, T), (T, T, C), (T, C, T), (T, C, C), (C, T, T), (C, T, C), (C, C, T), (C, C, C)\}$$

Possiamo pensare alla variabile aleatoria X come a una regola che a ogni terna in S associa il numero di teste che la compongono. Così per ogni valore 0, 1, 2, 3 possiamo calcolare la probabilità che X lo assuma:

$$\begin{aligned}P(X = 0) &= P(\{(C, C, C)\}) = \frac{1}{8} \\P(X = 1) &= P(\{(T, C, C), (C, T, C), (C, C, T)\}) = \frac{3}{8} \\P(X = 2) &= P(\{(T, T, C), (T, C, T), (C, T, T)\}) = \frac{3}{8} \\P(X = 3) &= P(\{(T, T, T)\}) = \frac{1}{8}\end{aligned}$$

Gli *eventi di interesse* associati a una variabile aleatoria X sono *eventi del tipo* $\{X > a\}$, $\{X < a\}$, $\{X \geq a\}$, $\{X \leq a\}$, $\{a < X < b\}$, $\{a \leq X < b\}$, $\{a < X \leq b\}$, $\{a \leq X \leq b\}$, $\{X = a\}$ e quelli ottenuto da questi componendoli con le operazioni di unione, intersezione, negazione. Infatti è sensato che uno si chieda: qual è la probabilità che X sia minore di un certo numero fissato x ? Cioè quanto vale $P(X < x)$? Oppure: qual è la probabilità che X sia maggiore di un certo numero fissato x ? Cioè quanto vale $P(X > x)$? Per rispondere,

osserviamo che se uno conosce $P(X \leq x)$, per ogni $x \in \mathbb{R}$, allora $P(X \geq x)$, piuttosto che $P(a < X \leq b)$ e ognuno di tutti i precedenti eventi sono facilmente calcolabili. Per esempio, per le regole di calcolo delle probabilità abbiamo:

$$P(X > x) = 1 - P(\underbrace{X \leq x}_{\text{complementare}}) = 1 - P(X \leq x)$$

e

$$P(x < X \leq y) = P(\{X \leq y\} \setminus \{X \leq x\}) = P(X \leq y) - P(X \leq x)$$

Quanto precede mostra che se conosciamo la funzione $F_X(x) := P(X \leq x)$, $x \in \mathbb{R}$, possiamo facilmente calcolare la probabilità di eventi associati a X . Per questa ragione alla funzione F_X si dà un nome particolare.

Definizione 3.1.2 (Funzione di ripartizione) La *funzione di ripartizione* F di una variabile aleatoria X è una funzione che a ogni $x \in \mathbb{R}$ associa il numero $F(x) := P(X \leq x)$.

La variabile aleatoria costruita nell'Esempio 3.1.1 assume un numero finito di modalità. Variabili di questo tipo si dicono *discrete*. Esistono poi le variabili aleatorie *continue* (per esempio il tempo di vita T di un componente elettronico). In questo capitolo noi vediamo le variabili aleatorie discrete.

3.2 Variabili aleatorie discrete

Definizione 3.2.1 (Variabili aleatorie discrete) Una variabile aleatoria X è detta *discreta* se può assumere una quantità finita o numerabile di valori: $I = \{x_1, x_2, \dots\}$.

Esempi di variabili aleatorie discrete sono: il numero di volte che bisogna lanciare una moneta prima di ottenere testa, il numero di successi in una sequenza di n prove di Bernoulli, il numero di teste che si ottengono lanciando tre monete (cfr. Esempio 3.1.1). Per una variabile discreta è possibile definire una *densità discreta* nel modo seguente:

Definizione 3.2.2 Sia X una variabile aleatoria discreta che assume i valori $I = \{x_1, x_2, \dots\}$. Allora la funzione $f_X(x_k) := P(X = x_k)$, $x_k \in I$, si chiama *densità discreta* della variabile aleatoria X .

Se f_X è la densità di una variabile aleatoria X a valori in I allora valgono le seguenti proprietà:

1. $0 \leq f_X(x) \leq 1$, in quanto la densità discreta è una probabilità;
2. $\sum_{k \in I} p_X(x_k) = 1$, in quanto necessariamente X deve assumere uno dei valori in I ;
3. $P(X \in B) = \sum_{k: x_k \in B} f_X(x_k)$;
4. se F_X è la funzione di ripartizione di X allora

$$F_X(x) = \sum_{k: x_k \leq x} p_X(x_k) \quad \forall x \in \mathbb{R};$$

5. se i valori che X assume costituiscono una successione di punti separati nella retta dei numeri reali, allora

$$f_X(x_k) = F_X(x_k) - F_X(x_{k-1}), \quad \forall x_k \in I;$$

Il punto 3. ci fa capire a cosa serve la densità discreta: ci permette di calcolare la probabilità che l'evento $\{X \in B\}$ si verifichi effettuando una semplice operazione algebrica.

I punti 4. e 5. mostrano come ottenere la funzione di ripartizione di una variabile aleatoria discreta dalla sua densità e viceversa. In particolare i punti 4. e 5. ci dicono che la funzione di ripartizione di una variabile aleatoria discreta è una funzione “a gradini”, che i gradini sono situati nei valori che X può assumere e che l'altezza del gradino corrispondente al punto $x_k \in I$ è proprio la densità discreta $f_X(x_k)$.

Nota 3.2.3 È interessante la motivazione euristica della parola “densità” utilizzata nel contesto delle variabili aleatorie discrete. Immaginiamo i valori che X può assumere come punti materiali su una retta. Allora attribuendo al generico punto x_k la massa $m_k := f_X(x_k)$, otteniamo una *distribuzione* di masse discrete sulla retta e f_X è proprio la densità di massa.

Esempio 3.2.4 (Segue Esempio 3.1.1) Sia X il numero di teste che si ottengono lanciando tre volte una moneta equa. La densità discreta di X è

$$f_X(x) = P(X = x) = \begin{cases} \frac{1}{8} & \text{se } x \in \{0, 3\} \\ \frac{3}{8} & \text{se } x \in \{1, 2\} \\ 0 & \text{se } x \notin \{0, 1, 2, 3\} \end{cases}$$

Per rappresentare graficamente l'andamento di questa densità usiamo un diagramma a barre. Un *diagramma a barre* è costruito disegnando in corrispondenza di ogni valore x_k una barra perpendicolare all'asse delle ascisse di lunghezza uguale alla densità $f_X(x_k)$, come in Figura 3.1. La funzione di ripartizione di X è

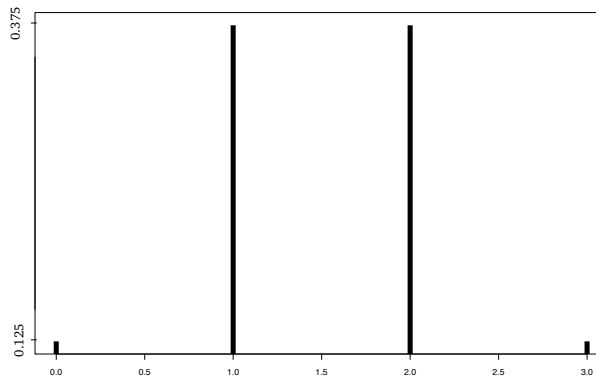


Figura 3.1: Densità f_X dell'Esempio 3.2.4

FUNZIONE DI
RIPARTIZIONE

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ f_X(0) = \frac{1}{8} & \text{se } 0 \leq x < 1 \\ f_X(0) + f_X(1) = \frac{1}{2} & \text{se } 1 \leq x < 2 \\ f_X(0) + f_X(1) + f_X(2) = \frac{7}{8} & \text{se } 2 \leq x < 3 \\ f_X(0) + f_X(1) + f_X(2) + f_X(3) = 1 & \text{se } x \geq 3 \end{cases}$$

Il grafico di F_X è rappresentato in Figura 3.2 (a). Osserviamo le seguenti proprietà generali della funzione di ripartizione di una variabile aleatoria discreta:

- a) A sinistra del valore minimo assunto da X F_X è nulla e a destra del valore più grande è pari a 1;
- b) F_X è funzione continua da destra;
- c) F_X è funzione monotona non decrescente.

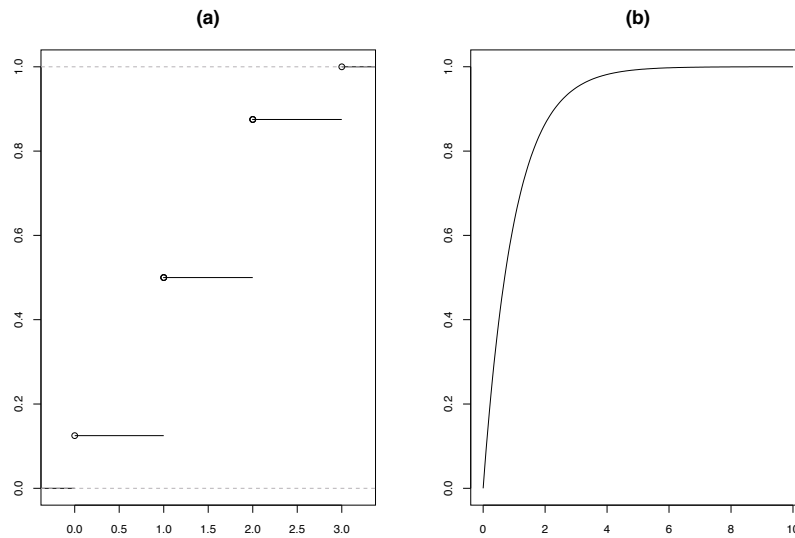


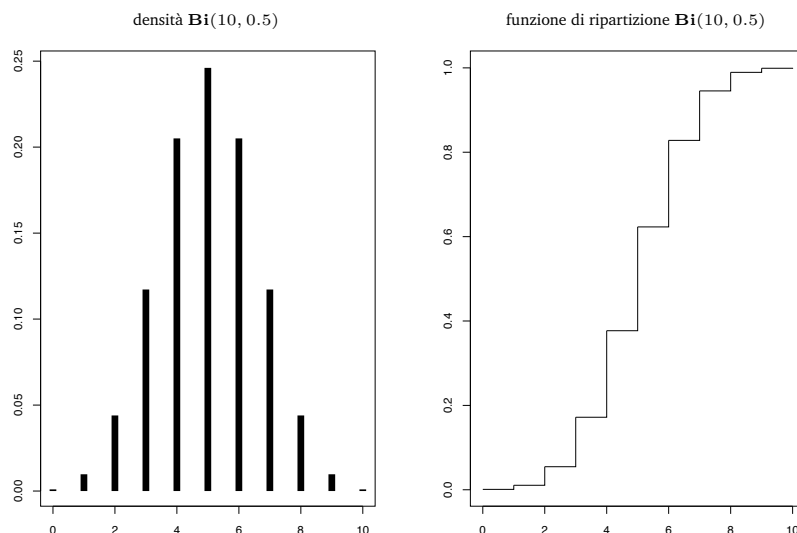
Figura 3.2: (a) f.d.r. F_X dell'Esempio 3.2.4,

3.3 Esempi di densità discrete notevoli

Vediamo ora in dettaglio alcuni esempi di densità discrete che sono importanti per le applicazioni.

3.3.1 Densità binomiale e bernoulliana

Riprendiamo lo schema delle prove di Bernoulli introdotte alla fine della Sezione 2.4.3. Consideriamo un esperimento aleatorio che consiste di n prove di Bernoulli, ciascuna delle quali può dare luogo solo a due risultati: successo o fallimento. Per convenzione usiamo “0” per indicare fallimento e “1” per indicare successo. L'esempio tipico è il lancio di una moneta.

Figura 3.3: $\text{Bi}(10, 0.5)$

Infine introduciamo gli eventi E_1 = “successo alla prima prova”, \dots , E_n = “successo alla n -esima prova”. Se assumiamo che *i)* tutte le prove siano caratterizzate dalla stessa probabilità di successo e *ii)* ciascuna prova non influenzi nessuna delle altre, cioè i risultati delle prove siano indipendenti, allora stiamo assumendo il seguente modello:

$$i) P(E_1) = \dots = P(E_n) = p \in (0, 1)$$

ii) E_1, \dots, E_n sono eventi indipendenti.

Segue dalle due precedenti ipotesi che ogni sequenza di k successi ed $n - k$ fallimenti ha probabilità $p^k(1 - p)^{n-k}$ e che la probabilità di ottenere esattamente k successi in n prove (con $k \leq n$) è

$$\binom{n}{k} p^k (1 - p)^{n-k}$$

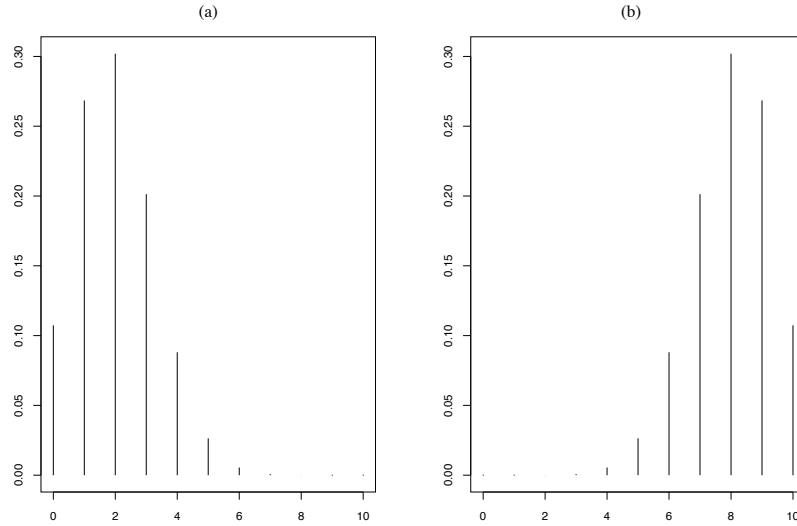
Abbiamo ora tutti gli strumenti per definire la variabile aleatoria discreta X che conta “il numero di successi ottenuti in n prove di Bernoulli”. Si vede subito che X può assumere solo i valori $0, 1, \dots, n$ ed è quindi una variabile aleatoria discreta. Inoltre, per quanto ottenuto qualche riga sopra, la sua densità è

DENSITA' VARIABILI BINOMIALE $f_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}$ se $k = 0, 1, \dots, n$

n = #tentativi
 p = probabilità successo singola prova

e prende il nome di *densità binomiale di parametri n e p* . Equivalentemente si dice che X è una variabile aleatoria *binomiale di parametri n e p* o ancora $X \sim \text{Bi}(n, p)$. La Figura 3.3 fornisce il diagramma a barre della densità ed il grafico della funzione di ripartizione di una variabile aleatoria $X \sim \text{Bi}(10, 0.5)$, mentre la Figura 3.4 mostra, mediante un diagramma a barre, l'andamento delle densità $\text{Bi}(10, 0.2)$ e $\text{Bi}(10, 0.8)$.

Sia $X \sim \text{Bi}(n, p)$; se $n = 1$ questa variabile rappresenta il numero di successi in una sola prova con probabilità di successo p , cioè X assume solo i valori 0 e 1, e la densità di X è

Figura 3.4: (a) $\text{Bi}(10, 0.2)$, (b) $\text{Bi}(10, 0.8)$

$f_X(k) = p^k(1-p)^{1-k}$ se $k = 0, 1$, cioè

$$f_X(k) = \begin{cases} 1-p & \text{se } k = 0 \\ p & \text{se } k = 1 \end{cases} \quad (3.3.1)$$

La densità (3.3.1) prende il nome di *densità bernoulliana di parametro p* ; equivalentemente si dice che X è una *bernoulliana di parametro p* o ancora $X \sim \text{Be}(p)$.

Per ragioni di comodità si dice che la variabile aleatoria costante $X \equiv 1$, cioè la variabile aleatoria che vale sempre 1, è bernoulliana di parametro 1 e che la variabile aleatoria costante $X \equiv 0$, cioè la variabile aleatoria che vale sempre 0, è bernoulliana di parametro 0.

Esempio 3.3.1 Riempendo a caso una schedina di totocalcio, qual è la probabilità di fare almeno 12?

Su una schedina del totocalcio sono elencate 14 partite e ogni partita può avere tre risultati “1”, “2” o “X”, ad indicare rispettivamente la vittoria della squadra ospitante, della squadra ospite o la parità. La probabilità di azzeccare una singola partita, scrivendo a caso uno dei simboli 1, 2 o X, è -almeno in prima approssimazione- uguale ad $1/3$. Inoltre, compilando a caso una schedina, azzeccare o meno il risultato di una certa partita non influenza la capacità di azzeccare le altre. Possiamo quindi schematizzare il nostro esperimento aleatorio con una successione di $n = 14$ prove di Bernoulli, con probabilità di successo nella singola prova $p = 1/3$. Sia Y il numero di partite azzeccate; allora $Y \sim \text{Bi}(14, 1/3)$ e

$$P(Y \geq 12) = \sum_{k=12}^{14} f_Y(k) = \sum_{k=12}^{14} \binom{14}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{14-k} = \frac{393}{4782969} \simeq 0.00008$$

Esercizio 3.3.2 Supponiamo che da un’urna contenente r biglie rosse e b biglie bianche estraiamo a caso una biglia, prendiamo nota del suo colore e la reinseriamo nell’urna. Quindi, ripetiamo questa procedura $n \geq 1$ volte e sia X il numero di biglie rosse estratte nelle n estrazioni. Verificare che $X \sim \text{Bi}(n, r/(r+b))$.

bah. se siete
comodi voi

Osservazione 3.3.3 Se abbiamo una sequenza di n prove tutte con la stessa probabilità di successo, ma non indipendenti, il numero di successi sulle n prove non è più regolato dal modello binomiale, come esemplificato di seguito:

Esercizio 3.3.4 (Modello ipergeometrico) Effettuo 3 estrazioni senza ripetizione da un'urna che contiene 20 palline numerate da 1 a 20. Sia X la variabile aleatoria che conta quante palline estratte esibiscono un numero minore o uguale a 4.

1. Determinate la probabilità di ottenere un numero minore o uguale a 4 alla prima, alla seconda, alla terza estrazione.

2. Determinare la densità discreta di X .

Soluzione

1. $p_1 = p_2 = p_3 = \frac{4}{20} = \frac{1}{5}$; infatti, per esempio, per la formula delle probabilità totali:

$$p_2 = P(S_2|S_1)P(S_1) + P(S_2|S_1^c)P(S_1^c) = \frac{3}{19} \times \frac{1}{5} + \frac{4}{19} \times \frac{4}{5} = \frac{1}{5}.$$

provare a fare
i conti di questa
roba qui

$$2. f_X(k) = \frac{\binom{4}{k} \binom{16}{3-k}}{\binom{20}{3}}, \text{ per } k = 0, 1, 2, 3. \blacksquare$$

3.3.2 Densità Geometrica

La densità geometrica fornisce un modello probabilistico per modellizzare l'istante di guasto X di un'apparecchiatura *non soggetta ad usura* ed inizialmente funzionante, ma che si può guastare per motivi contingenti e il cui funzionamento viene controllato agli tempi $1, 2, \dots$.

Sia X l'istante in cui l'apparecchiatura si guasta. La proprietà di assenza di usura di un'apparecchiatura si può esprimere con le probabilità nel seguente modo:

$$P(X = k + 1 | X > k) = P(X = 1), \quad k = 1, 2, \dots \quad (3.3.2)$$

La proprietà (3.3.2) significa che se un'apparecchiatura è funzionante al tempo $t = k$ allora la probabilità che si rompa nell'istante successivo $t = k + 1$ è uguale alla probabilità che si rompa al tempo $t = 1$. L'identità (3.3.2) è equivalente a

$$P(X > k + 1 | X > k) = P(X > 1), \quad k = 1, 2, \dots$$

e ci permette di determinare la densità di X se conosciamo $q := P(X > 1)$. Infatti:

$$q = P(X > 1) = P(X > k + 1 | X > k) = \frac{P(X > k + 1, X > k)}{P(X > k)} = \frac{P(X > k + 1)}{P(X > k)}$$

da cui $P(X > k + 1) = qP(X > k)$, $k = 1, 2, \dots$. Quindi

$$\begin{aligned} P(X > 2) &= qP(X > 1) = q^2 \\ P(X > 3) &= qP(X > 2) = q^3 \\ &\vdots \\ P(X > k + 1) &= qP(X > k) = q^{k+1} \end{aligned}$$

Segue che $F_X(k) = P(X \leq k) = 1 - P(X > k) = 1 - q^k$ e quindi:

$$P(X = k) = F_X(k) - F_X(k-1) = q^{k-1} - q^k = q^{k-1}(1 - q)$$

La quantità $p := 1 - q = P(X \leq 1)$ si chiama *intensità di guasto* e, in termini di p , possiamo scrivere

$$P(X = k) = p(1 - p)^{k-1}, \quad k = 1, 2, \dots$$

Notiamo che

$$P(X \in \mathbb{N}) = \sum_{k=1}^{+\infty} P(X = k) = \sum_{k=1}^{+\infty} p(1 - p)^{k-1} = p \sum_{k=0}^{+\infty} (1 - p)^k = p \frac{1}{1 - (1 - p)} = 1$$

Quindi X è una variabile aleatoria discreta a valori in \mathbb{N} con densità

$$f_X(k) = \begin{cases} p(1 - p)^{k-1} & \text{se } k = 1, 2, \dots \\ 0 & \text{altrove} \end{cases} \quad (3.3.3)$$

La densità (3.3.3) prende il nome di *densità geometrica di parametro p* , X è detta *variabile geometrica di parametro p* e si scrive $X \sim \text{Geom}(p)$.

Nota 3.3.5 (Proprietà di assenza di memoria) La proprietà (3.3.2) è nota anche come *proprietà di assenza di memoria* e abbiamo dimostrato che la densità geometrica è l'unico modello discreto sugli interi per cui vale la proprietà di assenza di memoria.

Nota 3.3.6 Consideriamo ora una sequenza di prove di Bernoulli con probabilità di successo nella singola prova pari a $p \in (0, 1)$ e in cui il numero delle prove non è fissato, cioè ripetiamo le prove fino a quando si verifica il primo successo. Siano E_1, E_2, \dots gli eventi $E_1 = \text{"successo alla prima prova"}$, \dots , $E_2 = \text{"successo alla seconda prova"}$, \dots . Siccome le prove sono bernoulliane, allora la probabilità di ottenere il primo successo alla k -esima prova, per $k = 1, 2, \dots$ è data da

$$P(E_1^c \cap \dots \cap E_{k-1}^c \cap E_k) = (1 - p)^{k-1}p$$

Pertanto, la variabile aleatoria discreta X che indica *"il numero di prove necessarie per osservare il primo successo, inclusa l'ultima"* ha densità

$$f_X(k) = (1 - p)^{k-1}p, \quad k = 1, 2, \dots \quad (3.3.4)$$

La densità (3.3.4) è proprio la densità geometrica di parametro p .

3.3.3 Densità di Poisson

Consideriamo il centralino di un numero verde. Questo in genere è costituito da un certo numero di linee alle quali rispondono degli operatori. Sia X il numero di chiamate che arrivano ad un certo operatore in un'ora. In un modello piuttosto semplificato possiamo pensare ad un grande numero n di utenti ognuno dei quali ha una probabilità molto piccola $p \in (0, 1)$ di chiamare il numero verde in questione per mettersi in contatto con l'operatore. Se assumiamo che i singoli utenti si mettano in contatto con l'operatore indipendentemente uno dall'altro, otteniamo che $X \sim \text{Bi}(n, p)$, con n un numero molto grande e p molto piccolo. Se il numero verde è organizzato razionalmente il numero delle linee è commisurato al bacino

nota trascurabile
secondo fra

di utenza, in modo tale che vi sia un'alta probabilità di trovare il numero verde libero. Una condizione perché ciò accada è che $\lambda := np$ sia un numero fissato e non eccessivamente grande. In questo caso possiamo scrivere $X \sim \mathbf{Bi}(n, \lambda/n)$, cioè

$$P(X = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Per capire cosa succede a $P(X = k)$ se n è grande osserviamo che, per $\lambda > 0$ fissato, al divergere di n ad infinito, abbiamo

$$\lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}, \quad (3.3.5)$$

Quindi, per n “grande”:

$$P(X = k) \simeq \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \dots, n, \quad \lambda = np \quad (3.3.6)$$

Tenendo conto di quanto detto sopra, introduciamo la densità

$$f(k) := \frac{e^{-\lambda} \lambda^k}{k!} \text{ se } k = 0, 1, 2, \dots, \quad \lambda > 0.$$

che prende il nome di *densità di Poisson di parametro λ* . Una variabile aleatoria con questa densità è detta *variabile di Poisson di parametro λ* e si scrive $X \sim \mathcal{P}(\lambda)$.

Esempio 3.3.7 Il numero di automobili X che attraversano la porta di un casello autostradale in un minuto è una variabile aleatoria di Poisson di parametro 3.2. La probabilità che in un minuto non passi nessuna automobile è

$$P(X = 0) = e^{-3.2} \simeq 0.041.$$

La probabilità che ne passino più di 2 è

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) = 1 - [f_X(0) + f_X(1) + f_X(2)] \\ &= 1 - e^{-3.2} \times \left(1 + 3.2 + \frac{3.2^2}{2!}\right) \approx 0.6200963. \end{aligned}$$

Nota 3.3.8 La formula (3.3.6) oltre che per introdurre la distribuzione di Poisson può essere utilizzata per calcolare valori approssimati di $P(X = k)$ quando $X \sim \mathbf{Bi}(n, p)$ con n grande e p piccolo in quanto evita il calcolo di coefficienti binomiali. Come indicato per esempio in Ross (2004), l'approssimazione risulta soddisfacente se $n \geq 100$ e $p \leq 0.05$ oppure $n \geq 100$ e $np \leq 10$.

Esempio 3.3.9 Un computer ha probabilità $p = 10^{-3}$ di ricevere un carattere errato. Sia X il numero di errori in un messaggio di 1000 caratteri. Per calcolare la probabilità che il computer riceva più di un errore in una trasmissione di 1000 caratteri, osserviamo che se gli errori avvengono indipendentemente, allora $X \sim \mathbf{Bi}(1000, 10^{-3})$. Usando l'approssimazione di Poisson con $\lambda = np = 1000 \cdot 10^{-3} = 1$, otteniamo $P(X > 1) = 1 - P(X \leq 1) = 1 - e^{-1} \times 1^0/0! - e^{-1} \times 1^1/1! \simeq 0.2642411$. Effettuando il calcolo esatto abbiamo

$$\begin{aligned} P(X > 1) &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - \binom{1000}{0} (10^{-3})^0 (1 - 10^{-3})^{1000} - \binom{1000}{1} (10^{-3})^1 (1 - 10^{-3})^{999} \\ &\simeq 0.2642410. \end{aligned}$$

posso approx
binomialae con poisson
se n grande e p
piccolo

3.4 Trasformazioni di variabili aleatorie discrete

Se X è una variabile aleatoria discreta e $g : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione reale, allora $Y := g(X)$ è ancora una variabile aleatoria discreta. Precisamente $Y := g(X)$ è il numero che si ottiene applicando la funzione g alla variabile X . Ci interessa ora capire come determinare la densità discreta della nuova variabile aleatoria $Y := g(X)$ a partire da X . Per essere più specifici sia X una variabile aleatoria discreta a valori in I con densità $f_X(x)$ e sia $g : I \rightarrow \mathbb{R}$. Se $Y := g(X)$, allora

$$P(Y = y) = P(g(X) = y) = P\left(\bigcup_{k: g(x_k)=y} \{X = x_k\}\right) = \sum_{k: g(x_k)=y} f_X(x_k)$$

Cioè:

$$f_Y(y) = \sum_{k: g(x_k)=y} f_X(x_k) \quad (3.4.1)$$

La Formula (3.4.1) fornisce la densità di $Y = g(X)$ a partire da quella di X . Inoltre, la Formula (3.4.1) dice anche che se X e W sono due variabili aleatorie discrete che hanno la stessa densità, lo stesso vale per $g(X)$ e $g(W)$.

Esempio 3.4.1 Partecipo ad un gioco così fatto: pago una posta iniziale di 3 €, viene lanciata 3 volte una moneta equilibrata e ogni volta che esce testa vinco 2 euro. Qual è la probabilità di vincere 1 € (al netto della posta iniziale)?

Siano X la variabile aleatoria che indica il numero di vittorie su 3 giocate e Y quella che indica la vincita accumulata dopo 3 giocate. Allora X ha densità $f_X(0) = 1/8$; $f_X(1) = f_X(2) = 3/8$ e $f_X(3) = 1/8$ e $Y = 2X - 3$. Segue che per la densità di Y abbiamo

y	-3	-1	1	3
$f_Y(y)$	1/8	3/8	3/8	1/8

In particolare:

$$P(\text{"vincere 1 €"}) = P(Y = 1) = f_Y(1) = 3/8 = 0.375.$$

Esempio 3.4.2 La probabilità di vincere giocando a una slot machine è $p = 0.2$ e per partecipare a 10 giocate si paga una posta iniziale di 10 €. Se a ogni giocata o si totalizza 0 o si vincono 2 €, qual è la probabilità di vincere 4 € (al netto della posta iniziale)?

Siano X la variabile aleatoria che indica il numero di vittorie su 10 giocate e Y quella che indica la vincita accumulata dopo 10 giocate. Allora $X \sim \text{Bi}(10, 0.2)$ e $Y = 2X - 10$. Inoltre, la densità di probabilità di Y è

$$f_Y(k) = \binom{10}{\frac{10+k}{2}} 0.2^{\frac{10+k}{2}} 0.8^{\frac{10-k}{2}}, \quad \text{se } k = 0, \pm 2, \pm 4, \pm 6, \pm 8, \pm 10$$

In particolare:

$$P(\text{"vincere 4 €"}) = P(Y = 4) = \binom{10}{\frac{10+4}{2}} 0.2^{\frac{10+4}{2}} \times 0.8^{\frac{10-4}{2}} = \binom{10}{7} 0.2^7 \times 0.8^3 \simeq 0.0008.$$

Esempio 3.4.3 Siano X una variabile aleatoria discreta con densità

x	-3	-2	1	2	3
$f_X(x)$	1/8	1/8	1/4	3/8	1/8

e $Y = X^2$. Allora Y ha densità:

y	1	4	9
$f_Y(y)$	1/4	$1/8 + 3/8 = 1/2$	$1/8 + 1/8 = 1/4$

3.5 Valore atteso di variabili aleatorie discrete

Lancio 6 volte un dado regolare e mi chiedo: quante volte mi aspetto esca la faccia numerata 5? Ragioniamo: sto ripetendo 6 volte una prova che può dare esattamente 6 distinti casi elementari tutti equiprobabili. Allora su 6 lanci, “mediamente” il 5 uscirà una volta: 1 è il *numero atteso* di lanci in cui esce 5.

Formalizziamo ora matematicamente il ragionamento precedente fornendo la nozione di *media o valore atteso* di una variabile aleatoria discreta.

Definizione 3.5.1 Sia X una variabile aleatoria discreta a valori in I e sia f_X la sua densità. Se

$$\sum_{x \in I} |x| f_X(x) < +\infty$$

si definisce *media di X o valore atteso di X* il numero

$$\mu = E(X) := \sum_{x \in I} x f_X(x),$$

altrimenti si dice che X non ammette valore atteso.

Per la media di una trasformazione $Y = g(X)$ di una variabile aleatoria discreta X con densità f_X a valori in I abbiamo che, se Y ammette media, allora:

$$E(Y) = \sum_{x \in I} g(x) f_X(x). \quad (3.5.1)$$

Prima di procedere con gli esempi, facciamo qualche osservazione sulla definizione appena data. Innanzi tutto osserviamo che se X è una variabile aleatoria discreta, non è detto che X abbia valore atteso. Questo dipende dalla convergenza della “serie” $\sum_{x \in I} |x| f_X(x)$. Ovviamente se X assume un numero finito di valori, cioè se I è un insieme finito, questa serie diventa una somma finita che è sicuramente convergente e la media di X in questo caso esiste. Il motivo per il quale si richiede la convergenza assoluta, invece della semplice convergenza, è essenzialmente tecnico e non lo discuteremo in questo contesto.

Il valore atteso di una variabile aleatoria X è un oggetto legato alla densità f_X di X piuttosto che alla funzione che definisce la variabile aleatoria X . Questo significa che due variabili aleatorie con la stessa densità hanno lo stesso valore atteso.

Se torniamo all’interpretazione della densità di una variabile aleatoria discreta, come densità di massa, vista nella Nota 3.2.3, abbiamo che il valore atteso di X può essere visto come il baricentro del sistema di masse descritto.

Seguono degli esempi.

Esempio 3.5.2 Sia X la variabile aleatoria che conta il numero di lanci su 6 in cui osservo 5, allora $X \sim \mathbf{Bi}(6, 1/6)$ e $E(X) = \sum_{k=0}^6 \binom{6}{k} k \frac{1}{6^k} \left(\frac{5}{6}\right)^{6-k} = 1$.

Esempio 3.5.3 Se X è uniforme su $\{1, \dots, n\}$, cioè $f_X(k) = 1/n$ per ogni $k = 1, \dots, n$, allora

$$E(X) = \frac{1 + \dots + n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

Esempio 3.5.4 Se $X \sim \mathbf{Be}(p)$ allora

$$E(X) = 1 \times p + 0 \times (1-p) = p$$

Esempio 3.5.5 Siano X un variabile aleatoria discreta con densità

x	-3	-2	1	2	3
$f_X(x)$	1/8	1/8	1/4	3/8	1/8

e $Y = X^2$. Allora

$$E(X) = -3 \times \frac{1}{8} - 2 \times \frac{1}{8} + 1 \times \frac{1}{4} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{3}{4} = 0.75$$

e

$$E(Y) = (-3)^2 \times \frac{1}{8} + (-2)^2 \times \frac{1}{8} + 1^2 \times \frac{1}{4} + 2^2 \times \frac{3}{8} + 3^2 \times \frac{1}{8} = \frac{36}{8} = 4.5.$$

Nella seconda colonna della Tabella 3.1 trovate le medie di alcune densità discrete notevoli:

Modello	Media	Varianza
Uniforme discreta su $1, \dots, n$	$E(X) = \frac{n+1}{2}$	$\text{Var}(X) = \frac{n^2-1}{12}$
Binomiale: $X \sim \mathbf{Bi}(n, p)$	$E(X) = np$	$\text{Var}(X) = np(1-p)$
Geometrica: $X \sim \text{Geom}(p)$	$E(X) = \frac{1}{p}$	$\text{Var}(X) = \frac{1-p}{p^2}$
Poisson: $X \sim \text{Pois}(\lambda)$	$E(X) = \lambda$	$\text{Var}(X) = \lambda$

Tabella 3.1: Media e varianze di alcune variabili aleatorie discrete “notevoli”

3.5.1 Proprietà del valore atteso

Elenchiamo alcune proprietà che discendono direttamente dalla definizione di valore atteso.

1. Se X è costante e pari a c allora $E(X) = c$;
2. se $Y = aX + b$ allora $E(Y) = aE(X) + b$;
3. se $X \geq 0$ allora $E(X) \geq 0$;
4. se $a \leq X \leq b$ allora $a \leq E(X) \leq b$;
5. $E(g(X) + h(X)) = E(g(X)) + E(h(X))$.

Esercizio 3.5.6 Verificate per esercizio queste proprietà.

La proprietà 1. ci dice che il valore atteso di una costante è la costante stessa; questa proprietà è talvolta chiamata proprietà di *coerenza* del valore atteso. La proprietà 2. ci dice come si comporta il valore atteso quando si effettuano operazioni lineari sulla variabile aleatoria sottostante; questa proprietà è detta proprietà di *linearità* del valore atteso. Infine, la proprietà 3. è detta *positività* del valore atteso e la 4. *internalità* del valore atteso.

Nota 3.5.7 Tutte le proprietà del valore atteso sopra enunciate valgono sia nel caso discreto che nel caso di variabili aleatorie continue del Capitolo 4, cioè in tutti i casi per cui in questo corso abbiamo definito il valore atteso. Questo ci autorizzerà nel seguito ad applicarle a tutte le variabili aleatorie che prenderemo in considerazione senza ulteriormente specificare se sono discrete o continue.

Nota 3.5.8 Vedremo altre proprietà del valore atteso, per esempio la media della somma di variabili aleatorie nel Capitolo 5.

3.6 Varianza di variabile aleatoria discreta

La media della variabile aleatoria non riassume tutte le proprietà qualitative di una variabile aleatoria, nel senso che ci sono variabili aleatorie che hanno la stessa media ma che sono qualitativamente molto differenti come esemplificato nel prossimo esempio.

Esempio 3.6.1 A e B giocano al lancio dei dadi e puntano entrambi su pari. Ma A è avverso al rischio e punta solo 1€, mentre B è “propenso al rischio” e punta 1000€. La vincita di A è una variabile aleatoria W discreta che assume solo i due valori -1 e 1 con probabilità $1/2$, e quella di B è una variabile aleatoria Z che assume i due valori -1000 e 1000 con probabilità $1/2$. Sia per A che per B il gioco è “equo”, cioè entrambi vincono mediamente 0 euro, cioè $E(W) = E(Z) = 0$ ma Z, W sono molto diverse fra loro. La differenza fondamentale fra W e Z è che mentre W assume valori vicini alla propria media, Z assume valori piuttosto lontani da $E(Z)$.

Per quantificare la “dispersione” di una variabile aleatoria X dalla sua media $E(X)$, prendiamo la trasformazione $(X - E(X))^2$ e ne consideriamo la media $E(X - E(X))^2$.

Definizione 3.6.2 Sia X una variabile aleatoria discreta che ha media $E(X)$. Se $E[(X - E(X))^2] < \infty$, allora la *varianza* di X è

$$\text{Var}(X) := E[(X - E(X))^2] = \sum_{x \in I} (x - \mu)^2 f_X(x).$$

La radice quadrata della varianza $\sqrt{\text{Var}(X)}$ prende il nome di *deviazione standard* di X .

Esempio 3.6.3 (Segue Esempio 3.6.1) Poiché $E(W) = E(Z) = 0$, allora

$$\text{Var}(W) = E((W - E(W))^2) = E(W^2) = \sum_{k \in \{-1, 1\}} k^2 f_W(k) = (-1)^2 \times \frac{1}{2} + 1^2 \times \frac{1}{2} = 1,$$

mentre per la varianza di Z si ha

$$\text{Var}(Z) = E((Z - E(Z))^2) = E(Z^2) = (-10^3)^2 \times \frac{1}{2} + (10^3)^2 \times \frac{1}{2} = 10^6.$$

Come già anticipato, $\text{Var}(Z)$ è (molto) più grande di $\text{Var}(Y)$ ad indicare che Z si discosta da $E(Z)$ (molto) più di quanto non faccia Y da $E(Y)$.

3.6.1 Proprietà della varianza

La varianza $\text{Var}(X)$ è una misura quadratica dello scostamento di X da $E(X)$ quindi, le proprietà elementari della varianza di seguito fornite dovrebbero risultare intuitive.

Proposizione 3.6.4 Sia X una variabile aleatoria, allora

1. Se X è costante e pari a c , non c'è dispersione e $\text{Var}(X) = 0$. Viceversa, se $\text{Var}(X) = 0$ allora X è costante.
2. Se X ammette varianza ed $\alpha \in \mathbb{R}$ allora $\text{Var}(\alpha X) = \alpha^2 \text{Var}(X)$.
3. Se X ammette varianza e $\beta \in \mathbb{R}$ allora $\text{Var}(X + \beta) = \text{Var}(X)$.
4. Se X ammette varianza allora $\text{Var}(X) = E(X^2) - E(X)^2$.

Dimostrazione * Utilizzeremo le proprietà della media precedentemente fornite.

1. Se $P(X = c) = 1$ allora $E(X) = c$ e $\text{Var}(X) = E((c - c)^2) = E(0) = 0$. Viceversa, non lo vediamo.
2. Poichè $E(\alpha X) = \alpha E(X)$ allora

$$\begin{aligned} \text{Var}(\alpha X) &= E((\alpha X - E(\alpha X))^2) = E((\alpha X - \alpha E(X))^2) \\ &= E(\alpha^2 (X - E(X))^2) = \alpha^2 E((X - E(X))^2) = \alpha^2 \text{Var}(X). \end{aligned}$$

3. Osserviamo che per la linearità del valore atteso $E(X + \beta) = E(X) + \beta$, quindi

$$\begin{aligned} \text{Var}(X + \beta) &= E((X + \beta - E(X + \beta))^2) \\ &= E((X + \beta - E(X) - \beta)^2) \\ &= E((X - E(X))^2) \\ &= \text{Var}(X). \end{aligned}$$

4. Se X ammette varianza allora

$$\begin{aligned} \text{Var}(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2X E(X) + E(X)^2) \\ &= E(X^2) - 2E(X E(X)) + E(E(X)^2) \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ &= E(X^2) - E(X)^2. \quad \blacksquare \end{aligned}$$

Osservazione 3.6.5 (sulle proprietà della varianza.) Il punto 1. afferma che le uniche variabili aleatorie con varianza nulla sono le costanti. Questo è in pieno accordo con il concetto intuitivo di varianza come misura di quanto una variabile aleatoria si discosta dalla propria media. Il punto 2. ci dice che la varianza è quadratica (mentre la media è lineare). Il punto 3. mostra come la varianza sia invariante per traslazioni. Infatti sommando ad una variabile aleatoria un numero, cioè traslandola, anche la media viene traslata dello stesso numero e lo scostamento della variabile dalla sua media non cambia. Il punto 4. mostra una formula molto utile nelle applicazioni e negli esercizi per calcolare la varianza.

Esempio 3.6.6 (Continuazione Esercizio 3.5.5) Sia X un variabile aleatoria discreta con densità

x	-3	-2	1	2	3
$f_X(x)$	1/8	1/8	1/4	3/8	1/8

Già sappiamo che $E(X) = 0.75$ e $E(X^2) = 4.5$. Quindi

$$\text{Var}(X) = E(X^2) - (E(X))^2 = 4.5 - 0.75^2 = 3.9375.$$

Esempio 3.6.7 Se X è variabile aleatoria uniforme discreta su $\{1, \dots, n\}$ sappiamo che $E(X) = (n+1)/2$ e

$$E(X^2) = \sum_{k=1}^n \frac{k^2}{n} = \frac{n(n+1)(2n+1)}{6n} = \frac{(n+1)(2n+1)}{6}$$

$$\text{Var}(X) = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}$$

3.6.2 Disuguaglianza di Chebychev

La successiva importante disuguaglianza, nota come *disuguaglianza di Chebychev*, precisa in che senso una variabile X con varianza “piccola” è concentrata intorno alla sua media.

Proposizione 3.6.8 (Disuguaglianza di Chebychev) Sia X una variabile aleatoria che ammette media $E(X) = \mu$ e varianza $\text{Var}(X) = \sigma^2 > 0$. Allora per ogni $\epsilon > 0$:

$$P(|X - E(X)| > \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}.$$

Dimostrazione

$$\begin{aligned} \sigma^2 &= \sum_{x:|x-\mu|\leq\epsilon} (x-\mu)^2 f_X(x) + \sum_{x:|x-\mu|>\epsilon} (x-\mu)^2 f_X(x) \\ &\geq \sum_{x:|x-\mu|>\epsilon} (x-\mu)^2 f_X(x) \\ &\geq \sum_{x:|x-\mu|>\epsilon} \epsilon^2 f_X(x) \\ &= \epsilon^2 \sum_{x:|x-\mu|>\epsilon} f_X(x) \\ &= \epsilon^2 P(|X - E(X)| > \epsilon) \end{aligned} \quad \blacksquare$$

3.7 Standardizzazione di una variabile aleatoria

In questa sezione ci occuperemo di una particolare trasformazione affine di una variabile aleatoria, detta standardizzazione.

Sia X una variabile aleatoria non costante che ha media $E(X) = \mu$ e varianza $\text{Var}(X) = \sigma^2$. Poichè X non è costante $\text{Var}(X) > 0$. Consideriamo ora la variabile aleatoria Y :

$$Y := \frac{X - \mu}{\sigma} \quad (3.7.1)$$

dove $\sigma = \sqrt{\sigma^2}$ è la deviazione standard di X . Allora

$$E(Y) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{E(X) - \mu}{\sigma} = 0,$$

e

$$\text{Var}(X) = \text{Var}\left(\frac{X - \mu}{\sigma}\right) = \frac{\text{Var}(X - \mu)}{\sigma^2} = \frac{\text{Var}(X)}{\sigma^2} = 1.$$

Quindi, qualunque siano la media e la varianza di X ci siamo ricondotti a una variabile aleatoria Y con media uguale a 0 (diremo che è centrata) e varianza uguale a 1. Per questo motivo Y è detta *standardizzata della variabile X* e l'operazione che trasforma la variabile X nella corrispondente variabile Y è detta *standardizzazione*. Inoltre la funzione di ripartizione F_X della variabile X è legata alla funzione di ripartizione F_Y della sua standardizzata Y dalla semplice relazione

$$F_X(t) = P(X \leq t) = P\left(\frac{X - \mu}{\sigma} \leq \frac{t - \mu}{\sigma}\right) = F_Y\left(\frac{t - \mu}{\sigma}\right).$$

L'operazione di standardizzazione gioca un ruolo fondamentale nel teorema di De Moivre Laplace e nel Teorema centrale del limite che vedremo nel Capitolo 6.

Coefficiente di variazione Un'altra misura della dispersione di X intorno al polo di riferimento media μ è dato da $\frac{\sigma}{\mu}$. La grandezza $\frac{\sigma}{\mu}$ è detto *coefficiente di variazione*. Esso è una misura relativa della dispersione: poichè μ, σ sono espresse nella stessa unità di misura, $\frac{\sigma}{\mu}$ è un numero puro; è utile usare i coefficienti di variazione anziché le deviazioni standard per confrontare variabili aleatorie non omogenee. Nella teoria delle misure, dove una variabile aleatoria X è una misura di un parametro fisico, la deviazione standard σ è detta “incertezza della misura”, la media μ “valore di misura” e il coefficiente di variazione “incertezza relativa”.

Capitolo 4

Variabili aleatorie continue

4.1 Introduzione¹

Una variabile aleatoria è *continua* quando può assumere tutti i valori di un insieme continuo di numeri reali, per esempio come un intervallo. Esempi di variabili continue sono la durata di funzionamento di un componente, il tempo d'attesa di un automobilista a un semaforo o la misura di una grandezza fisica come la temperatura, la corrente elettrica, la massa, ecc.

Rappresentare i possibili valori di una variabile aleatoria mediante un continuo numerico permette di includere nel modello livelli arbitrariamente elevati nella precisione delle osservazioni sperimentali. Se eseguiamo una misura di temperatura con un termometro e leggiamo sullo strumento $20\text{ }^{\circ}\text{C}$, questo numero rappresenta l'arrotondamento di un valore che in realtà è compreso tra $19.5\text{ }^{\circ}\text{C}$ e $20.5\text{ }^{\circ}\text{C}$; usando quel termometro, ogni misura di temperatura il cui valore appartenga a tale intervallo darà come lettura $20\text{ }^{\circ}\text{C}$. Analogamente, se ripetiamo l'esperimento usando un altro termometro più preciso e leggiamo $20.3\text{ }^{\circ}\text{C}$, intendiamo che questo è l'arrotondamento di un valore compreso tra $20.25\text{ }^{\circ}\text{C}$ e $20.35\text{ }^{\circ}\text{C}$. E così via. In linea di principio si potrebbe pensare di eseguire la misura con uno strumento che ha una precisione arbitrariamente elevata, ma si arriverebbe *sempre* alla conclusione che il valore misurato è compreso in un intervallo (a, b) di valori. Se teniamo presente che la misura della temperatura è sempre soggetta a "rumore" e quindi è da considerarsi una variabile aleatoria X , ci si rende conto che, anche se idealmente essa è rappresentabile da un numero reale, in pratica non è utile chiedersi quanto vale la probabilità che la misura assuma il valore x , ma piuttosto qual è la probabilità che essa assuma un valore in un intervallo (a, b) .

Le considerazioni precedenti sono valide per qualunque variabile continua X . Pertanto: *la descrizione delle probabilità associate ad X si ottiene assegnando opportunamente le probabilità che X assuma un valore in un generico intervallo di numeri reali.* A tale scopo è comodo ignorare lo spazio campionario dell'esperimento casuale e fissare l'attenzione sulla *funzione di ripartizione* di X , definita come nel caso discreto dalla formula

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}$$

Ad ogni scelta particolare della funzione di ripartizione corrisponde un diverso modello dell'esperimento, ma in ogni caso vale la formula di calcolo

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$$

¹ Estratto da M. Verri, M. AA 09/10 (VI) - Appunti di lezione, *Variabili aleatorie continue*

perché $\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$ e i due eventi $\{X \leq a\}$, $\{a < X \leq b\}$ sono incompatibili.

4.2 Variabili aleatorie assolutamente continue

Una classe rilevante di variabili aleatorie continue è quella delle variabili aleatorie assolutamente continue.

Definizione 4.2.1 (Variabili aleatorie assolutamente continue) Una variabile aleatoria X è detta assolutamente continua se esiste una funzione $f_X : \mathbb{R} \rightarrow \mathbb{R}^+$ integrabile, tale che per ogni $a < b$ vale la formula dell'area:

$$P(X \in (a, b]) = \int_a^b f_X(x) dx \quad (4.2.1)$$

f_X prende il nome di densità di X .

Per le variabili aleatorie assolutamente continue e le loro densità valgono le seguenti proprietà:

1. qualunque sia il risultato dell'esperimento, sicuramente X assume almeno un valore e quindi vale la proprietà di normalizzazione: $\int_{\mathbb{R}} f_X(x) dx = 1$;
2. se $-\infty < a < b < +\infty$ allora

$$\begin{aligned} P(X \in (a, b)) &= P(X \in (a, b]) = P(X \in [a, b)) = P(X \in [a, b]) = \\ &= F_X(b) - F_X(a) = \int_a^b f_X(x) dx \end{aligned}$$

3. La funzione di ripartizione di X è $F_X(x) = \int_{-\infty}^x f_X(s) ds$

4. se F_X è la funzione di ripartizione di X allora $f_X(x) = F'_X(x)$ per tutti gli $x \in \mathbb{R}$ per cui esiste $F'_X(x)$ (è conseguenza del teorema fondamentale del calcolo integrale).

Il punto 2. può essere opportunamente rafforzato nel modo seguente:

5. Se X è variabile aleatoria assolutamente continua con densità f_X e $B = B_1 \cup B_2 \cup \dots \cup B_n$ dove i B_k , $k = 1, 2, \dots, n$ sono intervalli disgiunti, allora

$$P(X \in B) = \int_B f_X(x) dx = \sum_{k=1}^n \int_{B_k} f_X(x) dx$$

Osservazione 4.2.2 Dal momento che $F_X(x) = \int_{-\infty}^x f_X(s) ds$, allora la funzione di ripartizione di una variabile aleatoria assolutamente continua è sempre:

1. monotona non decrescente (e strettamente crescente sugli x per cui la densità è strettamente positiva);
2. continua $\forall x \in \mathbb{R}$;
3. e tale che $\lim_{x \rightarrow -\infty} F_X(x) = 0$ e $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

Osservazione 4.2.3 Dal punto di vista operativo variabili aleatorie assolutamente continue e discrete sono simili, dal momento che se X è una variabile assolutamente continua, allora $P(X \in B)$ si calcola facendo l'integrale su B della densità, mentre, se X è discreta, si calcola $P(X \in B)$ facendo una somma sugli elementi di B .

D'altro canto, deriva dalla formula dell'area che nel caso assolutamente continuo 1) f_X NON è una probabilità (come nel caso discreto) ma è la funzione da integrare per ottenere una probabilità e 2) $P(X = x) = 0 \forall x \in \mathbb{R}$ (mentre nel caso discreto $P(X = x) > 0$, quando x è un valore assunto dalla discreta X). In questo senso, le variabili aleatorie assolutamente continue sono molto diverse dalle variabili aleatorie discrete.

Osservazione 4.2.4 Come abbiamo fatto nella Nota 3.2.3 ci interessa evidenziare la motivazione euristica della parola “densità”. Supponiamo che f_X sia la densità di una variabile aleatoria assolutamente continua X : questo significa che f_X attribuisce un numero $f_X(x)$ ad ogni $x \in \mathbb{R}$. Se ora immaginiamo l'asse reale come un materiale inhomogeneo in cui la densità di massa è f_X , allora la massa del segmento *infinitesimo* $(x, x + dx)$ è $f_X(x)dx$.

Esempio 4.2.5 Il tempo di vita in ore di un dato tipo di pile è una variabile aleatoria assolutamente continua con densità

$$f_X(x) = \begin{cases} \frac{100}{x^2} & \text{se } 100 < x < \infty \\ 0 & \text{altrove} \end{cases}$$

1. Calcolate la probabilità che una pila debba essere sostituita entro le 150 ore di attività.
2. Una radio per funzionare ha bisogno di 5 pile. Se le pile funzionano in modo indipendente, qual è la probabilità che esattamente 2 pile su 5 debbano essere sostituite entro le prime 150 ore di attività.

Soluzione

$$1. P(X \leq 150) = \int_{100}^{150} \frac{100}{x^2} dx = -\frac{100}{2x} \Big|_{100}^{150} = \frac{1}{3}$$

2. Sia $Y = \text{“}n^0 \text{ di pile su 5 da sostituite entro le prime 150 ore di attività”}$. Allora $Y \sim \text{Bi}(5, 1/3)$ e la probabilità cercata vale

$$P(Y = 2) = f_Y(2) = \binom{5}{2} \frac{1}{3^2} \left(\frac{2}{3}\right)^3 \quad \blacksquare$$

4.2.1 Trasformazioni di una variabile aleatoria continua

Una trasformazione $Y = g(X)$ di una variabile aleatoria assolutamente continua X può essere assolutamente continua, discreta, né continua né discreta. Per esempio, se X ha la densità f_X dell'Esempio 4.2.5 e

$$Y = \begin{cases} 0 & \text{se } X \leq 150 \\ 1 & \text{se } X > 150 \end{cases}$$

allora $Y \sim \text{Bernoulli}(1/3)$.

Ma se $y = g(x)$ è una funzione “regolare”, allora la funzione composta $Y = g(X)$ definisce un'altra variabile (assolutamente) continua.

Vediamo un metodo generale e semplice per determinare la densità f_Y di Y ; è basato sull'uso della funzione di ripartizione. Per questo lo chiamiamo *metodo della funzione di ripartizione*. In breve, il metodo parte dal calcolare la fdr di Y , F_Y , esprimendo l'evento $\{Y \leq y\}$ in termini di appartenenza di X a un opportuno insieme². Quindi, si controlla se la funzione F_Y è derivabile, e in caso positivo, si ottiene la densità di F_Y , derivando. Illustriamo il metodo in due importanti casi di trasformazione: 1. *trasformazione affine*: $Y = \alpha X + \beta, \alpha \neq 0$ e *trasformazione quadratica*: $Y = X^2$.

1. Trasformazione affine: $Y = \alpha X + \beta, \alpha \neq 0$

$$F_Y(y) = P(Y \leq y) = P(\alpha X + \beta \leq y) = \begin{cases} P\left(X \leq \frac{y - \beta}{\alpha}\right) & \text{se } \alpha > 0 \\ P\left(X \geq \frac{y - \beta}{\alpha}\right) & \text{se } \alpha < 0 \end{cases}$$

da cui

$$F_Y(y) = \begin{cases} F_X\left(\frac{y - \beta}{\alpha}\right) & \text{se } \alpha > 0 \\ 1 - F_X\left(\frac{y - \beta}{\alpha}\right) & \text{se } \alpha < 0 \end{cases}$$

Perciò, se X è assolutamente continua, si ha

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{|\alpha|} f_X\left(\frac{y - \beta}{\alpha}\right) \quad (4.2.2)$$

2. Trasformazione quadratica: $Y = X^2$

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = \begin{cases} 0 & \text{se } y \leq 0 \\ P(-\sqrt{y} \leq X \leq \sqrt{y}) & \text{se } y > 0 \end{cases}$$

da cui

$$F_Y(y) = \begin{cases} 0 & \text{se } y \leq 0 \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}) & \text{se } y > 0 \end{cases}$$

Perciò, se X è assolutamente continua, si ha

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} 0 & \text{se } y < 0 \\ \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}} & \text{se } y > 0 \end{cases}$$

Nell'unico punto $x = 0$ in cui F_X non è derivabile poniamo $f_Y(0) = 0$.

Esercizio 4.2.6 Sia X una variabile aleatoria assolutamente continua e non negativa. Determinate la densità di $Y = \sqrt{X}$.

²l'insieme delle controimmagini di $y = g(x)$

4.2.2 Indici di variabili aleatorie assolutamente continue

Se X è variabile assolutamente continua tale che

$$\int_{-\infty}^{+\infty} |x| f_X(x) dx < \infty$$

il *valore atteso* (o *media*) di X è la media pesata dei possibili valori di X :

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

Se l'integrale $\int_{-\infty}^{+\infty} x f_X(x) dx$ non converge assolutamente, allora si dice che X non ammette media.

Il motivo per il quale si richiede la convergenza assoluta, invece della semplice convergenza, è tecnico e non lo discuteremo in questo contesto.

La media $E(X)$ descrive il “centro” della distribuzione di probabilità di X allo stesso modo in cui in meccanica il baricentro di un filo pesante rappresenta la posizione del punto di equilibrio del filo stesso.

Il valore atteso di una trasformazione $Y = g(X)$, è la media dei valori di $g(X)$ pesati con la densità di X , cioè

$$E(Y) = E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx \quad (\text{se } E(Y) \text{ esiste}).$$

Tutte le proprietà viste per le medie delle variabili aleatorie discrete continuano a valere anche per le medie di variabili aleatorie assolutamente continue, per esempio:

2. Se $Y = aX + b$ allora $E(Y) = a E(X) + b$;
3. se $X \geq 0$ allora $E(X) \geq 0$;
4. se $a \leq X \leq b$ allora $a \leq E(X) \leq b$;

La *moda* di X è ogni valore x a cui corrisponde il massimo assoluto di $f_X(x)$. Il *quantile di ordine* λ (con $0 < \lambda < 1$) di X è un numero x_λ per cui $P(X < x_\lambda) \leq \lambda \leq P(X \leq x_\lambda)$ e quindi si ha

$$F_X(x_\lambda) = P(X \leq x_\lambda) = \lambda$$

cioè x_λ è quel valore di X a sinistra del quale l'area sotto la curva della densità vale λ . Essendo $F_X(x)$ una funzione invertibile, si può anche scrivere

$$x_\lambda = F_X^{-1}(\lambda)$$

I quantili $x_{0.25}$, $x_{0.50}$ e $x_{0.75}$ sono detti, rispettivamente, *primo*, *secondo* e *terzo quartile* di X . Il secondo quartile $x_{0.50}$ è la *mediana* di X . I quantili $x_{0.10}, \dots, x_{0.90}$ sono il *primo*, ..., *nono decile* di X .

La *varianza* di X è la media pesata³ degli scarti quadratici dei valori di X rispetto al valore atteso:

$$\text{Var}(X) = E((X - \mu)^2) = \int_{-\infty}^{+\infty} (x - \mu)^2 f_X(x) dx$$

³Se il supporto di X è un intervallo non limitato, X è dotato di varianza purché l'integrale improprio sia convergente.

e la *deviazione standard* è la radice quadrata positiva della varianza. Anche in questo caso, come nel caso discreto, valgono la disuguaglianza di Chebychev e tutte le proprietà della varianza di variabile aleatoria discreta. Qui ne ripetiamo alcune:

1. $\text{Var}(\alpha X) = \alpha^2 \text{Var}(X)$.
2. $\text{Var}(X + \beta) = \text{Var}(X)$.
3. $\text{Var}(X) = E(X^2) - E(X)^2$.

Esempio 4.2.7 La funzione $f_X(x) = e^{-x} \mathbf{1}_{(0,+\infty)}(x)$ è una densità di variabile aleatoria X a valori positivi. È infatti funzione integrabile, positiva e $\int_0^\infty e^{-x} dx = 1$. La funzione di ripartizione corrispondente è data da:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \begin{cases} 0 & \text{se } x \leq 0 \\ 1 - e^{-x} & \text{se } x > 0 \end{cases}$$

Riguardo a media e varianza abbiamo:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^{+\infty} x e^{-x} dx = \left[-(x+1) e^{-x} \right]_{x=0}^{x \rightarrow +\infty} = 1 \\ E(X^2) &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_0^{+\infty} x^2 e^{-x} dx = \left[-(x^2 + 2x + 2) e^{-x} \right]_{x=0}^{x \rightarrow +\infty} = 2 \\ \text{Var}(X) &= E(X^2) - \mu^2 = 2 - 1 = 1 \end{aligned}$$

Esercizio 4.2.8 Data la densità $f_X(x) = \frac{1}{M-m} \mathbf{1}_{[m,M]}(x)$, verificate che $E(X) = \frac{m+M}{2}$, $\text{Var}(X) = \frac{(M-m)^2}{12}$ e calcolate il quantile x_λ di ordine λ per ogni $0 < \lambda < 1$.

4.3 Esempi di densità continue notevoli

In questa sezione elenchiamo alcune delle densità continue più importanti per le applicazioni.

4.3.1 Densità uniforme continua

Il modello della densità uniforme risponde all'esigenza di modellizzare un punto "scelto a caso" nell'intervallo $(a, b]$. Per farlo estendiamo nel continuo l'idea di spazio finito uniforme. Sia X = un punto scelto a caso in $(a, b]$. Osserviamo che nel caso di uno spazio campionario S finito, l'uniformità è resa da

$$P(E) = \frac{\text{numerosità di } E}{\text{numerosità di } S} \quad \forall E \in S.$$

Così, nel continuo sostituiamo alla nozione di numerosità di un insieme quella di lunghezza di un intervallo e diciamo che un numero X è scelto a caso in $(0, 1]$ se

$$P(c < X \leq d) = \frac{\text{lunghezza di } (c, d]}{\text{lunghezza di } (a, b]} = \frac{d - c}{b - a}, \quad \forall a < c < d < b$$

La densità f_X per cui si ha $P(c < X \leq d) = (d - c)/(b - a)$ è

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{se } a < x < b \\ 0 & \text{altrove} \end{cases} \quad (4.3.1)$$

La densità (4.3.1) è detta *densità uniforme continua sull'intervallo* $(a, b]$, la variabile aleatoria X è detta *uniforme su* $(a, b]$ e si scrive $X \sim \mathcal{U}(a, b)$. La funzione di ripartizione di X è

$$F_X(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } a \leq x < b \\ 1 & \text{se } x \geq b \end{cases}$$

(che è funzione derivabile con continuità tranne nei punti a e b). Media e varianza di $X \sim \mathcal{U}(a, b)$ sono: $\mu = \frac{a+b}{2}$, (cioè il valore centrale) e $\sigma^2 = \frac{(b-a)^2}{12}$.

Accenno alla simulazione numerica da distribuzione continue.⁴ Un'importante applicazione della distribuzione uniforme è il suo utilizzo nella *simulazione numerica delle distribuzioni assolutamente continue*. Supponiamo che X sia una variabile aleatoria con funzione di ripartizione F_X continua e crescente nel supporto di X (e perciò invertibile con inversa F_X^{-1}). Posto $U = F_X(X)$, la variabile U ha come supporto l'intervallo $0 \leq u \leq 1$ e per u in tale intervallo si ha

$$P(U \leq u) = P(F_X(X) \leq u) = P(X \leq F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u$$

Ne segue che è $U \sim \mathcal{U}(0, 1)$. Si conclude che

$$\boxed{U \sim \mathcal{U}(0, 1) \quad \implies \quad F_X^{-1}(U) = X}$$

Questa proprietà viene usata per generare dati con la distribuzione di X avendo a disposizione un generatore casuale di numeri tra 0 ed 1: se u_1, \dots, u_n sono n numeri casuali tra 0 ed 1, i loro trasformati $x_i = F_X^{-1}(u_i)$ sono appunto numeri che seguono la distribuzione di X .

Non sempre però questo è il metodo più efficiente di simulazione di numeri casuali.

Esempio 4.3.1 Per $X \sim \mathcal{E}(\lambda)$ (si veda § 4.3.2 successivo), si ha $u = F_X(x) = 1 - e^{-\lambda x}$, $x > 0$, e risolvendo rispetto ad x si ottiene

$$x = F_X^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u)$$

⁴Estratto da Verri, M. AA 09/10 (VII) - Appunti di lezione, *Modelli continui*

Ad esempio, dal seguente campione di 10 numeri, generati casualmente in $(0, 1)$,

$$\begin{aligned} u_1 &= 0.74197 \\ u_2 &= 0.11107 \\ u_3 &= 0.36331 \\ u_4 &= 0.42561 \\ u_5 &= 0.84587 \\ u_6 &= 0.67538 \\ u_7 &= 0.20622 \\ u_8 &= 0.29741 \\ u_9 &= 0.43056 \\ u_{10} &= 0.55853 \end{aligned}$$

si può ottenere un campione di 10 numeri estratto da $\mathcal{E}(2)$ calcolando

$$\begin{aligned} x_1 &= -\frac{1}{2} \ln(1 - 0.74197) = 0.67734 \\ x_2 &= -\frac{1}{2} \ln(1 - 0.11107) = 0.05886 \\ &\dots \end{aligned}$$

4.3.2 Densità esponenziale

La densità esponenziale è l'analogo continuo della densità geometrica. Supponiamo di avere un'apparecchiatura *non soggetta ad usura* (per esempio i transistori) ed inizialmente funzionante, ma che si può guastare per motivi contingenti (per esempio sovraccarichi esterni, etc...). Sia T l'istante, in minuti secondi, in cui l'apparecchiatura si guasta. La probabilità che l'apparecchiatura sia ancora funzionante dopo s secondi è $P(T > s)$. Quindi, se $s \leq 0$, allora $P(T > s) = 1$. Supponiamo ora $s > 0$. Per l'assenza di usura, la probabilità che un'apparecchiatura di età t non si guasti nei prossimi s istanti è uguale alla probabilità che un'apparecchiatura nuova non si guasti nell'intervallo di tempo $(0, s]$. In formule:

$$P(T > t + s | T > t) = P(T > s), \quad \forall s, t > 0. \quad (4.3.2)$$

Ma allora

$$P(T > s) = \frac{P(\{T > t + s\} \cap \{T > t\})}{P(T > t)} = \frac{P(T > t + s)}{P(T > t)}$$

e quindi $P(T > t + s) = P(T > t)P(T > s)$. Se definiamo $\bar{F}(t) := P(T > t)$, per ogni $t \geq 0$, abbiamo che

$$\bar{F}(t + s) = \bar{F}(t)\bar{F}(s) \quad \forall t, s > 0$$

Una funzione⁵ che verifica questa *equazione funzionale* è $e^{\alpha t}$, dove $\alpha \in \mathbb{R}$. Quindi $P(T > t) = e^{\alpha t}$ e $P(T \leq t) = 1 - e^{\alpha t}$ per $t \geq 0$. Inoltre, poiché $P(T \leq t) \leq 1$, allora necessariamente $\alpha \leq 0$ e, per evitare situazioni banali, $\alpha < 0$. Quindi la funzione di ripartizione di T è data da:

$$F_T(t) = \begin{cases} 0 & \text{se } t < 0 \\ 1 - e^{-\lambda t} & \text{se } t \geq 0, \end{cases} \quad \text{con } \lambda > 0 \quad (4.3.3)$$

⁵In realtà l'unica funzione continua.

La variabile aleatoria T è assolutamente continua e la sua densità si ottiene derivando la funzione di ripartizione:

$$F'_T(t) = \begin{cases} 0 & \text{se } t < 0 \\ \lambda e^{-\lambda t} & \text{se } t > 0 \end{cases}$$

Pertanto $f_T(t) = \lambda e^{-\lambda t} \mathbf{1}_{[0, +\infty)}(t)$ è una densità per T . Questa densità è detta *densità esponenziale di parametro λ* e la variabile aleatoria T è detta *variabile esponenziale di parametro λ* . Si scrive anche $T \sim \mathcal{E}(\lambda)$.

La Figura 4.1 mostra l'andamento di densità e funzione di ripartizione $\mathcal{E}(\lambda)$ al variare di λ : al diminuire di λ aumenta la probabilità che la variabile aleatoria esponenziale assuma valori grandi. La funzione $\bar{F}(x) = e^{-\lambda x}$, per $x > 0$, rappresenta la *probabilità di sopravvivenza*

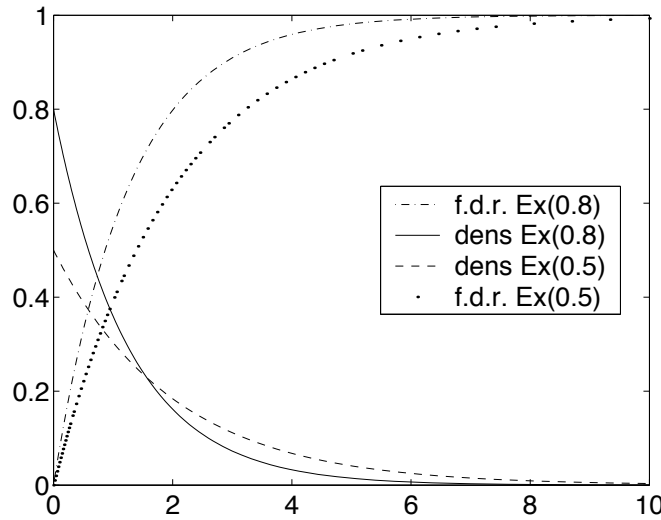


Figura 4.1: Densità e funzione di ripartizione $\mathcal{E}(\lambda)$

dell'attrezzatura. Media e varianza di $X \sim \mathcal{E}(\lambda)$ sono date da:

$$\begin{aligned} E(X) &= \int_0^{+\infty} x \lambda e^{-\lambda x} dx = - \int_0^{+\infty} x \cdot \frac{d}{dx} e^{-\lambda x} dx = x e^{-\lambda x} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx = \frac{1}{\lambda} \\ E(X^2) &= \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = - x^2 e^{-\lambda x} \Big|_0^{+\infty} + 2 \int_0^{+\infty} x e^{-\lambda x} dx = \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2} \\ \text{Var}(X) &= E(X^2) - E(X)^2 = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2. \end{aligned}$$

Osserviamo che se $X \sim \mathcal{E}(\lambda)$ e $Y = \alpha X$ con $\alpha > 0$, allora

$$f_Y(y) = \frac{1}{\alpha} f_X\left(\frac{y}{\alpha}\right) = \frac{\lambda}{\alpha} e^{-\frac{\lambda}{\alpha} y},$$

cioè, se nelle misurazioni del tempo, cambiamo l'unità di misura, la variabile così ottenuta è sempre esponenziale ma di parametro λ/α .

Osservazione 4.3.2 La proprietà (4.3.2) è anche detta *proprietà di assenza di memoria*. La densità esponenziale è l'unico modello continuo a valori in $(0, \infty)$ che gode di questa proprietà.

4.3.3 Densità gaussiana

La variabile gaussiana fornisce un utile modello probabilistico per gli errori che si commettono per esempio nei procedimenti di misurazione. Il ruolo fondamentale in probabilità della densità gaussiana standard sarà più chiaro quando verrà presentato il Teorema centrale del limite. Per ora limitiamoci a definirla e a descriverne qualche proprietà.

Definizione 4.3.3 Una variabile aleatoria X con densità

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\}, \quad \forall x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma^2 > 0 \quad (4.3.4)$$

è detta *gaussiana di parametri μ e σ^2* e si indica $X \sim \mathcal{N}(\mu, \sigma^2)$.

I parametri vengono indicati con le lettere greche μ e σ^2 e hanno una precisa interpretazione: μ è la media e σ^2 la varianza. Infatti, si può dimostrare che

$$E(X) = \mu, \quad \text{Var}(X) = \sigma^2.$$

Quindi, come messo in evidenza nella Figura 4.2, μ è un polo di riferimento e σ un indice della concentrazione (o dispersione) della densità $\mathcal{N}(\mu, \sigma^2)$ intorno a μ . La funzione di ripartizione

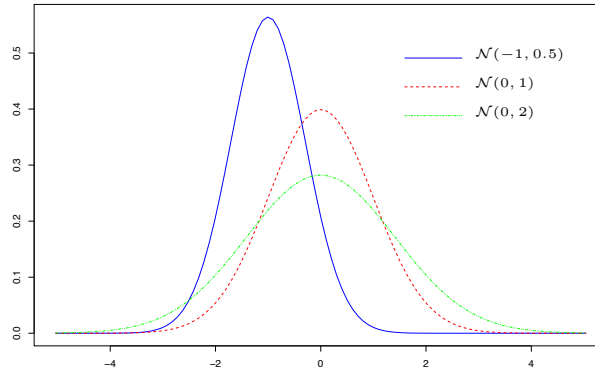


Figura 4.2: Grafico delle densità di probabilità $\mathcal{N}(0, 1)$, $\mathcal{N}(0, 2)$ e $\mathcal{N}(-1, 0.5)$

$\mathcal{N}(\mu, \sigma^2)$ è data da

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{t - \mu}{\sigma} \right)^2 \right\} dt$$

e non è possibile calcolarla analiticamente ma solo numericamente.

Definizione 4.3.4 La densità gaussiana di parametri $\mu = 0$ e $\sigma^2 = 1$ è detta *densità gaussiana standard* (o *normale standard*) e $Z \sim \mathcal{N}(0, 1)$ è detta *variabile aleatoria gaussiana standard*.

La densità gaussiana standard è convenzionalmente indicata con la lettera φ e la sua funzione di ripartizione con la lettera Φ , cioè

$$\begin{aligned} \varphi(x) &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \\ \Phi(z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx \end{aligned}$$

Trovate in Figura 4.3 **(a)** il grafico della funzione φ (che ha andamento a campana con punto

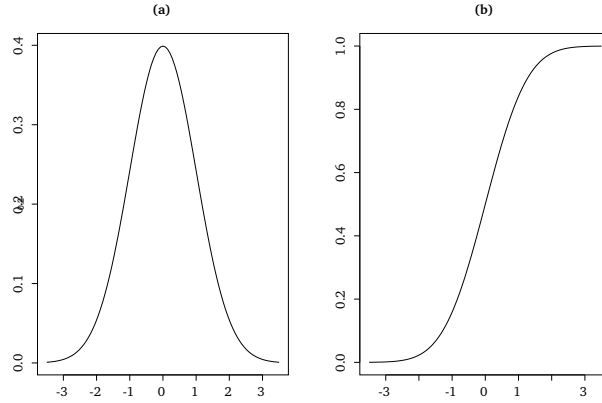


Figura 4.3: Densità **(a)** e funzione di ripartizione **(b)** $\mathcal{N}(0, 1)$

di massimo in 0 ed è simmetrico rispetto all'asse delle ordinate, ovvero φ è funzione “pari”, cioè $\varphi(-z) = \varphi(z) \forall z > 0$) e in Figura 4.3 **(b)** il grafico di Φ . In particolare notate che Φ è funzione strettamente crescente su \mathbb{R} .

I valori numerici di Φ sono tabulati tipicamente per $z \geq 0$. Se $z < 0$, $\Phi(z)$ si può ottenere usando la formula

$$\Phi(z) = 1 - \Phi(-z) \quad \forall z \in \mathbb{R} \quad (4.3.5)$$

che deriva dalla simmetria di φ . La Formula (4.3.5) traduce in termini della fdr Φ il fatto che l'area della coda di φ a sinistra di $-z$ è uguale all'area della coda di destra di φ a destra di z . Si veda, per esempio il grafico in Figura 4.4, dove l'area tra le due linee tratteggiate rappresenta graficamente $P(|Z| < z) = P(-z < Z < z)$, l'area a sinistra di $-z$ rappresenta $\Phi(-z)$ e l'area a destra $1 - \Phi(z)$.

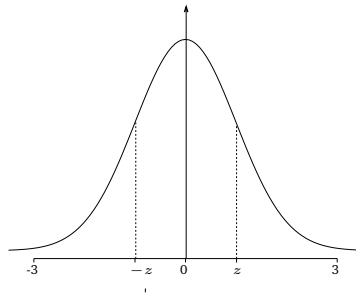


Figura 4.4: Rappresentazione grafica di $\Phi(-z)$, $P(-z < Z < z)$ e $1 - \Phi(z)$

Segue dalla Formula (4.3.5) che $\Phi(0) = 1/2$ e che

$$\begin{aligned} P(|Z| < z) &= P(|Z| \leq z) = 2\Phi(z) - 1, \\ P(|Z| > z) &= P(|Z| \geq z) = 2(1 - \Phi(z)), \quad \forall z > 0 \\ z_\lambda &= -z_{1-\lambda}, \quad 0 < \lambda < 1 \end{aligned}$$

dove z_λ è il quantile di ordine λ di Φ . Inoltre, per $z = 3$, abbiamo:

$$P(|Z| \geq 3) = 2(1 - \Phi(3)) \simeq 0.0026,$$

cioè $Z \sim \mathcal{N}(0, 1)$ ha probabilità trascurabile di assumere valori all'esterno dell'intervallo $[-3, 3]$.

Trasformazioni affini di gaussiane. Se $X \sim \mathcal{N}(\mu, \sigma^2)$ e $Y \sim aX + b$, con $a \neq 0$, allora $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$. Basta applicare la Formula (4.2.2) per le trasformazioni affini di variabili aleatorie assolutamente continue. In particolare, la standardizzata $Z = \frac{X - \mu}{\sigma}$ di una variabile aleatoria $X \sim \mathcal{N}(\mu, \sigma^2)$ ha densità gaussiana standard

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1),$$

da cui segue che se F_X è la funzione di ripartizione di $X \sim \mathcal{N}(\mu, \sigma^2)$ allora

$$F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right), \quad \forall x \in \mathbb{R}. \quad (4.3.6)$$

Il risultato in (4.3.6) permette di usare le tavole di Φ per calcolare probabilità di eventi connessi a $X \sim \mathcal{N}(\mu, \sigma^2)$ e quindi anche i quantili x_λ di $X \sim \mathcal{N}(\mu, \sigma^2)$ in termini dei quantili z_λ di $Z \sim \mathcal{N}(0, 1)$. Per i quantili abbiamo:

$$x_\lambda = \sigma z_\lambda + \mu, \quad \forall 0 < \lambda < 1$$

Esercizio 4.3.5 Se $X \sim \mathcal{N}(2, 9)$, allora

$$F_X(1.6) = \Phi\left(\frac{1.6 - 2}{\sqrt{9}}\right) = \Phi(-0.2) = 1 - \Phi(0.2) \simeq 1 - 0.5793 = 0.4207$$

$$x_{0.5} = 3 \times 0 + 2 = 2$$

$$X_{0.22} = 3z_{0.22} + 2 = -3z_{0.78} + 2 \simeq -3 \times 0.7722 + 2 = -0.3166.$$

4.3.4 Densità lognormale

La densità lognormale è usata per modellizzare tempi di attesa di un evento, quando non vale la proprietà di assenza di memoria. È anche usata in finanza per modellizzare prezzi di beni finanziari.

Definizione 4.3.6 Una variabile aleatoria X è lognormale di parametri $\mu \in \mathbb{R}$, $\sigma > 0$ se $Y = \ln X \sim \mathcal{N}(\mu, \sigma^2)$.

La densità lognormale è

$$f_X(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\}, \quad x > 0 \quad (4.3.7)$$

e si può dimostrare che $E(X) = e^{\mu + \sigma^2/2}$ e $\text{Var}(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$.

Esempio 4.3.7 La “magnitudo” Y di un terremoto che rilascia un’energia X (espressa in unità di 10^{12} erg) è definita dalla funzione $Y = (1/1.8) \log_{10} X$.

1. Se $Y \sim \mathcal{N}(4, 1)$, quanto vale l’energia media $E(X)$?

Deduciamo da $Y = (1/1.8) \log_{10} X$ che $X = e^{1.8 \log(10) \times Y}$ e quindi che X è lognormale di parametri $\mu = 1.8 \log(10) \times 4$, $\sigma = 1.8 \log(10)$. Pertanto:

$$E(X) = e^{1.8 \log(10) \times 4 + 1.8^2 \log(10)^2 / 2} = 8.5 \times 10^{10} \text{ (in unità } 10^{12} \text{ erg)}.$$

4.4 Approssimazione gaussiana della funzione di ripartizione binomiale⁶

Un risultato molto importante della teoria delle probabilità è il teorema di De Moivre-Laplace. Questo teorema afferma che, se standardizziamo una variabile aleatoria con densità binomiale di parametri n e p , la funzione di ripartizione della variabile così ottenuta converge, per $n \rightarrow +\infty$ e p fissato, alla funzione di ripartizione di una variabile aleatoria gaussiana standard. Vedremo nel Capitolo 6 che questo risultato è un caso particolare del Teorema centrale del limite, ma la sua formulazione e dimostrazione è stata fornita in modo indipendente e molto tempo prima. Diamo qui di seguito l’enunciato del teorema di De Moivre-Laplace e ne illustriamo il suo utilizzo con un esempio. Ricordiamo che n prove di Bernoulli sono n esperimenti, con due possibili risultati, successo e insuccesso, i risultati di ciascuna prova sono eventi tra loro indipendenti e infine in ogni singola prova è uguale la probabilità che si verifichi il successo. (Si veda la Sezione 3.3.1 sulle variabili aleatorie binomiali).

Teorema 4.4.1 (di De Moivre-Laplace) Sia S_n il numero di successi in n prove di Bernoulli, in ognuna delle quali il successo ha probabilità $p \in (0, 1)$. Allora,

$$P\left(a < \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \rightarrow \Phi(b) - \Phi(a), \quad \text{per } n \rightarrow +\infty, \quad \forall a < b,$$

dove Φ è la funzione di ripartizione di una gaussiana standard.

Nota 4.4.2 La variabile $(S_n - np)/\sqrt{np(1-p)}$ è la standardizzata di una variabile aleatoria binomiale. Infatti, S_n ha densità binomiale di parametri n e p e, np e $\sqrt{np(1-p)}$ sono, rispettivamente, la sua media e la sua deviazione standard.

Nota 4.4.3 Si noti che abbiamo due possibili approssimazioni per le probabilità collegate ad una densità binomiale. Possiamo utilizzare una approssimazione di Poisson se n è “grande” e p è “piccolo”, mentre si può vedere che vale un’approssimazione gaussiana se n è “grande” e p è “lontano” dai valori estremi 0 e 1. Esistono varie “ricette” per stabilire quanto n deve essere grande e p lontano da 0 e da 1. Per esempio, l’approssimazione gaussiana è buona se $n \geq 30$, $np > 5$ e $n(1-p) > 5$, oppure per $np(1-p) \geq 10$.

Esempio 4.4.4 Calcolare in modo approssimato la probabilità di ottenere in 100 lanci di una moneta equa un numero di teste compreso fra 45 e 55 (inclusi).

⁶Questa sezione è in parte tratta da Roussas, G.G. (1997) *A Course in Mathematical Statistics*, Academic Press.

Sia S_{100} la variabile aleatoria che conta il numero di teste nei 100 lanci. Allora $S_{100} \sim \mathbf{Bi}(100, 1/2)$ e la probabilità richiesta è

$$\begin{aligned}
 P(45 \leq S_{100} \leq 55) &= P(44 < S_{100} \leq 55) \\
 &= P\left(\frac{44 - \frac{100}{2}}{\sqrt{\frac{100}{4}}} < \frac{S_{100} - \frac{100}{2}}{\sqrt{100 \times \frac{1}{4}}} \leq \frac{55 - \frac{100}{2}}{\sqrt{\frac{100}{4}}}\right) \\
 &\simeq \Phi\left(\frac{55 - 50}{5}\right) - \Phi\left(\frac{44 - 50}{5}\right) \\
 &= \Phi(1) - \Phi(-1.2) \\
 &= \Phi(1) + \Phi(1.2) - 1 \simeq 0.841345 + 0.884930 - 1 \simeq 0.726275.
 \end{aligned}$$

D'altra parte, poiché S_{100} è variabile aleatoria discreta con funzione di ripartizione costante a tratti sull'intervallo $[k, k+1)$, per $k = 0, \dots, 100$, allora:

$$\begin{aligned}
 P(45 \leq S_{100} \leq 55) &= P(44 < S_{100} \leq 55) \\
 &= P(S_{100} \leq 55) - P(S_{100} \leq 44) \\
 &= P(S_{100} \leq 55.5) - P(S_{100} \leq 44.5) \\
 &\simeq \Phi\left(\frac{55.5 - 50}{5}\right) - \Phi\left(\frac{44.5 - 50}{5}\right) \\
 &= \Phi(1.1) - \Phi(-1.1) \\
 &= 2\Phi(1.1) - 1 \simeq 2 \times 0.864334 - 1 \simeq 0.728668.
 \end{aligned}$$

Nell'ultima equazione per calcolare un valore approssimato di $P(45 \leq S_{100} \leq 55) = P(S_{100} \leq 55) - P(S_{100} \leq 44)$, abbiamo apportato una *correzione di continuità* sostituendo a 55 il valore $55+0.5$ e a 44 il valore $44+0.5$. Calcolando ora esattamente $P(45 \leq S_{100} \leq 55)$ mediante la densità binomiale, otteniamo che $P(45 \leq S_{100} \leq 55) = 0.728747$. Quindi, senza la correzione di continuità l'approssimazione gaussiana produce un errore in percentuale pari a $(0.728747 - 0.726275)/0.728747 \simeq 0.34\%$, mentre, con la correzione di continuità, l'errore è $(0.728747 - 0.728668)/0.728747 \simeq 0.011\%$: l'introduzione della correzione di continuità ha ridotto l'errore di approssimazione di un fattore 31.

In generale, se n è grande e $S_n \sim \mathbf{Bi}(n, p)$, con $p \in (0, 1)$, la correzione di continuità si apporta nel seguente modo:

$$P(S_n \leq r) \simeq \Phi\left(\frac{r + 0.5 - np}{\sqrt{np(1-p)}}\right), \quad r = 0, 1, \dots$$

Infine, trovate in Figura 4.5 il grafico della funzione di ripartizione gaussiana standard Φ e della standardizzata di una variabile aleatoria $S_{20} \sim \mathbf{Bi}(20, 0.5)$, a confronto.

4.5 Accenno al caso di variabili aleatorie miste⁷

Una classe rilevante di variabili aleatorie continue è quella delle variabili aleatorie assolutamente continue. Non tutte le variabili aleatorie continue sono assolutamente continue, cioè

⁷Estratto da Verri, M. AA 09/10 (VI) - Appunti di lezione, *Variabili aleatorie continue*

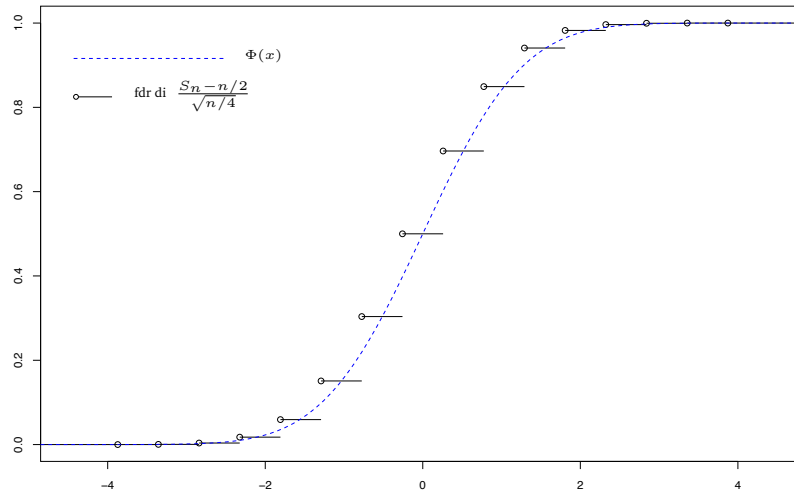


Figura 4.5: Fdr di $(S_n - np)/\sqrt{np(1-p)}$ e Φ a confronto per $n = 20$ e $p = 0.5$

dotate di densità continua. Ricorrendo all'analogia meccanica, oltre ai sistemi di particelle e ai fili, si possono anche considerare distribuzioni di massa “miste”, cioè costituite da un filo a cui sono sovrapposte una o più particelle di massa concentrata. **Vediamo un esempio di variabile aleatoria mista.**

Esempio 4.5.1 Studiamo la variabile $X = \text{“tempo d’attesa di un automobilista al semaforo”}$. Supponiamo che il semaforo sia “Rosso” (evento R) per un periodo T , “Verde” (evento V) per un altro periodo T , e così via. Se un automobilista arriva al semaforo, può trovare “Verde” (e allora passa senza aspettare) oppure “Rosso” (e allora attende fino a quando scatta il “Verde”). Di conseguenza $0 \leq X \leq T$. Calcoliamo la funzione di ripartizione di X . Abbiamo

$$F_X(x) = P(X \leq x) = 0 \quad \text{per } x < 0$$

perché è impossibile che X assuma valori negativi, e

$$F_X(x) = 1 \quad \text{per } x > T$$

perché è certo che l'automobilista aspetti meno di un tempo T . Inoltre si ha

$$F_X(0) = P(X = 0) = P(V)$$

perché quando il semaforo è “Verde” l'automobilista non aspetta. Infine, applicando la formula delle probabilità totali, si ha per $0 < x \leq T$

$$\begin{aligned} F_X(x) &= P(X \leq x|V) P(V) + P(X \leq x|R) P(R) \\ &= P(V) + P(X \leq x|R) P(R) \end{aligned}$$

perché $P(X \leq x|V) = 1$, essendo certo che l'automobilista aspetterà meno di un tempo $x > 0$ se arriva durante il “Verde”. Facciamo ora l'ulteriore ipotesi che l'istante d'arrivo dell'automobilista al semaforo sia del tutto casuale: essendo uguali le durate T del “Verde” e del “Rosso”,

ne segue $P(V) = P(R) = 1/2$; inoltre, la casualità dell'arrivo comporta che la distribuzione di X condizionata all'evento R segua il modello uniforme nel continuo, cioè che

$$P(X \leq x|R) = \frac{x}{T} \quad \text{per } 0 \leq x \leq T$$

In definitiva si ottiene

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{2} + \frac{x}{2T} & \text{se } 0 \leq x \leq T \\ 1 & \text{se } x > T \end{cases}$$

Questa funzione di ripartizione non è assolutamente continua perché $P(X = 0) = P(X \leq 0) = 1/2$. In effetti si tratta di una *funzione di ripartizione mista* (o combinata) perché corrisponde alla sovrapposizione di una massa di valore $1/2$, concentrata nell'origine, e di una massa di valore $1/2$ uniformemente diffusa nell'intervallo $0 \leq x \leq T$ (un filo "omogeneo"). Questa interpretazione meccanica ci permette anche di calcolare il tempo medio d'attesa dell'automobilista al semaforo: esso coincide con il baricentro del sistema misto "particella + filo", che è dato dalla media pesata dei baricentri della particella ($= 0$) e del filo ($= T/2$), dove i pesi sono rispettivamente le masse della particella ($= 1/2$) e del filo ($= 1/2$)

$$E(X) = \frac{1}{2} \times 0 + \frac{1}{2} \times \frac{T}{2} = \frac{T}{4}.$$

Capitolo 5

Variabili aleatorie multivariate

Nei Capitoli 3 e 4 abbiamo introdotto la variabile aleatoria per modellare il concetto di numero casuale. Spesso nelle applicazioni accade che sia necessario considerare simultaneamente più variabili aleatorie definite sullo stesso spazio campionario, cioè relative a uno stesso esperimento aleatorio. Consideriamo il seguente esempio.

Esempio 5.0.2 Siamo interessati a studiare l'affidabilità di una piccola rete di tre linee elettriche connesse secondo lo schema in Figura 5.1. I tempi di guasto di ciascuna linea sono

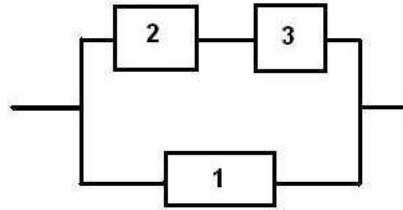


Figura 5.1: schema a blocchi di una rete di tre linee elettriche

modellati con tre variabili aleatorie esponenziali: $X_1 \sim \mathcal{E}(1)$, $X_2 \sim \mathcal{E}(2)$ e $X_3 \sim \mathcal{E}(3)$. Sia A l'evento

$$A = \text{"la rete funziona perfettamente al tempo } t\text{"}$$

Allora l'affidabilità della rete in un istante t è $P(A)$ che possiamo esprimere in termini delle X_1, X_2, X_3 nel seguente modo

$$\begin{aligned} P(A) &= P(\{X_1 > t\} \cup \{X_2 > t\} \cap \{X_3 > t\}) = \\ &= P(X_1 > t) + P(\{X_2 > t\} \cap \{X_3 > t\}) - P(\{X_1 > t\} \cap \{X_2 > t\} \cap \{X_3 > t\}). \end{aligned}$$

Vediamo che per valutare l'affidabilità della rete, dobbiamo saper valutare il comportamento congiunto delle tre X_1, X_2, X_3 . In particolare:

- se le tre linee elettriche funzionano in modo indipendente, allora gli eventi $\{X_1 > t\}, \{X_2 > t\}, \{X_3 > t\}$ sono indipendenti e la conoscenza delle densità di X_1, X_2, X_3 è

sufficiente per calcolare $P(A)$:

$$\begin{aligned} P(A) &= P(X_1 > t) + P(X_2 > t)P(X_3 > t) - P(X_1 > t)P(X_2 > t)P(X_3 > t) \\ &= e^{-t} + e^{-2t}e^{-3t} - e^{-t}e^{-2t}e^{-3t} = e^{-t} + e^{-5t} + e^{-6t}, \end{aligned}$$

Ma l'ipotesi di componenti indipendenti non è sempre giustificata. Per esempio, se le tre linee sono vicine fra loro e si trovano in una zona soggetta a rischio sismico, allora un terremoto di sufficiente intensità potrebbe metterle tutte fuori servizio, cioè per i guasti da terremoto non è possibile assumere linee indipendenti. Dobbiamo quindi procedere a modellare probabilisticamente il comportamento congiunto di X_1, X_2, X_3 e, in generale, di n va X_1, \dots, X_n se la rete è costituita da n linee elettriche. Prima di far ciò formalizziamo la nozione di “variabili aleatorie indipendenti”.

5.1 Variabili aleatorie indipendenti

Partiamo da una notazione: la scrittura $\{X_1 \in B_1, \dots, X_n \in B_n\}$ indica l'intersezione degli eventi $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$:

$$\{X_1 \in B_1, \dots, X_n \in B_n\} = \{X_1 \in B_1\} \cap \dots \cap \{X_n \in B_n\}.$$

Definizione 5.1.1 Siano X_1, \dots, X_n n variabili aleatorie definite su uno stesso spazio campionario. X_1, \dots, X_n sono *indipendenti* se per tutti gli insiemi di numeri reali¹ B_1, \dots, B_n si ha

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdots P(X_n \in B_n) \quad (5.1.1)$$

Osservazione 5.1.2 Se X_1, \dots, X_n sono indipendenti, allora anche $Y_1 = g_1(X_1), \dots, Y_n = g_n(X_n)$ sono indipendenti fra di loro, per qualunque scelta delle trasformazioni g_1, \dots, g_n .

Esempio 5.1.3 Consideriamo l'esperimento aleatorio consistente nel lanciare dieci volte due monete equilibrate e siano X il numero di teste nei dieci lanci della prima moneta e Y il numero di teste nella seconda. Chiaramente, per ogni $A, B \subset \{0, 1, \dots, 10\}$ gli eventi $\{X \in A\}$ e $\{Y \in B\}$ sono indipendenti perché $\{X \in A\}$ dipende solo dai risultati del lancio della prima moneta e $\{Y \in B\}$ dai risultati del lancio della seconda. Inoltre, $X \sim \mathbf{Bi}(10, 1/2)$ e $Y \sim \mathbf{Bi}(10, 1/2)$. Quindi, secondo la Definizione 5.1.1, le variabili aleatorie X, Y sono indipendenti e

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B) = \sum_{k \in A} \sum_{j \in B} \binom{10}{k} \binom{10}{j} \frac{1}{2^{20}}.$$

Per esempio,

$$P(X = 5, Y > 8) = \binom{10}{5} \frac{11}{2^{20}} \simeq 0.00264.$$

¹per cui si può calcolare $P(X_1 \in B_1), \dots, P(X_n \in B_n)$

Sia ora Z la variabile aleatoria che indica il numero totale di teste nei lanci di entrambe le monete. Allora

$$P(X = 5, Z \leq 8) = P(X = 5, Y \leq 3) = f_X(5)F_Y(3) \simeq 0.0423 ,$$

mentre, $P(X = 5)P(Z \leq 8) = f_X(5)F_Z(8) \simeq 0.06194 \neq 0.0423 = P(X = 5, Z \leq 8)$, dal momento che $Z \sim \text{Bi}(20, 0.5)$. Quindi, X, Z non sono indipendenti.

5.2 Coppie di variabili aleatorie

Analizziamo ora coppie di variabili aleatorie che riguardano lo stesso esperimento aleatorio e che in generale non sono indipendenti. Trattiamo coppie di variabili aleatorie, procediamo seguendo lo stesso ordine seguito per le variabili aleatorie.

Definizione 5.2.1 (Funzione di ripartizione congiunta) Sia (X, Y) una coppia di variabili aleatorie definite sullo stesso spazio campionario. La seguente funzione di due variabili:

$$F_{X,Y}(x, y) := P(X \leq x, Y \leq y), \quad x \in \mathbb{R}, y \in \mathbb{R}$$

si chiama *funzione di ripartizione congiunta* di X, Y .

La conoscenza di $F_{X,Y}$ permette di calcolare le probabilità di tutti gli eventi connessi alla coppia (X, Y) . Per esempio:

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1)$$

Possiamo dedurre da $F_{X,Y}$ la funzione di ripartizione *marginale* F_X di X e F_Y di Y . Abbiamo infatti che

$$F_X(x) = P(X \leq x) = P(X \leq x, Y < \infty) = F_{X,Y}(x, \infty) = F_{X,Y}(x, \infty)$$

dove $F_{X,Y}(x, \infty)$ sta per $\lim_{y \rightarrow +\infty} F_{X,Y}(x, y)$. Analogamente: $F_Y(y) = F_{X,Y}(\infty, y)$.

Quindi dalla funzione di ripartizione congiunta siamo in grado di calcolare tutte le funzioni di ripartizione marginali e quindi le probabilità di eventi riguardanti X e Y separatamente.

5.2.1 Coppie di variabili aleatorie discrete

Definizione 5.2.2 Siano X, Y due variabili aleatorie discrete X, Y definite su uno stesso spazio campionario, rispettivamente a valori in x_1, x_2, \dots e y_1, y_2, \dots . La funzione $f_{X,Y}(x_i, y_j) := P(X = x_i, Y = y_j)$ si chiama *densità congiunta discreta* di X, Y .

Osservazione 5.2.3 Quando si guarda al comportamento congiunto di due variabili aleatorie X, Y discrete, si usa anche l'espressione: *variabile aleatoria bivariata (X, Y) discreta*.

Esempio 5.2.4 (Densità trinomiale) Supponiamo di estrarre a caso senza ripetizione n oggetti da una popolazione che ne contiene 3 tipi diversi nelle seguenti proporzioni: p_1, p_2 e $p_3 = 1 - p_1 - p_2$. Sia X il numero di oggetti estratti di tipo 1 e Y quello di tipo 2. Allora X, Y sono variabili aleatorie discrete e la densità congiunta di X, Y è detta *trinomiale* di parametri n, p_1, p_2 .

Per scrivere esplicitamente la densità, possiamo pensare di estrarre gli elementi dalla popolazione uno alla volta. Poiché le n scelte sono indipendenti, la probabilità che la sequenza delle n estrazioni contenga n_1 elementi di tipo 1, n_2 elementi di tipo 2 e $n_3 = n - n_1 - n_2$ di tipo 3 (in un ordine prefissato) è $p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n - n_1 - n_2}$. Inoltre, il numero di modi differenti in cui l'ordine degli n oggetti può essere specificato è pari a

$$\binom{n}{n_1 n_2} := \frac{n!}{n_1! \times n_2! \times (n - n_1 - n_2)!}.$$

Segue che la probabilità di ottenere esattamente n_1 elementi di tipo 1, n_2 di tipo 2 e $n - n_1 - n_2$ elementi di tipo 3 è data da

$$P(X = n_1, Y = n_2) = \binom{n}{n_1 n_2} p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n - n_1 - n_2},$$

per $n_1, n_2 = 0, \dots, n$ e $n_1 + n_2 \leq n$.

Si osservi che se l'urna contiene solo 2 tipi diversi di oggetti 1 e 2, allora la variabile aleatoria che riassume l'esperimento dicendomi quanti oggetti di tipo 1 ho ottenuto è $X \sim \text{Bi}(n, p_1)$.

Proprietà della densità congiunta discreta Una densità congiunta discreta soddisfa le seguenti proprietà:

1. proprietà di non negatività: $0 \leq f_{X,Y}(x_i, y_j) \leq 1$;
2. proprietà di normalizzazione: $\sum_{i,j} f_{X,Y}(x_i, y_j) = 1$.

Alla luce delle precedenti due proprietà anche nel caso bivariato, si può dare una interpretazione meccanica della densità congiunta discreta $f_{X,Y}$: immaginando un sistema di particelle collocate sul piano nelle posizioni (x_i, y_j) e aventi masse $f_{X,Y}(x_i, y_j)$, allora la funzione $f_{X,Y}$ descrive come si distribuisce la massa unitaria tra le varie particelle del sistema.

3. La funzione di ripartizione congiunta di $F_{X,Y}$ è

$$F_{X,Y}(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f_{X,Y}(x_i, y_j) \quad \forall x, y;$$

4. per ogni regione del piano \mathbb{R}^2 vale la formula

$$P((X, Y) \in B) = \sum_{(x,y) \in B} f_{X,Y}(x, y);$$

5. la densità *marginale* di X e quella di Y sono date da

$$f_X(x_i) = P(X = x_i) = P(X = x_i, Y \in \mathbb{R}) = \sum_j P(X = x_i, Y = y_j) = \sum_j f_{X,Y}(x_i, y_j)$$

e

$$f_Y(y_j) = \sum_i f_{X,Y}(x_i, y_j)$$

(5.2.1)

Infine, per stabilire se due variabili aleatorie discrete sono indipendenti basta guardare alla loro densità congiunta:

6. due variabili aleatorie discrete X, Y sono indipendenti se e solo se la loro densità congiunta fattorizza nel prodotto delle densità marginali, cioè:

$$X \text{ e } Y \text{ sono indipendenti sse } f_{X,Y}(x_i, y_j) = f_X(x_i)f_Y(y_j) \quad \forall x_i, y_j$$

Osservazione 5.2.5 La densità congiunta discreta di una coppia di variabili aleatorie X, Y che assumono un numero finito di valori può essere rappresentata mediante una *tabella a doppia entrata* come quella dell'Esercizio 5.2.6. Coerentemente con le formule di marginalizzazione (5.2.1), nella rappresentazione mediante tabella, le densità marginali si ottengono facendo le somme lungo le righe e lungo le colonne.

Esercizio 5.2.6 Da un gruppo di 7 batterie, di cui 3 nuove, 2 usate ma funzionanti e 2 difettose, ne vengono scelte 3 a caso. Siano X e Y rispettivamente il numero di batterie nuove e usate (ma funzionanti) tra quelle scelte.

1. Determinare la densità congiunta di (X, Y) e le densità marginali di X e di Y .
 2. Stabilite se X e Y sono indipendenti.
 3. Le tre batterie scelte sono montate su di un apparecchio che funziona se nessuna di esse è difettosa. Determinare la probabilità che l'apparecchio funzioni.
1. La densità congiunta di X, Y e le densità marginali di X ed Y sono riportate nelle seguente tabella:

$X \setminus Y$	0	1	2	f_X
0	0	2/35	2/35	4/35
1	3/35	12/35	3/35	18/35
2	6/35	6/35	0	12/35
3	1/35	0	0	1/35
f_Y	10/35	20/35	5/35	1

2. Poiché per esempio $f_{X,Y}(0, 2) = \frac{2}{35} \neq \frac{4}{35} \times \frac{5}{35} = f_X(0)f_Y(2)$ allora X ed Y non sono indipendenti.
3. $P(X + Y = 3) = f_{X,Y}(3, 0) + f_{X,Y}(1, 2) + f_{X,Y}(2, 1) = \frac{1 + 3 + 6}{35} = \frac{2}{7}$ ■

Densità congiunta di n variabili aleatorie discrete Se consideriamo simultaneamente più di 2 variabili aleatorie discrete relative a uno stesso esperimento aleatorio, lavoreremo con la densità congiunta discreta di n variabili aleatorie discrete X_1, \dots, X_n , definita da

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) := P(X_1 = x_1, \dots, X_n = x_n)$$

Segue un esempio di densità congiunta discreta notevole.

Esempio 5.2.7 (Densità multinomiale) Supponiamo che una popolazione contenga oggetti di $k \geq 2$ tipi diversi e che la proporzione degli oggetti di tipo i nella popolazione sia p_i per $i = 1, \dots, k$ ($p_i > 0, \sum_{i=1}^k p_i = 1$). Inoltre, supponiamo che n oggetti siano scelti a caso dalla popolazione con reimmissione. Sia X_i il numero di oggetti di tipo i estratti, per $i = 1, \dots, k$

e sia \mathbf{X} il vettore aleatorio che ha componenti X_1, \dots, X_k . Allora il vettore aleatorio \mathbf{X} è discreto, la somma delle sue componenti è pari al numero di estrazioni ($X_1 + \dots + X_k = n$) e la sua densità è detta *multinomiale* di parametri n, p_1, \dots, p_k .

Per scrivere esplicitamente la densità, possiamo pensare di estrarre gli elementi dalla popolazione uno alla volta. Poiché le n scelte sono indipendenti, la probabilità che la sequenza delle n estrazioni contenga n_1 elementi di tipo 1, \dots , n_k elementi di tipo k (in un ordine prefissato) è $p_1^{n_1} \dots p_k^{n_k}$. Inoltre, il numero di modi differenti in cui l'ordine degli n oggetti può essere specificato è pari al numero di partizioni ordinate di classe (n_1, \dots, n_k) , cioè

$$\binom{n}{n_1 \dots n_k} := \frac{n!}{n_1! \times n_2! \times \dots \times n_k!}.$$

Segue che la probabilità di ottenere esattamente n_1 elementi di tipo 1, \dots , n_k elementi di tipo k è

$$P(X_1 = n_1, \dots, X_k = n_k) = \binom{n}{n_1 \dots n_k} p_1^{n_1} \dots p_k^{n_k}, \quad n_1, \dots, n_k = 0, \dots, n \text{ e } n_1 + \dots + n_k = n$$

5.3 Variabili aleatorie congiuntamente continue

Definizione 5.3.1 (Variabili aleatorie congiuntamente continue) Due variabili aleatorie X, Y definite su uno stesso spazio campionario sono *congiuntamente continue* se esiste una funzione $f_{X,Y}$ di due variabili a valori reali, non negativa e integrabile, tale che valga la *formula del volume*:

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x, y) \, dx dy \quad \forall a_1 \leq a_2 \text{ e } b_1 \leq b_2;$$

$f_{X,Y}$ prende il nome di *densità congiunta* di X, Y .

Osservazione 5.3.2 Per essere coerenti con la definizione di variabili aleatorie univariate assolutamente continue, di cui la precedente fornisce un'estensione, dovremmo parlare di *variabili aleatorie congiuntamente assolutamente continue*. Ma omettiamo l'avverbio assolutamente per semplificare la presentazione.

Osservazione 5.3.3 Quando si guarda al comportamento congiunto di due variabili aleatorie X, Y congiuntamente (assolutamente) continue, si usa anche l'espressione: *variabile aleatoria bivariata* (X, Y) *assolutamente continua*.

Per le variabili aleatorie X, Y congiuntamente continue e le loro densità valgono proprietà analoghe a quelle delle variabili aleatorie univariate continue. Ricordiamone alcune:

1. proprietà di normalizzazione: $\int_{\mathbb{R}} \int_{\mathbb{R}} f_{X,Y}(x, y) \, dx dy = 1$;
2. la funzione di ripartizione $F_{X,Y}$ è data da $F_{X,Y}(a, b) = \int_{-\infty}^a \int_{-\infty}^b f_{X,Y}(x, y) \, dx dy$;
3. viceversa, $f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$ per tutti gli (x, y) tali che esiste la derivata parziale;

4. le variabili aleatorie X, Y , separatamente prese, sono entrambe assolutamente continue e le loro densità *marginali* sono date da:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \quad \text{e} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

Infatti:

$$P(a < X \leq b) = P(a < X \leq b, Y < \infty) = \int_a^b \left[\int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \right] dx$$

e quindi abbiamo ritrovato per $P(a < X \leq b)$ la formula dell'area con $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$.

Infine, per stabilire se due variabili aleatorie congiuntamente continue sono indipendenti basta guardare alla loro densità congiunta:

5. due variabili aleatorie congiuntamente continue X, Y sono indipendenti se e solo se la loro densità congiunta fattorizza nel prodotto delle densità marginali, cioè:

$$X \text{ e } Y \text{ sono indipendenti se e solo se } f_{X,Y}(x_i, y_j) = f_X(x_i) f_Y(y_j) \quad \forall x_i, y_j$$

Esempio 5.3.4 Siano S e T due variabili aleatorie che descrivono i tempi di guasto, in minuti secondi, di due componenti elettronici e sia

$$f_{S,T}(s, t) = \begin{cases} e^{-s-t} & \text{se } s > 0 \text{ e } t > 0 \\ 0 & \text{altrove} \end{cases}$$

la loro densità congiunta. Allora le densità marginali di S, T sono rispettivamente:

$$f_S(s) = \int_0^{\infty} e^{-s-t} dt = e^{-s} \quad \text{se } s > 0 \text{ e } 0 \text{ altrove}$$

$$f_T(t) = \int_0^{\infty} e^{-s-t} ds = e^{-t} \quad \text{se } t > 0 \text{ e } 0 \text{ altrove,}$$

cioè $S \sim \mathcal{E}(1)$ e $T \sim \mathcal{E}(1)$. Inoltre, per ogni $s > 0, t > 0$:

$$f_{S,T}(s, t) = e^{-s-t} = e^{-s} e^{-t} = f_S(s) f_T(t)$$

da cui deduciamo che S e T sono indipendenti.

5.3.1 Densità congiunta uniforme

Una densità congiunta uniforme su una regione limitata R del piano Oxy fornisce un modello probabilistico per modellare un punto scelto a caso in R . Siano X, Y le coordinate di un punto scelto a caso in R . La sua densità congiunta uniforme di X, Y su R è data da

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{|R|} & \text{se } (x, y) \in R \\ 0 & \text{altrove} \end{cases} \quad (5.3.1)$$

dove $|R|$ rappresenta l'area della regione R . In questo modello uniforme si ha:

$$P(X, Y \in A) = \frac{|A|}{|R|}, \quad \forall A \subseteq R$$

dove $|A|$ rappresenta l'area della regione A . Cioè la probabilità di estrarre un punto in un'area A sottoinsieme di R è proporzionale all'area di A .

Esempio 5.3.5 (Densità uniforme sul cerchio) ² Siano (X, Y) le coordinate di un punto

²Tratto da Baldi P., 1998 Calcolo delle probabilità e statistica, McGraw-Hill

“scelto a caso” nel cerchio R di raggio r : $R = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq r^2\}$. Questo significa che le va X, Y hanno densità congiunta costante su R e nulla al di fuori di R :

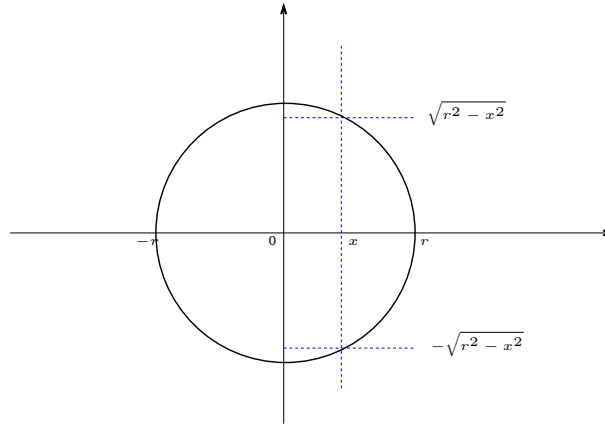
$$f_{X,Y}(x, y) = \begin{cases} a & (x, y) \in R \\ 0 & \text{altrove.} \end{cases}$$

Dalla proprietà di normalizzazione 1. segue che il valore della costante a deve essere tale che

$$\int_R a \, dx \, dy = 1$$

cioè a è il reciproco dell'area del cerchio R : $a = 1/(\pi r^2)$.

Per le densità marginali f_X, f_Y abbiamo:



$$f_X(x) = \int_{\mathbb{R}} \frac{1}{\pi r^2} \mathbf{1}_R(x, y) dy = \int_{-\sqrt{r^2 - x^2}}^{\sqrt{r^2 - x^2}} \frac{1}{\pi r^2} dy = \frac{2}{r^2 \pi} \sqrt{r^2 - x^2} \quad \text{se } x \in (-r, r).$$

Se invece $x \notin (-r, r)$ allora $f_X(x) = 0$. In definitiva,

$$f_X(x) = \frac{2}{r^2 \pi} \sqrt{r^2 - x^2} \mathbf{1}_{(-r, r)}(x).$$

Per motivi di simmetria vale anche che

$$f_Y(y) = \frac{2}{r^2 \pi} \sqrt{r^2 - y^2} \mathbf{1}_{(-r, r)}(y).$$

Nota 5.3.6 Impariamo da questo esempio che le densità marginali di una densità congiunta uniforme non sono necessariamente uniformi.

Esercizio 5.3.7 Le variabili aleatorie X, Y hanno densità congiunta

$$f_{X,Y}(x, y) = \begin{cases} k(x + y) & \text{se } 0 < x < 1, 0 < y < 1 \\ 0 & \text{altrove} \end{cases}$$

1. determinate il valore della costante k ;
2. determinate la densità marginale di X e quella di Y .
3. X, Y sono indipendenti?

Soluzione

1. Dalla proprietà di normalizzazione segue che il valore della costante k deve soddisfare

$$\int_0^\infty \int_0^\infty k(x+y) dx dy = 1$$

cioè k è il reciproco dell'integrale doppio: $\int_0^1 \int_0^1 (x+y) dx dy$:

$$\int_0^1 \int_0^1 (x+y) dx dy = \int_0^1 \left(x + \frac{1}{2}\right) dx = \frac{1}{2} + \frac{1}{2} = 1$$

quindi $k = 1$.

$$2. f_X(x) = \begin{cases} \int_0^1 (x+y) dy = x + \frac{1}{2} & 0 < x < 1 \\ 0 & \text{altrove} \end{cases}$$

Per ragioni di simmetria $f_Y(y) = (y + 1/2)$ per $0 < y < 1$ e 0 altrove.

3. Poiché $(x + 1/2)(y + 1/2) \neq (x + y)$, per “quasi” tutti $0 < x, y < 1$, allora X, Y non sono indipendenti. ■

5.4 Trasformazioni di variabili aleatorie

Se X_1, X_2, \dots, X_n sono variabili aleatorie definite sullo stesso spazio campionario e $g(x_1, x_2, \dots, x_n)$ è una funzione “regolare” di n variabili (a valori reali), allora la trasformazione $Z = g(X_1, X_2, \dots, X_n)$ definisce una nuova variabile aleatoria.

Noi non ci occuperemo di determinare se Z è discreta o continua, o di come è fatta la sua densità, ma ci limiteremo a studiarne il valore atteso. Ci soffermeremo in particolare su media e varianza di somme di variabili aleatorie. Inoltre, riporteremo dei risultati sulla densità di somme di variabili aleatorie indipendenti in alcuni casi semplici, ma “notevoli”, cioè di comune applicazione.

5.5 Valore atteso di trasformazioni di variabili aleatorie

Senza perdere in generalità, trattiamo del valore atteso di una trasformazione di due variabili aleatorie.

Siano X, Y due variabili aleatorie definite sullo stesso spazio campionario, g una funzione di due variabili a valori reali e $Z = g(X, Y)$. Se il valore atteso di Z esiste, allora

$$E(Z) = E(g(X, Y)) = \begin{cases} \sum_i \sum_j g(x_i, y_j) f_{X,Y}(x_i, y_j) & \text{nel caso discreto} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dx dy & \text{nel caso continuo} \end{cases} \quad (5.5.1)$$

Esercizio 5.5.1 (Continuazione Esercizio 5.3.7) Se $g(x, y) = xy$ e X, Y hanno densità congiunta:

$$f_{X,Y}(x, y) = \begin{cases} x + y & \text{se } 0 < x < 1, 0 < y < 1 \\ 0 & \text{altrove} \end{cases}$$

allora

$$E(XY) = \int_0^1 \int_0^1 xy(x+y) dx dy = \int_0^1 \left(\frac{y}{3} + \frac{y^2}{2}\right) dy = \frac{1}{3}$$

Notiamo in questo esempio che il calcolo di $E(XY)$ coinvolge l'uso della densità congiunta.

Il calcolo di $E(XY)$ si semplifica se X, Y sono indipendenti, come di seguito descritto.

Media del prodotto di variabili aleatorie indipendenti: Se X e Y sono indipendenti e ammettono media, allora anche XY ha media e

$$E(XY) = E(X) E(Y). \quad (5.5.2)$$

Dimostrazione Tralasciamo la questione dell'esistenza della media del prodotto e concentriamoci sul suo calcolo, verificando la proprietà (5.5.2) nel caso continuo. Se X e Y sono continue con densità f_X e f_Y , allora:

$$E(XY) = \int_{\mathbb{R}} \int_{\mathbb{R}} xy f_X(x) f_Y(y) dx dy = \int_{\mathbb{R}} x f_X(x) dx \cdot \int_{\mathbb{R}} y f_Y(y) dy = E(X) E(Y). \quad \blacksquare$$

Osservazione 5.5.2 Iterando il procedimento della dimostrazione appena sviluppata è immediato verificare che se X_1, \dots, X_n sono n variabili aleatorie indipendenti che ammettono media, allora anche $\prod_{i=1}^n X_i$ ammette media e $E(\prod_{i=1}^n X_i) = \prod_{i=1}^n E(X_i)$.

5.5.1 Somme di variabili aleatorie

Un'interessante trasformazione di n variabili aleatorie X_1, X_2, \dots, X_n è la *somma di variabili aleatorie*: $\sum_{j=1}^n X_j$.

Esempio 5.5.3 Per esempio siamo interessati a studiare una somma se un canale di comunicazione è disturbato e sovrappone un “rumore” Y al segnale trasmesso X . In tal caso, il segnale ricevuto è $Z = X + Y$.

Media della somma. Siano X, Y variabili aleatorie definite sul medesimo spazio campionario e che hanno media. Allora anche la somma $X + Y$ ha media e

$$E(X + Y) = E(X) + E(Y). \quad (5.5.3)$$

Dimostrazione della proprietà (5.5.3) Tralasciamo la questione dell'esistenza della media della somma e concentriamoci sul suo calcolo, verificando la proprietà (5.5.3) nel caso di variabili aleatorie congiuntamente continue:

$$\begin{aligned} E(X + Y) &= \int_{\mathbb{R}^2} (x + y) f_{X,Y}(x, y) dx dy \\ &= \int_{\mathbb{R}} x \left(\int_{\mathbb{R}} f_{X,Y}(x, y) dy \right) dx + \int_{\mathbb{R}} y \left(\int_{\mathbb{R}} f_{X,Y}(x, y) dx \right) dy \\ &= \int_{\mathbb{R}} x f_X(x) dx + \int_{\mathbb{R}} y f_Y(y) dy = E(X) + E(Y). \end{aligned}$$

La dimostrazione procede analogamente se X e Y sono variabili aleatorie discrete. \blacksquare

Osservazione 5.5.4 È importante osservare che la media di $X + Y$ dipende soltanto dalle densità marginali f_X, f_Y . In generale, la media della somma di $n \geq 2$ variabili aleatorie X_1, \dots, X_n è data dalla somma delle n medie:

$$\boxed{E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n)} \quad (5.5.4)$$

Varianza della somma. Occupiamoci ora del problema del calcolo della varianza della somma di variabili aleatorie.

Se X ha varianza $\text{Var}(X)$ (finita) e Y ha varianza $\text{Var}(Y)$ (finita), allora anche $X + Y$ ha varianza finita e

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2E[(X - E(X))(Y - E(Y))]. \quad (5.5.5)$$

Inoltre, se X e Y sono indipendenti allora

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \quad (5.5.6)$$

Dimostrazione delle Formule (5.5.5) e (5.5.6). Al solito tralasciamo la questione dell'esistenza della varianza della somma e concentriamoci sul suo calcolo.

$$\begin{aligned} \text{Var}(X + Y) &= E[((X - E(X)) + (Y - E(Y)))^2] \\ &= E[(X - E(X))^2 + (Y - E(Y))^2 + 2(X - E(X))(Y - E(Y))] \\ &= E[(X - E(X))^2] + E[(Y - E(Y))^2] + 2E[(X - E(X))(Y - E(Y))] \\ &= \text{Var}(X) + \text{Var}(Y) + 2E[(X - E(X))(Y - E(Y))]. \end{aligned}$$

e la proprietà (5.5.5) è dimostrata. Riguardo a (5.5.6), basta notare che se X, Y sono indipendenti, allora anche $X - E(X)$ e $Y - E(Y)$ sono indipendenti³ e

$$E[(X - E(X))(Y - E(Y))] = E[(X - E(X))]E[Y - E(Y)] = [E(X) - E(X)][E(Y) - E(Y)] = 0$$

■

Esercizio* 5.5.5 Si dimostri che la varianza della somma di n variabili aleatorie X, \dots, X_n è data da:

$$\text{Var}(X + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n E[(X_i - E(X_i))(X_j - E(X_j))]$$

Osservazione 5.5.6 In generale, la varianza della somma di $n \geq 2$ variabili aleatorie X_1, \dots, X_n indipendenti è data dalla somma delle n varianze:

$$\boxed{X_1, \dots, X_n \text{ indipendenti} \implies \text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)} \quad (5.5.7)$$

³si veda l'Osservazione 5.1.2

5.6 Covarianza, coefficiente di correlazione

Abbiamo visto che se X e Y sono variabili aleatorie con varianza finita allora:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{E}[(X - \text{E}(X))(Y - \text{E}(Y))]$$

L'ultimo addendo nella somma a destra è di per sè rilevante in probabilità. Quindi introduciamo la seguente definizione.

Definizione 5.6.1 Siano X, Y due variabili aleatorie definite sul medesimo spazio campionario e che ammettono varianza. Si definisce *covarianza* di X, Y il numero

$$\text{Cov}(X, Y) = \text{E}[(X - \text{E}(X))(Y - \text{E}(Y))].$$

Se $\text{Var}(X) > 0$ e $\text{Var}(Y) > 0$, si definisce *coefficiente di correlazione* di X, Y il numero:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$

Covarianza e coefficiente di correlazione godono delle proprietà elencate nella seguente proposizione.

Proposizione 5.6.2 Siano X, Y, Z variabili aleatorie con varianza finita e $a, b \in \mathbb{R}$. Allora

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
2. $\text{Cov}(aX + b, Y) = a \text{Cov}(X, Y)$;
3. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$;
4. $\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X) \text{E}(Y)$;
5. se X, Y sono indipendenti allora $\text{Cov}(X, Y) = 0$;
6. $|\rho_{X,Y}| \leq 1$ e $|\rho_{X,Y}| = 1$ se e solo se esistono $a, b \in \mathbb{R}$ tali che $P(Y = aX + b) = 1$. Inoltre in tal caso:

$$a = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \quad e \quad b = \text{E}(Y) - a \text{E}(X).$$

Dimostrazione Le proprietà 1.–5. seguono immediatamente dalle proprietà della media e la dimostrazione viene lasciata per esercizio al lettore.

La dimostrazione della proprietà 6. è mutuata da Ross (2004), pag. 329 e si basa sulle proprietà della varianza. Siano σ_1^2, σ_2^2 le varianze di X, Y , rispettivamente. Allora

$$\begin{aligned} 0 &\leq \text{Var}\left(\frac{X}{\sigma_1} + \frac{Y}{\sigma_2}\right) = \frac{\text{Var}(X)}{\sigma_1^2} + \frac{\text{Var}(Y)}{\sigma_2^2} + 2 \text{Cov}\left(\frac{X}{\sigma_1}, \frac{Y}{\sigma_2}\right) \\ &= \frac{\sigma_1^2}{\sigma_1^2} + \frac{\sigma_2^2}{\sigma_2^2} + 2 \frac{\text{Cov}(X, Y)}{\sigma_1 \sigma_2} \quad [\text{per il punto 2.}] \\ &= 2(1 + \rho_{X,Y}) \end{aligned}$$

da cui otteniamo $\rho_{X,Y} \geq -1$ Inoltre,

$$0 \leq \text{Var}\left(\frac{X}{\sigma_1} - \frac{Y}{\sigma_2}\right) = \frac{\text{Var}(X)}{\sigma_1^2} + \frac{\text{Var}(Y)}{\sigma_2^2} - 2 \frac{\text{Cov}(X, Y)}{\sigma_1 \sigma_2} = 2(1 - \rho_{X,Y})$$

e quindi $\rho_{X,Y} \leq 1$.

Per dimostrare la seconda parte della proprietà 6., osserviamo che $\rho_{X,Y} = 1$ se e solo se $\text{Var}(X/\sigma_1 - Y/\sigma_2) = 0$. Segue quindi dalle proprietà della varianza che

$$\rho_{X,Y} = 1 \text{ se e solo se } P\left(\frac{X}{\sigma_1} - \frac{Y}{\sigma_2} = \frac{E(X)}{\sigma_1} - \frac{E(Y)}{\sigma_2}\right) = 1.$$

Inoltre, $\rho_{X,Y} = 1$ se e solo se $\text{Cov}(X, Y) = \sigma_1\sigma_2$ e quindi $\rho_{X,Y} = 1$ se e solo se

$$Y = E(Y) + \frac{\text{Cov}(X, Y)}{\sigma_1^2}(X - E(X))$$

Invece, per $\rho_{X,Y} = -1$ valgono le seguenti equivalenze che completano la dimostrazione: $\rho_{X,Y} = -1$ se e solo se (sse) $\text{Cov}(X, Y) = -\sigma_1\sigma_2$ sse $\text{Var}(X/\sigma_1 + Y/\sigma_2) = 0$ sse $P(X/\sigma_1 + Y/\sigma_2 = E(X)/\sigma_1 + E(Y)/\sigma_2) = 1$ sse

$$Y = E(Y) + \frac{\text{Cov}(X, Y)}{\sigma_1^2}(X - E(X)). \quad \blacksquare$$

Nota 5.6.3 Il punto 6. della Proposizione 5.6.2 illustra un noto risultato della teoria della *regressione lineare*: esiste un legame di tipo lineare fra le variabili aleatorie X e Y (cioè $Y = aX + b$) se e solo se $\rho(X, Y) = \pm 1$, inoltre $\rho(X, Y) = -1$ implica $\text{Cov}(X, Y) < 0$ e $a < 0$ mentre $\rho(X, Y) = 1$ implica $\text{Cov}(X, Y) > 0$ e $a > 0$.

Nota 5.6.4 Attenzione a “comprendere correttamente” la proprietà 5. la quale afferma che se X, Y sono indipendenti allora la covarianza è nulla (anche se X, Y non sono indipendenti). In generale il viceversa è falso, come mostra il seguente controesempio:

Esempio 5.6.5 Sia X una variabile aleatoria discreta con densità uniforme su $\{-1, 0, 1\}$ e sia $Y = X^2$. Allora $E(X) = 0$ in quanto X è una variabile aleatoria simmetrica e $E(XY) = E(X^3) = (-1)^3/3 + 1^3/3 = 0$, da cui $\text{Cov}(X, Y) = 0$. Ma, chiaramente, X e Y non sono indipendenti.

Esercizio 5.6.6 (Continuazione dell'Esercizio 5.2.6) La densità congiunta di X, Y e le densità marginali di X ed Y sono riportate nelle seguente tabella:

$X \setminus Y$	0	1	2	f_X
0	0	2/35	2/35	4/35
1	3/35	12/35	3/35	18/35
2	6/35	6/35	0	12/35
3	1/35	0	0	1/35
f_Y	10/35	20/35	5/35	1

Deriviamo che

$$\begin{aligned}
 E(X) &= 1 \times \frac{18}{35} + 2 \times \frac{12}{35} + 3 \times \frac{3}{35} = \frac{9}{7} \\
 E(X^2) &= 1 \times \frac{18}{35} + 2^2 \times \frac{12}{35} + 3^2 \times \frac{3}{35} = \frac{93}{35} \\
 \text{Var}(X) &= \frac{93}{35} - \left(\frac{9}{7}\right)^2 = \frac{246}{245} \\
 E(Y) &= 1 \times \frac{20}{35} + 2 \times \frac{5}{35} = \frac{6}{7} \\
 E(Y^2) &= 1 \times \frac{20}{35} + 2^2 \times \frac{5}{35} = \frac{8}{7} \\
 \text{Var}(Y) &= \frac{8}{7} - \left(\frac{6}{7}\right)^2 = \frac{20}{49} \\
 E(XY) &= 1 \times 1 \times \frac{12}{35} + 1 \times 2 \times \frac{3}{35} + 2 \times 1 \times \frac{6}{35} = \frac{6}{7} \\
 \text{Cov}(X, Y) &= \frac{6}{7} - \frac{9}{7} \times \frac{6}{7} = -\frac{12}{49} \\
 \rho(X, Y) &= \frac{-\frac{12}{49}}{\sqrt{\frac{246}{245} \times \frac{20}{49}}} \simeq -0.3825
 \end{aligned}$$

Esercizio 5.6.7 (Continuazione dell'Esercizio 5.3.7) Se X, Y hanno densità congiunta:

$$f_{X,Y}(x, y) = \begin{cases} x + y & \text{se } 0 < x < 1, 0 < y < 1 \\ 0 & \text{altrove} \end{cases}$$

allora $E(XY) = \frac{1}{3}$ (cfr. Esercizio 5.5.1) e

$$\begin{aligned}
 E(Y) &= E(X) = \int_0^1 x \left(x + \frac{1}{2}\right) dx = \frac{7}{12} \\
 E(Y^2) &= E(X^2) = \int_0^1 x^2 \left(x + \frac{1}{2}\right) dx = \frac{5}{12} \\
 \text{Var}(X) &= \text{Var}(Y) = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144} \\
 \text{Cov}(X, Y) &= \frac{1}{3} - \frac{7}{12} \times \frac{7}{12} = -\frac{1}{144} \\
 \rho(X, Y) &= \frac{-\frac{1}{144}}{\sqrt{\frac{11}{144} \times \frac{11}{144}}} = -\frac{1}{11}.
 \end{aligned}$$

5.6.1 Riproducibilità⁴

Il calcolo della distribuzione di una somma di variabili aleatorie è un problema solitamente difficile. Ci sono tuttavia alcuni casi notevoli nei quali tale problema ha una soluzione molto semplice. Ad esempio, l'intuito suggerisce che, se $X \sim \mathbf{Bi}(n_1, p)$, $Y \sim \mathbf{Bi}(n_2, p)$ e X, Y sono

⁴Estratto da Verri, M. AA 09/10 (IX) - Appunti di lezione

indipendenti, allora $X + Y \sim \mathbf{Bi}(n_1 + n_2, p)$. Infatti, si può pensare a una serie di $n_1 + n_2$ prove bernoulliane, indipendenti e tutte con la stessa probabilità di successo p , in cui X conta il numero di successi sulle prime n_1 prove e Y il numero di successi sulle ultime n_2 . Quindi $X + Y$ è il numero di successi su tutte le $n_1 + n_2$ prove. Questo risultato è noto come *proprietà di riproducibilità* della densità binomiale. La densità binomiale e altre densità che soddisfano una proprietà di riproducibilità sono elencate nella Tabella 5.1, dove si intende sempre che X e Y sono variabili aleatorie indipendenti.

Modello	X	Y	$X + Y$
Binomiale	$X \sim \mathbf{Bi}(n_1, p)$	$\mathbf{Bi}(n_2, p)$	$X + Y \sim \mathbf{Bi}(n_1 + n_2, p)$
Poisson	$X \sim \text{Pois}(\lambda_1)$	$Y \sim \text{Pois}(\lambda_2)$	$X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$
Gaussiano	$X \sim \mathcal{N}(\mu_1, \sigma_1^2)$	$Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$	$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
Erlang	$X \sim \Gamma(n_1, \lambda)$	$Y \sim \Gamma(n_2, \lambda)$	$X + Y \sim \Gamma(n_1 + n_2, \lambda)$

Tabella 5.1: Riproducibilità di alcune variabili aleatorie “notevoli”

Nota 5.6.8 Naturalmente le precedenti proprietà di riproducibilità valgono anche quando sommiamo un numero $n \geq 2$ di variabili aleatorie indipendenti. In particolare se X_1, \dots, X_n oltre che essere indipendenti sono anche regolate dalla stessa densità abbiamo:

Modello	X_1, \dots, X_n	$X_1 + \dots + X_n$
Poisson	$X_i \sim \text{Pois}(\lambda), \quad i = 1, \dots, n$	$X_1 + \dots + X_n \sim \text{Pois}(n\lambda)$
Gaussiano	$X_i \sim \mathcal{N}(\mu, \sigma^2), \quad i = 1, \dots, n$	$X_1 + \dots + X_n \sim \mathcal{N}(n\mu, n\sigma^2)$
Erlang	$X_i \sim \Gamma(m, \lambda), \quad i = 1, \dots, n$	$X_1 + \dots + X_n \sim \Gamma(nm, \lambda)$

Esercizio 5.6.9 (Variabile binomiale come somma di bernoulliane indipendenti) Siano X_1, \dots, X_n n variabili aleatorie indipendenti con densità di Bernoulli di parametro $p \in (0, 1)$.

1. Dimostrate che $X_1 + \dots + X_n \sim \mathbf{Bi}(n, p)$.
2. Sfruttando il punto 1., dimostrate che la media di una va $\mathbf{Bi}(n, p)$ è np e la varianza è $np(1 - p)$.

Esercizio 5.6.10 Un certo prodotto si ottiene assemblando tre componenti. La lunghezza totale Y del prodotto è uguale alla somma delle lunghezze X_1, X_2 e X_3 dei suoi componenti. Si assuma che le lunghezze dei componenti siano variabili aleatorie indipendenti e $X_1 \sim \mathcal{N}(2, 0.01)$, $X_2 \sim \mathcal{N}(3, 0.04)$, $X_3 \sim \mathcal{N}(4, 0.04)$.

1. Calcolare la probabilità che la lunghezza di quel prodotto rispetti il limite di tolleranza del 2% intorno alla media.

Soluzione Per la proprietà di riproducibilità del modello gaussiano abbiamo che $Y = X_1 + X_2 + X_3 \sim \mathcal{N}(2 + 3 + 4, 0.01 + 0.04 + 0.04)$, cioè $Y \sim \mathcal{N}(9, 0.09)$. Pertanto:

$$P(|Y - \mu_Y| \leq 0.02\mu_Y) = P\left(\frac{|Y - 9|}{0.3} \leq \frac{0.02 \times 9}{0.03}\right) = 2\Phi(0.6) - 1 \simeq 2 \times 0.72575 - 1 = 0.4515. \blacksquare$$

5.6.2 Densità di Erlang

Nell'ultima riga della Tabella 5.1 viene indicato il modello di Erlang che non avevamo ancora introdotto.

Definizione 5.6.11 Una variabile aleatoria assolutamente continua X con densità

$$f(x) = \begin{cases} \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0, \end{cases} \quad n = 1, 2, \dots, \quad \lambda > 0 \quad (5.6.1)$$

è detta *densità di Erlang di parametri n e λ* e si indica $X \sim \Gamma(n, \lambda)$.

Se $n = 1$, allora $\frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} = \lambda e^{-\lambda x}$, cioè la densità esponenziale $\mathcal{E}(\lambda)$ è un caso particolare del modello di Erlang con $n = 1$.

Si può dimostrare che

1. una variabile aleatoria di Erlang $X \sim \Gamma(n, \lambda)$ ha valore atteso e varianza date da

$$E(X) = \frac{n}{\lambda}, \quad \text{Var}(X) = \frac{n}{\lambda^2};$$

2. Se X_1, X_2, \dots, X_n sono variabili aleatorie indipendenti di densità esponenziale di parametro $\lambda > 0$, allora la densità di $X_1 + \dots + X_n$ è Erlang di parametri n, λ . Da qui deriva la proprietà di riproducibilità riportata in Tabella 5.1 e l'applicazione del modello di Erlang nella "teoria delle code", per modellare il tempo di attesa dell' n -esimo arrivo quando ad ogni arrivo si riaggiorna l'orologio per l'attesa del successivo, cioè vale una proprietà di assenza di memoria, cosicché gli intertempi fra due arrivi successivi hanno densità esponenziale e sono indipendenti.

Infine, il simbolo $\Gamma(n, \lambda)$ usato per indicare la densità di Erlang è giustificato dal fatto che le densità di Erlang appartengono a una famiglia più grande di densità di variabili aleatorie assolutamente continue dette densità *Gamma*.

5.7 Densità bivariata gaussiana

Concludiamo la trattazione delle coppie di variabili, descrivendo brevemente il modello gaussiano bivariato.

Definizione 5.7.1 Sia (X, Y) una coppia di variabili aleatorie definite sullo stesso spazio campionario che ammette densità congiunta

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right]}$$

Questa densità è detta *densità gaussiana bivariata* e si scrive $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$.

Si può dimostrare che se $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$ allora

1. $aX + bY + c \sim \mathcal{N}(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2)$, per ogni a, b, c a patto che o $a \neq 0$ o $b \neq 0$;
2. le densità marginali di X, Y sono gaussiane con $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ e $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$;
3. il parametro ρ è il coefficiente di correlazione lineare di X, Y ;
4. se $\rho = 0$ allora $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ e quindi X, Y sono indipendenti.

Capitolo 6

Media campionaria e teoremi limite

6.1 Campione casuale e media campionaria

Consideriamo una popolazione di elementi, a ciascuno dei quali è associata una certa grandezza numerica e sulla quale si vogliono ottenere informazioni a partire da un campione di n dati x_1, \dots, x_n estratti da questa popolazione. Per esempio consideriamo la popolazione dei bambini italiani che frequentano i nidi di infanzia e siamo interessati ai redditi annuali dei loro genitori, oppure alla composizione della loro famiglia. Allora x_1 rappresenta il reddito in euro della famiglia del primo bambino estratto, x_2 quello del secondo bambino e così via. I dati x_1, \dots, x_n sono pensati come realizzazioni di n variabili aleatorie X_1, \dots, X_n indipendenti e identicamente distribuite (i.i.d.), cioè tutte aventi la stessa densità f . Questa ipotesi traduce un fatto sensato cioè che il campione x_1, \dots, x_n è selezionato in maniera casuale.

Definizione 6.1.1 *Un campione casuale lungo n estratto da una popolazione di densità f è un insieme X_1, \dots, X_n di variabili aleatorie i.i.d.. La media $\mu = E(X_1)$, che è uguale per tutte le X_i del campione, prende il nome di *media della popolazione* e la comune varianza σ^2 è detta *varianza della popolazione**

Si chiama *media campionaria* di un campione casuale X_1, \dots, X_n la quantità

$$\frac{X_1 + \dots + X_n}{n} \tag{6.1.1}$$

e la si indica con \bar{X}_n . La media campionaria dipende dal campione, quindi è una variabile aleatoria. Valore atteso e varianza di \bar{X} si possono facilmente calcolare. Applicando le proprietà di valore atteso e varianza abbiamo:

$$E(\bar{X}_n) = E\left(\frac{\sum_{j=1}^n X_j}{n}\right) = \frac{E\left(\sum_{j=1}^n X_j\right)}{n} = \frac{n \times \mu}{n} = \mu$$

e

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{\sum_{j=1}^n X_j}{n}\right) = \frac{n \text{Var}(X_1)}{n^2} = \frac{\sigma^2}{n}$$

Questo significa che la media campionaria \bar{X}_n è centrata nella media della popolazione μ con una varianza che è più piccola della varianza della popolazione σ^2 di un fattore $1/n$, cosicché la dispersione della media campionaria \bar{X}_n intorno a μ diminuisce man mano che aumentano le variabili aleatorie che campioniamo. Questo risultato è chiaramente descritto dai risultati sulle leggi dei grandi numeri.

6.2 Legge dei grandi numeri

Lanciamo un numero elevato n di volte una moneta (cioè consideriamo un esperimento ripetibile infinite volte) e consideriamo la frequenza relativa di testa negli n lanci. Questa frequenza relativa può essere rappresentata come una media campionaria \bar{X}_n di un campione casuale X_1, \dots, X_n , dove X_i vale 1 se il risultato della i -esima prova è testa, 0 altrimenti. Se la moneta non è truccata ci aspettiamo che, salvo in casi eccezionali, questa frequenza tenda a stabilizzarsi intorno a $1/2$, al crescere di n .

Più in generale in tutte le situazioni pratiche in cui è possibile eseguire molte repliche di un esperimento aleatorio si osserva tipicamente che la frequenza relativa *empirica* delle volte in cui un dato evento A si verifica tende a stabilizzarsi intorno a un valore costante, che viene interpretato come la probabilità $P(A)$ di A . Questo risultato noto come *stima frequentista della probabilità* trova una conferma teorica nella “Legge dei grandi numeri”. Si parla di Legge debole dei grandi numeri e di Legge forte dei grandi numeri. La prima è una conseguenza immediata della disuguaglianza di Chebychev:

Proposizione 6.2.1 (Legge debole dei grandi numeri) *Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti¹ ed identicamente distribuite (i.i.d.) con media μ e varianza σ^2 finite e sia \bar{X}_n la media campionaria di X_1, \dots, X_n . Allora, per ogni $\epsilon > 0$*

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \epsilon) = 0.$$

Dimostrazione Poiché le X_i sono i.i.d. e $E(\bar{X}_n) = \mu$ e $\text{Var}(\bar{X}_n) = \sigma^2/n$, allora segue dalla disuguaglianza di Chebychev che per ogni $\epsilon > 0$

$$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \rightarrow 0 \quad (n \rightarrow +\infty) \quad \blacksquare$$

la Legge debole dei grandi numeri mette in evidenza che, pur partendo da un esperimento aleatorio costituito da prove ripetute, di ciascuna delle quali poco si può predire (le prove sono indipendenti), facendo le medie di tali prove si ottiene un risultato che può essere predetto con un elevato grado di certezza. In realtà vale un risultato “più forte” la cui dimostrazione è più laboriosa.

Proposizione 6.2.2 (Legge forte dei grandi numeri) *Sia X_1, X_2, \dots una successione di variabili aleatorie i.i.d. con media finita μ . Allora*

$$P(\{\omega : \lim_{n \rightarrow +\infty} \bar{X}_n(\omega) = \mu\}) = 1.$$

¹Una successione X_1, X_2, \dots di variabili aleatorie sono *indipendenti* se per ogni $m \geq 2$, le variabili aleatorie X_1, \dots, X_m sono indipendenti

In pratica la legge forte applicata all'esempio dei lanci di una moneta dice che per "quasi tutte" le successioni di risultati X_1, X_2, \dots la frequenza relativa di testa \bar{X}_n converge al trucco p della moneta.

Osservazione 6.2.3 La Legge dei grandi numeri fornisce una giustificazione teorica alla scelta che si fa in statistica di "stimare", cioè approssimare, una media incognita μ con la media campionaria \bar{X}_n .

Esempio 6.2.4 (Istogramma) ² Un'altra applicazione della legge debole dei grandi numeri è in statistica descrittiva nell'uso degli istogrammi per approssimare una densità continua f . Sia X_1, \dots, X_n un campione casuale numeroso X_1, \dots, X_n estratto da f . Suddividiamo l'insieme dei possibili valori che ogni X può assumere in intervalli, detti *classi*, e, a ogni classe associamo un rettangolo, detto *canale*, che ha per base l'intervallo stesso e per altezza la percentuale di osservazioni che cadono nella classe divisa per l'ampiezza della classe. Per ogni classe I , in base alla teoria abbiamo

$$P(X \in I) = \int_I f(x) dx$$

e, se l'ampiezza Δx dell'intervallo I è abbastanza piccola,:

$$P(X \in I) \simeq f(x^0) \Delta x^0 \quad (6.2.1)$$

dove x^0 è il punto centrale di I . D'altro canto, per quanto detto prima sulla stima frequentista della probabilità, la frequenza relativa (%) di osservazioni in I fornisce una stima di $P(X \in I)$:

$$P(X \in I) \simeq \% \text{ di osservazioni in } I$$

da cui, sostituendo in (6.2.1), otteniamo

$$f(x) \simeq \frac{\% \text{ di osservazioni in } I}{\Delta x}$$

il che appunto significa approssimare la densità su I con il corrispondente canale e l'intera densità con l'unione dei canali, cioè con l'istogramma.

In laboratorio avete visto come costruire praticamente gli istogrammi (numero di canali, ampiezza dei canali...).

Esempio 6.2.5 (Metodo di integrazione Monte Carlo) Sia h una funzione continua su $[0, 1]$. Vogliamo calcolare in modo approssimato $\int_0^1 h(x) dx$. Esistono molte formule di quadratura, ma la tecnica Monte Carlo è una delle più semplici. Inoltre, anche se può non risultare il miglior metodo per funzioni su $[0, 1]$, si estende facilmente e diventa competitiva nel caso di integrali multidimensionali. Infatti, nei metodi numerici "tradizionali", l'errore di approssimazione dipende dalla dimensione, mentre ciò non accade nel caso del metodo Monte Carlo. I generatori di numeri casuali in ogni libreria di sistema producono valori le cui proprietà si avvicinano alle realizzazioni di variabili aleatorie i.i.d. con densità uniforme su $(0,1)$ e rendono implementabile il metodo Monte Carlo basato sul seguente corollario alla Legge forte dei grandi numeri:

²Estratto da Verri, M. AA 09/10 (X) - Appunti di lezione, *Il Teorema del limite centrale*

Corollario 6.2.6 Sia h una funzione su $[0, 1]$ con $\int_0^1 |h(x)| dx < +\infty$. Siano U_1, U_2, \dots variabili aleatorie i.i.d. con densità uniforme su $[0, 1]$. Allora

$$P \left(I_{1n} := \frac{1}{n} \sum_{j=1}^n h(U_j) \rightarrow \int_0^1 h(x) dx, n \rightarrow +\infty \right) = 1$$

Dimostrazione È sufficiente osservare che le variabili aleatorie $h(U_1), h(U_2), \dots$ sono i.i.d. con media finita $\int_0^1 h(x) dx$ ed applicare la Legge forte dei grandi numeri. ■

Il metodo Monte Carlo consiste nell'approssimare $\int_0^1 h(x) dx$ con I_{1n} per n “grande”. Per ogni n fissato, la bontà dell'approssimazione può essere valutata tramite

$$\text{Var}(I_{1n}) = \text{Var} \left(\frac{1}{n} \sum_{j=1}^n h(U_j) \right) = \frac{\int_0^1 h^2(x) dx - (\int_0^1 h(x) dx)^2}{n}.$$

Al fine di ridurre la varianza, il metodo delle “variabili antitetiche” approssima il valore dell'integrale mediante

$$I_{2n} := \frac{1}{2n} \sum_{i=1}^n (h(U_i) + h(1 - U_i)).$$

6.3 Teorema centrale del limite

Consideriamo n variabili aleatorie X_1, \dots, X_n i.i.d. con media μ e varianza σ^2 , entrambe finite. Abbiamo visto nella precedente sezione che per n “grande”, la media campionaria \bar{X}_n approssima in un opportuno senso la media μ :

$$\bar{X}_n \simeq \mu.$$

Se inoltre X_1, \dots, X_n sono gaussiane, allora è immediato verificare che $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, e quindi siamo in grado di valutare probabilisticamente la “dispersione” dei valori assunti da \bar{X}_n intorno a μ : ad esempio, osservando che

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim \mathcal{N}(0, 1),$$

otteniamo

$$P(|\bar{X}_n - \mu| \leq \delta) = \Phi\left(\frac{\sqrt{n}\delta}{\sigma}\right) - \Phi\left(-\frac{\sqrt{n}\delta}{\sigma}\right) = 2\Phi\left(\frac{\sqrt{n}\delta}{\sigma}\right) - 1$$

che si calcola usando le tavole della funzione di ripartizione gaussiana standard.

In questa sezione presenteremo una versione semplice del *Teorema centrale del limite* (o *Teorema del limite centrale*) il cui significato euristico è il seguente: la media campionaria di un numero n , sufficientemente grande, di variabili aleatorie i.i.d., di media μ e varianza σ^2 finite ha una funzione di ripartizione che è approssimativamente gaussiana di media μ e varianza σ^2/n .

Teorema 6.3.1 Sia X_1, X_2, \dots una successione di variabili aleatorie i.i.d. con media μ e varianza σ^2 , con $0 < \sigma^2 < +\infty$. Allora per ogni $x \in \mathbb{R}$:

$$\lim_{n \rightarrow +\infty} P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq x\right) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \Phi(x). \quad (6.3.1)$$

Il teorema può essere interpretato nel modo seguente: pur di prendere un numero elevato di variabili nella successione, la funzione di ripartizione di $\sqrt{n}(\bar{X}_n - \mu)/\sigma$, cioè della standardizzata della media campionaria \bar{X}_n , è approssimabile con quella gaussiana standard. Quindi, per quanto visto sulle standardizzate di variabili aleatorie gaussiane, approssimativamente \bar{X}_n ha funzione di ripartizione gaussiana di media μ e varianza σ^2/n . La bontà dell'approssimazione dipende dal numero di variabili aleatorie sommate e dalla forma della funzione di ripartizione delle variabili aleatorie di cui si fa la media.

Equivalentemente, l'enunciato del teorema centrale del limite può essere dato in termini di somme di variabili aleatorie i.i.d.. Infatti

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\sqrt{n}(\sqrt{n}(\bar{X}_n - \mu))}{\sqrt{n}\sigma} = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

cioè la standardizzata di \bar{X}_n coincide con quella di S_n . Quindi, sotto le ipotesi del teorema centrale del limite:

$$\lim_{n \rightarrow +\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \Phi(x).$$

Interpretazione “pratica” del Teorema centrale del limite³ Il teorema centrale del limite può essere generalizzato al caso di una successione di variabili aleatorie indipendenti ma non identicamente distribuite. La conclusione del teorema è sempre la stessa, cioè per n grande una \bar{X}_n ha approssimativamente fdr $\mathcal{N}(\mu, \sigma^2/n)$, se “nessuna delle X_i ha una variabilità⁴ preponderante sulle altre”. Questa condizione è certamente soddisfatta se le variabili X_i sono identicamente distribuite, perché in tal caso avranno tutte la stessa varianza. La versione più generale del Teorema centrale del limite gioca un ruolo importante nelle applicazioni; la sua interpretazione pratica è che *ogni effetto casuale tende a distribuirsi normalmente se è dato dalla sovrapposizione di numerosi fattori aleatori indipendenti che, singolarmente presi, hanno influenza trascurabile sul risultato finale*. Ad esempio, gli errori di misura si possono rappresentare come somma di un numero elevato di singoli termini (errori elementari), ciascuno dei quali è dovuto ad una causa, non dipendente dalle altre. Quali che siano le funzioni di ripartizione degli errori elementari, le peculiarità di queste non si manifestano nella somma di un gran numero di termini e la funzione di ripartizione della somma è vicina alla funzione di ripartizione gaussiana.

Esempio 6.3.2 L'altezza delle persone tende a distribuirsi normalmente intorno a un “valore medio” μ perché essa è dovuta alla somma delle lunghezze delle varie parti dello scheletro; la variabilità di ciascuna parte è piccola rispetto allo scostamento totale dell'altezza dalla media μ . Un analogo discorso vale per il peso delle persone.

³Estratto da Verri, M. AA 09/10 (X) - Appunti di lezione, *Il Teorema del limite centrale*

⁴leggi varianza

Poiché diverse variabili aleatorie di uso comune si possono rappresentare come somma di numerose variabili i.i.d., allora il teorema centrale del limite può essere usato per approssimare le vere funzioni di ripartizione di queste variabili.

Seguono alcuni esempi di applicazione del teorema centrale del limite.

Esempio 6.3.3 Nella Sezione 4.4 abbiamo discusso la possibilità di approssimare la funzione di ripartizione binomiale con quella gaussiana, sulla base del Teorema 4.4.1 di De Moivre Laplace. Effettivamente, il Teorema 4.4.1 di De Moivre Laplace è un caso particolare del teorema centrale del limite. In realtà esso rappresenta una prima versione del teorema centrale del limite. Infatti, una variabile aleatoria binomiale $\mathbf{Bi}(n, p)$ ha la stessa densità della somma di n variabili aleatorie i.i.d. di Bernoulli di parametro $p \in (0, 1)$.

Rimandiamo all'Esempio 4.4.4 per la discussione sulla bontà della approssimazione.

Invece, per quanto concerne la correzione di continuità, può essere utile ricordare qui come si apporta nel caso di una somma di variabili aleatorie indipendenti a valori interi (ma non necessariamente bernoulliane).

Se X_1, \dots, X_n sono variabili aleatorie i.i.d. discrete e a valori interi con comune media μ e comune varianza $\sigma^2 > 0$, $S_n = \sum_{j=1}^n X_j$ ed n è grande, la *correzione di continuità* si apporta nel seguente modo:

$$P(S_n \leq r) \simeq \Phi\left(\frac{r + 0.5 - n\mu}{\sqrt{n\sigma^2}}\right), \quad \forall r \text{ intero.}$$

Esempio 6.3.4 Sia X una variabile aleatoria di Poisson di parametro $\lambda = 100$. Calcolare un valore approssimato di $P(X < 110)$.

La variabile aleatoria $X \sim \mathcal{P}(100)$ ha la stessa densità della somma di 100 variabili aleatorie Y_1, \dots, Y_{100} i.i.d. $\sim \mathcal{P}(1)$; queste variabili aleatorie sono discrete a valori interi e hanno media e varianza pari a 1. Quindi, per il teorema centrale del limite, la fdr $\mathcal{P}(100)$ si può approssimare con la fdr $\mathcal{N}(100, 100)$. Inoltre, l'approssimazione è migliore con la correzione di continuità. In particolare:

$$\begin{aligned} P(X < 110) &= P(X \leq 109) = P\left(\sum_{j=1}^{100} Y_j \leq 109\right) = P\left(\sum_{j=1}^{100} Y_j \leq 109.5\right) = \\ &= P\left(\frac{\sum_{j=1}^{100} Y_j - 100}{10} \leq \frac{109.5 - 100}{10}\right) \simeq \Phi\left(\frac{109.5 - 100}{10}\right) \simeq 0.8289 \end{aligned}$$

Senza la correzione di continuità, un valore approssimato di $P(X < 110)$ è dato da $\Phi((109 - 100)/10) \simeq 0.8159$. (Il valore esatto di $P(X < 110)$ è 0.82944.)

Esempio 6.3.5 Siano U_1, \dots, U_{147} variabili aleatorie indipendenti e uniformi sull'intervallo $(0, 2)$ e $S = \sum_{j=1}^{147} U_j$. Calcolare un valore approssimato di $P(S < 161)$.

In quanto somma di variabili aleatorie i.i.d. assolutamente continue, anche S è assolutamente continua da cui $P(S < 161) = P(S \leq 161)$. Inoltre $E(S) = 147 \times E(U_1) = 147$ e $\text{Var}(S) = 147 \text{Var}(U_1) = 147/3 = 49$. Per il teorema centrale del limite, la funzione di ripartizione di $(S - E(S))/\sqrt{\text{Var } S}$ converge a Φ . Quindi, $P(S < 161) = F_S(161) \simeq \Phi((161 - 147)/\sqrt{49}) = \Phi(2) \simeq 0.9772$. (Qui non serve la correzione di continuità perché S è già continua!)

Capitolo 7

Test diagnostici, valutazione di riconoscitori e curve ROC

7.1 Test diagnostici

In un test clinico un individuo di una certa popolazione viene sottoposto a un'analisi di laboratorio (*test*) per sapere se ha o meno una certa malattia. Il risultato del test può essere *negativo*, a indicare che l'individuo è *sano* [rispetto a quella malattia], oppure *positivo*, a indicare che l'individuo è *malato*. Tuttavia tutti i test utilizzati in pratica non sono completamente affidabili, nel senso che può accadere che

- (a) sottoponendo un individuo sano al test, il test fornisca un risultato positivo (*falso positivo*)
- (b) sottoponendo un individuo malato al test, il test dia un risultato negativo (*falso negativo*).

Ovviamente un test è “buono” se rende minime le probabilità di osservare falsi positivi o falsi negativi. Così, per valutare la bontà di un test, prima di applicarlo su larga scala, lo si verifica su individui di cui si conosce lo stato di salute. Supponiamo di sottoporre un individuo a un test clinico, e siano M l'evento “l'individuo è malato”, M^c l'evento “l'individuo è sano”, Y l'evento “il test è positivo” e N l'evento “il test è negativo”. Le grandezze $\alpha = P(Y|M)$ e $\beta = P(N|M^c)$ sono note nella letteratura epidemiologica rispettivamente come *sensibilità* e *specificità* del test. In un buon test queste grandezze devono essere quanto più possibile prossime ad uno. In realtà fra le due vi è un trade-off. Descriveremo nella Sezione ?? come valutare la bontà del test e come confrontare test diversi sulla base di specificità e sensibilità.

Altre grandezze che si utilizzano per valutare la bontà di un test clinico/diagnostico sono il *potere predittivo positivo del test* o *PPV* (dall'inglese Positive Predictive Value) o *precisione del test*, definito come la probabilità che un individuo risultato positivo al test sia un “vero positivo” e il *potere predittivo negativo del test* o *NPV* (Negative Predictive Value) o *accuratezza del test*, dato dalla probabilità che un individuo risultato negativo al test sia un “vero negativo”. In termini di probabilità condizionate, abbiamo

$$PPV = P(M|Y) \quad \text{e} \quad NPV = P(M^c|N)$$

Sia infine $p = P(M)$ la frequenza relativa della malattia nella popolazione.

Un'applicazione diretta delle formule delle probabilità totali e di Bayes ci permette di vedere come le grandezze appena introdotte sono legate fra di loro. Infatti,

$$PPV = P(M|Y) = \frac{P(Y|M)P(M)}{P(Y|M)P(M) + P(Y|M^c)P(M^c)} = \frac{\alpha p}{\alpha p + (1 - \alpha)(1 - p)}$$

$$NPV = \frac{\beta(1 - p)}{\beta(1 - p) + (1 - \alpha)p}.$$

Quindi, il potere predittivo di un test dipende non solo dalle caratteristiche intrinseche del test date da specificità e sensibilità, ma anche dalla incidenza della malattia nella popolazione, misurata con p . Nei casi di screening p è molto piccola e quindi anche $P(M|I)$ è piccola, cosicché il test usato su una popolazione sana dà *quasi sempre* falsi positivi e veri negativi. Tanto per fare un esempio pratico consideriamo la metodica “ELISA” per la rilevazione degli anticorpi relativi al retrovirus HIV, in cui $\alpha = 0.993$ e $\beta = 0.9999$. Inoltre, nel '95 si stimava che gli individui che avevano sviluppato anticorpi relativi all'HIV in Italia fossero lo 0.0025% della popolazione totale, cioè $p = 0.0025\%$. Segue che

$$PPV = \frac{0.993 \times 0.000025}{0.993 \times 0.000025 + (1 - 0.9999) \times (1 - 0.000025)} \simeq 0.2 = 20\%$$

questo significa che se si effettuasse ELISA “a tappeto” su tutta la popolazione italiana l'80% circa dei positivi sarebbero falsi positivi! Per ovviare a questo inconveniente nella pratica si restringe la popolazione da esaminare alla cosiddetta “popolazione a rischio” (elevando in questo modo p) e si consiglia a chi è risultato positivo al test ELISA di sottoporsi ad un altro test, più costoso, ma anche più accurato.

7.1.1 Stima della sensibilità e specificità

La sensibilità α e la specificità β di un test sono stimate utilizzando il test su individui dei quali si conosce lo stato di salute. Scelto a caso un campione casuale di m individui malati e un secondo campione di n individui sani nella popolazione gli stimatori di α e β sono dati da:

$$\hat{\alpha} = \frac{n^0 \text{ di veri positivi}}{n^0 \text{ di malati}} = \frac{n^0 \text{ di veri positivi}}{m} \quad (7.1.1)$$

$$\hat{\beta} = \frac{n^0 \text{ di veri negativi}}{n^0 \text{ di sani}} = \frac{n^0 \text{ di veri negativi}}{n} \quad (7.1.2)$$

Gli stimatori $\hat{\alpha}$ e $\hat{\beta}$ sono entrambi proporzioni campionarie relative binomiali, pertanto i rispettivi errori standard (stimati) sono dati da $SD_{\hat{\alpha}} = \frac{\hat{\alpha}(1 - \hat{\alpha})}{m}$ e $SD_{\hat{\beta}} = \frac{\hat{\beta}(1 - \hat{\beta})}{n}$. Inoltre, un intervallo di confidenza per esempio 95% bilatero asintotico per α è $IC(\alpha) = \hat{\alpha} \pm 1.96 \times SD_{\hat{\alpha}}$ e per β è $IC(\beta) = \hat{\beta} \pm 1.96 \times SD_{\hat{\beta}}$. Le dimensioni campionarie richieste per una stima accurata di α e β sono ottenute applicando le regolette spiegate in classe per determinarne la numerosità campionaria necessaria per ottenere intervalli di confidenza (bilateri) di lunghezza minore o uguale a un livello specificato.

D'altro canto, queste proporzioni sono spesso prossime a 1 cosicché gli intervalli appena dati per grandi campioni potrebbero portare a risultati di stima non accurati. È così preferibile usare o intervalli di confidenza esatti per una proporzione di grande popolazione, oppure intervalli di confidenza basati sull'approssimazione di Poisson della legge Binomiale¹.

¹cfr. per esempio, Lawrence M. Leemis, Kishor S. Trivedi, (1996) A Comparison of Approximate Interval Estimators for the Bernoulli Parameter, *The American Statistician*, vol. 50, pp 63-68