

LEZIONI DI TELECOMUNICAZIONI

Introduzione alla Teoria dei Fenomeni Aleatori

Francesco Palmieri

*Dipartimento di Ingegneria dell'Informazione
Seconda Università di Napoli*

*Non è la materia che genera il pensiero,
è il pensiero che genera la materia.*

Giordano Bruno

Indice

Prefazione	9
Introduzione	11
Nota bibliografica	12
1 Introduzione alla Probabilità	15
1.1 Il modello probabilistico	15
1.2 Eventi	20
1.3 Il concetto di probabilità	25
1.4 Definizione assiomatica di probabilità	29
1.5 Spazi campione discreti e finiti	33
1.5.1 Eventi elementari equiprobabili	35
1.5.2 Il numero di successi in un campione	40
1.6 Spazi continui uniformi	43
1.7 Probabilità condizionata	45
1.7.1 La legge della probabilità totale	48
1.7.2 Il teorema di Bayes	49
1.8 Indipendenza	50
1.9 Indipendenza condizionata	53
1.10 Prove indipendenti	53
1.10.1 Prove di Bernoulli	55
1.10.2 Approssimazioni per la legge binomiale	57
1.11 La legge dei grandi numeri	62
1.12 Problemi	64

2	La variabile aleatoria	69
2.1	Introduzione	69
2.2	Definizione di variabile aleatoria	70
2.3	La funzione di distribuzione cumulativa	75
2.4	La funzione di densità di probabilità	81
2.5	Cdf e pdf notevoli	85
2.5.1	La variabile aleatoria uniforme	85
2.5.2	La variabile aleatoria esponenziale	87
2.5.3	La variabile aleatoria di Laplace	89
2.5.4	La variabile aleatoria di Poisson	90
2.5.5	La variabile aleatoria di Rayleigh	91
2.5.6	La variabile aleatoria binomiale	92
2.6	Funzione di massa di probabilità	94
2.7	La variabile aleatoria Gaussiana	96
2.8	I momenti della variabile aleatoria	100
2.8.1	La media	101
2.8.2	La moda e la mediana	105
2.8.3	La media come operatore	107
2.8.4	I momenti di ordine superiore	107
2.8.5	I momenti centrali	107
2.9	Disugliaglianza di Chebyshev	110
2.10	Problemi	113
3	Due variabili aleatorie	117
3.1	Introduzione	117
3.2	La funzione cumulativa congiunta	119
3.3	La funzione di densità congiunta	122
3.3.1	Variabili discrete e miste	124
3.3.2	La variabile aleatoria gaussiana bidimensionale	128
3.4	Variabili aleatorie indipendenti	130
3.5	Cdf e pdf condizionate	131
3.5.1	Momenti congiunti	137
3.5.2	Variabili aleatorie indipendenti, scorrelate, ortogonali	139
3.5.3	Due variabili aleatorie gaussiane	141

3.6	Momenti condizionati	141
4	Più variabili aleatorie	143
4.1	Introduzione	143
4.1.1	Variabili condizionate	145
4.1.2	Variabili indipendenti	147
4.2	Momenti	148
4.2.1	La media	149
4.2.2	La matrice di correlazione	149
4.3	Momenti centrali	151
4.3.1	La matrice di covarianza	151
4.4	La variabile gaussiana multidimensionale	152
4.4.1	Due insiemi di variabili gaussiane	154
4.5	Stima MMSE	156
4.5.1	Il caso gaussiano	159
5	Funzioni di variabili aleatorie	163
5.1	Introduzione	163
5.2	Funzioni di una variabile	164
5.2.1	Il metodo della cdf	167
5.2.2	Il metodo della pdf	172
5.3	Funzioni di due variabili	178
5.3.1	Metodo della cdf	180
5.3.2	Il metodo della pdf	185
5.3.3	Variabili discrete	186
5.4	Funzioni di n variabili	187
5.4.1	Metodo della cdf	187
5.4.2	Metodo della pdf	188
5.4.3	Il teorema centrale del limite	190
5.5	Problemi	191
A	La CDF Gaussiana	193
B	La funzione di Dirac	195
B.1	La funzione di Dirac come limite	198
B.2	Derivate dell'impulso di Dirac	199

8

INDICE

C Lemma inversione matrici

201

Bibliografia

203

Prefazione

Queste note sono il compendio di lezioni da me tenute in corsi introduttivi per le Lauree in Ingegneria Elettronica, Informatica e delle Telecomunicazioni presso l'Università di Napoli Federico II e presso la Seconda Università di Napoli. I corsi a cui sono state rivolte queste lezioni sono quelli del nuovo ordinamento degli studi universitari, introdotto recentemente in Italia. Nonostante esistano nella letteratura molti testi di riferimento sulla teoria della probabilità e della variabile aleatoria, nella fase di preparazione delle lezioni, ho riscontrato una certa carenza di riferimenti bibliografici, in lingua italiana, che fossero orientati alla nuova organizzazione didattica e in particolare agli studenti dei primi anni del settore dell'informazione.

Queste note sono state scritte nella speranza di riuscire a introdurre gli studenti di ingegneria, o di altri curricula scientifici, all'utilizzo di alcuni importanti strumenti metodologici di analisi e di sintesi mediante l'approccio probabilistico.

Lo scrivente è fermamente convinto che l'insegnamento dei corsi universitari "brevi" del nuovo ordinamento non debba trascurare gli aspetti di base delle materie, che sono bagaglio esclusivo degli studi universitari. La preparazione sui fondamenti deve armonizzarsi con i temi più "professionalizzanti," senza sacrificare gli aspetti essenziali del rigore metodologico. La velocità dei cambiamenti tecnologici impone di formare dei laureati che sappiano affrontare nella loro vita professionale in maniera sistematica l'analisi e la risoluzione di problemi sempre diversi.

Ovviamente questi appunti forniscono solo una traiettoria di

riferimento e quindi solo una base di studio su cui costruire ulteriori approfondimenti. Molti eccellenti testi sulla Teoria della Probabilità sono disponibili nella letteratura corrente. Ne ho suggeriti alcuni nella bibliografia.

Il tono delle lezioni è piuttosto discorsivo e il livello di rigore matematico è mantenuto al minimo. Per una lettura proficua delle note è sufficiente che si abbia conoscenza delle principali nozioni di calcolo matematico e di teoria degli insiemi. L'enfasi delle lezioni è sui numerosi esempi presentati sui vari argomenti. Inoltre, per ogni derivazione matematica, per maggiore leggibilità, ho cercato di riportare tutti i passaggi.

Pregherei i lettori di essere pazienti se nel testo sono ancora presenti diverse distrazioni o errori. Negli ultimi anni ho reso disponibili queste note agli studenti dei nostri corsi in varie versioni aggiornandole continuamente e cercando di correggere tempestivamente alcuni errori segnalati da attenti lettori. Sarò molto grato a chiunque volesse segnalarmene di ulteriori a

francesco.palmieri@unina2.it

per la compilazione di un eventuale errata-corrige.

Ringrazio la CUES per la fiducia accordatami nel pubblicare queste dispense in forma rilegata. Spero che esse possano risultare utili anche agli studenti di corsi affini e di altre Università.

Desidero inoltre ringraziare tutti gli studenti che hanno pazientemente ascoltato le mie lezioni e i colleghi che hanno riletto il manoscritto e che mi hanno fornito preziosi suggerimenti.

I ringraziamenti più importanti vanno alla mia famiglia, alla quale ho sottratto molto del tempo che è stato dedicato alla preparazione di queste note.

Francesco Palmieri

Aversa, 1 gennaio 2010

Introduzione

La probabilità è la scienza che tratta di eventi *casuali*, ovvero di tutto ciò la cui occorrenza appare non prevedibile. Chiunque legga queste note, anche senza basi scientifiche, ha la nozione intuitiva di come alcuni eventi siano più o meno *probabili*. Per esempio nel lancio di una moneta si ha una certa confidenza nel fatto che, se la moneta non è truccata, ci sarà il 50% di probabilità che il risultato sia testa e altrettanto che sia croce. Oppure, nel tentare la fortuna alle corse dei cavalli il giocatore cerca di puntare sul cavallo che ritiene gli porterà un guadagno. La sua decisione deriva da una valutazione, più o meno sistematica, delle probabilità di vittoria dei vari cavalli, cosa che fa anche l'allibratore fissando le proprie quotazioni.

Queste nozioni intuitive hanno stimolato nel diciassettesimo secolo i primi tentativi di scrittura di una teoria matematica della probabilità che studiasse in maniera sistematica tutti i fenomeni che appaiono imprevedibili. Oggi, la teoria della probabilità costituisce uno strumento di riferimento indispensabile in moltissimi campi della scienza moderna quali lo studio di fenomeni fisici, l'analisi, la predizione e la sintesi di sistemi economici, lo studio dei canali e dei sistemi di telecomunicazione, e molti altri.

In particolare, per il settore dell'Informazione e per gli studenti di Ingegneria Elettronica, Ingegneria Informatica, Ingegneria delle Telecomunicazioni, Ingegneria dell' Informazione, Informatica, Scienze dell' Informazione, è di fondamentale importanza sapere affrontare dei problemi pratici utilizzando la modellistica probabilistica. Nella gestione e nel progetto di sistemi che trattano

l'informazione, i modelli probabilistici di sorgenti e canali costituiscono oggi la metodologia operativa più potente per quantificare i flussi informativi e la loro integrità conseguente la trasmissione.

La nozione stessa di informazione è collegata al concetto probabilistico di *imprevedibilità*: gli eleganti risultati della Teoria dell'Informazione, principalmente associati al lavoro di C. E. Shannon degli anni quaranta, ci forniscono la struttura teorica di riferimento per la sintesi e l'analisi di codificatori e sistemi di trasmissione in presenza di distorsioni e disturbi.

Nei moderni sistemi di telecomunicazioni, la maggior parte dei problemi di analisi e di sintesi sono studiati mediante modelli probabilistici. Ad esempio, i canali di comunicazione telefonica su fili e su mezzo elettromagnetico, quali ad esempio quelli cellulari, sono progettati sulla base di misure e modelli statistici che tengono in conto i comportamenti aleatori di disturbi e distorsioni introdotti dalle trasmissioni. Anche per le moderne comunicazioni multimediali gli strumenti analitici e progettuali di supporto sono basati su modelli quantitativi di natura probabilistica.

Queste note sono organizzate in più capitoli e delle appendici. Il primo capitolo è dedicato ai fondamenti della teoria della probabilità, mentre nel secondo viene introdotta la variabile aleatoria monodimensionale. Nel terzo capitolo si estende la teoria della variabile aleatoria al caso bidimensionale e nel quarto a quello n -dimensionale. Infine, il quinto capitolo è dedicato al calcolo della densità di variabili aleatorie che sono il risultato di trasformazioni deterministiche.

Nota bibliografica

La teoria della probabilità è ricchissima di riferimenti bibliografici. Uno dei testi più classici è quello di Gnedenko (1992), tradotto in italiano dal russo. Molto usati nella facoltà scientifiche sono anche i testi di Parzen (1960) e Feller (1968). Il testo di Papoulis (1965), nella traduzione italiana (1973) e nelle edizioni successive (1984,

1991), è divenuto il principale testo di riferimento nelle Facoltà di ingegneria. Da segnalare anche (Helstrom, 1984) e (Peebles, 1980) per la loro buona organizzazione didattica. Esercizi sulla probabilità sono disseminati nei testi a cui si è fatto riferimento, ma va segnalato in particolare l'ottimo manuale della Schaum (H. Hsu, 1996), da cui alcuni degli esercizi qui riportati sono stati ispirati.

Capitolo 1

Introduzione alla Probabilità

In questo capitolo vengono introdotti i principali elementi della teoria della probabilità. Dopo la definizione di cosa si intende per esperimento, spazio campione ed eventi ad esso collegati, viene introdotto il concetto di probabilità. La definizione segue l'approccio assiomatico da cui le proprietà fondamentali vengono derivate. Vengono inoltre presentate la definizione di probabilità condizionata, la legge della probabilità totale, il teorema di Bayes e la legge dei grandi numeri. Particolare attenzione è dedicata agli spazi campione discreti di dimensione finita.

1.1 Il modello probabilistico

La teoria della probabilità è stata introdotta per fornire un modello matematico per quei fenomeni che esibiscono un comportamento apparentemente imprevedibile o *aleatorio*. Esistono infatti molte situazioni fisiche in cui uno stesso esperimento ripetuto più volte può generare risultati diversi. Si pensi al lancio di una moneta o di un dado. Nel modello probabilistico si suppone che l'esperimento, nonostante venga ripetuto sotto le stesse condizioni, possa produrre un risultato fortuito e non prevedibile. In effetti, in un esper-

imento come quello del dado, la aleatorietà del risultato va attribuita alle condizioni sempre leggermente diverse sotto le quali esso viene eseguito. L'approccio probabilistico viene introdotto per ignorare tutti i dettagli fisici che contribuiscono al risultato dell'esperimento - che sarebbe comunque molto difficile tenere in conto. Si pensi ai complessi meccanismi fisici che contribuiscono al risultato del lancio del dado: elasticità, rugosità e geometria delle superfici, energia impressa al lancio, peso e distribuzione della massa nel dado, ecc. Quindi anche se la realtà di un esperimento segue delle leggi fisiche ben determinate (dette *deterministiche*, o *determinate*), l'esperimento mostra all'osservatore un risultato che appare imprevedibile, o *aleatorio* (*random*).

La figura 1.1 mostra la dicotomia tra l'esperimento vero e proprio, appartenente alla realtà fisica, e il modello matematico di riferimento.

Nel creare il modello matematico di un esperimento, bisogna per primo identificare l'insieme dei possibili risultati, assegnare loro una notazione e organizzarli in un insieme che chiameremo *Spazio Campione* \mathcal{S} , la cui dimensione $n_{\mathcal{S}}$ può essere finita o infinita. La corrispondenza tra l'esperimento e lo spazio campione è la parte più critica nello studio del fenomeno in quanto è necessario identificare ciò che è importante evidenziare nell'esperimento in una maniera che sia chiaramente distinguibile e manipolabile nel modello matematico. Negli esempi che seguono sarà evidente come la struttura di tale spazio possa essere diversa a seconda del tipo di esperimento.

Esempio 1.1 Una moneta è lanciata una volta.

I possibili risultati dell'esperimento sono $H = Testa$ e $C = Croce$. Lo spazio campione è: $\mathcal{S} = \{T, C\}$, $n_{\mathcal{S}} = 2$.

Esempio 1.2 Un dado a 6 facce numerate da 1 a 6 è lanciato una volta.

Lo spazio campione è: $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, $n_{\mathcal{S}} = 6$.

Esempio 1.3 Una moneta è lanciata due volte. Gli $n_{\mathcal{S}} = 4$ possibili risultati dell'esperimento, sono gli elementi dello spazio campione $\mathcal{S} = \{(T, T), (T, C), (C, T), (C, C)\}$.

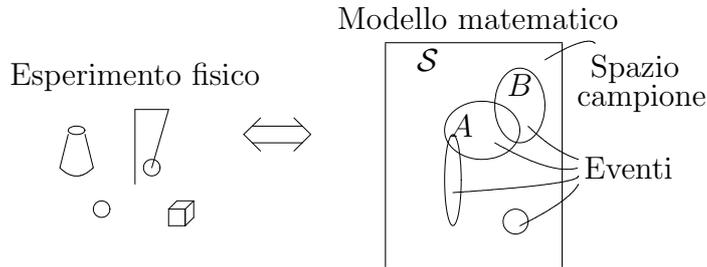


Figura 1.1: L'esperimento fisico e il modello probabilistico.

Esempio 1.4 Una dado a sei facce è lanciato due volte. Gli $n_S = 36$ possibili risultati dell'esperimento, sono tutte le possibili coppie di numeri da 1 a 6: $\mathcal{S} = \{(i, j), i, j = 1, \dots, 6\}$.

Esempio 1.5 Il bit "0" dopo essere stato trasmesso su un canale che introduce errori aleatori, può essere ricevuto correttamente come "0", o erroneamente come "1". Lo spazio campione di questo esperimento binario è: $\mathcal{S} = \{1, 0\}$, $n_S = 2$.

Esempio 1.6 Si giocano tredici partite di calcio. Il risultato di ognuna può essere: X=pareggio, 1=vincita della squadra di casa, 2=vincita della squadra ospite. Lo spazio campione \mathcal{S} sarà costituito da tutte le $n_S = 3^{13} = 1594323$ possibili combinazioni dei tre simboli su 13 posti (l'enumerazione esplicita sarebbe troppo lunga e priva di reale utilità).

Esempio 1.7 Si misuri la temperatura ambiente con un termometro la cui scala va da -20 a 60 gradi centigradi. Lo spazio campione è il sottoinsieme dell'asse reale: $\mathcal{S} = [-20, 60] \subset \mathcal{R}$, $\dim(\mathcal{S}) = \text{lunghezza dell'intervallo} = 80$.

Esempio 1.8 Da un'urna contenente palle numerate da 1 a 90, si estrae senza rimpiazzo una sequenza di sei palle. Lo spazio campione \mathcal{S} è costituito da tutte le $n_S = 90 \cdot 89 \cdot 87 \cdot 86 \cdot 86 \cdot 85 = 4.4828 \cdot 10^{11}$ possibili disposizioni (senza ripetizioni) di 6 dei 90 numeri. Se invece l'ordine è irrilevante (come nel popolare SuperEnalotto) la dimensione dello spazio campione si riduce al numero di combinazioni (senza ripetizioni) di 6 dei

90 numeri ed è ottenuta dividendo il numero di sopra per 6!. Ovvero, $n_S = \binom{90}{6} = 622614630$.

Esempio 1.9 Da un mazzo di 40 carte napoletane tre giocatori estraggono una carta ciascuno. Lo spazio campione è costituito da tutte le $n_S = 40 \cdot 39 \cdot 38 = 59280$ possibili terne (senza ripetizioni).

Esempio 1.10 Una sequenza di N bit è emessa da un computer. Lo spazio campione \mathcal{S} è costituito dalle $n_S = 2^N$ possibili sequenze binarie. Per esempio, se $N = 3$, $\mathcal{S} = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$, $n_S = 8$.

Esempio 1.11 Una sequenza di N bit è emessa da un computer ed è trasmessa su un canale con errori. L'esperimento consiste nel contare il numero di errori. Lo spazio campione è quindi: $\mathcal{S} = \{0, 1, 2, \dots, N\}$, $n_S = N + 1$.

Esempio 1.12 Le 40 carte di un mazzo di carte napoletane vengono distribuite a quattro giocatori (come nello Scopone Scientifico). Lo spazio campione è costituito da tutte le possibili mani. L'ordine della sequenza è importante per determinare la distribuzione delle carte tra i vari giocatori, ma non lo è ovviamente per ogni giocatore. Per contare il numero di possibili configurazioni, osserviamo che se si dovesse tenere conto dell'ordine completo avremmo $40!$ possibili permutazioni. Questo numero va diviso per $(10!)^4$ poiché l'ordine delle carte per ogni giocatore è irrilevante. Avremo pertanto: $n_S = \frac{40!}{(10!)^4} = 4.7054 \cdot 10^{21}$. Questo numero è il coefficiente multinomiale $\binom{40}{10 \ 10 \ 10 \ 10}$ (vedi un qualunque testo di analisi matematica per maggiori dettagli sul calcolo combinatorio).

Esempio 1.13 Un arciere scaglia una freccia contro un bersaglio a simmetria circolare di raggio 1 metro. La posizione a cui si conficca la freccia è il risultato dell'esperimento. Lo spazio campione può essere il cerchio di raggio 1 metro descritto in coordinate polari: $\mathcal{S} = \{(\rho, \theta), 0 \leq \rho < 1, -\pi < \theta \leq \pi\}$, $\dim(\mathcal{S}) = \text{area del cerchio} = \pi$. Si noti come questo modello non preveda che la freccia possa cadere al di fuori del bersaglio. Tale eventualità non è inclusa nello spazio campione e costituisce un evento impossibile. Una definizione più generale dello spazio campione per questo esperimento è abbastanza semplice e la si lascia al lettore per esercizio.

Esempio 1.14 Si valuti altezza di un filo d'erba scelto a caso. Teoricamente tale esperimento ha uno spazio campione pari a tutti i valori reali positivi: $\mathcal{S} = \mathcal{R}^+$, $\dim(\mathcal{S}) = \infty$. Poiché fili d'erba di dimensione superiore a qualche decimetro sono praticamente introvabili, vedremo nel seguito come sarà la attribuzione di una misura di probabilità allo spazio a tener conto di ciò.

Esempio 1.15 Si valuti l'età espressa in anni di una persona scelta a caso. Teoricamente tale esperimento ha uno spazio campione che corrisponde a tutti i numeri naturali (zero incluso): $\mathcal{S} = \mathcal{N}$, $n_{\mathcal{S}} = \infty$. Poiché persone di età superiore ai 110 anni sono (almeno per ora) introvabili, vedremo nel seguito come sarà la attribuzione di una misura di probabilità agli elementi dello spazio a tener conto di ciò.

E' utile sottolineare come la denominazione degli elementi dello spazio campione sia del tutto arbitraria. La cosa importante nel costruire il modello matematico è associare correttamente i risultati dell'esperimento a degli elementi distinguibili dello spazio campione. L'etichetta da associare ad essi è generalmente dettata da esigenze mnemoniche.

Lo studio di un esperimento e del modello probabilistico ad esso associato può essere piuttosto diverso a seconda della natura dello spazio campione. E' pertanto utile introdurre la seguente classificazione:

- *Spazio campione discreto e finito*: insieme numerabile di cardinalità $n_{\mathcal{S}}$ finita. Costituiscono esempi tipici: un qualunque insieme finito di numeri interi; l'insieme delle lettere dell'alfabeto; un insieme finito di oggetti; ecc.
- *Spazio campione discreto e infinito*: insieme numerabile di cardinalità infinita. Esempi tipici sono l'insieme dei numeri interi, o l'insieme dei numeri naturali.
- *Spazio campione continuo finito*: insieme associabile a sottoinsiemi di \mathcal{R}^d a cui è associabile un misura finita $\dim(\mathcal{S})$. Si pensi per esempio a intervalli dell'asse reale ($d = 1$), o a regioni limitate del piano ($d = 2$).

- *Spazio campione continuo e infinito*: insieme associabile a sottoinsiemi infiniti di \mathcal{R}^d . Esempi tipici sono l'insieme di tutti i numeri reali ($\mathcal{S} = \mathcal{R}$), o l'insieme dei numeri reali positivi ($\mathcal{S} = \mathcal{R}^+$), o l'insieme dei punti del piano ($\mathcal{S} = \mathcal{R}^2$), o l'insieme dei punti del primo quadrante del piano ($\mathcal{S} = \mathcal{R}^+ \times \mathcal{R}^+$).
- *Spazio campione misto*: insieme ibrido costituito da sottoinsiemi continui e discreti di dimensionalità finita o infinita.

Dei vari tipi di spazi campione, ci concentreremo in questo capitolo principalmente sugli spazi campione discreti e omogenei, rimandando una più approfondita discussione sugli spazi campione continui e misti ai capitoli sulle variabili aleatorie.

1.2 Eventi

Gli elementi dello spazio campione possono essere raggruppati e studiati secondo la teoria classica degli insiemi. Ogni sottoinsieme dello spazio campione è detto *Evento*. Quindi tutti i possibili sottoinsiemi di \mathcal{S} sono i possibili eventi associabili al modello di un esperimento. Denotando con $\xi \in \mathcal{S}$ l'elemento dello spazio campione associato al risultato di un esperimento, diremo che un evento $A \subseteq \mathcal{S}$ si è verificato, se $\xi \in A$. L'evento $A = \mathcal{S}$ che coincide con lo spazio campione è detto *evento certo*. Analogamente l'insieme vuoto $A = \emptyset$ è detto *evento impossibile*. Ad ogni evento è possibile associare una cardinalità, se si tratta di un evento discreto, o una dimensione se si tratta di un insieme continuo.

Con riferimento agli esperimenti descritti nella sezione precedente, presentiamo qui alcuni esempi di eventi denotandoli con le lettere maiuscole A, B, C, \dots . Si noti come negli esempi sia spesso utile descrivere il risultato dell'esperimento a parole prima di procedere alla sua formulazione matematica.

Esempio 1.1 (cont.):

$$A = \{\text{il risultato del lancio è Testa}\} = \{T\}$$

$$B = \{\text{il risultato del lancio è Testa o Croce}\} = \{T\} \cup \{C\} = \{T, C\} = \mathcal{S}$$

$$C = \{\text{la moneta si piazza di taglio}\} = \{\} = \emptyset$$

(questa eventualità non è prevista nello spazio campione, pertanto l'evento C corrisponde all'insieme vuoto, ovvero all'evento impossibile).

Esempio 1.2 (cont.):

$$A = \{\text{il risultato del lancio è 4}\} = \{4\}$$

$$B = \{\text{il risultato del lancio è un numero maggiore o uguale a 3}\} = \{3, 4, 5, 6\}$$

$$C = \{\text{il risultato del lancio è un numero minore o uguale a 6}\} = \{1, 2, 3, 4, 5, 6\} = \mathcal{S}$$

(tutti i risultati dell'esperimento appartengono all'evento C che quindi coincide con lo spazio campione, ovvero con l'evento certo)

$$D = \{\text{il risultato del lancio è un numero dispari}\} = \{1, 3, 5\}$$

In questo esperimento $A \subset B$. Quindi il verificarsi dell'evento A implica il verificarsi dell'evento B . Non vale ovviamente il viceversa.

Esempio 1.3 (cont.):

$$A = \{\text{I due lanci danno la sequenza Testa-Croce}\} = \{(T, C)\}$$

$$B = \{\text{I due lanci totalizzano una Testa e una Croce}\} = \{(T, C), (C, T)\}$$

$$C = \{\text{Il risultato del primo lancio è Testa}\} = \{(T, C), (T, T)\}$$

$$D = \{\text{I due lanci producono la stessa faccia}\} = \{(T, T), (C, C)\}$$

Esempio 1.4 (cont.):

$$A = \{\text{I due lanci danno la sequenza 6-2}\} = \{(6, 2)\}$$

$$B = \{\text{I due lanci totalizzano una somma pari a 8}\} = \{(2, 6), (6, 2), (4, 4), (3, 5), (5, 3)\}$$

$$C = \{\text{I due lanci totalizzano una somma minore o uguale a 4}\} = \{(1, 3), (3, 1), (2, 2), (1, 2), (2, 1), (1, 1)\}$$

$$D = \{\text{I due lanci danno lo stesso numero}\} = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$$

$$E = \{\text{I due lanci danno almeno un 6}\} = \{(1, 6), (2, 6), (3, 6), (4, 6), (5, 6), (6, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5)\}$$

Esempio 1.5 (cont.):

$$A = \{\text{il bit è ricevuto correttamente}\} = \{0\}$$

$$B = \{\text{il bit è ricevuto erroneamente}\} = \{1\}$$

$$C = \{\text{il bit non è ricevuto (cancellazione)}\} = \emptyset$$

(tale eventualità non è prevista nello spazio campione, pertanto l'evento coincide con l'insieme vuoto, ovvero con l'evento impossibile)

Esempio 1.6 (cont.):

$$A = \{\text{tutte le partite terminano in parità}\} = \underbrace{\{(X, X, \dots, X)\}}_{13}$$

$$B = \{\text{tutte le partite terminano in parità eccetto la seconda}\} \\ = \underbrace{\{(X, 1, X, \dots, X)\}}_{13}, \underbrace{\{(X, 2, X, \dots, X)\}}_{13}$$

$$C = \{\text{tutte le partite, eccetto una, terminano in parità}\} \\ = \underbrace{\{(1, X, \dots, X)\}}_{13}, \underbrace{\{(2, X, \dots, X)\}}_{13}, \underbrace{\{(X, 1, \dots, X)\}}_{13}, \underbrace{\{(X, 2, \dots, X)\}}_{13}, \dots, \underbrace{\{(X, \dots, X, 2)\}}_{13}$$

26

Esempio 1.7 (cont.):

$$A = \{\text{la temperatura è superiore ai 30 gradi}\} = [30, 60]$$

$$B = \{\text{la temperatura è di 20 gradi con un margine di errore di un grado}\} = [19, 21]$$

$$C = \{\text{la temperatura è esattamente 20 gradi}\} = \{20\} \text{ (questo è un evento un pò particolare poiché si tratta di un valore discreto su un insieme continuo; vedremo in seguito come trattare un tale evento visto che la dimensione di } C \text{ è zero)}$$

$$D = \{\text{la temperatura è di 80 gradi con un margine di errore di un centesimo di grado}\} = \emptyset \text{ (evento impossibile).}$$

Esempio 1.8 (cont.): Si supponga che l'ordine dei numeri sia irrilevante.

$$A = \{\text{si verifica la sestina } 2, 23, 28, 50, 78, 90\} = \{(2, 23, 28, 50, 78, 90)\}$$

$$B = \{\text{esce la quaterna } 15, 47, 48, 86\} \\ = \underbrace{\{\text{tutte le sestine che contengono i numeri } 15, 47, 48, 86\}}_{86 \cdot 85 / 2 = 3655}$$

(basta mettere da parte i quattro numeri della quaterna e contare le combinazioni risultanti delle rimanenti carte su due posti)

Esempio 1.9 (cont.):

$$A = \{\text{i tre giocatori hanno rispettivamente un asso di bastoni (1B), un sette di spade (7S) e un sei di coppe (6C)}\} = \{(1B, 7S, 6C)\}$$

$$B = \{\text{i tre giocatori hanno tutti un asso e nessuno l'asso di ori (1O)}\} \\ = \{(1S, 1C, 1B), (1S, 1B, 1C), (1C, 1S, 1B), (1C, 1B, 1S), (1B, 1S, 1C), (1B, 1C, 1S)\}$$

Esempio 1.10 (cont.): (si assuma $N = 3$)

$$A = \{\text{la sequenza rappresenta il numero 5 in codice binario}\} = \{(1, 0, 1)\}$$

$$B = \{\text{la sequenza è costituita da tutti bit uguali}\} = \{(0, 0, 0), (1, 1, 1)\}$$

$$C = \{\text{la sequenza contiene un solo bit pari a 0}\} = \{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}$$

$$D = \{\text{la sequenza inizia con il prefisso 10}\} = \{(1, 0, 0), (1, 0, 1)\}$$

$$E = \{\text{la sequenza contiene un numero dispari di 1}\} \\ = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}$$

$$F = \{\text{la sequenza contiene almeno due 1}\} \\ = \{(1, 1, 0), (0, 1, 1), (1, 0, 1), (1, 1, 1)\}$$

Esempio 1.11 (cont.):

$$A = \{\text{la sequenza ricevuta non contiene errori}\} = \{0\}$$

$$B = \{\text{la sequenza ricevuta è tutta errata}\} = \{N\}$$

$$C = \{\text{la sequenza contiene al più due errori}\} = \{0, 1, 2\}$$

Esempio 1.12 (cont.):

$$A = \{\text{il giocatore 1 ha tutti i sette}\}$$

$$B = \{\text{i quattro giocatori ricevono carte di un solo colore}\}$$

$$C = \{\text{i giocatori ricevono un sette per ciascuno}\}$$

$$D = \{\text{il giocatore 3 riceve una scala completa}\}$$

(si tralascia per brevità la enumerazione completa degli elementi di ogni evento; il lettore studi per esercizio la struttura di ognuno degli eventi e ne valuti la relativa cardinalità usando i risultati dell'analisi combinatoria)

Esempio 1.13 (cont.):

$$A = \{\text{l'arciere colpisce il "bull's eye," definito come il cerchio più interno di raggio 1 centimetro}\} = \{(\rho, \theta), 0 \leq \rho < 0.01, -\pi < \theta \leq \pi\}$$

$$B = \{\text{l'arciere manca il bersaglio}\} = \emptyset$$

$$C = \{\text{la freccia si conficca ad una distanza dal centro superiore a 20 centimetri}\} = \{(\rho, \theta), 0.20 \leq \rho < 1, -\pi < \theta \leq \pi\}$$

Esempio 1.14 (cont.):

$$A = \{\text{il filo d'erba è più basso di un centimetro}\} = [0, 0.01]$$

$$B = \{\text{il filo d'erba è più alto di un metro}\} = [1, \infty[$$

$$B = \{\text{il filo d'erba misura 10 centimetri con una tolleranza di 1 centimetro}\} = [0.09, 0.11]$$

Esempio 1.15 (cont.):

$$A = \{\text{l'età osservata è di 90 anni}\} = \{90\}$$

$$B = \{\text{l'età osservata è maggiore o uguale di 10 anni}\} = \{10, 11, 12, \dots\}$$

$$B = \{\text{l'età osservata è tra i 30 e i 35 anni (estremi esclusi)}\} = \{31, 32, 33, 34\}$$

1.3 Il concetto di probabilità

L'idea di probabilità consiste nell'associare ad ogni evento dello spazio campionario una *misura*, o un *peso*, che descriva quanto tale evento sia frequente nelle osservazioni di quell'esperimento.

Denotando con m il numero di volte in cui si effettua un esperimento e con m_A il numero di volte in cui si verifica l'evento A , si definisce *frequenza relativa* di A il rapporto

$$f_m(A) = \frac{m_A}{m}. \quad (1.1)$$

Nonostante, la frequenza relativa rifletta intuitivamente alcune delle caratteristiche che vorremmo attribuire alla definizione di probabilità, se adottassimo $f_m(A)$ come la nostra definizione, avremmo

immediatamente un problema: ogni volta che si effettua una sequenza di m prove dell'esperimento si ottiene un valore diverso, anche se magari solo leggermente diverso. Questo forse vuol dire che la attribuzione di una misura di probabilità non è univoca?

Per vederci più chiaro, consideriamo esplicitamente i risultati del semplice esperimento del lancio di una moneta (Tabella 1.1).

	m_T	m_C	$f_m(T) = \frac{m_T}{m}$	$f_m(C) = \frac{m_C}{m}$
Primi 100 lanci	47	53	0.47	0.53
Secondi 100 lanci	51	49	0.51	0.49
Terzi 100 lanci	53	47	0.53	0.47
300 lanci combinati	151	149	0.503	0.497

Tabella 1.1: Risultati del lancio di una moneta

Nell'esperimento sono state effettuate tre sessioni di 100 lanci e sono state calcolate le frequenze relative. Si noti come i valori siano leggermente diversi nelle tre sessioni e come i risultati combinati mostrino la tendenza a dare $f_m(T) = f_m(C) = 0.5$. I risultati dell'esperimento sembrano suggerire che la moneta non è truccata e che se effettuassimo un numero sempre più grande di prove le frequenze relative di entrambe Testa e Croce tenderebbero a $\frac{1}{2}$.

Prima di proseguire, consideriamo anche un altro tipico esperimento di lancio di un dado. Tabella 1.2 mostra i risultati di tre sessioni di lanci di un dado a sei facce. Anche qui in ogni sessione le occorrenze non sono esattamente distribuite sui sei possibili risultati, anche se è evidente la tendenza che le frequenze relative mostrano verso il valore $\frac{1}{6} \simeq 0.17$.

L'elemento che cattura la nostra attenzione nei risultati di questi esperimenti, è proprio il valore asintotico delle frequenze relative: esso sembra essere un invariante rispetto alla imprevedibilità dei risultati. E' infatti proprio a tale valore asintotico che attribuiremo il significato di probabilità.

La quantificazione probabilistica degli eventi deve tendere a modellare la realtà fisica sottostante in una maniera tale da estrarre dalla

	primi 100 lanci	secondi 100 lanci	terzi 100 lanci	300 lanci combi- nati
m_1	15	14	20	49
m_2	18	17	18	53
m_3	23	18	13	54
m_4	13	20	14	47
m_5	15	17	19	51
m_6	16	14	16	46
$f_m(1)$	0.15	0.14	0.20	0.16
$f_m(2)$	0.18	0.17	0.18	0.18
$f_m(3)$	0.23	0.18	0.13	0.18
$f_m(4)$	0.13	0.20	0.14	0.16
$f_m(5)$	0.15	0.17	0.19	0.17
$f_m(6)$	0.16	0.14	0.16	0.15

Tabella 1.2: Risultati dei lanci di un dado

variabilità delle osservazioni ciò che è invariante, e che assume pertanto valore di legge fisica.

Negli esperimenti del dado e della moneta la attribuzione di una misura di probabilità ai possibili risultati degli esperimenti, deve riflettere il fatto che le facce del dado e della moneta hanno uguale tendenza a mostrarsi come risultato, ovvero che il dado e la moneta non sono truccati.

Si può pertanto definire probabilità di un evento A , il limite della frequenza relativa delle sue occorrenze al crescere del numero di prove

$$P(A) \stackrel{\text{def}}{=} \lim_{m \rightarrow \infty} \frac{m_A}{m} = \lim_{m \rightarrow \infty} f_m(A). \quad (1.2)$$

La questione della convergenza di $\{f_m(A)\}$ a $P(A)$ è in realtà piuttosto insidiosa, almeno per chi si è appena avvicinato alla teoria della probabilità. La frequenza relativa $f_m(A)$, che per m sufficientemente grande, può essere considerata una *stima* di $P(A)$, può convergere a $P(A)$ anche in maniera piuttosto bizzarra, ovvero in maniera non necessariamente monotona. A tale proposito, si osservi come nell'esperimento della moneta, la sequenza di frequenze relative ottenute con $m = 100, 200, 300$, partendo dalle seconde 100 prove, poi raggruppando le prime 100 con le seconde 100 e poi combinandole tutte, sia $\{f_{100}(C), f_{200}(c), f_{300}(C)\} = \{0.51, 0.53, 0.50\}$ (il raggruppamento utilizzato è del tutto arbitrario).

La definizione di probabilità qui discussa, che segue il cosiddetto *approccio frequentista*, nonostante rifletta la nozione più intuitiva di probabilità, riserva notevoli insidie dal punto di vista formale. Per maggiore agilità didattica si preferisce pertanto intraprendere la strada alternativa dell'approccio assiomatico alla definizione di probabilità, rimandando le questioni relative alla armonizzazione dei due punti di vista a successivi approfondimenti.

1.4 Definizione assiomatica di probabilità

Le esigenze che motivano la formalizzazione della teoria della probabilità sulla base degli assiomi sono piuttosto intuitive: assegnare una misura di probabilità ad ogni sottoinsieme (evento) di \mathcal{S} rispettando dei semplici vincoli costitutivi. La probabilità deve essere positiva e limitata e tale da soddisfare una proprietà di additività che consenta di attribuire in maniera consistente una misura a tutti gli eventi dello spazio campione. Inoltre, le proprietà derivanti dovranno risultare coerenti con l'approccio frequentista descritto nella sezione precedente.

L'approccio assiomatico è basato su soli tre vincoli costitutivi: da tre assiomi è possibile far discendere tutta la formalizzazione matematica della teoria della probabilità ¹.

Dato uno spazio campione \mathcal{S} e degli eventi A e B ad esso associati, gli assiomi della probabilità sono:

ASSIOMA 1: $P(A) \geq 0 \forall A \subseteq \mathcal{S}$ (la probabilità di ogni evento è un valore positivo).

ASSIOMA 2: $P(\mathcal{S}) = 1$ (la probabilità dello spazio campione, o dell'evento certo, è pari a uno).

ASSIOMA 3: $P(A \cup B) = P(A) + P(B), \forall A, B : A \cap B = \emptyset$ (la probabilità dell'unione di due eventi disgiunti è pari alla somma delle loro probabilità).

Dimostreremo ora alcune importanti proprietà della probabilità in maniera costruttiva, partendo cioè dai soli tre assiomi appena enunciati.

¹La sufficienza dei tre assiomi non è più garantita quando si trattano spazi campione discreti di dimensione infinita e spazi continui. In tali casi, come vedremo in seguito, sarà necessario imporre qualche ulteriore vincolo.

Proprietà 1.1

$$P(\emptyset) = 0 \tag{1.3}$$

(la probabilità dell'evento impossibile è zero).

Prova: Poiché $\mathcal{S} \cup \emptyset = \mathcal{S}$ e $\mathcal{S} \cap \emptyset = \emptyset$, usando gli Assiomi 2 e 3, risulta $P(\mathcal{S}) = P(\mathcal{S}) + P(\emptyset) = 1$, da cui discende l'asserto.

Proprietà 1.2 Dato un evento $A \subseteq \mathcal{S}$,

$$P(\overline{A}) = 1 - P(A) \tag{1.4}$$

(la probabilità dell'evento complemento è pari a uno meno la probabilità dell'evento).

Prova: Poiché $\mathcal{S} = A \cup \overline{A}$ e $A \cap \overline{A} = \emptyset$, usando gli Assiomi 2 e 3, risulta $P(\mathcal{S}) = P(A) + P(\overline{A}) = 1$, da cui discende l'asserto.

Proprietà 1.3

$$0 \leq P(A) \leq 1 \tag{1.5}$$

(la probabilità di un evento è un numero positivo compreso tra zero e uno).

Prova: La prima disuguaglianza coincide con il primo assioma. La seconda è immediata conseguenza della Proprietà 1.2 e del primo assioma.

Proprietà 1.4 Per ogni coppia di eventi A e $B \subseteq \mathcal{S}$

$$P(A \cap \overline{B}) = P(A) - P(A \cap B). \tag{1.6}$$

Prova: La dimostrazione è immediata se si osserva la Figura 1.2. Poiché $A = (A \cap \overline{B}) \cup (A \cap B)$, con $(A \cap \overline{B})$ e $(A \cap B)$ eventi disgiunti, dal terzo assioma abbiamo che $P(A) = P(A \cap \overline{B}) + P(A \cap B)$, da cui discende l'asserto.

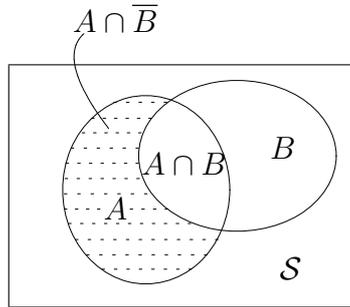


Figura 1.2: Il diagramma relativo alla proprietà 1.4.

Proprietà 1.5 Per ogni coppia di eventi A e $B \subseteq \mathcal{S}$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (1.7)$$

(la probabilità dell'unione di due eventi è uguale alla somma delle loro probabilità meno la probabilità della loro intersezione).

Prova: La dimostrazione anche qui è immediata se dal diagramma di Figura 1.2 si osserva come la unione di A e B può essere espressa come la unione di due eventi disgiunti, ovvero $A \cup B = B + (A \cap \bar{B})$. Dal terzo assioma risulta che $P(A \cup B) = P(B) + P(A \cap \bar{B})$. Utilizzando Proprietà 1.4 si ottiene il risultato cercato. Si noti che questa proprietà può essere vista come la generalizzazione del terzo assioma al caso di eventi non disgiunti.

Proprietà 1.6 Se l'evento $A \subseteq B$

$$P(A) \leq P(B) \quad (1.8)$$

(se un primo evento è incluso in un secondo, la probabilità del primo è minore o uguale a quella del secondo).

Prova: Si osservi il diagramma in Figura 1.3. L'evento B può essere espresso come somma di due eventi disgiunti, ovvero $B = A + (\bar{A} \cap B)$. Pertanto, usando rispettivamente il terzo e il primo assioma, abbiamo che $P(B) = P(A) + P(\bar{A} \cap B) \geq P(A)$, che è il risultato cercato.

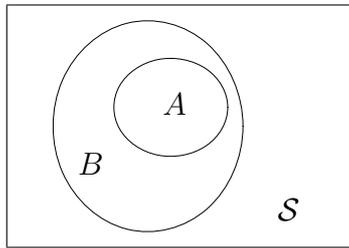


Figura 1.3: Il diagramma relativo alla proprietà 1.6.

Proprietà 1.7 *Dato un insieme A_1, A_2, \dots, A_n di eventi disgiunti*

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \quad (1.9)$$

(la probabilità dell'unione di eventi disgiunti è uguale alla somma delle probabilità dei singoli eventi).

Prova: Il risultato si può provare per induzione utilizzando il terzo assioma. Si consideri la sequenza di eventi $S_1 = A_1$, $S_2 = A_1 \cup A_2, \dots, S_i = \bigcup_{j=1}^i A_j, \dots$. Poiché $S_i = S_{i-1} \cup A_i$. Si ha che $P(S_1) = P(A_1)$, $P(S_2) = P(S_1) + P(A_2), \dots, P(S_i) = P(S_{i-1}) + P(A_i)$. Pertanto $P(S_n) = P(S_{n-1}) + P(A_n) = P(S_{n-2}) + P(A_{n-1}) + P(A_n) = \dots = \sum_{i=1}^n P(A_i)$.

Gli assiomi enunciati, e le proprietà da essi derivate, sono sufficienti a costruire tutta la teoria della probabilità per esperimenti aventi uno spazio campione finito. Per includere nella teoria anche gli spazi campione a dimensione infinita, è necessario modificare il terzo assioma ai fini di estendere il risultato di Proprietà 1.7 a una sequenza di infiniti eventi disgiunti. E' infatti impossibile dimostrare rigorosamente tale proprietà a partire dal solo terzo assioma. Pertanto, nei testi classici di teoria della probabilità, in qualche caso si aggiunge il seguente ulteriore assioma ai tre già enunciati, o semplicemente lo si sostituisce al terzo:

ASSIOMA 3': Per qualunque sequenza infinita di eventi mutuamente esclusivi $A_1, A_2, \dots, A_n, \dots$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (1.10)$$

(la probabilità dell'unione di infiniti eventi disgiunti è pari alla somma delle probabilità degli eventi).

In realtà, qualche altro problema formale emerge ancora quando si considerano spazi campione non numerabili e di dimensione infinita. In tali casi potrebbero esistere dei sottoinsiemi dello spazio campione a cui non è possibile assegnare una probabilità in maniera consistente con gli assiomi. In tali casi si può risolvere il problema semplicemente escludendo tali insiemi dai possibili eventi. In altre parole si considerano eventi solo i sottoinsiemi dello spazio campione a cui è possibile assegnare una probabilità in maniera consistente con gli assiomi. Il significato di queste considerazioni sarà più chiaro quando parleremo di variabili aleatorie continue in successive lezioni.

NOTAZIONI: Per economia di notazione, utilizzeremo nel seguito indistintamente le notazioni:

$$\begin{array}{ll} A \cup B & \text{oppure} \quad A + B \\ A \cap B & \text{oppure} \quad AB \end{array}$$

per denotare rispettivamente unione e intersezione.

1.5 Spazi campione discreti e finiti

La formalizzazione matematica degli assiomi della probabilità e delle proprietà derivate, diventa immediatamente comprensibile se ci si concentra su spazi campione discreti e di dimensione finita. In tal caso lo spazio campione è descritto da una collezione di *eventi elementari*:

$$\mathcal{S} = \{a_1, a_2, \dots, a_n\}. \quad (1.11)$$

Un evento elementare a_i , è un sottoinsieme dello spazio campione non ulteriormente divisibile (si pensi agli esempi del dado e della moneta). Assegnare delle probabilità a tutti i possibili sottoinsiemi (eventi) di \mathcal{S} può essere fatto assegnando una misura di probabilità a tutti i suoi eventi elementari, visto che tutti gli altri eventi possono essere ottenuti come unioni di essi.

Pertanto, denotando le probabilità degli n eventi elementari con

$$\Pi = \{p_1, p_2, \dots, p_n\}, \quad (1.12)$$

dove $p_i = P(a_i)$, per gli assiomi della probabilità bisognerà imporre che $0 \leq p_i \leq 1$, $i = 1, \dots, n$ e che $\sum_{i=1}^n p_i = 1$.

Esempio 1.1 (cont.): Torniamo all'esempio binario della moneta: $\mathcal{S} = \{T, C\}$, $\Pi = \{p_1, p_2\} = \{P(T), P(C)\}$. Poiché $P(T) + P(C) = 1$, si ha che $\Pi = \{p_1, (1 - p_1)\}$. La caratterizzazione probabilistica dell'esperimento dipende da p_1 e, in generale in una moneta truccata, una delle due facce potrebbe mostrarsi in percentuale più spesso dell'altra. Se invece $p_1 = P(T) = 0.5 = P(C)$, la moneta non è truccata e le due facce hanno la stessa probabilità di mostrarsi come risultato dell'esperimento.

Esempio 1.2 (cont.): Nell'esempio del dado abbiamo: $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, $\Pi = \{p_1, \dots, p_6\} = \{P(1), \dots, P(6)\}$. Poiché $P(1) + \dots + P(6) = 1$, si ha che la caratterizzazione probabilistica dell'esperimento dipende da cinque valori di probabilità, poiché il sesto è imposto dal vincolo. Ad esempio, $\Pi = \{p_1, p_2, p_3, p_4, p_5, (1 - \sum_{i=1}^5 p_i)\}$ e in generale alcune delle facce potrebbero mostrarsi in percentuali diverse. Altrimenti, se $p_i = P(i) = \frac{1}{6}$, il dado non è truccato e le sei facce hanno uguale probabilità di mostrarsi come risultato dell'esperimento. Calcoliamo ora la probabilità di alcuni eventi collegati all'esperimento del dado non truccato usando le proprietà.

$P(B) = Pr\{\text{il risultato del lancio è un numero maggiore o uguale a } 3\} = Pr\{\{3\} \cup \{4\} \cup \{5\} \cup \{6\}\} = Pr\{3\} + Pr\{4\} + Pr\{5\} + Pr\{6\} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3}$. Qui abbiamo usato la additività della probabilità per eventi disgiunti.

$$P(C) = Pr\{\text{il risultato del lancio è un numero minore o uguale a 6}\} = Pr\{\text{il risultato del lancio è un numero qualunque}\} = Pr\{\text{evento certo}\} = 1.$$

$$P(D) = Pr\{\text{il risultato del lancio o è un numero dispari}\} = Pr\{\{1\} \cup \{3\} \cup \{5\}\} = Pr\{1\} + Pr\{3\} + Pr\{5\} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

$$P(E) = Pr\{\text{il risultato del lancio o è un numero dispari, o è maggiore o uguale a 3}\} = Pr\{B \cup D\} = P(B) + P(D) - Pr\{B \cap D\} = P(B) + P(D) - (Pr\{3\} + Pr\{5\}) = \frac{2}{3} + \frac{1}{2} - (\frac{1}{6} + \frac{1}{6}) = \frac{5}{6}.$$

1.5.1 Eventi elementari equiprobabili

Se le probabilità degli eventi elementari sono tali che

$$p_i = \frac{1}{n}, \quad i = 1, \dots, n, \quad (1.13)$$

diremo che gli eventi elementari sono *equiprobabili*, oppure che *le probabilità sono identicamente distribuite*, oppure che *gli eventi elementari sono identicamente distribuiti*.

Nel caso di spazi campione con eventi equiprobabili, è abbastanza agevole calcolare la probabilità di eventi non elementari. Infatti, poiché qualunque evento può essere descritto come unione di eventi elementari, che sono ovviamente disgiunti e hanno uguale probabilità, la formula per la probabilità di un evento A è

$$P(A) = \frac{n_A}{n}, \quad (1.14)$$

dove n_A è la cardinalità (dimensione) dell'evento A ed n è la cardinalità dello spazio campione. Descritta a parole la formula è

$$Pr\{\text{evento } A\} = \frac{\text{numero di modi in cui l'evento } A \text{ può verificarsi}}{\text{numero di possibili risultati dell'esperimento}}. \quad (1.15)$$

Esaminiamo ora alcuni esempi di calcolo in cui la probabilità è valutata usando equazione (1.15).

Esempio 1.2 (cont.): Consideriamo il solito dado a sei facce non truccato (eventi elementari equiprobabili) e ricalcoliamo la probabilità di alcuni eventi ad esso associati usando equazione (1.15):

$$P(A) = Pr\{\text{il risultato del lancio è } 4\} = \frac{1}{6}$$

$$P(B) = Pr\{\text{il risultato del lancio è un numero maggiore o uguale a } 3\} = \frac{4}{6} = \frac{2}{3}$$

$$P(C) = Pr\{\text{il risultato del lancio è un numero minore o uguale a } 6\} = \frac{6}{6} = 1$$

$$P(D) = Pr\{\text{il risultato del lancio è un numero dispari}\} = \frac{3}{6} = \frac{1}{2}$$

Esempio 1.3 (cont.): Torniamo all'esperimento del doppio lancio di una moneta non truccata:

$$P(A) = Pr\{\text{I due lanci danno la sequenza Testa-Croce}\} = \frac{1}{4}$$

$$P(B) = Pr\{\text{I due lanci totalizzano una Testa e una Croce}\} = \frac{2}{4} = \frac{1}{2}$$

$$P(C) = Pr\{\text{Il risultato del primo lancio è Testa}\} = \frac{2}{4} = \frac{1}{2}$$

$$P(D) = Pr\{\text{I due lanci producono la stessa faccia}\} = \frac{2}{4} = \frac{1}{2}$$

Esempio 1.4 (cont.): Torniamo all'esperimento del doppio lancio di un dado non truccato:

$$P(A) = Pr\{\text{I due lanci danno la sequenza } 6-2\} = \frac{1}{36}$$

$$P(B) = Pr\{\text{I due lanci totalizzano una somma pari a } 8\} = \frac{5}{36}$$

$$P(C) = Pr\{\text{I due lanci totalizzano una somma minore o uguale a } 4\} = \frac{6}{36} = \frac{1}{6}$$

$$P(D) = Pr\{\text{I due lanci danno lo stesso numero}\} = \frac{6}{36} = \frac{1}{6}$$

$$P(E) = Pr\{\text{I due lanci danno almeno un } 6\} = \frac{11}{36}$$

Esempio 1.5 (cont.): Torniamo alla trasmissione del bit “0” sul canale con errori supponendo che la probabilità di ricevere il bit “0” è uguale a quella di ricevere il bit “1”:

$$P(A) = Pr\{\text{il bit è ricevuto correttamente}\} = \frac{1}{2}$$

$$P(B) = Pr\{\text{il bit è ricevuto erroneamente}\} = \frac{1}{2}$$

$$P(C) = Pr\{\text{il bit non è ricevuto (cancellazione)}\} = P(\emptyset) = 0$$

Esempio 1.6 (cont.): Si consideri di nuovo l’esperimento delle 13 partite (Totocalcio), assumendo che tutti i risultati siano equiprobabili.

$$P(A) = Pr\{\text{tutte le partite terminano in parità}\} = \frac{1}{1594323}$$

$$P(B) = Pr\{\text{tutte le partite terminano in parità eccetto che per la seconda}\} = \frac{2}{1594323}$$

$$P(C) = Pr\{\text{tutte le partite, eccetto una, terminano in parità}\} = \frac{26}{1594323}$$

$$P(D) = Pr\{\text{si verifica una delle due colonne giocate}\} = \frac{2}{1594323}$$

Esempio 1.8 (cont.): Si consideri nuovamente l’esperimento della estrazione senza rimpiazzo di sei palle numerate da 1 a 90. Si supponga che l’ordine in cui i numeri vengono estratti sia irrilevante (Enalotto):

$$P(A) = Pr\{\text{si verifica la sestina } 2, 23, 28, 50, 78, 90\} = \frac{1}{622614630}$$

$$P(B) = Pr\{\text{esce la quaterna } 15, 47, 48, 86\} = \frac{3655}{622614630}$$

Esempio 1.9 (cont.): Torniamo all’esperimento in cui tre giocatori estraggono ognuno una carta da un mazzo di 40 carte napoletane:

$$P(A) = Pr\{\text{i tre giocatori hanno rispettivamente un asso di bastoni (1B), un sette di spade (7S) e un sei di coppe (6C)}\} = \frac{1}{59280}$$

$$P(B) = Pr\{\text{i tre giocatori hanno tutti un asso e nessuno l’asso di ori (1O)}\} = \frac{6}{59280}$$

Esempio 1.10 (cont.): Torniamo all'esperimento in cui un calcolatore emette una sequenza di tre bit. Si supponga che tutte le stringhe siano equiprobabili:

$$P(A) = Pr\{\text{la sequenza rappresenta il numero 5 in codice binario}\} = \frac{1}{8}$$

$$P(B) = Pr\{\text{la sequenza è costituita da tutti bit uguali}\} = \frac{2}{8} = \frac{1}{4}$$

$$P(C) = Pr\{\text{la sequenza contiene un solo bit pari a 0}\} = \frac{3}{8}$$

$$P(D) = Pr\{\text{la sequenza inizia con il prefisso 10}\} = \frac{2}{8} = \frac{1}{4}$$

$$P(E) = Pr\{\text{la sequenza contiene un numero dispari di 1}\} = \frac{4}{8} = \frac{1}{2}$$

$$P(F) = Pr\{\text{la sequenza contiene almeno due 1}\} = \frac{4}{8} = \frac{1}{2}$$

Esempio 1.12 (cont.): Si consideri nuovamente l'esperimento in cui quattro giocatori ricevono 10 carte ciascuno da un mazzo di 40 carte napoletane:

$P(A) = Pr\{\text{il giocatore 1 ha tutti i sette}\}$. Per calcolare tale probabilità si osservi che il numero di modi in cui la partizione delle 40 carte può assegnare i quattro sette al primo giocatore è pari al coefficiente multinomiale $\binom{36}{6 \ 10 \ 10 \ 10}$. Questo è facilmente visualizzabile se si mette da parte la quaterna dei 4 sette e si distribuiscono le rimanenti 36 carte tra i 4 giocatori nella partizione 6,10,10,10. Pertanto: $P(A) = \frac{\binom{36}{6 \ 10 \ 10 \ 10}}{\binom{40}{10 \ 10 \ 10 \ 10}} = \frac{36! \ 10!}{40! \ 6!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{40 \cdot 39 \cdot 38 \cdot 37}$. Come verifica al conteggio si può anche proseguire come segue. Si consideri l'insieme delle 40 carte ordinate con i 4 sette in prima posizione nei primi 4 posti dell'insieme di 10 carte relativo al primo giocatore. Per ottenere le configurazioni incluse in A si possono permutare tutte le 36 carte ($36!$), piazzare i 4 sette a piacere sul primo insieme di 10 carte ($\binom{10}{4}$) e permutare i 4 sette tra loro ($4!$). In questo caso $n_S = 40!$ e $n_A = 36! \binom{10}{4} 4!$. Quindi: $P(A) = \frac{36! \binom{10}{4} 4!}{40!} = \frac{36! 10!}{40! 6!}$.

$P(B) = Pr\{\text{i quattro giocatori ricevono carte di un solo colore}\}$. Per contare il numero di modi in cui tale evento può verificarsi (ogni giocatore ha una scala completa), consideriamo la partizione che assegna al primo tutti gli ori, al secondo tutte le coppe, al terzo

tutti i bastoni, al quarto tutte le spade, esattamente nell'ordine da 1 a 10. Questa è evidentemente una sola delle $40!$ permutazioni delle 40 carte, visto che l'ordine è totalmente fissato. Per generare le possibili permutazioni incluse nell'evento B , permutiamo i colori tra i giocatori ($4!$) e permutiamo le carte all'interno delle dieci carte di ognuno ($10!$ per ognuno). Otteniamo pertanto $4! \cdot 10! \cdot 10! \cdot 10! \cdot 10!$ permutazioni, che è il numero cercato. Quindi: $P(B) = \frac{4!(10!)^4}{40!}$.

$P(C) = Pr\{\text{i giocatori ricevono un sette per ciascuno}\}$. Si mettano da parte i quattro sette. Le restanti 36 carte vanno distribuite tra i quattro giocatori nella partizione 9,9,9,9. Anche i 4 sette possono essere permutati a piacere. Quindi: $P(C) = \frac{4! \binom{36}{9,9,9,9}}{\binom{40}{10,10,10,10}} = \frac{4! \cdot 10^4}{40 \cdot 39 \cdot 38 \cdot 37}$. Un approccio alternativo al conteggio, che può essere effettuato per verifica, è il seguente. Si consideri l'insieme delle 40 carte ordinate, come nello studio dell'evento B : ovvero ogni giocatore ha una scala ordinata con il sette in settima posizione. Per ottenere le configurazioni incluse in C , possiamo: permutare i colori dei 4 giocatori ($4!$); cambiare il posto a ogni sette nell'ambito del sottoinsieme di ogni giocatore ($10 \cdot 10 \cdot 10 \cdot 10$); possiamo infine permutare a piacere tutte le rimanenti 36 carte ($36!$). Quindi la probabilità di C può essere espressa come il rapporto $\frac{4! \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 36!}{40!}$ che è lo stesso risultato già trovato usando i coefficienti multinomiali.

$P(D) = Pr\{\text{il giocatore 3 riceve una scala completa}\}$. Anche qui, come nello studio dell'evento B , partiamo da una scala completa ordinata a testa, ovvero consideriamo la partizione che assegna al primo tutti gli ori, al secondo tutte le coppe, al terzo tutti i bastoni, al quarto tutte le spade, esattamente nell'ordine da 1 a 10. Poiché non è specificato di quale colore deve essere la scala del terzo giocatore, gli si può assegnare uno di quattro colori (4). Poiché l'ordine non è importante, le sue 10 carte possono essere permutate a piacere ($10!$). Inoltre tutte le rimanenti 30 carte possono essere permutate a piacere ($30!$). Quindi: $P(D) = \frac{4 \cdot 10! \cdot 30!}{40!}$.

1.5.2 Il numero di successi in un campione

Un problema classico della teoria della probabilità, che merita attenzione particolare, è il seguente:

Un'urna contiene M palle di cui M_b bianche ($M_b < M$) e $M - M_b$ rosse. Un campione di n palle è estratto, con o senza rimpiazzo, dalle M palle (se l'estrazione è senza rimpiazzo, $n \leq M$, altrimenti non c'è vincolo su n). Sia k un intero tra 0 e n ($0 \leq k \leq n$). Qual'è la probabilità che il campione contiene esattamente k palle bianche? Ogni palla bianca è detta "successo" dell'esperimento (se l'estrazione è senza rimpiazzo $k \leq M_b$).

La rilevanza di questo problema è nelle numerose applicazioni che ad esso possono essere associate. Un esempio tipico è quello dell'impianto manifatturiero che produce M unità di prodotto. Si supponga che M_b di M unità di prodotto siano difettose. Se si campiona la produzione esaminando solo $n < M$ esemplari, qual'è la probabilità che k di essi siano difettosi? La risposta a questo quesito fornisce utili relazioni tra la percentuale di prodotti difettosi e la probabilità che il campione estratto contenga dei campioni difettosi. Questo consente di progettare delle strategie di campionamento che stimino la bontà della produzione da un campione limitato.

Consideriamo quindi prima la risoluzione del problema *senza rimpiazzo*. Si supponga di numerare tutte le palle da 1 a M raggruppando le M_b palle bianche nella prima parte dell'insieme. L'esperimento è: si estrae un campione di n palle senza rimpiazzo dalle M palle dell'urna. Lo spazio campione del problema è $\mathcal{S} = \{\text{insieme di tutte le } n\text{-ple estratte dagli } M \text{ elementi}\}$. La dimensione dello spazio campione è $n_{\mathcal{S}} = (M)_n$, ovvero è pari al numero di disposizioni di n elementi estratti senza rimpiazzo da un'urna di M elementi. Si supponga che tutte le configurazioni estratte siano equiprobabili. L'evento di cui vogliamo calcolare la probabilità è $E = \{\text{l}'n\text{-upla contiene esattamente } k \text{ palle bianche}\}$. Qui ci sono i vincoli che $n \leq M$ e che $k \leq M_b$, altrimenti la probabilità è nulla. Seguiamo un approccio costruttivo al conteggio della dimensione di

E . Supponiamo inizialmente che il campione di n palle contenga le k palle bianche nelle prime k posizioni. Il numero di modi in cui le k palle bianche possono essere state estratte senza rimpiazzo dalle M_b palle bianche è $(M_b)_k$. Il numero di modi in cui le $(n - k)$ palle rosse possono essere state estratte senza rimpiazzo dalle $(M - M_b)$ palle rosse è $(M - M_b)_{n-k}$. Infine, poiché le k palle bianche possono essere permutate sul campione di n palle in $\binom{n}{k}$ modi, la dimensione di E è $\binom{n}{k}(M_b)_k(M - M_b)_{n-k}$. La probabilità cercata è quindi

$$P(E(k, n; M_b, M)) = \binom{n}{k} \frac{(M_b)_k (M - M_b)_{n-k}}{(M)_n}, \quad (1.16)$$

$$k = 0, \dots, n; \quad n = 1, \dots, M.$$

Equazione (1.16) in forma estesa è

$$P(E(k, n; M_b, M)) = \frac{n!}{k!(n-k)!} \frac{(M_b)!}{(M_b-k)!} \frac{(M-n)!}{M!} \frac{(M-M_b)!}{(M-M_b-n+k)!}. \quad (1.17)$$

Ricordando che: $\binom{a}{b} = \frac{(a)_b}{b!}$, equazione (1.16) assume la forma più compatta

$$P(E(k, n; M_b, M)) = \frac{\binom{M_b}{k} \binom{M-M_b}{n-k}}{\binom{M}{n}}, \quad k = 0, \dots, n; \quad n = 1, \dots, M. \quad (1.18)$$

Il caso del campionamento *con rimpiazzo* può essere studiato con ragionamento analogo, semplicemente considerando le disposizioni con rimpiazzo. La probabilità cercata è pertanto

$$P(F(k, n; M_b, M)) = \binom{n}{k} \frac{(M_b)^k (M - M_b)^{n-k}}{(M)^n}, \quad k = 0, \dots, n. \quad (1.19)$$

E' utile introdurre il parametro

$$p = \frac{M_b}{M}, \quad (1.20)$$

che rappresenta la frazione di palle bianche nell'insieme di dimensione M .

Le probabilità di cui sopra, possono essere espresse come

$$P(E) = \binom{n}{k} p^k (1-p)^{n-k} \alpha, \quad (1.21)$$

$$P(F) = \binom{n}{k} p^k (1-p)^{n-k}, \quad (1.22)$$

dove α ha la seguente espressione

$$\alpha = \frac{(1-\frac{1}{M_b})(1-\frac{2}{M_b})\cdots(1-\frac{k-1}{M_b})(1-\frac{1}{M-M_b})(1-\frac{2}{M-M_b})\cdots(1-\frac{n-k-1}{M-M_b})}{(1-\frac{1}{M})(1-\frac{2}{M})\cdots(1-\frac{n-1}{M})}. \quad (1.23)$$

La espressione di α può essere verificata per diretta sostituzione ed è lasciata al lettore per esercizio.

Equazione (1.22) è un caso particolare di un risultato molto generale detto *legge binomiale*. Maggiori dettagli saranno dati in seguito. Si noti comunque come equazioni (1.22) e (1.21) coincidano asintoticamente. Infatti quando $M \rightarrow \infty$ abbiamo che $\alpha \rightarrow 1$. Quindi la differenza tra i risultati relativi a estrazioni con o senza rimpiazzo diventa sempre più piccola quando M diventa grande rispetto a n .

Esempio 1.16 Si consideri una lotteria che vende M biglietti e conferisce m premi. La domanda è: se un giocatore compra n biglietti, che probabilità ha di vincere? **Soluzione:** L'insieme degli M biglietti è suddiviso in m vincenti e $M-m$ perdenti. Il problema può essere risolto in analogia all'esempio dell'estrazione da un insieme di palle bianche e rosse, su riportato. Si noti che l'estrazione è da associare agli n biglietti acquistati e non agli m vincenti (nonostante la estrazione degli m biglietti vincenti avvenga successivamente all'aquisto degli n biglietti). L'evento vincita consiste nella presenza di uno o più biglietti vincenti nel campione di n biglietti acquistati. Per semplicità di calcolo conviene calcolare la probabilità di non vincere P_p , poiché la probabilità di vincere P_v può essere scritta come $P_v = 1 - P_p$. L'evento perdita è costituito da tutte le possibili combinazioni di n biglietti che non contengono alcun

biglietto vincente. Pertanto, usando equazione (1.16) con $M_b = m$ e $k = 0$, si ottiene

$$P_v = 1 - P_p = 1 - \frac{(M - m)_n}{(M)_n}. \quad (1.24)$$

Esempio numerico: $M = 1000$, $m = 5$, $n = 2$, $P_v = 1 - \frac{995 \cdot 994}{1000 \cdot 999} = 0.01$.

1.6 Spazi continui uniformi

Le definizioni fondamentali e gli assiomi della teoria della probabilità si applicano in maniera abbastanza agevole a spazi campione discreti, dove ad ogni evento viene associata, in maniera diretta o indiretta, una misura di probabilità. Nel caso di spazi campione continui la assegnazione di tale misura presenta qualche insidia in più a causa della struttura densa dei sottoinsiemi dello spazio campione. Si consideri l'esempio 1.7 della misura della temperatura ambiente. Se assegnassimo ad ogni numero reale contenuto nell'intervallo un valore finito di probabilità, avremmo che la somma di tali valori sarebbe infinita, vista la densità dell'insieme dei numeri reali. Avremmo pertanto violato il secondo assioma il quale impone che la somma dei contributi delle probabilità su tutto lo spazio campione sia pari a uno. Inoltre la assegnazione di probabilità deve possedere la proprietà di sovrapposizione imposta dal terzo assioma.

Una assegnazione coerente di probabilità a spazi continui deve pertanto seguire una strada diversa da quella usata per spazi campione discreti. Anche se tale costruzione verrà presentata nel capitolo sulle variabili aleatorie continue, è possibile in maniera intuitiva e geometrica, in analogia agli esempi della sezione precedente, risolvere alcuni problemi che si riferiscono a *spazi continui finiti* su cui la *probabilità è distribuita in maniera uniforme*. In tali casi si può adottare la formula intuitiva

$$Pr\{\text{evento } A\} = \frac{\text{dimensione dell'insieme } A}{\text{dimensione dello spazio campione}}. \quad (1.25)$$

Una giustificazione di tale espressione sarà data nei capitoli riguardanti la variabile aleatoria.

Esempio 1.7 (cont.): Torniamo all'esempio delle temperature. Si assuma che la temperatura sia uniformemente distribuita in $\mathcal{S} = [-20, 60]$. La dimensione dello spazio campione è la lunghezza dell'intervallo, ovvero: $\dim\mathcal{S} = 80$. Quindi valgono i seguenti risultati.

$$P(A) = Pr\{\text{la temperatura misurata è tra } 36.5 \text{ e } 36.6\} = \frac{0.1}{80} = 125 \cdot 10^{-5}$$

$$P(B) = Pr\{\text{la temperatura misurata è inferiore ai } 40 \text{ gradi}\} = \frac{60}{80} = 0.75$$

$$P(C) = Pr\{\text{la temperatura misurata è maggiore di } -25 \text{ gradi}\} = \frac{80}{80} = 1$$

$$P(D) = Pr\{\text{la temperatura misurata è esattamente di } 30 \text{ gradi}\} = \frac{0}{80} = 0$$

(il significato di questo risultato sarà chiarito ulteriormente quando parleremo di variabili aleatorie continue)

Esempio 1.13 (cont.): Torniamo all'esempio dell'arciere. Si assuma che la freccia abbia uguale probabilità di conficcarsi su qualunque posizione del bersaglio e che non sia possibile che cada al di fuori di esso. La dimensione dello spazio campione è la dimensione del bersaglio: $\dim\mathcal{S} = \pi$. Quindi valgono i seguenti risultati.

$$P(A) = Pr\{\text{la freccia si conficca esattamente al centro}\} = \frac{0}{\pi} = 0$$

$$P(B) = Pr\{\text{la freccia si conficca entro un raggio di } 1 \text{ centimetro dal centro}\} = \frac{\pi(0.01)^2}{\pi} = 0.0001.$$

$$P(C) = Pr\{\text{la freccia si conficca in un anello definito dai raggi } 0.5 \text{ e } 0.6 \text{ metri}\} = \frac{\pi(0.6)^2 - \pi(0.5)^2}{\pi} = 0.11.$$

1.7 Probabilità condizionata

La definizione di probabilità condizionata è di fondamentale importanza poiché consente di caratterizzare la probabilità di eventi, una volta che altri eventi ad essi collegati si sono verificati. Dandone una definizione non rigorosa possiamo dire che:

Dati due eventi A e B , per probabilità condizionata dell'evento A , dato l'evento B , denotata come $P(A|B)$, si intende la probabilità che A si verifichi, sotto l'assunzione che B si è verificato. In altre parole $P(A|B)$ rappresenta la valutazione della probabilità di A quando si è in possesso dell'informazione che B si è verificato.

Per apprezzare la definizione intuitiva appena data di probabilità condizionata, si pensi ad un caso di eventi discreti dove si utilizza invece della definizione assiomatica di probabilità, quella frequentista. Si supponga di osservare un grande numero N di occorrenze di un fenomeno aleatorio nel quale siano definiti gli eventi A e B . Siano N_A ed N_B il numero di volte in cui si verificano gli eventi A e B rispettivamente. Sia inoltre N_{AB} il numero di volte in cui entrambi gli eventi A e B si verificano simultaneamente. Associando alle frequenze il significato di probabilità, abbiamo che

$$p(A) = \frac{N_A}{N}, \quad p(B) = \frac{N_B}{N}, \quad p(AB) = \frac{N_{AB}}{N}. \quad (1.26)$$

La probabilità condizionata di A dato B , è la frequenza con cui si verificano insieme A e B nell'ambito delle occorrenze di B . In simboli

$$p(A|B) = \frac{N_{AB}}{N_B}. \quad (1.27)$$

Ovviamente la definizione non ha senso se N_B è zero. In tale caso diremo che la probabilità condizionata non è definita. Per arrivare ad una definizione più precisa delle probabilità condizionata, notiamo che equazione (1.27) può essere riscritta come

$$p(A|B) = \frac{\frac{N_{AB}}{N}}{\frac{N_B}{N}} = \frac{p(AB)}{p(B)}, \quad (1.28)$$

ovvero che la probabilità di A condizionata da B può essere espressa come il rapporto della probabilità dell'evento congiunto AB e della probabilità marginale di B . L'espressione così trovata è quella che utilizzeremo come definizione formale di probabilità condizionata.

Siano A e B due eventi dello spazio campione \mathcal{S} . Sia definita su tale spazio una misura di probabilità $P(\cdot)$. La probabilità condizionata dell'evento A dato l'evento B , denotata con $P(A|B)$, è definita come

$$P(A|B) = \frac{P(AB)}{P(B)}, \quad \text{se } P(B) > 0. \quad (1.29)$$

Altrimenti se $P(B) = 0$, $P(A|B)$ non è definita.

Cerchiamo ora di comprendere meglio il significato di probabilità condizionata mediante alcuni esempi.

Esempio 1.17 Si consideri una popolazione di persone di sesso maschile o femminile il cui colore dei capelli può essere biondo, o bruno. Lo spazio campione è dato dalle quattro possibili combinazioni $\{(c.biondi, maschio), (c.biondi, femmina), (c.bruni, maschio), (c.bruni, femmina)\}$. Le probabilità relative siano assegnate e pari a $\{\frac{1}{12}, \frac{1}{3}, \frac{1}{3}, \frac{1}{4}\}$. Ci si chiede quale sia la probabilità che dato un soggetto femmina il colore dei suoi capelli sia biondo. Si noti che la probabilità condizionata cercata è diversa dalla probabilità dell'evento congiunto (c.biondi, femmina) (nel nostro caso pari a $\frac{1}{3}$). Si osservi prima che la probabilità marginale $Pr\{\text{femmina}\} = \frac{1}{4} + \frac{1}{3} = \frac{7}{12}$. Pertanto, $Pr\{\text{c.biondi}|\text{femmina}\} = \frac{Pr\{\text{c.biondi, femmina}\}}{Pr\{\text{femmina}\}} = \frac{\frac{1}{3}}{\frac{7}{12}} = \frac{4}{7}$.

Esempio 1.18 Si consideri una famiglia con due figli. Ogni figlio può essere un maschio (M), o una femmina (F). Lo spazio campione è: $\mathcal{S} = \{MM, MF, FM, FF\}$. Si supponga che ad ogni evento elementare dello spazio campione sia assegnata una probabilità di $\frac{1}{4}$, ovvero ogni configurazione dei due sessi sia equiprobabile. Si calcolino le seguenti quantità: 1. la probabilità che il primo figlio sia maschio; 2. la probabilità che il secondo figlio sia maschio; 3. la probabilità che entrambi

i figli siano maschi; 4. la probabilità che il secondo figlio sia femmina dato che il primo figlio è maschio; 5. la probabilità che entrambi siano maschi dato che almeno uno è maschio.

Soluzione: Si definiscano gli eventi A e B come segue: $A = \{\text{il primo figlio è maschio}\}$; $B = \{\text{il secondo figlio è maschio}\}$. Gli eventi A e B possono anche essere esplicitamente enumerati come segue: $A = \{MM, MF\}$; $B = \{MM, FM\}$. Quindi: 1. $P(A) = \frac{2}{4} = \frac{1}{2}$; 2. $P(B) = P(A) = \frac{1}{2}$; 3. $P(AB) = \frac{1}{4}$; 4. $P(\bar{B}|A) = 1 - P(B|A) = 1 - \frac{P(AB)}{P(A)} = 1 - \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$; 5. L'evento $C = \{\text{almeno un figlio è maschio}\}$ può essere scritto come $C = A + B$. La probabilità cercata è $P(AB|A+B) = \frac{P((AB)(A+B))}{P(A+B)} = \frac{P(AB)}{P(A+B)} = \frac{P(AB)}{P(A)+P(B)-P(AB)} = \frac{\frac{1}{4}}{(\frac{1}{2}+\frac{1}{2}-\frac{1}{4})} = \frac{1}{3}$ (si noti come questo risultato sia in effetti un pò meno intuitivo del precedente).

Immedie conseguenze della definizione di probabilità condizionata sono le seguenti utili proprietà.

Proprietà 1.8 Se A e B sono disgiunti, ovvero $AB = \emptyset$, $P(A|B) = P(B|A) = 0$. Si dice che in tal caso i due eventi sono *incompatibili*.²

Prova: Immediata dalla definizione poiché $P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(\emptyset)}{P(B)} = 0$. Analogamente per $P(B|A)$.

Proprietà 1.9 Se $A \subseteq B$, $P(A|B) \geq P(A)$.

Prova: Immediata dalla definizione poiché $P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} \geq P(A)$.

Proprietà 1.10 Se $B \subseteq A$, $P(A|B) = 1$.

Prova: Immediata dalla definizione poiché $P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B)}{P(B)} = 1$.

²Si faccia attenzione alla importante differenza tra incompatibilità e indipendenza. L'indipendenza sarà introdotta in seguito e si verificherà come eventi incompatibili sono anche indipendenti, ma non necessariamente il viceversa.

1.7.1 La legge della probabilità totale

C'è un risultato importante che discende direttamente dalla definizione di probabilità condizionata. Siano A_1, A_2, \dots, A_N un insieme di eventi mutuamente esclusivi ed esaustivi

$$\bigcup_{i=1}^N A_i = \mathcal{S}, \quad A_i A_j = \emptyset, \quad i \neq j, \quad (1.30)$$

dove \mathcal{S} denota lo spazio campionario. Dato un evento qualunque B , è possibile esprimere la sua come la somma pesata delle probabilità condizionate $\{P(B|A_i), i = 1, \dots, N\}$

$$P(B) = \sum_{i=1}^N P(B|A_i)P(A_i). \quad (1.31)$$

Tale risultato è detto *legge della probabilità totale*. La dimostrazione è immediata se si osserva che $B = BA_1 + BA_2 + \dots + BA_N$, come è mostrato in figura 1.4, e che gli eventi $\{BA_1, BA_2, \dots, BA_N\}$ sono disgiunti. Pertanto dal terzo assioma e dalla definizione di probabilità condizionata

$$\begin{aligned} P(B) &= P(BA_1) + P(BA_2) + \dots + P(BA_N) \\ &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_N)P(A_N), \end{aligned} \quad (1.32)$$

che è il risultato enunciato.

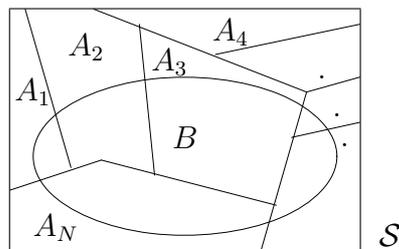


Figura 1.4: Il diagramma di Venn per la legge della probabilità totale.

1.7.2 Il teorema di Bayes

C'è un importante corollario della legge della probabilità totale se si interpreta l'evento B in equazione (1.31) come conseguenza degli eventi A_1, A_2, \dots, A_N . In altre parole, se agli eventi A_1, A_2, \dots, A_N viene attribuito il significato di *cause* di B , la legge della probabilità totale può essere letta come la scomposizione della probabilità di B secondo la probabilità delle sue cause. Poiché dalla definizione di probabilità condizionata è possibile invertire l'ordine del condizionamento, ovvero

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}, \quad (1.33)$$

la probabilità di ogni evento condizionante, dato B , può essere espressa in funzione delle probabilità marginali delle cause e delle probabilità condizionate di B dato A_i , $i = 1, \dots, N$, ovvero

$$P(A_i|B) = \frac{P(BA_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^N P(B|A_i)P(A_i)}, \quad i = 1, \dots, N. \quad (1.34)$$

Pertanto la probabilità $P(A_i|B)$ può essere interpretata come la probabilità che sia stato l'evento A_i a causare l'evento B . La espressione (1.34) è detta *Formula o Teorema di Bayes*.

Esempio 1.19 Si consideri il seguente esperimento di diagnostica. Un paziente mostra il sintomo $B = \{\text{mal di testa}\}$. Si supponga che le possibili cause del mal di testa possano essere solo $E = \{\text{encefalopatia}\}$, $I = \{\text{influenza}\}$, $S = \{\text{stress}\}$ e che tali eventi siano disgiunti ed esaustivi, ovvero che non possano coesistere e che non esistano altre patologie possibili (è solo un esempio ovviamente): $EI = ES = IS = \emptyset$, $E + I + S = \mathcal{S} = \{\text{evento certo}\}$. La legge della probabilità totale ci consente di scrivere: $P(B) = P(B|E)P(E) + P(B|I)P(I) + P(B|S)P(S)$, ovvero scomporre la probabilità di occorrenza del mal di testa secondo le probabilità condizionate $\{P(B|E), P(B|I), P(B|S)\}$, che esprimono la relazione specifica di dipendenza delle cause con l'effetto, e le probabilità $\{P(E), P(I), P(S)\}$ che esprimono la probabilità marginale delle tre patologie. Dalla conoscenza di questi sei parametri, il teorema di Bayes ci consente di risalire alla probabilità che ogni patologia sia la causa del mal di testa. Con riferimento ad un esempio numerico in

cui $\{P(B|E), P(B|I), P(B|S)\} = \{0.5, 0.6, 0.9\}$ e $\{P(E), P(I), P(S)\} = \{0.01, 0.2, 0.79\}$, abbiamo che

$$\begin{aligned} P(E|B) &= \frac{P(B|E)P(E)}{P(B|E)P(E) + P(B|I)P(I) + P(B|S)P(S)} = 0.0060 \\ P(I|B) &= \frac{P(B|I)P(I)}{P(B|E)P(E) + P(B|I)P(I) + P(B|S)P(S)} = 0.1435 \\ P(S|B) &= \frac{P(B|S)P(S)}{P(B|E)P(E) + P(B|I)P(I) + P(B|S)P(S)} = 0.8505 \end{aligned}$$

Le tre probabilità trovate sono anche dette probabilità “a posteriori,” mentre le tre probabilità $\{P(E), P(I), P(S)\}$ sono dette probabilità “a priori.”

1.8 Indipendenza

La nozione di eventi dipendenti e indipendenti è di importanza fondamentale nella teoria della probabilità. In effetti l'analisi di molti fenomeni aleatori si basa spesso su forti assunzioni di indipendenza di cause ed effetti che conferiscono una struttura probabilistica al modello matematico di riferimento.

Se un evento A è tale che la sua probabilità condizionata dall'evento B è la stessa della probabilità di A senza condizionamento, diremo che A è indipendente da B . Ovviamente la probabilità condizionata deve esistere e quindi l'evento B deve avere probabilità non nulla. Vale lo stesso ragionamento se esaminiamo la probabilità condizionata di B dato A . E' possibile pertanto dare la seguente definizione di indipendenza.

Siano A e B due eventi appartenenti allo stesso spazio campione S a probabilità non nulla. Diremo che A è indipendente da B se $P(A|B) = P(A)$. Poiché è possibile scrivere

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (1.35)$$

ne consegue che anche $P(B|A) = P(B)$. Inoltre, poiché $P(A|B) = \frac{P(AB)}{P(B)}$, abbiamo che $P(AB) = P(A)P(B)$.

La definizione data è piuttosto intuitiva, ma richiede che sia A che B siano eventi a probabilità non nulla. Pertanto la definizione più generale è:

Due eventi A e B si dicono indipendenti (a volte si dice “statisticamente indipendenti”) se

$$P(AB) = P(A)P(B). \quad (1.36)$$

Quindi due eventi sono indipendenti se la probabilità della loro intersezione è fattorizzabile nelle due probabilità marginali. In tale definizione le probabilità marginali possono anche essere nulle.

Esempio 1.4 (cont.): Per comprendere meglio il significato della definizione di indipendenza si ritorni all’esperimento del lancio dei due dadi. La domanda è se il risultato del lancio del primo dado influisca probabilisticamente su quella del secondo dado quando tutte le coppie sono assunte equiprobabili (dadi non truccati). La risposta è immediata se calcoliamo la probabilità condizionata $Pr\{i|j\} = \frac{Pr\{(i,j)\}}{Pr\{j\}} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$. Questo risultato vale per ogni i e j ed è uguale alla probabilità marginale del secondo lancio. Quindi i due eventi sono indipendenti (in effetti l’esperimento prevedeva due lanci indipendenti, quindi il risultato non fa altro che confermare l’ipotesi). La stessa conclusione poteva essere raggiunta osservando che la probabilità congiunta si fattorizza, ovvero $Pr\{(i,j)\} = \frac{1}{36} = \frac{1}{6} \frac{1}{6} = Pr\{i\}Pr\{j\}$.

Esempio 1.17 (cont.): Torniamo all’esempio in cui una popolazione di persone di sesso maschile o femminile può avere capelli di colore biondo o bruno. La domanda che ci poniamo è se c’è dipendenza probabilistica tra sesso e colore dei capelli quando la assegnazione di probabilità è quella data. La soluzione è piuttosto semplice poichè basta confrontare, per esempio, la probabilità condizionata $Pr\{c.biondi|femmina\} = \frac{12}{21}$ con quella non condizionata $Pr\{c.biondi\} = \frac{1}{12} + \frac{1}{3} = \frac{5}{12}$. Pertanto gli eventi $A = \{biondi\}$, $B = \{femmina\}$ non sono indipendenti (in questo esempio è più probabile che una femmina sia bionda). Le altre probabilità condizionate possono essere facilmente calcolate in maniera analoga.

Se si è in presenza di un numero di eventi superiore a due, la definizione di indipendenza va generalizzata. Adottando l'approccio basato sulla probabilità condizionata, è necessario considerare tutti i possibili raggruppamenti condizionanti. Per esempio, si consideri un insieme di tre eventi A , B e C . Per sancire la indipendenza di A dagli altri due bisogna accertare che

$$P(A|B) = P(A), \quad P(A|C) = P(A), \quad P(A|BC) = P(A), \quad (1.37)$$

o equivalentemente che

$$P(AB) = P(A)P(B), \quad P(AC) = P(A)P(C), \quad P(ABC) = P(A)P(BC). \quad (1.38)$$

Ciò ovviamente non implica che B e C siano indipendenti. Per avere che B sia indipendente da A e da C , bisogna imporre che

$$P(BA) = P(B)P(A), \quad P(BC) = P(B)P(C), \quad P(BAC) = P(B)P(AC). \quad (1.39)$$

Analogamente affinché C sia indipendente da A e da B bisogna avere che

$$P(CA) = P(C)P(A), \quad P(CB) = P(C)P(B), \quad P(CAB) = P(C)P(AB). \quad (1.40)$$

Pertanto l'insieme dei tre eventi A , B e C costituisce un insieme di eventi (completamente) indipendenti se e solo se

$$P(AB) = P(A)P(B), \quad P(AC) = P(A)P(C), \quad P(BC) = P(B)P(C), \quad (1.41)$$

$$P(ABC) = P(A)P(B)P(C). \quad (1.42)$$

Quindi se solo le equazioni (1.41) sono soddisfatte, i tre eventi sono solo indipendenti a coppie.

L'estensione al caso di n eventi è abbastanza intuitiva: dati n eventi A_1, A_2, \dots, A_n , essi sono indipendenti se per ogni insieme di interi $1 \leq k_1, k_2, \dots, k_l \leq n$ si ha che

$$P(A_{k_1}, A_{k_2}, \dots, A_{k_l}) = P(A_{k_1})P(A_{k_2}) \cdots P(A_{k_l}). \quad (1.43)$$

Dati n eventi, sono necessarie $\binom{n}{2}$ relazioni tra le coppie; $\binom{n}{3}$ relazioni tra le terne, $\binom{n}{4}$ relazioni tra le quaterne, eccetera. Quindi il numero totale di condizioni da imporre tra n eventi è

$$\sum_{i=2}^n \binom{n}{i} = \sum_{i=0}^n \binom{n}{i} - \binom{n}{1} - \binom{n}{0} = 2^n - n - 1. \quad (1.44)$$

1.9 Indipendenza condizionata

Il concetto di indipendenza probabilistica tra eventi può essere anche applicato alla probabilità condizionata.

Siano A , B e C tre eventi. Sia $P(AB|C)$ la probabilità dell'intersezione di A e B , condizionata all'evento C . Diremo che A è indipendente da B , condizionatamente a C , se

$$P(AB|C) = P(A|C)P(B|C). \quad (1.45)$$

E' importante enfatizzare che la indipendenza condizionata non implica l'indipendenza. Infatti due eventi possono essere condizionalmente indipendenti, ma non indipendenti. Maggiori dettagli sull'indipendenza condizionata saranno discussi negli esempi.

1.10 Prove indipendenti

Nelle note introduttive alla teoria della probabilità, abbiamo fatto riferimento al risultato di esperimenti e caratterizzato lo spazio campione relativo. Anche se la probabilità degli eventi associati all'esperimento serve a modellare la frequenza di occorrenza di tali eventi in prove ripetute dello stesso esperimento, il modello probabilistico di riferimento resta quello di una singola prova. In altre parole la probabilità degli eventi associati al modello sono quelle di una singola esecuzione dell'esperimento. La importante distinzione tra prove indipendenti ripetute dello stesso esperimento e la frequenza relativa *misurata* da n prove, è sottile e genera spesso delle ambiguità di interpretazione nella teoria della probabilità. La questione sarà meglio chiarita quando parleremo della legge dei grandi numeri. Per ora definiamo il seguente esperimento composto.

Si consideri un esperimento a cui corrisponde lo spazio campione \mathcal{S} . L'esperimento composto da n prove di tale esperimento corrisponde ad un nuovo spazio campione costituito da tutte le n -ple appartenenti a \mathcal{S}^n , ovvero al prodotto cartesiano di \mathcal{S} con se stesso n volte. Denotando con $\xi_1 \in \mathcal{S}, \xi_2 \in \mathcal{S}, \dots, \xi_n \in \mathcal{S}$ i risultati delle n prove,

$$(\xi_1, \xi_2, \dots, \xi_n) \in \mathcal{S}^n = \underbrace{\mathcal{S} \times \mathcal{S} \times \dots \times \mathcal{S}}_n. \quad (1.46)$$

Un esperimento è costituito da n prove *indipendenti* se i risultati delle varie prove non si influenzano a vicenda. Più formalmente, dato un insieme di n eventi E_1, E_2, \dots, E_n , ognuno risultato della singola prova ($E_i \in \mathcal{S}$), per ogni l interi $1 \leq k_1, k_2, \dots, k_l \leq n, l = 1, \dots, n$ si ha che

$$Pr\{E_{k_1}, E_{k_2}, \dots, E_{k_l}\} = Pr\{E_{k_1}\}Pr\{E_{k_2}\} \cdot \dots \cdot Pr\{E_{k_l}\}. \quad (1.47)$$

Esempio 1.1 (cont.): Nell'esperimento della moneta, ricordiamo come lo spazio campione fosse $\mathcal{S} = \{T, C\}$. Se di lanci ne vengono effettuati n in maniera indipendente, si ha che lo spazio campione delle n prove diventa il prodotto cartesiano $\mathcal{S}^n = \underbrace{\mathcal{S} \times \mathcal{S} \times \dots \times \mathcal{S}}_n$. Ad esempio, se

$n = 4$, abbiamo

$$\begin{aligned} \mathcal{S}^4 = \{ & (T, T, T, T), \\ & (T, T, T, C), \\ & (T, T, C, T), \\ & (T, T, C, C), \\ & (T, C, T, T), \\ & (T, C, T, C), \\ & (T, C, C, T), \\ & (T, C, C, C), \\ & (C, T, T, T), \\ & (C, T, T, C), \\ & (C, T, C, T), \end{aligned}$$

$$\begin{aligned}
& (C, T, C, C), \\
& (C, C, T, T), \\
& (C, C, T, C), \\
& (C, C, C, T), \\
& (C, C, C, C) \}. \tag{1.48}
\end{aligned}$$

L'indipendenza probabilistica porta alla fattorizzabilità della probabilità congiunta degli eventi nei vari lanci. Il calcolo delle seguenti probabilità è quindi immediato

$$Pr\{(T, C, C, T)\} = Pr\{T\}Pr\{C\}Pr\{C\}Pr\{T\} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16}; \tag{1.49}$$

$$Pr\{(T, \cdot, C, \cdot)\} = Pr\{T\}Pr\{C\} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}; \tag{1.50}$$

$$Pr\{(T, T, \cdot, \cdot)\} = Pr\{T\}Pr\{T\} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}. \tag{1.51}$$

1.10.1 Prove di Bernoulli

Quando i risultati dell'esperimento base di un insieme di n prove indipendenti appartengono ad uno spazio campione binario, si parla di *prove di Bernoulli*. Se quindi i due eventi elementari dello spazio campione di ogni prova vengono denominati "successo" o "insuccesso", parliamo di *probabilità di successo* p di *probabilità di insuccesso* $q = 1 - p$. E' facile ora andare a valutare la probabilità che in n prove indipendenti si verificano k ($0 \leq k \leq n$) successi. Infatti, poiché la probabilità di una qualunque configurazione di n successi e di $n - k$ insuccessi è

$$p^k(1 - p)^{n-k}, \tag{1.52}$$

sommando su tutte le $\binom{n}{k}$ possibili configurazioni di k successi in n prove, abbiamo che

$$Pr\{k \text{ successi in } n \text{ prove}\} = \binom{n}{k} p^k (1 - p)^{n-k}. \tag{1.53}$$

La quantità trovata è anche detta *legge binomiale* di parametro p e denotata come

$$b(k, n; p) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (1.54)$$

Si noti che

$$\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p+q)^n = 1, \quad (1.55)$$

dove si è fatto uso della formula di Newton $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$, e del fatto che $p = 1 - q$.

Esempio 1.1 (cont.): Nell'esempio della moneta, detta p la probabilità di avere Testa (successo), abbiamo i seguenti risultati

$$Pr\{5 \text{ Teste in } 6 \text{ lanci}\} = \binom{6}{5} p^5 (1-p) = 6p^5 (1-p)$$

$$Pr\{6 \text{ Teste in } 6 \text{ lanci}\} = \binom{6}{6} p^6 (1-p)^0 = p^6$$

$$Pr\{0 \text{ Teste in } 6 \text{ lanci}\} = \binom{6}{0} p^0 (1-p)^6 = (1-p)^6$$

$$\begin{aligned} Pr\{\text{almeno } 4 \text{ Teste in } 6 \text{ lanci}\} &= \sum_{k=4}^6 \binom{6}{k} p^k (1-p)^{6-k} = \binom{6}{4} p^4 (1-p)^2 \\ &+ \binom{6}{5} p^5 (1-p) + \binom{6}{6} p^6 = 15p^4 (1-p)^2 + 6p^5 (1-p) + p^6 = \\ &= 10p^6 - 24p^5 + 15p^4. \end{aligned}$$

$$\begin{aligned} Pr\{\text{al più } 1 \text{ Testa in } 6 \text{ lanci}\} &= \sum_{k=0}^1 \binom{6}{k} p^k (1-p)^{6-k} = \binom{6}{0} p^0 (1-p)^6 + \\ &+ \binom{6}{1} p (1-p)^5 = (1-p)^6 + 6p(1-p)^5. \end{aligned}$$

$$Pr\{\text{almeno } 5 \text{ Croci in } 6 \text{ lanci}\} = Pr\{\text{al più } 1 \text{ Testa in } 6 \text{ lanci}\}.$$

$$Pr\{1 \text{ Croce in } 6 \text{ lanci}\} = Pr\{5 \text{ Teste in } 6 \text{ lanci}\}.$$

Esempio 1.20 Si consideri una sorgente binaria che emette simboli dell'alfabeto binario "1" e "0", ripetutamente con probabilità p e $1-p$. Si supponga che la detta sorgente emetta una sequenza di n bit in maniera indipendente. La probabilità degli eventi associati alla sequenza può essere calcolata facilmente usando la fattorizzabilità della probabilità congiunta derivante dall'ipotesi di indipendenza. I risultati sono del tutto analoghi al caso del lancio della moneta. Si consideri per esempio $n = 6$.

$$Pr\{1, 0, 0, 0, 1, 1\} = p^3(1-p)^3$$

$$Pr\{\text{la stringa contiene 2 bit pari a 1}\} = \binom{6}{2}p^2(1-p)^4$$

$$\begin{aligned} Pr\{\text{la stringa contiene un numero pari di 1}\} &= Pr\{\text{la stringa non contiene alcun 1}\} \\ &+ Pr\{\text{la stringa contiene 2 bit pari a 1}\} \\ &+ Pr\{\text{la stringa contiene 4 bit pari a 1}\} \\ &+ Pr\{\text{la stringa contiene 6 bit pari a 1}\} \\ &= (1-p)^6 + \binom{6}{2}p^2(1-p)^4 + \binom{6}{4}p^4(1-p)^2 + p^6. \end{aligned}$$

Esempio 1.21 Un insieme di n punti viene distribuito aleatoriamente e in maniera uniforme sull'intervallo $[a, b]$. La probabilità che nell'intervallo $[t_1, t_2]$ cadano un certo numero di punti, dove $a \leq t_1 < t_2 \leq b$, può essere valutata usando la distribuzione binomiale, in analogia alle prove ripetute di Bernoulli. Assumendo che

$$p = Pr\{\text{un punto cade nell'intervallo } [t_1, t_2]\} = \frac{t_2 - t_1}{b - a}, \quad (1.56)$$

e che ogni punto viene generato in maniera indipendente, l'evento che quantifica il numero di punti che cadono nell'intervallo è equivalente all'evento che descrive il numero di successi in un esperimento di prove ripetute. Pertanto

$$Pr\{\text{in } n \text{ prove } m \text{ punti cadono in } [t_1, t_2]\} = \binom{n}{m}p^m(1-p)^{n-m}. \quad (1.57)$$

1.10.2 Approssimazioni per la legge binomiale

L'espressione (1.54) della legge binomiale può essere difficile da valutare, numericamente specialmente quando n è un valore elevato e p è piccolo. Il calcolo infatti comporta la valutazione di numeri piccoli (probabilità) che sono il risultato del rapporto di numeri molto grandi (si pensi alla definizione del coefficiente binomiale con i fattoriali). In genere il problema dell'accuratezza numerica in calcoli che coinvolgono quantità combinatorie si presenta anche avendo a disposizione un calcolatore, poiché è comunque necessario concepire algoritmi numericamente affidabili. Fortunatamente, la valutazione della legge binomiale può essere eseguita più

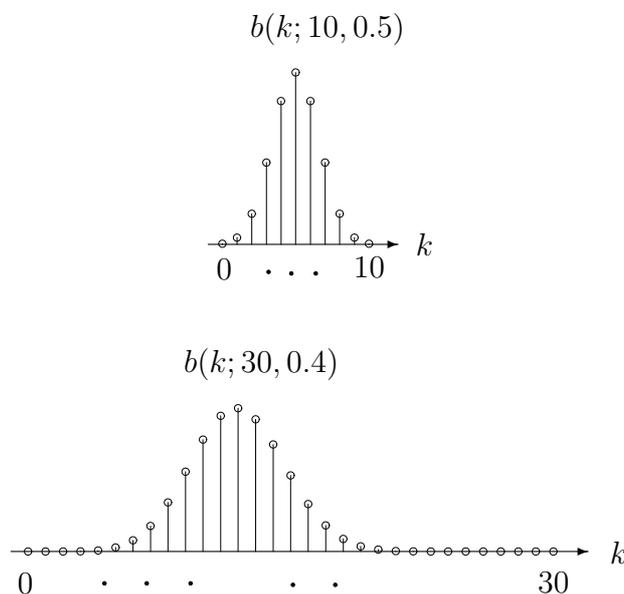


Figura 1.5: La legge binomiale per due valori di n e p .

facilmente facendo ricorso a delle espressioni approssimate. Figura 1.5 mostra l'andamento della distribuzione binomiale per $n = 10$, $p = 0.5$ e per $n = 30$, $p = 0.4$. Si noti come l'andamento sia simmetrico per $p = \frac{1}{2}$ e come, anche per $p \neq \frac{1}{2}$, specialmente per n elevato, l'andamento tenda a diventare simmetrico e sia assimilabile ad una campana. Queste considerazioni hanno portato alle seguenti approssimazioni.

Approssimazione di Poisson

La legge binomiale, quando p è piccola, n è elevato e k è piccolo relativamente a n , può essere approssimata dalla espressione

$$b(k, n; p) \simeq \frac{(np)^k e^{-np}}{k!}. \quad (1.58)$$

Il nome deriva dal fatto che l'espressione approssimante, rispetto al parametro $x = np$, è la cosiddetta distribuzione di Poisson. Tale espres-

sione sarà riconsiderata in un capitolo seguente a proposito delle variabili aleatorie. Il calcolo della approssimazione di Poisson è più semplice del calcolo di $b(k, n; p)$, anche se è comunque necessario calcolare il rapporto tra un esponenziale e un fattoriale.

Dimostrazione: L'approssimazione di Poisson si dimostra usando la formula di Stirling per l'approssimazione al fattoriale

$$n! \simeq (2\pi)^{\frac{1}{2}} n^{n+\frac{1}{2}} e^{-n}, \quad n \gg 1. \quad (1.59)$$

L'errore nella valutazione di $n!$ nella formula di Stirling è dell'ordine di $1/(12n)$ ed è piccolo già per $n = 10$. Si ricordi inoltre che

$$\left(1 - \frac{x}{n}\right)^n \simeq e^{-x}, \quad \text{per } n \gg |x|. \quad (1.60)$$

Ora poiché np è piccola rispetto a n ed $n \gg k$, possiamo scrivere

$$(1-p)^{n-k} = \left(1 - \frac{np}{n}\right)^{n-k} \simeq \left(1 - \frac{np}{n}\right)^n \simeq e^{-np}. \quad (1.61)$$

Quindi la legge binomiale diventa

$$\begin{aligned} b(k, n; p) &= \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \\ &\simeq \frac{(2\pi)^{\frac{1}{2}} n^{n+\frac{1}{2}} e^{-n}}{(2\pi)^{\frac{1}{2}} (n-k)^{n-k+\frac{1}{2}} e^{-(n-k)} k!} \frac{(np)^k}{n^k} e^{-np} \\ &= \left(\frac{n-k}{n}\right)^{-(n-k+\frac{1}{2})} \frac{1}{k!} e^{-k} (np)^k e^{-np}. \end{aligned} \quad (1.62)$$

Applicando di nuovo la (1.60), abbiamo

$$\left(\frac{n-k}{n}\right)^{(n-k+\frac{1}{2})} = \left(1 - \frac{k}{n}\right)^{(n-k+\frac{1}{2})} \simeq e^{-k}, \quad (1.63)$$

che sostituita nell'espressione precedente ci fornisce il risultato cercato.

Teorema di De Moivre-Laplace

Un'altra utile espressione per approssimare la legge binomiale è

$$b(k, n; p) \simeq \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}, \quad q = 1-p, \quad (1.64)$$

dove si è assunto che n sia grande e che $npq \gg 1$ e $|k - np| < L\sqrt{\mu pq}$ con $L \simeq 2,34$, ovvero si è abbastanza vicini alla media. Questa formula, che è nota come *teorema di De Moivre-Laplace*, non richiede il calcolo di alcun fattoriale - a differenza della approssimazione di Poisson. Il teorema di De Moivre-Laplace mostra come la distribuzione binomiale per n sufficientemente grande abbia un andamento simile alla *funzione di distribuzione gaussiana*

$$\mathcal{N}(x; \mu, \sigma) \triangleq \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (1.65)$$

con $x = k$, $\mu = np$, $\sigma^2 = npq$. Il legame è utile sia dal punto di vista teorico che computazionale. Infatti, la dimostrazione della legge dei grandi numeri, che seguirà, si basa su tale approssimazione. Mediante il teorema di De Moivre-Laplace il calcolo di probabilità del tipo

$$Pr\{k_1 \leq k \leq k_2\} = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k (1-p)^{n-k}, \quad (1.66)$$

può essere facilmente approssimato come

$$Pr\{k_1 \leq k \leq k_2\} \simeq \int_{k_1}^{k_2} \mathcal{N}(x; np, \sqrt{npq}) dx, \quad (1.67)$$

in quanto

$$\begin{aligned} Pr\{k_1 \leq k \leq k_2\} &= \sum_{k=k_1}^{k_2} b(k, n, p) \simeq \sum_{k=k_1}^{k_2} \mathcal{N}(k; np, \sqrt{npq}) \\ &\simeq \sum_{k=k_1}^{k_2} \int_k^{k+1} \mathcal{N}(x; np, \sqrt{npq}) dx \\ &\simeq \int_{k_1}^{k_2} \mathcal{N}(x; np, \sqrt{npq}) dx, \end{aligned} \quad (1.68)$$

dove si è supposto che l'area sottesa nell'intervallo $[k, k+1]$ sia circa pari a $\mathcal{N}(x; np, \sqrt{npq})$, con la funzione quasi costante. L'integrale in (1.67) si valuta riscrivendo

$$\int_{k_1}^{k_2} \mathcal{N}(x; np, \sqrt{npq}) dx = \phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \phi\left(\frac{k_1 - np}{\sqrt{npq}}\right), \quad (1.69)$$

dove

$$\phi(z) = \int_{-\infty}^z \mathcal{N}(x; 0, 1) dx. \quad (1.70)$$

I valori di ϕ , o di altre funzioni ad essa collegate, quali le funzioni *erf*, o *erfc*, sono tabellati o valutati numericamente (vedi Appendice A). Maggiori dettagli sull'uso della gaussiana saranno forniti nel capitolo sulla variabile aleatoria.

La prova della formula di De Moivre-Laplace è piuttosto tecnica e si preferisce rimandare il lettore al testo di Feller (1968) per la dimostrazione completa. Si noti comunque come nella funzione gaussiana, il valore μ , nel nostro caso pari a np , rappresenti l'ascissa centrale della funzione (media) e come il valore σ^2 , nel nostro caso pari a npq , ne determini la larghezza (varianza).

Esempio 1.22 In una stringa binaria di lunghezza $n = 32$, i bit assumono in maniera indipendente il valore 1 o 0, con probabilità rispettivamente $p = 0.2$ e $q = 1 - p = 0.8$. Valutiamo la probabilità che la stringa contenga: a. esattamente 10 bit pari a 1; b. che la stringa contenga un numero di bit pari a 1 compreso tra 10 e 20. Lo scopo dell'esempio è evidenziare la difficoltà computazionale della legge binomiale e l'utilità delle approssimazioni di Poisson e di De Moivre-Laplace.

Soluzione:

a.

$$Pr\{k = 10\} = \binom{32}{10} (0.2)^{10} (0.8)^{22} = \frac{32!}{22!10!} (0.2)^{10} (0.8)^{22} = 0.0487. \quad (1.71)$$

Usando l'approssimazione di Poisson abbiamo

$$Pr\{k = 10\} \simeq \frac{(32 \cdot 0.2)^{10} e^{-32 \cdot 0.2}}{10!} = \frac{6.4^{10} e^{-6.4}}{10!} = 0.0528. \quad (1.72)$$

L'approssimazione di De Moivre-Laplace è invece

$$Pr\{k = 10\} \simeq \frac{1}{\sqrt{2\pi \cdot 32 \cdot 0.2 \cdot 0.8}} e^{-\frac{(10 - 32 \cdot 0.2)^2}{2 \cdot 32 \cdot 0.2 \cdot 0.8}} = \frac{1}{\sqrt{2\pi \cdot 5.12}} e^{-\frac{3.6^2}{10.24}} = 0.0497. \quad (1.73)$$

La difficoltà nel calcolo dei fattoriali e delle potenze è evidente. Anche per la formula della binomiale esatta, i valori riportati sono la nostra

migliore stima ottenuta con Matlab in doppia precisione. L'approssimazione gaussiana fornisce sicuramente lo strumento più semplice per ottenere un rapido risultato.

b.

$$Pr\{10 \leq k \leq 20\} = \sum_{i=10}^{20} \binom{32}{i} (0.2)^i (0.8)^{32-i}. \quad (1.74)$$

Usando la approssimazione di Poisson abbiamo

$$Pr\{10 \leq k \leq 20\} \simeq \sum_{i=10}^{20} \frac{(32 \cdot 0.2)^i e^{-32 \cdot 0.2}}{i!}. \quad (1.75)$$

Il calcolo ripetuto delle potenze e dei fattoriali richiederebbe un piccolo algoritmo che tralasciamo di implementare per brevità e che comunque potrebbe dare problemi di precisione. L'approssimazione gaussiana fornisce invece semplicemente

$$\begin{aligned} Pr\{10 \leq k \leq 20\} &\simeq \Phi\left(\frac{20 - 32 \cdot 0.2}{\sqrt{32 \cdot 0.2 \cdot 0.8}}\right) - \Phi\left(\frac{10 - 32 \cdot 0.2}{\sqrt{32 \cdot 0.2 \cdot 0.8}}\right) \\ &= \Phi(6.0104) - \Phi(1.5910) \\ &\simeq 1 - 0.9441 = 0.0559, \end{aligned} \quad (1.76)$$

dove i valori delle $\Phi()$ sono stati ottenuti dalla tavola riportata in Appendice A. Si noti che il valore $\Phi(6.0104)$ è maggiore di 0.9999, e che non essendo disponibile nella tavole, viene considerato pari a 1. Il valore piccolo della probabilità è dovuto al fatto che il centro della distribuzione è attorno a 6.4 e che i valori $k = 10$ e $k = 20$ si trovano nella coda destra della distribuzione.

1.11 La legge dei grandi numeri

Per concludere questo capitolo, riportiamo di seguito uno dei risultati fondamentali della teoria della probabilità che serve a collegare in maniera sistematica le misure statistiche sui risultati di un esperimento con la definizione stessa di probabilità. Abbiamo introdotto il concetto di probabilità all'inizio del capitolo seguendo l'idea che essa costituisca un invariante rispetto alla variabilità riscontrata negli esperimenti. La

legge dei grandi numeri chiarisce questa questione dimostrando come, al divergere della dimensione di un campione in cui si verifichi l'occorrenza di un evento, la frequenza relativa si avvicini alla probabilità supposta.

Teorema 1.1 *Si consideri un esperimento E e sia A un evento ad esso collegato con probabilità $p = Pr\{A\}$. Su n prove indipendenti dello stesso esperimento, sia k il numero di volte in cui A si è verificato. Vale il seguente risultato*

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} Pr \left\{ \left| \frac{k}{n} - p \right| \leq \epsilon \right\} = 1. \quad (1.77)$$

Il teorema, noto come *legge debole dei grandi numeri*, dimostra come la frequenza relativa di un evento tenda in probabilità a $Pr\{A\}$; ovvero, il numero percentuale di occorrenze dell'evento tende ad avvicinarsi al suo valore teorico, al tendere all'infinito del numero di prove. La convergenza in probabilità significa che in pratica all'evento

$$\left\{ \left| \frac{k}{n} - p \right| \leq \epsilon \right\}, \quad (1.78)$$

compete una probabilità sempre più vicina a uno al crescere di n . In altre parole, per ϵ sufficientemente piccolo e n sufficientemente grande

$$Pr\{A\} \simeq p. \quad (1.79)$$

Prova: La prova più semplice della legge debole dei grandi numeri si basa sul teorema di De Moivre-Laplace. Si consideri l'esperimento in cui l'occorrenza dell'evento A sia associata a un "successo" e che la probabilità di tale successo sia p . Sia assegnato un valore $\epsilon > 0$, piccolo a piacere e si considerino i due numeri, $k_1 = n(p - \epsilon)$ e $k_2 = n(p + \epsilon)$. Sia k un intero tale che

$$k_1 \leq k \leq k_2. \quad (1.80)$$

Dividendo per n e sottraendo p si ottiene

$$-\epsilon \leq \frac{k}{n} - p \leq \epsilon. \quad (1.81)$$

Le due disuguaglianze esprimono lo stesso evento

$$\{k_1 \leq k \leq k_2\} = \left\{ \left| \frac{k}{n} - p \right| \leq \epsilon \right\}, \quad (1.82)$$

la cui probabilità è

$$\begin{aligned}
 Pr \left\{ \left| \frac{k}{n} - p \right| \leq \epsilon \right\} &= Pr \{k_2 \leq k \leq k_1\} \\
 &\simeq \phi \left(\frac{k_2 - np}{\sqrt{npq}} \right) - \phi \left(\frac{k_1 - np}{\sqrt{npq}} \right) \\
 &= \phi \left(\frac{\epsilon n}{\sqrt{npq}} \right) - \phi \left(\frac{-\epsilon n}{\sqrt{npq}} \right) \\
 &= 2\phi \left(\frac{\epsilon n}{\sqrt{npq}} \right) - 1, \tag{1.83}
 \end{aligned}$$

dove abbiamo usato il teorema di De Moivre-Laplace. Poiché per ogni ϵ finito

$$\lim_{n \rightarrow \infty} \frac{\epsilon n}{\sqrt{npq}} = \infty, \tag{1.84}$$

si ha che

$$\lim_{n \rightarrow \infty} \phi \left(\frac{\epsilon n}{\sqrt{npq}} \right) = 1. \tag{1.85}$$

Pertanto il teorema è dimostrato.

Esistono diverse estensioni della legge dei grandi numeri. E' possibile infatti dimostrare che la convergenza è anche più forte che la convergenza in probabilità, ovvero che la frequenza relativa converge con probabilità uno. Tale estensione detta *legge forte dei grandi numeri*, dovuta a Borel, viene tralasciata per brevità. Si preferisce rimandare il lettore ai testi classici sull'argomento per gli approfondimenti (ad esempio Gnedenko, 1962).

1.12 Problemi

Problema 1.1 Si consideri il lancio di due monete truccate con spazi campione $\mathcal{S}_1 = \{T, C\}$, $\mathcal{S}_2 = \{T, C\}$ e distribuzioni $\Pi_1 = \{0.4, 0.6\}$, $\Pi_2 = \{0.1, 0.9\}$.

- (a) Calcolare la probabilità che si ottengano due Teste;
- (b) Calcolare la probabilità che si ottenga almeno una Testa.

Problema 1.2 Si consideri il lancio di due dadi truccati con spazi campione $\mathcal{S}_1 = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{S}_2 = \{1, 2, 3, 4, 5, 6\}$ e distribuzioni $\Pi_1 = \Pi_2 = \{0.1, 0.1, 0.1, 0.2, 0.4, 0.1\}$.

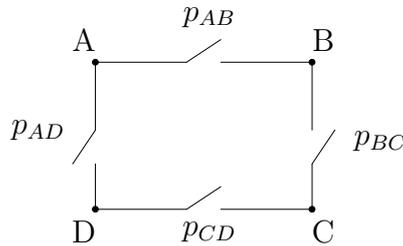
- (a) Calcolare la probabilità che si ottengano due sei;
 (b) Calcolare la probabilità che si ottenga almeno una somma pari a cinque.

Problema 1.3 Si consideri il lancio di due dadi truccati, ognuno avente spazio campione $S = \{1, 2, 3, 4, 5, 6\}$ e probabilità $\Pi = \{\frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{7}{12}\}$. Si valuti la probabilità che il risultato sia ≥ 10 .

Problema 1.4 Si consideri il lancio di tre monete non truccate. Calcolare la probabilità di avere almeno due teste.

Problema 1.5 Una fabbrica ha prodotto 10 calcolatori di cui due sono certamente difettosi, ma non sa quali. Calcolare la probabilità che provando solo 3 calcolatori si riescano a individuare entrambe i calcolatori difettosi.

Problema 1.6 Una rete di quattro calcolatori (A,B,C,D) è collegata secondo la seguente configurazione ad anello. Due calcolatori sono



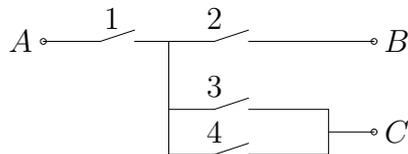
collegati se esiste almeno un percorso chiuso tra di essi. I collegamenti sui vari rami sono interrotti in maniera indipendente e con probabilità $p_{AB} = p_{BC} = p_{CD} = p_{DA} = 0.1$, come descritto in figura. Calcolare (a)

- $Pr\{B \text{ collegato a } D\}$;
 (b) $Pr\{B \text{ collegato a } C\}$;
 (c) $Pr\{\text{tutti i calcolatori sono connessi}\}$;
 (suggerimento: elencare tutte le possibili configurazioni)

Problema 1.7 L'accesso a un provider Internet è inaffidabile perché nell'1% dei tentativi la richiesta di connessione viene rifiutata. Calcolare:

- (a) la probabilità che per ottenere la connessione si debba provare esattamente tre volte;
- (b) la probabilità che per ottenere la connessione si debba tentare più di due volte.

Problema 1.8 Nel circuito di figura i quattro interruttori operano in maniera indipendente e sono aperti con probabilità $\{p_i\}_{i=1}^4$. Calcolare:



- (a) $Pr\{A \text{ e } B \text{ sono connessi}\}$;
- (b) $Pr\{B \text{ e } C \text{ sono connessi}\}$;
- (b) $Pr\{A, B \text{ e } C \text{ sono connessi}\}$;

Problema 1.9 La fabbrica A produce telefonini difettosi nella percentuale dell'1%, mentre la fabbrica B solo nella percentuale dello 0.5%. Il mercato dei telefonini è ripartito tra i due produttori nella misura del 30% per A e 70% per B. Calcolare la probabilità che un telefonino acquistato e trovato difettoso sia stato prodotto dalla fabbrica A.

Problema 1.10 Un dattilografo sbaglia in media un carattere ogni 100. Calcolare la probabilità che su una lettera di 1000 caratteri scritta da tale dattilografo ci siano più di due errori.

Problema 1.11 Un collegamento cellulare è inaffidabile perché il 10% delle conversazioni in media viene interrotta prima di essere terminata. Se 10 utenti sono collegati, calcolare:

- (a) la probabilità che nessun utente porti a termine la conversazione;
- (b) la probabilità che almeno un utente porti a termine la conversazione

Problema 1.12 La stringa binaria (b_1, b_2, b_3) assume valori nell'insieme

$$\{000, 001, 010, 011, 100, 101, 110, 111\}, \quad (1.86)$$

con probabilità

$$\{0.01, 0.03, 0.2, 0.5, 0.1, 0.1, 0.03, 0.03\}. \quad (1.87)$$

- (a) Valutare $Pr\{b_1 = 1\}$
- (b) Valutare $Pr\{b_2 = 1\}$
- (c) b_1 e b_2 sono indipendenti? spiegare perché.

Capitolo 2

La variabile aleatoria

In questo capitolo, dopo una introduzione con degli esempi volti a spiegare le motivazioni che hanno indotto alla formalizzazione della variabile aleatoria, viene introdotta la definizione formale. In particolare vengono presentate la variabile aleatoria continua e discreta monodimensionale. Vengono introdotte la funzione di distribuzione cumulativa e la funzione di densità. Vari esempi di distribuzioni notevoli sono presentati tra cui la variabile aleatoria Gaussiana. Vengono inoltre introdotti i momenti della variabile aleatoria e vari esempi.

2.1 Introduzione

Nella pratica ingegneristica e nello studio di fenomeni fisici, siano essi di natura tecnica, scientifica, economica, o altro, è molto comune dover *misurare*, o modellare una variabile fisica. L'imperfezione degli strumenti di analisi, o la parziale imprevedibilità del fenomeno, impongono che la formalizzazione matematica preveda che l'incertezza venga tenuta in conto e opportunamente descritta. La formalizzazione si focalizza su fatto che la variabile assume valori definiti su uno spazio campione ordinato (l'asse reale o sottoinsiemi di esso) con probabilità diverse. Per meglio motivare l'introduzione degli strumenti analitici che saranno discussi in seguito, cominciamo con un esempio.

Esempio 2.1 Si supponga di volere quantificare l'altezza in centimetri

degli studenti del corso di Teoria dei Fenomeni Aleatori. Si tratta ovviamente di una quantità fisica non costante: ci sono studenti alti, bassi, alcuni molto alti e altri decisamente bassi. Si decide pertanto di elaborare una piccola statistica. Per prima cosa, anche se l'altezza è associabile ad un numero reale, decidiamo di discretizzare i valori in intervalli di 5 cm. Rileviamo la tabella 2.1, che per migliore visibilità viene tradotta nel diagramma a barre di figura 2.1. L'altezza di ogni barra è proporzionale al numero di persone trovate in quell'intervallo. Il diagramma si chiama *istogramma* e i vari intervalli *bin*. Quest'ultimo termine proviene dall'inglese "bin", che significa "contenitore": l'istogramma rappresenta i livelli di riempimento dei vari bin. La figura mostra come i risultati si concentrino prevalentemente attorno ai valori 155-180 e che il diagramma rappresenta una funzione di "densità." Si noti come la tendenza dei dati rilevati sia di concentrarsi al centro con un solo picco piuttosto smussato. La figura mostra anche una possibile curva di densità ottenuta dalla interpolazione dei dati discretizzati. Nella figura 2.2 sono riportati gli stessi dati in un formato un pò diverso: ad ogni intervallo è associata una barra che ha una altezza proporzionale al numero di persone che hanno una altezza *minore o uguale* al valore in ascissa. Il diagramma è detto *istogramma cumulativo*. Sulla stessa figura è anche tracciata una curva interpolante che rappresenta la funzione di densità cumulativa.

Questo esempio serve a introdurre la formalizzazione della variabile aleatoria che viene caratterizzata mediante una funzione di densità, o con una funzione cumulativa di densità, come quelle mostrate nelle due figure. Le due funzioni tengono conto in maniera compatta come la probabilità si distribuisce sui vari valori della grandezza fisica in oggetto.

I dettagli relativi alle importanti differenze tra gli istogrammi e le funzioni di densità e cumulativa, hanno a che fare con le differenze esistenti tra la statistica e la probabilità e saranno discusse in maggiore dettaglio in uno dei capitoli seguenti.

2.2 Definizione di variabile aleatoria

Si supponga di considerare un esperimento il cui risultato sia un valore numerico. Tale esperimento può essere modellato come in figura 2.3, ovvero come la associazione di ogni elemento ξ di uno spazio campione \mathcal{S}

Altezza (cm)	numero	cumulativo
140 – 144	1	1
145 – 149	0	1
155 – 159	2	3
160 – 164	10	13
165 – 169	13	26
170 – 174	24	50
175 – 179	12	62
180 – 184	7	69
185 – 189	3	72
190 – 194	0	72
195 – 199	1	73
200 – 204	0	73
205 – 209	1	74
210 – 214	0	74
215 – 219	0	74

Tabella 2.1: I risultati della rilevazione delle altezze.

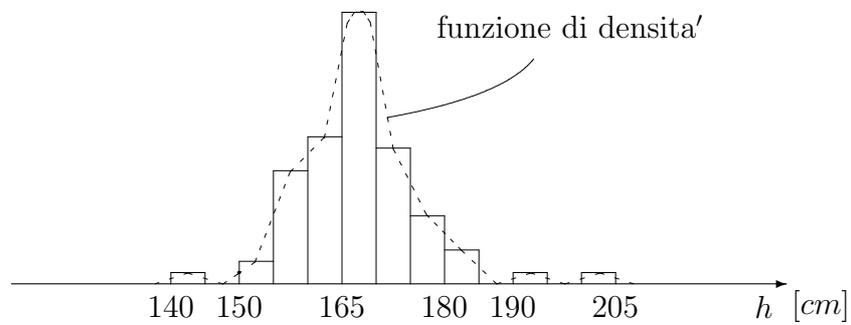


Figura 2.1: L'istogramma della rilevazione delle altezze e la relativa funzione di densità

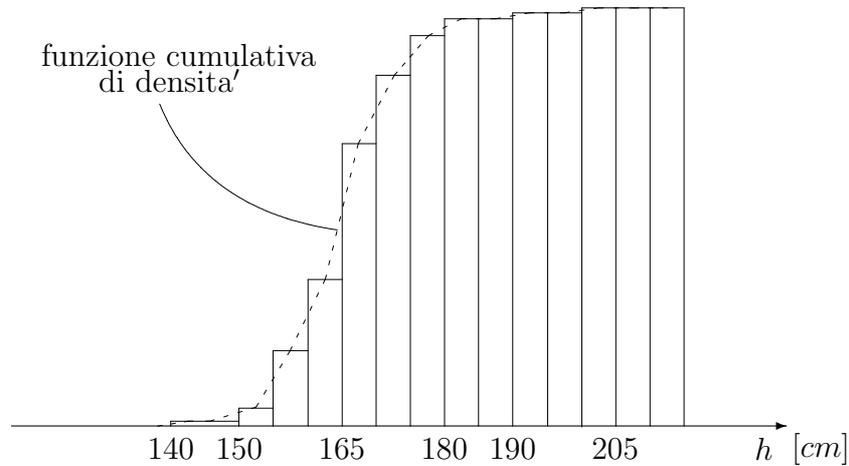


Figura 2.2: L'istogramma cumulativo della rilevazione delle altezze e la relativa funzione di densità cumulativa

a un valore sull'asse ordinato dei numeri reali. I valori possibili sono tutti i numeri reali o sottoinsiemi continui o discreti di essi. Tale associazione è detta *variabile aleatoria*. Più formalmente,

Dato uno spazio campione \mathcal{S} , a cui è associata una misura di probabilità Π , si definisce variabile aleatoria la funzione X che associa ad ogni elemento dello spazio campione $\xi \in \mathcal{S}$ un numero reale:

$$X : \xi \in \mathcal{S} \longleftrightarrow x \in \mathcal{A} \subseteq \mathcal{R}. \quad (2.1)$$

Si noti come la variabile aleatoria (che è una funzione) si denoti con la lettera maiuscola X , mentre il suo valore x si denoti con la lettera minuscola. Useremo talvolta anche la notazione $X(\xi)$. L'aspetto cruciale della definizione della variabile aleatoria è che l'insieme di definizione \mathcal{A} , che è un sottoinsieme dell'asse reale, è un insieme *ordinato*. Ciò consente di definire relazioni d'ordine tra valori della variabile aleatoria, cosa che non sarebbe facilmente formalizzabile direttamente nello spazio campione.

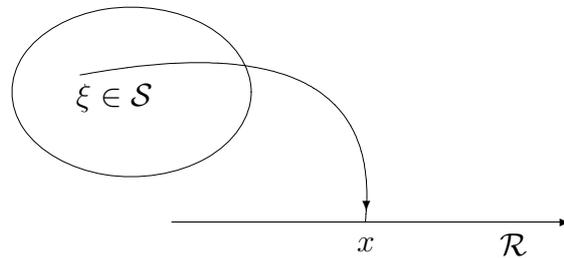


Figura 2.3: La definizione di variabile aleatoria

Se l'insieme di definizione \mathcal{A} è un insieme continuo, o discreto, parleremo rispettivamente di *variabile aleatoria continua*, o *discreta*. Come vedremo in seguito, è anche possibile definire variabili aleatorie miste, ovvero caratterizzate da insiemi di definizione formati dall'unione di sottoinsiemi continui e discreti.

Ci sono moltissime situazioni di interesse fisico-ingegneristico dove la formalizzazione della variabile aleatoria risulta molto utile. Vediamo alcuni esempi.

Esempio 2.2 Si riprenda l'esperimento del dado a sei facce del capitolo precedente. Lo spazio campione è $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. Agli elementi dello spazio campione è stata associata la variabile a sei valori $X \in \mathcal{A} = \{1, 2, 3, 4, 5, 6\}$. I valori ordinati sull'asse reale sono rappresentati in figura 2.4. Si tratta di una variabile aleatoria discreta.

Esempio 2.3 Si riprenda l'esperimento del lancio della moneta. Lo spazio campione è $\mathcal{S} = \{T, C\}$. Ai due risultati dell'esperimento si associa la variabile $X \in \mathcal{A} = \{0, 1\}$ i cui valori sono rappresentati in figura 2.5. Si tratta di una variabile aleatoria discreta.

Esempio 2.4 Si consideri la variabile aleatoria T_0 che descrive la durata in secondi di una conversazione telefonica. L'insieme di definizione \mathcal{A} è l'insieme dei numeri reali positivi come mostrato in figura 2.6. La variabile aleatoria è continua.

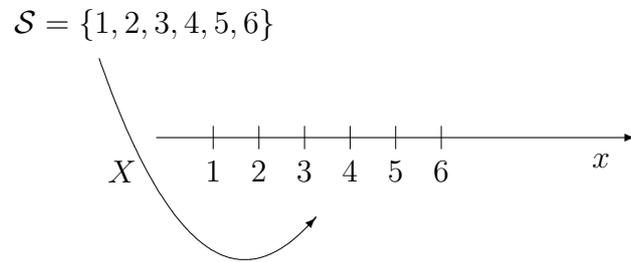


Figura 2.4: La variabile aleatoria associata all'esperimento del dado

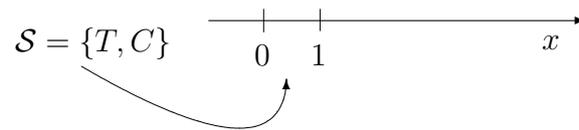


Figura 2.5: La variabile aleatoria associata all'esperimento della moneta

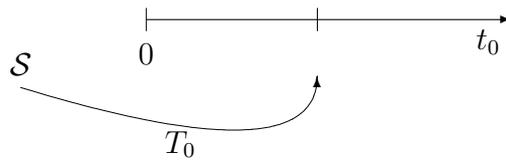


Figura 2.6: La variabile aleatoria associata alla durata di una conversazione telefonica

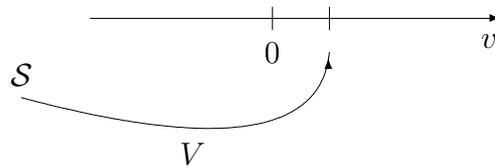


Figura 2.7: La variabile aleatoria associata alla tensione misurata ai capi di un resistore rumoroso

Esempio 2.5 Si supponga di misurare in un certo istante di tempo la tensione V ai capi di un resistore rumoroso. Tale valore, misurato in Volt, sarà un numero reale che può essere sia negativo che positivo. Pertanto $V \in \mathcal{A} = \mathcal{R}$. La variabile aleatoria è continua e il suo insieme di definizione è mostrato in figura 2.7.

La descrizione probabilistica della variabile aleatoria deve consentirci di quantificare il modo in cui i vari risultati dell'esperimento si addensano sull'insieme di definizione. Figura 2.8 mostra alcune realizzazioni di una variabile aleatoria e come esse si siano distribuite non uniformemente sull'insieme di definizione. La formalizzazione probabilistica deve consentire di misurare tale distribuzione e valutare la probabilità di eventi del tipo $\{x_1 < X \leq x_2\}$, $\{X \leq x_2\}$, $\{x_1 < X\}$, $\{X > 0\}$, $\{X < +\infty\}$, $\{X = x_1\}$, con $x_1, x_2 \in \mathcal{A}$.

2.3 La funzione di distribuzione cumulativa

Si definisce *funzione di distribuzione cumulativa*, o *cdf* (*cumulative distribution function*)

$$F_X(x) \triangleq Pr\{X \leq x\}, \quad \forall \quad -\infty < x < \infty \quad . \quad (2.2)$$

Ovvero, dato un valore $x \in \mathcal{A}$, la cdf esprime la probabilità che il risultato dell'esperimento sia un numero minore o uguale a x . Anche se tale

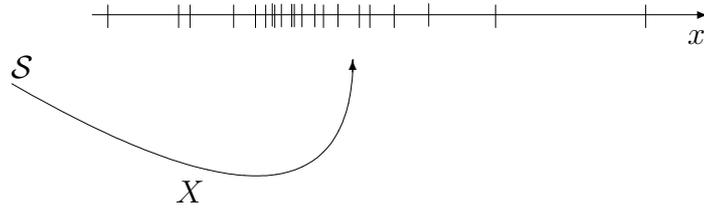


Figura 2.8: Alcune realizzazioni di un variabile aleatoria

definizione può risultare in primo approccio un pò artificiosa, mostreremo come essa consenta una compatta e rigorosa descrizione di come la variabile aleatoria si distribuisce su \mathcal{A} . Si pensi all'esempio della distribuzione delle altezze presentato all'inizio del capitolo. Esaminiamo prima alcune importanti proprietà della funzione di distribuzione.

Proprietà 2.1 $F_X(x) \geq 0, \forall x$ ed è tale che

$$F_X(-\infty) = 0, \quad F_X(+\infty) = 1. \quad (2.3)$$

Prova: La proprietà di positività deriva dal primo assioma della probabilità. Inoltre $F_X(+\infty) = Pr\{X \leq +\infty\} = 1$, è la probabilità dell'evento certo. Analogamente, $F_X(-\infty) = Pr\{X \leq -\infty\} = 0$, è la probabilità dell'evento impossibile.

Proprietà 2.2 La cdf è una funzione non decrescente, ovvero:

$$F_X(x_1) \leq F_X(x_2), \quad \text{se } x_1 < x_2. \quad (2.4)$$

Prova: Con $x_1 < x_2$, abbiamo che

$$\{X \leq x_1\} \subseteq \{X \leq x_2\}. \quad (2.5)$$

Da Proprietà 1.6 abbiamo il risultato.

Proprietà 2.3 La cdf è continua a destra, ovvero

$$F_X(x^+) = F_X(x). \quad (2.6)$$

Prova: Si noti che $\forall \epsilon > 0$,

$$\{X \leq x + \epsilon\} = \{X \leq x\} \cup \{x < X \leq x + \epsilon\}. \quad (2.7)$$

Poichè gli eventi sono disgiunti,

$$Pr\{X \leq x + \epsilon\} = Pr\{X \leq x\} + Pr\{x < X \leq x + \epsilon\}. \quad (2.8)$$

Andando al limite con $\epsilon \rightarrow 0$ per valori positivi, abbiamo che

$$\lim_{\epsilon \rightarrow 0^+} Pr\{X \leq x + \epsilon\} = F_X(x^+) = F_X(x) + 0, \quad (2.9)$$

poiché l'insieme $\{x < X \leq x + \epsilon\}$ tende all'insieme vuoto.¹

Proprietà 2.4

$$Pr\{x_1 < X \leq x_2\} = F_X(x_2) - F_X(x_1), \quad \forall x_1 < x_2. \quad (2.10)$$

Questa proprietà è di estrema utilità e giustifica anche la nomenclatura di funzione *cumulativa*. Infatti la conoscenza della funzione consente il calcolo immediato, per differenza, della probabilità che la variabile sia contenuta in un certo intervallo.

Prova: Con $x_1 < x_2$, è possibile decomporre l'evento $\{X \leq x_2\}$ nella seguente unione di eventi disgiunti

$$\{X \leq x_2\} = \{X \leq x_1\} \cup \{x_1 < X \leq x_2\}. \quad (2.11)$$

Dal terzo assioma e dalla definizione di cdf abbiamo che

$$F_X(x_2) = F_X(x_1) + Pr\{x_1 < X \leq x_2\}, \quad (2.12)$$

che dimostra la proprietà.

Proprietà 2.5

$$Pr\{X = x\} = F_X(x) - F_X(x^-). \quad (2.13)$$

¹In effetti per usare i limiti sulle probabilità abbiamo implicitamente utilizzato il terzo assioma modificato. Per maggior dettagli, vedi (Papoulis, 1965).

Prova: Si consideri l'evento $\{x - \epsilon < X \leq x\}$ con $\epsilon > 0$. Dalla proprietà 2.4, abbiamo che

$$Pr\{x - \epsilon < X \leq x\} = F_X(x) - F_X(x - \epsilon). \quad (2.14)$$

Facendo tendere a zero ϵ otteniamo il risultato.

Una conseguenza notevole dell'ultima proprietà è che per ogni cdf continua in x ,

$$Pr\{X = x\} = 0. \quad (2.15)$$

Tale proprietà è di fondamentale importanza e formalizza in maniera rigorosa come in una variabile aleatoria di tipo continuo, la struttura densa dell'asse reale porti ad una probabilità nulla per ogni punto dell'insieme di definizione. Invece, se in un punto $x = x_0$ la cdf ha una discontinuità di prima specie,

$$Pr\{X = x_0\} = F_X(x_0) - F_X(x_0^-) = F_X(x_0^+) - F_X(x_0^-). \quad (2.16)$$

Pertanto l'entità del salto è proprio pari alla probabilità che la variabile assuma quel valore. Ci si ricordi che la cdf è comunque sempre continua a destra. Vediamo ora alcuni esempi.

Esempio 2.3: Torniamo all'esempio della moneta con spazio campione $\mathcal{S} = \{T, C\}$ a cui viene associata la variabile $X \in \mathcal{A} = \{0, 1\}$. Supponiamo che la distribuzione di probabilità su tali valori sia $\Pi = \{\frac{1}{2}, \frac{1}{2}\}$. Figura 2.9 mostra l'andamento della cdf. Si notino i salti di $\frac{1}{2}$ a $x = 0$ e $x = 1$.

Esempio 2.2: Riconsideriamo l'esempio del dado a sei facce in cui lo spazio campione è $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. La variabile aleatoria è $X \in \mathcal{A} = \{1, 2, 3, 4, 5, 6\}$. Supponiamo che la distribuzione di probabilità su tali valori sia uniforme, ovvero $\Pi = \{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}$. Figura 2.10 mostra l'andamento della cdf. Si notino i salti di $\frac{1}{6}$ a $x = 1, 2, 3, 4, 5, 6$. Calcoliamo ora la probabilità che il risultato dell'esperimento sia compreso in un certo intervallo usando la cdf. Ad esempio: $Pr\{\text{il risultato è maggiore o uguale a } 3\} = Pr\{3 \leq X < \infty\} = Pr\{3 \leq X \leq 6\} = Pr\{2 < X \leq 6\} = F_X(6) - F_X(2) = 1 - \frac{2}{6} = \frac{2}{3}$. Si noti come nel caso di variabili aleatorie discrete particolare attenzione vada riservata alla inclusione o alla esclusione degli estremi.

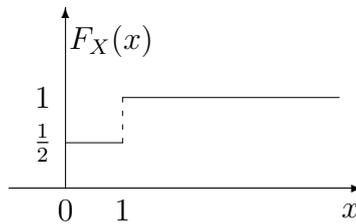


Figura 2.9: La cdf della variabile aleatoria associata all'esperimento della moneta

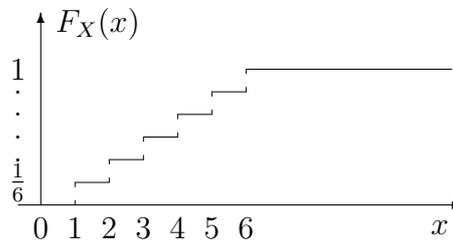


Figura 2.10: La cdf della variabile aleatoria associata all'esperimento del dado

Esempio 2.4: Torniamo all'esempio della conversazione telefonica. Si assuma che l'andamento della cdf sia quello riportato in figura 2.11. Si noti come la cdf sia continua e che, poiché i valori ammissibili della variabile aleatoria sono solo positivi, la funzione è identicamente nulla per valori negativi di x . Dalla proprietà 2.4 si ha che: $Pr\{\text{la conversazione dura un tempo compreso tra 10 e 60 secondi}\} = Pr\{10 < T_0 \leq 60\} = F_{T_0}(60) - F_{T_0}(10)$. Oppure: $Pr\{\text{la conversazione dura più di 30 secondi}\} = Pr\{T_0 > 30\} = Pr\{30 < T_0 \leq +\infty\} = 1 - F_{T_0}(30)$.

Esempio 2.5: Consideriamo l'esempio della tensione ai capi del resistore rumoroso. Si assuma che l'andamento della cdf sia quello riportato in figura 2.12. Si noti come la cdf sia continua e che poiché i valori ammissibili sono tutti i valori reali, la funzione ha valori diversi da zero

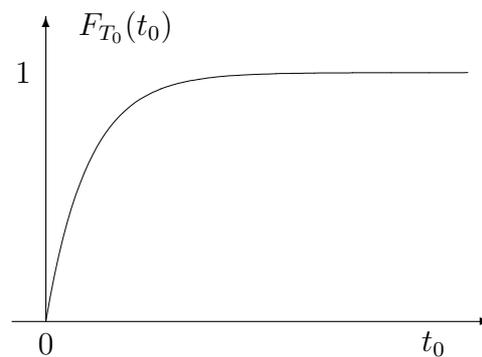


Figura 2.11: La cdf della durata della conversazione telefonica T_0

su tutto l'asse reale. La probabilità che il valore della tensione sia compreso in un determinato intervallo si calcola usando le proprietà come nell'esempio precedente. Ovvero: $Pr\{\text{la tensione è compresa tra } v_1 \text{ e } v_2\} = Pr\{v_1 < V \leq v_2\} = F_V(v_2) - F_V(v_1)$. Si noti come la inclusione o la esclusione degli estremi, in questo caso di variabile aleatoria continua, sia irrilevante ai fini del risultato finale.

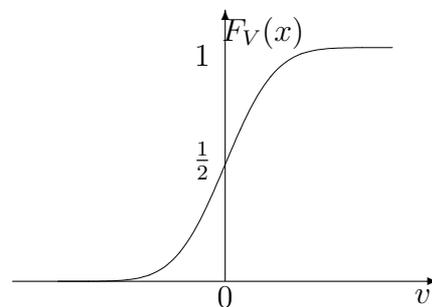


Figura 2.12: La cdf della tensione V ai capi del resistore rumoroso

2.4 La funzione di densità di probabilità

La definizione della cdf di una variabile aleatoria consente una prima descrizione di come i valori della variabile si distribuiscono sull'asse reale. Abbiamo visto dagli esempi come la stessa definizione si applichi sia a variabili aleatorie continue che discrete. Quest'ultima caratteristica la rende particolarmente interessante, anche se a volte non si intuisce solo dal grafico della cdf come le realizzazioni della variabile aleatoria si addensano sull'asse reale. La definizione di funzione di densità di probabilità serve proprio a fornire una descrizione più immediata di ciò, anche se qualche difficoltà formale in più emerge rispetto alla cdf, specialmente nel caso delle variabili aleatorie di tipo discreto. Vedremo in seguito come tali problemi vengano facilmente superati mediante l'utilizzo della funzione generalizzata di Dirac.

Si definisce *funzione di densità di probabilità* o *pdf* (*probability density function*) di una variabile aleatoria X , la derivata della funzione di distribuzione cumulativa

$$f_X(x) \triangleq \frac{d}{dx} F_X(x). \quad (2.17)$$

Si noti che $F_X(x)$ potrebbe essere discontinua e non derivabile. A differenza dell'esempio 2.5 in cui la cdf è continua e derivabile, negli esempi 2.3 e 2.2 abbiamo delle cdf che presentano delle discontinuità e nell'esempio 2.4, la cdf è continua, ma con derivata discontinua.

La definizione della funzione impulso di Dirac, descritta in Appendice B, ci consente comunque di effettuare la operazione di derivazione accettando che il risultato contenga degli impulsi, ovvero sia una funzione generalizzata. Figure 2.13, 2.14, 2.15 e 2.16 mostrano le pdf relative ai quattro esempi. Si noti come la presenza degli impulsi di Dirac nelle pdf di tipo discreto rifletta il fatto che la *massa* della distribuzione di probabilità, pari all'area degli impulsi, sia tutta concentrata in quei punti.

Esaminiamo ora le principali proprietà della pdf.

Proprietà 2.6 $f_X(x) \geq 0 \forall x$.

Prova: La positività della pdf discende immediatamente dalla monotonicità della cdf.

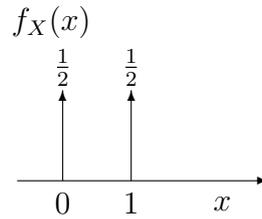


Figura 2.13: La pdf della variabile aleatoria associata all'esperimento della moneta

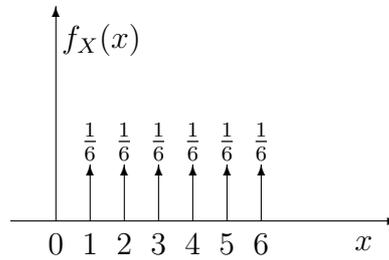


Figura 2.14: La pdf della variabile aleatoria associata all'esperimento del dado

Proprietà 2.7

$$\int_a^b f_X(x) dx = F_X(b) - F_X(a). \quad (2.18)$$

Prova: Questa proprietà discende direttamente dalla definizione di pdf e dal teorema fondamentale del calcolo integrale.

In effetti in quest'ultima proprietà abbiamo assunto che la derivata esista ovunque nell'intervallo (a, b) . La definizione di funzione di Dirac ci consente di operare in questo modo, anche se qualche attenzione in più dobbiamo riservarla ai casi in cui impulsi sono presenti agli estremi dell'intervallo. In tali casi è necessario distinguere tra limite destro e limite sinistro di F_X come spiegato a proposito delle sue proprietà.

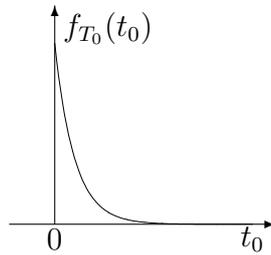


Figura 2.15: La pdf dell'esperimento della conversazione telefonica

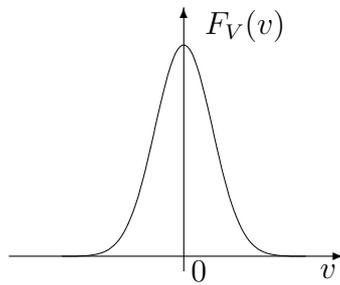


Figura 2.16: La pdf della tensione ai capi di un resistore rumoroso

Quest'ultima proprietà implica altre due proprietà (corollari) fondamentali:

Proprietà 2.8

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1. \quad (2.19)$$

L'area sottesa dalla pdf è unitaria.

Prova: Immediata da proprietà 2.7 e dalle proprietà asintotiche della cdf: $F_X(+\infty) = 1$ e $F_X(-\infty) = 0$.

Proprietà 2.9

$$Pr\{a < X \leq b\} = \int_a^b f_X(x) dx. \quad (2.20)$$

L'area sottesa dalla pdf nell'intervallo $]a, b]$ è proprio pari alla probabilità che la variabile aleatoria ha di trovarsi in quell'intervallo. Tale proprietà, mostrata graficamente in Figura 2.17 è proprio quella che ci consente calcolare probabilità di uso comune a partire dalla conoscenza della pdf.

Prova: Immediata conseguenza di proprietà 2.4 e 2.7.

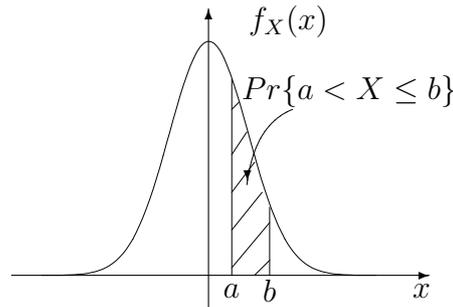


Figura 2.17: Il calcolo della probabilità $Pr\{a < X \leq b\}$ dalla pdf

Si noti come, dalla definizione, la funzione di densità possa essere scritta come

$$f_X(x) = \lim_{\Delta x \rightarrow 0} \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{Pr\{x < X \leq x + \Delta x\}}{\Delta x}. \quad (2.21)$$

Pertanto il valore della pdf in x può essere interpretato come il rapporto tra la probabilità che X assuma valori in un intorno di x , normalizzata alla dimensione dell'intorno. Da tale proprietà si vede anche come sia possibile che per una variabile aleatoria continua in x , con pdf diversa da zero, si abbia che $Pr\{X = x\} = 0$. Infatti la probabilità

$$Pr\{x < X \leq x + \Delta x\} = f_X(x)\Delta x, \quad (2.22)$$

tende a zero al tendere di Δx a zero, se in x non c'è un impulso di Dirac.

Queste considerazioni sono importanti per comprendere come la densità di probabilità in un punto x non rappresenti affatto la probabilità che ha la variabile aleatoria di assumere quel valore.

2.5 Cdf e pdf notevoli

Le sezioni che seguono discutono alcune cdf e pdf notevoli di uso comune di tipo continuo, discreto e misto.

2.5.1 La variabile aleatoria uniforme

Caso continuo: La variabile aleatoria uniforme di tipo continuo ha la caratteristica di avere una densità costante in un intervallo $[a, b]$ fissato. Ovvero

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{altrove.} \end{cases} \quad (2.23)$$

La corrispondente funzione di distribuzione cumulativa è pertanto

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases} \quad (2.24)$$

Figura 2.18 mostra l'andamento di pdf e cdf. Si noti come la discontinuità agli estremi della pdf non crei ambiguità e che comunque l'inclusione o l'esclusione degli estremi sia irrilevante ai fini dell'integrazione di $f_X(x)$. L'altezza della pdf è uguale a $\frac{1}{b-a}$ per soddisfare la condizione $\int_{-\infty}^{+\infty} f_X(x)dx = 1$.

Avendo discusso la variabile aleatoria uniforme di tipo continuo, siamo ora in condizioni di giustificare più rigorosamente la formula euristica (1.25) data nel capitolo precedente. Infatti la probabilità che una variabile aleatoria ha di cadere nell'intervallo $[x_1, x_2] \subseteq [a, b]$ è

$$Pr\{x_1 < X \leq x_2\} = \int_{x_1}^{x_2} \frac{1}{b-a} dx = \frac{x_2 - x_1}{b-a} = \frac{\dim[x_1, x_2]}{\dim[a, b]}. \quad (2.25)$$

Caso discreto: La variabile aleatoria uniforme di tipo discreto ha una massa di probabilità concentrata su un numero discreto e finito di punti, tipicamente a spaziatura uniforme (non necessariamente). Figura 2.19 mostra la pdf e la cdf relativa. In formule

$$f_X(x) = \frac{1}{N} \sum_{k=0}^{N-1} \delta(x - \alpha - k\Delta), \quad (2.26)$$

$$F_X(x) = \frac{1}{N} \sum_{k=0}^{N-1} u(x - \alpha - k\Delta), \quad (2.27)$$

dove Δ è la spaziatura tra i punti e α è la posizione del primo punto.

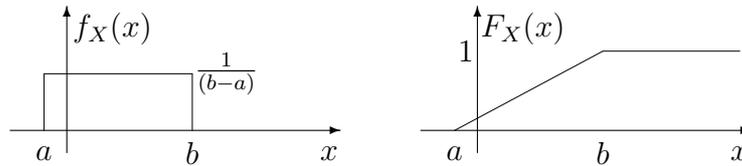


Figura 2.18: La pdf e la cdf per la variabile aleatoria uniforme di tipo continuo

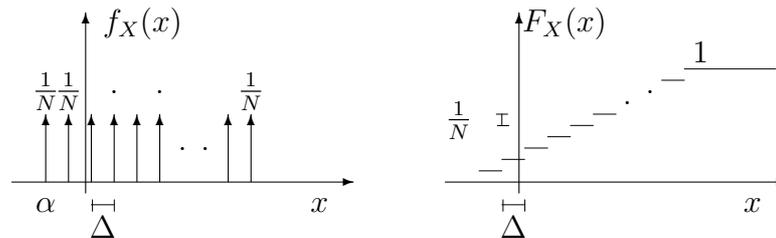


Figura 2.19: La pdf e la cdf per la variabile aleatoria uniforme di tipo discreto

Esempio 2.6 La dimensione X di una fessura misurata in mm è una variabile aleatoria distribuita uniformemente nell'intervallo $[0, 3]$. Si valuti: (a) la probabilità che la fessura sia più piccola di 0.1 mm; (b) la probabilità che la fessura sia o più piccola 0.01 mm o maggiore di 2 mm. **Soluzione:** La densità di probabilità è costante con valore $1/3$. (a) $Pr\{X \leq 0.1\} = \int_0^{0.1} \frac{1}{3} dx = \frac{1}{30}$; (b) $Pr\{\{X \leq 0.01\} \cup \{X > 2\}\} = Pr\{X \leq 0.01\} + Pr\{X > 2\} = \frac{1}{300} + \frac{1}{3} = \frac{101}{300}$.

Esempio 2.7 Un evento atmosferico può verificarsi ad un'ora qualunque delle 24 ore di una giornata. Qual'è la probabilità che tale evento si verichi nelle prossime 2 ore? **Soluzione:** L'ora a cui si verifica l'evento in questione è una variabile aleatoria uniformemente distribuita sui 24 valori discreti $\{1, 2, \dots, 24\}$. La massa di probabilità su ogni punto è ovviamente $\frac{1}{24}$. La probabilità che l'evento si verifichi nelle prossime due ore, se non ho informazioni su quelle precedenti, è semplicemente $\frac{2}{24} = \frac{1}{12}$.

2.5.2 La variabile aleatoria esponenziale

Caso continuo: La variabile aleatoria esponenziale monolatera ha una pdf decrescente con legge esponenziale. I valori assunti possono essere solo positivi. La pdf e la cdf sono rispettivamente

$$f_X(x) = \alpha e^{-\alpha x} u(x), \quad (2.28)$$

$$F_X(x) = (1 - e^{-\alpha x}) u(x), \quad (2.29)$$

con $\alpha > 0$. Figura 2.20 ne mostra gli andamenti.

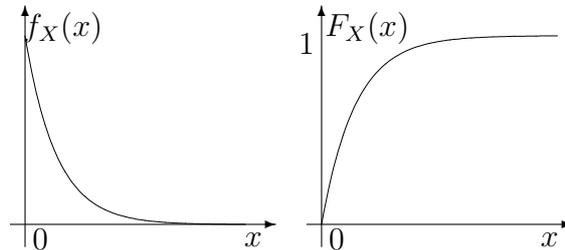


Figura 2.20: La pdf e la cdf per la variabile aleatoria esponenziale

Caso discreto: la variabile aleatoria esponenziale di tipo discreto segue lo stesso andamento della variabile continua ed è del tipo

$$f_X(x) = \sum_{k=0}^{\infty} \beta e^{-\gamma k} \delta(x - k\Delta), \quad (2.30)$$

con $\gamma > 0$. Il valore di β è legato alla proprietà 2.8 che richiede che $f_X(x)$ sia una valida funzione di densità di probabilità. Poiché

$$\int_{-\infty}^{+\infty} f_X(\xi) d\xi = \sum_{k=0}^{\infty} \beta e^{-\gamma k} = \beta \frac{1}{1 - e^{-\gamma}}, \quad (2.31)$$

la condizione $\int_{-\infty}^{+\infty} f_X(\xi) d\xi = 1$ diventa

$$\beta = 1 - e^{-\gamma}. \quad (2.32)$$

Esempio 2.8 Ad un server di rete, il tempo T (misurato in msec) che intercorre tra l'arrivo di un pacchetto e il successivo è distribuito secondo una pdf esponenziale con parametro $\alpha = 5$. Calcolare la probabilità che il tempo di attesa sia: (a) inferiore a 1 msec; (b) esattamente pari a 1 msec; (c) compreso nell'intervallo 1-10 msec. **Soluzione:** (a) $Pr\{X \leq 1\} = Pr\{0 < X \leq 1\} = \int_0^1 5e^{-5\xi} d\xi = -e^{-5\xi} \Big|_0^1 = 1 - e^{-5} = 0.9933$; (b) $Pr\{X = 1\} = 0$; (c) $Pr\{1 < X \leq 10\} = \int_1^{10} 5e^{-5\xi} d\xi = -e^{-5\xi} \Big|_1^{10} = e^{-5} - e^{-50} = 0.0067$.

Esempio 2.9 Ad un casello autostradale, sia t il tempo che intercorre tra l'arrivo di un'auto e la successiva. Un orologio, che ha la massima risoluzione in secondi, ne misura il tempo T in secondi interi. Assumendo di sapere solo che la variabile discreta T è distribuita secondo una pdf discreta esponenziale con parametro $\gamma = \frac{1}{5}$, calcolare la probabilità che il tempo t tra un'auto e la successiva sia: (a) inferiore a un secondo; (b) tra 3 e 5 secondi; (c) maggiore di un'ora; (d) tra 1.5 e 2.5 secondi. **Soluzione:** La pdf di T è di tipo discreto e assume valori in $\{0, 1, 2, 3, 4, \dots\}$. L'orologio scatta al valore successivo al termine di ogni secondo, pertanto la probabilità su $T = j$ è la probabilità che $j < t < j + 1$, $j = 0, 1, \dots$. Pertanto: (a) $Pr\{t < 1\} = Pr\{T = 0\} = (1 - e^{-\gamma})e^{-\gamma} = 0.1484$; (b) $Pr\{3 < t < 5\} = Pr\{\{T = 3\} \cup \{T = 4\}\} = (1 - e^{-\gamma}) \sum_{i=3}^4 e^{-\gamma i} = (1 - e^{-\gamma}) (e^{-\gamma 3} + e^{-\gamma 4}) = 0.1809$; (c) $Pr\{t > 1 \text{ ora}\} = Pr\{T \geq 3600\} = 1 - Pr\{0 \leq T \leq 3599\} = 1 - (1 - e^{-\gamma}) \sum_{i=0}^{3599} e^{-\gamma i} = 1 - (1 - e^{-\gamma}) \frac{1 - e^{-\gamma 3600}}{1 - e^{-\gamma}} = 1 - (1 - e^{-\gamma 3600}) = 2.0322 \cdot 10^{-313} \simeq 0$; (d) la risposta esatta a questo quesito è impossibile, in quanto è nota solo la probabilità su $T = 1$ e su $T = 2$, che

corrispondono rispettivamente a $Pr\{1 < t < 2\}$ e $Pr\{2 < t < 3\}$. Pertanto è possibile solo limitare superiormente la probabilità cercata: $Pr\{1.5 < t < 2.5\} \leq Pr\{T = 1\} + Pr\{T = 2\}$.

2.5.3 La variabile aleatoria di Laplace

Caso continuo: La variabile aleatoria esponenziale bilatera, detta anche *variabile aleatoria di Laplace*, segue lo stesso andamento della variabile aleatoria esponenziale, ma può assumere anche valori negativi distribuiti simmetricamente rispetto a quelli positivi. In formule

$$f_X(x) = \frac{\alpha}{2} e^{-\alpha|x|}, \quad (2.33)$$

$$F_X(x) = \frac{1}{2} e^{\alpha x} u(-x) + \left(1 - \frac{1}{2} e^{-\alpha x}\right) u(x). \quad (2.34)$$

Figura 2.21 ne mostra gli andamenti relativi.

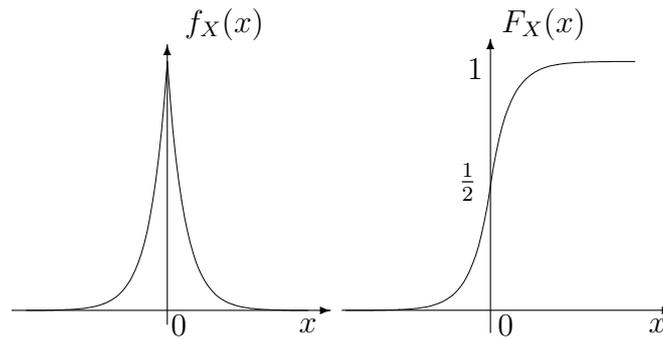


Figura 2.21: La pdf e la cdf per la variabile aleatoria di Laplace

Caso discreto: La estensione al caso discreto è lasciata al lettore per esercizio.

Esempio 2.10 Il valore istantaneo X di un segnale vocale è distribuito secondo una pdf di Laplace. Calcolare la probabilità che il segnale mandi

in saturazione un amplificatore che opera linearmente nell'intervallo $[-x_0, x_0]$, con $x_0 > 0$. **Soluzione:** $Pr\{\text{l'amplificatore satura}\} = Pr\{\{X > x_0\} \cup \{X < -x_0\}\} = 1 - Pr\{-x_0 < X \leq x_0\} = 1 - \int_{-x_0}^{x_0} \frac{\alpha}{2} e^{-\alpha|\xi|} d\xi = 1 - \alpha \int_0^{x_0} e^{-\alpha\xi} d\xi = 1 + e^{-\alpha\xi} \Big|_0^{x_0} = e^{-\alpha x_0}$.

2.5.4 La variabile aleatoria di Poisson

La variabile aleatoria di Poisson è di tipo discreto. La densità di probabilità può essere scritta come

$$f_X(x) = e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} \delta(x - k), \quad a > 0, \quad (2.35)$$

oppure come

$$Pr\{X = k\} = e^{-a} \frac{a^k}{k!}, \quad k \geq 0 \quad a > 0. \quad (2.36)$$

Si noti che vale la seguente relazione ricorsiva

$$Pr\{X = k\} = Pr\{X = k - 1\} \frac{a}{k}. \quad (2.37)$$

Ovvero la sequenza $\{Pr\{X = k\}\}$ è crescente o decrescente a seconda del valore di $\frac{a}{k}$, e resta costante quando $a = k$. Pertanto, se $a < 1$, la probabilità è massima per $k = 0$. Se $a > 1$, ma non è un intero, allora il massimo si ha per k pari all'intero immediatamente minore di a . Invece se $a > 1$ ed è un intero, si hanno due massimi in corrispondenza di $k = a$ e $k = a - 1$. Figura 2.22 mostra l'andamento della pdf e della cdf per $a = 3$.

Esempio 2.11 Il numero di pacchetti X in arrivo ad un server in un minuto è distribuito secondo una pdf di Poisson con parametro $a = 100$. Si valuti la probabilità che in un minuto: (a) non ci sia alcun arrivo; (b) ci siano esattamente 10 arrivi; (c) ci siano più di 50 arrivi. **Soluzione:** (a) $Pr\{X = 0\} = e^{-a} = 3.7201 \cdot 10^{-44}$; (b) $Pr\{X = 10\} = e^{-a} \frac{a^{10}}{10!} = e^{-100} \frac{100^{10}}{10!}$ (per eseguire i calcoli, conviene utilizzare la formula ricorsiva (2.37), per evitare di dover calcolare rapporti tra quantità troppo grandi); (c) $Pr\{X > 50\} = e^{-a} \sum_{i=51}^{\infty} \frac{a^i}{i!} =$

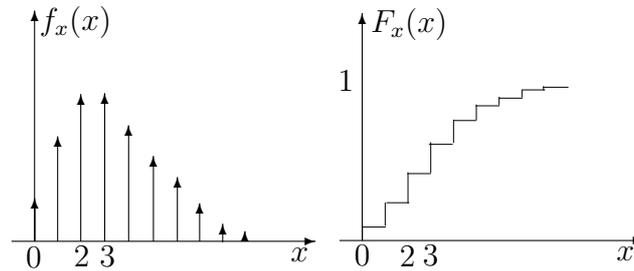


Figura 2.22: La pdf e la cdf per la variabile aleatoria di Poisson con $a = 3$

$1 - Pr\{X \leq 50\} = 1 - e^{-a} \sum_{i=0}^{50} \frac{a^i}{i!} = 1 - e^{-100} \sum_{i=0}^{50} \frac{100^i}{i!}$. È necessario valutare l'espressione al calcolatore, o in maniera approssimata, in quanto non è disponibile una espressione in forma chiusa. Suggeriamo al lettore di utilizzare ancora la formula ricorsiva (2.37) per progettare un algoritmo numericamente stabile.

2.5.5 La variabile aleatoria di Rayleigh

La variabile aleatoria di Rayleigh è di tipo continuo. La pdf è descritta dalla espressione

$$f_X(x) = \frac{x}{\alpha^2} e^{-\frac{x^2}{2\alpha^2}} u(x). \quad (2.38)$$

Figura 2.23 ne mostra l'andamento. Il lettore verifichi che equazione (2.38) è una valida pdf e valuti la cdf corrispondente.

Esempio 2.12 Un arciere scaglia una freccia verso un bersaglio circolare di diametro D_0 . La freccia colpisce il bersaglio ad una distanza D dal centro, modellata come una variabile aleatoria distribuita secondo una pdf di tipo Rayleigh con parametro α . Calcolare la probabilità che: (a) la freccia cada entro una distanza d_0 dal centro; (b) la freccia cada all'esterno del bersaglio. **Soluzione:** (a) $Pr\{D \leq d_0\} = \int_0^{d_0} \frac{x}{\alpha^2} e^{-\frac{x^2}{2\alpha^2}} dx = -e^{-\frac{x^2}{2\alpha^2}} \Big|_0^{d_0} = 1 - e^{-\frac{d_0^2}{2\alpha^2}}$; (b) dal risultato in (a), $Pr\{D > D_0\} = \int_{D_0}^{\infty} \frac{x}{\alpha^2} e^{-\frac{x^2}{2\alpha^2}} dx = e^{-\frac{D_0^2}{2\alpha^2}}$.

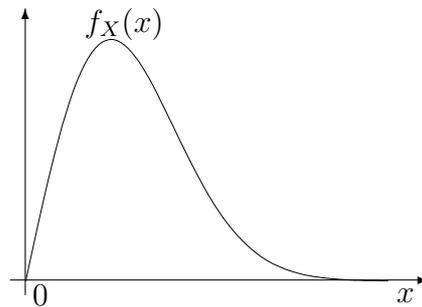


Figura 2.23: La pdf per la variabile aleatoria di Rayleigh

2.5.6 La variabile aleatoria binomiale

La variabile aleatoria binomiale è di tipo discreto ed è descritta dalla seguente formula

$$f_X(x) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta(x-k). \quad (2.39)$$

Equivalentemente la pdf può essere descritta come

$$Pr\{X = k\} = b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad (2.40)$$

dove si riconosce la legge binomiale descritta nel primo capitolo. Figura 2.24 mostra la pdf per $n = 7$ e $p = 0.5$. Usando la definizione di coefficiente binomiale abbiamo che vale la seguente relazione ricorsiva

$$\begin{aligned} Pr\{X = k\} &= \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \\ &= \frac{n!(n-k+1)}{(n-k+1)!k(k-1)!} p^{k-1} (1-p)^{-1} (1-p)^{n-k+1} \\ &= \frac{n-k+1}{k} \frac{p}{(1-p)} Pr\{X = k-1\}. \end{aligned} \quad (2.41)$$

Quindi la sequenze è crescente quando:

$$\frac{n-k+1}{k} \frac{p}{(1-p)} > 1, \quad (2.42)$$

ovvero quando $k < (n+1)p$. Viceversa è decrescente quando $k > (n+1)p$. Se $(n+1)p$ è un intero, abbiamo due massimi per $k = (n+1)p - 1$ e $k = (n+1)p$ (vedi la figura 2.24 per un esempio). Altrimenti il massimo è attinto per $k = [(n+1)p]$ ($[\cdot]$ indica la parte intera).

Un caso particolare degno di nota è la cosiddetta *variabile di Bernoulli*, che si ottiene per $n = 1$. Tale variabile è quella associabile all'esperimento binario della moneta truccata. In particolare:

$$Pr\{X = 0\} = (1-p); \quad Pr\{X = 1\} = p. \quad (2.43)$$

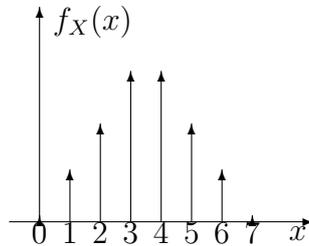


Figura 2.24: La pdf per la variabile aleatoria binomiale con $n = 7$ e $p = 0.5$

Esempio 2.13 I 5 calcolatori di un ufficio tentano simultaneamente l'accesso ad un servizio web. L'accesso, a causa di traffico interferente, non è sempre garantito e ha una probabilità di successo solo del 70%. Calcolare la probabilità che: (a) tutti i calcolatori ottengano l'accesso; (b) tre ottengano l'accesso; (c) almeno due ottengano l'accesso.

Soluzione: Il numero X di calcolatori che ottiene l'accesso al servizio è una variabile aleatoria binomiale con $p = 0.7$. Pertanto: (a) $Pr\{X = 5\} = \binom{5}{5}p^5(1-p)^0 = p^5 = 0.7^5 = 0.1681$; (b) $Pr\{X = 3\} = \binom{5}{3}p^3(1-p)^2 = 10 \cdot 0.7^3 \cdot 0.3^2 = 0.3087$; (c) $Pr\{X \geq 2\} = 1 - Pr\{X \leq 1\} = 1 - \binom{5}{0}p^0(1-p)^5 - \binom{5}{1}p^1(1-p)^4 = 0.9692$.

2.6 Funzione di massa di probabilità

In una variabile aleatoria discreta la probabilità è concentrata in un insieme discreto di punti. E' utile associare alla probabilità presente in ogni punto il concetto di *massa di probabilità*. In tal caso si parla di *funzione di massa di probabilità*, o più semplicemente *funzione di massa*, denotata come *pmf* (*probability mass function*). Per la variabile aleatoria che assume valori sull'insieme discreto di punti $\{x_k\}$, con probabilità

$$Pr\{X = x_k\}, \quad (2.44)$$

si definisce *funzione di massa di probabilità*, o *pmf* (*probability mass function*)

$$p_X(x) = \sum_k Pr\{X = x_k\}d(x - x_k), \quad (2.45)$$

dove $d(\cdot)$ è l'*impulso unitario* (o *delta discreta*) definito come

$$d(x) = \begin{cases} 1 & x = 0 \\ 0 & \text{altrove.} \end{cases} \quad (2.46)$$

Si noti la notevole differenza esistente tra l'impulso unitario $d(\cdot)$ e l'impulso di Dirac $\delta(\cdot)$. Figura 2.25 mostra la differente notazione grafica. Si noti inoltre che la funzione $d(\cdot)$ è tale che:

$$\int_{-\infty}^{\infty} d(x - x_k)dx = 0, \quad (2.47)$$

in contrasto alla proprietà fondamentale della delta di Dirac. La pmf della variabile aleatoria discreta può anche essere descritta semplicemente dalla sequenza

$$p_X[k] = Pr\{X = x_k\}, \quad (2.48)$$

che però non contiene la indicazione esplicita dei valori $\{x_k\}$. Figura 2.26 mostra la pmf relativa alla distribuzione binomiale.

Si ricordi che l'insieme di definizione della variabile aleatoria discreta può anche essere infinito e che le probabilità devono essere tali che

$$\sum_k p_X[k] = 1. \quad (2.49)$$

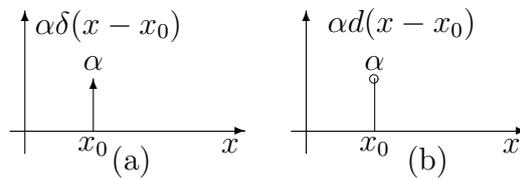


Figura 2.25: (a) La delta di Dirac; (b) la delta discreta

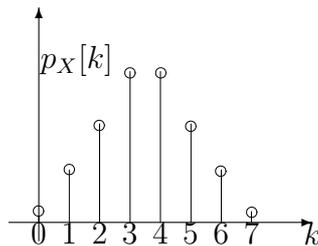


Figura 2.26: La pmf corrispondente alla variabile aleatoria binomiale di figura 2.24

Mediante l'utilizzo della funzione di Dirac la pmf può anche essere scritta come

$$f_X(x) = \sum_k P_X[k] \delta(x - x_k). \quad (2.50)$$

Il lettore apprezzi la differenza formale tra le tre possibili rappresentazioni della variabile aleatoria discreta:

$$f_X(x), \quad p_X[k], \quad p_X(x), \quad (2.51)$$

in cui la prima è una funzione generalizzata, la seconda è una sequenza e la terza è una funzione costituita da impulsi discreti.

Distribuzioni miste: La formalizzazione della pdf mediante le funzioni di Dirac ci consente una descrizione unica, sia per variabili aleatorie che hanno dei sottoinsiemi di definizione discreti, sia per variabili aleatorie definite su insiemi densi. Pertanto quando è necessario manipolare

variabili aleatorie miste si preferisce utilizzare la formalizzazione con le funzioni generalizzate.

Esempio 2.14 Una tensione V espressa in Volt è una variabile aleatoria distribuita secondo una densità del tipo di quella riportata in figura 2.16, quindi definita su tutto l'asse reale delle tensioni. Un voltmetro misura il valore della tensione su una scala che va da v_1 a v_2 . Quando il valore della tensione è al di fuori di tale intervallo, il voltmetro riporta i valori di fondo-scala. Si valuti la densità di probabilità della tensione misurata. **Soluzione:** Si denoti con Y la variabile aleatoria che rappresenta la tensione misurata. In formule

$$Y = \begin{cases} v_1 & \text{se } V < v_1 \\ V & \text{se } v_1 \leq V < v_2 \\ v_2 & \text{se } V \geq v_2. \end{cases} \quad (2.52)$$

La variabile aleatoria risultante è di tipo misto, poiché la tensione misurata è uguale a quella effettiva per i valori inclusi nell'intervallo (v_1, v_2) , ma assume i valori discreti di fondo-scala con probabilità non nulla. Pertanto

$$f_Y(y) = p_1\delta(y - v_1) + f_V(y)(u(y - v_1) - u(y - v_2)) + p_2\delta(y - v_2), \quad (2.53)$$

dove

$$p_1 = Pr\{V < v_1\} = \int_{-\infty}^{v_1} f_V(x)dx, \quad p_2 = Pr\{V > v_2\} = \int_{v_2}^{+\infty} f_V(x)dx. \quad (2.54)$$

Figura 2.27 mostra la densità risultante.

2.7 La variabile aleatoria Gaussiana

Un'attenzione particolare merita la variabile aleatoria *Gaussiana*, detta anche variabile aleatoria *normale*. In un gran numero di applicazioni tale distribuzione emerge naturalmente, anche in virtù di uno dei risultati fondamentali della teoria della probabilità che è il *Teorema centrale del limite*, di cui parleremo nel seguito di queste note. La pdf di una variabile aleatoria gaussiana è definita come

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \triangleq \mathcal{N}(x; \mu, \sigma). \quad (2.55)$$

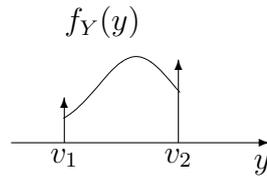


Figura 2.27: La pdf della tensione misurata con gli effetti di fondo-scala

La figura 2.28 ne mostra il tipico andamento *a campana*. Il parametro μ rappresenta il punto di massimo attorno al quale è centrata la curva mentre σ rappresenta il parametro da cui dipende la larghezza della curva². Si noti come 2σ rappresenti la larghezza della campana all'ordinata pari a $e^{-\frac{1}{2}} = 0.61$ del valore massimo.

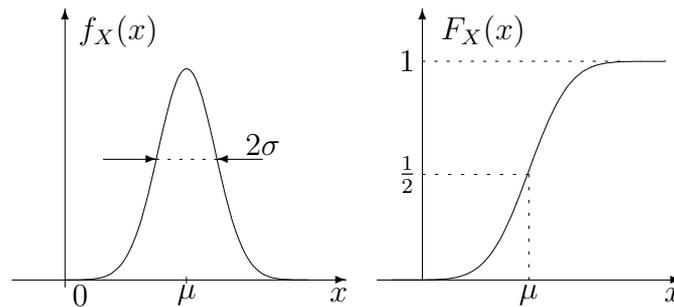


Figura 2.28: La pdf e la cdf della variabile aleatoria gaussiana

La funzione di distribuzione cumulativa non è esprimibile in forma chiusa e viene tabellata, o calcolata con degli algoritmi di integrazione numerica. Per rendere i calcoli sulla variabile aleatoria gaussiana più

²Il parametro σ è detto *deviazione standard*. La deviazione standard elevata al quadrato σ^2 è detta *varianza*. La definizione generale di deviazione standard e varianza sarà data in seguito.

agevoli, si definisce la densità *normale unitaria*

$$\mathcal{N}(x; 0, 1) \triangleq \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad (2.56)$$

ovvero la densità gaussiana centrata in zero e con parametro $\sigma = 1$. La cdf della normale unitaria è

$$\Phi(x) \triangleq \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi. \quad (2.57)$$

La cdf di una generica gaussiana è facilmente espressa in funzione della $\Phi(x)$

$$F_x(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\xi-\mu)^2}{2\sigma^2}} d\xi = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \Phi\left(\frac{x-\mu}{\sigma}\right). \quad (2.58)$$

La probabilità che la variabile aleatoria cada in un intervallo generico $]a, b]$ si esprime pertanto come

$$\begin{aligned} Pr\{a < X \leq b\} &= \int_a^b \mathcal{N}(\xi; \mu, \sigma) d\xi \\ &= F_X(b) - F_X(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \end{aligned} \quad (2.59)$$

La funzione $\Phi(x)$ è tabellata in appendice A per alcuni valori di x . Si noti come sia sufficiente tabellare la funzione solo per valori positivi in quanto

$$\Phi(-x) = 1 - \Phi(x). \quad (2.60)$$

Più comune nella letteratura della teoria dell'informazione è la funzione *quantile*

$$Q(x) \triangleq \int_x^{+\infty} \mathcal{N}(\xi; 0, 1) d\xi, \quad (2.61)$$

legata alla funzione $\Phi(x)$ dalla relazione: $Q(x) = 1 - \Phi(x)$. Anche la funzione $Q(x)$, è riportata nelle tabelle solo per $x > 0$, visto che

$$Q(-x) = 1 - Q(x). \quad (2.62)$$

La probabilità che la variabile aleatoria cada in un intervallo generico $]a, b]$ si esprime pertanto come

$$\begin{aligned} Pr\{a < X \leq b\} &= \int_a^b \mathcal{N}(\xi; \mu, \sigma) d\xi \\ &= F_X(b) - F_X(a) = Q\left(\frac{a-\mu}{\sigma}\right) - Q\left(\frac{b-\mu}{\sigma}\right). \end{aligned} \quad (2.63)$$

Figura 2.29 mostra graficamente le definizioni di $\Phi(x)$, $Q(x)$, e il calcolo della probabilità.

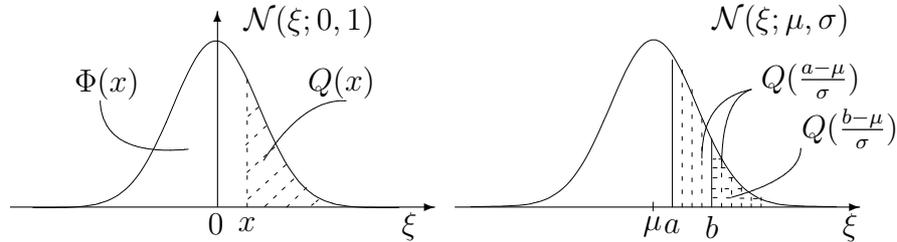


Figura 2.29: (a) La definizione delle funzioni $\Phi(x)$ e $Q(x)$; (b) Il calcolo di $Pr\{a < X \leq b\}$ mediante la funzione quantile.

Ai fini della manipolazione della densità gaussiana, nella letteratura sono definite anche le funzioni *erf* (*error function*) e *erfc* (*complementary error function*)

$$erf(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi, \quad (2.64)$$

$$erfc(x) \triangleq 1 - erf(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-\xi^2} d\xi. \quad (2.65)$$

Per dimostrare l'ultima uguaglianza basta ricordare che

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\eta^2} d\eta = 1. \quad (2.66)$$

Dalla simmetria della funzione

$$\begin{aligned} 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\eta^2} d\eta - \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi \\ &= \frac{2}{\sqrt{\pi}} \int_0^{+\infty} e^{-\xi^2} d\xi - \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi \\ &= \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-\xi^2} d\xi. \end{aligned} \quad (2.67)$$

Anche le relazioni con le funzioni Φ e Q sono facilmente ottenibili

$$\Phi(x) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right) = 1 - \frac{1}{2} \operatorname{erfc} \left(\frac{x}{\sqrt{2}} \right), \quad (2.68)$$

$$Q(x) = \frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right) = \frac{1}{2} \operatorname{erfc} \left(\frac{x}{\sqrt{2}} \right), \quad (2.69)$$

$$\operatorname{erf}(x) = 2\Phi(\sqrt{2}x) - 1 = 1 - 2Q(\sqrt{2}x), \quad (2.70)$$

$$\operatorname{erfc}(x) = 2 \left(1 - \Phi(\sqrt{2}x) \right) = 2Q(\sqrt{2}x). \quad (2.71)$$

Usando le funzioni erf e erfc abbiamo

$$\begin{aligned} \Pr\{a < X \leq b\} &= \frac{1}{2} \operatorname{erf} \left(\frac{b - \mu}{\sqrt{2}\sigma} \right) - \frac{1}{2} \operatorname{erf} \left(\frac{a - \mu}{\sqrt{2}\sigma} \right) \\ &= \frac{1}{2} \operatorname{erfc} \left(\frac{a - \mu}{\sqrt{2}\sigma} \right) - \frac{1}{2} \operatorname{erfc} \left(\frac{b - \mu}{\sqrt{2}\sigma} \right). \end{aligned} \quad (2.72)$$

Il lettore potrà pertanto utilizzare le funzioni per cui dispone di un più agevole accesso ai valori numerici (tavole o algoritmi).

Esempio 2.15 Si consideri una variabile aleatoria X che rappresenta la temperatura ambiente espressa in gradi centigradi. Si assuma che essa sia distribuita secondo una densità gaussiana $\mathcal{N}(x; 20, 5)$. Valutiamo le seguenti probabilità: 1) $\Pr\{X > 30\} = \int_{30}^{+\infty} \mathcal{N}(x; 20, 5)dx = \Phi(\infty) - \Phi\left(\frac{30-20}{5}\right) = 1 - \Phi(2) = 0.0228$; 2) $\Pr\{19 < X \leq 21\} = \int_{19}^{21} \mathcal{N}(x; 20, 5)dx = Q\left(\frac{19-20}{5}\right) - Q\left(\frac{21-20}{5}\right) = Q(-0.2) - Q(0.2) = 1 - Q(0.2) - Q(0.2) = 1 - 2Q(0.2) = 1 - 2 \cdot 0.4207 = 0.1586$; 3) $\Pr\{10 < X \leq 20\} = \int_{10}^{20} \mathcal{N}(x; 20, 5)dx = Q\left(\frac{10-20}{5}\right) - Q\left(\frac{20-20}{5}\right) = Q(-2) - Q(0) = 1 - Q(2) - 0.5 = 0.5 - Q(2) = 0.5 - 0.0228 = 0.4772$ (i valori della funzione quantile possono essere ottenuti mediante tavole o calcolatrici).

2.8 I momenti della variabile aleatoria

La funzione cumulativa, o la funzione di densità di una variabile aleatoria, costituiscono una caratterizzazione completa di come la probabilità si distribuisce sull'asse reale. Nonostante ciò, è spesso utile caratterizzare una variabile aleatoria mediante un piccolo insieme di parametri che ne rappresentino alcune caratteristiche essenziali.

2.8.1 La media

Si definisce *media* di una variabile aleatoria X

$$\mu \triangleq E[X] = \int_{-\infty}^{+\infty} \xi f_X(\xi) d\xi. \quad (2.73)$$

La media rappresenta il baricentro della funzione $f_X(x)$ ed è anche detta *valore atteso* o, nella terminologia anglo-sassone, *expectation* della variabile aleatoria (da cui la notazione $E[.]$). Per una variabile aleatoria di tipo discreto che assume valori $\{x_k\}$ secondo la funzione di massa $p_X(k)$ la definizione diventa

$$E[X] = \sum_k x_k p_X[k]. \quad (2.74)$$

La definizione (2.73) è comunque la più generale in quanto essa è applicabile anche alle densità di tipo misto, purché si siano utilizzate le funzioni generalizzate di Dirac. Vediamo alcuni esempi di calcolo della media.

Variabile aleatoria uniforme: Si supponga che la variabile aleatoria X sia di tipo continuo e uniformemente distribuita nell'intervallo $[a, b]$. La media è

$$E[X] = \int_a^b \frac{1}{b-a} \xi d\xi = \frac{1}{b-a} \left. \frac{\xi^2}{2} \right|_a^b = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{b+a}{2}. \quad (2.75)$$

La media è quindi il punto medio dell'intervallo di definizione. Nel caso discreto, si assuma che gli N valori assunti dalla variabile siano uniformemente spazati e pari a $\{\alpha + \Delta k, k = 0, \dots, N-1\}$. La media è

$$E[X] = \frac{1}{N} \sum_{k=0}^{N-1} (\alpha + \Delta k) = \alpha + \frac{\Delta}{N} \sum_{k=0}^{N-1} k = \alpha + \frac{\Delta}{N} \frac{(N-1)N}{2} = \alpha + \frac{\Delta(N-1)}{2}, \quad (2.76)$$

dove abbiamo fatto uso del risultato notevole: $\sum_{k=1}^n k = \frac{n(n+1)}{2}$.

Variabile aleatoria esponenziale: Si supponga che la variabile aleatoria X sia di tipo continuo e distribuita secondo una pdf esponenziale

monolatera. La media è

$$\begin{aligned} E[X] &= \alpha \int_0^{+\infty} \xi e^{-\alpha\xi} d\xi = \alpha \left(\left. \frac{\xi e^{-\alpha\xi}}{-\alpha} \right|_0^{+\infty} - \int_0^{+\infty} \frac{e^{-\alpha\xi}}{-\alpha} d\xi \right) \\ &= 0 + \int_0^{+\infty} e^{-\alpha\xi} d\xi = \left. \frac{e^{-\alpha\xi}}{-\alpha} \right|_0^{+\infty} = \frac{1}{\alpha}. \end{aligned} \quad (2.77)$$

Variabile aleatoria di Bernoulli: La variabile aleatoria X può assumere solo i due valori discreti 0 e 1 con probabilità $(1-p)$ e p . La media è pertanto

$$E[X] = 0 \cdot (1-p) + 1 \cdot p = p. \quad (2.78)$$

Variabile aleatoria di Laplace: La media della variabile aleatoria di Laplace è facilmente ottenuta per simmetria

$$E[X] = \frac{\alpha}{2} \int_{-\infty}^{+\infty} \xi e^{-\alpha|\xi|} d\xi = 0, \quad (2.79)$$

poiché la funzione integranda è una funzione dispari, visto che ξ è una funzione dispari e $e^{-\alpha|\xi|}$ è una funzione pari.

Variabile aleatoria gaussiana: Dalla definizione abbiamo che la media è

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \xi e^{-\frac{(\xi-\mu)^2}{2\sigma^2}} d\xi. \quad (2.80)$$

Cambiando variabile all'interno dell'integrale $x = \frac{\xi-\mu}{\sigma}$, abbiamo

$$\begin{aligned} E[X] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma x + \mu) e^{-\frac{x^2}{2}} dx \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx + \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx. \end{aligned} \quad (2.81)$$

Poiché la funzione integranda del primo integrale è dispari, abbiamo che il primo integrale è nullo mentre il secondo vale uno per la proprietà fondamentale delle pdf e pertanto

$$E[X] = \mu. \quad (2.82)$$

Quindi il parametro μ della variabile aleatoria gaussiana è proprio la sua media.

Variabile aleatoria binomiale: Dalla definizione e dalla pmf della variabile aleatoria binomiale, la media si può scrivere come

$$E[X] = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \quad (2.83)$$

Dalla definizione di coefficiente binomiale abbiamo

$$k \binom{n}{k} = k \frac{n!}{(n-k)!k!} = n \frac{(n-1)!}{(n-k)!(k-1)!} = n \binom{n-1}{k-1}. \quad (2.84)$$

La media si può pertanto esprimere come

$$E[X] = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}. \quad (2.85)$$

Ponendo $m = k - 1$ all'interno della sommatoria e usando la formula di Newton per la potenza di un binomio abbiamo

$$E[X] = np \sum_{m=0}^{n-1} \binom{n-1}{m} p^m (1-p)^{n-1-m} = np(p+1-p)^{n-1} = np. \quad (2.86)$$

Si noti come la media sia un valore compreso tra 0 e n e come essa non sia necessariamente uno dei valori assunti dalla variabile aleatoria.

Variabile aleatoria di Poisson: Dalla definizione di media e di pmf di Poisson abbiamo

$$\begin{aligned} E[X] &= e^{-\lambda} \sum_{k=0}^{+\infty} k \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{+\infty} k \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{m=0}^{+\infty} \frac{\lambda^m}{m!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned} \quad (2.87)$$

Variabile aleatoria di Rayleigh: La media della variabile è data dalla espressione

$$E[X] = \int_0^{+\infty} \xi \frac{\xi}{\alpha^2} e^{-\frac{\xi^2}{2\alpha^2}} d\xi. \quad (2.88)$$

Si noti che la primitiva di $\frac{\xi}{\alpha^2} e^{-\frac{\xi^2}{2\alpha^2}}$ è $-e^{-\frac{\xi^2}{2\alpha^2}}$. Pertanto, integrando per parti abbiamo

$$E[X] = -\xi e^{-\frac{\xi^2}{2\alpha^2}} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\frac{\xi^2}{2\alpha^2}} d\xi = 0 + \sqrt{2\pi\alpha^2} \int_0^{+\infty} \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{\xi^2}{2\alpha^2}}. \quad (2.89)$$

L'integrale ha come funzione integranda una pdf gaussiana a media nulla e parametro σ ed è esteso a metà dominio di definizione. Pertanto esso vale $\frac{1}{2}$ e

$$E[X] = \sqrt{\frac{\pi}{2}} \alpha. \quad (2.90)$$

Proprietà generali della media

Proprietà 2.10 La media di una pdf pari è nulla. Più in generale, se la pdf è simmetrica attorno ad un valore m , la media è pari a m .

Prova: La media è l'integrale della funzione $xf_X(x)$. Esso è nullo se $f_X(x)$ è pari (simmetrica attorno allo zero) poiché la funzione integranda è dispari. Più in generale, se la funzione è simmetrica attorno al punto $x = m$, ovvero $f_X(m-x) = f_X(m+x)$, abbiamo

$$E[X] = \int_{-\infty}^{+\infty} xf_X(x)dx = \int_{-\infty}^{+\infty} (m-y)f_X(m-y)dy, \quad (2.91)$$

dove abbiamo operato la sostituzione $x = m - y$.

$$E[X] = m \int_{-\infty}^{+\infty} f_X(m-y)dy - \int_{-\infty}^{+\infty} yf_X(m-y)dy = m \cdot 1 - 0 = m, \quad (2.92)$$

dove il primo integrale è pari a uno poiché trattasi dell'integrale della pdf su tutto l'asse reale, e il secondo è nullo poiché la funzione integranda è dispari.

Proprietà 2.11 La media di una variabile aleatoria che assume valori solo nell'intervallo (a, b) , può essere espressa direttamente in funzione della cdf come

$$E[X] = xF_X(x) \Big|_a^b - \int_a^b F_X(\xi)d\xi. \quad (2.93)$$

Prova: Diretta conseguenza della definizione e della integrazione per parti.

Andando al limite, la proprietà vale anche se $a = -\infty$ e/o $b = +\infty$

$$E[X] = \lim_{a, b \rightarrow \mp\infty} \left(xF_X(x)|_a^b - \int_a^b F_X(\xi)d\xi \right). \quad (2.94)$$

2.8.2 La moda e la mediana

La media rappresenta il baricentro di una distribuzione. Esistono comunque altri parametri per caratterizzare il *centro* di una distribuzione.

La moda mo di una variabile aleatoria è il valore in cui la pdf (o la pmf), attinge il massimo.

Si noti che in una distribuzione si può avere anche più di un punto di massimo poichè una densità di probabilità può anche essere *multi-modale*, ovvero presentare più di un picco. Tali densità non sono state ancora considerate in queste note e una discussione su di esse è rinviata a successivi approfondimenti.

La mediana di una variabile aleatoria è quel valore m , tale che

$$Pr\{X \leq m\} = Pr\{X > m\} = \frac{1}{2}. \quad (2.95)$$

La condizione può essere espressa in forma integrale

$$\int_{-\infty}^m f_X(\xi)d\xi = \int_m^{+\infty} f_X(\xi)d\xi = \frac{1}{2}. \quad (2.96)$$

Si noti come in densità simmetriche attorno alla media, la mediana coincide con la media. Inoltre se la distribuzione è anche unimodale la moda coincide con la media. La figura 2.30 mostra tre diverse pdf con le rispettive medie, mode e mediane. Va notato come in variabili aleatorie discrete l'equazione caratterizzante la mediana potrebbe non avere soluzione. E' abbastanza semplice immaginare una suddivisione in due sottoinsiemi ordinati dei valori della pmf tale da non totalizzare esattamente $\frac{1}{2}$ per parte. Si veda il terzo esempio nella figura 2.30. In tali casi si può prendere come mediana il punto in cui la massa di probabilità

puo' essere divisa in parte a destra e in parte a sinistra. Nei casi in cui si totalizza esattamente $\frac{1}{2}$ dividendo i punti dello spazio campione in due, si può assumere come mediana il punto medio anche se non incluso nello spazio campione. Esaminiamo alcuni esempi.

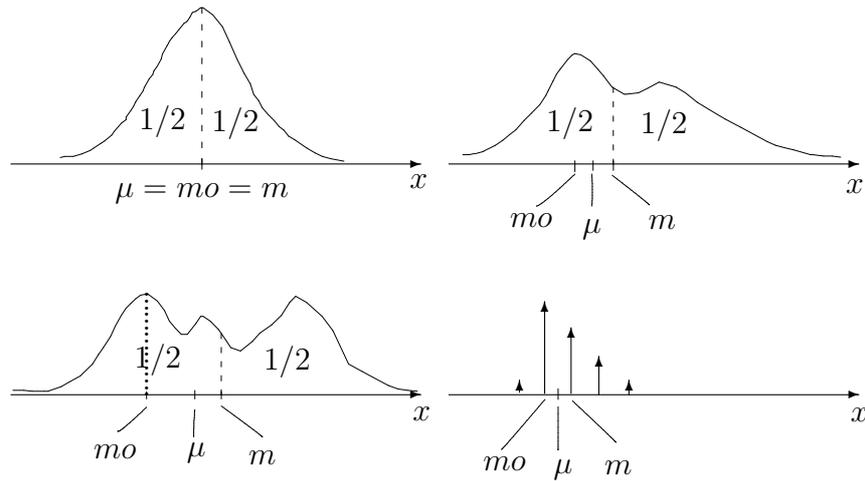


Figura 2.30: Varie pdf e le relative medie, mode e mediane.

Variabile aleatoria gaussiana: La moda, la mediana e la media coincidono e sono pari a μ . Questo è immediato poiché la densità di probabilità è simmetrica e unimodale.

Variabile aleatoria esponenziale: La moda è $mo = 0$, la mediana può essere valutata dall'equazione

$$\int_0^m \alpha e^{-\alpha\xi} d\xi = \int_m^{+\infty} \alpha e^{-\alpha\xi} d\xi. \quad (2.97)$$

Ovvero

$$\frac{e^{-\alpha\xi}}{-\alpha} \Big|_0^m = \frac{e^{-\alpha\xi}}{-\alpha} \Big|_m^{+\infty} \quad ; \quad e^{-\alpha m} - 1 = -e^{-\alpha m} \quad ; \quad m = \frac{1}{\alpha} \ln 2. \quad (2.98)$$

Si noti il risultato in contrasto con la media che è pari a $\frac{1}{\alpha}$.

Variabile aleatoria binomiale: Si consideri il caso con $n = 7$ e $p = 0.5$. Come si può evincere da figura 2.24, per simmetria, la mediana è 3 o 4. In questo caso, così come si fa di solito in situazioni analoghe, si assume che la mediana sia il punto medio $m = 3.5$. Anche la moda è pari a 3 o 4 e la assunzione sul punto medio può essere analoga.

2.8.3 La media come operatore

La media rappresenta la somma dei valori assunti dalla variabile aleatoria pesati secondo la loro misura di probabilità. Tale operazione può essere considerata come operatore $E[\cdot]$ e applicata ad altre funzioni di X . Ad esempio sia $Y = g(X)$ una variabile aleatoria funzione di X . La media di Y è

$$E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx. \quad (2.99)$$

2.8.4 I momenti di ordine superiore

I momenti di ordine superiore della variabile aleatoria X sono le medie delle potenze intere di X , ovvero si definisce momento di ordine n (≥ 1) di X

$$\mu^n \triangleq E[X^n] = \int_{-\infty}^{+\infty} \xi^n f_X(\xi)d\xi. \quad (2.100)$$

Tale definizione include la media per $n = 1$, che per semplicità denoteremo con μ piuttosto che con μ^1 . Di particolare rilevanza è il momento del secondo ordine $\mu^2 = E[X^2]$ detto *media quadratica*.

2.8.5 I momenti centrali

La media di una variabile aleatoria rappresenta il baricentro della sua distribuzione. Pertanto è utile considerare i momenti di ordine superiore della variabile dopo aver rimosso la media, ovvero *attorno alla media*.

Si definisce momento centrale di ordine n ($n > 1$), di una variabile aleatoria X ,

$$\sigma_x^n \triangleq E[(X - \mu)^n]. \quad (2.101)$$

La varianza e la deviazione standard

Assume significato particolare il momento centrale del secondo ordine, che è denominato *varianza* della variabile aleatoria

$$\sigma_x^2 = VAR[X] = E[(X - \mu)^2]. \quad (2.102)$$

Tale parametro rappresenta una misura di dispersione attorno alla media μ trattandosi della media della funzione $(X - \mu)^2$ che è la distanza euclidea al quadrato di X dalla media. Tale parametro assume particolare importanza nelle applicazioni, specialmente per variabili aleatorie aventi distribuzioni unimodali e simmetriche. La radice quadrata della varianza è anche detta *deviazione standard*

$$\sigma_x = \sqrt{VAR[X]} = \sqrt{E[(X - \mu)^2]}, \quad (2.103)$$

La deviazione standard è utile per misurare la dispersione nelle stesse unità dimensionali della variabile aleatoria X . Si noti come la varianza coincida con il momento del secondo ordine quando la variabile ha media nulla.

La seguente formula generale lega la media quadratica $E[X^2]$ alla varianza, anche quando la media non è nulla

$$\sigma_x^2 = E[X^2] - \mu^2. \quad (2.104)$$

Prova: La prova è immediata conseguenza della definizione

$$\begin{aligned} \sigma_x^2 &= \int_{-\infty}^{+\infty} (\xi - \mu)^2 f_X(\xi) d\xi \\ &= \int_{-\infty}^{+\infty} \xi^2 f_X(\xi) d\xi + \mu^2 \int_{-\infty}^{+\infty} f_X(\xi) d\xi - 2\mu \int_{-\infty}^{+\infty} \xi f_X(\xi) d\xi \\ &= E[X^2] + \mu^2 - 2\mu^2 = E[X^2] - \mu^2. \end{aligned}$$

Variabile aleatoria uniforme: I momenti di ordine superiore della variabile aleatoria uniforme di tipo continuo nell'intervallo $[a, b]$, sono facilmente calcolabili. Infatti dalla definizione, abbiamo

$$E[X^n] = \int_a^b \frac{1}{b-a} \xi^n d\xi = \frac{1}{b-a} \frac{\xi^{n+1}}{n+1} \Big|_a^b = \frac{b^{n+1} - a^{n+1}}{(b-a)(n+1)}. \quad (2.105)$$

Tipico il caso in cui la distribuzione è simmetrica attorno all'origine con $a = -b = \frac{\Delta}{2}$ e $n = 2$

$$E[X^2] = \frac{b^2}{3} = \frac{\Delta^2}{12}. \quad (2.106)$$

In tal caso il risultato è anche la varianza della variabile aleatoria poiché la media è nulla. La deviazione standard è pertanto $\frac{\Delta}{2\sqrt{3}}$.

Per il caso discreto non esiste una forma chiusa per qualunque valore di n . E' possibile comunque trovare una espressione per alcuni valori di n . Valutiamo il momento del secondo ordine

$$E[X^2] = \frac{1}{N} \sum_{k=0}^{N-1} (\alpha + k\Delta)^2. \quad (2.107)$$

Dai due seguenti risultati notevoli

$$\sum_{k=0}^n k = \frac{n(n+1)}{2}, \quad (2.108)$$

$$\sum_{k=0}^n k^2 = \frac{n(n+1)(2n+1)}{6}, \quad (2.109)$$

abbiamo che

$$\begin{aligned} E[X^2] &= \frac{1}{N} \sum_{k=0}^{N-1} \alpha^2 + \frac{2\alpha\Delta}{N} \sum_{k=0}^{N-1} k + \frac{\Delta^2}{N} \sum_{k=0}^{N-1} k^2 \\ &= \alpha^2 + \frac{2\alpha\Delta}{N} \frac{(N-1)N}{2} + \frac{\Delta^2}{N} \frac{(N-1)N(2(N-1)+1)}{6} \\ &= \alpha^2 + \alpha\Delta(N-1) + \frac{\Delta^2}{6}(2(N-1)^2 + (N-1)). \end{aligned} \quad (2.110)$$

Interessante anche qui è il caso della variabile aleatoria simmetrica attorno all'origine con $\alpha = -\frac{(N-1)\Delta}{2}$. Il momento del secondo ordine, che coincide con la varianza è

$$E[X^2] = \frac{(N^2-1)\Delta^2}{12}. \quad (2.111)$$

Variabile aleatoria gaussiana: Si consideri la variabile aleatoria gaussiana a media nulla e parametro σ . I momenti di ordine n sono definiti come

$$E[X^n] = \int_{-\infty}^{\infty} \xi^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\xi^2}{2\sigma^2}} d\xi. \quad (2.112)$$

C'è un modo semplice per arrivare ad una espressione chiusa per qualunque ordine n . Si noti prima di tutto come per n dispari la funzione integranda sia una funzione dispari. Pertanto $E[X^n] = 0, \forall n$ dispari. Si riscriva ora la condizione

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\xi^2}{2\sigma^2}} d\xi = 1, \quad (2.113)$$

dopo la sostituzione $\alpha = \frac{1}{2\sigma^2}$, come

$$\int_{-\infty}^{\infty} e^{-\alpha\xi^2} d\xi = \sqrt{\frac{\pi}{\alpha}}. \quad (2.114)$$

Derivando k volte ambo i membri dell'equazione rispetto ad α (visto che le condizioni per cui la derivata può essere invertita con il segno di integrale sono soddisfatte) abbiamo che

$$\int_{-\infty}^{\infty} \xi^{2k} e^{-\alpha\xi^2} d\xi = \frac{1 \cdot 3 \cdot \dots \cdot (2k-1)}{2^k} \sqrt{\frac{\pi}{\alpha^{2k+1}}}. \quad (2.115)$$

Risostituendo $\alpha = \frac{1}{2\sigma^2}$ e usando la definizione (2.112)

$$E[X^n] = \begin{cases} 1 \cdot 3 \cdot \dots \cdot (n-1)\sigma^n & \text{per } n \text{ pari,} \\ 0 & \text{per } n \text{ dispari.} \end{cases} \quad (2.116)$$

Si noti come per $n = 2$ abbiamo $E[X^2] = \sigma^2$. Ecco dimostrato ciò che era stato anticipato, ovvero che σ^2 e σ sono proprio varianza e deviazione standard della variabile aleatoria gaussiana. E' evidente come nel caso di variabile aleatoria a media non nulla, σ^2 rappresenti la media quadratica della variabile a media nulla $(X - \mu)$.

2.9 Disugliaglianza di Chebyshev

La varianza rappresenta un misura di dispersione attorno alla media. In particolare, se la variabile aleatoria è gaussiana possiamo valutare la probabilità che la variabile aleatoria assuma un valore che dista dalla media di una quantità maggiore di un dato valore α

$$Pr\{|X - \mu| \geq \alpha\} = 2Q\left(\frac{\alpha}{\sigma}\right), \quad (2.117)$$

che riscritta con $\alpha = k\sigma$, fornisce la utile relazione

$$Pr\{|X - \mu| \geq k\sigma\} = 2Q(k). \quad (2.118)$$

Si ricordi come la funzione $Q(\cdot)$ sia una funzione decrescente del suo argomento. E' interessante approfondire su questo argomento perché è possibile generalizzare la formula con una maggiorazione che vale per qualunque densità mediante la

Teorema 2.1 Disuguaglianza di Chebyshev: *Per qualunque variabile aleatoria X che ha varianza finita σ_X^2 e media μ ,*

$$Pr\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2}. \quad (2.119)$$

La maggiorazione di Chebyshev ci consente di garantire che, indipendentemente dalla distribuzione, solo una percentuale più piccola di $\frac{1}{k^2}$ dei valori della variabile aleatoria cade al di fuori dell'intervallo $[\mu - k\sigma, \mu + k\sigma]$. Si noti come sia necessario che $k > 1$ affinché la maggiorazione abbia una qualche utilità: se $k < 1$ avremmo la inutile maggiorazione con un numero maggiore di uno.

Prova: La prova è piuttosto semplice e si può ottenere partendo dalla definizione di varianza e tramite le seguenti maggiorazioni

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (\xi - \mu)^2 f_X(\xi) d\xi \geq \int_{|\xi - \mu| \geq k\sigma} (\xi - \mu)^2 f_X(\xi) d\xi \\ &\geq k^2 \sigma^2 \int_{|\xi - \mu| \geq k\sigma} f_X(\xi) d\xi = k^2 \sigma^2 Pr\{|X - \mu| \geq k\sigma\}. \end{aligned} \quad (2.120)$$

Le maggiorazioni che si effettuano per arrivare alla disuguaglianza di Chebyshev ci consentono di ottenere un risultato ancora più generale che consente di legare la probabilità di trovarsi ad una data distanza dalla media a tutti i momenti centrali di ordine superiore. Il risultato generale è noto come

Teorema 2.2 Disuguaglianza di Bienaymé: *Per qualunque variabile aleatoria X ,*

$$Pr\{|X - a| \geq \epsilon\} \leq \frac{E[|X - a|^n]}{\epsilon^n}, \quad (2.121)$$

dove a è un punto qualunque dell'asse reale.

Prova: Si consideri la variabile aleatoria Y che assume solo valori positivi, ovvero $f_Y(y) = 0 \forall y < 0$. Sia α una costante positiva assegnata, dimostriamo prima che

$$Pr\{Y \geq \alpha\} \leq \frac{E[Y]}{\alpha}. \quad (2.122)$$

Tale risultato è anche noto come *disuguaglianza di Markov*. La prova è immediata poiché

$$E[Y] = \int_0^{\infty} \xi f_Y(\xi) d\xi > \int_{\alpha}^{\infty} \xi f_Y(\xi) d\xi \geq \alpha \int_{\alpha}^{\infty} f_Y(\xi) d\xi = \alpha Pr\{Y \geq \alpha\}. \quad (2.123)$$

Ponendo $Y = |X - a|^n$, e $\alpha = \epsilon^n$, abbiamo

$$Pr\{|X - a|^n \geq \epsilon^n\} \leq \frac{E[|X - a|^n]}{\epsilon^n}. \quad (2.124)$$

Pertanto

$$Pr\{|X - a| \geq \epsilon\} \leq \frac{E[|X - a|^n]}{\epsilon^n}. \quad (2.125)$$

La disuguaglianza di Chebyshev si ottiene per $a = \mu$, $n = 2$ e $\epsilon = k\sigma$. Altre maggiorazioni con i momenti centrali si ottengono con $a = \mu$.

Esempio 2.16 Si consideri un impianto di produzione di un manufatto che deve misurare 1.5 metri. Il controllo qualità valuta che la lunghezza media è di 1.5 metri e che la deviazione standard è di 3 centimetri. Si stimi la probabilità che la misura di un pezzo scelto a caso sia al di fuori dell'intervallo [1.4 1.6] metri. **Soluzione:** Per utilizzare la disuguaglianza di Chebyshev, poniamo $k\sigma = 0.1$ con $\sigma = 0.03$. Pertanto $k = 10/3$ e si ha

$$Pr\{|X - 1.5| \geq 0.1\} \leq \frac{9}{100} = 0.09. \quad (2.126)$$

Al più il nove per cento dei pezzi sarà al fuori dell'intervallo. La conoscenza della densità di probabilità ci consentirebbe una valutazione più precisa. Infatti se X fosse modellabile come una variabile gaussiana avremmo

$$Pr\{|X - 1.5| \geq 0.1\} = 2Q\left(\frac{0.1}{0.03}\right) = 2Q\left(\frac{10}{3}\right) = 8.58 \cdot 10^{-4}. \quad (2.127)$$

Si noti come il valore relativo alla maggiorazione di Chebyshev può essere anche molto conservativo rispetto al reale valore della probabilità cercata.

2.10 Problemi

Problema 2.1 Una moneta truccata ha spazio campione $S = \{0, 1\}$ e probabilità $\Pi = \{0.3, 0.7\}$. Calcolare cdf e pdf della variabile aleatoria X pari alla somma aritmetica dei risultati di tre lanci della moneta.

Problema 2.2 Si consideri un esperimento il cui risultato è uno dei valori $\{-1, 0, 1\}$ con probabilità $\{\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\}$ rispettivamente. La variabile aleatoria X è la somma dei risultati di due prove indipendenti dell'esperimento. Calcolare (a) Pdf e cdf di X ; (b) $Pr\{X > 0\}$.

Problema 2.3 Le conversazioni telefoniche di Concetta durano in media 60 secondi e la loro durata è modellata come una variabile aleatoria esponenziale. La fatturazione delle telefonate prevede che alla risposta si paghino 100 lire, e che ogni 30 secondi si paghi uno "scatto" di 50 lire. Sia C la variabile aleatoria che descrive il costo della telefonata. (a) Si valuti cdf e pdf di C ; (b) quanto spende in media Concetta?

Problema 2.4 Una tensione V , distribuita secondo una densità gaussiana a media nulla e varianza $\sigma^2 = 4 \text{ Volt}^2$, viene misurata due volte in maniera indipendente. Calcolare (a) la probabilità che in entrambe le misure la tensione sia superiore ai 10 Volt; (b) la probabilità che nella prima misura sia tra -1 e 1 Volt, e nella seconda misura sia tra -2 e +2 Volt.

Problema 2.5 Un pacchetto trasmesso da un computer all'istante $t_0 = 0$ impiega 1 microsecondo a raggiungere un server che impiega un tempo aleatorio pari a T secondi per gestirlo e ritrasmetterlo. Si assuma che T sia distribuito secondo una pdf esponenziale con media pari a 100 microsecondi. Il computer destinatario riceverà il pacchetto dopo un ulteriore ritardo di propagazione di 10 microsecondi. (a) Valutare la pdf e la cdf dell'istante T_a di arrivo a destinazione del pacchetto; (b) Tale pacchetto è da considerarsi perso se raggiunge il destinatario dopo un certo tempo massimo. Calcolare la probabilità di perdita se il ritardo massimo è di 400 microsecondi.

Problema 2.6 Esprimere in termini di cdf e pdf le seguenti probabilità $Pr\{a \leq X \leq b\}$; $Pr\{X \geq b\}$; $Pr\{X > a\}$.

Problema 2.7 Il controllo qualità di una fabbrica di resitori da 100Ω ci garantisce che il valore reale delle resistenze prodotte è in media esatto, ma oscilla attorno al valore nominale con una deviazione standard $\sigma = 1.5\Omega$. (a) Qual è il valore di tolleranza ($\pm\Delta$) che posso garantire al 10% senza fare ipotesi sulla distribuzione? (b) Qual'è il valore di tolleranza che posso garantire al 10% se assumo che la resistenza sia distribuita secondo una pdf gaussiana?

Problema 2.8 Sia X un campione di un segnale gaussiano a media nulla e deviazione standard $\sigma = 10$. Il segnale può essere scalato di un fattore α prima di essere inviato in un dispositivo avente la classica caratteristica di saturazione

$$Y = \phi(\alpha X) = \begin{cases} -1 & \alpha X \leq -1 \\ \alpha X & -1 < \alpha X \leq 1 \\ 1 & \alpha X > 1 \end{cases} \quad (2.128)$$

Determinare il valore del guadagno α che limita la saturazione al 5%.

Problema 2.9 Una variabile aleatoria di tipo misto X ha la seguente pdf

$$f_X(x) = g(x) + \beta\delta(x - 5), \quad (2.129)$$

dove

$$g(x) = \begin{cases} \alpha e^{-x} & 0 < x \leq 3 \\ 0 & \text{altrove.} \end{cases} \quad (2.130)$$

Determinare α e β tali che la mediana di X sia uguale a 2.

Problema 2.10 La *variabile aleatoria di Erlang* ha una pdf del tipo

$$f_X(x) = \frac{\alpha^n}{(n-1)!} x^{n-1} e^{-\alpha x} u(x) = \text{Erl}(\alpha, n). \quad (2.131)$$

La pdf di Erlang generalizza la pdf esponenziale che si ottiene per $n = 1$.

(a) Dimostrare per induzione che la cdf è

$$F_X(x) = \left(1 - e^{-\alpha x} \sum_{j=0}^{n-1} \frac{(\alpha x)^j}{j!} \right) u(x); \quad (2.132)$$

- (b) Dimostrare che la media è $E[X] = \frac{n}{\alpha}$;
 (c) Dimostrare che i momenti (non centrali) sono dati dalla formula

$$E[X^j] = \frac{(n+j-1)!}{\alpha^j (n-1)!}; \quad (2.133)$$

- (d) Dimostrare che la varianza è

$$VAR[X] = \frac{n}{\alpha^2}. \quad (2.134)$$

Problema 2.11 La *variabile aleatoria di Cauchy* ha pdf

$$f_X(x) = \frac{1}{\pi\beta \left(1 + \frac{x^2}{\beta^2}\right)}. \quad (2.135)$$

- (a) Dimostrare che $f_X(x)$ è una valida pdf;
 (b) Dimostrare che la varianza è infinita.

Problema 2.12 La *variabile aleatoria discreta geometrica* ha una pmf

$$Pr\{X = n\} = pq^{n-1} = p_X[n], \quad n = 1, 2, 3, \dots \quad (2.136)$$

- (a) Si verifichi che $p_X[n]$ è una valida pmf;
 (b) Dimostrare che la media è $E[X] = \frac{1}{p}$;
 (c) Dimostrare che la varianza è $VAR[X] = \frac{q}{p^2}$.

Capitolo 3

Due variabili aleatorie

In questo capitolo viene introdotta la formalizzazione probabilistica congiunta per due variabili aleatorie mediante cdf e pdf congiunte. Vengono inoltre definite le pdf e cdf condizionate, i momenti congiunti e i momenti condizionati.

3.1 Introduzione

Nello studio dei fenomeni aleatori, è spesso necessario tenere in conto più di una grandezza fisica e caratterizzarne le relazioni di dipendenza reciproca. Tali variabili vanno studiate congiuntamente ai fini della caratterizzazione delle loro relazioni di dipendenza in termini di cdf, pdf e momenti congiunti.

Lo schema di riferimento è quello di figura 3.1 in cui ad ogni elemento dello spazio campione viene associata una coppia di variabili reali.

Ogni valore di tale coppia può essere visualizzato come un punto su un piano. Nel modello più comune, le due variabili X e Y sono i valori assunti dalle due coordinate cartesiane. Altre associazioni sono ovviamente possibili, come una coppia di variabili polari, una coppia di variabili curvilinee, ecc. La caratterizzazione congiunta delle due variabili aleatorie sarà quindi una funzione di x e di y . Il vettore $\mathbf{Z} = (X, Y)$ delle due variabili aleatorie è detto *variabile aleatoria bidimensionale*.

Per meglio motivare la introduzione delle due variabili aleatorie consideriamo un esempio.

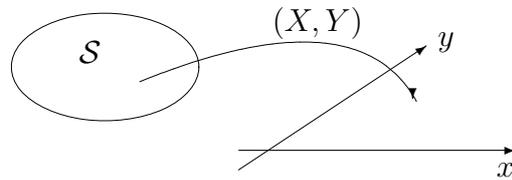


Figura 3.1: La coppia di variabili aleatorie X e Y

Esempio 3.1 Si consideri l'esperimento di misura congiunta dell'altezza e dell'età di una popolazione di individui. Sia X la variabile aleatoria che rappresenta l'altezza di un individuo scelto a caso, misurata in metri, e Y la sua età espressa in anni. La coppia (x, y) sul piano corrisponde a un punto dello spazio campione.

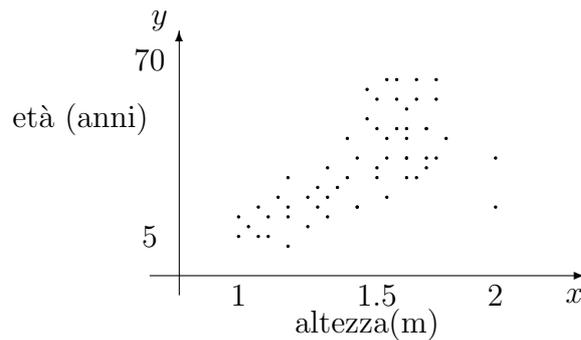


Figura 3.2: Il piano altezza-età con alcuni punti relativi ad una popolazione finita.

Figura 3.2 mostra alcuni punti misurati su una popolazione finita di individui. Come si può facilmente evincere, i punti sembrano addensarsi attorno ad una curva, con qualche eccezione. Ciò significa che la relazione tra le due variabili non può essere totalmente determinata (o deterministica), ma necessita di un modello che tenga conto della variabilità riscontrata. E' pertanto ragionevole modellare le due grandezze

come due variabili aleatorie che hanno una mutua *dipendenza statistica* (la dipendenza statistica tra due variabili aleatorie sarà più precisamente definita nel seguito). Pertanto, con il solito ragionamento secondo il quale al crescere del numero di misure all'infinito otteniamo uno spazio campione di tipo continuo, possiamo pensare ad una misura di probabilità (densità) definita sul piano (x, y) .

3.2 La funzione cumulativa congiunta

La caratterizzazione congiunta delle due variabili è definita su domini del piano mediante una funzione di distribuzione cumulativa (cdf) congiunta ed una funzione di densità (pdf) congiunta. Le definizioni delle due cdf di X e di Y , prese separatamente e dette pertanto *cdf marginali*, sono già state date nel capitolo precedente e sono

$$F_X(x) = Pr\{X \leq x\}, \quad F_Y(y) = Pr\{Y \leq y\}. \quad (3.1)$$

La *funzione cumulativa di distribuzione congiunta* delle due variabili aleatorie è definita come

$$F_{XY}(x, y) \triangleq Pr\{X \leq x, Y \leq y\}. \quad (3.2)$$

Tale funzione, in un generico punto (x, y) del piano, rappresenta la probabilità congiunta che le variabili siano minori di x e di y rispettivamente. Figura 3.3 mostra il sottoinsieme del piano cartesiano a cui fa riferimento la definizione della cdf congiunta.

Si immagini, in analogia al caso monodimensionale, una funzione che tende a uno per x e y tendenti all'infinito e che tende a zero per x e y tendenti a meno infinito. Dalla definizione è immediato vedere come valgano le seguenti relazioni

$$F_{XY}(x, \infty) = Pr\{X \leq x, Y \leq \infty\} = Pr\{X \leq x\} = F_X(x); \quad (3.3)$$

$$F_{XY}(\infty, y) = Pr\{X \leq \infty, Y \leq y\} = Pr\{Y \leq y\} = F_Y(y); \quad (3.4)$$

$$F_{XY}(\infty, \infty) = Pr\{X \leq \infty, Y \leq \infty\} = 1; \quad (3.5)$$

$$F_{XY}(x, -\infty) = Pr\{X \leq x, Y \leq -\infty\} = 0; \quad (3.6)$$

$$F_{XY}(-\infty, y) = Pr\{X \leq -\infty, Y \leq y\} = 0; \quad (3.7)$$

$$F_{XY}(-\infty, -\infty) = 0. \quad (3.8)$$

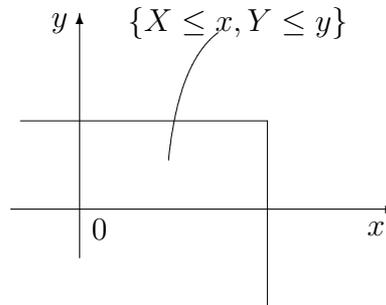


Figura 3.3: Il piano delle due variabili aleatorie X e Y e il sottoinsieme $\{X \leq x, Y \leq y\}$.

La cdf congiunta delle due variabili aleatorie può essere utilizzata, in analogia al caso monodimensionale, per valutare la probabilità che la coppia (X, Y) corrisponda ad un punto appartenente a certi domini ortogonali del piano (x, y) . I seguenti esempi possono essere facilmente interpretati con l'aiuto delle figure.

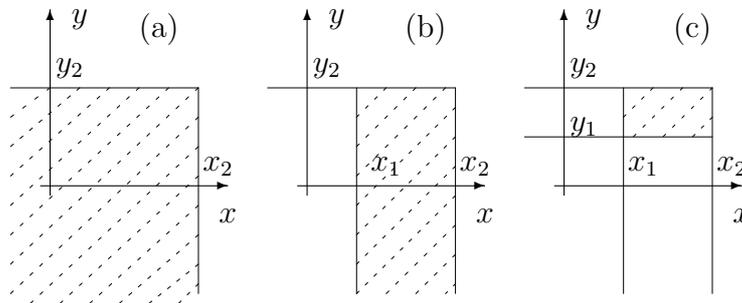


Figura 3.4: Gli insiemi per il calcolo delle probabilità mediante la cdf congiunta.

Dalla definizione di cdf congiunta abbiamo che

$$Pr\{X \leq x_2, Y \leq y_2\} = F_{XY}(x_2, y_2),$$

$$Pr\{X > x_2, Y > y_2\} = 1 - F_{XY}(x_2, y_2), \quad (3.9)$$

come è possibile osservare in figura 3.4(a).

Per calcolare la probabilità $Pr\{x_1 < X \leq x_2, Y \leq y_2\}$, si noti, con l'aiuto di figura 3.4(b), come valga la seguente decomposizione

$$\{X \leq x_2, Y \leq y_2\} = \{X \leq x_1, Y \leq y_2\} + \{x_1 < X \leq x_2, Y \leq y_2\}. \quad (3.10)$$

Pertanto, poiché gli eventi sono disgiunti, abbiamo che

$$\begin{aligned} Pr\{x_1 < X \leq x_2, Y \leq y_2\} \\ &= Pr\{X \leq x_2, Y \leq y_2\} - Pr\{X \leq x_1, Y \leq y_2\} \\ &= F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2). \end{aligned} \quad (3.11)$$

Analogamente:

$$\begin{aligned} Pr\{X \leq x_2, y_1 < Y \leq y_2\} \\ &= Pr\{X \leq x_2, Y \leq y_2\} - Pr\{X \leq x_2, Y \leq y_1\} \\ &= F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1). \end{aligned} \quad (3.12)$$

Calcoliamo ora la probabilità $Pr\{x_1 < X \leq x_2, y_1 < Y \leq y_2\}$. Si osservi con l'aiuto di figura 3.4(c) come valga la seguente decomposizione

$$\begin{aligned} \{x_1 < X \leq x_2, Y < y_2\} \\ &= \{x_1 < X \leq x_2, y_1 < Y \leq y_2\} + \{x_1 < X \leq x_2, Y \leq y_1\}. \end{aligned} \quad (3.13)$$

Pertanto usando equazione (3.11) abbiamo

$$\begin{aligned} Pr\{x_1 < X \leq x_2, y_1 < Y \leq y_2\} \\ &= F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1). \end{aligned} \quad (3.14)$$

Il calcolo della probabilità che la coppia di variabili aleatorie (X, Y) appartenga a domini di forma diversa da quelli esaminati, richiede in genere delle espressioni più complesse, a meno che tali domini non siano esprimibili come unione di domini del tipo considerato. In generale, per tali domini, si preferisce utilizzare la pdf congiunta che andremo qui di seguito a definire.

La proprietà fondamentale della cdf congiunta, così come si è visto nel caso monodimensionale, è la monotonicità. Infatti, la cdf, fissato x ,

è non decrescente in y . Analogamente, fissato y , è non decrescente in x . Infatti per ogni $\Delta y > 0$, abbiamo che

$$\begin{aligned} F_{XY}(x, y + \Delta y) &= Pr\{X \leq x, Y \leq y + \Delta y\} \\ &= Pr\{X \leq x, Y \leq y\} + Pr\{X \leq x, y < Y \leq y + \Delta y\} \\ &= F_{XY}(x, y) + Pr\{X \leq x, y < Y \leq y + \Delta y\} \\ &\geq F_{XY}(x, y). \end{aligned} \quad (3.15)$$

Analogamente

$$F_{XY}(x + \Delta x, y) \geq F_{XY}(x, y). \quad (3.16)$$

3.3 La funzione di densità congiunta

Si definisce *funzione di densità di probabilità congiunta*, o *pdf congiunta*, delle due variabili aleatorie X e Y

$$f_{XY}(x, y) \triangleq \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}. \quad (3.17)$$

Poiché la cdf congiunta è una funzione non decrescente, ovvero fissato x è non decrescente in y e viceversa, abbiamo che

$$f_{XY}(x, y) \geq 0. \quad (3.18)$$

La interpretazione della densità, analogamente al caso monodimensionale, si può ottenere osservando dalla definizione di derivate parziali che

$$f_{XY}(x, y) = \lim_{\Delta x \rightarrow 0, \Delta y \rightarrow 0} \frac{F_{XY}(x + \Delta x, y + \Delta y) - F_{XY}(x, y + \Delta y) - F_{XY}(x + \Delta x, y) + F_{XY}(x, y)}{\Delta x \Delta y}. \quad (3.19)$$

Quindi da equazione (3.14) abbiamo che

$$\lim_{\Delta x \rightarrow 0, \Delta y \rightarrow 0} \frac{Pr\{x < X \leq x + \Delta x, y < Y \leq y + \Delta y\}}{\Delta x \Delta y} = f_{XY}(x, y). \quad (3.20)$$

Pertanto il volume sotteso dalla funzione di densità di probabilità in un intorno rappresenta la probabilità che la coppia di variabili assuma valori in tale insieme

$$Pr\{x < X \leq x + dx, y < Y \leq y + dy\} = f_{XY}(x, y) dx dy. \quad (3.21)$$

Generalizzando, dato un dominio \mathcal{D} del piano (x, y) , abbiamo che:

$$Pr\{(X, Y) \in \mathcal{D}\} = \iint_{\mathcal{D}} f_{XY}(\xi, \eta) d\xi d\eta. \quad (3.22)$$

ovvero, la probabilità che la coppia di variabili ha di cadere in un dato dominio è pari al volume sotteso dalla pdf congiunta su tale dominio. La proprietà appena enunciata, che consente dalla conoscenza della pdf di calcolare la probabilità di qualunque evento sul piano (x, y) , ci consente anche di legare la pdf alla cdf: basta considerare la probabilità dell'evento $\{X \leq x, Y \leq y\}$ per concludere che

$$F_{XY}(x, y) = Pr\{X \leq x, Y \leq y\} = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(\xi, \eta) d\xi d\eta. \quad (3.23)$$

Da qui discende che

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(\xi, \eta) d\xi d\eta = 1, \quad (3.24)$$

ovvero, il volume sotteso dalla pdf congiunta è pari a uno. Derivando la cdf congiunta rispetto a x e y rispettivamente, abbiamo da equazione (3.23) che

$$\begin{aligned} \frac{\partial F_{XY}(x, y)}{\partial x} &= \int_{-\infty}^y f_{XY}(x, \eta) d\eta, \\ \frac{\partial F_{XY}(x, y)}{\partial y} &= \int_{-\infty}^x f_{XY}(\xi, y) d\xi. \end{aligned} \quad (3.25)$$

A partire dalle definizioni date, è possibile legare la pdf congiunta alle pdf marginali. Avendo già dimostrato che $F_X(x) = F_{XY}(x, \infty)$ e che $F_Y(y) = F_{XY}(\infty, y)$, abbiamo che le cdf marginali possono essere legate alla pdf congiunta dalle seguenti relazioni

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x \int_{-\infty}^{\infty} f_{XY}(\xi, \eta) d\xi d\eta, \\ F_Y(y) &= \int_{-\infty}^{\infty} \int_{-\infty}^y f_{XY}(\xi, \eta) d\xi d\eta. \end{aligned} \quad (3.26)$$

Si noti che il legame tra la cdf congiunta e quelle marginali è già stato discusso all'inizio del capitolo.

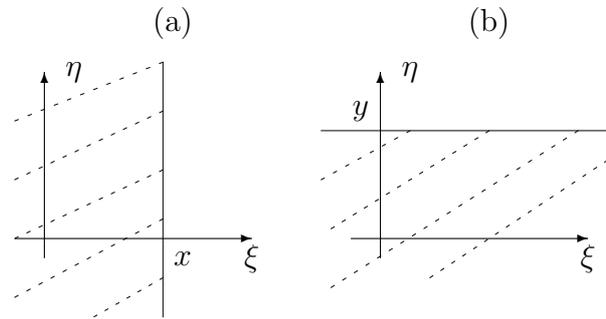


Figura 3.5: Gli insiemi per legare la cdf marginale e la pdf congiunta.

Le proprietà appena enunciate possono essere facilmente interpretate guardando alle figure 3.5(a) e (b). Ovvero, la cdf marginale di X in x , si ottiene integrando la pdf congiunta a sinistra della ascissa x . Analogamente la cdf marginale di Y in y si ottiene integrando la pdf congiunta al di sotto dell'ordinata y .

Derivando ancora le equazioni ottenute rispetto a x e a y rispettivamente, abbiamo il legame tra le pdf marginali e la pdf congiunta

$$\begin{aligned} f_X(x) &= \frac{\partial F_X(x)}{\partial x} = \int_{-\infty}^{\infty} f_{XY}(x, \eta) d\eta, \\ f_Y(y) &= \frac{\partial F_Y(y)}{\partial y} = \int_{-\infty}^{\infty} f_{XY}(\xi, y) d\xi. \end{aligned} \quad (3.27)$$

Come si vede in figura 3.6, la pdf marginale di X in x si ottiene integrando la pdf congiunta lungo la retta verticale di ascissa x . Analogamente la pdf marginale di Y in y si ottiene integrando la pdf congiunta lungo la retta orizzontale avente ordinata y .

3.3.1 Variabili discrete e miste

Nella discussione sulle cdf e sulle pdf congiunte non abbiamo fatto menzione al fatto che le cdf potrebbero essere discontinue e non derivabili. Anche per le coppie di variabili aleatorie, analogamente alla singola vari-

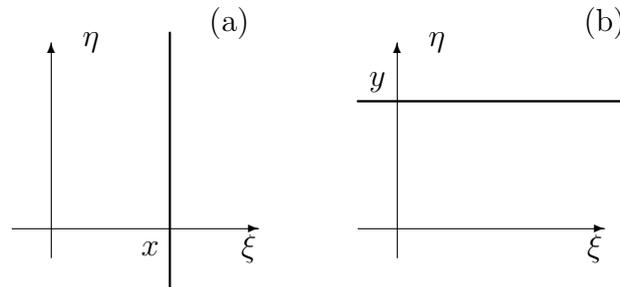


Figura 3.6: Gli insiemi per legare la pdf congiunta e le pdf marginali.

abile, è possibile utilizzare le funzioni di Dirac e includere in una formulazione unica variabili aleatorie continue, discrete e miste. Nel caso di variabili discrete o miste, maggiore attenzione va dedicata alle masse di probabilità concentrate in punti, o in sottoinsiemi degeneri. Alcuni degli esempi che saranno riportati nel seguito chiariranno la metodologia. Si noti solo che una funzione di Dirac bidimensionale è tale che

$$\delta(x - x_0, y - y_0) = \begin{cases} \infty & (x, y) = (x_0, y_0) \\ 0 & \text{altrove,} \end{cases} \quad (3.28)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi, \eta) \delta(\xi - x_0, \eta - y_0) d\xi d\eta = f(x_0, y_0). \quad (3.29)$$

Esempio 3.2 Variabile aleatoria uniforme di tipo continuo su un dominio rettangolare. Si consideri la pdf congiunta:

$$f_{XY}(x, y) = \begin{cases} K & A_1 < x \leq A_2 \text{ e } B_1 < y \leq B_2 \\ 0 & \text{altrove,} \end{cases} \quad (3.30)$$

riportata in figura 3.7.

Poiché il volume sotteso dalla pdf è pari a $(A_2 - A_1)(B_2 - B_1)K$ e deve essere pari a uno, il valore della costante K deve essere pari a

$$K = \frac{1}{(A_2 - A_1)(B_2 - B_1)}. \quad (3.31)$$

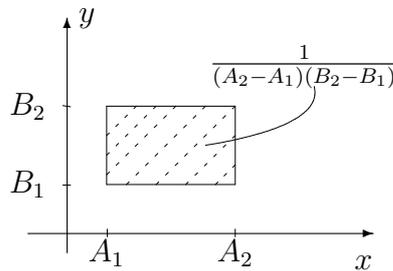


Figura 3.7: La pdf congiunta per la variabile aleatoria uniforme di tipo continuo

Le seguenti probabilità sono facilmente ottenibili tracciando graficamente i domini corrispondenti:

a) $Pr\{X > A_1\} = 1.$

b) $Pr\{a < X \leq b\}$ con $A_1 < a < b < A_2$. Tale probabilità è pari al volume sotteso dal dominio definito dall'evento. Ovvero, $Pr\{a < X \leq b\} = (b - a)(B_2 - B_1)K = \frac{(b-a)}{(A_2-A_1)}.$

c) $Pr\{a < X \leq b, c < Y \leq d\}$ con $A_1 < a < b < A_2$ e $B_1 < c < d < B_2$. Analogamente al caso precedente abbiamo: $Pr\{a < X \leq b, c < Y \leq d\} = (b - a)(d - c)K = \frac{(b-a)(d-c)}{(A_2-A_1)(B_2-B_1)}.$

Esempio 3.3 Una variabile aleatoria discreta. Si consideri la variabile aleatoria discreta definita sui punti di in figura 3.8. Su tutti i punti indicati nel piano (x, y) è contenuta la stessa massa di probabilità. Trattandosi di 35 punti, la massa su ogni punto è pari a $1/35$. Formalmente la pdf ha una delta di Dirac con area pari a $1/35$ su ogni punto.

Le seguenti probabilità sono facilmente ottenibili tracciando graficamente i domini relativi sul piano (x, y) e verificando il numero di punti inclusi.

a) $Pr\{X \geq 5\} = 15/35 = 3/7.$

b) $Pr\{3 < X \leq 4, y > 5\} = 1/35.$

c) $Pr\{2 < X \leq 5, 1.5 \leq Y < 5\} = 9/35.$

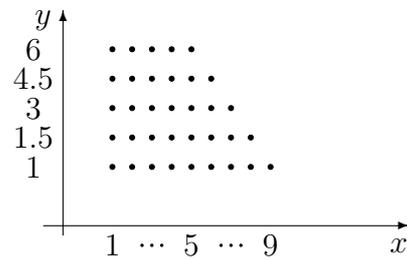


Figura 3.8: La pdf congiunta per una variabile aleatoria di tipo discreto

Esempio 3.4 Una variabile aleatoria con una massa concentrata su un sottospazio lineare. Si consideri una pdf bidimensionale diversa da zero dappertutto, eccetto che sulla retta tracciata in figura 3.9.

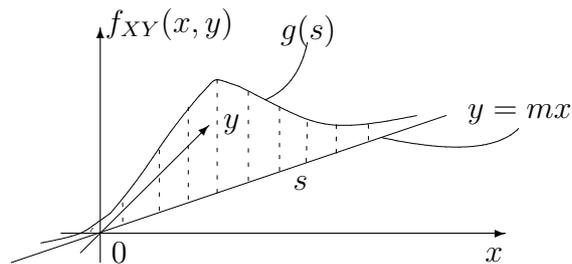


Figura 3.9: La pdf congiunta per una variabile aleatoria con massa concentrata su una retta

La figura mostra come la pdf sia una “sottile” funzione definita sul dominio lineare costituito dalla retta. Se la retta ha equazione $y = mx$, la coordinata lineare sulla retta è $s = x\sqrt{1 + m^2}$. Quindi la pdf può essere scritta formalmente come

$$f_{XY}(x, y) = g(s)\delta(x, y - mx). \quad (3.32)$$

Ovviamente la funzione $g(s)$ è una pdf monodimensionale e deve soddisfare il vincolo

$$\int_{-\infty}^{\infty} g(s) ds = 1. \quad (3.33)$$

Le probabilità di eventi corrispondenti a vari domini del piano possono essere facilmente ottenute per integrazione bidimensionale.

3.3.2 La variabile aleatoria gaussiana bidimensionale

La variabile aleatoria gaussiana assume anche nel caso bidimensionale un particolare significato applicativo. La tipica forma a campana si sviluppa in due dimensioni sul piano (x, y) .

Due variabile aleatorie X e Y si dicono congiuntamente gaussiane se la loro pdf congiunta è data dall'espressione

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \cdot \exp \left[-\frac{1}{2(1-r^2)} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2r(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right]. \quad (3.34)$$

Le pdf marginali sono le due gaussiane monodimensionali

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left[-\frac{(x-\mu_1)^2}{2\sigma_1^2} \right], \\ f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left[-\frac{(y-\mu_2)^2}{2\sigma_2^2} \right]. \end{aligned} \quad (3.35)$$

La dimostrazione è immediata riscrivendo l'esponente come

$$\begin{aligned} &\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2r(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} = \\ &= \left[\frac{x-\mu_1}{\sigma_1} - \frac{r(y-\mu_2)}{\sigma_2} \right]^2 + \frac{(1-r^2)(y-\mu_2)^2}{\sigma_2^2}. \end{aligned} \quad (3.36)$$

Quindi la pdf marginale di Y è

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \cdot \exp \left[-\frac{(y-\mu_2)^2}{2\sigma_2^2} \right] \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2(1-r^2)} \left[\frac{x-\mu_1}{\sigma_1} - \frac{r(y-\mu_2)}{\sigma_2} \right]^2 \right] dx. \quad (3.37)$$

Effettuando nell'integrale il cambio di variabili

$$z = \frac{1}{\sqrt{1-r^2}} \left(\frac{x-\mu_1}{\sigma_1} - r \frac{y-\mu_2}{\sigma_2} \right), \quad (3.38)$$

abbiamo che

$$f_Y(y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \cdot \exp \left[-\frac{(y-\mu_2)^2}{2\sigma_2^2} \right] \int_{-\infty}^{\infty} \exp \left[-\frac{z^2}{2} \right] dz \sigma_1 \sqrt{1-r^2}. \quad (3.39)$$

Riconoscendo l'integrale come l'area sottesa da una gaussian unitaria

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left[-\frac{(y-\mu_2)^2}{2\sigma_2^2} \right], \quad (3.40)$$

che è il risultato cercato. Analogamente si può ottenere $f_X(x)$.

Da quanto sopra esposto, i parametri μ_1 , μ_2 , σ_1 e σ_2 rappresentano medie e varianze (marginali) delle due variabili aleatorie X e Y . Il parametro r , di cui parleremo più approfonditamente nel seguito, è quello che collega le due variabili ed è detto *coefficiente di correlazione*. Per ora si noti che se $r = 0$,

$$f_{XY}(x, y) = f_X(x)f_Y(y). \quad (3.41)$$

L'andamento della gaussiana bidimensionale è ancora quello di una funzione a campana centrata alla coordinata (μ_1, μ_2) . In figura 3.10 sono mostrate le curve di livello relative. Si noti come esse siano delle ellissi con assi principali ruotati. Infatti la equazione $f_{XY}(x, y) = C$, può essere riscritta come

$$\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2r(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} = C_1, \quad (3.42)$$

dove C_1 è una costante. I valori μ_1 e μ_2 determinano la posizione centrale delle ellissi, mentre i parametri σ_1 , σ_2 e r ne determinano l'orientamento e il rapporto tra gli assi principali.

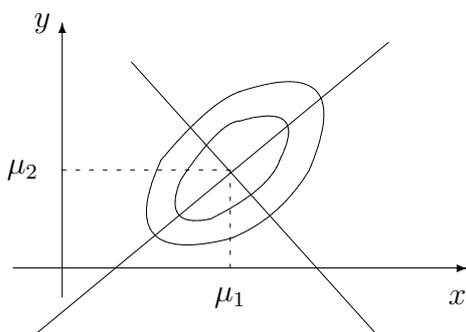


Figura 3.10: Le curve di livello di una pdf gaussiana bidimensionale

3.4 Variabili aleatorie indipendenti

Si ricordi che due eventi A e B si dicono indipendenti se $P(AB) = P(A)P(B)$. Analogamente, due variabili aleatorie si dicono *indipendenti* se gli eventi $\{X \leq x\}$ e $\{Y \leq y\}$ sono indipendenti. Pertanto se X e Y sono indipendenti,

$$F_{XY}(x, y) = Pr\{x \leq X, Y \leq y\} = F_X(x)F_Y(y). \quad (3.43)$$

Inoltre se le pdf esistono, abbiamo che

$$f_{XY}(x, y) = f_X(x)f_Y(y). \quad (3.44)$$

In altre parole la indipendenza di due variabili aleatorie comporta la separabilità delle funzioni di distribuzione e di densità congiunte. La definizione si applica ovviamente anche alle densità discrete dove la funzione di massa di probabilità bidimensionale può essere scomposta nel prodotto di due funzioni monodimensionali. Se si vuole calcolare la probabilità di un evento definito in un dominio separabile, e le variabili aleatorie sono indipendenti, abbiamo che

$$\begin{aligned} Pr\{x_1 < X \leq x_2, y_1 < Y < y_2\} &= \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{XY}(\xi, \eta) d\xi d\eta \\ &= \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_X(\xi) f_Y(\eta) d\xi d\eta = \int_{x_1}^{x_2} f_X(\xi) d\xi \int_{y_1}^{y_2} f_Y(\eta) d\eta \\ &= Pr\{x_1 < X \leq x_2\} Pr\{y_1 < Y < y_2\}. \end{aligned} \quad (3.45)$$

3.5 Cdf e pdf condizionate

La caratterizzazione probabilistica di una variabile aleatoria può essere condizionata dalla occorrenza di un altro evento. Si pensi all'esempio, più volte riportato, della variabile aleatoria che descrive la temperatura ambiente. La distribuzione della temperatura può essere molto diversa se si tratta di una temperatura misurata in un interno o un esterno, o magari in una giornata di bello o brutto tempo. Tali eventi condizionanti possono essere tenuti in conto mediante la definizione di cdf e pdf condizionate.

Si consideri un fenomeno aleatorio descritto dallo spazio campione \mathcal{A} e si denoti con A un generico evento. Si consideri inoltre una variabile aleatoria X , e si denoti con x il generico valore reale da essa assunto. Si definisce *funzione di distribuzione cumulativa di X condizionata all'evento A*

$$F_X(x|A) \triangleq Pr\{X \leq x|A\}. \quad (3.46)$$

Analogamente si definisce *funzione di densità di probabilità condizionata ad A*

$$f_X(x|A) \triangleq \frac{d}{dx} F_X(x|A). \quad (3.47)$$

Per le definizioni date valgono tutte le considerazioni fatte per la variabile aleatoria non condizionata. La sola differenza con la variabile aleatoria non condizionata, è la presenza dell'evento condizionante A che è tenuto in conto mediante il condizionamento nella probabilità che definisce la cdf. Si noti inoltre come formalmente lo spazio campione \mathcal{A} giochi lo stesso ruolo della variabile aleatoria X , mentre l'evento A , che rappresenta il generico elemento dello spazio \mathcal{A} , va visto in analogia al generico valore x assunto da X . Si ricordi che la definizione di probabilità condizionata all'evento A si richiede che $Pr\{A\} \neq 0$ per avere senso.

E' possibile definire anche una cdf congiunta (di tipo misto) come

$$F_X(x, A) = Pr\{X \leq x, A\} = Pr\{X \leq x|A\}Pr\{A\} = F_X(x|A)Pr\{A\}. \quad (3.48)$$

Analogamente, è naturale definire una pdf congiunta (mista) del tipo

$$f_X(x, A) = \frac{d}{dx} F_X(x, A) = \left(\frac{d}{dx} F_X(x|A) \right) Pr\{A\} = f_X(x|A)Pr\{A\}. \quad (3.49)$$

Le relazioni con la cdf e la pdf non condizionata, per la regola della probabilità totale sono

$$\begin{aligned} F_X(x) &= \sum_{A \in \mathcal{A}} F_X(x, A) = \sum_{A \in \mathcal{A}} F_X(x|A)Pr\{A\}, \\ f_X(x) &= \sum_{A \in \mathcal{A}} f_X(x, A) = \sum_{A \in \mathcal{A}} f_X(x|A)Pr\{A\}, \end{aligned} \quad (3.50)$$

dove si è inteso di sommare (o integrare) su tutto lo spazio campione \mathcal{A} condizionante. La relazione inversa tra la pdf e la cdf è ovviamente

$$F_X(x|A) = \int_{-\infty}^x f_X(\xi|A)d\xi. \quad (3.51)$$

Esempio 3.5 Un impianto manifatturiero produce resistori da 1000Ω utilizzando due macchine, $M1$ e $M2$. I valori dei resistori prodotti dalla prima macchina sono in media più precisi di quelli prodotti dalla seconda. In particolare, entrambi le macchine producono valori distribuiti secondo delle distribuzioni gaussiane a media $r_0 = 1000\Omega$, ma il primo con deviazione standard $\sigma_1 = 10\Omega$ e il secondo con deviazione standard $\sigma_2 = 20\Omega$. Inoltre la prima macchina produce solo un terzo del volume totale, mentre il rimanente è affidato alla seconda macchina. Si valuti la probabilità che un resistore scelto a caso dalla fabbrica abbia un valore compreso tra 999Ω e 1001Ω . **Soluzione:** La formalizzazione del problema può essere fatta mediante delle densità di probabilità condizionate. L'evento condizionante è l'essere stato prodotto dalla prima o dalla seconda macchina. Sia $\mathcal{A} = \{M1, M2\}$, con $Pr\{M1\} = 1/3$, $Pr\{M2\} = 2/3$. Pertanto, denotando con R la variabile aleatoria che descrive il valore di resistenza di un pezzo scelto a caso, abbiamo

$$f_R(r|M1) = \mathcal{N}(r; r_0, \sigma_1); \quad f_R(r|M2) = \mathcal{N}(r; r_0, \sigma_2). \quad (3.52)$$

Pertanto la densità non condizionata di R è

$$\begin{aligned} f_R(r) &= f_R(r|M1)Pr\{M1\} + f_R(r|M2)Pr\{M2\} \\ &= \mathcal{N}(r; r_0, \sigma_1)\frac{1}{3} + \mathcal{N}(r; r_0, \sigma_2)\frac{2}{3}. \end{aligned} \quad (3.53)$$

La probabilità cercata è

$$\begin{aligned}
 Pr\{999 < R \leq 1001\} &= \int_{999}^{1001} f_R(\xi) d\xi \\
 &= \int_{999}^{1001} f_R(\xi|M1) d\xi Pr\{M1\} + \int_{999}^{1001} f_R(\xi|M2) d\xi Pr\{M2\} \\
 &= \int_{999}^{1001} \mathcal{N}(\xi; 1000, 10) d\xi \frac{1}{3} + \int_{999}^{1001} \mathcal{N}(\xi; 1000, 20) d\xi \frac{2}{3} \\
 &= \left(2\Phi\left(\frac{1}{10}\right) - 1\right) \frac{1}{3} + \left(2\Phi\left(\frac{1}{20}\right) - 1\right) \frac{2}{3} = 0.0531.
 \end{aligned} \tag{3.54}$$

Esempio 3.6 Si consideri la variabile aleatoria che descrive la vita di una lampadina. Si supponga che tale variabile sia distribuita secondo una pdf esponenziale con vita media pari a 1000 ore. Si valuti la probabilità che la lampadina osservata in condizioni di funzionamento dopo 10 ore, si fulmini nell'ora successiva. **Soluzione:** La pdf della variabile che rappresenta la vita della lampadina espressa in ore sia

$$f_V(v) = \alpha e^{-\alpha v} u(v), \tag{3.55}$$

con $\alpha = 1/1000$. Consideriamo la densità di V condizionata all'evento $\{V > v_0\}$, dove $v_0 = 10$

$$\begin{aligned}
 f_V(v|V > v_0) &= \frac{d}{dv} F_V(v|V > v_0) = \frac{d}{dv} Pr\{V \leq v|V > v_0\} \\
 &= \frac{d}{dv} \left(\frac{Pr\{V \leq v, V > v_0\}}{Pr\{V > v_0\}} \right) \\
 &= \frac{1}{Pr\{V > v_0\}} \frac{d}{dv} Pr\{V \leq v, V > v_0\}.
 \end{aligned} \tag{3.56}$$

La probabilità $Pr\{V > v_0\}$ è calcolata dalla pdf non condizionata di V

$$Pr\{V > v_0\} = \int_{v_0}^{\infty} \alpha e^{-\alpha v} dv = e^{-\alpha v_0}. \tag{3.57}$$

La probabilità congiunta $Pr\{V \leq v, V > v_0\}$ è invece

$$\begin{aligned}
 Pr\{V \leq v, V > v_0\} &= \begin{cases} 0 & \text{se } v \leq v_0 \\ Pr\{v_0 < V \leq v\} & \text{se } v > v_0 \end{cases} \\
 &= \begin{cases} 0 & \text{se } v \leq v_0 \\ \int_{v_0}^v \alpha e^{-\alpha \xi} d\xi & \text{se } v > v_0 \end{cases} \\
 &= \begin{cases} 0 & \text{se } v \leq v_0 \\ e^{-\alpha v_0} - e^{-\alpha v} & \text{se } v > v_0. \end{cases}
 \end{aligned} \tag{3.58}$$

Eseguendo la derivata e sostituendo, abbiamo

$$f_V(v|V > v_0) = \begin{cases} 0 & \text{se } v \leq v_0 \\ \alpha e^{-\alpha(v-v_0)} & \text{se } v > v_0. \end{cases} \quad (3.59)$$

La pdf condizionata è mostrata in figura 3.11 insieme alla densità non condizionata.

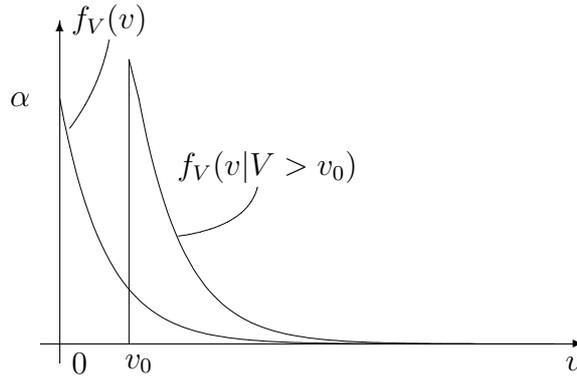


Figura 3.11: La pdf condizionata e la pdf non condizionata dell'esempio

La probabilità cercata è pertanto:

$$\begin{aligned} Pr\{V \leq v_0 + 1 | V > v_0\} &= \int_{-\infty}^{v_0+1} f_V(\xi | V > v_0) d\xi = \int_{v_0}^{v_0+1} \alpha e^{-\alpha(\xi-v_0)} d\xi \\ &= -e^{-\alpha(\xi-v_0)} \Big|_{v_0}^{v_0+1} = 1 - e^{-\alpha} = 1 - e^{-0.001} = 9.99 \cdot 10^{-4}. \end{aligned} \quad (3.60)$$

Un importante tipo di condizionamento per una variabile aleatoria X , è quello basato sull'evento

$$\{Y \leq y\}, \quad (3.61)$$

dove Y è un'altra variabile aleatoria. La cdf condizionata diventa

$$F_X(x|Y \leq y) = Pr\{X \leq x | Y \leq y\} = \frac{Pr\{X \leq x, Y \leq y\}}{Pr\{Y \leq y\}} = \frac{F_{XY}(x, y)}{F_Y(y)}, \quad (3.62)$$

dove F_{XY} è la cdf congiunta di X e Y e F_Y è la cdf marginale di Y . Vista la analogia alla definizione di probabilità condizionata, si definisce *cdf di X condizionata ad Y*

$$F_{X|Y}(x|y) \triangleq Pr\{X \leq x | Y \leq y\} = \frac{F_{XY}(x, y)}{F_Y(y)}. \quad (3.63)$$

La pdf condizionata all'evento $\{Y \leq y\}$ è definita in maniera analoga come

$$f_X(x|Y \leq y) = \left(\frac{d}{dx} F_{X|Y}(x|y) \right). \quad (3.64)$$

Un altro esempio tipico di evento condizionante è

$$\{y < Y \leq y + \Delta y\}, \quad (3.65)$$

dove Δy è un valore positivo reale. In tal caso

$$Pr\{y < Y \leq y + \Delta y\} = F_Y(y + \Delta y) - F_Y(y), \quad (3.66)$$

$$\begin{aligned} F_X(x|y < Y \leq y + \Delta y) &= \frac{Pr\{X \leq x, y < Y \leq y + \Delta y\}}{Pr\{y < Y \leq y + \Delta y\}} \\ &= \frac{F_{XY}(x, y + \Delta y) - F_{XY}(x, y)}{F_Y(y + \Delta y) - F_Y(y)}. \end{aligned} \quad (3.67)$$

La pdf corrispondente è pertanto

$$f_X(x|y < Y \leq y + \Delta y) = \frac{d}{dx} F_X(x|y < Y \leq y + \Delta y). \quad (3.68)$$

Ancora un altro evento condizionante di particolare interesse è

$$\{Y = y\}. \quad (3.69)$$

Tale evento merita attenzione particolare visto che spesso, e in particolare nel caso di una variabile aleatoria di tipo continuo, $Pr\{Y = y\} = 0$. Quindi la definizione mediante la probabilità condizionata data all'inizio del paragrafo non è applicabile. La definizione che si adotta è invece

$$F_X(x|Y = y) = \lim_{\Delta y \rightarrow 0} F_X(x|y < Y \leq y + \Delta y), \quad (3.70)$$

ammesso che tale limite esista. Tale funzione può essere riscritta come

$$\begin{aligned} F_X(x|Y=y) &= \lim_{\Delta y \rightarrow 0} \frac{F_{XY}(x, y + \Delta y) - F_{XY}(x, y)}{F_Y(y + \Delta y) - F_Y(y)} = \frac{\frac{d}{dy} F_{XY}(x, y)}{\frac{d}{dy} F_Y(y)} \\ &= \frac{\int_{-\infty}^x f_{XY}(\xi, y) d\xi}{f_Y(y)}, \end{aligned} \quad (3.71)$$

dove si è implicitamente assunto che $f_Y(y) \neq 0$. Derivando rispetto a x si ottiene la pdf corrispondente

$$f_X(x|Y=y) = \frac{f_{XY}(x, y)}{f_Y(y)}. \quad (3.72)$$

Tale espressione è spesso denominata *pdf di X condizionata da Y* e denotata come $f_{X|Y}(x|y)$. La definizione e i risultati su esposti consentono di scrivere la formula

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) dy, \quad (3.73)$$

che è spesso denominata *legge della probabilità totale per le pdf*. Anche la formula

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) f_Y(y)}{f_X(x)}. \quad (3.74)$$

è denominata *legge di Bayes per le pdf*, dove ovviamente $f_{Y|X}(y|x) \triangleq f_Y(y|X=x)$.

Nota: A causa delle possibili ambiguità che potrebbero emergere dalla definizione di cdf e pdf condizionata a riguardo dell'evento condizionante (ciò si verifica spesso nella letteratura sull'argomento), si consiglia al lettore di utilizzare comunque la notazione esplicita sul condizionamento.

Se un evento A è indipendente dalla variabile X , abbiamo ovviamente che tutte le cdf e le pdf condizionate coincidono con quelle non condizionate

$$F_X(x|A) = F_X(x); \quad f_X(x|A) = f_X(x), \quad (3.75)$$

e pertanto

$$F_X(x, A) = F_X(x)Pr\{A\}; \quad f_X(x, A) = f_X(x)Pr\{A\}. \quad (3.76)$$

Più in particolare, se la variabile aleatoria Y è indipendente da X , abbiamo che

$$F_X(x|y) = F_X(x); \quad f_X(x|y) = f_X(x), \quad (3.77)$$

e pertanto

$$\begin{aligned} F_{XY}(x, y) &= F_{X|Y}(x|y)F_Y(y) = F_X(x)F_Y(y); \\ f_{XY}(x, y) &= f_{X|Y}(x|y)f_Y(y) = f_X(x)f_Y(y). \end{aligned} \quad (3.78)$$

3.5.1 Momenti congiunti

Date due variabili aleatorie X e Y , si definiscono *momenti congiunti di ordine* $l = n + k$ le medie

$$m^{nk} \triangleq E[X^n Y^k] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \xi^n \eta^k f_{XY}(\xi, \eta) d\xi d\eta. \quad (3.79)$$

Per $l = 1$, i momenti sono le medie di X e Y

$$m^{10} = E[X] = \mu_x; \quad m^{01} = E[Y] = \mu_y. \quad (3.80)$$

Il momento del secondo ordine, con $n = k = 1$, è detto *correlazione tra X e Y*

$$R_{xy} = E[XY]. \quad (3.81)$$

Dimostreremo in seguito la seguente disuguaglianza notevole

$$|R_{xy}| \leq \sqrt{E[X^2]} \sqrt{E[Y^2]}. \quad (3.82)$$

Si definiscono *momenti centrali congiunti di ordine* $l = n + k$ le medie

$$\begin{aligned} \eta^{nk} &\triangleq E[(X - \mu_x)^n (Y - \mu_y)^k] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\xi - \mu_x)^n (\beta - \mu_y)^k f_{XY}(\xi, \beta) d\xi d\beta. \end{aligned} \quad (3.83)$$

I momenti centrali congiunti di ordine $l = 2$, con $(n, k) = (2, 0)$ e $(n, k) = (0, 2)$ sono proprio le varianze di X e di Y

$$\eta^{20} = E[(X - \mu_x)^2] = \sigma_x^2; \quad \eta^{02} = E[(Y - \mu_y)^2] = \sigma_y^2. \quad (3.84)$$

Il momento centrale congiunto di ordine due, con $n = k = 1$ è detto *covarianza di X e Y*

$$C_{xy} = E[(X - \mu_x)(Y - \mu_y)]. \quad (3.85)$$

Applicando la definizione di media, e separando gli integrali abbiamo che

$$E[(X - \mu_x)(Y - \mu_y)] = E[XY] - \mu_x E[Y] - E[X]\mu_y + \mu_x \mu_y. \quad (3.86)$$

Si ottiene pertanto l'uguaglianza notevole

$$C_{xy} = E[XY] - E[X]E[Y] = m^{11} - \mu_x \mu_y. \quad (3.87)$$

Dimostreremo in seguito anche la disuguaglianza notevole

$$|E[(X - \mu_x)(Y - \mu_y)]| \leq \sqrt{E[(X - \mu_x)^2]} \sqrt{E[(Y - \mu_y)^2]}. \quad (3.88)$$

Si definisce *coefficiente di correlazione tra X e Y* il rapporto

$$\rho_{xy} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sqrt{\sigma_x^2} \sqrt{\sigma_y^2}}. \quad (3.89)$$

Dalla disuguaglianza (3.88), abbiamo che il coefficiente di correlazione soddisfa la relazione

$$-1 \leq \rho_{xy} \leq 1. \quad (3.90)$$

In particolare, è possibile dimostrare che la condizione $\rho = 1$ è equivalente alla uguaglianza

$$\frac{X - E[X]}{\sqrt{\sigma_x^2}} = \frac{Y - E[Y]}{\sqrt{\sigma_Y^2}}. \quad (3.91)$$

Analogamente, la condizione $\rho = -1$ è equivalente alla uguaglianza

$$\frac{X - E[X]}{\sqrt{\sigma_x^2}} = -\frac{Y - E[Y]}{\sqrt{\sigma_Y^2}}. \quad (3.92)$$

Le relazioni precedenti, mostrano come il coefficiente di correlazione misuri l'*entità della dipendenza lineare tra X e Y*. Infatti nel caso limite per $\rho = \pm 1$, le due variabili sono linearmente dipendenti (predicibili) l'una dall'altra.

Prova di (3.82), (3.88), (3.91), (3.92): Si consideri la quantità $h(t) = E[(tX + Y)^2] = t^2 E[X^2] - 2tE[XY] + E[Y^2]$. Ovviamente

$h(t) \geq 0 \quad \forall t$. Ciò impone che la equazione $h(t) = 0$ abbia una sola o nessuna soluzione, ovvero che il suo discriminante sia minore o uguale a zero

$$E[XY]^2 - E[X^2]E[Y^2] \leq 0, \quad (3.93)$$

da cui disuguaglianza (3.82). Applicando lo stesso ragionamento alle variabili $X' = X - \mu_x$ e $Y' = Y - \mu_y$, abbiamo che

$$E[(X - \mu_x)(Y - \mu_y)]^2 - E[(X - \mu_x)^2]E[(Y - \mu_y)^2] \leq 0, \quad (3.94)$$

da cui disuguaglianza (3.88). Il caso relativo all'uguaglianza corrisponde alla equazione $E[(tX + Y)^2] = 0$, che implica che $X = -\frac{1}{t}Y$ ¹, ovvero che esiste una relazione di proporzionalità tra X e Y . La equazione è soddisfatta quando $E[XY]^2 - E[X^2]E[Y^2] = 0$ e per $t = -E[XY]/E[X^2]$. Pertanto

$$X = \frac{E[XY]}{E[X^2]}Y. \quad (3.95)$$

Poichè $E[XY] = \pm\sqrt{E[X^2]E[Y^2]}$, abbiamo che

$$X = \pm \frac{E[X^2]}{\sqrt{E[X^2]} + \sqrt{E[Y^2}}}Y. \quad (3.96)$$

Analogamente, se usiamo la stessa procedura sulle variabili $X - \mu_x$ e $Y - \mu_y$, abbiamo che

$$X - \mu_x = \pm \frac{E[(X - \mu_x)^2]}{\sqrt{E[(X - \mu_x)^2]} + \sqrt{E[(Y - \mu_y)^2}}}(Y - \mu_y), \quad (3.97)$$

che corrisponde a equazioni (3.91) e (3.92), dove il segno positivo e il segno negativo valgono se $\rho_{xy} = 1$ o $\rho_{xy} = -1$ rispettivamente.

3.5.2 Variabili aleatorie indipendenti, scorrelate, ortogonali

Abbiamo già visto che due variabili aleatorie X e Y sono indipendenti se e solo se

$$f_{XY}(x, y) = f_X(x)f_Y(y), \quad (3.98)$$

¹Ciò in effetti vale *quasi ovunque*. Il lettore tralasci questo aspetto per ora.

ovvero se la loro pdf congiunta è fattorizzabile nel prodotto delle due pdf marginali.

Due variabili aleatorie si dicono *scorrelate* se

$$E[XY] = E[X]E[Y], \quad (3.99)$$

ovvero se la correlazione è fattorizzabile nel prodotto delle due medie marginali. Per due variabili aleatorie scorrelate abbiamo che la covarianza è nulla. Infatti

$$C_{xy} = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0. \quad (3.100)$$

Ovviamente due variabili aleatorie scorrelate hanno coefficiente di correlazione nullo.

Altri tipi di decorrelazione di ordine $l = n + k$ sono definibili imponendo la fattorizzabilità dei momenti congiunti:

$$E[X^n Y^k] = E[X^n]E[Y^k]. \quad (3.101)$$

per qualche n e k specificato. Si noti come l'indipendenza implichi la fattorizzabilità di tutti i momenti congiunti. Infatti

$$\begin{aligned} m_{nk} = E[X^n Y^k] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \xi^n \eta^k f_{XY}(\xi, \eta) d\xi d\eta \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \xi^n \eta^k f_X(\xi) f_Y(\eta) d\xi d\eta \\ &= \int_{-\infty}^{\infty} \xi^n f_X(\xi) d\xi \int_{-\infty}^{\infty} \eta^k f_Y(\eta) d\eta \\ &= E[X^n]E[Y^k]. \end{aligned} \quad (3.102)$$

Si faccia attenzione perché in generale non vale il viceversa: due variabili aleatorie per le quali alcuni momenti congiunti sono fattorizzabili, non sono necessariamente indipendenti. L'indipendenza implica ovviamente che anche i momenti centrali corrispondenti sono fattorizzabili

$$\eta_{nk} = E[(X - \mu_x)^n (Y - \mu_y)^k] = E[(X - \mu_x)^n] E[(Y - \mu_y)^k] = \sigma_x^n \sigma_y^k. \quad (3.103)$$

Quindi, se due variabili aleatorie sono indipendenti, sono anche scorrelate, ma in generale non è vero il contrario. E' opportuno enfatizzare che l'indipendenza è una condizione più forte della decorrelazione.

Due variabili aleatorie X e Y si dicono *ortogonali* se

$$E[XY] = 0. \quad (3.104)$$

Ciò implica che due variabili aleatorie scorrelate sono anche ortogonali se almeno una di esse è a media nulla.

3.5.3 Due variabili aleatorie gaussiane

Nella densità di probabilità congiunta della variabile aleatoria gaussiana descritta in equazione (3.34), il parametro r , rappresenta proprio il coefficiente di correlazione tra X e Y (il lettore lo dimostri per esercizio). Pertanto, per $r = 0$, abbiamo che:

$$f_{XY}(x, y) = \mathcal{N}(x; \mu_x; \sigma_x^2) \mathcal{N}(y; \mu_y; \sigma_y^2), \quad (3.105)$$

ovvero la pdf congiunta si fattorizza nel prodotto delle due pdf marginali. Ciò immediatamente implica che per due variabili aleatorie congiuntamente gaussiane, la decorrelazione implica l'indipendenza. Quindi, *per due variabili aleatorie congiuntamente gaussiane indipendenza e decorrelazione sono equivalenti*.

3.6 Momenti condizionati

E' possibile definire momenti di una variabile aleatoria quando essa è condizionata da altri eventi. Abbiamo già visto come sia possibile definire delle cdf e delle pdf condizionate. Abbiamo anche sottolineato che spesso conviene fare esplicita menzione all'evento condizionante. Quindi data una variabile aleatoria x e un evento A , si definisce momento di X di ordine n

$$E[X^n|A] \triangleq \int_{-\infty}^{+\infty} \xi^n f_X(\xi|A) d\xi. \quad (3.106)$$

Ovviamente, se l'evento A e la variabile X sono indipendenti, ne consegue che $f_X(x|A) = f_X(x)$ e quindi $E[X^n|A] = E[X^n]$. Analogamente al caso non condizionato, possono essere definiti i momenti centrali e i momenti congiunti condizionati.

Capitolo 4

Più variabili aleatorie

In questo capitolo viene introdotta la formalizzazione relativa a più di due variabili aleatorie. I concetti introdotti nel capitolo precedente vengono generalizzati al caso di n variabili aleatorie. Vengono introdotte la cdf e la pdf multidimensionale e i momenti relativi. E' inoltre introdotta la gaussiana multidimensionale, la matrice di correlazione e la matrice di covarianza. Particolare enfasi è attribuita alla formalizzazione matriciale. In coda al capitolo viene introdotta brevemente la stima a minimo errore quadratico medio.

4.1 Introduzione

I concetti introdotti nel capitolo precedente, a proposito della coppia di variabili aleatorie, possono essere generalizzati ad un insieme di n variabili aleatorie

$$X_1, X_2, \dots, X_n. \quad (4.1)$$

La n -pla è utilmente associata ad un vettore¹

$$\underline{X} = (X_1, X_2, \dots, X_n)^T. \quad (4.2)$$

Pertanto la *variabile aleatoria n -dimensionale* \underline{X} è

$$\underline{X} : \xi \in \mathcal{S} \rightarrow \underline{x} = (x_1, x_2, \dots, x_n)^T \in \mathcal{I} \subseteq \mathcal{R}^n. \quad (4.3)$$

¹Quando non specificato, per convenzione, un vettore è un vettore colonna.

Quindi una variabile aleatoria multi-dimensionale è una associazione tra lo spazio campione e un sottoinsieme di \mathcal{I} di \mathcal{R}^n su cui è definita una misura di probabilità. Si definisce *funzione cumulativa*, o *cdf congiunta* delle n variabili aleatorie

$$F_{\underline{X}}(\underline{x}) = F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = Pr\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\}. \quad (4.4)$$

Le proprietà della cdf n -dimensionale sono del tutto analoghe a quelle già discusse per due variabili aleatorie. Per esempio, le seguenti identità sono facilmente dimostrabili

$$F_{\underline{X}}(+\infty, x_2, \dots, x_n) = F_{X_2 \dots X_n}(x_2, \dots, x_n), \quad (4.5)$$

$$F_{\underline{X}}(x_1, +\infty, \dots, +\infty) = F_{X_1}(x_1), \quad (4.6)$$

$$F_{\underline{X}}(-\infty, x_2, \dots, x_n) = 0, \quad (4.7)$$

$$F_{\underline{X}}(+\infty, +\infty, \dots, +\infty) = 1, \quad (4.8)$$

$$F_{\underline{X}}(+\infty, -\infty, \dots, +\infty) = 0. \quad (4.9)$$

La *pdf congiunta* è la derivata n -esima della cdf congiunta

$$f_{\underline{X}}(\underline{x}) = f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = \frac{\partial^n F_{\underline{X}}(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}. \quad (4.10)$$

La pdf soddisfa la proprietà fondamentale

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{\underline{X}}(\xi_1, \xi_2, \dots, \xi_n) d\xi_1 d\xi_2 \dots d\xi_n = 1, \quad (4.11)$$

ed è tale che la probabilità che la \underline{X} cada in un dominio $\mathcal{D} \subset \mathcal{R}^n$ è

$$Pr\{\underline{X} \in \mathcal{D}\} = \int_{\mathcal{D}} f_{\underline{X}}(\underline{x}) d\underline{x}. \quad (4.12)$$

Per ottenere la cdf dalla pdf basta integrare

$$F_{\underline{X}}(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_{\underline{X}}(\xi_1, \xi_2, \dots, \xi_n) d\xi_1 d\xi_2 \dots d\xi_n. \quad (4.13)$$

Nel descrivere un insieme di n variabili aleatorie congiuntamente, è possibile definire cdf e pdf congiunte relative a sottoinsiemi di esse. Per esempio se $n = 3$, abbiamo le seguenti cdf e pdf

$$F_{X_1}(x_1), F_{X_2}(x_2), F_{X_3}(x_3); f_{X_1}(x_1), f_{X_2}(x_2), f_{X_3}(x_3); \quad (4.14)$$

$$\begin{aligned} &F_{X_1X_2}(x_1, x_2), F_{X_1X_3}(x_1, x_3), F_{X_2X_3}(x_2, x_3); \\ &f_{X_1X_2}(x_1, x_2), f_{X_1X_3}(x_1, x_3), f_{X_2X_3}(x_2, x_3); \end{aligned} \quad (4.15)$$

$$F_{X_1X_2X_3}(x_1, x_2, x_3); f_{X_1X_2X_3}(x_1, x_2, x_3). \quad (4.16)$$

Le (4.14) sono dette cdf e pdf *marginali*, o del *I ordine*. Le (4.15) e (4.16) sono dette cdf e pdf rispettivamente del *II* e *III* ordine.

Da una cdf o pdf congiunta n -dimensionale è possibile ottenere tutte le cdf e le pdf di ordine inferiore. Ecco alcuni esempi per $n = 3$

$$F_{X_1X_2}(x_1, x_3) = F_{\underline{X}}(x_1, +\infty, x_3); \quad (4.17)$$

$$F_{X_2}(x_2) = F_{\underline{X}}(+\infty, x_2, +\infty); \quad (4.18)$$

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{\underline{X}}(x_1, x_2, x_3) dx_2 dx_3; \quad (4.19)$$

$$f_{X_1X_2}(x_1, x_2) = \int_{-\infty}^{+\infty} f_{\underline{X}}(x_1, x_2, x_3) dx_3. \quad (4.20)$$

4.1.1 Variabili condizionate

Anche per la variabile aleatoria multidimensionale è possibile definire funzioni cumulative e densità condizionate. Dalla definizione di un evento condizionante A , le definizioni di cdf e pdf di \underline{X} condizionate ad A sono

$$F_{\underline{X}}(\underline{x}|A) = Pr\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n|A\}, \quad (4.21)$$

$$f_{\underline{X}}(\underline{x}|A) = \frac{\partial^n F_{\underline{X}}(x_1, x_2, \dots, x_n|A)}{\partial x_1 \partial x_2 \cdots \partial x_n}. \quad (4.22)$$

Ovviamente l'evento condizionante può essere definito in termini di altre variabili aleatorie, o anche mediante sottoinsiemi delle stesse n variabili aleatorie descritte. Ad esempio la densità congiunta di X_1 e X_2 condizionata a X_3 e X_4 è

$$f_{X_1X_2}(x_1, x_2|x_3, x_4). \quad (4.23)$$

Si ricordi che nella definizione di densità condizionata si è implicitamente fatto riferimento all'evento condizionante $\{X_3 = x_3, X_4 = x_4\}$, con tutte le attenzioni del caso dovute al fatto che tale evento ha probabilità nulla

per variabili aleatorie di tipo continuo. Generalizzando le considerazioni fatte a proposito della coppia di variabili aleatorie, valgono le relazioni

$$f_{X_1X_2}(x_1, x_2|x_3, x_4) = \frac{f_{X_1X_2X_3X_4}(x_1, x_2, x_3, x_4)}{f_{X_3X_4}(x_3, x_4)}; \quad (4.24)$$

$$f_{X_1}(x_1|x_2, x_3) = \frac{f_{X_1X_2X_3}(x_1, x_2, x_3)}{f_{X_2X_3}(x_2, x_3)}; \quad (4.25)$$

$$f_{X_1\dots X_n}(x_1, \dots, x_n) = f_{X_n}(x_n|x_1, \dots, x_{n-1})f_{X_1\dots X_{n-1}}(x_1, \dots, x_{n-1}). \quad (4.26)$$

Iterando l'ultima relazione otteniamo la *regola a catena*

$$f_{X_1\dots X_n}(x_1, \dots, x_n) = f_{X_n}(x_n|x_1, \dots, x_{n-1})f_{X_{n-1}}(x_{n-1}|x_1, \dots, x_{n-2}) \cdots f_{X_2}(x_2|x_1)f_{X_1}(x_1). \quad (4.27)$$

Inoltre, per ottenere densità marginali, ovvero *eliminare* delle variabili condizionate, basta integrare, come nei seguenti esempi

$$f_{X_1X_3}(x_1, x_3|x_5) = \int_{-\infty}^{+\infty} f_{X_1X_2X_3}(x_1, x_2, x_3|x_5)dx_2. \quad (4.28)$$

$$f_{X_3}(x_3|x_1) = \int_{-\infty}^{+\infty} f_{X_3X_4}(x_3, x_4|x_1)dx_4. \quad (4.29)$$

Per *eliminare* variabili condizionanti, bisogna utilizzare anche la relazione condizionante. Ad esempio, per ottenere

$$f_{X_1X_2}(x_1, x_2|x_3) \quad \text{da} \quad f_{X_1X_2}(x_1, x_2|x_3, x_4), \quad (4.30)$$

riscriviamo

$$\begin{aligned} f_{X_1X_2}(x_1, x_2|x_3) &= \frac{f_{X_1X_2X_3}(x_1, x_2, x_3)}{f_{X_3}(x_3)} \\ &= \frac{\int_{-\infty}^{+\infty} f_{X_1X_2X_3X_4}(x_1, x_2, x_3, x_4)dx_4}{f_{X_3}(x_3)} \\ &= \frac{\int_{-\infty}^{+\infty} f_{X_1X_2}(x_1, x_2|x_3, x_4)f_{X_3X_4}(x_3, x_4)dx_4}{f_{X_3}(x_3)} \\ &= \int_{-\infty}^{+\infty} f_{X_1X_2}(x_1, x_2|x_3, x_4)f_{X_4}(x_4|x_3)dx_4, \end{aligned} \quad (4.31)$$

che è la relazione cercata. Il lettore verifichi la seguente relazione per esercizio

$$\begin{aligned} f_{X_1X_2}(x_1, x_2|x_4) \\ = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X_1X_2}(x_1, x_2|x_3, x_4, x_5)f_{X_3X_5}(x_3, x_5|x_4)dx_3dx_5. \end{aligned} \quad (4.32)$$

4.1.2 Variabili indipendenti

Ricordiamo come n variabili aleatorie vengano caratterizzate mediante la probabilità congiunta degli eventi

$$\{X_1 \leq x_1\}, \{X_2 \leq x_2\}, \dots, \{X_n \leq x_n\}. \quad (4.33)$$

Quindi le variabili sono indipendenti se tali eventi sono indipendenti. Ricordiamo che in generale n eventi sono indipendenti se e solo se le probabilità di tutti gli eventi congiunti fino all'ordine n sono fattorizzabili. Rispetto alle cdf e alle pdf, diremo che n variabili aleatorie sono indipendenti, se tutte le cdf, o le pdf, fino all'ordine n sono fattorizzabili. Ad esempio, affinché X_1, X_2, X_3 siano indipendenti è necessario che

$$\begin{cases} f_{X_1 X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2); \\ f_{X_1 X_3}(x_1, x_3) = f_{X_1}(x_1)f_{X_3}(x_3); \\ f_{X_2 X_3}(x_2, x_3) = f_{X_2}(x_2)f_{X_3}(x_3); \end{cases} \quad (4.34)$$

$$f_{X_1 X_2 X_3}(x_1, x_2, x_3) = f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3). \quad (4.35)$$

E' opportuno enfatizzare che nel caso delle variabili aleatorie, la sola relazione (4.35) è sufficiente a garantire l'indipendenza. Viceversa, la fattorizzazione delle sole pdf di ordine inferiore non implica necessariamente l'indipendenza. Ciò si vede facilmente notando che le pdf di ordine inferiore possono essere ottenute per integrazione della pdf congiunta di ordine massimo. Pertanto

Date n variabili aleatorie X_1, X_2, \dots, X_n , condizione necessaria e sufficiente per l'indipendenza è

$$f_{\underline{X}}(\underline{x}) = f_{X_1}(x_1)f_{X_2}(x_2) \cdots f_{X_n}(x_n). \quad (4.36)$$

L'indipendenza può anche essere riferita a gruppi di variabili aleatorie. Ad esempio, le variabili X_1, X_2, X_3 sono indipendenti dalle variabili X_4, X_5 se

$$f_{X_1 X_2 X_3 X_4 X_5}(x_1, x_2, x_3, x_4, x_5) = f_{X_1 X_2 X_3}(x_1, x_2, x_3)f_{X_4 X_5}(x_4, x_5). \quad (4.37)$$

Ciò ovviamente non esclude che le variabili all'interno di ogni gruppo possano essere mutuamente dipendenti. Come ulteriore esempio, i gruppi

di variabili $\{X_1\}$, $\{X_2, X_3\}$, $\{X_4, X_5, X_6\}$ sono mutuamente indipendenti se

$$\begin{aligned} f(x_1, x_2, x_3, x_4, x_5) &= f(x_1)f(x_2, x_3)f(x_4, x_5, x_6), \\ f(x_1, x_2, x_3) &= f(x_1)f(x_2, x_3), \\ f(x_1, x_4, x_5, x_6) &= f(x_1)f(x_4, x_5, x_6), \\ f(x_2, x_3, x_4, x_5, x_6) &= f(x_2, x_3)f(x_4, x_5, x_6), \end{aligned} \quad (4.38)$$

dove si è utilizzata una notazione abbreviata per le pdf. Comunque la relazione

$$f(x_1, x_2, x_3, x_4, x_5) = f(x_1)f(x_2, x_3)f(x_4, x_5, x_6), \quad (4.39)$$

è sufficiente poiché implica le altre.

L'indipendenza può anche essere definita con riferimento alle pdf condizionate. Ad esempio, le variabili $\{X_1, X_2\}$ sono indipendenti da $\{X_3, X_4, X_5\}$ se

$$f(x_1, x_2 | x_3, x_4, x_5) = f(x_1, x_2). \quad (4.40)$$

Si noti che questa relazione implica anche che

$$f(x_3, x_4, x_5 | x_1, x_2) = f(x_3, x_4, x_5). \quad (4.41)$$

4.2 Momenti

Dato un insieme di k variabili aleatorie

$$X_{i_1}, X_{i_2}, \dots, X_{i_k} \quad i_l \in \{1, 2, \dots, n\}, \quad l = 1, \dots, k, \quad (4.42)$$

estratte dalle n variabili $\{X_1, X_2, \dots, X_n\}$, si definisce *momento congiunto di ordine*

$$m = e(i_1) + e(i_2) + \dots + e(i_k), \quad (4.43)$$

delle k variabili aleatorie

$$\begin{aligned} E[X_{i_1}^{e(i_1)} X_{i_2}^{e(i_2)} \dots X_{i_k}^{e(i_k)}] \\ = \int x_{i_1}^{e(i_1)} x_{i_2}^{e(i_2)} \dots x_{i_k}^{e(i_k)} f_{X_{i_1} X_{i_2} \dots X_{i_k}}(x_{i_1} x_{i_2} \dots x_{i_k}) dx_{i_1} dx_{i_2} \dots dx_{i_k}, \end{aligned} \quad (4.44)$$

dove $e(i_1), e(i_2), \dots, e(i_k)$ sono degli esponenti interi.

Ad esempio

$$E[X_3], E[X_1^2 X_2^6], E[X_1 X_2 X_3], E[X_1^8 X_3 X_5 X_9^2], E[X_1 X_2], \quad (4.45)$$

sono momenti di ordine 1, 8, 3, 12, 2.

4.2.1 La media

Per ogni variabile X_i

$$\mu_i = E[X_i], \quad (4.46)$$

è la *media marginale* della i -esima variabile. La media della variabile aleatoria n -dimensionale \underline{X} è l'insieme delle medie delle singole variabili, che organizzate in un vettore definiscono il *vettore medio* o la *media n -dimensionale*

$$\underline{\mu} = E[\underline{X}] = (E[X_1], E[X_2], \dots, E[X_n])^T = (\mu_1, \mu_2, \dots, \mu_n)^T. \quad (4.47)$$

4.2.2 La matrice di correlazione

Gli n^2 momenti del secondo ordine ottenibili dalle n variabili aleatorie sono

$$r_{ij} = E[X_i X_j], \quad i, j = 1, \dots, n. \quad (4.48)$$

Il valore $r_{ij} = E[X_i X_j]$, per $i \neq j$, è la *correlazione* tra X_i e X_j ed è chiaramente tale che

$$r_{ij} = r_{ji}. \quad (4.49)$$

Per $i = j$,

$$r_{ii} = E[X_i^2], \quad (4.50)$$

è il *valore quadratico medio* di X_i . Tutte le r_{ij} possono essere organizzate nella matrice

$$\underline{\underline{R}}_X = E[\underline{X} \underline{X}^T] = \begin{bmatrix} E[X_1^2] & E[X_1 X_2] & \dots & E[X_1 X_n] \\ E[X_2 X_1] & E[X_2^2] & \dots & E[X_2 X_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_n X_1] & E[X_n X_2] & \dots & E[X_n^2] \end{bmatrix}, \quad (4.51)$$

detta *matrice di correlazione di \underline{X}* . La matrice di correlazione soddisfa le seguenti proprietà.

Proprietà 4.1 La matrice di correlazione è simmetrica, ovvero

$$\underline{R}_X = \underline{R}_X^T. \quad (4.52)$$

Prova: La simmetria delle correlazioni implica immediatamente la simmetria della matrice.

Proprietà 4.2 La matrice di correlazione è definita positiva, ovvero

$$\underline{z}^T \underline{R}_X \underline{z} \geq 0 \quad \forall \underline{z} \in \mathcal{R}^n. \quad (4.53)$$

Prova: La matrice di correlazione è definita positiva poiché

$$\underline{z}^T \underline{R}_X \underline{z} = \underline{z}^T E[\underline{X} \underline{X}^T] \underline{z} = E[\underline{z}^T \underline{X} \underline{X}^T \underline{z}] = E[(\underline{z}^T \underline{X})^2] \geq 0 \quad \forall \underline{z} \in \mathcal{R}^n. \quad (4.54)$$

Le variabili aleatorie X_i e X_j , $i \neq j$, sono *scorrelate* se

$$E[X_i X_j] = E[X_i] E[X_j]. \quad (4.55)$$

Se tutte le n variabili aleatorie sono scorrelate, la matrice di correlazione diventa

$$\underline{R}_X = E[\underline{X} \underline{X}^T] = \begin{bmatrix} E[X_1^2] & \mu_1 \mu_2 & \cdot & \cdot & \cdot & \mu_1 \mu_n \\ \mu_2 \mu_1 & E[X_2^2] & \cdot & \cdot & \cdot & \mu_1 \mu_n \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mu_n \mu_1 & \mu_n \mu_2 & \cdot & \cdot & \cdot & E[X_n^2] \end{bmatrix}, \quad (4.56)$$

che può essere espressa in forma matriciale compatta come

$$\underline{R}_X = \text{diag}(E[X_1^2] - \mu_1^2, \dots, E[X_n^2] - \mu_n^2) + \underline{\mu} \underline{\mu}^T. \quad (4.57)$$

Le variabili aleatorie X_i e X_j , $i \neq j$, sono *ortogonali* se

$$E[X_i X_j] = 0. \quad (4.58)$$

Se tutte le n variabili aleatorie sono ortogonali, la matrice di correlazione diventa

$$\begin{aligned} \underline{R}_X = E[\underline{X} \underline{X}^T] &= \begin{bmatrix} E[X_1^2] & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & E[X_2^2] & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & E[X_n^2] \end{bmatrix} \quad (4.59) \\ &= \text{diag}(E[X_1^2], \dots, E[X_n^2]). \quad (4.60) \end{aligned}$$

Si ricordi che due variabili aleatorie scorrelate sono anche ortogonali se almeno una di esse è a media nulla. Quindi per n variabili aleatorie scorrelate a media nulla la matrice di correlazione è diagonale.

4.3 Momenti centrali

I momenti centrali congiunti sono i momenti delle variabili aleatorie dopo che ad esse è stata rimossa la media. Ovvero: dato un insieme di k variabili aleatorie

$$X_{i_1}, \dots, X_{i_k} \quad i_l \in \{1, \dots, n\}, \quad l = 1, \dots, k, \quad (4.61)$$

estratte dalle n variabili $\{X_1, \dots, X_n\}$, e aventi medie

$$\mu_{i_1}, \dots, \mu_{i_k}, \quad (4.62)$$

si definisce *momento centrale congiunto di ordine*

$$m = e(i_1) + \dots + e(i_k), \quad (4.63)$$

delle k variabili aleatorie

$$\begin{aligned} & E[(X_{i_1} - \mu_{i_1})^{e(i_1)} \dots (X_{i_k} - \mu_{i_k})^{e(i_k)}] \\ &= \int (X_{i_1} - \mu_{i_1})^{e(i_1)} \dots (X_{i_k} - \mu_{i_k})^{e(i_k)} f_{X_{i_1} \dots X_{i_k}}(x_{i_1} \dots x_{i_k}) dx_{i_1} \dots dx_{i_k}, \end{aligned} \quad (4.64)$$

dove $e(i_1), \dots, e(i_k)$ sono degli esponenti interi.

Ad esempio

$$\begin{aligned} & E[(X_1 - \mu_1)^2 (X_2 - \mu_2)^6], \\ & E[(X_1 - \mu_1)(X_2 - \mu_2)(X_3 - \mu_3)], E[(X_1 - \mu_1)(X_2 - \mu_2)], \end{aligned} \quad (4.65)$$

sono momenti centrali di ordine 8, 3 e 2.

Si noti che i momenti centrali di ordine uno sono tutti nulli

$$E[(X_i - \mu_i)] = E[X_i] - \mu_i = 0. \quad (4.66)$$

4.3.1 La matrice di covarianza

Il valore

$$c_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)], \quad i \neq j, \quad (4.67)$$

è la *covarianza tra X_i e X_j* . Il valore

$$\sigma_i^2 = c_{ii} = E[(X_i - \mu_i)^2], \quad (4.68)$$

è la *varianza di* X_i . Il legame tra la covarianza e la correlazione è stato già ottenuto nel capitolo sullo studio di due variabili aleatorie. Esso è comunque immediatamente ottenuto dalla definizione e dalla linearità della media ed è

$$c_{ij} = r_{ij} - \mu_i \mu_j, \quad \forall i, j. \quad (4.69)$$

Tutte le covarianze e le varianze possono essere organizzate nella *matrice di covarianza*

$$\begin{aligned} \underline{\underline{C}}_X &= E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})^T] & (4.70) \\ &= \begin{bmatrix} E[(X_1 - \mu_1)^2] & \cdot & \cdot & \cdot & E[(X_1 - \mu_1)(X_n - \mu_n)] \\ E[(X_2 - \mu_2)(X_1 - \mu_1)] & \cdot & \cdot & \cdot & E[(X_2 - \mu_2)(X_n - \mu_n)] \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ E[(X_n - \mu_n)(X_1 - \mu_1)] & \cdot & \cdot & \cdot & E[(X_n - \mu_n)^2] \end{bmatrix}. \end{aligned}$$

La matrice di covarianza è anch'essa simmetrica e definita positiva. Il legame con la matrice di correlazione è immediata conseguenza di (4.69)

$$\underline{\underline{C}}_X = E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})^T] = E[\underline{X} \underline{X}^T] - \underline{\mu} \underline{\mu}^T = \underline{\underline{R}}_X - \underline{\mu} \underline{\mu}^T. \quad (4.71)$$

Se le n variabili aleatorie sono scorrelate, la matrice di covarianza diventa ovviamente diagonale

$$\begin{aligned} \underline{\underline{C}}_X &= \begin{bmatrix} E[(X_1 - \mu_1)^2] & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & E[(X_2 - \mu_2)^2] & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & E[(X_n - \mu_n)^2] \end{bmatrix} \\ &= \text{diag}(\sigma_1^2, \dots, \sigma_n^2). & (4.72) \end{aligned}$$

4.4 La variabile gaussiana multidimensionale

La variabile aleatoria n -dimensionale ha una pdf descritta dalla seguente formula

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2} |\underline{\underline{C}}_X|^{1/2}} \exp\left(-\frac{1}{2}(\underline{x} - \underline{\mu})^T \underline{\underline{C}}_X^{-1}(\underline{x} - \underline{\mu})\right) = \mathcal{N}(\underline{x}; \underline{\mu}, \underline{\underline{C}}_X). \quad (4.73)$$

4.4. LA VARIABILE GAUSSIANA MULTIDIMENSIONALE 153

Si noti che la formula generalizza la gaussiana monodimensionale e che la varianza monodimensionale è stata rimpiazzata dalla matrice di covarianza.

Le curve di livello della pdf $f_{\underline{X}}(\underline{x})$ nello spazio n -dimensionale sono degli iper-ellissoidi descritti dall'equazione quadratica

$$\left((\underline{x} - \underline{\mu})^T \underline{C}_X^{-1} (\underline{x} - \underline{\mu}) \right) = K, \quad (4.74)$$

per vari valori di K . L'orientamento degli iper-ellissoidi dipende dalla autostruttura della matrice di covarianza

$$\underline{C}_X = \underline{Q} \underline{\Lambda}_X \underline{Q}^T, \quad (4.75)$$

dove \underline{Q} contiene gli autovettori di \underline{C}_X e $\underline{\Lambda}_X = \text{diag}(\lambda_1, \dots, \lambda_n)$ gli autovettori. Si noti che, poichè la matrice è definita positiva, $\lambda_i \geq 0$, $i = 1, \dots, n$.

Il caso monodimensionale si ottiene per $n = 1$ e $\underline{C}_X = \sigma^2$. Il caso bidimensionale, discusso nel capitolo scorso, per

$$\underline{C}_X = \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}. \quad (4.76)$$

Se le variabili sono scorrelate

$$\underline{C}_X = \text{diag}(\sigma_1^2, \dots, \sigma_n^2). \quad (4.77)$$

Pertanto

$$(\underline{x} - \underline{\mu})^T \underline{C}_X^{-1} (\underline{x} - \underline{\mu}) = (\underline{x} - \underline{\mu})^T \text{diag}\left(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_n^2}\right) (\underline{x} - \underline{\mu}) \quad (4.78)$$

$$= \sum_{i=1}^n \frac{1}{\sigma_i^2} (x_i - \mu_i)^2; \quad (4.79)$$

$$|\underline{C}_X| = \prod_{i=1}^n \sigma_i^2. \quad (4.80)$$

La pdf diventa

$$f_{\underline{X}}(\underline{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right) = \prod_{i=1}^n \mathcal{N}(x_i; \mu_i, \sigma_i). \quad (4.81)$$

Quest'ultima equazione mostra la fattorizzazione della pdf n -dimensionale nel prodotto di n gaussiane monodimensionali quando le variabili sono scorrelate. Pertanto vale l'importante risultato:

n variabili aleatorie gaussiane scorrelate sono anche indipendenti.

Quando si ha a che fare con variabili aleatorie congiuntamente gaussiane, indipendenza e decorrelazione sono proprietà equivalenti.

4.4.1 Due insiemi di variabili gaussiane

Nel risolvere problemi di stima, è utile ricavare delle relazioni notevoli quando si è in presenza di due insiemi di variabili aleatorie gaussiane tra di loro dipendenti. Consideriamo un insieme di $n + m$ variabili aleatorie congiuntamente gaussiane

$$X_1, X_2, \dots, X_n, \quad Y_1, Y_2, \dots, Y_m. \quad (4.82)$$

Tutte le variabili possono essere racchiuse in un unico vettore

$$\underline{Z} = (X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m)^T, \quad (4.83)$$

caratterizzato da una pdf gaussiana $(n + m)$ -dimensionale

$$f_{\underline{Z}}(\underline{z}) = \mathcal{N}(\underline{z}; \underline{\mu}_Z, \underline{\underline{C}}_Z), \quad (4.84)$$

in cui $\underline{\mu}_Z$ e $\underline{\underline{C}}_Z$ rappresentano rispettivamente media e covarianza di \underline{Z} . Le variabili \underline{X} e \underline{Y} sono anch'esse marginalmente gaussiane

$$f_{\underline{X}}(\underline{x}) = \mathcal{N}(\underline{x}; \underline{\mu}_X, \underline{\underline{C}}_X), \quad f_{\underline{Y}}(\underline{y}) = \mathcal{N}(\underline{y}; \underline{\mu}_Y, \underline{\underline{C}}_Y), \quad (4.85)$$

co medie e covarianze $\underline{\mu}_X$, $\underline{\mu}_Y$, $\underline{\underline{C}}_X$ e $\underline{\underline{C}}_Y$. La struttura della variabile \underline{Z} implica la seguente struttura della matrice di covarianza

$$\begin{aligned} \underline{\underline{C}}_Z &= E[(\underline{Z} - \underline{\mu}_Z)(\underline{Z} - \underline{\mu}_Z)^T] \\ &= E \left[\begin{pmatrix} \underline{X} - \underline{\mu}_X \\ \underline{Y} - \underline{\mu}_Y \end{pmatrix} \begin{pmatrix} \underline{X}^T - \underline{\mu}_X^T & \underline{Y}^T - \underline{\mu}_Y^T \end{pmatrix} \right] \\ &= \begin{bmatrix} \underline{\underline{C}}_X & \underline{\underline{C}}_{XY} \\ \underline{\underline{C}}_{YX} & \underline{\underline{C}}_Y \end{bmatrix}, \end{aligned} \quad (4.86)$$

4.4. LA VARIABILE GAUSSIANA MULTIDIMENSIONALE 155

dove \underline{C}_{XY} è la matrice di mutua covarianza

$$\underline{C}_{XY} = E[(\underline{X} - \underline{\mu}_X)(\underline{Y} - \underline{\mu}_Y)^T] = (E[(\underline{Y} - \underline{\mu}_Y)(\underline{X} - \underline{\mu}_X)^T])^T = \underline{C}_{YX}^T. \quad (4.87)$$

L'obiettivo è ricavare una espressione per la pdf di \underline{X} dato \underline{Y} , che può essere riscritta come

$$f_{\underline{X}}(\underline{x}|\underline{y}) = \frac{f_{\underline{X}\underline{Y}}(\underline{x}, \underline{y})}{f_{\underline{Y}}(\underline{y})} = \frac{f_{\underline{Z}}(\underline{z})}{f_{\underline{Y}}(\underline{y})}. \quad (4.88)$$

Più esplicitamente (tralasciando la doppia sottolineatura delle matrici per economia di notazione)

$$f_{\underline{X}}(\underline{x}|\underline{y}) = \frac{(2\pi)^{\frac{m}{2}} |C_Y|^{1/2}}{(2\pi)^{\frac{(n+m)}{2}} |C_Z|^{1/2}} \frac{\exp \left[-\frac{1}{2} \begin{pmatrix} \underline{x}^T - \underline{\mu}_X^T & \underline{y}^T - \underline{\mu}_Y^T \end{pmatrix} C_Z^{-1} \begin{pmatrix} \underline{x} - \underline{\mu}_X \\ \underline{y} - \underline{\mu}_Y \end{pmatrix} \right]}{\exp \left[-\frac{1}{2} (\underline{y}^T - \underline{\mu}_Y^T) C_Y^{-1} (\underline{y} - \underline{\mu}_Y) \right]}. \quad (4.89)$$

La seguente relazione matriciale può essere verificata direttamente effettuando la moltiplicazione blocco per blocco

$$\begin{aligned} & \begin{bmatrix} I & -C_{XY}C_Y^{-1} \\ 0 & I \end{bmatrix} C_Z \begin{bmatrix} I & 0 \\ -C_Y^{-1}C_{YX} & I \end{bmatrix} \\ &= \begin{bmatrix} I & -C_{XY}C_Y^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} C_X & C_{XY} \\ C_{YX} & C_Y \end{bmatrix} \begin{bmatrix} I & 0 \\ -C_Y^{-1}C_{YX} & I \end{bmatrix} \\ &= \begin{bmatrix} C_X - C_{XY}C_Y^{-1}C_{YX} & 0 \\ 0 & C_Y \end{bmatrix}. \end{aligned} \quad (4.90)$$

Da questa relazione immediatamente segue che

$$|C_Z| = |C_X - C_{XY}C_Y^{-1}C_{YX}| |C_Y|. \quad (4.91)$$

Invertendo ambo i membri di formula (4.90), abbiamo

$$\begin{aligned} & \begin{bmatrix} I & 0 \\ C_Y^{-1}C_{YX} & I \end{bmatrix}^{-1} C_Z^{-1} \begin{bmatrix} I & -C_{XY}C_Y^{-1} \\ 0 & I \end{bmatrix}^{-1} \\ &= \begin{bmatrix} (C_X - C_{XY}C_Y^{-1}C_{YX})^{-1} & 0 \\ 0 & C_Y^{-1} \end{bmatrix}. \end{aligned} \quad (4.92)$$

Pertanto

$$C_Z^{-1} = \begin{bmatrix} I & 0 \\ C_Y^{-1}C_{YX} & I \end{bmatrix} \begin{bmatrix} (C_X - C_{XY}C_Y^{-1}C_{YX})^{-1} & 0 \\ 0 & C_Y^{-1} \end{bmatrix} \begin{bmatrix} I & -C_{XY}C_Y^{-1} \\ 0 & I \end{bmatrix}. \quad (4.93)$$

Sostituendo equazioni (4.91) e (4.93) in (4.89)

$$f_{\underline{X}}(\underline{x}|\underline{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}|C_X - C_{XY}C_Y^{-1}C_{YX}|^{1/2}} \cdot \exp \left[-\frac{1}{2} (\underline{x}^T - \underline{x}_0^T) (C_X - C_{XY}C_Y^{-1}C_{YX})^{-1} (\underline{x} - \underline{x}_0) \right], \quad (4.94)$$

dove

$$\underline{x}_0 = \underline{\mu}_X + C_{XY}C_Y^{-1}(\underline{y} - \underline{\mu}_Y) = E[\underline{X}|\underline{Y} = \underline{y}], \quad (4.95)$$

è la media condizionata di \underline{X} dato $\underline{Y} = \underline{y}$. La matrice all'esponente è la covarianza di \underline{X} condizionata a $\underline{Y} = \underline{y}$

$$C_X - C_{XY}C_Y^{-1}C_{YX} = E[(\underline{X} - \underline{\mu}_X)(\underline{X} - \underline{\mu}_X)^T | \underline{Y} = \underline{y}]. \quad (4.96)$$

Tali relazioni saranno utili per ottenere lo stimatore a minimo errore quadratico medio quando le variabili in gioco sono congiuntamente gaussiane.

4.5 Stima MMSE

La formalizzazione della variabile aleatoria multipla ci consente di impostare in una elegante maniera il problema della stima di una o più variabili aleatorie dalla conoscenza del valore assunto da altre variabili aleatorie ad esse statisticamente collegate. In altre parole, se in un esperimento fisico è nota la caratterizzazione probabilistica congiunta delle variabili in gioco, è possibile ricavare la funzione ottima per stimare il valore di alcune di esse dalla conoscenza delle altre.

Esaminiamo per ora solo il caso in cui si voglia stimare una variabile aleatoria monodimensionale (desiderata) D dall'osservazione di altre n variabili aleatorie X_1, X_2, \dots, X_n ad essa statisticamente collegate. Figura 4.1 mostra lo schema di riferimento.

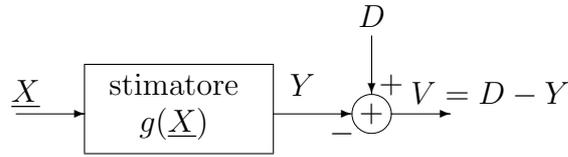


Figura 4.1: Il problema della stima di D da \underline{X} .

Per le $n + 1$ variabili aleatorie X_1, X_2, \dots, X_n, D , sia nota la pdf congiunta

$$f_{\underline{X}D}(x_1, \dots, x_n, d). \quad (4.97)$$

Vogliamo trovare la funzione (stimatore) che sia in grado di stimare D da \underline{X} secondo un criterio che minimizzi una misura dell'errore $V = D - g(\underline{X})$. Il criterio consiste nella minimizzazione dell'errore quadratico medio

$$R_{mq} = E[V^2] = E[(D - g(\underline{X}))^2]. \quad (4.98)$$

Il problema, noto come *Minimum Mean Square Error* (MMSE, minimo errore quadratico medio), è trovare la funzione $g_0(\underline{X})$ che minimizza R_{mq} . Ovvero

$$g_0(\underline{X}) = \arg \min_g E[(D - g(\underline{X}))^2]. \quad (4.99)$$

Per ottenere la soluzione, riscriviamo l'espressione della funzione di costo nella maniera seguente

$$\begin{aligned} R_{mq} &= \int_{\chi} \int_{\mathcal{D}} (d - g(\underline{x}))^2 f_{\underline{X}D}(x_1, \dots, x_n, d) dd d\underline{x} \\ &= \int_{\chi} f_{\underline{X}}(x_1, \dots, x_n) \int_{\mathcal{D}} (d - g(\underline{x}))^2 f_D(d|x_1, \dots, x_n) dd d\underline{x}, \end{aligned} \quad (4.100)$$

dove χ e \mathcal{D} sono i domini di definizione di \underline{x} e d . Notiamo che l'integrale è minimizzato se l'integrale interno è minimizzato visto che la pdf di \underline{X} è fissata ed è positiva. Il criterio può essere riscritto come

$$\begin{aligned} g_0(\underline{X}) &= \arg \min_g \int_{\mathcal{D}} (d - g(\underline{x}))^2 f_D(d|x_1, \dots, x_n) dd \\ &= \arg \min_g E[(D - g(\underline{X}))^2 | \underline{X} = \underline{x}]. \end{aligned} \quad (4.101)$$

Quindi la minimizzazione della media quadratica di V è equivalente alla minimizzazione della media quadratica di V condizionata da $\underline{X} =$

\underline{x} , $\forall \underline{x}$. La seguente semplice manipolazione della media condizionata di $E[V^2|\underline{X} = \underline{x}]$ porta alla soluzione.

$$\begin{aligned}
& E[(D - g(\underline{X}))^2|\underline{X} = \underline{x}] \\
&= \int_{\mathcal{D}} d^2 f_D(d|x_1, \dots, x_n) dd - 2g(\underline{x}) \int_{\mathcal{D}} df_D(d|x_1, \dots, x_n) dd \\
&+ g(\underline{x})^2 \int_{\mathcal{D}} f_D(d|x_1, \dots, x_n) dd \\
&= \left(g(\underline{x}) - \int_{\mathcal{D}} d f_D(d|x_1, \dots, x_n) dd \right)^2 \\
&+ \int_{\mathcal{D}} d^2 f_D(d|x_1, \dots, x_n) dd - \left(\int_{\mathcal{D}} df_D(d|x_1, \dots, x_n) dd \right)^2. \quad (4.102)
\end{aligned}$$

Poiché gli ultimi due termini dell'ultima espressione non dipendono da g , la minimizzazione si realizza quando il primo termine si azzerava, ovvero per

$$g_0(\underline{x}) = E[D|\underline{X} = \underline{x}] = \int_{\mathcal{D}} d f_D(d|x_1, \dots, x_n) dd. \quad (4.103)$$

Pertanto:

Il miglior stimatore di D da $X = x$ è la media di D condizionata a $\underline{X} = \underline{x}$.

In generale lo stimatore è una funzione non lineare di \underline{x} , la cui complessità dipende dalla distribuzione condizionata di D dato X . Vedremo comunque che in alcuni esempi, lo stimatore a media condizionata può diventare una funzione lineare di \underline{x} . La forma di equazione (4.102) è detta *forma canonica dell'errore quadratico medio*. E' utile ricavare anche il valore quadratico medio dell'errore in condizioni di ottimalità

$$\begin{aligned}
R_{mq0} &= E[(D - g_0(\underline{X}))^2] \\
&= E[D^2] + E[g_0(\underline{X})^2] - 2E[Dg_0(\underline{X})] \\
&= E[D^2] + E[E[D|\underline{X} = \underline{x}]g_0(\underline{X})] - 2E[Dg_0(\underline{X})] \\
&= E[D^2] + E[E[Dg_0(\underline{X})|\underline{X} = \underline{x}]] - 2E[Dg_0(\underline{X})] \\
&= E[D^2] - E[Dg_0(\underline{X})]. \quad (4.104)
\end{aligned}$$

Una interessante interpretazione di questa espressione può essere data osservando che uno stimatore nullo darebbe errore pari a $E[D^2]$ (nessuna

stima). Quindi l'errore che si riesce a ottenere con lo stimatore ottimo è tanto migliore quanto maggiore è la correlazione $E[Dg_0(\underline{X})]$ che può essere ottenuta tra D e l'uscita dello stimatore.

4.5.1 Il caso gaussiano

Il risultato del problema assume una forma speciale quando le variabili \underline{X} e D sono congiuntamente gaussiane. Basta usare i risultati ottenuti nella sezione 4.4.1 a riguardo della media di variabili gaussiane condizionate ad altre variabili gaussiane per ottenere lo stimatore ottimo

$$g_0(\underline{x}) = \mu_D + \underline{C}_{DX} \underline{C}_X^{-1} (\underline{x} - \underline{\mu}_X). \quad (4.105)$$

Lo stimatore ottimo è una funzione *lineare*² di \underline{x} . Il valore quadratico medio dell'errore in condizioni di ottimalità è

$$R_{mq0} = \text{var}[D] - \underline{C}_{DX} \underline{C}_X^{-1} \underline{C}_{XD}. \quad (4.106)$$

Si noti che la matrice di mutua covarianza $\underline{C}_{DX} = \underline{C}_{XD}^T$ è qui un vettore riga.

Esempio 4.1 Consideriamo il seguente esempio scalare. Si supponga che sia necessario stimare una variabile aleatoria D da una variabile scalare osservata $X = D + W$, dove W è una variabile di *disturbo* a media nulla e indipendente da D . La variabile W modella una sorta di errore aleatorio nell'osservare D . Si supponga che sia D che W siano gaussiane a media nulla e varianza σ_D^2 e σ_W^2 . Vogliamo trovare lo stimatore a minimo errore quadratico medio per D . La media di X è nulla poiché $E[X] = E[D + W] = E[D] + E[W]$. Applicando il risultato della sezione precedente abbiamo

$$g_0(x) = C_{DX} C_X^{-1} x. \quad (4.107)$$

Più specificamente

$$C_{DX} = E[DX] = E[D(D + W)] = E[D^2] + E[D]E[W] = E[D^2] = \sigma_D^2. \quad (4.108)$$

²In effetti si tratta in generale di una funzione affine poichè compare anche una costante. Lo stimatore è ovviamente lineare se le medie sono nulle.

$$C_X = E[X^2] = E[(D+W)^2] = E[D^2] + E[W^2] + 2E[DW] = \sigma_D^2 + \sigma_W^2. \quad (4.109)$$

Quindi

$$g_0(x) = \frac{\sigma_D^2}{\sigma_D^2 + \sigma_W^2} x. \quad (4.110)$$

Stimare D consiste in scalare X di una quantità che dipende dal rapporto tra le varianze di D e di X . L'errore quadratico medio in condizioni di ottimo è

$$\begin{aligned} R_{mq0} &= E[D^2] - E[Dg_0(X)] = E[D^2] - \frac{E[DX]\sigma_D^2}{\sigma_D^2 + \sigma_W^2} \\ &= \sigma_D^2 - \frac{\sigma_D^4}{\sigma_D^2 + \sigma_W^2} = \frac{\sigma_D^2 \sigma_W^2}{\sigma_D^2 + \sigma_W^2}. \end{aligned} \quad (4.111)$$

Esempio 4.2 Questo esempio rappresenta una situazione classica nella teoria della probabilità. Si voglia stimare una variabile aleatoria costante, ma non nota, D da campioni rumorosi di essa. Ovvero, le osservazioni in \underline{X} siano ottenute come

$$\underline{X} = D\underline{e} + \underline{W}, \quad (4.112)$$

dove $\underline{e} = (1, \dots, 1)^T$ e \underline{W} è un vettore aleatorio gaussiano a media nulla e componenti indipendenti aventi varianza σ_W^2 . Pertanto la matrice di covarianza di \underline{W} è

$$\underline{C}_{\underline{W}} = \sigma_W^2 \underline{I}. \quad (4.113)$$

Si assuma che D e \underline{W} siano indipendenti tra loro. Vogliamo stimare D da \underline{X} .

La soluzione richiede il calcolo esplicito di C_{DX} e C_X^{-1} . In particolare il vettore di mutua covarianza, che a causa delle medie nulle è anche la mutua correlazione, diventa

$$\underline{C}_{DX} = E[D\underline{X}^T] = E[D(D\underline{e} + \underline{W})^T] = \sigma_D^2 \underline{e}^T + E[D\underline{W}^T] = \sigma_D^2 \underline{e}^T. \quad (4.114)$$

La matrice di covarianza di \underline{X} , che è anche la matrice di correlazione, diventa

$$\underline{C}_{\underline{X}} = E[\underline{X}\underline{X}^T] = E[(D\underline{e} + \underline{W})(D\underline{e} + \underline{W})^T] = E[D^2] \underline{e} \underline{e}^T + \sigma_W^2 \underline{I}. \quad (4.115)$$

Lo stimatore ottimo è quindi

$$g_0(\underline{x}) = \sigma_D^2 \underline{e}^T (E[D^2] \underline{e} \underline{e}^T + \sigma_W^2 \underline{I})^{-1} \underline{x}. \quad (4.116)$$

Per ottenere una versione più esplicita dello stimatore, è necessario eseguire l'inversione della matrice in parentesi. A tale scopo, è possibile usare il *Lemma dell'inversione delle matrici* descritto in Appendice C. Ponendo

$$\begin{cases} B^{-1} = \sigma_W^{-2} I; \\ C = \underline{e}; \\ D^{-1} = \sigma_D^{-2}; \end{cases} \quad (4.117)$$

abbiamo

$$\begin{aligned} (E[D^2] \underline{e} \underline{e}^T + \sigma_W^2 I)^{-1} &= \sigma_W^{-2} I - \sigma_W^{-2} I \underline{e} (\sigma_D^{-2} + \underline{e}^T \sigma_W^{-2} I \underline{e})^{-1} \underline{e}^T \sigma_W^{-2} I \\ &= \sigma_W^{-2} I - \sigma_W^{-4} \underline{e} \left(\sigma_D^{-2} + \frac{n}{\sigma_W^2} \right)^{-1} \underline{e}^T. \end{aligned} \quad (4.118)$$

Lo stimatore diventa

$$\begin{aligned} g_0(\underline{x}) &= \frac{\sigma_D^2}{\sigma_W^2} \left(1 - \frac{n\sigma_D^2}{\sigma_W^2 + n\sigma_D^2} \right) \underline{e}^T \underline{x} \\ &= \frac{\sigma_D^2}{\sigma_W^2 + n\sigma_D^2} \underline{e}^T \underline{x} = \frac{\sigma_D^2}{\sigma_W^2 + n\sigma_D^2} \sum_{i=1}^n x_i. \end{aligned} \quad (4.119)$$

Lo stimatore è simile alla media aritmetica, ma include l'informazione sulla varianza di D e di W nel coefficiente di scala. Se il rumore W ha una varianza molto piccola rispetto a quella di D , ovvero $\sigma_W^2 \ll \sigma_D^2$, abbiamo che

$$g_0(\underline{x}) \simeq \frac{1}{n} \sum_{i=1}^n x_i, \quad (4.120)$$

ovvero lo stimatore coincide approssimativamente alla media aritmetica.

Capitolo 5

Funzioni di variabili aleatorie

In questo capitolo vengono introdotte le principali tecniche per caratterizzare le variabili aleatorie dopo che esse siano state trasformate secondo delle funzioni. Viene trattato prima il caso della trasformazione di una variabile in un'altra, e successivamente quello della trasformazione di due o più variabili in un'altra variabile. Viene inoltre presentato un risultato fondamentale nella teoria della probabilità che descrive la distribuzione asintotica della somma di più variabili aleatorie: il teorema centrale del limite.

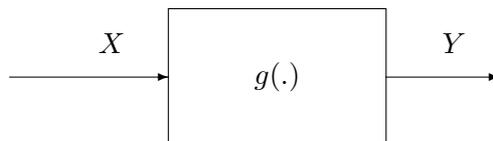
5.1 Introduzione

Nel trattamento e nello studio dei segnali di informazione, mediante la formulazione probabilistica, è necessario avere a disposizione le metodologie che consentano di caratterizzare le variabili in gioco, dopo che esse siano state trasformate secondo delle funzioni note. Si pensi ad un segnale aleatorio trasformato da un amplificatore distortore. L'uscita del sistema è una variabile aleatoria dalle caratteristiche diverse da quelle dell'ingresso che è necessario saper caratterizzare per potere valutare l'effetto distortore e amplificante introdotto dal sistema. Come altro esempio si pensi a un sistema di comunicazione in cui una variabile aleatoria rappresenta un segnale di informazione. Se tale segnale è con-

taminato da un disturbo additivo, modellato anch'esso come una variabile aleatoria, è cruciale per lo studio del segnale saper caratterizzare la variabile aleatoria risultante e le relazioni reciproche tra le componenti. In particolare è necessario saper valutare per il segnale risultante pdf o cdf, correlazioni mutue tra le varie componenti, eccetera. Gli esempi di funzioni di variabili aleatorie nelle applicazioni sono tantissimi. Il nostro scopo è presentare le principali tecniche di manipolazione con riferimento ad esempi notevoli di trasformazioni tra una o più variabili che consentano allo studente di acquisire la metodologia relativa.

5.2 Funzioni di una variabile

Cominciamo con il caso più semplice di una singola variabile aleatoria trasformata mediante una funzione data in un'altra variabile aleatoria. Più precisamente si consideri una variabile aleatoria X e una funzione deterministica $g(X)$, che applicata ad ogni realizzazione di X , ci fornisce una nuova variabile Y come mostrato in figura.



Più formalmente, per ogni realizzazione x di X ,

$$x \in \mathcal{A} \longleftrightarrow y = g(x) \in \mathcal{B}, \quad (5.1)$$

dove \mathcal{A} è l'insieme di definizione di x e \mathcal{B} quello di y . Il problema che ci poniamo in questa sezione è la caratterizzazione di Y , note la caratterizzazione di X e la funzione g . Il problema del calcolo della pdf o della cdf di Y richiede innanzitutto un esame attento del dominio \mathcal{A} di definizione di X e del codominio \mathcal{B} su cui assume valori la variabile Y . Potremmo infatti avere a che fare con una variabile X di tipo continuo e di una Y continua o discreta; oppure con una X variabile discreta e Y variabile discreta anch'essa; oppure con entrambi X e Y variabili miste. Anche le strutture di \mathcal{A} e \mathcal{B} possono essere a limitate o illimitate e la

funzione g continua o discontinua, invertibile o non invertibile. In tutti questi casi dovremo imparare a caratterizzare la variabile aleatoria Y . Cominciamo col mostrare qualche esempio.

Esempio 5.1 Si consideri la variabile aleatoria X gaussiana a media nulla e varianza σ^2 , e la funzione $g(X) = X^2$. Il dominio di definizione di X è chiaramente tutto l'asse reale ($\mathcal{A} = \mathcal{R}$), mentre il dominio di Y è il solo semiasse positivo ($\mathcal{B} = \mathcal{R}^+$). La funzione non è invertibile ed è a due valori, in quanto ad ogni valore di $y > 0$ corrispondono i due valori $\pm\sqrt{y}$. La pdf di Y sarà nulla sul semiasse negativo e dovrà contenere tutta la massa di probabilità sul semiasse positivo. Vedremo in seguito come trovare tale pdf esplicitamente.

Esempio 5.2 Si consideri una variabile aleatoria gaussiana X a media nulla e varianza σ^2 , e la trasformazione $Y = \text{sgn}(X)$. Il dominio di definizione di X è tutto l'asse reale ($\mathcal{A} = \mathcal{R}$), mentre il dominio di Y è l'insieme discreto $\mathcal{B} = \{-1, 1\}$. La variabile Y è di tipo discreto a due valori con tutta la massa di probabilità concentrata in due punti. Il calcolo della pdf di Y è immediato ed è

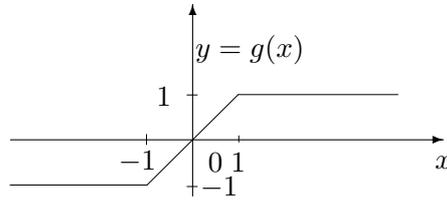
$$f_Y(y) = \frac{1}{2}\delta(y+1) + \frac{1}{2}\delta(y-1), \quad (5.2)$$

in quanto tutti valori positivi di X (probabilità 0.5) corrispondono a $Y = 1$ e quelli negativi a $Y = -1$.

Esempio 5.3 Si consideri una variabile aleatoria X gaussiana a media nulla e varianza σ^2 e la trasformazione

$$y = g(x) = \begin{cases} -1 & x < -1 \\ x & -1 < x < 1 \\ 1 & x > 1. \end{cases} \quad (5.3)$$

Il dominio di definizione di X è tutto l'asse reale ($\mathcal{A} = \mathcal{R}$), mentre il dominio di Y è l'intervallo $\mathcal{B} = [-1, 1]$. Sui punti $y = -1$ e $y = 1$ si concentra una parte della massa di probabilità di Y , in quanto per tutti i valori di $X > 1$, la funzione dà il valore saturato $Y = 1$, mentre per i valori di $X < -1$ si ha $Y = -1$. Per gli altri valori la trasformazione è la funzione identità e pertanto la distribuzione resta invariata. La figura seguente mostra la trasformazione $y = g(x)$.



La densità dell'uscita è facilmente scritta come

$$f_Y(y) = Pr\{X < -1\}\delta(y + 1) + \mathcal{N}(y; 0, \sigma^2)[u(y + 1) - u(y - 1)] + Pr\{X > 1\}\delta(y - 1). \quad (5.4)$$

Esempio 5.4 Si consideri la variabile aleatoria discreta X che descrive il valore numerico del risultato del lancio di un dado non truccato. La funzione $g(X)$ sia una funzione di parità, ovvero

$$Y = g(X) = \begin{cases} 0 & \text{per } X \text{ pari} \\ 1 & \text{per } X \text{ dispari.} \end{cases} \quad (5.5)$$

Il dominio di X è l'insieme discreto $\mathcal{A} = \{1, 2, 3, 4, 5, 6\}$, mentre il dominio di Y è l'insieme $\mathcal{B} = \{0, 1\}$. Entrambi le variabili sono di tipo discreto, quindi basta enumerare i valori assunti dalla funzione e calcolarne la probabilità. La massa di probabilità di X è distribuita uniformemente come $\Pi_X = \{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}$. La densità di Y è di calcolo immediato per semplice enumerazione poiché

$$\begin{aligned} Pr\{Y = 0\} &= Pr\{X = 2\} + Pr\{X = 4\} + Pr\{X = 6\} = 3\frac{1}{6} = \frac{1}{2} \\ Pr\{Y = 1\} &= Pr\{X = 1\} + Pr\{X = 3\} + Pr\{X = 5\} = 3\frac{1}{6} = \frac{1}{2}. \end{aligned} \quad (5.6)$$

Gli esempi riportati sono piuttosto semplici e mostrano come in alcuni casi sia abbastanza immediato ottenere la distribuzione di Y con una tecnica diretta e intuitiva. Sfortunatamente per generiche funzioni il calcolo può presentare qualche insidia in più. Esistono vari approcci per ottenere la distribuzione di Y . Essi risultano più o meno agili a seconda della tipologia di variabili e di trasformazione. Il lettore dopo un po' di pratica, imparerà a selezionare la tecnica più appropriata, che magari gli è più congeniale e che gli consente di raggiungere il risultato più rapidamente.

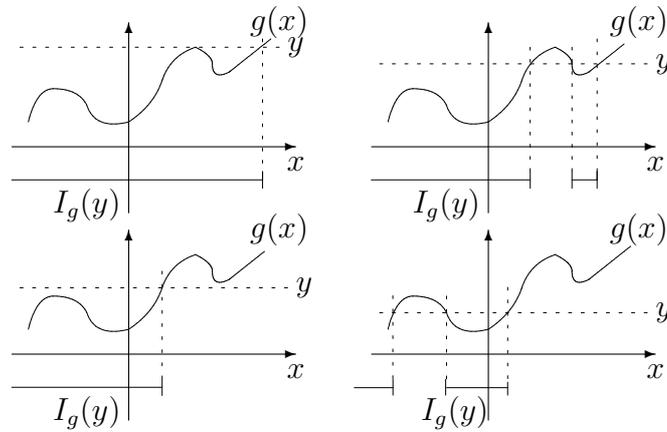


Figura 5.1: L'insieme $I_g(y)$ per il metodo della cdf.

5.2.1 Il metodo della cdf

Uno dei metodi più generali per trovare la distribuzione di probabilità di Y è cercare di ottenere la sua cdf partendo dalla definizione

$$F_Y(y) = Pr\{Y \leq y\} = Pr\{g(X) \leq y\}. \quad (5.7)$$

Poiché è nota la distribuzione di X , bisogna cercare di esprimere l'evento $\{g(X) \leq y\}$, come evento su X . Se la funzione fosse invertibile ovunque, avremmo ovviamente $\{g(X) \leq y\} = \{X \leq g^{-1}(y)\}$. In generale ciò non è vero e pertanto bisogna esprimere l'evento come

$$\{g(X) \leq y\} = \{X \in I_g(y)\}, \quad (5.8)$$

dove $I_g(y)$ è l'insieme del dominio di definizione di X su cui è verificata la disuguaglianza $g(X) \leq y$. La cdf di Y è pertanto

$$F_Y(y) = Pr\{X \in I_g(y)\}. \quad (5.9)$$

La figura 5.1 mostra per una generica funzione $g(x)$, come varia l'insieme $I_g(y)$ in funzione di y . Si noti come l'insieme possa anche essere non semplicemente connesso. Per ogni valore di y , bisogna valutare la $Pr\{X \in I_g(y)\}$. Vediamo come applicare il metodo ad alcuni esempi.

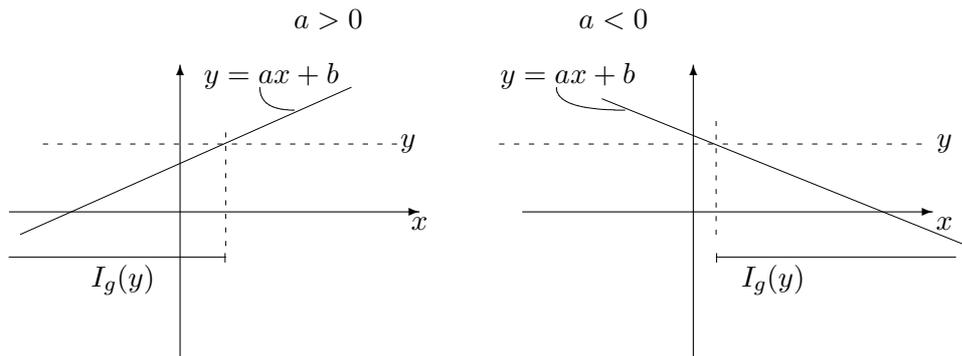
Esempio 5.5 Consideriamo una variabile aleatoria X di tipo continuo, definita su tutto l'asse reale e la funzione $Y = aX + b$, con a e b due costanti reali. Vogliamo calcolare la distribuzione di Y . Si tratta di un caso molto semplice in quanto la funzione è invertibile. Usando il metodo della cdf, per $a > 0$, possiamo scrivere

$$\begin{aligned} F_Y(y) &= Pr\{Y \leq y\} = Pr\{X \in I_g(y)\} \\ &= Pr\{aX + b \leq y\} = Pr\left\{X \leq \frac{y-b}{a}\right\} = F_X\left(\frac{y-b}{a}\right). \end{aligned} \quad (5.10)$$

Se invece $a < 0$ abbiamo

$$\begin{aligned} F_Y(y) &= Pr\{Y \leq y\} = Pr\{X \in I_g(y)\} \\ &= Pr\{aX + b \leq y\} = Pr\left\{X > \frac{y-b}{a}\right\} = 1 - F_X\left(\frac{y-b}{a}\right). \end{aligned} \quad (5.11)$$

La figura seguente mostra schematicamente la funzione e l'insieme $I_g(y)$ rispettivamente per $a > 0$ e per $a < 0$.



Per $a > 0$ la pdf è

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right), \quad (5.12)$$

mentre per $a < 0$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right). \quad (5.13)$$

Sinteticamente, per ogni a abbiamo

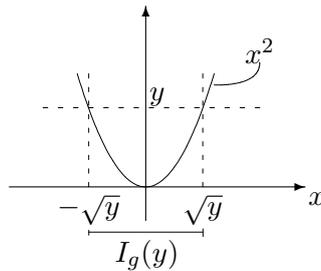
$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right). \quad (5.14)$$

Quindi la trasformazione lineare trasla e contrae (se $|a| > 1$), o espande (se $|a| < 1$), la distribuzione.

Esempio 5.6 Riprendiamo l'esempio della funzione $g(X) = X^2$. La variabile aleatoria X sia di tipo continuo e avente una pdf definita su tutto l'asse reale. La cdf di Y può essere scritta come

$$F_Y(y) = Pr\{Y \leq y\} = Pr\{X^2 \leq y\} = Pr\{X \in I_g(y)\}. \quad (5.15)$$

Ora se $y < 0$, l'insieme $I_g(y)$ è l'insieme vuoto e pertanto $F_Y(y) = 0$. Diversamente per $y \geq 0$ abbiamo $I_g(y) = [-\sqrt{y}, \sqrt{y}]$. La figura seguente mostra la trasformazione $y = x^2$ e l'insieme $I_g(y)$ per $y > 0$.



La cdf di Y è pertanto

$$F_Y(y) = \begin{cases} 0 & \text{per } y < 0 \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}) & \text{per } y \geq 0. \end{cases} \quad (5.16)$$

Notiamo che la cdf di Y è continua su tutto l'asse y e in particolare in $y = 0$. Pertanto nella pdf, che calcoliamo per semplice derivazione, non ci saranno delta di Dirac.

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} 0 & \text{per } y < 0 \\ \frac{d}{dy} F_X(\sqrt{y}) - \frac{d}{dy} F_X(-\sqrt{y}) & \text{per } y \geq 0. \end{cases} \quad (5.17)$$

$$f_Y(y) = \begin{cases} 0 & \text{per } y < 0 \\ \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})], & \text{per } y \geq 0. \end{cases} \quad (5.18)$$

Se la pdf di X è simmetrica attorno allo zero ($f_X(x) = f_X(-x)$; $F_X(-x) = 1 - F_X(x)$), le espressioni per cdf e pdf si semplificano

$$\begin{aligned} F_Y(y) &= \begin{cases} 0 & \text{per } y < 0 \\ 2F_X(\sqrt{y}) - 1 & \text{per } y \geq 0 \end{cases} \\ f_Y(y) &= \begin{cases} 0 & \text{per } y < 0 \\ \frac{1}{\sqrt{y}} f_X(\sqrt{y}) & \text{per } y \geq 0 \end{cases} \end{aligned} \quad (5.19)$$

Il caso particolare di X gaussiana a media μ e varianza σ^2 , corrisponde alle seguenti cdf e pdf.

$$F_Y(y) = \begin{cases} 0 & \text{per } y < 0 \\ \Phi\left(\frac{\sqrt{y}-\mu}{\sigma}\right) - \Phi\left(\frac{-\sqrt{y}-\mu}{\sigma}\right) & \text{per } y \geq 0, \end{cases} \quad (5.20)$$

dove la funzione Φ è la cdf della gaussiana normalizzata. Inoltre, $f_Y(y) = 0$ per $y < 0$, e

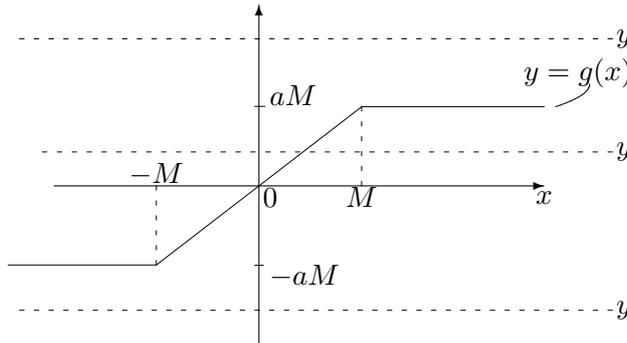
$$f_Y(y) = \frac{1}{2\sqrt{2\pi y}\sigma} \left(e^{-\frac{(\sqrt{y}-\mu)^2}{2\sigma^2}} + e^{-\frac{(-\sqrt{y}-\mu)^2}{2\sigma^2}} \right), \quad (5.21)$$

per $y \geq 0$.

Esempio 5.7 Consideriamo ora l'amplificatore lineare con saturazione descritto dalla relazione

$$Y = g(X) = \begin{cases} -aM & \text{per } -\infty < X < -M \\ aX & \text{per } -M < X < M \\ aM & \text{per } M < X < \infty, \end{cases} \quad (5.22)$$

con X variabile aleatoria di tipo continuo definita su tutto l'asse reale. La figura seguente mostra la funzione e come valutare l'insieme $I_g(y)$ per vari valori di y .



Per $y > aM$, $I_g(y) = \mathcal{R}$, e pertanto

$$F_Y(y) = Pr\{Y \leq y\} = 1. \quad (5.23)$$

Per $-aM < y < aM$, $I_g(y) = \{X \leq \frac{y}{a}\}$ e

$$F_Y(y) = Pr\{Y \leq y\} = Pr\{X \leq \frac{y}{a}\} = F_X\left(\frac{y}{a}\right). \quad (5.24)$$

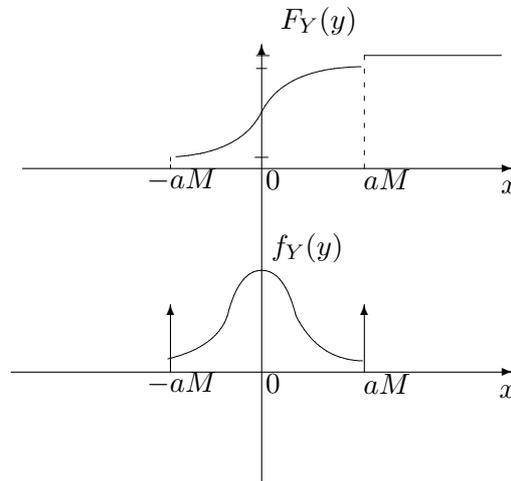
Per $y < -aM$, $I_g(y) = \phi$, e pertanto

$$F_Y(y) = Pr\{Y \leq y\} = 0. \quad (5.25)$$

In sintesi

$$F_Y(y) = \begin{cases} 1 & \text{per } y \geq aM \\ F_X\left(\frac{y}{a}\right) & \text{per } -aM \leq y < aM \\ 0 & \text{per } y < -aM. \end{cases} \quad (5.26)$$

Si noti che se la densità di X è diversa da zero su tutto l'asse reale, $F_Y(y)$ è discontinua con dei salti in $y = aM$ e $y = -aM$, in quanto $F_X(M) < 1$ e $F_X(-M) > 0$. La pdf di Y ha pertanto delle delta di Dirac in quei punti. Questo conferma che una massa di probabilità si concentra sui valori di saturazione $\pm aM$. La pdf è calcolata esplicitamente mediante derivazione. Attenzione deve essere riservata ai punti discontinuità dove le delta di Dirac assumono area pari alla differenza tra il limite destro e quello sinistro di $F_Y(y)$. La figura seguente mostra un tipica cdf e la pdf derivata.



L'espressione analitica è ottenuta facilmente

$$\begin{aligned}
 f_Y(y) &= F_Y(-aM^+)\delta(y+aM) + \left(\frac{d}{dy}F_Y(y)\right)[u(y+aM) - u(y-aM)] \\
 &\quad + (1 - F_Y(aM^-))\delta(y-aM) \\
 &= F_X(M)\delta(y+aM) + \frac{1}{a}F_x\left(\frac{y}{a}\right)[u(y+aM) - u(y-aM)] \\
 &\quad + (1 - F_X(M))\delta(y-aM).
 \end{aligned} \tag{5.27}$$

5.2.2 Il metodo della pdf

La caratterizzazione di Y può anche essere direttamente ottenuta in riferimento alla sua pdf in funzione della pdf di X . Tale tecnica è però limitata al caso in cui la funzione $g(x)$ sia derivabile.

Teorema 5.1 Sia X una variabile aleatoria avente pdf $f_X(x)$. La pdf $f_Y(y)$ della variabile $Y = g(X)$, può essere ottenuta come segue.

Per i valori di y in cui l'equazione $y = g(x)$ non ammette soluzione, $f_Y(y) = 0$.

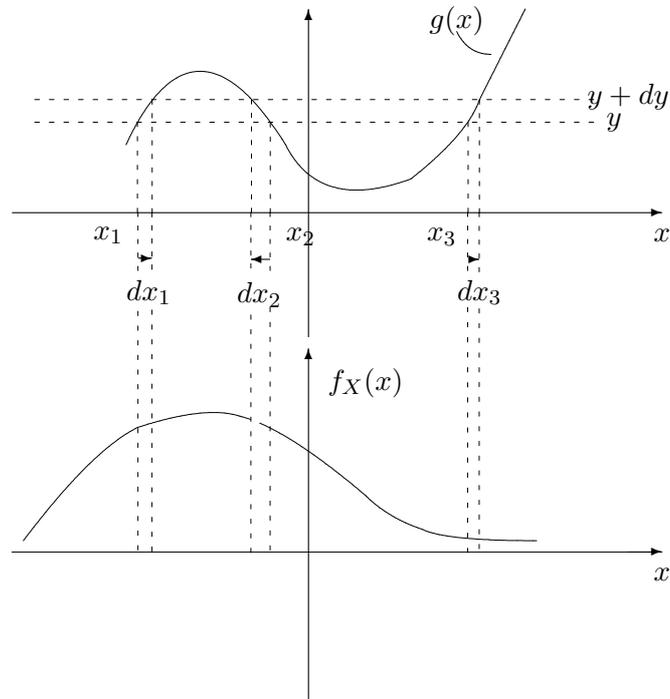
Per i valori di y in cui l'equazione $y = g(x)$ ammette le n soluzioni x_1, x_2, \dots, x_n , la densità è

$$f_Y(y) = \frac{f_X(x_1)}{|g'(x_1)|} + \frac{f_X(x_2)}{|g'(x_2)|} + \dots + \frac{f_X(x_n)}{|g'(x_n)|}, \tag{5.28}$$

dove $g'(x_i)$ indica la derivata prima di g in x_i , $i = 1, \dots, n$.

Si noti che nella espressione della pdf, le radici dell'equazione $y = g(x)$ dipendono da y , anche se ciò non è direttamente specificato. Infatti, a diversi valori di y possono corrispondere insiemi diversi di soluzioni. Inoltre, affinché il metodo sia applicabile, la funzione g deve essere derivabile per ogni y nelle radici x_1, \dots, x_n . Alcuni esempi mostreranno l'efficacia del metodo.

Prova: Per dimostrare il teorema facciamo riferimento alla funzione di figura, che per il valore di y tracciato, ha le tre radici reali x_1, x_2, x_3 .



La pdf di Y è tale che

$$f_Y(y) dy = Pr\{y < Y \leq y + dy\}. \quad (5.29)$$

Quindi per calcolare la pdf di Y , bisogna valutare la probabilità che X assuma i valori tali che

$$Pr\{y < Y \leq y + dy\} = Pr\{y < g(X) \leq y + dy\}. \quad (5.30)$$

Questi valori saranno negli intorni delle tre radici x_1, x_2, x_3 , ovvero

$$\begin{aligned} Pr\{y < Y \leq y + dy\} &= Pr\{x_1 < X \leq x_1 + dx_1\} \\ &+ Pr\{x_2 + dx_2 < X \leq x_2\} + Pr\{x_3 < X \leq x_3 + dx_3\} \\ &= f_X(x_1)dx_1 + f_X(x_2)|dx_2| + f_X(x_3)dx_3. \end{aligned} \quad (5.31)$$

Poiché $g'(x_i)dx_i = dy$, $i = 1, 2, 3$, abbiamo

$$f_Y(y)dy = \frac{f_X(x_1)}{|g'(x_1)|}dy + \frac{f_X(x_2)}{|g'(x_2)|}dy + \frac{f_X(x_3)}{|g'(x_3)|}dy, \quad (5.32)$$

che dimostra il risultato del teorema.

Esempio 5.8 Ritorniamo all'esempio della funzione lineare $Y = aX + b$, con a e b due costanti reali arbitrarie. Il metodo della pdf è facilmente applicabile in quanto per ogni y , l'equazione $y = ax + b$, ammette l'unica soluzione $x_1 = \frac{y-b}{a}$. Inoltre poiché $g'(x) = a$, la pdf di Y secondo il teorema 5.1 è

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right). \quad (5.33)$$

Si tratta ovviamente dello stesso risultato ottenuto con il metodo della cdf. In questo caso il metodo della pdf è sicuramente più rapido.

Esempio 5.9 Ritorniamo all'esempio della funzione $Y = X^2$ per ritrovare con il metodo della pdf la densità di Y . Per $y < 0$, la equazione $y = x^2$, non ammette soluzione pertanto $f_Y(y) = 0$. Per $y \geq 0$, la equazione $y = x^2$ ammette le due soluzioni $x_1 = \sqrt{y}$, $x_2 = -\sqrt{y}$. Poiché $g'(x) = 2x$, la formula del teorema 5.1 ci fornisce il risultato

$$\begin{aligned} f_Y(y) &= \frac{f_X(x_1)}{|g'(x_1)|} + \frac{f_X(x_2)}{|g'(x_2)|} \\ &= \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})], \quad \text{per } y \geq 0. \end{aligned} \quad (5.34)$$

Esempio 5.10 Valutiamo la densità di $Y = g(X) = |X|$ (raddrizzatore a doppia semi-onda). La variabile X sia di tipo continuo distribuita secondo la densità $f_X(x)$. Per $y < 0$, la equazione $y = |x|$ non ammette soluzione, quindi $f_Y(y) = 0$. Per $y \geq 0$, la equazione $y = |x|$ ammette le radici $x_1 = y$, $x_2 = -y$. Poiché $g'(x) = \text{sgn}(x)$, la formula del teorema 5.1 fornisce il risultato

$$f_Y(y) = \frac{f_X(x_1)}{|\text{sgn}(x_1)|} + \frac{f_X(x_2)}{|\text{sgn}(x_2)|} = f_X(y) + f_X(-y), \quad \text{per } y \geq 0. \quad (5.35)$$

Esempio 5.11 Consideriamo l'esempio della funzione iperbolica $Y = g(X) = \frac{a}{X}$. Sia X una variabile aleatoria con densità $f_X(x)$. Per ogni valore di y , la equazione $y = \frac{a}{x}$, fornisce l'unica soluzione $x_1 = \frac{a}{y}$. Poiché

$$g'(x) = -\frac{a}{x^2}, \quad (5.36)$$

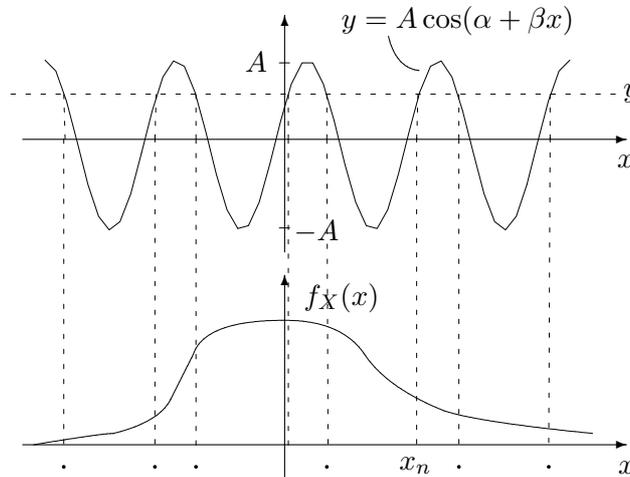
La formula del teorema 5.1 fornisce il risultato

$$f_Y(y) = \frac{f_X(x_1)}{\left| -\frac{a}{x_1^2} \right|} = \frac{|a|}{y^2} f_X\left(\frac{a}{y}\right). \quad (5.37)$$

Esempio 5.12 Un esempio tipico è il calcolo della densità della variabile Y ottenuta dalla funzione

$$Y = A \cos(\alpha + \beta X), \quad (5.38)$$

con $A > 0$ e α e β costanti reali arbitrarie. Un segmento della funzione è mostrato nella seguente figura



Il dominio di Y è l'intervallo $\mathcal{B} = [-A, A]$, pertanto, per $y < -A$ e per $y > A$, $f_Y(y) = 0$. Per $-A < y < A$, l'equazione $y = A \cos(\alpha + \beta x)$ ammette le infinite soluzioni fornite dalla formula

$$x_n = \frac{1}{\beta} \left[\cos^{-1} \left(\frac{y}{A} \right) - \alpha \right], \quad (5.39)$$

in quanto la funzione \cos^{-1} fornisce infinite soluzioni (vedi figura): se θ è l'argomento principale di $\cos^{-1} \frac{y}{A}$, le soluzioni sono

$$\frac{1}{\beta}(-\theta + 2\pi n - \alpha), \quad \frac{1}{\beta}(\theta + 2\pi n - \alpha), \quad n = \dots, -1, 0, 1, \dots \quad (5.40)$$

La derivata di $g(x)$ è $g'(x) = -A\beta \sin(\alpha + \beta x)$, il cui valore assoluto, valutato nelle soluzioni x_n , fornisce sempre

$$\begin{aligned} |g'(x_n)| &= A|\beta| |\sin(\alpha + \beta x_n)| = A|\beta| \sqrt{1 - \cos^2(\alpha + \beta x_n)} \\ &= A|\beta| \sqrt{1 - \left(\frac{y}{A}\right)^2} \\ &= |\beta| \sqrt{A^2 - y^2}. \end{aligned} \quad (5.41)$$

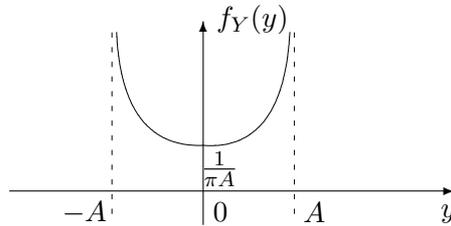
La formula del teorema 5.1 ci fornisce l'espressione

$$f_Y(y) = \frac{1}{|\beta| \sqrt{A^2 - y^2}} \sum_n f_X(x_n). \quad (5.42)$$

Variabile X uniforme: Calcoliamo esplicitamente la pdf di Y quando $\beta = 1$ e X è una variabile aleatoria uniforme in $[-\pi, \pi]$. In tal caso sono solo due le soluzioni dell'equazione $y = g(x)$ che corrispondono a valori della pdf di X diversi da zero. Trattandosi di una pdf uniforme $f_X(x_n) = \frac{1}{2\pi}$ per entrambi. Abbiamo pertanto

$$f_Y(y) = \begin{cases} \frac{1}{\pi \sqrt{A^2 - y^2}} & \text{per } -A < y < A \\ 0 & \text{altrove.} \end{cases} \quad (5.43)$$

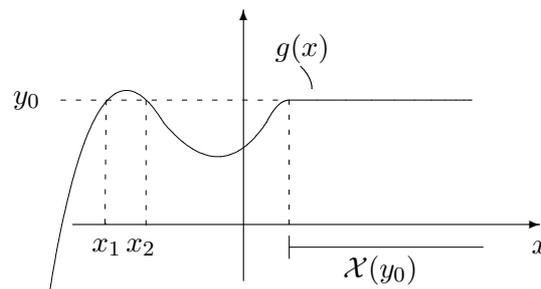
Il tipico andamento della funzione è riportato in figura.



Si noti che nonostante la funzione vada a infinito in $\pm A$, non ci sono impulsi di Dirac e pertanto $Pr\{Y = A\} = Pr\{Y = -A\} = 0$.

Il teorema 5.1 per il calcolo diretto della pdf è limitato al caso in cui l'equazione $y = g(x)$ presenta un insieme discreto di soluzioni x_1, x_2, \dots

Ciò richiede che la funzione sia localmente monotona. E' possibile generalizzare il risultato includendo anche funzioni che hanno tratti costanti per cui la equazione $y = g(x)$, per certi valori di $y = y_0$ ha come soluzioni tutti i punti di un intervallo (anche non semplicemente connesso) $\mathcal{X}(y)$. La figura seguente mostra un tale esempio in cui per $y = y_0$, l'equazione $y = g(x)$ ammette le soluzioni x_1, x_2 e tutti i punti dell'intervallo $\mathcal{X}(y_0)$. La pdf di Y in quel punto conterrà una delta di Dirac con area pari alla $Pr\{X \in \mathcal{X}(y_0)\}$.

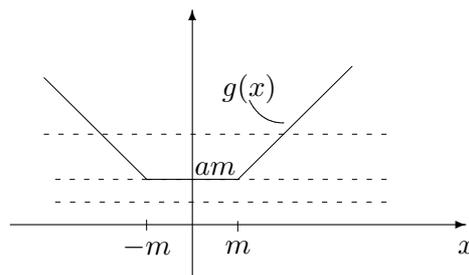


Discutiamo la estensione del teorema 5.1 con riferimento a un esempio.

Esempio 5.13 Si consideri la trasformazione

$$g(x) = \begin{cases} aX & X > m \\ am & -m < X < m \\ -aX & X < -m. \end{cases} \quad (5.44)$$

La funzione è riportata nella seguente figura.



Si tratta della funzione raddrizzatrice a doppia semionda, ma con un tratto costante per valori di x al di sotto del minimo m ($a, m > 0$). Per $y > am$, la equazione $y = g(x)$ ammette le due soluzioni $x_1 = \frac{y}{a}$ e $x_2 = -\frac{y}{a}$. Poichè in tale intervallo $g'(x) = a \operatorname{sgn}(x)$, la formula del teorema 5.1 fornisce

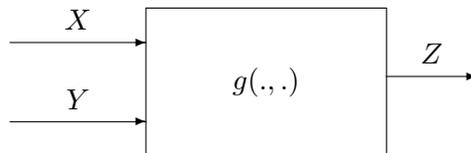
$$f_Y(y) = \frac{1}{|a|} \left[f_X\left(\frac{y}{a}\right) + f_X\left(-\frac{y}{a}\right) \right], \quad y > am. \quad (5.45)$$

Per $y = am$, le soluzioni dell'equazione $y = g(x)$ sono tutti i punti dell'intervallo $\mathcal{X}(y_0) = \{-m < x < m\}$. Pertanto in $y = am$ ci sarà una delta di Dirac di area pari a $\Pr\{-m < X < m\}$. Per $y < am$, l'equazione $y = g(x)$ non ammette alcuna soluzione, per cui $f_Y(y) = 0$. Sinteticamente, la pdf di Y si scrive

$$f_Y(y) = \left(\int_{-m}^m f_X(\xi) d\xi \right) \delta(y-am) + \frac{1}{|a|} \left[f_X\left(\frac{y}{a}\right) + f_X\left(-\frac{y}{a}\right) \right] u(y-am). \quad (5.46)$$

5.3 Funzioni di due variabili

Consideriamo ora il caso di due variabili aleatorie X e Y sulle quali è applicata una funzione deterministica che restituisce una nuova variabile aleatoria Z . Lo schema di riferimento è mostrato in figura.



Più formalmente, per ogni realizzazione (x, y) di (X, Y) ,

$$(x, y) \in \mathcal{A} \longleftrightarrow z = g(x, y) \in \mathcal{B}, \quad (5.47)$$

dove \mathcal{A} è l'insieme di definizione della variabile bidimensionale (x, y) e \mathcal{B} quello di z . Il problema è la caratterizzazione di Z , note la caratterizzazione congiunta di X e Y e la funzione g . Anche qui, come nel

caso della funzione di una variabile, è necessario innanzitutto esaminare la struttura dei domini \mathcal{A} e \mathcal{B} . Infatti X e Y possono essere discrete, continue o miste e la funzione continua o discontinua. Anche la invertibilità della funzione rispetto ad ognuna delle due variabili di ingresso è da tenere in conto. L'obiettivo è la determinazione della cdf $F_Z(z)$ o della pdf $f_Z(z)$, note la funzione g e la densità congiunta $f_{XY}(x, y)$. Cominciamo con qualche semplice esempio.

Esempio 5.14 Si considerino due variabili X e Y binarie, ovvero $X \in \{0, 1\}$ e $Y \in \{0, 1\}$. La funzione g sia la funzione XOR tra X e Y

$$Z = X \oplus Y, \quad (5.48)$$

e X e Y siano indipendenti e aventi distribuzioni $\Pi_X = \{0.1, 0.9\}$ e $\Pi_Y = \{0.3, 0.7\}$. Il dominio di definizione di Z è ancora l'insieme binario, $Z \in \{0, 1\}$, e serve calcolare la distribuzione Π_Z . La soluzione è facilmente ottenuta per enumerazione

$$\begin{aligned} Pr\{Z = 0\} &= Pr\{X = 0, Y = 0\} + Pr\{X = 1, Y = 1\} \\ &= Pr\{X = 0\}Pr\{Y = 0\} + Pr\{X = 1\}Pr\{Y = 1\} \\ &= 0.1 \cdot 0.3 + 0.9 \cdot 0.7 = 0.66 \\ Pr\{Z = 1\} &= 1 - Pr\{Z = 0\} = 0.34 \end{aligned} \quad (5.49)$$

Esempio 5.15 Si consideri il prodotto aritmetico $Z = X Y$ di due variabili indipendenti X e Y , dove X è una variabile a due valori $X \in \{0, 1\}$ con distribuzione $\Pi_X = \{0.5, 0.5\}$, e Y è una variabile di tipo continuo avente densità $f_Y(y)$. La densità di Z è calcolabile osservando che o $Z = Y$, per $X = 1$, o è zero per $X = 0$. Quindi

$$f_Z(z) = f_Y(z)Pr\{X = 1\} + \delta(y)Pr\{X = 0\} = \frac{1}{2} (f_Y(z) + \delta(y)). \quad (5.50)$$

Gli esempi esposti sono piuttosto semplici e il loro svolgimento intuitivo. E' necessario derivare delle metodologie più sistematiche per la caratterizzazione di Z .

5.3.1 Metodo della cdf

Ricordiamo che la cdf di Z è per definizione

$$F_Z(z) = Pr\{Z \leq z\} = Pr\{g(X, Y) \leq z\}. \quad (5.51)$$

Analogamente al caso della funzione di una variabile, definendo l'insieme $I_g(z)$ come l'insieme nel dominio di (X, Y) in cui è verificata la disuguaglianza, l'evento che definisce la cdf viene riscritto come evento su (X, Y)

$$\{g(X, Y) \leq z\} = \{(X, Y) \in I_g(z)\}. \quad (5.52)$$

Pertanto

$$F_Z(z) = Pr\{(X, Y) \in I_g(z)\} = \int_{I_g} f_{XY}(x, y) dx dy. \quad (5.53)$$

La natura della funzione g ovviamente determina la struttura dell'insieme $I_g(z)$ per ogni valore della variabile z . Tale insieme può essere continuo, discreto e non semplicemente connesso. Vedremo la applicazione del metodo direttamente con riferimento a degli esempi.

La pdf di Z è ottenibile per derivazione dalla cdf. Quando le variabili sono di tipo continuo è possibile esprimere la pdf direttamente in funzione della pdf congiunta di (X, Y) . Ricordando che la pdf è definita come

$$f_Z(z) dz = Pr\{z < Z \leq z + dz\}, \quad (5.54)$$

indicando l'insieme $\Delta I_g(z)$ come l'insieme nel piano (x, y) tale che

$$\{z < Z \leq z + dz\} = \{(X, Y) \in \Delta I_g(z)\}, \quad (5.55)$$

abbiamo che

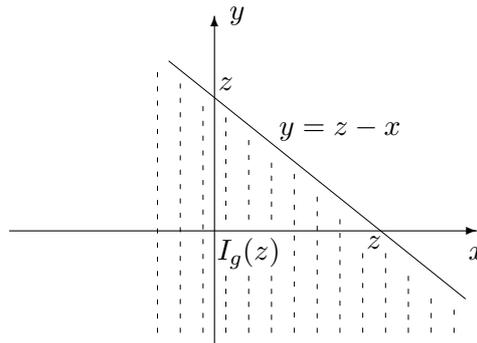
$$f_z(z) dz = \int_{\Delta I_g(z)} f_{XY}(x, y) dx dy. \quad (5.56)$$

Esempio 5.16 (Somma di due variabili) L'esempio più tipico della funzione di due variabile è la somma $Z = X + Y$, dove X e Y sono due variabili aleatorie continue. Si pensi alla applicazione del canale

rumoroso, dove una delle componenti è il segnale e l'altra il rumore. La cdf di Z si valuta come

$$F_Z(z) = Pr\{Z \leq z\} = Pr\{X + Y \leq z\} = Pr\{(X, Y) \in I_g(z)\}, \quad (5.57)$$

dove l'insieme $I_g(z)$ è il semipiano del piano (x, y) delimitato dalla retta $y = z - x$ mostrato in figura



La cdf di Z è pertanto

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{XY}(x, y) dx dy. \quad (5.58)$$

Derivando ¹ si ottiene la pdf

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(z - y, y) dy. \quad (5.60)$$

Quando le variabili X e Y sono indipendenti, ovvero $f_{XY}(x, y) = f_X(x)f_Y(y)$, la pdf di Z si scrive come

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - y)f_Y(y) dy. \quad (5.61)$$

¹Ricordiamo che se $f(x)$ è una funzione avente primitiva $F(x)$, si ha che

$$\frac{\partial}{\partial b} \int_a^b f(\xi) d\xi = \frac{\partial}{\partial b} (F(b) - F(a)) = \frac{\partial}{\partial b} F(b) = f(b). \quad (5.59)$$

Nel caso dell'esempio la procedura si applica solo alla prima variabile di $f_{XY}(x, y)$.

Riconosciamo che l'integrale è la *convoluzione* tra le due densità. Pertanto si ha il risultato notevole: *la pdf della somma di due variabili aleatorie indipendenti è la convoluzione delle due densità.*

I metodi noti dalla teoria dei segnali per il calcolo della convoluzione possono essere applicati in questo caso al calcolo della pdf di Z . Per esempio se le due pdf sono uniformi, ovvero $f_X(x) = \frac{1}{2a}\Pi\left(\frac{x}{2a}\right)$ e $f_Y(y) = \frac{1}{2a}\Pi\left(\frac{y}{2a}\right)$, si ha

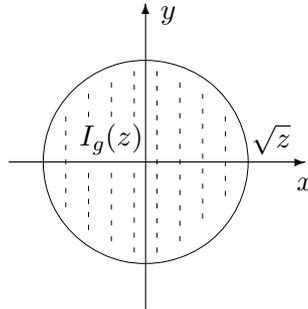
$$f_Z(z) = \frac{1}{2a}\Lambda\left(\frac{z}{2a}\right). \quad (5.62)$$

Si ricorda che le funzioni Π e Λ sono le funzioni rettangolare e triangolare rispettivamente.

Esempio 5.17 Valutiamo ora la densità di $Z = X^2 + Y^2$. La cdf di Z si scrive come

$$F_Z(z) = Pr\{Z \leq z\} = Pr\{X^2 + Y^2 \leq z\} = \int_{I_g(z)} f_{XY}(x, y) dx dy, \quad (5.63)$$

dove $I_g(z) = \{(x, y) : x^2 + y^2 \leq z\}$. Se $z < 0$, l'insieme è vuoto e quindi $F_Z(z) = 0$. Se $z > 0$, $I_g(z)$ è l'insieme dei punti del piano (x, y) all'interno del cerchio di raggio \sqrt{z} come mostrato in figura



E' interessante andare a calcolare la cdf e la pdf di Z quando X e Y sono indipendenti, gaussiane a media nulla e aventi la stessa varianza σ^2 , ovvero

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \quad (5.64)$$

Calcolando l'integrale in coordinate polari mediante la solita sostituzione $x = r \cos \theta$, $y = r \sin \theta$, si ha

$$\begin{aligned} F_Z(z) &= \frac{1}{2\pi\sigma^2} \int_{-\pi}^{\pi} \int_0^{\sqrt{z}} r e^{-\frac{r^2}{2\sigma^2}} dr d\theta \\ &= \frac{1}{\sigma^2} \int_0^{\sqrt{z}} r e^{-\frac{r^2}{2\sigma^2}} dr = -e^{-\frac{r^2}{2\sigma^2}} \Big|_0^{\sqrt{z}} \\ &= 1 - e^{-\frac{z}{2\sigma^2}}, \quad z \geq 0. \end{aligned} \quad (5.65)$$

La pdf è quindi la densità esponenziale

$$f_Z(z) = \frac{1}{2\sigma^2} e^{-\frac{z}{2\sigma^2}} u(z). \quad (5.66)$$

Esempio 5.18 Consideriamo la funzione $Z = \sqrt{X^2 + Y^2}$. L'esempio è interessante poiché Z è il modulo di un vettore le cui coordinate cartesiane sono X e Y . La cdf di Z si scrive come

$$F_Z(z) = Pr\{Z \leq z\} = Pr\{\sqrt{X^2 + Y^2} \leq z\} = Pr\{(X, Y) \in I_g(z)\}, \quad (5.67)$$

dove $I_g(z) = \{(x, y) : \sqrt{x^2 + y^2} \leq z\}$ è il cerchio di raggio z . Se $z < 0$ l'insieme è vuoto, pertanto $F_Z(z) = 0$. Altrimenti per $z > 0$,

$$f_Z(z) = \int_{I_g(z)} f_{XY}(x, y) dx dy. \quad (5.68)$$

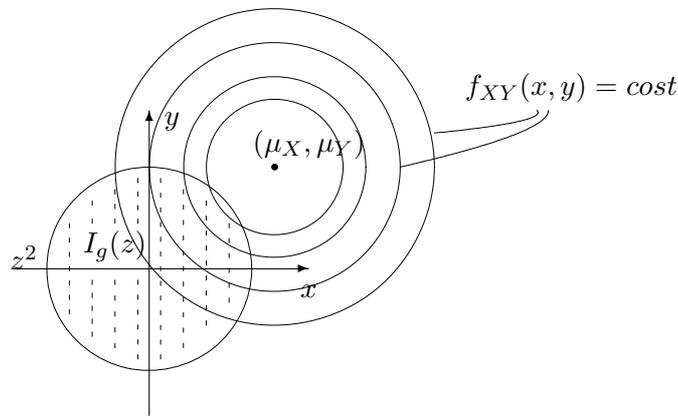
Se le variabili X e Y sono indipendenti e gaussiane a media nulla con la stessa varianza σ^2 , analogamente all'esempio precedente eseguendo l'integrale in coordinate polari abbiamo per $z > 0$

$$\begin{aligned} F_Z(z) &= \frac{1}{2\pi\sigma^2} \int_{I_g(z)} e^{-\frac{x^2+y^2}{2\sigma^2}} dx dy \\ &= \frac{1}{2\pi\sigma^2} \int_{-\pi}^{\pi} \int_0^z r e^{-\frac{r^2}{2\sigma^2}} dr d\theta \\ &= \frac{1}{\sigma^2} \int_0^z r e^{-\frac{r^2}{2\sigma^2}} dr = -e^{-\frac{r^2}{2\sigma^2}} \Big|_0^z \\ &= 1 - e^{-\frac{z}{2\sigma^2}}, \quad z \geq 0. \end{aligned} \quad (5.69)$$

La pdf ottenuta per derivazione è la densità di Rayleigh

$$f_Z(z) = \frac{z}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}} u(z). \quad (5.70)$$

E' interessante generalizzare il risultato quando le due variabili aleatorie X e Y sono ancora delle gaussiane indipendenti, hanno la stessa varianza, ma non la stessa media. Se μ_X e μ_Y sono le due medie, bisogna integrare sul cerchio di raggio z una densità non centrata nell'origine come mostrato nelle curve di livello di figura



La cdf si calcola per $z > 0$ come segue

$$\begin{aligned} F_Z(z) &= \frac{1}{2\pi\sigma^2} \int_{I_g(z)} e^{-\frac{(x-\mu_X)^2+(y-\mu_Y)^2}{2\sigma^2}} dx dy \\ &= \frac{1}{2\pi\sigma^2} \int_{-\pi}^{\pi} \int_0^z e^{-\frac{(r \cos \theta - \mu_X)^2 + (r \sin \theta - \mu_Y)^2}{2\sigma^2}} r dr d\theta \\ &= \int_0^z \frac{r}{2\pi\sigma^2} e^{-\frac{r^2 + \mu_X^2 + \mu_Y^2}{2\sigma^2}} \int_{-\pi}^{\pi} e^{\frac{r}{\sigma^2}(\mu_X \cos \theta + \mu_Y \sin \theta)} d\theta dr \quad (5.71) \end{aligned}$$

Un risultato più esplicito è ottenibile parametrizzando le medie in coordinate polari definendo A e ϕ tali che $\mu_X = A \cos \phi$ e $\mu_Y = A \sin \phi$. Pertanto

$$F_Z(z) = \int_0^z \frac{r}{2\pi\sigma^2} e^{-\frac{r^2 + \mu_X^2 + \mu_Y^2}{2\sigma^2}} \int_{-\pi}^{\pi} e^{\frac{r}{\sigma^2}(\mu_X \cos \theta + \mu_Y \sin \theta)} d\theta dr$$

$$\begin{aligned}
&= \int_0^z \frac{r}{2\pi\sigma^2} e^{-\frac{r^2+A^2}{2\sigma^2}} \int_{-\pi}^{\pi} e^{\frac{Ar}{\sigma^2}(\cos\phi\cos\theta+\sin\phi\sin\theta)} d\theta dr \\
&= \int_0^z \frac{r}{2\pi\sigma^2} e^{-\frac{r^2+A^2}{2\sigma^2}} \int_{-\pi}^{\pi} e^{\frac{Ar}{\sigma^2}\cos(\theta-\phi)} d\theta dr \\
&= \int_0^z \frac{r}{2\pi\sigma^2} e^{-\frac{r^2+A^2}{2\sigma^2}} \int_0^{2\pi} e^{\frac{Ar}{\sigma^2}\cos\theta} d\theta dr,
\end{aligned} \tag{5.72}$$

dove l'ultimo passaggio dipende dal fatto che la funzione integranda è periodica di periodo 2π ed è integrata su tutto il periodo. Questo rende l'integrale indipendente da ϕ . Ricordando ora la definizione di funzione di Bessel modificata di ordine zero

$$I_0(\xi) = \frac{1}{2\pi} \int_0^{2\pi} e^{\xi \cos\theta} d\theta, \tag{5.73}$$

abbiamo che la cdf di Z è

$$F_Z(z) = \int_0^z \frac{r}{\sigma^2} e^{-\frac{r^2+A^2}{2\sigma^2}} I_0\left(\frac{Ar}{\sigma^2}\right) dr. \tag{5.74}$$

La pdf, ottenuta per semplice derivazione, è la pdf di Rice

$$f_Z(z) = \frac{z}{\sigma^2} e^{-\frac{z^2+A^2}{2\sigma^2}} I_0\left(\frac{Az}{\sigma^2}\right) u(z). \tag{5.75}$$

Si noti come la pdf di Rayleigh sia un caso particolare della Rice.

5.3.2 Il metodo della pdf

Un metodo alternativo per il calcolo di $f_Z(z)$ è basato sulle trasformazioni di pdf fissando le variabili X e Y in maniera progressiva. L'idea è la seguente.

Supponiamo che la variabile X sia fissata al valore $X = x_0$. La aleatorietà di Z è quindi dipendente solo dalla aleatorietà di Y . Pertanto $Z = g(x_0, Y)$ può essere derivata utilizzando il metodo della pdf rispetto a Y . Più in particolare, se y_1, y_1, \dots, y_m sono le soluzioni dell'equazione $z = g(x_0, y)$, abbiamo che

$$f_Z(z|X = x_0) = \frac{f_{Y|X}(y|X = x_0)}{\left| \frac{\partial g(x_0, y)}{\partial y} \right|_{y=y_1}} + \dots + \frac{f_{Y|X}(y|X = x_0)}{\left| \frac{\partial g(x_0, y)}{\partial y} \right|_{y=y_m}}. \tag{5.76}$$

Ora devo supporre che il calcolo precedente mi fornisca un risultato per tutti i valori di $x_0 \in \mathcal{X}$, ovvero di tutta la $f_{Z|X}(z|x)$. Pertanto poiché

$$f_{ZX}(z, x) = f_{Z|X}(z|x)f_X(x), \quad (5.77)$$

il risultato si ottiene integrando

$$f_Z(z) = \int_{\mathcal{X}} f_{Z|X}(z|x)f_X(x) dx. \quad (5.78)$$

La tecnica esposta non è detto che fornisca una risposta semplice al problema. La complessità del calcolo dipenderà dal tipo di funzione. Per spiegare meglio la tecnica ripetiamo il caso del calcolo della pdf della somma di due variabili aleatorie.

Esempio 5.19 Sia $Z = X + Y$. Fissato $X = x_0$ abbiamo $Z = x_0 + Y$. L'equazione $z = x_0 + y$ ammette l'unica soluzione $y = z - x_0$. Quindi

$$f_Z(z|X = x_0) = \frac{f_{Y|X}(y|X = x_0)}{\frac{\partial g(x_0, y)}{\partial y}} \Bigg|_{y=z-x_0} = f_{Y|X}(z - x_0|X = x_0). \quad (5.79)$$

Pertanto $f_Z(z|x) = f_{Y|X}(z - x|x)$ e

$$f_Z(z) = \int_{\mathcal{X}} f_{Y|X}(z - x|x)f_X(x) dx, \quad (5.80)$$

che è il risultato cercato. Ovviamente quando X e Y sono indipendenti, $f_{Y|X}(y|x) = f_Y(y)$ e si ottiene la convoluzione

$$f_Z(z) = \int_{\mathcal{X}} f_Y(z - x)f_X(x) dx = (f_Y * f_X)(z). \quad (5.81)$$

5.3.3 Variabili discrete

Le insidie nel calcolo della cdf o della pdf di Z sono legate chiaramente alla natura della funzione g e della densità di X e Y . Nel caso di X e Y variabili discrete la pmf di Z è di calcolo immediato per semplice enumerazione. Ovvero se i valori discreti assunti dalle variabili X e Y

sono $\{(x_n, y_m)\}_{nm}$, i corrispondenti valori di Z sono $z_{nm} = g(x_n, y_m)$, che costituiscono lo spazio campione discreto di Z . La massa di probabilità che si concentra su questi punti è immediatamente ottenuta dalle probabilità di occorrenza delle coppie (x_n, y_m) . Si noti che a più coppie può corrispondere lo stesso valore di Z , pertanto le probabilità relative vanno sommate opportunamente.

5.4 Funzioni di n variabili

Le tecniche mostrate per la trasformazione da due variabili a una variabile aleatoria possono facilmente essere generalizzate al caso multidimensionale. Date n variabili, X_1, X_2, \dots, X_n , si vuole studiare la distribuzione di

$$Z = g(X_1, \dots, X_n), \quad (5.82)$$

dove g è una funzione deterministica n -dimensionale. Più formalmente, per ogni realizzazione (x_1, \dots, x_n) di (X_1, \dots, X_n) ,

$$(x_1, \dots, x_n) \in \mathcal{A} \longleftrightarrow z = g(x_1, \dots, x_n) \in \mathcal{B}, \quad (5.83)$$

dove \mathcal{A} è l'insieme di definizione della variabile n -dimensionale (x_1, \dots, x_n) e \mathcal{B} quello di z . Il problema è la caratterizzazione di Z , note la caratterizzazione congiunta di X_1, \dots, X_n e la funzione g . Anche qui, come nei casi esaminati in precedenza, è necessario innanzitutto esaminare la struttura dei domini \mathcal{A} e \mathcal{B} . Infatti X_1, \dots, X_n possono essere discrete, continue o miste e la funzione continua o discontinua. Anche la invertibilità della funzione rispetto ad ognuna delle n variabili di ingresso è da tenere in conto. L'obiettivo è la determinazione della cdf $F_Z(z)$ o della pdf $f_Z(z)$, note la funzione g e la densità congiunta $f_{X_1 \dots X_n}(x_1, \dots, x_n)$.

5.4.1 Metodo della cdf

Ricordiamo che la cdf di Z è per definizione

$$F_Z(z) = Pr\{Z \leq z\} = Pr\{g(X_1, \dots, X_n) \leq z\}. \quad (5.84)$$

Analogamente al caso della funzione di due variabile, definendo l'insieme $I_g(z)$ come l'insieme nel dominio di (X_1, \dots, X_n) in cui è verificata la

disuguaglianza, l'evento che definisce la cdf viene riscritto come evento su (X_1, \dots, X_n)

$$\{g(X_1, \dots, X_n) \leq z\} = \{(X_1, \dots, X_n) \in I_g(z)\}. \quad (5.85)$$

Pertanto

$$F_Z(z) = Pr\{(X_1, \dots, X_n) \in I_g(z)\} = \int_{I_g} f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n. \quad (5.86)$$

La pdf di Z è ottenibile per derivazione dalla cdf. Quando le variabili sono di tipo continuo è possibile esprimere la pdf direttamente in funzione della pdf congiunta di (X_1, \dots, X_n) . Ricordando che la pdf è definita come

$$f_Z(z) dz = Pr\{z < Z \leq z + dz\}, \quad (5.87)$$

indicando l'insieme $\Delta I_g(z)$ come l'insieme nello spazio (x_1, \dots, x_n) tale che

$$\{z < Z \leq z + dz\} = \{(X_1, \dots, X_n) \in \Delta I_g(z)\}, \quad (5.88)$$

abbiamo che

$$f_Z(z) dz = \int_{\Delta I_g(z)} f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n. \quad (5.89)$$

5.4.2 Metodo della pdf

La tecnica della pdf presentata per due variabili aleatorie è facilmente generalizzata al caso di n variabili. Supponiamo che tutte le variabili, eccetto la prima, siano state fissate a $X_2 = x_{20}, \dots, X_n = x_{n0}$. Quindi la variabilità di Z dipende solo dalla variabilità di X_1 . Siano $x_1^1, \dots, x_1^{m_1}$ le soluzioni dell'equazione

$$z = g(x_1, x_{20}, \dots, x_{n0}). \quad (5.90)$$

Pertanto

$$\begin{aligned} & f_{Z|X_2 \dots X_n}(z|X_2 = x_{20}, \dots, X_n = x_{n0}) \\ &= \frac{f_{X_1|X_2 \dots X_n}(x_1|X_2 = x_{20}, \dots, X_n = x_{n0})}{\frac{\partial g(x_1, x_{20}, \dots, x_{n0})}{\partial x_1}} \Bigg|_{x_1 = x_1^1} \\ &+ \dots \\ &+ \frac{f_{X_1|X_2 \dots X_n}(x_1|X_2 = x_{20}, \dots, X_n = x_{n0})}{\frac{\partial g(x_1, x_{20}, \dots, x_{n0})}{\partial x_1}} \Bigg|_{x_1 = x_1^{m_1}}. \end{aligned} \quad (5.91)$$

Ora devo supporre che il calcolo precedente mi fornisca tutta la $f_{Z|X_2 \dots X_n}(z|x_2, \dots, x_n)$. Pertanto poiché

$$f_{Z X_2 \dots X_n}(z, x_2, \dots, x_n) = f_{Z|X_2 \dots X_n}(z|x_2, \dots, x_n) f_{X_2 \dots X_n}(x_2, \dots, x_n), \quad (5.92)$$

il risultato si ottiene integrando

$$f_Z(z) = \int_{\mathcal{X}_2 \times \dots \times \mathcal{X}_n} f_{Z|X_2 \dots X_n}(z|x_2, \dots, x_n) f_{X_2 \dots X_n}(x_2, \dots, x_n) dx_2 \cdots dx_n. \quad (5.93)$$

Come sottolineato in precedenza, non è detto che la procedura porti ad un facile risultato. Nei casi specifici il lettore dovrà adottare la tecnica che gli risulterà più agevole o più congeniale.

Esempio 5.20 Esaminiamo ora come primo esempio il caso tipico della sovrapposizione di n variabili aleatorie. Sia $Z = X_1 + \dots + X_n$. Fissati $X_2 = x_2, \dots, X_n = x_n$ abbiamo $Z = x_1 + x_2 + \dots + x_n$. L'equazione $z = x_1 + x_2 + \dots + x_n$ ammette l'unica soluzione $x_1 = z - x_2 - \dots - x_n$. Quindi

$$\begin{aligned} & f_{Z|X_2 \dots X_n}(z|X_2 = x_2, \dots, X_n = x_n) \\ &= \left. \frac{f_{X_1|X_2 \dots X_n}(x_1|X_2 = x_2, \dots, X_n = x_n)}{\frac{\partial g(x_1, x_2, \dots, x_n)}{\partial x_1}} \right|_{x_1 = z - x_2 - \dots - x_n} \\ &= f_{X_1|X_2 \dots X_n}(z - x_2 - \dots - x_n|X_2 = x_2, \dots, X_n = x_n). \end{aligned} \quad (5.94)$$

$$f_Z(z) = \int_{\mathcal{X}_2 \times \dots \times \mathcal{X}_n} f_{X_1|X_2 \dots X_n}(z - x_2 - \dots - x_n|x_2, \dots, x_n) f_{X_2 \dots X_n}(x_2, \dots, x_n) dx_2 \cdots dx_n. \quad (5.95)$$

che è il risultato cercato. Per l'importante caso con X_1, \dots, X_n indipendenti abbiamo

$$f_Z(z) = \int_{\mathcal{X}_2 \times \dots \times \mathcal{X}_n} f_{X_1}(z - x_2 - \dots - x_n) f_{X_2}(x_2) \cdots f_{X_n}(x_n) dx_2 \cdots dx_n. \quad (5.96)$$

L'espressione ottenuta è suscettibile di una ulteriore manipolazione. Operiamo il seguente cambio di variabili

$$\begin{aligned} y_n &= x_n \\ y_{n-1} &= x_{n-1} + x_n \\ &\dots \\ y_3 &= x_3 + \dots + x_n \\ y_2 &= x_2 + \dots + x_n. \end{aligned} \quad (5.97)$$

Ovvero

$$\begin{aligned}
 x_n &= y_n \\
 x_{n-1} &= y_{n-1} - y_n \\
 &\dots \\
 x_3 &= y_3 - y_4 \\
 x_2 &= y_2 - y_3.
 \end{aligned} \tag{5.98}$$

lo Jacobiano della trasformazione è

$$J \begin{pmatrix} x_2 & x_3 & \dots & x_n \\ y_2 & y_3 & \dots & y_n \end{pmatrix} = \begin{vmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{vmatrix} = 1. \tag{5.99}$$

Quindi l'integrale si può riscrivere come

$$\begin{aligned}
 f_Z(z) &= \int_{-\infty}^{\infty} f_{X_1}(z - y_2) \int_{-\infty}^{\infty} f_{X_2}(y_2 - y_3) \dots \\
 &\dots \int_{-\infty}^{\infty} f_{X_{n-1}}(y_{n-1} - y_n) f_{X_n}(y_n) dy_n \dots dy_3 dy_2.
 \end{aligned} \tag{5.100}$$

Il risultato è la convoluzione delle n pdf

$$f_Z(z) = (f_{X_1} * f_{X_2} * \dots * f_{X_N})(z). \tag{5.101}$$

5.4.3 Il teorema centrale del limite

Nello studio dei sistemi fisici con il metodo probabilistico è molto comune dover studiare l'effetto sovrapposto di più variabili. Per esempio, potrebbe essere importante studiare la distribuzione della variabile

$$Z = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_i), \tag{5.102}$$

dove le X_i sono delle variabili aleatorie con medie μ_i . Il fattore $1/n$ è stato introdotto per evitare che la somma diverga. Abbiamo visto sopra come sia possibile, dalla conoscenza della densità congiunta di X_1, \dots, X_n , valutare il risultato esatto (a meno di eventuali difficoltà di calcolo). La cosa interessante è che, sotto lasche ipotesi di indipendenza tra le X_i , e indipendentemente da quali siano le distribuzioni delle X_i , la densità di

Z tende sempre ad una gaussiana! Il risultato, noto come *teorema centrale del limite*, ha un'importanza fondamentale e giustifica l'assunzione di gaussianità per variabili aleatorie connesse a tanti fenomeni fisici. Istogrammi misurati in situazioni pratiche assomigliano molto spesso a delle curve a campana (si veda anche l'esempio riportato sulla copertina). Pertanto quando si ha a che fare con la somma di un congruo numero di variabili, si può affidabilmente supporre che il risultato di distribuisca, almeno approssimativamente, secondo una pdf gaussiana.

Vista la natura introduttiva di queste note, tralasciamo la prova, rimandando il lettore ad uno dei tanti testi di riferimento esistenti nella letteratura. Si noti che, quando le variabili sono indipendenti, il risultato della somma è la convoluzione delle densità. Il teorema centrale del limite quindi predice che la convoluzione di tante funzioni produce un risultato che è una funzione che ha una forma che tende ad una funzione gaussiana

5.5 Problemi

Problema 5.1 Si valuti la densità di $Z = |X|$, dove X è una variabile aleatoria uniforme in $[-\Delta, \Delta]$.

Problema 5.2 Si valuti la densità di $Z = |X|$, dove X è una variabile aleatoria gaussiana con media nulla e varianza σ_X^2 .

Problema 5.3 Si valuti la densità di $Z = aX + bY$, dove X e Y sono variabili aleatorie indipendenti aventi densità $f_X(x)$ e $f_Y(y)$ e a, b sono due costanti reali. Suggerimento: valutare prima le densità di aX e bY .

Problema 5.4 Valutare la densità della somma $Z = X + Y$ di due variabili aleatorie esponenziali indipendenti aventi varianze σ_1^2 e σ_2^2 rispettivamente.

Problema 5.5 Valutare la densità della somma $Z = X_1 + X_2 + X_3$ di tre variabili aleatorie uniformi indipendenti in $[-\Delta, \Delta]$.

Problema 5.6 Si studi la somma di n variabili aleatorie discrete indipendenti aventi tutto spazio campione $\{-a, a\}$ e probabilità uniforme $\{1/2, 1/2\}$.

Problema 5.7 Dimostrare che la somma di n variabili aleatorie gaussiane indipendenti è ancora una variabile aleatoria gaussiana.

Appendice A

La CDF Gaussiana

La Funzione di Distribuzione Cumulativa (CDF) Gaussiana $\Phi(x)$ è stata definita nel Capitolo 2. I valori di $\Phi(x)$, o della funzione quantile $Q(x) = 1 - \Phi(x)$, sono disponibili nella maggior parte delle moderne calcolatrici e in software di calcolo numerico quali Matlab. Per agevolare lo studente negli esercizi, riportiamo nella pagina seguente una tabella che contiene i valori della $\Phi(x)$ per $x \in [0 \ 3.69]$. Si ricordi che $\Phi(-x) = 1 - \Phi(x)$. I valori di x sono quelli che corrispondono alle prime quattro cifre significative di $\Phi(x)$. Ovviamente maggiore precisione, anche per valori che eccedono $x = 3.69$, possono essere ottenuti usando una calcolatrice, o software quale Matlab, imponendo la precisione desiderata.

I valori in tabella sono stati ottenuti con Matlab usando la funzione `erfc(x)` nell'istruzione `phi=1-0.5* erfc(x/sqrt(2))`.

Appendice B

La funzione di Dirac

La funzione di *Dirac*, o funzione *delta*, o funzione *impulso ideale*, denotata con il simbolo $\delta(x)$, è una di quelle funzioni speciali che vengono dette *funzioni generalizzate*. Si tratta infatti di una funzione che è definita solo in un punto, nel quale assume valore infinito, ma area unitaria. Tale funzione, che può essere inserita nel contesto della teoria matematica delle distribuzioni (Lighthill, 1959), è un comodo artificio matematico che ci consente di manipolare e derivare funzioni discontinue e non differenziabili e a superare alcuni problemi formali quando bisogna considerare valori puntuali di funzioni continue.

La funzione *delta*, o *delta di Dirac*, o *impulso di Dirac* è tale che per ogni funzione $f(x)$ continua in $x = 0$:

$$\int_{x_1}^{x_2} f(x)\delta(x)dx = \begin{cases} f(0) & \text{se } x_1 \leq 0 \leq x_2 \\ 0 & \text{altrove} \end{cases} \quad (\text{B.1})$$

Dalla definizione discendono immediatamente le seguenti proprietà:

Proprietà B.1

$$\int_{-\infty}^{+\infty} \delta(x)dx = 1. \quad (\text{B.2})$$

Questa proprietà mostra come l'area sottesa dalla delta sia unitaria. La prova è immediata conseguenza della definizione per $f(x) = 1$.

Proprietà B.2 (Proprietà del Campionamento) Per ogni funzione $f(x)$ continua in x_0 ,

$$\int_{-\infty}^{+\infty} f(x)\delta(x-x_0)dx = f(x_0). \quad (\text{B.3})$$

La prova è immediata conseguenza della definizione semplicemente operando un cambio di variabili nell'integrale.

Proprietà B.3 (Proprietà del cambiamento di scala) Per ogni $f(x)$ continua in $x = 0$,

$$\int_{-\infty}^{+\infty} f(x)\delta(\alpha x)dx = \frac{1}{|\alpha|}f(0). \quad (\text{B.4})$$

La prova di questa proprietà, è immediata se si osserva che:

$$\int_{-\infty}^{+\infty} f(x)\delta(\alpha x)dx = \int_{-\infty}^{+\infty} f\left(\frac{x}{\alpha}\right)\delta(\xi)\frac{d\xi}{|\alpha|}. \quad (\text{B.5})$$

Applicando la proprietà del campionamento si ottiene il risultato.

Proprietà B.4

$$\delta(\alpha x) = \frac{1}{|\alpha|}\delta(x). \quad (\text{B.6})$$

Questa è una immediata conseguenza della proprietà del cambiamento di scala per $f(x) = 1$.

Proprietà B.5

$$\int_{-\infty}^x \alpha\delta(\xi-x_0)d\xi = \alpha u(x-x_0), \quad (\text{B.7})$$

dove la funzione discontinua $u(x)$ è la cosiddetta funzione gradino o scalino, definita come:

$$u(x) = \begin{cases} 1 & \text{per } x \geq 0 \\ 0 & \text{altrove} \end{cases} \quad (\text{B.8})$$

(si veda figura B.1).

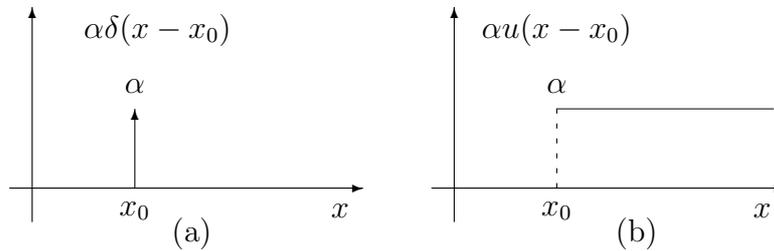


Figura B.1: (a) $\alpha\delta(x - x_0)$; (b) $\alpha u(x - x_0)$

La prova è diretta conseguenza della proprietà B.1.

Proprietà B.6

$$\frac{d}{dx}u(x - x_0) = \delta(x - x_0). \quad (\text{B.9})$$

La prova è immediata conseguenza della definizione di funzione a gradino, in quanto:

$$\int_{-\infty}^x \delta(\xi - x_0) d\xi = \begin{cases} 0 & x < x_0 \\ 1 & x \geq x_0 \end{cases} = u(x - x_0). \quad (\text{B.10})$$

Si noti come quest'ultima proprietà, consenta di esprimere analiticamente la derivata di funzioni discontinue e in particolare con discontinuità di prima specie. Infatti, come nell'esempio del gradino, per funzioni che presentano un salto di discontinuità, la derivata è una delta di Dirac centrata in quel punto e avente area pari alla altezza del salto.

Come già discusso, la delta ha area finita, ma valore infinito nel punto di applicazione, ovvero la funzione delta è un impulso:

$$\delta(x) = \begin{cases} +\infty & x = 0 \\ 0 & \text{altrove.} \end{cases} \quad (\text{B.11})$$

La figura B.1(a), mostra la funzione delta disegnata come un impulso con la freccia rivolta verso l'alto. La notazione serve a ricordare che il valore in quel punto è infinito, anche se la sua area è finita. Tipicamente, per comodità di lettura, disegneremo una freccia di lunghezza proporzionale alla sua area. Figura B.1(a) mostra la funzione $f(x) = \alpha\delta(x - x_0)$ e figura B.1(b) il suo integrale.

B.1 La funzione di Dirac come limite

La funzione di Dirac può anche essere ottenuta come limite di funzioni ad area unitaria. Ovvero come limite di funzioni che diventano infinitamente strette mantenendo area unitaria. L'esempio più semplice si ottiene usando la funzione rettangolare:

$$\Pi\left(\frac{x}{\tau}\right) = \begin{cases} 1 & -\frac{\tau}{2} \leq x < \frac{\tau}{2} \\ 0 & \text{altrove.} \end{cases} \quad (\text{B.12})$$

Normalizzando l'area a uno e andando al limite abbiamo:

$$\lim_{\tau \rightarrow 0} \frac{1}{\tau} \Pi\left(\frac{x}{\tau}\right) = \delta(x). \quad (\text{B.13})$$

Altro esempio notevole è l'impulso gaussiano:

$$g(x; \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}, \quad (\text{B.14})$$

che ha area unitaria e il parametro σ che ne controlla la larghezza. Al limite abbiamo:

$$\lim_{\sigma \rightarrow 0} g(x; \sigma) = \delta(x). \quad (\text{B.15})$$

L'impulso triangolare definito come:

$$\Lambda\left(\frac{x}{\tau}\right) = \begin{cases} \frac{x}{\tau} & -\tau < x < 0 \\ -\frac{x}{\tau} & 0 < x < \tau \\ 0 & \text{altrove} \end{cases}, \quad (\text{B.16})$$

normalizzato ad area unitaria ci da anch'esso al limite la funzione di Dirac:

$$\lim_{\tau \rightarrow 0} \frac{1}{\tau} \Lambda\left(\frac{x}{\tau}\right) = \delta(x). \quad (\text{B.17})$$

Anche l'impulso *sinc*, definito come:

$$\text{sinc}(\tau x) = \frac{\sin \pi \tau x}{\pi \tau x}, \quad (\text{B.18})$$

può essere usato per ottenere la delta di Dirac. Infatti l'area da esso sottesa è pari a $\frac{1}{\tau}$, la larghezza del lobo principale $\frac{2}{\tau}$ e la larghezza dei lobi laterali $\frac{1}{\tau}$. Pertanto normalizzando l'area e restringendo il lobo principale abbiamo che:

$$\lim_{\tau \rightarrow \infty} \tau \text{sinc}(\tau x) = \delta(x). \quad (\text{B.19})$$

Si noti che nel caso della funzione sinc, le funzioni intermedie prima del limite sono ad area unitaria, ma hanno anche valori negativi. Maggiori dettagli sulle funzioni gaussiana, triangolare, sinc possono essere reperite su qualunque testo di teoria dei segnali. Per le prove dei risultati asintotici si veda Lighthill (1959).

B.2 Derivate dell'impulso di Dirac

E' anche possibile definire le *derivate dell'impulso di Dirac* $\delta^{(n)}(x)$, $n = 1, 2, \dots$, tramite una proprietà analoga a quella usata per definire δ .

La funzione *derivata n -esima della delta di Dirac* $\delta^{(n)}(x)$, è tale che per ogni funzione $f(x)$ derivabile in $x = 0$ fino all'ordine n -esimo:

$$\int_{-\infty}^{\infty} f(x)\delta^{(n)}(x)dx = (-1)^n \left. \frac{d^n f(x)}{dx^n} \right|_{x=0}. \quad (\text{B.20})$$

Quindi formalmente la derivata della delta di Dirac applicata in un punto ci consente di ottenere tramite una operazione di prodotto e integrazione il valore della derivata n -esima della funzione in quel punto. La ragione del segno $(-1)^n$ ha a che fare con la utilizzabilità della $\delta^{(n)}(x)$ nel contesto della trasformata di Laplace. Per maggiori dettagli sulla delta di Dirac si veda Lighthill (1959).

Appendice C

Lemma inversione matrici

Nel calcolo matriciale il seguente risultato è noto come *il lemmadell'inversione delle matrici* (MIL - Matrix Inversion Lemma), o come formula di Sherman-Morrison-Woodbury.

Teorema C.1 Siano A e B due matrici $m \times m$ invertibili legate dalla relazione:

$$A = B^{-1} + CD^{-1}C^T, \quad (\text{C.1})$$

dove C è una matrice $m \times n$ e D è una matrice $n \times n$ invertibile. L'inversa di A può essere scritta nel seguente modo:

$$A^{-1} = B - BC(D + C^T BC)^{-1}C^T B. \quad (\text{C.2})$$

Prova: L'espressione può essere verificata direttamente accertando che $AA^{-1} = I$, ovvero

$$\begin{aligned} AA^{-1} &= (B^{-1} + CD^{-1}C^T) (B - BC(D + C^T BC)^{-1}C^T B) \\ &= B^{-1}B + CD^{-1}C^T B \\ &\quad - B^{-1}BC(D + C^T BC)^{-1}C^T B - CD^{-1}C^T BC(D + C^T BC)^{-1}C^T B \\ &= I + CD^{-1}C^T B \\ &\quad - C(D + C^T BC)^{-1}C^T B - CD^{-1}C^T BC(D + C^T BC)^{-1}C^T B \\ &= I + CD^{-1} [I - D(D + C^T BC)^{-1} - C^T BC(D + C^T BC)^{-1}] C^T B \\ &= I + CD^{-1} [I - (D + C^T BC)(D + C^T BC)^{-1}] C^T B \\ &= I. \end{aligned} \quad (\text{C.3})$$

Altre forme simili del lemma sono disponibili nella letteratura sul calcolo matriciale. Per una discussione sull'argomento si rimanda il lettore all'articolo di Tylavsky e Sohie (1986).

Bibliografia

- W. Feller, 1968.** *An Introduction to Probability Theory and Its Applications*, John Wiley & Sons.
- B. V. Gnedenko, 1992** *Teoria della Probabilità*, Traduzione dal russo: *Kurs teorii verojatnostej*, Edizioni Mir, Mosca e Editori Riuniti.
- C. W. Helstrom, 1984** *Probability and Stochastic Processes for Engineers*, Macmillan.
- M.J. Lighthill, 1959** *An Introduction to Fourier Analysis and Generalized Functions*, Cambridge University Press.
- H. Hsu, 1996** *Schaum's Outlines on Probability, Random Variables, & Random Processes*, McGraw-Hill.
- A. Papoulis, 1962.** *The Fourier Integral and Its Applications*, McGraw-Hill.
- A. Papoulis, 1965.** *Probability, Random Variables, and Sochastic Processes*, McGraw-Hill.
- A. Papoulis, 1973.** *Probabilità, Variabili Aleatorie e Processi Stocastici*, traduzione italiana dal (Papoulis, 1965), Boringhieri.
- A. Papoulis, 1984.** *Probability, Random Variables, and Sochastic Processes*, McGraw-Hill, IIa edizione.
- A. Papoulis, 1991.** *Probability, Random Variables, and Sochastic Processes*, McGraw-Hill, IIIa edizione.
- E. Parzen, 1960.** *Modern Probability Theory and Its Applications*, John Wiley & Sons.

- P. Z. Peebles, Jr., 1980.** *Probability, Random Variables, and Signal Principles*, McGraw-Hill.
- D. J. Tylavsky e G. R. Sohie, 1986.** "Generalizations of the Matrix Inversion Lemma," *Proceedings of IEEE*, Vol. 74, N. 7, July.