Capitolo 5

Elementi di calcolo delle probabilità

Come si è visto nel capitolo 3 la misura di una grandezza fisica è sempre affetta da errori il cui valore è, entro certi limiti, inconoscibile ovvero c'è un certo grado di incertezza da associare al valore misurato di ogni grandezza fisica. In matematica le variabili che si comportano in questo modo sono dette *aleatorie* o *casuali* e il *Calcolo delle Probabilità* fornisce gli strumenti matematici e logici con cui trattarle. Quindi considerando le misure delle grandezze fisiche come variabili aleatorie utilizzeremo il calcolo delle probabilità, i cui principi basici sono esposti in questo capitolo, per la costruzione della cosiddetta *Teoria delle Incertezze di Misura*.

Si ritiene che il calcolo delle probabilità nasca nel XVII secolo da uno scambio di corrispondenza tra Fermat e Pascal su come risolvere un problema posto da un giocatore d'azzardo sul lancio dei dati[1]. Nel XVII secolo Lagrange lega la probabilità agli errori di misura e nel 1812 Laplace scrive il primo trattato matematico sulla probabilità[2].

In questo capitolo forniamo una prima trattazione di base del calcolo delle probabilità con l'intento di introdurre i concetti necessari alla trattazione delle incertezze di misura. Per l'approfondimento del calcolo della probabilità si possono consultare i testi riportati in bibliografia ([8],[3]) nei quali si trovano anche numerosi esercizi, che sono un indispensabile complemento per la comprensione del calcolo delle probabilità.

La probabilità, nell'uso comune che si fa di questo termine, riguarda situazioni nelle quali ci si trova in condizioni di incertezza sia per effettuare delle scelte sia di prendere delle decisioni. Tipiche affermazioni qualitative nel linguaggio comune sono:

- è probabile che nel pomeriggio pioverà
- è probabile che se ho lasciato la mia auto in divieto di sosta in centro, troverò una multa.
- è probabile che il governo cadrà in autunno.

Il calcolo delle probabilità permette di fare *valutazioni quantitative* su eventi che hanno esiti incerti.

5.1 Definizioni di probabilità

La probabilità è la misura quantitativa della possibilità che un evento si verifichi. Per convenzione la probabilità di un evento è un numero compreso tra 0 e 1, dove 0 indica l'evento

[1] Su questo problema, noto come il "problema di De Méré" si trova un'ampia documentazione in rete.

[2] Il titolo originale del trattato di Laplace, reperibile in rete, è: "Théorie analytique des probabilités".

impossibile e 1 l'evento certo. L'esempio classico è il lancio di una moneta perfettamente simmetrica (ma con la possibilità di distinguere il lato testa da quello croce). Data la simmetria le probabilità di testa e croce sono uguali e valgono 0.5.

Mentre il concetto di probabilità è intuitivo, la sua definizione, a parte quella originaria detta classica, dovuta a Laplace, è complessa al punto che esistono diverse scuole di pensiero. Schematicamente la probabilità di un evento può essere assegnata secondo:

- Definizione classica o combinatoria (Laplace 1812)
- Definizione frequentista (von Mises, 1920)
- Definizione soggettiva(de Finetti 1937)

5.1.1 Probabilità: definizione classica o combinatoria

Questa è la definizione più naturale e riguarda eventi ideali oppure con caratteristiche molto prossime a quelle ideali come il lancio di un dado perfettamente cubico[3] o l'estrazione di una certa carta da un mazzo ben mescolato o l'estrazione di un numero da un bussolotto o .... La definizione classica della probabilità si formula nel seguente modo:

> *La probabilità di un evento è il rapporto tra il numero dei casi favorevoli e il numero dei casi possibili, purché questi ultimi siano ugualmente possibili*

Si noti che in questa definizione si utilizza il concetto di probabilità per definire la probabilità, infatti l'espressione "eventi ugualmente possibili" equivale ad affermare che sono eventi con la stessa probabilità. Si utilizza quindi il concetto di probabilità per definire la probabilità!

La definizione combinatoria della probabilità è inapplicabile se l'insieme dei casi possibili è infinito.

Questa definizione di probabilità è detta anche "combinatoria" , e nei casi in cui questa definizione è applicabile, l'esecuzione dei calcoli pratici è notevolmente facilitata dall'utilizzo di quella branca della matematica detta *calcolo combinatorio* i cui fondamenti sono esposti nell'appendice B.

5.1.2 Probabilità: definizione frequentista

Consideriamo un esperimento ripetibile, come ad esempio il lancio di un dado o di una moneta o l'estrazione di una carta da un mazzo di 52 carte, e tra i possibili esiti dell'esperimento definiamo un *evento favorevole* (per esempio l'uscita del 6 nel lancio del dado, l'uscita di testa nel lancio della moneta, l'uscita di una carta di cuori nell'estrazione della carta dal mazzo). Definiamo *frequenza*[4] il rapporto tra k, numero di volte in cui l'evento favorevole si è verificato, e n, numero totale delle prove.

La definizione frequentista della probabilità è:

> *La probabilità P di un evento è il limite a cui tende la frequenza al tendere all'infinito del numero delle prove effettuate*

---

[3] Un dado *reale* perfettamente cubico non esiste è solo un'astrazione.

[4] In alcuni testi di statistica il rapporto k/n è detto frequenza relativa, mentre il termine frequenza assoluta (impropriamente) indica il numero di volte (k) che l'evento favorevole si è presentato nelle n prove.

In formule

$$P = \lim_{n\to\infty} \frac{k}{n} \tag{5.1}$$

La definizione (5.1) oltre ad essere molto intuitiva è ampiamente utilizzata in vari campi sia della scienza sia delle assicurazioni (vita, responsabilità civile,...) sia del controllo di qualità dei prodotti industriali.

Il grande merito di questa definizione è l'aver stabilito la corretta relazione tra la frequenza e la probabilità, concetti molto diversi, essendo la prima una quantità calcolata a posteriori, cioè dopo aver effettuato l'esperimento, la seconda una quantità, definita a priori.

Nella concezione classica o combinatoria la probabilità è stabilita a priori, prima di guardare i dati. Nella concezione frequentista il valore della probabilità è stimato a posteriori, dall'esame dei dati. La definizione frequentista (5.1) è soggetta ad alcune critiche; per utilizzarla occorre che 1) le prove che originano gli eventi siano illimitatamente ripetibili e 2) le prove successive devono svolgersi sempre nelle medesime condizioni. Entrambe queste condizioni possono essere verificate solo in modo approssimativo negli esperimenti reali. Inoltre questa definizione non si applica a tutta quella classe di eventi che non sono ripetibili ma per i quali è lecito domandarsi quale sia il loro livello di incertezza. Ad esempio potremo chiederci qual è la probabilità che una particolare stella esploda nei prossimi 30 giorni, oppure qual è la probabilità che un miliardo di anni fa ci fosse vita su Marte, oppure qual è la probabilità che domani a Roma piova. A queste domande e a tutte quelle per le quali è impensabile anche ipotizzare di ripetere all'infinito un esperimento, la definizione frequentista della probabilità non è in grado di dare una risposta.

5.1.3 Probabilità: definizione soggettivista

Come detto all'inizio di questo capitolo il concetto primitivo di probabilità è legato all'incertezza sul verificarsi di un evento ovvero a quanta informazione è nota sulla modalità con cui l'evento accade. Consideriamo ad esempio un soggetto che voglia stimare il valore della probabilità che nel lancio di un dado bene equilibrato esca il "6". Con solo questa informazione, il soggetto stimerà 1/6 questa la probabilità. Supponiamo ora che un altro soggetto riesca a vedere, dopo il lancio, una delle facce laterali del dado; per questo secondo soggetto, in possesso di maggiore informazione rispetto al primo, la probabilità del "6" varrà 0 oppure 1/4 in funzione del numero visto. Quindi allo stesso evento soggetti in possesso di informazioni diverse danno valutazioni differenti della probabilità dell'evento: *la probabilità non è una caratteristica intrinseca di un evento ma il suo valore è condizionato dalla quantità di informazioni sull'evento* Basandoci su quanto detto possiamo introdurre la definizione soggettiva, detta anche bayesiana, della probabilità:

> La Probabilità di un evento è la misura del *grado di fiducia* che un individuo coerente attribuisce, in base alle sue informazioni, al verificarsi dell'evento.

Una definizione più operativa della probabilità soggettiva è basata sul concetto di *scommessa coerente* e si formula nel seguente modo:

> *La probabilità è il prezzo p che un individuo coerente ritiene equo pagare per ricevere 1 se l'evento si verifica*

Per coerenza si intende che il soggetto può assumere indifferentemente, ovvero senza vantaggio, il ruolo dello scommettitore oppure quello di chi accetta la scommessa. In altre parole la coerenza consiste nell'accettazione della scommessa inversa: ricevere p e pagare 1 se l'evento si verifica.

5.2 Teoria Assiomatica della probabilità

Il calcolo della probabilità è stato organizzato in una teoria matematica rigorosa, nota come Teoria Assiomatica della Probabilità, da Kolgomorov nel 1933. Nella teoria assiomatica della probabilità si postula l'esistenza di uno spazio Ω che contiene tutti gli eventi elementari $\{E_i\}$ che possono verificarsi. La probabilità di un evento $P(E)$ è una funzione a valori reali degli eventi che appartengono allo spazio Ω, che soddisfa i seguenti assiomi:

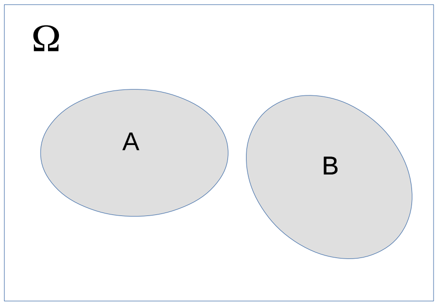
1. Positività: La Probabilità di un evento E_i è un numero positivo o nullo
$$P(E_i) \geq 0$$
2. Certezza (o Unitarietà) : La Probabilità dell'evento certo e quindi dello Spazio Campionario Ω è sempre 1:
$$P(\Omega) = 1$$
3. Unione: Siano E_i e E_j due eventi mutuamente esclusivi (o incompatibili), allora la probabilità della loro unione è la somma delle singole probabilità di E_i e E_j:
$$P(E_i \cup E_j) = P(E_i) + P(E_j) \quad \text{con} \quad E_i \cap E_j = \varnothing$$

Partendo da questi tre assiomi si possono dedurre tutte le regole e i teoremi che regolano il calcolo delle probabilità.

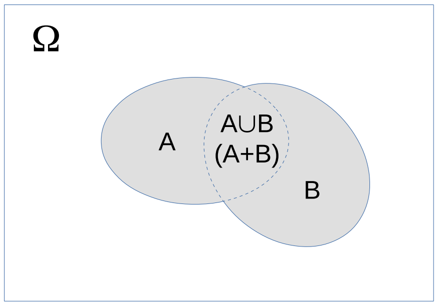
La teoria assiomatica della probabilità per quanto fornisca un modo rigoroso per la deduzione di regole e teoremi, non fornisce alcuna modalità su come assegnare il valore della probabilità agli eventi. In altre parole nella teoria assiomatica della probabilità non esiste una definizione operativa[5] su come calcolare il valore della probabilità degli eventi e tale valore deve essere assegnato in modo esterno alla teoria.

[5]Sul significato di definizione operativa vedi il paragrafo 1.2.

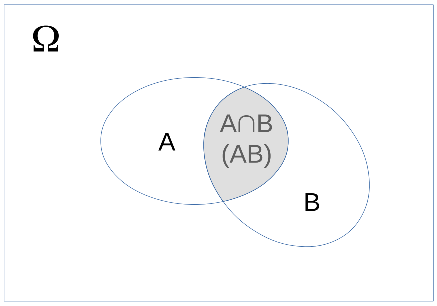
5.3 Probabilità e diagrammi di Wenn



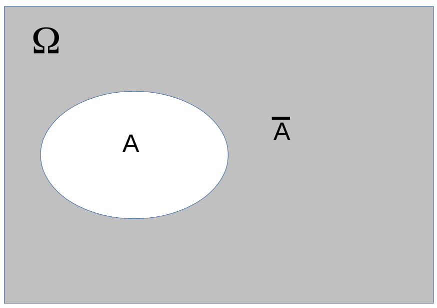
a)



b)



c)



d)

I diagrammi di Wenn forniscono un modo grafico per rappresentare i concetti fondamentali del calcolo delle probabilità, In questa rappresentazione un evento è indicato come una figura chiusa su un piano. Nella figura a) il rettangolo indicato con Ω rappresenta l'insieme di tutti i possibili esiti di un fenomeno aleatorio (o eventi), A e B rappresentano due possibili realizzazioni del fenomeno e con $\varnothing$ si indica l'evento impossibile che non può essere disegnato. Ad esempio supponiamo che Ω rappresenti tutti i possibili esiti del lancio di un dado, A sia l'uscita del numero "3" e B l'evento: uscita del numero "1" oppure del "6". I due eventi sono mutualmente esclusivi e per questo motivo le figure indicate con A e B non si intersecano. Nella figura b) è mostrata una situazione differente: gli eventi A e B non sono mutualmente esclusivi, ad esempio nel lancio di un dado A sia l'uscita del numero "3" oppure del numero "2" e B sia l'uscita di un numero dispari. Il verificarsi dell'evento A *oppure* dell'evento B si indica con $A \cup B$ (notazione insiemistica) oppure con $A + B$ (notazione algebrica, che preferiremo). Nella figura c) è evidenziato il verificarsi dell'evento A *e* dell'evento B che si indica con $A \cap B$ (notazione insiemistica) oppure con AB (notazione algebrica). Usando l'esempio precedente, l'evento AB consiste nell' uscita del numero "3". Nella figura d) è mostrato l'evento A e il suo complementare $\overline{A}$. Come si intuisce dalla figura, l'evento $\overline{A}$ è l'insieme di tutti gli eventi che non sono A. Ovviamente un evento e il suo complementare sono mutualmente esclusivi.

Indicando con $P(x)$ la probabilità che si verifichi l'evento x e riferendosi ai diagrammi di Wenn a lato, avremo, oltre alle proprietà sempre valide: $P(\Omega) = 1$, $P(\varnothing) = 0$, che:

Figura a)

$$P(A + B) = P(A) + P(B)$$
$$P(A \cup B = \varnothing) = 0$$

che deriva direttamente dal terzo postulato di Kolgomorov

Figura b) e c)

$$P(A + B) = P(A) + P(B) - P(AB) \tag{5.2}$$

Questa relazione, oltre ad essere evidente dal diagramma di Wenn, si dimostra osservando che $A + B = A + \overline{A}B$ da cui

$$P(A + B) = P(A) + P(\overline{A}B) \tag{5.3}$$

poiché A e $\overline{A}B$ sono eventi incompatibili. Inoltre essendo $B = AB + \overline{A}B$, si ha $P(B) = P(AB) + P(\overline{A}B)$ o anche $P(\overline{A}B) = P(B) - P(AB)$. Sostituendo questa ultima relazione nella (5.3) si ottiene la (5.2)

Figura d)

$$P(A + \overline{A}) = P(A) + P(\overline{A}) = 1 \quad \text{oppure} \quad P(A + \overline{A}) = P(\Omega) = 1$$

relazione che deriva direttamente dai postulati di Kolgomorov

5.4 Probabilità condizionata

La probabilità condizionata è la probabilità del verificarsi di un certo evento, sapendo o supponendo che si è verificato un altro evento. Se A è l'evento di cui si vuole la probabilità e B è l'evento noto o già avvento o che si presume sia avvenuto, "la probabilità condizionata di A dato B" si indica con $P(A|B)$. Questa probabilità è quindi quella che si verifichi l'evento $A \cap B$ una volta che si sia verificato B. Essendosi ridotto lo spazio degli eventi all'evento B potremo scrivere:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{5.4}$$

Esempio1. Consideriamo il lancio di un dado. L'evento A sia l'uscita del numero "3" la cui probabilità è $1/6$ (un caso favorevole su 6 possibili). L'evento B sia l'uscita di un numero dispari la cui probabilità è $P(B) = 1/2$ (3 casi favorevoli su 6 possibili). La probabilità dell'evento $A \cap B$ è: $P(A \cap B) = 1/6$ (l'evento A è totalmente contenuto in B) e infine la probabilità che accada A una volta che si sia realizzato B si ottiene applicando la (5.4): $P(A|B) = P(A \cap B)P(B) = (1/6)/(1/2) = 1/3$.

Esempio 2. Supponiamo che una scatola contenga 7 lampadine opache tra cui due sono "fulminate". Per trovare le lampadine fulminate le provate una ad una. Qual è la probabilità che le prime due lampadine provate siano quelle fulminate?

Siano F_1 e F_2 rispettivamente gli eventi prova della lampadina fulminata prima e seconda. Si deve calcolare la probabilità dell'evento $F_1 \cap F_2$ (o anche F_1F_2). La probabilità di selezionare per prima una delle due lampadine fulminate è $P(F_1) = 2/7$. La probabilità di selezionare la seconda lampadina fulminata avendo già provato la prima, è $P(F_2|F_1) = 1/6$. Per la (5.4) abbiamo:

$$P(F_1F_2) = P(F_2|F_1)P(F_1) = \frac{2}{7} \cdot \frac{1}{6} = \frac{1}{21} = 4.8\%$$

5.4.1 Eventi indipendenti e teorema della probabilità composta

Due eventi A e B sono detti *indipendenti* o *scorrelati* se il verificarsi dell'uno non influisce sul presentarsi dell'altro, ovvero se

$$P(A|B) = P(A) \quad \text{o equivalentemente} \quad P(B|A) = P(B) \tag{5.5}$$

Esempio. Estrazione di due numeri, con reinserimento, da un urna con 30 palline numerate. Consideriamo i due eventi A: *estrazione del numero 30* e B: *estrazione del numero 18*. Si vuole la probabilità che in due estrazioni successive si verifichi l'evento A "e" l'evento B ovvero la probabilità dell'evento AB. I due eventi A e B sono *indipendenti* quindi per la (5.4):

$$P(AB) \equiv P(A \cap B) = P(A|B)P(B) = P(A)P(B) = \frac{1}{30} \cdot \frac{1}{30} = 1.11 \times 10^{-3}$$

Partendo dalle relazioni (5.5) possiamo formulare il *teorema della probabilità composta*:

> Se A, B, C ..., sono eventi tra loro indipendenti, la probabilità dell'evento $ABC\ldots$ è data dal prodotto delle probabilità dei singoli eventi. In formule:
>
> $$P(ABC\ldots) = P(A)P(B)P(C)\ldots$$

Esempio 1. Giocando tre volte alla *roulette* un numero (non importa quale), qual è la probabilità di vincere tutte e tre le volte? Alla *roulette* la probabilità che esca un certo numero è $1/37$. I tre eventi descritti sono evidentemente indipendenti quindi la probabilità cercata è: $(1/37)(1/37)(1/37) = 2 \times 10^{-5}$.
Esempio 2. In un sacchetto ci sono 5 biglie rosse e 7 nere. Qual è la probabilità di estrarre in due estrazioni una biglia rossa e una biglia nera, avendo cura di reinserire la biglia dopo la prima estrazione? La probabilità di estrarre la biglia rossa è $5/12$ e quella di estrarre la nera $7/12$, essendo gli eventi indipendenti[6] la probabilità cercata è il prodotto delle due:$35/144$.

5.5 Il Teorema di Bayes

Ovviamente nell'evento $A \cap B$ il ruolo degli eventi A e B si può scambiare (l'operatore $\cap$ è commutativo), per cui $P(A \cap B) = P(B \cap A)$. Quindi possiamo scrivere una relazione analoga alla (5.4)

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{5.6}$$

Dalle due relazioni (5.4) e (5.6) otteniamo:

$$P(A|B)P(B) = P(B|A)P(A) \tag{5.7}$$

La (5.7) è una delle forme in cui viene scritto il *Teorema di Bayes* anche noto come *formula di Bayes*. Un altro utile modo di scrivere il Teorema di Bayes è il seguente:

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

Esempio. Supponiamo che una persona abbia due monete in tasca. Una " normale", con testa e croce nei due lati, l'altra "truccata" in cui entrambi i lati riportano testa (indistinguibile dalla testa della prima moneta). La persona prende una delle due monete a caso, la lancia ed esce testa. Qual è la probabilità che la moneta lanciata sia quella normale?

Soluzione. Indichiamo con N la moneta normale, con F quella truccata e con T l'uscita di testa. Le probabilità condizionate utili alla probabilità cercata sono:

$$\begin{aligned}
P(T|N) =& \frac{1}{2}, \quad \text{probabilità che esca testa se la moneta è normale} \\
P(T|F) =& 1, \quad \text{probabilità che esca testa se la moneta è truccata} \\
P(N) =& \frac{1}{2}, \quad \text{probabilità di scegliere la moneta normale} \\
P(F) =& \frac{1}{2}, \quad \text{probabilità di scegliere la moneta truccata}
\end{aligned}$$

[6] Si noti che senza il reinserimento i due eventi descritti non sono più indipendenti

La probabilità che esca testa è:

$$P(T) = P(T|N)P(N) + P(T|F)P(F) = \frac{1}{2} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{3}{4}$$

Usando il teorema di Bayes si ottiene la soluzione:

$$P(N|T) = \frac{P(T|N)P(N)}{P(T)} = \frac{(1/2) \cdot (1/2)}{3/4} = \frac{1}{3}$$

La formula di Bayes è molto utilizzata nell'ambito medico per valutare la "bontà" dei test clinici. Riportiamo una tipica applicazione della formula di Bayes in ambito clinico legata al test sull'infezione da HIV. Un test del sangue per rilevare la presenza del virus dell'HIV ha le seguenti caratteristiche:

- Con sangue infetto il test dà un risultato positivo nel 99.7% dei casi
- Con sangue sano il test dà un risultato negativo nel 92.6% dei casi
- Il 0.5% della popolazione è infettata dal HIV.

Se un cittadino "*preso a caso*" nella popolazione è positivo al test qual è la probabilità che sia realmente infettato dal HIV?

Definiamo B: test positivo e A: soggetto infetto. Si vuole P(A|B)?

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})} \\ &= \frac{0.997 \times 0.005}{0.997 \times 0.005 + (1 - 0.926) \times 0.995} = 6.3\% \end{aligned}$$

Il fatto che la probabilità abbia un valore sorprendentemente piccolo è sostanzialmente dovuto al fatto che la frazione dei cittadini infetti su cui si fa il test è molto piccola (lo 0.5%). Il risultato sarebbe stato completamente differente se il test fosse stato eseguito su un gruppo di cittadini a rischio.

5.6 Distribuzioni di probabilità

La probabilità con cui una variabile aleatoria assume i suoi valori è descritta dalla *Distribuzione di probabilità*. Ad esempio nel lancio di due dadi la variabile aleatoria X somma dei numeri usciti è un numero compreso tra 2 e 12 e la probabilità con cui X assume i valori compresi in questo intervallo è data dal numero di modi in cui X si ottiene dalla somma di due interi compresi tra 1 e 6 diviso per tutte le $6 \times 6 = 36$ possibilità[7]. Calcoliamo esplicitamente qualche valore della distribuzione: i valori $X = 2$ e $X = 12$ si ottengono in un solo modo e quindi $P(1) = P(12) = 1/36 = 0.0278$, mentre $X = 7$ si ottiene delle 6 coppie $(1, 6)$, $(2, 5)$, $(3, 4)$, $(4, 3)$, $(5, 2)$, $(6, 1)$, e quindi $P(7) = 6/36 = 0.167$ Nella figura 5.1 è mostrata in forma di istogramma la rappresentazione grafica della distribuzione di probabilità della somma della variabile aleatoria somma delle uscite di due dadi.

[7] Più esattamente si tratta delle disposizioni con ripetizione (vedi l'appendice B) di due oggetti (i dadi) su 6 posti (i risultati del lancio) $D'_{6,2} = 6^2$

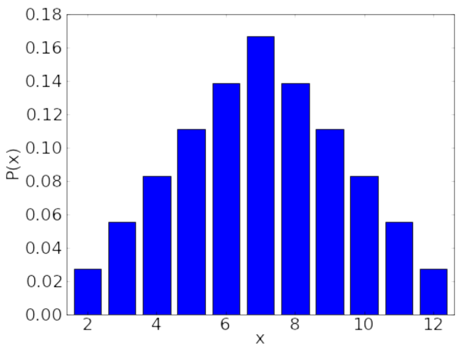


Figura 5.1: Istogramma della Distribuzione della probabilità della variabile aleatoria X: somma delle uscite di due dadi. Ad ogni valore della variabile corrisponde una barra di altezza proporzionale alla probabilità di uscita della variabile

Riassumendo, per una variabile aleatoria discreta X che possa assumere valori $x_1, x_2, \ldots, x_n$ (eventualmente anche con $n = \infty$) diremo che la sua distribuzione di probabilità è rappresentata dall'insieme dei valori di probabilità p_i delle x_i, in formula

$$P(x_i) = p_i \tag{5.8}$$

Se $X = \{x_1, \ldots, x_n\}$ è un insieme completo di eventi mutualmente esclusivi, ovvero la variabile X non può assumere altri valori oltre a quelli indicati, allora è valida la condizione di normalizzazione:

$$\sum_i p_i = 1$$

5.6.1 Distribuzioni di probabilità - Variabili continue

Una variabile continua X può assumere un'infinità non numerabile di valori x. Quindi, diversamente da quanto accade per le variabili discrete, la probabilità che una variabile continua assuma esattamente uno specifico valore reale è nulla. Per una variabile continua X è possibile assegnare una probabilità all'evento *la variabile X è contenuta nell'intervallo* $x, x+\Delta x$, che scriveremo $P(x < X < x+\Delta x)$. Se l'intervallo Δx è infinitesimo allora anche la probabilità è infinitesima e potremo scrivere $dP = (dP/dx)dx = f(x)dx$. La funzione $f(x)$ è detta **funzione di densità di probabilità**, spesso abbreviata con *pdf*[8].

Proprietà della funzione densità di probabilità. La funzione densità di probabilità gode delle seguenti proprietà

• La funzione densità di probabilità è definita positiva (o nulla) per ogni x nel dominio di definizione (Ω).

$$f(x) \geq 0 \qquad \text{per ogni } x \in \Omega$$

• La funzione densità di probabilità è *normalizzata* ovvero il suo integrale esteso a tutto il dominio di definizione della variabile aleatoria vale 1.

$$\int_\Omega f(x)dx = 1$$

[8]L'acronimo deriva dal termine inglese *probability density function*

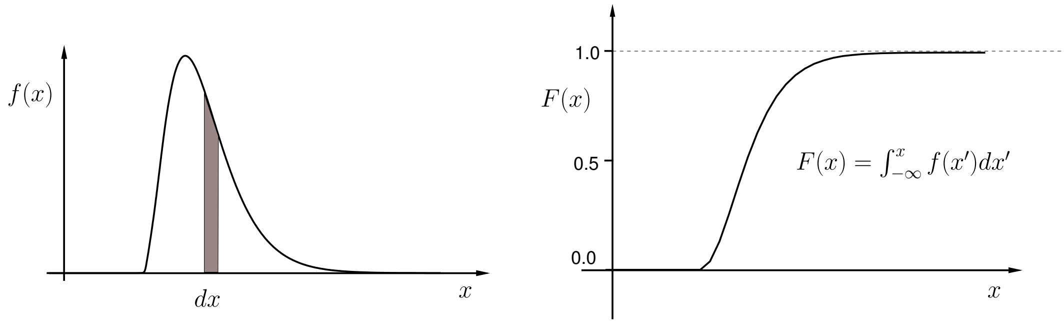


Figura 5.2: Esempio di funzione di distribuzione di densità di probabilità (pdf) per una variabile continua x (grafico di sinistra). Il grafico di destra è la funzione cumulativa di probabilità relativa alla pdf mostrata a sinistra. Si noti che il valore di $F(x)$ per $x \to \infty$ tende a 1 (certezza).

Oltre alla *pdf* per le variabili aleatorie continue si definisce la *funzione di distribuzione cumulativa* $F(x)$ definita definita come la probabilità che la variabile x assuma un qualsiasi valore minore di un determinato valore X

$$F(x) \equiv P(X < x) = \int_{-\infty}^{X} f(x)dx \tag{5.9}$$

5.7 Proprietà delle distribuzioni di probabilità

Le distribuzioni di probabilità sono individuate completamente dalla loro forma matematica, tuttavia possono essere caratterizzate in modo sintetico da un numero limitato di parametri che indicano le loro proprietà essenziali. Tra questi:

1. un parametro che indichi la posizione (il "centro") della distribuzione (valore medio)
2. un parametro che indichi come i valori si accumulano attorno al centro della distribuzione (varianza)
3. un parametro che sia legato alla simmetria della distribuzione
4. un parametro che sia legato alla forma della distribuzione

Lo strumento matematico che utilizzeremo per ottenere i valori di questi parametri è il *Valore Atteso*. (I parametri 3 e 4 verranno descritti nel paragrafo 5.10).

5.7.1 Valore atteso

Consideriamo la variabile aleatoria X, se supponiamo che sia una variabile discreta assumerà i valori $x_1, x_2, \ldots, x_n$ con le probabilità $p_1, p_2, \ldots, p_n$, se invece supponiamo che sia continua avrà una *pdf* $f(x)$. Sia $g(X)$ una generica funzione di X, definiamo *valore atteso* di $g(X)$ l'espressione:

$$\mathbb{E}[g(X)] = \sum_i g(x_i)p_i \qquad\qquad \mathbb{E}[g(X)] = \int_\Omega g(x)f(x)\,dx \tag{5.10}$$

Dove la prima espressione è valida per variabili discrete e la seconda per quelle continue. Una importante proprietà dell'*operatore* valore atteso è la *linearità*. Ovvero se $g(X)$ e $h(x)$ sono due generiche funzioni e a e b sono due costanti, allora vale la relazione

$$\mathbb{E}[ag(X) + bh(X)] = a\,\mathbb{E}[g(X)] + b\,\mathbb{E}[h(X)]$$

Il valore atteso di una costante è la costante stessa:

$$\mathbb{E}[c] = c \tag{5.11}$$

come si deduce dalla definizione stessa di valore atteso. Come corollario della (5.11) vale la relazione: $\mathbb{E}[\mathbb{E}[x]] = \mathbb{E}[x]$.

5.7.2 Valore Medio e Varianza

Un importante caso particolare della (5.10) si ha quando la funzione $g(x)$ coincide con la variabile stessa ($g(x) = x$). Il valore atteso di x è detto anche *Valore Medio* della variabile aleatoria x e spesso è indicato con μ. Per distribuzioni di probabilità discrete e continue abbiamo rispettivamente le seguenti definizioni di *Valore Medio*:

$$\mathbb{E}(X) \equiv \mu = \sum_i x_i p_i \qquad\qquad \mathbb{E}(X) \equiv \mu = \int_\Omega x f(x)\, dx \tag{5.12}$$

Dove, come nella relazione precedente, la prima espressione è valida per variabili discrete e la seconda per quelle continue.

Varianza. Il valore atteso dello *scarto quadratico* definito dalla espressione $(X - \mu)^2$, è detto *Varianza* ($\mathbb{V}\mathrm{ar}[X]$ oppure σ^2).

$$\begin{aligned} \mathbb{V}\mathrm{ar}[X] \equiv \sigma^2 =& \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2] \\ =& \mathbb{E}[X^2] - 2\mu\,\mathbb{E}[X] + \mu^2 = \mathbb{E}[X^2] - \mu^2 \end{aligned} \tag{5.13}$$

In parole: *la varianza della variabile aleatoria X è pari al valore atteso della variabile al quadrato meno il valore quadrato del suo valore atteso.*

La quantità σ, radice quadrata della varianza, è detta *deviazione standard* della variabile X.

Alcune proprietà della varianza. Dalla definizione (5.13) si deduce che la l'operazione di valutare la varianza non è lineare e in particolare, se a e b sono delle costanti, valgono le relazioni

$$\mathbb{V}\mathrm{ar}[aX] = a^2\, \mathbb{V}\mathrm{ar}[x] \tag{5.14}$$

e

$$\mathbb{V}\mathrm{ar}[aX + b] = a^2\, \mathbb{V}\mathrm{ar}[x] \tag{5.15}$$

Le dimostrazioni sono lasciate come facile esercizio per il lettore.

Il valore medio e la deviazione standard di una distribuzione di probabilità discreta o continua danno due importanti informazioni sulla distribuzione: il valore medio indica genericamente (e generalmente!) la zona dove la probabilità è più grande e la deviazione standard indica quanto è ampia questa zona[9] .

Si deve notare che il valore medio e la varianza di una distribuzione *non sempre esistono*. Vedi ad esempio il paragrafo 5.9.5

Esempi di calcolo del valore medio e della varianza (e deviazione standard) di distribuzioni di probabilità discrete e continue saranno date nei prossimi paragrafi.

[9] Questa affermazione è valida approssimativamente in molti casi e serve per dare un'idea di come vanno in genere le cose, ma si possono fare controesempi che la contraddicono

5.8 Distribuzioni di variabili discrete

Nei paragrafi che seguono saranno illustrate alcune distribuzioni discrete con le loro caratteristiche più importanti. Le distribuzioni illustrate sono quelle che maggiormente ricorrono nello studio dei fenomeni fisici.

5.8.1 Il Processo di Bernoulli

Consideriamo un evento che abbia due modalità di presentarsi usualmente dette successo e insuccesso. Esempio classico di questo tipo di eventi è il lancio di una moneta nel quale si hanno i due esiti: Testa (diciamo successo) e Croce (insuccesso). Sia p la probabilità di successo e $q = 1 - p$ quella dei insuccesso. Associamo all'evento la variabile a due valori X: $X = 1$ per successo e $X = 0$ per insuccesso. L'evento con due esiti esclusivi è detto *esperimento di Bernoulli* oppure *prova di Bernoulli* e la variabile X, associata all'esperimento, è detta *variabile di Bernulli.*
Con l'espressione *processo di Bernoulli* si intende una successione (anche infinita) di esperimenti di Bernoulli di uguale parametro p, tra loro indipendenti.

Distribuzione di Bernoulli. La distribuzione di probabilità della variabile a due valori X di Bernoulli è ovviamente:

$$\begin{cases} P(X=0) = 1 - p \\ P(X=1) = p \end{cases} \tag{5.16}$$

Calcoliamo valore medio e varianza della variabile di Bernoulli X:

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1-p) = p \qquad \mathbb{E}[X^2] = 1^2 \cdot p + 0^2 \cdot (1-p) = p$$

$$\mathbb{V}\text{ar}[X] = \mathbb{E}[X^2] - [\mathbb{E}[X]]^2 = p - p^2 = p(1-p) = pq$$

5.8.2 Distribuzione binomiale

La distribuzione Binomiale dà la probabilità del numero k di successi in N processi di Bernoulli. Si consideri ad esempio un evento che si presenta con due sole modalità successo o insuccesso (un processo di Bernoulli come testa o croce, probabilità che esca "6" nel lancio di un dado, ...). Sia p la probabilità di successo e (ovviamente) $q = 1-p$ quella di insuccesso. Ad esempio nel lancio di una moneta $p = q = 1/2$, per l'uscita di "6" nel lancio di un dado (successo) $p = 1/6$ e $q = 1 - 1/6 = 5/6$ per l'insuccesso. Prendiamo in considerazione N prove ripetute di un evento che si può presentare con due modalità e sia p la probabilità della modalità che riteniamo successo; supponiamo inoltre che le N prove siano indipendenti l'una dalle altre; vogliamo calcolare la probabilità di ottenere k successi in N prove. Vista l'indipendenza delle prove possiamo applicare la regole delle probabilità composte per cui la probabilità di avere k successi nelle prime k prove e $N - k$ insuccessi nelle rimanenti prove è data da:

$$\underbrace{p \cdot p \ldots p}_{k \text{ volte}} \cdot \underbrace{(1-p) \ldots \cdot (1-p)}_{N-k \text{ volte}} = p^k (1-p)^{N-k}$$

Poiché non siamo interessati all'ordine con cui otteniamo i k successi, dovremo contare tutte le configurazioni in cui si hanno k successi in N prove, e queste sono le combinazioni[10] di

[10] si veda in proposito l'appendice B.

N oggetti di classe k. Otteniamo infine l'espressione della *distribuzione binomiale*:

$$\mathcal{B}_N(k) = \binom{N}{k} p^k q^{N-k} = \frac{N!}{(N-k)!k!} p^k q^{N-k} \tag{5.17}$$

Normalizzazione. Verifichiamo che la distribuzione binomiale è normalizzata.

$$\sum_{k=0}^{N} \mathcal{B}_N(k) = \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k} = (p+q)^n = 1 \tag{5.18}$$

Valore medio.

$$\begin{aligned}\mu =& \sum_{k=0}^{N} k\mathcal{B}_N(k) = \sum_{k=0}^{N} k \binom{N}{k} p^k q^{N-k} = \sum_{k=1}^{N} \frac{N!}{(N-k)!(k-1)!} p^k q^{N-k} \\ =& Np \sum_{j=0}^{M} \frac{M!}{(M-j)!j!} p^j q^{M-j} = Np \end{aligned} \tag{5.19}$$

dove sono state fatte le posizioni $j = k-1$, $M = N-1$ e si è usata la proprietà di normalizzazione (5.18) della binomiale.

Varianza.

$$\begin{aligned}\mathbb{E}[k^2] =& \sum_{k=0}^{N} k^2\mathcal{B}_N(k) = \sum_{k=0}^{N} k^2 \binom{N}{k} p^k q^{N-k} = \sum_{k=1}^{N} \frac{N!k}{(N-k)!(k-1)!} p^k q^{N-k} \\ =& Np \sum_{j=0}^{M} (j+1) \frac{M!}{(M-j)!j!} p^j q^{M-j} = Np(Mp+1) = Np(Np-p+1) \end{aligned} \tag{5.20}$$

dove sono state usate le stesse posizioni utilizzate per il calcolo del valore medio. Infine con l'uso della (5.13) otteniamo:

$$\mathbb{V}\text{ar}[k] = Np(Np-p+1) - N^2p^2 = Npq \tag{5.21}$$

Somma di variabili binomiali. La somma di due variabili binomiali è ancora una variabile binomiale.

5.8.3 Distribuzione di Poisson o poissoniaina

In molti esperimenti si osservano eventi che accadono in un certo intervallo temporale, o in una certa area, o all'interno di un volume, o di una lunghezza eccetera. Gli eventi a cui ci riferiamo devono essere indipendenti tra loro e se accadono nel tempo la loro probabilità di successo deve essere una costante nell'unità di tempo. Esempi tipici di questi eventi sono il numero di decadimenti radioattivi di una certa sorgente in un intervallo di durata prefissata; oppure il numero di automobili che passano in una strada di traffico non congestionato in un orario prefissato della giornata; oppure il numero di raggi cosmici che attraversano la superficie di un contatore di radiazione ionizzante in un dato intervallo temporale. Ciascuno degli eventi indicati è caratterizzato da un numero intero (gli eventi accaduti) e questo numero appartiene ad una distribuzione di Poisson. La forma matematica della distribuzione di Poisson è:

$$P_\mu(k) = \frac{\mu^k}{k!}\mathrm{e}^{-\mu} \tag{5.22}$$

dove μ è il parametro legato alla probabilità di successo di cui si è appena accennato. La distribuzione di Poisson si ottiene come limite della distribuzione binomiale per $N \to \infty$ e per $p \to 0$ ma con il prodotto $Np = \mu$ che resta finito.

Normalizzazione. Verifichiamo che la distribuzione di Poisson (5.22) è normalizzata. Infatti:

$$\sum P_\mu(k) = \sum_{k=0}^{\infty} \frac{\mu^k}{k!}\mathrm{e}^{-\mu} = \mathrm{e}^{-\mu}\sum_{k=0}^{\infty}\frac{\mu^k}{k!}$$

Essendo l'espressione nel simbolo di sommatoria lo sviluppo in serie di Taylor dell'esponenziale e^{μ}, l'enunciato è dimostrato.

Valore Medio. Per definizione il valore medio di una variabile aleatoria con una distribuzione di probabilità discreta $P(k)$ è $\sum kP(k)$. Nel caso di una distribuzione di Poisson si ha

$$\sum kP_\mu(k) = \sum_{k=0}^{\infty} k\,\frac{\mu^k}{k!}\mathrm{e}^{-\mu} = \mu\sum_{k=1}^{\infty}\frac{\mu^{k-1}}{(k-1)!}\mathrm{e}^{-\mu} = \mu\sum_{j=0}^{\infty}\frac{\mu^j}{j!}\mathrm{e}^{-\mu} = \mu$$

Dove si è posto $j = k - 1$.

Varianza. Per definizione la varianza è il valore atteso dello scarto quadratico: $\mathbb{V}\mathrm{ar}[k] = \mathbb{E}[(k-\mu)^2] = \mathbb{E}[k^2] - \mu^2$. Nel caso della distribuzione di Poisson:

$$\mathbb{E}[k^2] = \sum_{k=0}^{\infty} k^2\,\frac{\mu^k}{k!}\mathrm{e}^{-\mu} = k\sum_{k=1}^{\infty}\mu\frac{\mu^{k-1}}{(k-1)!}\mathrm{e}^{-\mu}$$

e ponendo $j = k - 1$:

$$\mathbb{E}(k^2) = \mu\sum_{j=0}^{\infty}(j+1)\frac{\mu^j}{j!}\mathrm{e}^{-\mu} = \mu\left(\sum_{j=0}^{\infty} j\frac{\mu^j}{j!}\mathrm{e}^{-\mu} + \sum_{j=0}^{\infty}\frac{\mu^j}{j!}\mathrm{e}^{-\mu}\right) = \mu^2 + \mu$$

Da questo risultato si ottiene immediatamente:

$$\mathbb{V}\mathrm{ar}[k] = \mathbb{E}[k^2] - (\mathbb{E}[k])^2 = \mu$$

Caratteristica della distribuzione di Poisson è che il valore medio e varianza hanno lo stesso valore.

La distribuzione di Poisson come limite della binomiale. La distribuzione di Poisson si può ottenere come limite della distribuzione binomiale quando il numero delle prove N tende all'infinito, la probabilità di successo p tende a zero mentre il prodotto Np, che rappresenta il valore medio della binomiale, rimane finito. Dimostriamo questa affermazione con il seguente esempio. Consideriamo eventi casuali che possano accadere in un intervallo di tempo ΔT. Ad esempio il numero di raggi cosmici che fanno scattare un contatore di particelle ionizzanti nell'intervallo di tempo ΔT. Dividiamo l'intervallo ΔT in n intervallini temporali di durata $\Delta T/n$ e supponiamo che:

- La probabilità che un evento capiti in ciascuno intervallino di durata $\Delta T/n$ sia costante.
- La probabilità che si verifichino due o più eventi nello stesso intervallino sia nulla.
- Gli eventi siano indipendenti tra loro.

Avendo diviso l'intervallo di tempo ΔT in n parti e assumendo le ipotesi elencate, l'arrivo dei raggi cosmici nell'intervallo ΔT può essere considerato come n processi di Bernoulli con probabilità di successo $p = \lambda\Delta T/n$ dove λ è una costante con le dimensioni dell'inverso di un tempo. Quindi la probabilità di avere k eventi nell'intervallo ΔT è data dalla binomiale con parametri n e p:

$$\mathcal{B}_{n,p}(k) = \binom{n}{k}\left(\frac{\lambda\Delta T}{n}\right)^k\left(1-\frac{\lambda\Delta T}{n}\right)^{n-k}$$

Se passiamo al limite per $n \to \infty$ si avrà che $p \to 0$, tuttavia il prodotto $np = \lambda\Delta T$ rimane costante. Il limite della binomiale in queste condizioni è la distribuzione di Poisson, infatti:

$$\begin{aligned}
&\lim_{n\to\infty}\frac{n(n-1)\dots(n-k+1)}{k!}\left(\frac{\Delta T\lambda}{n}\right)^k\left(1-\frac{\Delta T\lambda}{n}\right)^{n-k}\\
=&\frac{(\Delta T\lambda)^k}{k!}\lim_{n\to\infty}\frac{n(n-1)\dots(n-k+1)}{n^k}\left(1-\frac{\Delta T\lambda}{n}\right)^{-k}\left(1-\frac{\Delta T\lambda}{n}\right)^{n} \qquad (5.23)\\
=&\frac{(\Delta T\lambda)^k}{k!}\lim_{n\to\infty}1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\Delta T\lambda}{n}\right)^{-k}\left(1-\frac{\Delta T\lambda}{n}\right)^{n} \qquad (5.24)
\end{aligned}$$

nel passaggio dalla (5.23) alla (5.24), nella prima frazione nel limite, è stato messo in evidenza n^k tra i k termini del numeratore: $(n(n-1)\dots(n-k+1))$ semplificandolo con il denominatore. Per $n \to \infty$ tutti i termini della (5.24) tendono a 1 tranne l'ultimo. Per calcolare il limite a cui tende l'ultimo termine, poniamo $y = -n/\Delta T\lambda$, e calcoliamo:

$$\lim_{n\to\infty}\left(1-\frac{\Delta T\lambda}{n}\right)^n = \lim_{y\to\infty}\left[\left(1+\frac{1}{y}\right)^y\right]^{-\Delta T\lambda} = \mathrm{e}^{-\Delta T\lambda}$$

dove si è fatto uso del limite notevole $\lim_{x\to\infty}(1+1/x)^x =$e. Riprendendo la (5.24) otteniamo proprio l'espressione della distribuzione di Poisson:

$$\frac{(\Delta T\lambda)^k}{k!}\mathrm{e}^{-\Delta T\lambda} = \frac{\mu^k}{k!}\mathrm{e}^{-\mu}$$

dove nell'ultimo passaggio si è posto $\mu = \lambda\Delta T$.

Esercizio. L'intensità al livello I del mare dei raggi cosmici penetranti è $I = 1.2 \times 10^2$ particelle /(m^2 s). Un rivelatore di raggi cosmici ha una superficie $A = 10^2\,$cm^2 e viene tenuto acceso per vari intervalli temporali tutti della durata di $\Delta t = 6\,$s.

Determinare quale distribuzione di probabilità seguono i conteggi dei vari intervalli.

Soluzione. Il numero dei conteggi del rivelatore segue la distribuzione di Poisson, infatti sono soddisfatte le condizioni che la caratterizzano:

- Il passaggio di ogni raggio cosmico è un evento (largamente) indipendente dal passaggio di tutti gli altri – eventi indipendenti –
- Il numero degli eventi registrati dal contatore per unità tempo è costante. La "pioggia" di raggi cosmici è uniforme

- due eventi non possono arrivare allo stesso tempo
- la probabilità di contare un raggio è proporzionale alla lunghezza dell'intervallo di tempo

Il valore medio e deviazione standard di questa distribuzione di Poisson sono:

$$\mu = I \cdot A \cdot \Delta t = (1.2 \times 10^2) \times (10^2 \times 10^{-4}) \times 6 = 7.2, \qquad \sigma = \sqrt{\mu} = 2.7$$

Infine la distribuzione è:

$$P(k) = \frac{7.2^k}{k!}\mathrm{e}^{-7.2}$$

Somma di variabili di Poisson. La somma di due variabili di Poisson indipendenti di valori medi μ_1 e μ_2 è ancora una variabile di Poisson di valore medio $\mu_1+\mu_2$. Per dimostrare questa affermazione consideriamo due variabili di Poisson n_1 e n_2 di valori medi μ_1 e μ_2. Le variabili n_1 e n_2 hanno distribuzioni di probabilità $P_{\mu_1}(k) = \mu_1^k \mathrm{e}^{-\mu_1}/k!$ e $P_{\mu_2}(k) = \mu_2^k \mathrm{e}^{-\mu_2}/k!$. Indichiamo con N la loro somma:

$$N = n_1 + n_2$$

Poiché N si ottiene da tutti i valori di $n_1 < N$ sommati a $n_2 = N - n_1$, la probabilità di ottenere il valore N è data da:

$$P(N) = \sum_{n_1=0}^{n_1=N} P_{\mu_1}(n_1)P_{\mu_2}(n_2 = N - n_1)$$

essendo le due variabili n_1 e n_2 indipendenti. Sostituendo l'espressione esplicita delle distribuzioni si ha:

$$P(N) = \sum_{n_1=0}^{n_1=N} \frac{\mu_1^{n_1}}{n_1!}\mathrm{e}^{-\mu_1} \cdot \frac{\mu_2^{(N-n_1)}}{(N-n_1)!}\mathrm{e}^{-\mu_2} = \mathrm{e}^{-(\mu_1+\mu_2)}\frac{1}{N!}\sum_{n_1=0}^{n_1=N} \frac{N!}{(N-n_1)!n_1!}\mu_1^{n_1}\mu_2^{N-n_1} =$$

$$= \mathrm{e}^{-(\mu_1+\mu_2)}\frac{1}{N!}\sum_{n_1=0}^{n_1=N} \binom{N}{n_1}\mu_1^{n_1}\mu_2^{N-n_1} = \frac{(\mu_1+\mu_2)^N}{N!}\mathrm{e}^{-(\mu_1+\mu_2)}$$

che è proprio una distribuzione di Poisson con valore medio $\mu_1 + \mu_2$. Questo risultato si ottiene in modo ancora più semplice con l'uso delle *funzioni generatrici dei momenti delle distribuzioni* come è mostrato nell'appendice C.

5.9 Distribuzioni di variabili continue

In questo paragrafo saranno descritte alcune fra le più comuni distribuzioni di probabilità di variabili aleatorie continue assieme alle loro principali caratteristiche e applicazioni.

5.9.1 Distribuzione uniforme

La variabile aleatoria con una distribuzione uniforme o distribuzione rettangolare in un intervallo a–b è tale che tutti gli intervalli della stessa lunghezza nell'intervallo di definizione sono equiprobabili. In altre parole la *pdf* è costante in a–b (vedi la figura (5.3)). La forma del distribuzione uniforme è:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{per } a < x < b \\ 0 & \text{altrove} \end{cases} \tag{5.25}$$

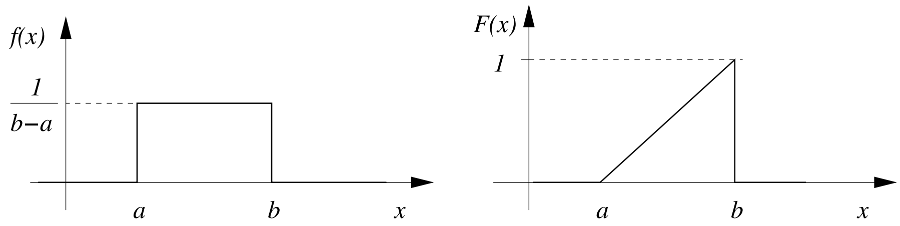


Figura 5.3: La distribuzione uniforme $f(x)$ e la relativa distribuzione cumulativa $F(x)$.

Normalizzazione. La verifica che distribuzione uniforme è normalizzata è immediata.

Valore medio. Per ragioni di simmetria il valore medio della distribuzione uniforme con l'espressione data dalla (5.25) è : $\mu = (a+b)/2$

Varianza. Per ottenere la varianza è necessario calcolare il seguente integrale:

$$\mathbb{E}[x^2] = \int_a^b x^2 \frac{dx}{b-a} = \frac{1}{3}(b^2 + a^2 + ab)$$

e tenendo conto dell'ultima delle (5.13) si ha:

$$\mathbb{V}\text{ar}[x] = \frac{(b-a)^2}{12}$$

Distribuzione cumulativa .La distribuzione cumulativa per la distribuzione uniforme è:

$$F(X) = \int_{-\infty}^{x} f(x')dx' = \int_a^x \frac{dx'}{b-a} = \frac{x-a}{b-a}$$

Esercizio. Una persona arriva ad una fermata di autobus diretti alla stazione ferroviaria. La persona sa che i bus passano ogni 15 minuti senza però conoscere la tabelle degli orari. Sapendo che se il bus non passa entro 5 minuti perde il treno, calcolare la probabilità che la persona riesca a prendere il treno.
Soluzione. La pdf dei tempi di arrivo del bus, per le conoscenze della persona è una costante e vale $(1/15)\,\text{min}^{-1}$. La probabilità che il bus arrivi in $5\,\text{min}$ è quindi $\int_0^5 (1/15)dt = 1/3$.

Applicazione. La distribuzione del valore "vero" di una grandezza misurata con uno strumento digitale è uniforme all'interno dell'intervallo definito dall'ultima cifra del *display* (quella meno significativa). Quindi l'incertezza dovuta <u>alla sola interizzazione della grandezza</u> è $1/\sqrt{12}$ del valore dell'ultima cifra.

5.9.2 Distribuzione triangolare

La distribuzione $f(x)$ di probabilità triangolare della variabile aleatoria continua $x \in (a,b)$ con $a < c < b$, è definita come:

$$f(x) = \begin{cases} \dfrac{2}{(c-a)(b-a)}(x-a) & \text{per } a < x < c \\ \dfrac{2}{(b-a)(b-c)}(b-x) & \text{per } c < x < b \\ 0 & \text{altrove} \end{cases} \tag{5.26}$$

Un esempio di questa distribuzione è mostrato nella figura 5.4. Valore medio della distribuzione triangolare è:

$$\overline{x} = \int_a^b f(x)\,dx = \int_a^c x\frac{2(x-a)}{(c-a)(b-a)}dx + \int_c^b x\frac{2(b-x)}{(b-a)(b-c)}\,dx = \tag{5.27}$$

$$= \frac{(c-a)(a+2c)}{3(b-a)} + \frac{(b-c)(b+2c)}{3(b-a)} = \frac{a+b+c}{3} \tag{5.28}$$

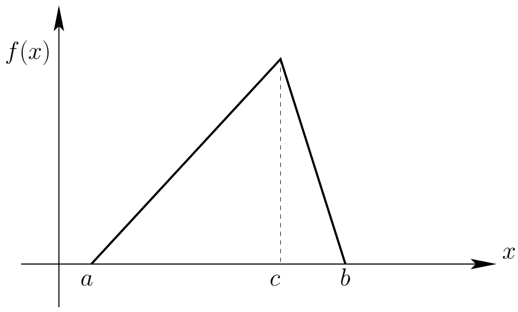


Figura 5.4: Distribuzione triangolare generica.

Nel caso particolare, ma molto comune, in cui la distribuzione triangolare sia simmetrica (con la forma di un triangolo isoscele) allora $c = (a+b)/2$ e la distribuzione (5.26) diviene:

$$f(x) = \begin{cases} \frac{4}{(b-a)^2}(x-a) & \text{per } a < x < c \\ \frac{4}{(b-a)^2}(b-x) & \text{per } c < x < b \\ 0 & \text{altrove} \end{cases} \tag{5.29}$$

In questo caso il valore medio della x è:

$$\overline{x} = \frac{a+b+(a+b)/2}{3} = \frac{a+b}{2}$$

Per il calcolo della varianza operiamo una traslazione della distribuzione centrandola sullo zero dell'asse delle ascisse: $x \to x - (a+b)/2$. Ovviamente questa operazione non influisce sul valore della varianza. Per semplificare la notazione poniamo inoltre $\Delta = b = -a$ per cui $b - a = 2\Delta$. La (5.29) si riduce a

$$f(x) = \begin{cases} \frac{1}{\Delta^2}(x+\Delta) & \text{per } -\Delta < x < 0 \\ \frac{1}{\Delta^2}(\Delta - x) & \text{per } 0 < x < \Delta \\ 0 & \text{altrove} \end{cases} \tag{5.30}$$

Tenendo conto che $\mathbb{E}[x] = 0$, la varianza di una distribuzione triangolare simmetrica è data da:

$$\mathbb{V}\text{ar}[x] = \mathbb{E}[x^2] = \int_{-\Delta}^{0} \frac{x^2}{\Delta^2}(x+\Delta) + \int_0^{\Delta} \frac{x^2}{\Delta^2}(\Delta - x) = \frac{\Delta^2}{6} = \frac{(b-a)^2}{24}$$

5.9.3 Distribuzione gaussiana o normale

La distribuzione di densità di probabilità gaussiana o normale della variabile aleatoria continua x si scrive come:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5.31}$$

dove μ e σ sono due parametri costanti. Come vedremo nel seguito il parametro μ è il valore medio della x e σ^2 è la sua varianza. Per brevità la distribuzione normale (5.31) spesso è indicata con $N(\mu, \sigma)$. Nella figura (5.5) sono riportati i tipici andamenti a campana della gaussiana per alcuni valori di μ e σ.

La variabile z definita come:

$$z = \frac{x - \mu}{\sigma} \tag{5.32}$$

è detta *variabile normale standardizzata*; è facile rendersi conto che la variabile z è una variabile gaussiana con $\mu = 0$ e $\sigma = 1$, la cui distribuzione è:

$$f(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} \tag{5.33}$$

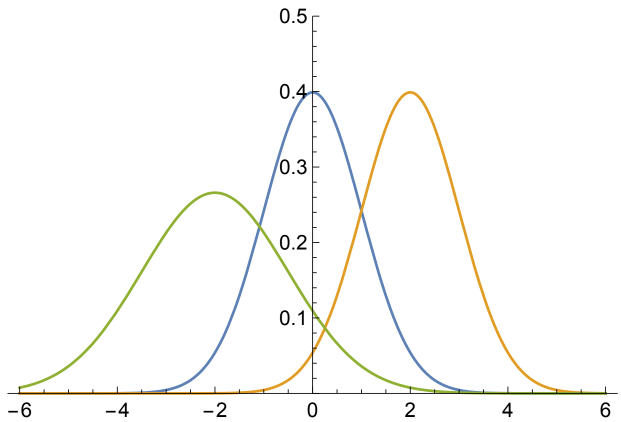


Figura 5.5: Tre distribuzioni di probabilità normali. Partendo da sinistra la prima con $\mu = -2$ e $\sigma = 1.5$ la seconda con $\mu = 0$ e $\sigma = 1.0$ (variabile normale standardizzata) e la terza con $\mu = 2$ e $\sigma = 1.0$. Si noti che la la distribuzione normale è simmetrica e centrata sul suo valore medio. La larghezza della curva è determinata del valore del parametro σ (*deviazione standard*).

Si può dimostrare che la (5.31) è normalizzata[11]

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

Si lasciano come semplici esercizi di calcolo integrale, trovare valore medio e varianza della x, che risultano essere:

$$\mathbb{E}[x] = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu, \quad \mathbb{V}\mathrm{ar}[x] = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2$$

[11] Per verificare la normalizzazione è necessario calcolare l'integrale $\int_{-\infty}^{+\infty} \mathrm{e}^{-x^2}$. Un modo ingegnoso ed elegante per il calcolo di questo integrale si può trovare in [4] oppure in [9].

Distribuzione cumulativa. La funzione cumulativa di probabilità per la distribuzione normale è usualmente indicata con $\Phi(x)$ e vale:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x'-\mu)^2}{2\sigma^2}} dx' \tag{5.34}$$

l'integrale non può essere espresso attraverso funzioni elementari ed esistono tabelle nelle quali sono riportati i valori numerici dell'integrale [12]

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} dz$$

Calcolo della probabilità per la distribuzione normale

Nella teoria degli errori in vari casi è necessario calcolare la probabilità che la variabile normale x sia contenuta in un certo intervallo di estremi a e b (con $a < b$). Utilizzando la (5.31) avremo

$$P(a < x < b) = \frac{1}{\sigma\sqrt{2\pi}} \int_{a}^{b} \mathrm{e}^{-(x-\mu)^2/2\sigma^2} dx \tag{5.35}$$

Per il calcolo di questo integrale, che dipende sia da μ sia da σ, procediamo ad un cambiamento di variabili utilizziamo la variabile normale standardizzata definita nella (5.32). Dalla definizione di z ricaviamo che $dx = \sigma dz$ e l'integrale precedente prende la forma

$$P(a < x < b) = \int_{z_a}^{z_b} \frac{\mathrm{e}^{-z^2/2}}{\sqrt{2\pi}} dz \tag{5.36}$$

dove gli estremi di integrazione sono dati da:

$$z_a = \frac{a-\mu}{\sigma} \qquad\qquad z_b = \frac{b-\mu}{\sigma}$$

Poiché le tabelle forniscono la funzione cumulativa di probabilit'a $\Phi(x)$, definita nella (5.34), la probabilità cercata è data da:

$$P(a < x < b) = \Phi(z_b) - \Phi(z_a) = \Phi\left(\frac{a-\mu}{\sigma}\right) - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

Con questo modo di operare si ottengono le probabilità che la variabile normale x si contenuta in un intervallo di 2, 4 e 6 σ centrato attorno al valore medio (vedi la figura 5.6)

- $P(\mu - \sigma < x < \mu + \sigma) = P(-1 < z < 1) = 0.6826$

---

[12] In letteratura si trovano diverse funzioni che si basano sull'integrale della gaussiana; il loro numero e l'uso di una simbologia spesso non coerente può generare confusione. Per fare un po' di chiarezza ecco un elenco di queste funzioni con la loro definizione matematica:

- Funzione cumulativa di probabilità $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} dz$
- Funzione degli errori: $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-z^2} dz$
- Funzione degli errori complementare: $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \mathrm{e}^{-z^2} dz$

E' facile verificare che vale la relazione:

$$\Phi(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$$

- $P(\mu - 2\sigma < x < \mu + 2\sigma) = P(-2 < z < 2) = 0.9545$

- $P(\mu - 3\sigma < x < \mu + 3\sigma) = P(-3 < z < 3) = 0.9973$

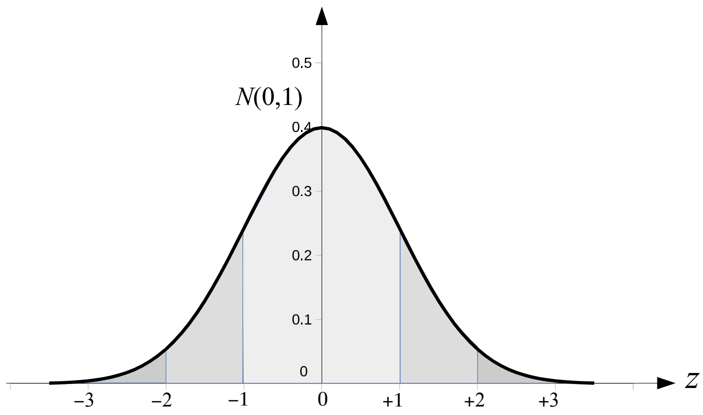


Figura 5.6: Distribuzione della variabile normale standardizzata z. L'area in colore più chiaro indica $P(-1 < z < 1) = 0.6826$, con l'aggiunta dell'area di colore intermedio si ottiene $P(-2 < z < 2) = 0.9545$ e infine aggiungendo l'area di colore più scuro si ottiene $P(-3 < z < 3) = 0.9973$.

5.9.4 Distribuzione t di Student

Con distribuzione *t di Student* (o *distribuzione t*) si indica un qualunque membro di una famiglia di distribuzioni continue di particolare utilità quando si voglia stimare il valore medio di una grandezza distribuita in modo normale ma di varianza non nota. L'espressione esplicita della distribuzione *t di Student* è:

$$f(t) = \frac{1}{\sqrt{\nu}B(1/2, \nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{5.37}$$

Nella *t di Student* compare il parametro ν, reale e positivo, detto *numero di gradi di libertà*. Il termine $\sqrt{\nu}B(1/2, \nu/2)$, nel quale compare la funzione speciale Beta[13], è necessario per normalizzare la distribuzione ($\int f(t)dt = 1$).

Valore medio. Il valore medio della *t di Student* è zero (tranne quando $\nu \leq 1$ nel qual caso è indefinito) in quanto la funzione è simmetrica $f(t) = f(-t)$.

Varianza.

$$\mathbb{V}\text{ar}[t] = \frac{\nu}{\nu - 2} \qquad \text{Esiste solo per } \nu > 2$$

[13]La funzione Beta, detta anche integrale di Eulero di prima specie, è una funzione speciale definita come:

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$$

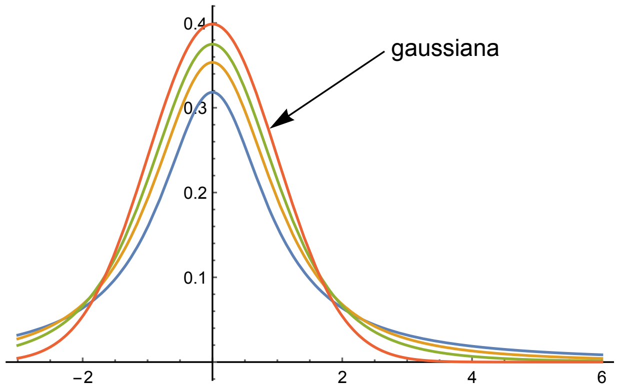


Figura 5.7: Quattro distribuzioni di probabilità t di Student con valore medio nullo e varianza pari 1. Le curve si distinguano per il valore del numero ν dei gradi di libertà: $\nu = 1$ (curva più bassa), $\nu = 2$,$\nu = 4$ e $\nu = \infty$ (curva più alta). La curva con $\nu = \infty$ coincide con una gaussiana con lo stesso valore medio e stessa varianza.

Per $\nu \leq 2$ l'integrale $\int t^2 f(t)dt$ diverge.

La distribuzione t di Student riveste particolare importanza in statistica, infatti descrive la distribuzione di probabilità dell'importante variabile

$$t = \frac{\overline{x} - \mu}{s\sqrt{n}}$$

dove $\overline{x}$ è la media aritmetica di n realizzazioni diella variabile aleatoria x distribuita in modo normale e s è la stima della varianza di x pari a

$$s = \sqrt{\frac{\sum(x-\mu)^2}{n-1}}$$

Torneremo sull'utilizzo della distribuzione t di Student in statistica nel capitolo 6.

5.9.5 Distribuzione di Cauchy o Breit-Wigner o lorenziana

Questa distribuzione è usata in fisica per descrivere i cosiddetti fenomeni di risonanza in fisica atomica e in fisica nucleare e subnucleare. Nei testi di fisica è citata come Breit-Wigner. La densità di probabilità di questa distribuzione è:

$$f(x) = \frac{\Gamma}{\pi} \frac{1}{(x - x_o)^2 + \Gamma^2} \tag{5.38}$$

dove x_o e Γ sono dei parametri reali; il primo rappresenta il centro di simmetria della distribuzione e il secondo, definito positivo, è legato alla larghezza della curva.

La funzione cumulativa è:

$$F(x) = \int_{-\infty}^{x} \frac{\Gamma}{\pi} \frac{1}{(x' - x_o)^2 + \Gamma^2} dx' = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{x - x_0}{\Gamma}$$

La distribuzione di Cauchy (5.38) non ha valore medio ne varianza, infatti come si può facilmente verificare, l'integrale con cui si calcolano il valore medio e la varianza divergono.

La distribuzione di Cauchy è spesso citata come esempio di distribuzione patologica proprio perché priva di valore medio e di varianza.

Diamo alcuni esempi di variabili che seguono la distribuzione di Cauchy. Il rapporto di due variabili gaussiane segue la distribuzione di Cauchy. Consideriamo un angolo θ come variabile aleatoria con una distribuzione uniforme tra $-\pi/2$ $\pi/2$ allora la sua tangente $\tan\theta$ ha una distribuzione di Cauchy. La somma di N variabili di Cauchy è ancora una una variabile di Cauchy. Questa ultima affermazione potrebbe essere erroneamente interpretata come una contraddizione del teorema del limite centrale che sarà esposto nel prossimo capitolo.

5.9.6 Distribuzione del χ^2

Una distribuzioni di probabilità che ha un importante ruolo nella valutazione della qualità dei risultati che si ottengono con l'analisi statistica (argomento che verrà trattato nel capitolo 8) è quella nota come la distribuzione del χ^2 (pronunciare "chi quadro").

Consideriamo n variabili casuali z_i $(i = 1, \ldots, n)$, dove ciascuna delle z_i sia distribuita in modo normale con valore medio nullo e varianza unitaria ($\mathcal{N}(0,1)$). Supponiamo inoltre le variabili z_i siano tra loro indipendenti.

La variabile casuale χ^2 definita come somma dei quadrati delle z_i (variabili normali standard):

$$\chi^2 = \sum_{i=1}^{n} z_i^2 \tag{5.39}$$

ha una distribuzione di densità di probabilità detta *distribuzione del χ^2 con n gradi di libertà* la cui distribuzione[14]la è :

$$f(x|\,n) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{(n/2-1)} \mathrm{e}^{-x/2} \tag{5.40}$$

dove si è sostituito x a χ^2 per maggiore chiarezza grafica.

Riassumendo: *La variabile casuale χ^2 definita come la somma dei quadrati di n variabili indipendenti gaussiane standard z_i^2 $(i = 1, \ldots, n)$ ha una distribuzione di probabilità data dalla (5.40).* Come appare evidente dalla (5.39) e dalla (5.40) la distribuzione della variabile χ^2 ha come unico parametro il numero intero n che viene detto *numero di gradi di libertà*. Nella figura 5.8 è mostrata una distribuzione del χ^2 per alcuni valori del numero di gradi di libertà.

Valore medio. Dalla definizione di valore medio e utilizzando l'integrazione per parti si ha

$$\mathbb{E}[\chi^2] = \int_0^\infty x f(x|\,n)dx = \int_0^\infty \frac{x}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{(n/2-1)} \mathrm{e}^{-x/2} dx = n \tag{5.41}$$

Il valore medio del χ^2 è pari al *numero di gradi di libertà* come può essere anche intuito esaminando la sua definizione in (5.39).

Varianza.

$$\mathbb{V}\mathrm{ar}[\chi^2] = \int_0^\infty (x-n)^2 f(x|\,n)dx = \int_0^\infty \frac{(x-n)^2}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{(n/2-1)} \mathrm{e}^{-x/2} dx = 2n \tag{5.42}$$

[14]L'espressione (5.40) della *pdf* del χ^2 contiene, tra le costanti di normalizzazione, la funzione speciale $\Gamma(x)$. La $\Gamma(x)$ è l'estensione del fattoriale a variabili reali (e complesse). Se n è un intero positivo allora $\Gamma(n) = (n-1)!$; in generale $\Gamma(x) = \int_0^\infty t^{x-1}\mathrm{e}^{-t}dt$.

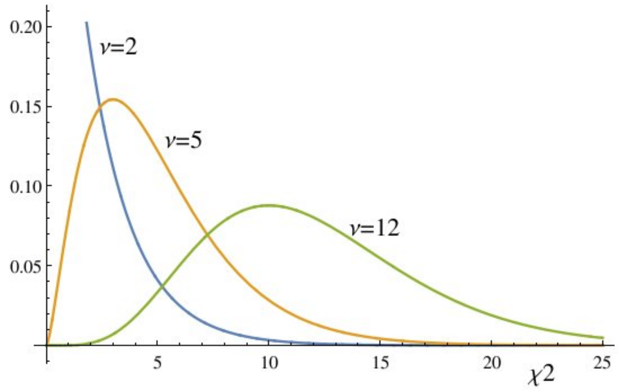


Figura 5.8: Distribuzioni di probabilità della variabile χ^2 per alcuni valori del numero dei gradi di libertà

Il χ^2 ridotto. Il χ^2 ridotto è una nuova variabile casuale definita come il rapporto tra il χ^2 e il numero di gradi di libertà

$$\tilde{\chi}^2 = \frac{\chi^2}{n} \tag{5.43}$$

Dalla questa definizione si deduce immediatamente che:

- il valore medio del χ^2 ridotto è 1: $\mathbb{E}[\tilde{\chi}^2] = 1$
- la varianza del χ^2 ridotto diminuisce come l'inverso del numero dei gradi di libertà: $\mathbb{V}\text{ar}[\tilde{\chi}^2] = 2/n$

5.10 Momenti di variabili aleatorie

Il *momento di ordine n* di una variabile aleatoria, indicato con m_n', è definito come il valore atteso della potenza n-esima della variabile; per variabili aleatorie discrete e continue si ha rispettivamente:

$$m_n' = \sum_k k^n P_n \qquad m_n' = \int_\Omega x^n f(x)\, dx$$

Il momento del primo ordine non è altro che il valore medio della variabile: $m_1' = \mu$.

Momenti centrali di variabili aleatorie. Il *momento centrale di ordine n* m_n di una variabile aleatoria è definito come il valore atteso della potenza n-esima dello scarto della variabile dal suo valore medio; per variabili aleatorie discrete e continue si ha rispettivamente:

$$m_n = \sum_k (k-\mu)^n P_n \qquad m_n = \int_\Omega (x-\mu)^n f(x)\, dx$$

dove μ è il valore medio della variabile. E' facile verificare che $m_1 = 0$ e $m_2 = \sigma^2$. I momenti delle distribuzione possono essere ottenuti tramite le *funzioni generatici dei momenti* che sono introdotte nell'appendice C.

I momenti centrali di ordine 3 e 4 vengono utilizzati per definire due parametri che caratterizzano la forma delle distribuzioni: l'*indice di asimmetria* e la *curtosi*.

Indice di Asimmetria o Skewness. Il parametro con cui si caratterizza l'asimmetria di una distribuzione di probabilità rispetto al suo valore medio è detto *Indice di Asimmetria o Skewness*, è indicato con γ_1 ed è definito dalla relazione:

$$\gamma_1 = \frac{m_3}{\sigma^3} = \mathbb{E}\left[\left(\frac{x-\mu}{\sigma}\right)^3\right] = \frac{1}{\sigma^3}(\mathbb{E}[x^3] - 3\mu\sigma^2 - \mu^3)$$

Nella figura 5.9 sono mostrate due distribuzioni con *skewness* positiva e negativa confrontate con una distribuzione simmetrica e quindi con *skewness* nulla.

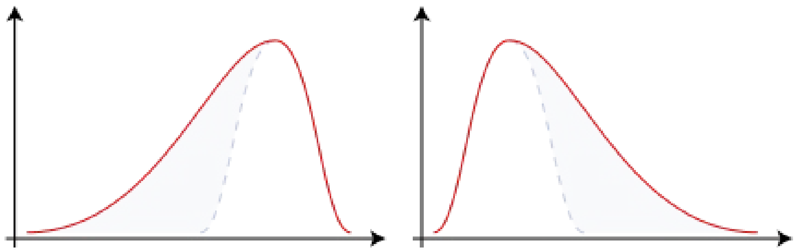


Figura 5.9: Distribuzioni con indice di simmetria (*skewness*) diverso da zero. A sinistra una distribuzione con *skewness* negativa, a destra una distribuzione con *skewness* positiva. La curva tratteggiata nelle figure rappresenta una distribuzione simmetrica (una normale) con *skewness* nulla.

Curtosi o Kurtosis. Questo parametro il cui nome è derivato dal greco $\kappa\upsilon\rho\tau\omega\zeta$ (deformato), indica di quanto una distribuzione di probabilità si "allontana" nella forma dalla distribuzione normale. Il parametro curtosi (o coefficiente di curtosi), è definito dalla espressione:

$$\beta_2 = \frac{m_4}{\sigma^4} - 3$$

Poiché per una gaussiana $m_4/\sigma^4 = 3$, le distribuzioni con $\beta_2 < 0$ sono più "piatte" della gaussiana, quelle con $\beta_2 > 0$ più "appuntite" della gaussiana (vedi la figura 5.10).

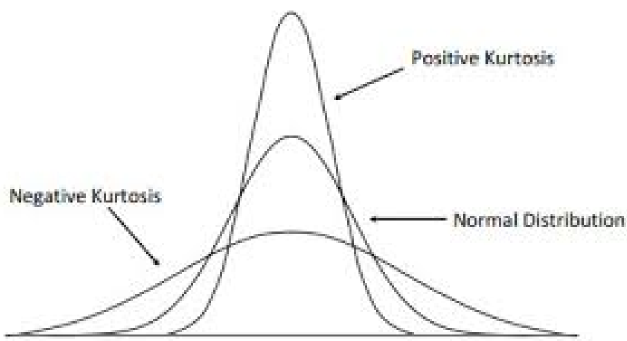


Figura 5.10: Distribuzioni con indice di curtosi negativo, nullo e positivo

Oltre ai parametri già descritti nei paragrafi precedenti una distribuzione di probabilità è caratterizzata da ulteriori parametri tra i quali:

- il *Valore Modale o Moda*: che rappresenta il valore della variabile aleatoria al quale corrisponde il massimo della densità di probabilità. Se la distribuzione è simmetrica, la moda coincide con il valore medio. Se la *pdf* ha più di un massimo si dice che la distribuzione è multimodale.

- la *Mediana*: è il valore $x_{0.5}$ della variabile aleatoria che divide a metà la probabilità. Cioè $P(x < x_{0.5}) = P(x > x_{0.5}) = 0.5$. Se la distribuzione è simmetrica la mediana coincide con il valore medio.

Nella figura 5.11 è mostrata una generica *pdf* con l'indicazione della posizione della moda della mediana e del valore medio.

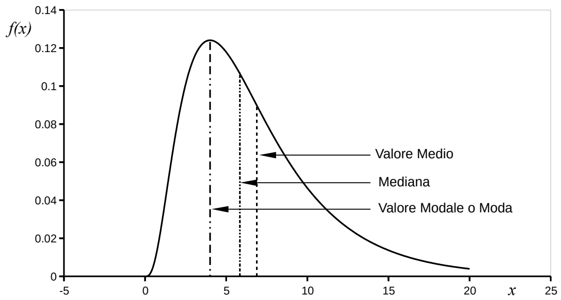


Figura 5.11: Generica distribuzione di probabilità in cui sono indicati i valori della moda della mediana e del valore medio

5.11 Cenni alle Distribuzioni Multivariate

Fino ad ora ci siamo occupati di variabili aleatorie singole, tuttavia in generale un evento può dipendere da più di una variabile aleatoria. In questo caso le distribuzioni di probabilità discrete e continue vengono dette *multivariate*.

Per introdurre i concetti relativi alle distribuzioni multivariate consideriamo il seguente esempio sul lancio contemporaneo di due dadi riconoscibili per esempio dal colore. Prendiamo in considerazione le due variabili aleatorie n_1, valore del dado del colore prescelto, e n_t somma delle uscite dei due dadi. Per ogni coppia delle possibili realizzazioni dell'evento (n_1, n_t) è facile calcolarne la probabilità $P(n_1, n_t)$ che si realizzi. La tabella 5.1 riporta le $P(n_1, n_t)$ in funzione della coppia di valori di n_1 e n_t. La tabella 5.1 rappresenta un esempio di una distribuzione multivariata, in particolare *bivariata*, delle due variabili aleatorie discrete n_1 e n_t.

In generale si potranno avere distribuzioni multivariate di variabili tutte discrete, tutte continue oppure mescolate in parte discrete e in parte continue.

Distribuzioni Marginali. Da una distribuzione multivariata è possibile ottenere la distribuzione di ogni singola variabile sommando sulla probabilità di quelle restanti. Con questa operazione si ottengono le distribuzioni dette *marginali*. Così tornando all'esempio, appena fatto le distribuzioni marginali di n_1 e n_t si ottengono sommando su tutte le modalità di realizzazione della variabile che non interessa: così

$$P(n_1) = \sum_{n_t} P(n_1, n_t) \qquad P(n_t) = \sum_{n_1} P(n_1, n_t)$$

Tabella 5.1: Distribuzione bivariata discreta. La prima riga indica i valori di n_t e la prima colonna i valori di n_1. Per il significato dei simboli vedi il testo.

	2	3	4	5	6	7	8	9	10	11	12	
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0	6/36
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	6/36
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	6/36
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	6/36
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	6/36
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	6/36
	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1

L'ultima colonna e l'ultima riga della tabella 5.1 riportano rispettivamente i valori delle distribuzioni marginali $P(n_1)$ e $P(n_t)$.

Una volta determinate le distribuzioni marginali è possibile calcolare i valori medi delle variabili n_1 e n_t:

$$\mathbb{E}[n_1] = \sum_{n_1=1}^{6} n_1 P(n_1) = \frac{7}{2}, \qquad \mathbb{E}[n_t] = \sum_{n_t=2}^{12} n_t P(n_t) = 7 \tag{5.44}$$

5.11.1 Covarianza e Coefficiente di correlazione

Osservando la tabella 5.1 si nota che c'è una certa correlazione tra le due variabili, si nota cioè che a valori gradi di n_1 aumenta la probabilità di avere valori grandi di n_t e viceversa. Questa osservazione qualitativa viene resa quantitativa introducendo i parametri detti *covarianza* e *coefficiente di correlazione*.

Definiamo covarianza $\mathbb{C}\mathrm{ov}[x, y]$ di due variabili aleatorie x e y l'espressione:

$$\mathbb{C}\mathrm{ov}[x, y] = \mathbb{E}[(x - \mu_x)(y - \mu_y)] \tag{5.45}$$

dove μ_x e μ_y indicano rispettivamente i valori attesi di x e di y.

Il coefficiente di correlazione è definito dalla relazione:

$$\rho(x, y) = \frac{\mathbb{C}\mathrm{ov}[x, y]}{\sigma_x \sigma_y} \tag{5.46}$$

Si dimostra che il coefficiente di correlazione è compreso tra -1 e 1 [15]:

$$-1 \le \rho \le 1 \tag{5.47}$$

Come esempio calcoliamo la covarianza tra le variabili n_1 e n_t utilizzate nell'esempio precedente. Applicando la definizione covarianza, tenendo conto dei valori medi di n_1 e n_t dati

[15] Per dimostrare la (5.47) introduciamo le variabili standardizzate $x^* = (x - \mu_x)/\sigma_x$ e $y^* = (y - \mu_x)/\sigma_y$; dalla loro definizione risulta $\mathbb{E}[x^*] = \mathbb{E}[y^*] = 0$ e $\mathbb{V}\mathrm{ar}[x^*] = \mathbb{E}[x^{*2}] = \mathbb{V}\mathrm{ar}[y^*] = \mathbb{E}[y^{*2}] = 1$. E' un facile esercizio dimostrare che $\rho(x^*, y^*) = \rho(x, y)$. Calcoliamo il valore atteso della quantità $(x^* - y^*)^2$, che essendo un quadrato, è positivo o nullo:

$$0 \le \mathbb{E}[(x^* - y^*)^2] = \mathbb{E}[x^{*2}] + \mathbb{E}[y^{*2}] - 2\,\mathbb{E}[x^* y^*] = 2 - 2\,\mathbb{E}[x^* y^*]$$

da cui si ottiene $\mathbb{E}[x^* y^*] \le 1$. Calcolando il valore atteso di $(x^* + y^*)^2$ si ottiene $\mathbb{E}[x^* y^*] \ge -1$, relazione che completa la dimostrazione della (5.47).

dalla (5.44) e che $n_t = n_1 + n_2$, si ha

$$\mathbb{C}\text{ov}[n_1, n_t] = \mathbb{E}[(n_1 - \frac{7}{2})(n_t - 7)] = \mathbb{E}[n_1 n_t - 7n_1 - \frac{7}{2}n_t + \frac{49}{2}]$$
$$= \mathbb{E}[n_1(n_1 + n_2)] - 7\,\mathbb{E}[n_1] - \frac{7}{2}\,\mathbb{E}[n_t] + \frac{49}{2} \quad = \mathbb{E}[n_1^2] + \mathbb{E}[n_1 n_2] - \frac{49}{2}$$

Il valore aspettato del prodotto $n_1 n_2$ è pari al prodotto del dei valori aspettati in quanto n_1 e n_2 sono tra loro indipendenti: $\mathbb{E}[n_1 n_2] = \mathbb{E}[n_1]\,\mathbb{E}[n_2] = 49/2$. Il valore aspettato di n_1^2 è $\mathbb{E}[n_1^2] = \sum_{k=1}^{6} k^2 P(k) = (1/6)\sum_{k=1}^{6} k^2 = 91/6$. Per il calcolo precedente si è sfruttata la formula che dà la somma dei primi n interi al quadrato[16]: $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$. Sostituendo i valori calcolati otteniamo per la covarianza di n_1 e n_t:

$$\mathbb{C}\text{ov}[n_1, n_t] = \frac{91}{6} - \frac{49}{4} = 2.917 \tag{5.48}$$

Per calcolare il coefficiente di correlazione dobbiamo calcolare le varianze di n_1 e n_t. Si ha:

$$\mathbb{V}\text{ar}[n_1] = \mathbb{E}[n_1^2] - \frac{49}{4} = 2.917$$
$$\mathbb{V}\text{ar}[n_t] = \mathbb{E}[(n_1 + n_2)^2] - 49 = \mathbb{E}[n_1^2] + \mathbb{E}[n_2^2] + 2\,\mathbb{E}[n_1 n_2] - 49 = \frac{91}{3} - \frac{49}{2} = 5.833$$

Siamo ora in grado di calcolare il coefficiente di correlazione ρ tra n_1 e n_t:

$$\rho = \frac{\mathbb{C}\text{ov}[n_1, n_t]}{\sqrt{\mathbb{V}\text{ar}[n_1]}\sqrt{\mathbb{V}\text{ar}[n_t]}} = \frac{2.917}{1.708 \times 2.415} = 0.585$$

questo valore misura in modo quantitativo la tendenza che le due variabili aleatorie n_1 e n_t hanno di crescere (e diminuire) contemporaneamente. Si lascia come facile esercizio il calcolo della covarianza delle due variabili n_1 e n_2 che come si può facilmente prevedere sarà nullo.

Distribuzioni multivariate continue. Nel caso di variabili aleatorie continue la funzione di densità di probabilità sarà una una funzione di più variabili. Estendendo le proprietà delle *pdf* mono-dimensionali e limitandoci al caso di due sole variabili aleatorie X e Y, la probabilità che $x_1 < X < x_2$ e che $y_1 < Y < y_2$ è data da:

$$P(x_1 < X < x_2,\ y_1 < Y < y_2) = \int_{x_1}^{x_2} dx \int_{y_1}^{y_2} dy\ f(x, y)$$

Dovrà inoltre valere la proprietà di normalizzazione della $f(x, y,)$:

$$\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy\ f(x, y) = 1$$

[16] Uno dei modi per ottenere questa formula parte dalla uguaglianza $\sum_{k=1}^{n} k^3 = \sum_{k=1}^{n}(k+1)^3 - (n+1)^3$. Sviluppando il cubo di $(k+1)$

$$\sum_{k=1}^{n} k^3 = \sum_{k=1}^{n} k^3 + 3\sum_{k=1}^{n} k^2 + \sum_{k=1}^{n} k + \sum_{k=1}^{n} 1 - (n+1)^3$$

eliminando la sommatoria dei cubi e isolando la sommatoria dei quadrati si ottiene facilmente $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$.

Distribuzioni marginali. Analogamente a quanto fatto nell'esempio se paragrafo precedente possiamo introdurre le *distribuzioni marginali* $f_x(x)$ e $f_y(y)$ per le variabili aleatorie continue

$$f_x(x) = \int_{-\infty}^{+\infty} dy\ f(x,y); \qquad f_y(y) = \int_{-\infty}^{+\infty} dx\ f(x,y)$$

La distribuzione marginale è la distribuzione della variabile x (y) indipendentemente dal valore assunto dall'altra variabile y (x).

Covarianza per variabili continue. Analogamente a quanto visto nel paragrafo precedente, definiamo covarianza delle variabili aleatorie x e y la quantità:

$$\mathbb{C}\mathsf{ov}[x,y] = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy\ (x-\mu_x)(y-\mu_y) f(x,y) \tag{5.49}$$

Coefficiente di correlazione per variabili continue. Nel caso di variabili aleatorie continue, per il coefficiente di correlazione è sempre valida la relazione (5.46).

5.11.2 Varianza della somma di variabili aleatorie

L'introduzione della covarianza (relazioni (5.45) o (5.49)), permette di completare l'elenco delle proprietà dell'operatore varianza descritte nel paragrafo 5.7.2, aggiungendo l'espressione della varianza della combinazione lineare di due variabili aleatorie. Consideriamo quindi la variabile aleatoria $t = ax + by$ dove x e y sono variabili aleatorie e a e b sono delle costanti e calcoliamone la varianza:

$$\begin{aligned}
\mathbb{V}\mathsf{ar}[t] =& \mathbb{V}\mathsf{ar}[ax+by] = \mathbb{E}[(ax+by-a\mu_x-b\mu_y)^2] = \\
=& \mathbb{E}[(ax-a\mu_x)^2 + (by-b\mu_y)^2 + 2ab(x-\mu_x)(y-\mu_y)] = \\
=& a^2\,\mathbb{V}\mathsf{ar}[x] + b^2\,\mathbb{V}\mathsf{ar}[y] + 2ab\,\mathbb{E}[(x-\mu_x)(y-\mu_y)] = \\
=& a^2\,\mathbb{V}\mathsf{ar}[x] + b^2\,\mathbb{V}\mathsf{ar}[y] + 2ab\,\mathbb{C}\mathsf{ov}[x,y]
\end{aligned} \tag{5.50}$$

La varianza della somma di due variabili tiene quindi conto della eventuale correlazione tra le due variabili come è indicato dalla presenza del termine di covarianza. Nel caso in cui le due variabili x e y siano *indipendenti* allora $\mathbb{C}\mathsf{ov}[x,y] = 0$ e la (5.50) si riduce a:

$$\mathbb{V}\mathsf{ar}[t] = \mathbb{V}\mathsf{ar}[ax+by] = a^2\,\mathbb{V}\mathsf{ar}[x] + b^2\,\mathbb{V}\mathsf{ar}[y] \tag{5.51}$$

5.11.3 Distribuzione bivariata gaussiana

L'espressione matematica della distribuzione normale bivariata con valori medi μ_x, μ_y deviazioni standard σ_x, σ_y e coefficiente di correlazione ρ è data da:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right)\right] \tag{5.52}$$

Utilizzando le variabili standardizzate $z_x = (x-\mu_x)/\sigma_x$ e $z_y = (y-\mu_y)/\sigma_y$, la distribuzione precedente diviene:

$$f(z_x, z_y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{z_x^2 + z_y^2 - 2\rho z_x z_y}{2(1-\rho^2)}\right] \tag{5.53}$$

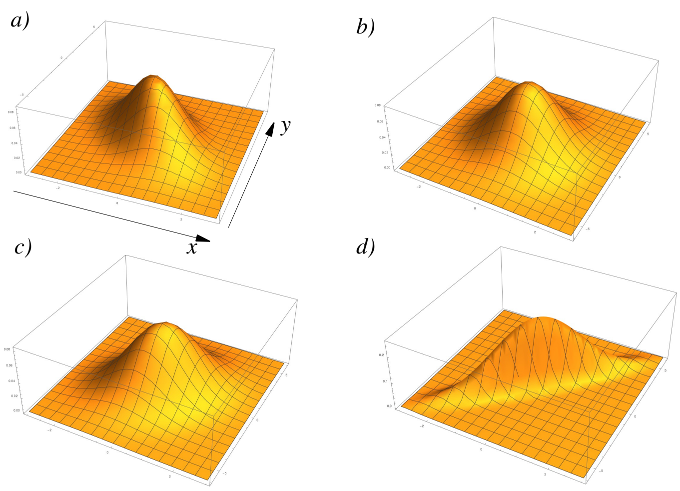


Figura 5.12: Quattro gaussiane bivariate con $\mu_x = \mu_y = 0$, $\sigma_x = \sigma_y = 1$ e *a*) $\rho = -0.4$, *b*) $\rho = 0$, *c*) $\rho = 0.3$, *d*) $\rho = 0.95$.

Quattro esempi di gaussiane bivariate (5.53) sono mostrati nella figura 5.12 dove si può notare il notevole effetto del valore del coefficiente di correlazione sulla forma "a campana" delle distribuzioni. Questa osservazione è rilevante poiché in varie circostanze sperimentali le incertezze sulle grandezze misurate sono fortemente correlate ($\rho \simeq 1$) con un possibile importante effetto sulla valutazione finale delle incertezze da assegnare alle variabili come vedremo in dettaglio in un prossimo capitolo.

5.11.4 Distribuzione multinomiale

Una distribuzione discreta multivariata che è opportuno conoscere è quella detta multinomiale. Infatti la distribuzione multinomiale descrive da un punto di vista statistico gli istogrammi che sono uno strumento fondamentale nell'analisi degli esprimenti di fisica.

La distribuzione multinomiale è una generalizzazione della distribuzione binomiale. Mentre nella binomiale l'evento si presenta in due modalità esclusive, nella multinomiale, l'evento si può presentare in m modalità esclusive ciascuna con la sua probabilità p_i, $(i = 1, \ldots, m)$ con $\sum p_i = 1$. Se N sono le prove ripetute la distribuzione multinomiale ha la seguente espressione:

$$P(k_1, k_2, \ldots, k_m) = \frac{N!}{k_1!k_2!\ldots k_m!} p_1^{k_1} p_2^{k_2} \ldots p_m^{k_m} s \tag{5.54}$$

Valore medio. Il valore medio di ogni variabile k_i:

$$\mathbb{E}[k_i] = Np_i$$

Varianza.

$$\mathbb{V}\text{ar}[k_i] = Np_i(1 - p_i)$$

5.12 Trasformazioni di variabili aleatorie

Consideriamo una variabile continua aleatoria x con *pdf* $f_x(x)$. Se $y = g(x)$ è una funzione continua di x allora anche y è una variabile aleatoria continua cha avrà una *pdf* $f_y(y)$ che ci proponiamo di calcolare. Supponiamo inizialmente che $g(X)$ sia strettamente[17] monotòna come la curva mostrata in figura 5.13(b). La *pdf* della variabile y si ottiene osservando che $f_x(x)dx$, ovvero la probabilità che la variabile abbia un valore compreso nell'intervallo dx attorno a x deve essere uguale a $f_y(y)dy$ probabilità che la variabile dipendente $y = g(x)$ sia compresa nell'intervallo dy attorno a y. Potremo quindi scrivere la seguente equazione:

$$f_x(x)dx = f_y(y)dy \tag{5.55}$$

Dalla (5.55) otteniamo la relazione: $f_y(y) = f_x(x)|dx/dy|$, il modulo è necessario perché

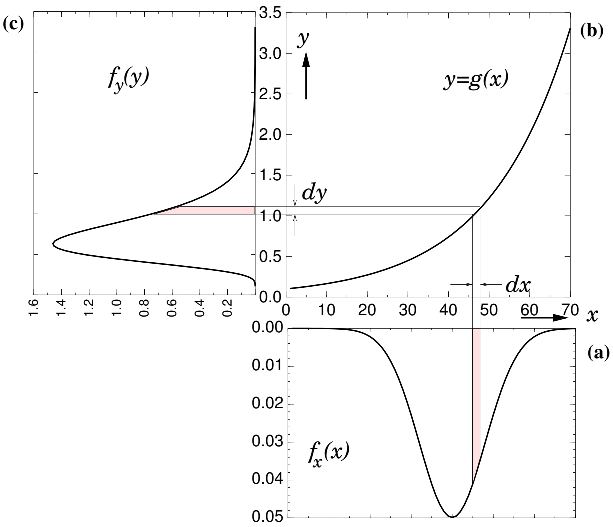


Figura 5.13: Esempio di trasformazioni di variabili aleatorie. Nel grafico (a) è mostrata la distribuzione normale $\mathcal{N}(40, 8)$ della variabile x (si noti che la curva è disegnata invertita, con l'asse delle ascisse crescente verso il basso). Nel grafico (b) è mostrato l'andamento della variabile y legata alla x dalla relazione funzionale: $y = 0.1\mathrm{e}^{0.05\,x}$. Nel grafico (c) è mostrata infine la distribuzione di probabilità della y. Si noti come la la relazione non lineare tra x e y trasformi un distribuzione simmetrica in un'altra non-simmetrica

qualsiasi *pdf* è definita positiva mentre il segno di dx/dy è legato all'andamento è crescente o discendente della funzione $g(x)$. Poiché la $g(x)$ è invertibile (essendo strettamente monotòna) potremo scrivere $x = g^{-1}(y)$, da cui otteniamo infine:

$$f_y(y) = f_x[g^{-1}(y)] \left| \frac{dg^{-1}(y)}{dy} \right| \tag{5.56}$$

Ad esempio calcoliamo la *pdf* della variabile y che dipende dalla variabile casuale x secondo la relazione $y = a\mathrm{e}^{bx}$. Supponiamo che la x abbia una distribuzuione normale $\mathcal{N}(\mu, \sigma)$. Invertiamo la relazione tra x e y: $x = (1/b) \ln y/a$ da cui $dx/dy = 1/b|y|$. Infine

[17] Una funzione $f(x)$ è detta strettamente monotòna se per $x_2 > x_1$, tra $f(x_1)$ e $f(x_2)$ esiste una relazione d'ordine stretto: $f(x_1) < f(x_2)$ oppure $f(x_1) > f(x_2)$. Questa condizione implica che la funzione $f(x)$ è invertibile

applicando la (5.56) si ottiene:

$$f_y(y) = \frac{1}{\sqrt{2\pi}\,\sigma b}\frac{1}{y}\exp\left[-\frac{\left(\ln\frac{y}{a}-\mu\right)^2}{2\sigma^2}\right]$$

Gli andamenti delle funzioni utilizzate in questo esempio sono riportati nella figura 5.13 per valori numerici dei parametri fissati in modo arbitrario.

5.12.1 Funzioni non monotone

Se la dipendenza funzionale $g(x)$ è monotòna a tratti, come ad esempio è mostrato nella figura 5.14, allora possiamo applicare il ragionamento del paragrafo precedente a ciascuno dei tratti in cui la realzione tra x e y è monotòna. Come esempio consideriamo una dipendenza funzionale del tipo mostrato nel grafico 5.14 che indichiamo con $y = g(x)$. Sia x_o il punto che separa i due tratti monotòni della $g(x)$. Allora per $x > x_o$ avremo $x = g_1^{-1}(y)$ mentre per $x < x_o$ avremo $x = g_2^{-1}(y)$. Utilizzando la regola della probabilità totale, potremo scrivere

$$f_y(y)dy = f_x(x_1)dx_1 + f(x_2)dx_2$$

da cui otteniamo il risultato finale:

$$f_y(y) = f_x(g_1^{-1}(y))\left|\frac{dx_1}{dy}\right| + f(g_2^{-1}(y))\left|\frac{dx_2}{dy}\right|$$

La relazione precedente può essere estesa al caso in cui siano presenti N tratti in cui la relazione $g(x)$ è monotòna e si generalizza in:

$$f_y(y) = \sum_i^N f_x(g_i^{-1}(y))\left|\frac{dx_i}{dy}\right|$$

Un esempio in cui si applica quanto qui esposto è il modo in cui si può ricavare la distribuzione

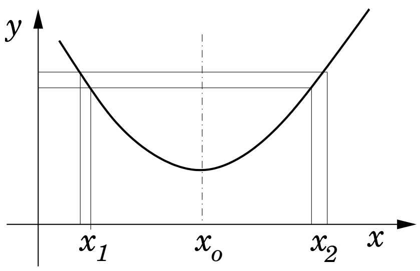


Figura 5.14: funzione monotòna a tratti

del χ^2 per un grado di libertà. Infatti per un grado di libertà $\chi^2 = z^2$ dove z è la variabile gaussiana standard. Si lascia come esercizio ricavare questa distribuzione.

Distribuzione di probabilità della funzione cumulativa. La relazione matematica che definisce la funzione cumulativa di probabilità introdotta nel paragrafo 5.6.1 e che qui riportiamo,

$$F(x) = \int_{-\infty}^{x} f(x')dx' \tag{5.57}$$

può essere interpretata come una dipendenza funzionale tra F e x ed è quindi lecito chiedersi quale sia la distribuzione della variabile F. Una rappresentazione grafica esemplificativa di tale relazione è mostrata nella figura 5.2 (grafico di destra). Di conseguenza la variabile F è una variabile aleatoria con una sua *pdf* che ci proponiamo di calcolare. Differenziando la (5.57) si ottiene:

$$1 \cdot dF = d \int_{-\infty}^{x} f_x(x')dx' = f_x(x)dx \tag{5.58}$$

confrontando questa equazione con la (5.55) si deduce che $f_F(F) = 1$.

In altre parole otteniamo l'importante risultato che la F, interprtata come variabile aleatoria dipendente da x attraverso la (5.57), ha una distribuzione di probabilità uniforme compresa tra 0 e 1 *qualsiasi* sia la distribuzione di x. Questa proprietà della funzione cumulativa è sfruttata, in alcuni casi, nel metodo di calcolo detto "montecarlo" per generare eventi distribuiti con una determinata distribuzione $f_x(x)$. Come esempio supponiamo che x sia distribuita in modo esponenziale $f_x(x) = a\mathrm{e}^{-xa}$, $x \geq 0$. La funzione cumulativa di probabilità è:

$$F = \int_0^x a\mathrm{e}^{-xa} = 1 - \mathrm{e}^{-xa}$$

Invertendo la equazione precedente si ottiene:

$$x = -\frac{1}{a}\ln(1-F)$$

quindi per la (5.57) assegnando a F valori estratti da una distribuzione uniforme tra 0 e 1 le x così ottenute avranno la distribuzione esponenziale $a\mathrm{e}^{-xa}$.