CAPITOLO 1 – Analisi combinatoria

1.1 Introduzione

1.2 Il principio fondamentale del calcolo combinatorio

1.3 Permutazioni

1.4 Disposizioni e combinazioni

1.1 Introduzione

Problema. Un sistema di comunicazione consiste di n antenne allineate. Il sistema è detto funzionante se non vi sono 2 antenne difettose consecutive. Sapendo che esattamente m delle n antenne sono difettose, qual è la probabilità che il sistema sia funzionante?

Ad esempio, se $n = 4$ e $m = 2$ le possibili configurazioni sono 6:

$$
\begin{array}{lcccll}
 & 0 & 1 & 1 & 0: & \text{sistema funzionante} \\
 & 0 & 1 & 0 & 1: & \text{sistema funzionante} \\
(1 = \text{antenna funzionante}) & 1 & 0 & 1 & 0: & \text{sistema funzionante} \\
(0 = \text{antenna difettosa}) & 0 & 0 & 1 & 1: & \text{sistema non funzionante} \\
 & 1 & 0 & 0 & 1: & \text{sistema non funzionante} \\
 & 1 & 1 & 0 & 0: & \text{sistema non funzionante}
\end{array}
$$

Molti problemi del calcolo delle probabilità si risolvono semplicemente calcolando il numero di modi in cui avviene un dato evento; sarà questo l'argomento dell'*analisi combinatoria*.

1.2 Il principio fondamentale del calcolo combinatorio

Principio fondamentale del calcolo combinatorio. Si realizzino 2 esperimenti. Si supponga che il primo esperimento abbia m esiti possibili, e che per ognuno di questi il secondo esperimento abbia n esiti possibili. Allora, se sequenze distinte di esiti dei due esperimenti producono esiti finali distinti, i due esperimenti hanno in tutto mn esiti possibili.

Dimostrazione. Elenchiamo tutti gli esiti dei due esperimenti:

$$
m \text{ righe: } \quad \left\{ \begin{array}{llll} (1,1) & (1,2) & \ldots & (1,n) \quad \leftarrow n \text{ elementi} \\ (2,1) & (2,2) & \ldots & (2,n) \quad \leftarrow n \text{ elementi} \\ \vdots & \vdots & & \vdots \\ (m,1) & (m,2) & \ldots & (m,n) \quad \leftarrow n \text{ elementi} \end{array} \right.
$$

dove diciamo che l'esito finale è (i,j) se il primo esperimento ha prodotto esito i e il secondo esperimento ha prodotto esito j. L'insieme dei possibili esiti consiste quindi di m righe, ognuna contenente n elementi. Vi sono in tutto mn esiti possibili.

Esempio. In un centro studi vi sono 3 laboratori, ed in ognuno di essi 10 computer. Si vuole scegliere un laboratorio ed uno dei computer in esso presenti. Quante sono le scelte possibili?

Soluzione. Si può vedere la scelta del laboratorio come l'esito del primo esperimento e la scelta successiva di uno dei computer in esso presenti come l'esito del secondo esperimento; per il principio fondamentale del calcolo combinatorio vi sono in tutto $3 \times 10 = 30$ scelte possibili.

Esempio. Si lanciano a caso due dadi da gioco. Quanti sono i risultati possibili?

Soluzione. Il lancio del primo dado dà l'esito del primo esperimento, e il lancio del secondo dado dà l'esito del secondo esperimento; per il principio fondamentale del calcolo combinatorio vi sono in tutto $6 \times 6 = 36$ risultati possibili.

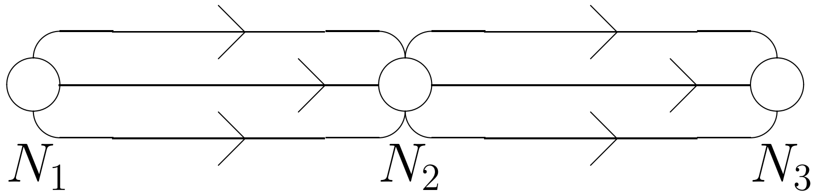
Notiamo che sequenze distinte di esiti dei due esperimenti producono esiti finali distinti; in altri termini (i, j) è un risultato distinto da (j, i).

Esempio. Si lancia a caso un dado da gioco. Se esce un numero pari si lancia nuovamente il dado. Se esce un numero dispari si lancia un dado truccato, in cui il 6 è stato modificato in 5. Quanti sono i risultati possibili?
Soluzione. Suddividendo il calcolo secondo le due alternative relative al primo lancio, per il principio fondamentale del calcolo combinatorio vi sono in tutto $3\times6+3\times5 = 33$ risultati possibili.

Esempio. Quante sono le funzioni booleane definite su $\{0, 1\}$?
Soluzione. Per ogni $x \in \{0, 1\}$ risulta $f(x)$ uguale a 0 o 1, con $f(0)$ corrispondente all'esito del primo esperimento e $f(1)$ all'esito del secondo esperimento. Vi sono pertanto $2 \times 2 = 4$ funzioni siffatte.

x	$f(x) = 0$	$f(x) = x$	$f(x) = \overline{x}$	$f(x) = 1$
0	0	0	1	1
1	0	1	0	1

Esercizio. Si cosideri il seguente grafo orientato. Quanti sono i percorsi distinti possibili che portano dal nodo N_1 al nodo N_3?



Soluzione. $3 \times 3 = 9$.

Esercizio. Si effettuano due esperimenti. Il primo ha m possibili esiti. Se il primo esperimento produce l'esito $i \in \{1, 2, \ldots, m\}$, il secondo può avere n_i possibili esiti. Supponendo che sequenze distinte di esiti dei due esperimenti producano esiti finali distinti, quanti sono gli esiti finali possibili?

Soluzione. $n_1 + n_2 + \ldots + n_m$.

Principio fondamentale (generalizzato) del calcolo combinatorio. Si realizzino r esperimenti. Si supponga che il primo esperimento abbia n_1 esiti possibili, e che per ognuno di questi il secondo esperimento abbia n_2 esiti possibili, e ancora che per ognuno degli esiti dei primi 2 esperimenti il terzo esperimento abbia n_3 esiti possibili, ecc. Allora, se sequenze distinte di esiti degli r esperimenti producono esiti finali distinti, allora gli r esperimenti producono in tutto $n_1 \cdot n_2 \cdots n_r$ esiti possibili.

Esempio. Quante sono le targhe automobilistiche formate da 7 caratteri, di cui 3 numeri e 4 sono lettere (scelte tra le 26 lettere dell'alfabeto anglosassone)?
Soluzione. Per la versione generalizzata del principio fondamentale del calcolo combinatorio vi sono $10 \times 10 \times 10 \times 26 \times 26 \times 26 \times 26 = 456\,976\,000$ targhe possibili.

Esempio. Nell'esempio precedente quante targhe vi sarebbero escludendo le ripetizioni tra numeri e lettere?
Soluzione. Vi sarebbero $10 \times 9 \times 8 \times 26 \times 25 \times 24 \times 23 = 258\,336\,000$ targhe.

Esempio. Si lancia a caso una moneta per n volte. Quante sono le sequenze di risultati possibili?
Soluzione. Ognuno degli n esperimenti consistenti nel lancio della moneta ha 2 possibili esiti, e quindi i risultati possibili sono $2 \times 2 \times \cdots \times 2 = 2^n$.
Ad esempio, per $n = 3$ si hanno 8 risultati: $ccc, cct, ctc, ctt, tcc, tct, ttc, ttt$.

Esercizio. In quanti modi si possono scegliere r oggetti in un insieme di n oggetti, tenendo conto dell'ordine delle scelte?
Soluzione. $n(n-1)(n-2)\cdots(n-r+1)$.

Esempio. Quanti sono i risultati possibili nel gioco del Superenalotto (tenendo conto dell'ordine delle estrazioni)?
Soluzione. Nella prima estrazione vi sono 90 possibili risultati, nella seconda ve ne sono 89, e così via nella sesta ve ne sono 85, quindi i risultati possibili sono in tutto $90 \times 89 \times 88 \times 87 \times 86 \times 85 = 448\,282\,533\,600$.

1.3 Permutazioni

In quanti modi si possono ordinare le lettere a, b, c? Per il principio fondamentale del calcolo combinatorio i casi possibili sono $3 \times 2 \times 1 = 6$: abc, acb, bac, bca, cab, cba. Ciascuno di questi ordinamenti prende il nome di *permutazione*.

Le permutazioni distinte di n oggetti sono $n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1 = n!$

$$
\begin{array}{ll}
0! = 1 & 10! = 3\,628\,800 \\
1! = 1 & 11! = 39\,916\,800 \\
2! = 2 & 12! = 479\,001\,600 \\
3! = 6 & 13! = 6\,227\,020\,800 \\
4! = 24 & 14! = 87\,178\,291\,200 \\
5! = 120 & 15! = 1\,307\,674\,368\,000 \\
6! = 720 & 16! = 20\,922\,789\,888\,000 \\
7! = 5\,040 & 17! = 355\,687\,428\,096\,000 \\
8! = 40\,320 & 18! = 6\,402\,373\,705\,728\,000 \\
9! = 362\,880 & 19! = 121\,645\,100\,408\,832\,000
\end{array}
$$

Esempio. In un centro di calcolo 10 computer eseguono simultaneamente 10 programmi, di cui 4 sono scritti in Linguaggio C e 6 in Java. Viene poi stilato l'elenco dei programmi nell'ordine di terminazione delle esecuzioni. Sapendo che le 10 esecuzioni hanno avuto termine in istanti diversi,
(a) quanti sono i possibili elenchi dei programmi?
(b) quanti sono i possibili elenchi, se i programmi in C e Java compaiono in 2 liste, con precedenza alla lista dei programmi in C?

Soluzione. (a) Ad ogni elenco corrisponde una possibile permutazione di 10 oggetti, quindi la risposta è $10! = 3\,628\,800$.
(b) Vi sono $4!$ elenchi dei programmi in C e $6!$ elenchi dei programmi in Java; per il principio fondamentale del calcolo combinatorio vi sono quindi $4!\cdot 6! = 24\cdot 720 = 17\,280$ diversi elenchi se i programmi in C e Java compaiono in 2 liste, con precedenza alla lista dei programmi in C.

Esempio. Quanti sono gli anagrammi di S T A T I S T I C A ?

Soluzione. Se le 10 lettere da permutare fossero distinte vi sarebbero $10! = 3\,628\,800$ permutazioni possibili. Tuttavia le lettere non sono distinte: se permutiamo le lettere S tra di loro, le lettere T tra di loro, le lettere A tra di loro, e le lettere I tra di loro, si ottiene comunque la stessa parola. Il numero di anagrammi distinti è quindi

$$\frac{10!}{2! \cdot 3! \cdot 2! \cdot 2!} = \frac{3\,628\,800}{2 \cdot 6 \cdot 2 \cdot 2} = 75\,600.$$

Permutazioni di oggetti non tutti distinti

Un ragionamento analogo a quello svolto nell'esempio precedente mostra che vi sono

$$\frac{n!}{n_1!\, n_2! \cdots n_r!}$$

permutazioni distinte di n oggetti presi da r categorie, dei quali n_1 sono identici fra loro, n_2 sono identici fra loro e distinti dai precedenti, $\ldots$, n_r sono identici fra loro e distinti dai precedenti, con

$$n = n_1 + n_2 + \cdots + n_r.$$

Esempio. In un centro di calcolo 10 computer eseguono simultaneamente 10 programmi, di cui 4 sono scritti in Linguaggio C e 6 in Java. Viene poi stilato l'elenco dei programmi nell'ordine di terminazione delle esecuzioni. Quanti sono i possibili elenchi dei programmi se è indicato solo il loro linguaggio?

Soluzione. Gli elenchi possibili sono

$$\frac{10!}{4!\,6!} = \frac{3\,628\,800}{24 \cdot 720} = 210.$$

Esempio. Quanti sono i vettori booleani di dimensione n costituiti da k bit pari a 1 e da $n-k$ bit pari a 0?

Soluzione. I possibili vettori siffatti sono $\dfrac{n!}{k!\,(n-k)!}$.

Per $n = 4$ e $k = 2$ i vettori sono $\dfrac{4!}{2!\,(4-2)!} = 6$: 0011, 0101, 0110, 1001, 1010, 1100.

Esempio. Quanti sono i risultati possibili nel gioco del Superenalotto (senza tener conto dell'ordine delle estrazioni)?

Soluzione. Come visto in precedenza, se si tiene conto dell'ordine delle estrazioni, i possibili risultati sono $90 \times 89 \times 88 \times 87 \times 86 \times 85 = 448\,282\,533\,600$. Dividendo per il numero di possibili permutazioni dei numeri estratti, pari a $6! = 720$, si ottiene il numero dei risultati possibili nel gioco del Superenalotto senza tener conto dell'ordine delle estrazioni:

$$\frac{90 \times 89 \times 88 \times 87 \times 86 \times 85}{6!} = \frac{448\,282\,533\,600}{720} = 622\,614\,630.$$

1.4 Disposizioni e combinazioni

Quanti insiemi di r oggetti si possono formare a partire da n oggetti distinti?

Per risolvere questo problema va specificato se gli insiemi da formare sono *ordinati* (in cui l'ordine è rilevante) o *non ordinati* (dove l'ordine non è rilevante).

Nel caso di insiemi *ordinati* le sequenze da formare si dicono **disposizioni semplici** se non sono ammesse ripetizioni, altrimenti si dicono **disposizioni con ripetizioni**.

Per il principio fondamentale del calcolo combinatorio
– il numero di disposizioni semplici di n oggetti raggruppati in r classi è

$$D_{n,r} = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!} = (n)_r$$

dove $(n)_r := n(n-1)(n-2)\cdots(n-r+1)$ è detto *fattoriale discendente*;
– il numero di disposizioni con ripetizioni di n oggetti raggruppati in r classi è

$$D'_{n,r} = n \cdot n \cdot n \cdots n = n^r.$$

Esempio. Quante parole di lunghezza 2 si possono formare da un alfabeto di 4 lettere
(a) se le lettere non possono ripetersi?
(b) se le lettere possono ripetersi?

Soluzione. (a) Si tratta di disposizioni semplici di $n = 4$ oggetti raggruppati in $r = 2$ classi, quindi $D_{4,2} = (4)_2 = 4 \cdot 3 = 12$.
(ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc)
(b) Si tratta di disposizioni con ripetizioni di $n = 4$ oggetti raggruppati in $r = 2$ classi, quindi $D'_{4,2} = 4^2 = 16$.
(aa, ab, ac, ad, ba, bb, bc, bd, ca, cb, cc, cd, da, db, dc, dd)

Esercizio. Da un'urna che contiene n biglie numerate da 1 a n si effettuano k estrazioni a caso. Quante sono le distinte sequenze ottenibili se le estrazioni
(a) si effettuano senza reinserimento?
(b) si effettuano con reinserimento?

Soluzione. (a) $(n)_k = n(n-1)(n-2)\cdots(n-k+1)$; (b) n^k.

Esempio. Quante sono le operazioni distinte definite su 2 variabili booleane?
Soluzione. Essendo $x_1 \in \{0,1\}$ e $x_2 \in \{0,1\}$, vi sono 4 coppie (x_1, x_2) distinte. Ad ognuna di queste si può assegnare valore 0 o 1. Si tratta quindi di determinare il numero di disposizioni con ripetizioni di 2 oggetti (i valori 0 e 1) in quattro classi (le coppie (x_1, x_2)); vi sono pertanto $D'_{2,4} = 2^4 = 16$ operazioni distinte.

x_1	x_2	0	$x_1 \wedge x_2$	$x_1 \wedge \overline{x_2}$	x_1	$\overline{x_1} \wedge x_2$	x_2	$x_1 \oplus x_2$	$x_1 \vee x_2$
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1

x_1	x_2	$\overline{x_1} \wedge \overline{x_2}$	$\overline{x_1 \oplus x_2}$	$\overline{x_2}$	$x_1 \vee \overline{x_2}$	$\overline{x_1}$	$\overline{x_1} \vee x_2$	$\overline{x_1} \vee \overline{x_2}$	1
0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1

$(\wedge = \text{AND}, \vee = \text{OR}, \oplus = \text{OR ESCLUSIVO})$

Consideriamo il problema di determinare quanti insiemi non ordinati di r oggetti si possono formare a partire da n oggetti distinti.

Nel caso di insiemi *non ordinati* le sequenze si dicono **combinazioni semplici** se non sono ammesse ripetizioni, altrimenti si dicono **combinazioni con ripetizioni**.

In generale, dato che $(n)_r = n(n-1)\cdots(n-r+1)$ rappresenta il numero di scelte di r oggetti tra n, tenendo conto dell'ordine nel quale questi vengono selezionati, e dato che ogni insieme di r oggetti viene in tal modo contato $r!$ volte, si ha che il numero di sottoinsiemi di r oggetti che si possono formare da un insieme di n oggetti è

$$\frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{(n-r)!\,r!} = \frac{(n)_r}{r!}.$$

Per $r = 0, 1, \dots, n$ definiamo

$$\binom{n}{r} = \frac{n!}{(n-r)!\,r!} = \frac{(n)_r}{r!} = \frac{n(n-1)\cdots(n-r+1)}{r!}.$$

Se $r < 0$ o se $r > n$ si pone

$$\binom{n}{r} = 0.$$

Il numero $C_{n,r}$ di combinazioni semplici di n oggetti raggruppati in r classi è

$$C_{n,r} = \binom{n}{r} = \frac{n!}{(n-r)!\,r!} = \frac{(n)_r}{r!} = \frac{n(n-1)\cdots(n-r+1)}{r!}.$$

Questo denota il numero di sottoinsiemi di dimensione r che si possono formare con gli elementi di un insieme di dimensione n senza tener conto dell'ordine della selezione.

Il numero $C'_{n,r}$ di combinazioni con ripetizioni di n oggetti raggruppati in r classi è

$$C'_{n,r} = \binom{n+r-1}{r} = \frac{(n+r-1)!}{(n-1)!\,r!} = \frac{(n+r-1)_r}{r!} = \frac{(n+r-1)(n+r-2)\cdots n}{r!}.$$

Esempio. Quante combinazioni di 4 oggetti in gruppi di 2 si possono formare
(a) nel caso di combinazioni semplici? (gli oggetti non possono ripetersi)
(b) nel caso di combinazioni con ripetizioni? (gli oggetti possono ripetersi)

Soluzione. (a) $C_{4,2} = (4)_2/2! = (4 \cdot 3)/2 = 6.$
(ab, ac, ad, bc, bd, cd)
(b) $C'_{4,2} = (5)_2/2! = (5 \cdot 4)/2 = 10.$
(aa, ab, ac, ad, bb, bc, bd, cc, cd, dd)

Esercizio. Una classe di tango argentino ha 22 studenti, 10 donne e 12 uomini. In quanti modi si possono formare 5 coppie?

Soluzione. Vi sono $\binom{10}{5}$ modi di selezionare 5 persone da 10 donne, $\binom{12}{5}$ modi di selezionare 5 persone da 12 uomini, e 5! modi di accoppiare 5 donne con 5 uomini, quindi la soluzione è $\binom{10}{5}\binom{12}{5} 5! = 252 \cdot 792 \cdot 120 = 23\,950\,080.$

Esempio. In quanti modi si può scegliere un insieme di 3 nodi dai 20 nodi di un grafo?

Soluzione. $C_{20,3} = \binom{20}{3} = \frac{(20)_3}{3!} = \frac{20 \cdot 19 \cdot 18}{3 \cdot 2 \cdot 1} = 1140.$

Esempio. (a) Quanti comitati composti da 2 donne e 3 uomini si possono formare da un gruppo di 5 donne e 7 uomini? (b) Quanti sono i comitati se 2 uomini che hanno litigato rifiutano di sedere insieme nel comitato?

Soluzione. (a) Ci sono $\binom{5}{2}$ possibili insiemi con 2 donne, e $\binom{7}{3}$ possibili insiemi di 3 uomini; segue allora dal principio fondamentale del calcolo combinatorio che vi sono

$$\binom{5}{2}\binom{7}{3} = \frac{5 \cdot 4}{2 \cdot 1} \cdot \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 350 \quad \text{comitati possibili formati da 2 donne e 3 uomini.}$$

(b) Se due uomini rifiutano di far parte insieme nel comitato, vi sono $\binom{2}{0}\binom{5}{3}$ insiemi di 3 uomini che non contengono nessuno dei 2 litiganti, e $\binom{2}{1}\binom{5}{2}$ insiemi di 3 uomini che contengono esattamente 1 dei 2 litiganti, e quindi vi sono $\binom{2}{0}\binom{5}{3} + \binom{2}{1}\binom{5}{2} = 30$ gruppi di 3 uomini senza i 2 litiganti. In totale vi sono $30 \cdot \binom{5}{2} = 300$ comitati.

Tavola (di Tartaglia-Newton) dei coefficienti binomiali

$n\backslash k$	0	1	2	3	4	5	6	7	somma
0	1	0	0	0	0	0	0	0	1
1	1	1	0	0	0	0	0	0	2
2	1	2	1	0	0	0	0	0	4
3	1	3	3	1	0	0	0	0	8
4	1	4	6	4	1	0	0	0	16
5	1	5	10	10	5	1	0	0	32
6	1	6	15	20	15	6	1	0	64
7	1	7	21	35	35	21	7	1	128

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{(n)_k}{k!}$$

$$\binom{n}{0} = \frac{n!}{0!\,n!} = \frac{n!}{n!} = 1; \qquad \binom{n}{n} = \frac{n!}{n!\,0!} = \frac{n!}{n!} = 1;$$

$$\binom{n}{1} = \frac{n!}{1!\,(n-1)!} = \frac{n\,(n-1)!}{(n-1)!} = n; \qquad \binom{n}{n-1} = \frac{n!}{(n-1)!\,1!} = \frac{n\,(n-1)!}{(n-1)!} = n;$$

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n!}{(n-k)!\,k!} = \binom{n}{n-k}.$$

Formula di ricorrenza dei coefficienti binomiali

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r} \qquad 1 \leq r \leq n.$$

Dimostrazione analitica

$$\begin{aligned}
\binom{n-1}{r-1} + \binom{n-1}{r} &= \frac{(n-1)!}{(r-1)!\,(n-r)!} + \frac{(n-1)!}{r!\,(n-r-1)!} \\
&= \frac{(n-1)!}{(r-1)!\,(n-r)\,(n-r-1)!} + \frac{(n-1)!}{r\,(r-1)!\,(n-r-1)!}
\end{aligned}$$

$$= \left[\frac{1}{n-r} + \frac{1}{r}\right] \frac{(n-1)!}{(r-1)!\,(n-r-1)!} = \frac{n}{r\,(n-r)}\,\frac{(n-1)!}{(r-1)!\,(n-r-1)!} = \binom{n}{r}.$$

Esercizio. Scrivere un programma che usi la formula di ricorrenza dei coefficienti binomiali per ottenere la tavola di Tartaglia-Newton per $\binom{n}{r}$, $0 \leq r \leq n \leq N$, con N assegnato. Quali sono le condizioni iniziali da usare?

Formula di ricorrenza dei coefficienti binomiali

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r} \qquad 1 \leq r \leq n.$$

Dimostrazione combinatoria

Si basa sul fatto che i sottoinsiemi di r elementi di un insieme di n oggetti sono $\binom{n}{r}$. Consideriamo un insieme di n oggetti e fissiamo l'attenzione su uno di essi, che chiamiamo *oggetto 1*. Vi sono $\binom{n-1}{r-1}\binom{1}{1}$ sottoinsiemi di r elementi che contengono l'oggetto 1. Inoltre vi sono $\binom{n-1}{r}\binom{1}{0}$ sottoinsiemi di r elementi che non contengono l'oggetto 1. La somma dei termini $\binom{n-1}{r-1} + \binom{n-1}{r}$ fornisce il numero $\binom{n}{r}$ di sottoinsiemi di r elementi.

Teorema del binomio

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k\, y^{n-k}, \qquad n \geq 1.$$

$\binom{n}{k}$ è detto *coefficiente binomiale* perché interviene nello sviluppo del binomio.

Esercizio. Dimostrare il teorema del binomio per induzione su n.

Esempio. Quanti sono i sottoinsiemi di un insieme di n elementi?

Soluzione. Poiché vi sono $\binom{n}{k}$ sottoinsiemi di dimensione k, dal teorema del binomio per $x = y = 1$ si ha

$$\sum_{k=0}^{n} \binom{n}{k} = (1+1)^n = 2^n.$$

(Si veda anche l'ultima colonna della tavola dei coefficienti binomiali).

Nella somma $\displaystyle\sum_{k=0}^{n}\binom{n}{k} = 2^n$ è incluso il caso $k = 0$ che corrisponde all'insieme vuoto.
Quindi il numero di sottoinsiemi non vuoti di un insieme di n elementi è

$$\sum_{k=1}^{n}\binom{n}{k} = 2^n - 1.$$

Analogamente, il numero di sottoinsiemi costituiti da almeno 2 elementi è

$$\sum_{k=2}^{n}\binom{n}{k} = 2^n - n - 1.$$

Esercizio. Utilizzare il teorema del binomio per dimostrare che

$$\sum_{i=0}^{n}(-1)^i\binom{n}{i} = 0, \qquad n > 0.$$

Proposizione. Per ogni $n \geq k \geq 0$ sussiste la seguente identità:

$$\sum_{j=k}^{n} \binom{j}{k} = \binom{n+1}{k+1}.$$

Dimostrazione. Procedendo per induzione su n, notiamo che per $n = k$ l'identità è valida, essendo

$$\binom{k}{k} = \binom{k+1}{k+1} = 1.$$

Supponendo valida l'identità per n, vediamo che essa sussiste anche per $n+1$. Si ha

$$\sum_{j=k}^{n+1} \binom{j}{k} = \sum_{j=k}^{n} \binom{j}{k} + \binom{n+1}{k} = \binom{n+1}{k+1} + \binom{n+1}{k}.$$

Ricordando la formula di ricorrenza $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$ si ottiene $\sum_{j=k}^{n+1} \binom{j}{k} = \binom{n+2}{k+1}$, da cui segue immediatamente la tesi.

È possibile mostrare che i sottoinsiemi di un insieme di n elementi sono 2^n anche nel seguente modo: assegnamo ad ogni elemento dell'insieme il valore 0 o 1. Ad ogni assegnazione corrisponde il sottoinsieme (unico) i cui elementi sono quelli con il valore pari a 1. Le possibili assegnazioni sono $\{(x_1, x_2, \ldots, x_n) : x_i \in \{0, 1\}, i = 1, 2, \ldots, n\}$. Poiché vi sono 2^n assegnazioni possibili, tanti sono i possibili sottoinsiemi.

Esempio. Per $n = 4$ vi sono $2^4 = 16$ sottoinsiemi di $\{a, b, c, d\}$:

$$
\begin{array}{ll}
(0,0,0,0) : \{\} & (1,0,0,0) : \{a\} \\
(0,0,0,1) : \{d\} & (1,0,0,1) : \{a,d\} \\
(0,0,1,0) : \{c\} & (1,0,1,0) : \{a,c\} \\
(0,0,1,1) : \{c,d\} & (1,0,1,1) : \{a,c,d\} \\
(0,1,0,0) : \{b\} & (1,1,0,0) : \{a,b\} \\
(0,1,0,1) : \{b,d\} & (1,1,0,1) : \{a,b,d\} \\
(0,1,1,0) : \{b,c\} & (1,1,1,0) : \{a,b,c\} \\
(0,1,1,1) : \{b,c,d\} & (1,1,1,1) : \{a,b,c,d\}
\end{array}
$$

Esercizio. Calcolare quanti sono i vettori $(x_1, \ldots, x_k)$ nei quali
(a) ogni x_i è un intero positivo tale che $1 \leq x_i \leq n$;
(b) ogni x_i è un intero positivo tale che $1 \leq x_i \leq n$, ed inoltre ogni x_i è diverso da ciascun intero $x_1, x_2, \ldots, x_{i-1}$;
(c) ogni x_i è un intero positivo tale che $1 \leq x_i \leq n$, ed inoltre $x_1 < x_2 < \cdots < x_k$;
(d) ogni x_i è un intero positivo tale che $1 \leq x_i \leq n$, ed inoltre $x_1 \leq x_2 \leq \cdots \leq x_k$.

Soluzione. (a) $D'_{n,k} = n^k$; (b) $D_{n,k} = (n)_k$; (c) $C_{n,k} = \binom{n}{k}$; (d) $C'_{n,k} = \binom{n+k-1}{k}$.

Esercizio. Scrivere un programma che, per valori di k e di n assegnati,
(i) generi tutte le sequenze dei 4 casi dell'esercizio precedente;
(ii) conti quante sequenze sono state generate;
(iii) valuti le quantità n^k, $(n)_k$, $\binom{n}{k}$, $\binom{n+k-1}{k}$;
(iv) verifichi che i numeri di sequenze generate nei 4 casi corrispondano alle quantità valutate al punto (iii).

Esercizio. Dimostrare le seguenti uguaglianze:

$$\sum_{i_1=1}^{n}\sum_{\substack{i_2=1\\ i_2\neq i_1}}^{n}\cdots\sum_{\substack{i_k=1\\ i_k\neq i_r\ \forall r<k}}^{n} 1=(n)_k,\qquad \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n} 1=n^k,$$

$$\sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1-1}\cdots\sum_{i_k=1}^{i_{k-1}-1} 1=\binom{n}{k},\qquad \sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1}\cdots\sum_{i_k=1}^{i_{k-1}} 1=\binom{n+k-1}{k}.$$

Esercizio. Dimostrare la seguente uguaglianza (detta di Vandermonde):

$$\binom{n+m}{k}=\sum_{i=0}^{k}\binom{n}{i}\binom{m}{k-i}=\binom{n}{0}\binom{m}{k}+\binom{n}{1}\binom{m}{k-1}+\cdots+\binom{n}{k}\binom{m}{0}.$$

Suggerimento: Considerare n biglie nere e m biglie bianche. In quanti modi si può formare un sacchetto di k biglie?

Esercizio. Un lotto di 4 computer distinti deve essere assegnato a 2 laboratori.
(a) In quanti modi può essere fatto?
(b) E se ogni laboratorio deve ricevere almeno un computer?

Soluzione. (a) Assegnando k computer al primo laboratorio e $4-k$ al secondo laboratorio, l'assegnazione dei computer si può effettuare in $\binom{4}{k}$ modi; si ha quindi:

$$\sum_{k=0}^{4} \binom{4}{k} = 2^4 = 16;$$

$\{\}\{1234\}$, $\{1\}\{234\}$, $\{2\}\{134\}$, $\{3\}\{124\}$, $\{4\}\{123\}$, $\{12\}\{34\}$, $\{13\}\{24\}$, $\{14\}\{23\}$, $\{23\}\{14\}$, $\{24\}\{13\}$, $\{34\}\{12\}$, $\{123\}\{4\}$, $\{124\}\{3\}$, $\{134\}\{2\}$, $\{234\}\{1\}$, $\{1234\}\{\}$.

(b) Analogamente si ha:

$$\sum_{k=1}^{3} \binom{4}{k} = 2^4 - \binom{4}{0} - \binom{4}{4} = 14;$$

sono le stesse assegnazioni del caso (a) tranne $\{\}\{1234\}$ e $\{1234\}\{\}$.

CAPITOLO 2 – Assiomi della probabilità

2.1 Introduzione

2.2 Spazio campionario ed eventi

2.3 Assiomi della probabilità

2.4 Alcune semplici proprietà

2.5 Spazi campionari con esiti equiprobabili

2.6 La probabilità come funzione di insieme continua

2.7 La probabilità come misura della fiducia

2.1 Introduzione

In questo capitolo introduciamo il concetto di

probabilità di un evento

e quindi mostriamo come le probabilità possano essere calcolate in certe situazioni.

Preliminarmente avremo però bisogno di definire i concetti di

spazio campionario

e di

evento di un esperimento.

2.2 Spazio campionario ed eventi

Chiameremo *esperimento* qualunque fenomeno il cui risultato non possa essere previsto con certezza. Sebbene l'esito dell'esperimento non sia noto a priori, supponiamo che l'insieme di tutti i possibili esiti lo sia. Definiamo questo insieme *spazio campionario* dell'esperimento e lo denotiamo con S; i suoi elementi sono detti *eventi elementari*.

Esempio. Nell'esperimento che consiste nel lancio di 2 dadi, lo spazio campionario consiste di $D'_{6,2} = 6^2 = 36$ elementi:

$$S = \{(i, j) : i, j = 1, 2, 3, 4, 5, 6\},$$

dove (i, j) indica che il primo dado mostra il numero i e l'altro dado il numero j.

Esempio. Se l'esito dell'esperimento è l'ordine di arrivo di una competizione sportiva con n partecipanti, lo spazio campionario consiste di $n!$ elementi:

$$\begin{aligned} S &= \{\text{tutte le } n! \text{ permutazioni di } (1, 2, \ldots, n)\} \\ &= \{(\omega_1, \omega_2, \ldots, \omega_n) : \forall i,\ \omega_i \in \{1, 2, \ldots, n\},\ \omega_i \notin \{\omega_1, \omega_2, \ldots, \omega_{i-1}\}\} \end{aligned}$$

Esempio. Se l'esperimento consiste nel lanciare successivamente n monete, lo spazio campionario è costituito da $D'_{2,n} = 2^n$ elementi:

$$S = \{\omega_1\omega_2\cdots\omega_n : \forall i,\ \omega_i \in \{c,t\}\}; \qquad (S \text{ è finito})$$

$$\text{per } n=3: \qquad S = \{ccc, cct, ctc, ctt, tcc, tct, ttc, ttt\}$$

Esempio. Un esperimento consiste nel lanciare ripetutamente una moneta. Consideriamo come esito dell'esperimento il numero d'ordine del lancio in cui compare testa per la prima volta. Lo spazio campionario è l'insieme degli interi non negativi:

$$S = \{n : n = 1, 2, \ldots\} \qquad (S \text{ è infinito numerabile})$$

Esempio. Se l'esperimento consiste nel misurare il tempo di vita di un'apparecchiatura elettronica, lo spazio campionario consiste nell'insieme dei numeri reali non negativi:

$$S = \{x \in \mathbb{R} : 0 \le x < \infty\} \qquad (S \text{ è infinito non numerabile})$$

Un sottoinsieme A dello spazio campionario sarà detto *evento*. Un evento è quindi un insieme di possibili esiti di un esperimento. Se l'esito di un esperimento è contenuto in A, diremo che l'evento A si è verificato.

Esempio. Nell'esperimento del lancio di 2 dadi, l'evento

$$A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

si verifica quando la somma dei 2 dadi è 7.

Esempio. Nell'esperimento del lancio di 3 monete, l'evento

$$A = \{ccc, cct, ctc, ctt\}$$

si verifica quando al primo lancio esce croce.

Esempio. Nell'esperimento che consiste nel misurare il tempo di vita di un'apparecchiatura elettronica, l'evento

$$A = \{x \in \mathbb{R} : 0 \leq x \leq 5\}$$

si verifica quando l'apparecchiatura non dura più di 5 ore.

Operazioni tra eventi

• Dati due eventi A e B, definiamo il nuovo evento $A \cup B$, detto *unione* di A e B, formato da tutti gli esiti dell'esperimento che stanno in A o in B o in entrambi.

• Analogamente, dati due eventi A e B, definiamo il nuovo evento $A \cap B$, detto *intersezione* di A e B, formato da tutti gli esiti dell'esperimento che sono sia in A che in B. (Talora $A \cap B$ si indica con AB).

• Per ogni evento A definiamo il nuovo evento $\overline{A}$, detto *complementare* di A, formato da tutti gli esiti dell'esperimento che non sono in A. (Talvolta $\overline{A}$ si indica con A^c).

Esempio. Nell'esperimento del lancio di 2 monete, con $S = \{cc, ct, tc, tt\}$, se

$$A = \{cc, ct\} = \{\text{croce al primo lancio}\},$$

$$B = \{cc, tt\} = \{\text{nei due lanci si ha lo stesso risultato}\},$$

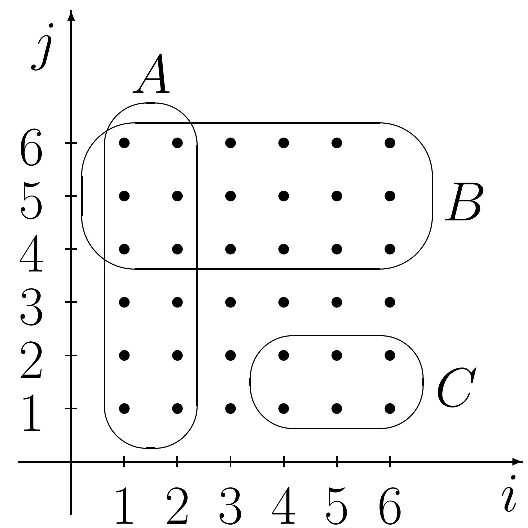
si ha

$$A \cup B = \{cc, ct, tt\}, \qquad A \cap B = \{cc\}, \qquad \overline{A} = \{tc, tt\}, \qquad \overline{B} = \{ct, tc\}.$$

- Il risultato di qualunque esperimento appartiene certamente allo spazio campione; pertanto S viene detto *evento certo*.
- Un evento si dice *impossibile*, e si indica con $\emptyset$, se non contiene esiti dell'esperimento. ($\emptyset$ corrisponde all'insieme vuoto).
- Due eventi A e B si dicono *incompatibili* se $A \cap B = \emptyset$.
- $A_1, A_2, \ldots$ si dicono *a due a due incompatibili* se $A_i \cap A_j = \emptyset$ per ogni $i \neq j$.
- Più eventi (in numero finito o infinito) si dicono *necessari* se la loro unione è S.

Esempio. Nell'esperimento del lancio di due dadi sia

$$A = \{(i, j): i = 1, 2;\ j = 1, 2, \ldots, 6\},$$
$$B = \{(i, j): i = 1, 2, \ldots, 6;\ j = 4, 5, 6\},$$
$$C = \{(i, j): i = 4, 5, 6;\ j = 1, 2\}.$$
$$A \cap B = \{(i, j): i = 1, 2;\ j = 4, 5, 6\} \neq \emptyset,$$
$$A \cap C = \emptyset, \quad B \cap C = \emptyset.$$



Esempio. Nell'esperimento dell'estrazione di una biglia da un'urna contenente n biglie numerate da 1 a n, sia $A_k = \{1, 2, \dots, k\}$, per $k = 1, 2, \dots, n$. Gli eventi $A_1, A_2, \dots, A_n$ sono necessari, ma non incompatibili, essendo

$$A_1 \cup A_2 \cup \cdots \cup A_n = \{1, 2, \dots, n\} \equiv S, \qquad A_i \cap A_j = A_{\min(i,j)} \neq \emptyset.$$

Esempio. Nell'esperimento che consiste nello scegliere a caso un vettore booleano di lunghezza n, indichiamo con A_k l'evento costituito dai vettori aventi esattamente k bit pari a $\mathbf{1}$, per $k = 0, 1, \dots, n$. Gli eventi $A_0, A_1, \dots, A_n$ sono necessari ed incompatibili. Infatti risulta

$$A_0 \cup A_1 \cup \cdots \cup A_n = S, \qquad A_i \cap A_j = \emptyset \quad \forall i \neq j,$$

essendo $S = \{(\omega_1, \omega_2, \dots, \omega_n) : \omega_i \in \{\mathbf{0}, \mathbf{1}\}\ \forall i\}$ e

$$A_k = \{(\omega_1, \omega_2, \dots, \omega_n) \in S : \sum_{i=1}^{n} I[\omega_i = \mathbf{1}] = k\},$$

dove $I[P] = 1$ se P è una proposizione vera, e $I[P] = 0$ altrimenti.

- Se $A_1, A_2, \ldots$ sono eventi, si definiscono l'unione e l'intersezione di questi come

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots, \qquad \bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \cdots;$$

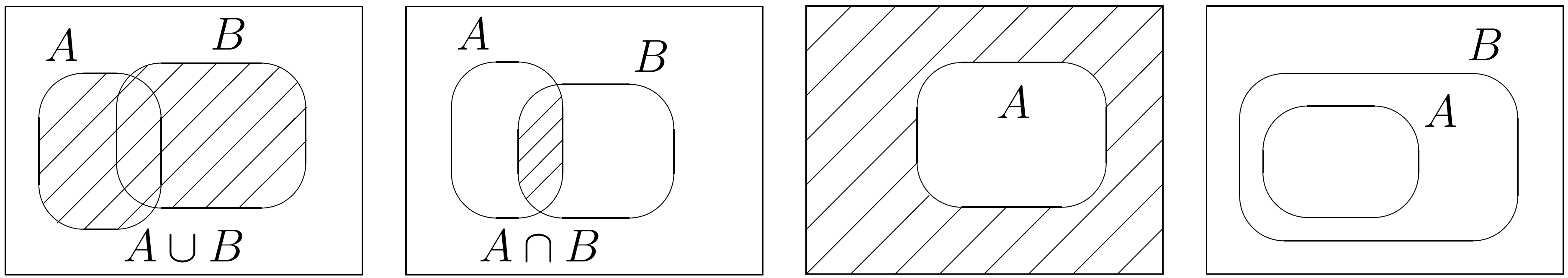
$\cup_{i=1}^{\infty} A_i$ è l'evento formato da tutti gli esiti che sono compresi in almeno uno degli eventi $A_1, A_2, \ldots$;
$\cap_{i=1}^{\infty} A_i$ è l'evento formato da tutti gli esiti che sono compresi in tutti gli eventi $A_1, A_2, \ldots$

- Per ogni evento A risulta

$$A \cup \overline{A} = S, \qquad A \cap \overline{A} = \emptyset, \qquad A \cup S = S, \qquad A \cap S = A.$$

- Dati due eventi A e B, se tutti gli esiti di A sono anche in B, allora diciamo che A è contenuto in B, oppure che A implica B, e scriviamo $A \subset B$.
- Se $A \subset B$ e $B \subset A$, diciamo che A e B coincidono, e scriviamo $A = B$.
- Risulta $A \subset B$ se e solo se $\overline{B} \subset \overline{A}$.

Diagrammi di Venn



$\overline{A}$ $\qquad$ $A \subset B$

Proprietà

- commutative:

$$A \cup B = B \cup A, \qquad A \cap B = B \cap A;$$

- associative:

$$(A \cup B) \cup C = A \cup (B \cup C), \qquad (A \cap B) \cap C = A \cap (B \cap C);$$

- distribuitive:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \qquad (A \cap B) \cup C = (A \cup C) \cap (B \cup C).$$

- formule di De Morgan:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}, \qquad \overline{A \cap B} = \overline{A} \cup \overline{B},$$

valide anche per un insieme finito di eventi $A_1, A_2, \ldots, A_n$:

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A}_i, \qquad \overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A}_i.$$

Esercizio. Siano A e B eventi distinti; stabilire se le seguenti identità sono vere:
(i) $A = (A \cap B) \cup (A \cap \overline{B})$;
(ii) $A \cap \overline{B} = \overline{B \cup \overline{A}}$;
(iii) $\overline{A} \cup (A \cap B) = \overline{B \cap \overline{A}}$;
(iv) $A \cup \overline{(A \cup B)} = A \cap \overline{B}$;
(v) $(A \cap B) \cup \overline{(A \cup B)} = (A \cup \overline{B}) \cap (B \cup \overline{A})$.
Soluzione. (i) vera; (ii) vera; (iii) falsa; (iv) falsa; (v) vera.

La classe degli eventi

Abbiamo visto in precedenza che un sottoinsieme A dello spazio campionario è detto evento. Più precisamente, la *classe degli eventi* $\mathcal{F}$ è una famiglia di sottoinsiemi di S tale che

(i) $S \in \mathcal{F}$;

(ii) se $A \in \mathcal{F}$ allora $\overline{A} \in \mathcal{F}$;

(iii) se $A_1, A_2, \ldots \in \mathcal{F}$ allora $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$.

Da tali proprietà segue che $\mathcal{F}$ è una σ-algebra (sigma-algebra) di eventi.

Inoltre risulta:

(iv) se $A_1, A_2, \ldots \in \mathcal{F}$ allora $A_1 \cap A_2 \cap \cdots \in \mathcal{F}$.

Esempio. Alcuni esempi di classi degli eventi:

- $\mathcal{F} = \{\emptyset, \mathcal{S}\}$;
- $\mathcal{F} = \{\emptyset, A, \overline{A}, S\}$;
- $\mathcal{F} = \mathcal{P}(S)$ (insieme delle parti di S), con $|S| = n$ e $|\mathcal{P}(S)| = 2^n$.

2.3 Assiomi della probabilità

La probabilità di un evento può essere definita in termini della frequenza relativa.

Supponiamo che un esperimento, il cui spazio campionario è S, venga ripetuto varie volte sotto le medesime condizioni. Per ogni evento E dello spazio campionario S, definiamo $n(E)$ come *frequenza assoluta*, ossia il numero di volte che si è verificato E nelle prime n ripetizioni dell'esperimento. Notiamo che risulta $0 \leq n(E) \leq n$. Allora $P(E)$, la *probabilità* dell'evento E, è definita come

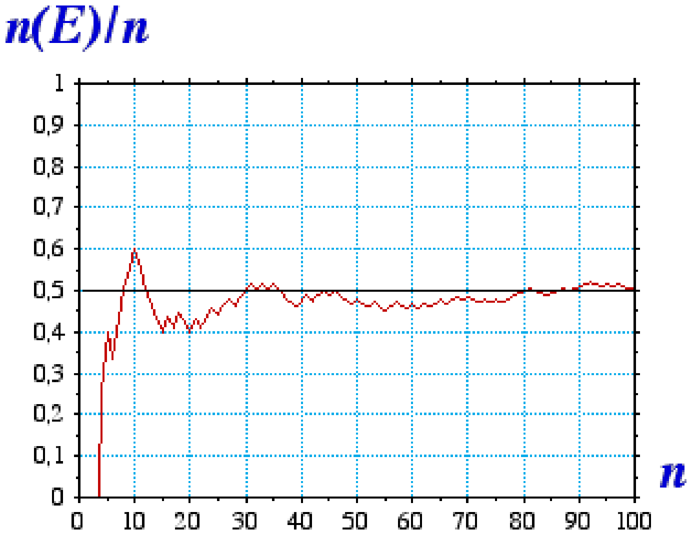
$$P(E) = \lim_{n\to\infty} \frac{n(E)}{n}$$

Cioè, $P(E)$ è definita come limite della *frequenza relativa* $n(E)/n$, ossia limite della proporzione del numero di volte che E si verifica. Notiamo che risulta $0 \leq n(E)/n \leq 1$.

Quindi $P(E)$ è la frequenza limite di E.

Esempio. Lancio di una moneta ripetuto 100 volte; $S = \{c, t\}$; $E = \{t\}$

n	ω_n	$n(E)$	$n(E)/n$
1	c	0	0
2	c	0	0
3	c	0	0
4	t	1	0,25
5	t	2	0,4
6	c	2	0,33
7	t	3	0,43
8	t	4	0,5
9	t	5	0,56
10	t	6	0,6



La definizione

$$P(E) = \lim_{n\to\infty} \frac{n(E)}{n}$$

presenta un serio inconveniente:

- come possiamo sapere se $n(E)/n$ converge al limite ad una costante, che sia la stessa per ogni possibile successione di ripetizioni dell'esperimento?

Per esempio, supponiamo che l'esperimento consista nel lancio di una moneta.

- Come possiamo sapere che la proporzione di teste ottenute nei primi n lanci converga ad un valore quando n diventa grande?
- Inoltre, anche sapendo che questa converga, come possiamo sapere che ripetendo un'altra volta la serie di esperimenti, otterremo ancora lo stesso limite nella proporzione di teste?

È più ragionevole assumere un insieme di *assiomi* più semplici ed intuitivi per definire la probabilità.

Consideriamo un esperimento il cui spazio campionario sia S. Per ogni evento A dello spazio S supponiamo che esista un numero reale $P(A)$, definito come ***probabilità*** di A, con $P : \mathcal{F} \to \mathbb{R}$, il quale soddisfi i seguenti 3 assiomi.

Assioma 1.
Per ogni $A \in \mathcal{F}$ si ha

$$0 \leq P(A) \leq 1$$

Assioma 2.

$$P(S) = 1$$

Assioma 3. (Additività numerabile)
Per ogni successione di eventi $A_1, A_2, \ldots$ a due a due incompatibili (ossia tali che $A_i \cap A_j = \emptyset$ quando $i \neq j$), si ha

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

Proposizione.

$$P(\emptyset) = 0$$

Dimostrazione. Consideriamo una successione di eventi $A_1, A_2, \ldots$, dove $A_1 = S$, $A_2 = A_3 = \ldots = \emptyset$; allora, essendo gli eventi $A_1, A_2, \ldots$ a due a due incompatibili ed essendo

$$S = \bigcup_{i=1}^{\infty} A_i$$

dall'Assioma 3 ricaviamo

$$P(S) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = P(S) + \sum_{i=2}^{\infty} P(\emptyset)$$

il che implica $\sum_{i=2}^{\infty} P(\emptyset) = 0$ e quindi

$$P(\emptyset) = 0$$

perciò l'evento impossibile ha probabilità 0 di verificarsi.

Proposizione. (Additività finita) Per ogni collezione finita $A_1, A_2, \ldots, A_n$ di eventi a due a due incompatibili (ossia tali che $A_i \cap A_j = \emptyset$ quando $i \neq j$),

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

Dimostrazione. Consideriamo la successione $A_1, A_2, \ldots, A_n, \emptyset, \emptyset, \ldots$; essendo gli eventi di tale successione a due a due incompatibili, dall'Assioma 3 segue che

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\left(\bigcup_{i=1}^{n} A_i\right) \cup \emptyset \cup \emptyset \cup \ldots\right) = \sum_{i=1}^{n} P(A_i) + P(\emptyset) + P(\emptyset) + \ldots$$

Ricordando che $P(\emptyset) = 0$, si ottiene

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

da cui segue la tesi.

Esempio. Nell'esperimento del lancio di una moneta si ha $S = \{t, c\}$; occorre fissare 2 reali p_1 e p_2, in modo che

$$P(\{t\}) = p_1, \qquad P(\{c\}) = p_2.$$

Dall'Assioma 1 segue

$$0 \leq p_1 \leq 1, \qquad 0 \leq p_2 \leq 1.$$

Dall'Assioma 2 e dalla proprietà di additività finita si ha

$$1 = P(S) = P(\{t\} \cup \{c\}) = P(\{t\}) + P(\{c\}) = p_1 + p_2.$$

Ad esempio, se testa e croce si verificano con uguale probabilità avremo

$$P(\{t\}) = \frac{1}{2}, \qquad P(\{c\}) = \frac{1}{2};$$

se invece abbiamo la sensazione che la moneta sia truccata in modo da mostrare testa con probabilità doppia rispetto a quella di croce, avremo

$$P(\{t\}) = \frac{2}{3}, \qquad P(\{c\}) = \frac{1}{3}.$$

Esempio. Nel lancio di un dado, con $S = \{1, 2, 3, 4, 5, 6\}$, se supponiamo che le sei facce siano equiprobabili, allora dovremo porre

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}.$$

Dalla proprietà di additività finita segue che la probabilità che in lancio del dado si ottenga un numero pari vale

$$P(\{2, 4, 6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}.$$

Più in generale, se il dado ha n facce, con $S = \{1, 2, \dots, n\}$, e se si suppone che le n facce siano equiprobabili, allora avremo

$$P(\{1\}) = P(\{2\}) = \ldots = P(\{n\}) = \frac{1}{n}.$$

La probabilità che in un lancio del dado si ottenga un numero compreso tra 1 e k è

$$P(\{1, 2, \dots, k\}) = P(\{1\}) + P(\{2\}) + \ldots + P(\{k\}) = \frac{k}{n}, \qquad (1 \le k \le n).$$

2.4 Alcune semplici proprietà

Proposizione. Per ogni evento A

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

Dimostrazione. Dall'Assioma 2 e dalla proprietà di additività finita, con A e $\overline{A}$ eventi incompatibili, segue

$$1 = P(S) = P(A \cup \overline{A}) = P(A) + P(\overline{A}),$$

da cui si giunge alla tesi.

Esempio. Se la probabilità di ottenere 2 volte testa lanciando due monete è $1/4$, allora la probabilità di ottenere al più una volta testa deve essere $3/4$.

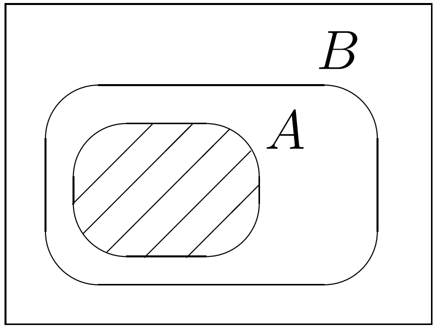
Esempio. Se la probabilità di ottenere n volte croce lanciando n monete è $1/2^n$, allora la probabilità di ottenere almeno una volta testa deve essere $1 - 1/2^n$.

Proposizione. Se $A \subset B$, allora

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

Dimostrazione. Essendo $A \subset B$, abbiamo che B può essere espresso come $B = A \cup (\overline{A} \cap B)$, con $A \cap (\overline{A} \cap B) = \emptyset$.

Dalla proprietà di additività finita segue
$P(B) = P(A \cup (\overline{A} \cap B)) = P(A) + P(\overline{A} \cap B)$,
da cui si ha $P(B) \geq P(A)$, essendo $P(\overline{A} \cap B) \geq 0$.



$A \subset B$

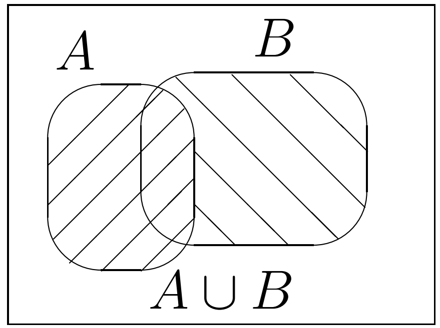
Esempio. Nel lancio di un dado, la probabilità che si ottenga 1 è minore della probabilità che si ottenga un numero dispari. Infatti, $A = \{1\} \subset \{1, 3, 5\} = B$.

Esempio. Nell'esperimento consistente del lancio di 3 monete, la probabilità che si ottenga 1 testa è minore della probabilità che si ottenga almeno 1 testa, avendosi $A = \{cct, ctc, tcc\} \subset \{cct, ctc, ctt, tcc, tct, ttc, ttt\} = B$.

Proposizione.

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

Dimostrazione. Notiamo che $A \cup B$ può essere espresso come unione di due eventi incompatibili A e $\overline{A} \cap B$. Grazie alla proprietà di additività finita otteniamo



$$P(A \cup B) = P(A \cup (\overline{A} \cap B)) = P(A) + P(\overline{A} \cap B).$$

Inoltre, essendo $B = (A \cap B) \cup (\overline{A} \cap B)$, con $A \cap B$ e $\overline{A} \cap B$ eventi incompatibili, applicando nuovamente la proprietà di additività finita abbiamo

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

o, equivalentemente, che

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

che completa la dimostrazione.

Esempio. Uno studente deve sottoporsi a due test. Con probabilità 0,5 supererà il primo test; con probabilità 0,4 supererà il secondo test; con probabilità 0,3 li supererà entrambi. Quanto vale la probabilità che non supererà nessuno dei due test?

Soluzione. Sia B_i l'evento che lo studente superi il test i-esimo, $i = 1, 2$. La probabilità che superi almeno un test sarà quindi pari a

$$P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1 \cap B_2) = 0{,}5 + 0{,}4 - 0{,}3 = 0{,}6.$$

La probabilità che lo studente non supererà nessuno dei due test è dunque

$$P(\overline{B_1} \cap \overline{B_2}) = P(\overline{B_1 \cup B_2}) = 1 - P(B_1 \cup B_2) = 1 - 0{,}6 = 0{,}4.$$

Esercizio. Dimostrare le seguenti relazioni, per eventi A_1 e A_2 qualsiasi:

$$P(A_1 \cup A_2) \leq P(A_1) + P(A_2), \qquad P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1.$$

Proposizione.

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

Dimostrazione. Ricordando che $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$, si ha

$$P(A \cup B \cup C) = P((A \cup B) \cup C) = P(A \cup B) + P(C) - P((A \cup B) \cap C),$$

e ancora

$$P(A \cup B \cup C) = P(A) + P(B) - P(A \cap B) + P(C) - P((A \cup B) \cap C).$$

Per la legge distributiva si ha

$$P(A \cup B \cup C) = P(A) + P(B) - P(A \cap B) + P(C) - P((A \cap C) \cup (B \cap C)),$$

da cui segue

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

ossia la tesi.

Proposizione. Principio di inclusione/esclusione

La probabilità dell'unione di n eventi può esprimersi al seguente modo:

$$\begin{aligned} P(A_1 \cup A_2 \cup \ldots \cup A_n) &= \sum_{i=1}^{n} P(A_i) - \sum_{i_1<i_2} P(A_{i_1} \cap A_{i_2}) + \ldots \\ &+ (-1)^{r+1} \sum_{i_1<i_2<\ldots<i_r} P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_r}) \\ &+ \ldots + (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n). \end{aligned}$$

Per $n = 2$: $P(A_1 \cup A_2) = \sum_{i=1}^{2} P(A_i) - P(A_1 \cap A_2)$.

Per $n = 3$: $P(A_1 \cup A_2 \cup A_3) = \sum_{i=1}^{3} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + P(A_1 \cap A_2 \cap A_3)$.

La seguente formula esprime in modo compatto il principio di inclusione/esclusione:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<i_2<\ldots<i_r} P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_r}).$$

La sommatoria $\sum_{i_1<i_2<\ldots<i_r} P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_r})$ è calcolata per tutti gli $\binom{n}{r}$ possibili sottoinsiemi di dimensione r dell'insieme $\{1, 2, \ldots, n\}$.

Il numero di termini a 2° membro nella formula del principio di inclusione/esclusione è

$$\sum_{r=1}^{n} \binom{n}{r} = 2^n - 1.$$

$2^n - 1 = 3$ per $n = 2$: $\sum_{i=1}^{2} P(A_i) - P(A_1 \cap A_2)$.

$2^n - 1 = 7$ per $n = 3$: $\sum_{i=1}^{3} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + P(A_1 \cap A_2 \cap A_3)$.

2.5 Spazi campionari con esiti equiprobabili

In molti esperimenti è naturale assumere che tutti gli esiti dello spazio campionario siano equiprobabili, con S insieme finito: $S = \{1, 2, \ldots, N\}$. Allora si ipotizza che

$$P(\{1\}) = P(\{2\}) = \ldots = P(\{N\})$$

il che implica

$$P(\{i\}) = \frac{1}{N}, \qquad i = 1, 2, \ldots, N,$$

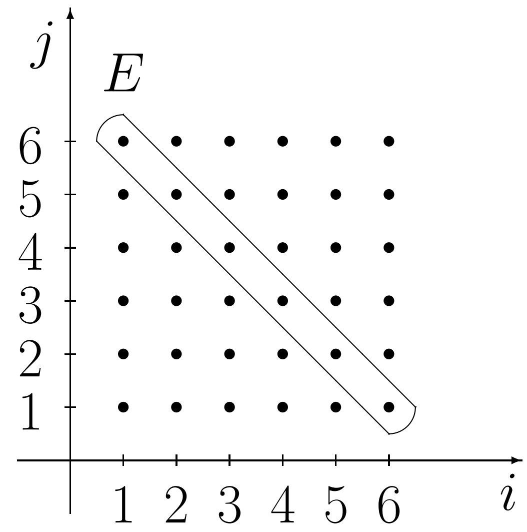
essendo $1 = P(S) = P(\{1\} \cup \{2\} \cup \ldots \cup \{N\}) = P(\{1\}) + P(\{2\}) + \ldots + P(\{N\})$.
Per la proprietà di additività avremo perciò che per ogni evento A

$$P(A) = \frac{|A|}{|S|} = \frac{\text{numero di elementi di } A}{\text{numero di elementi di } S} \qquad \text{(definizione classica di probabilità)}$$

Se assumiamo che tutti gli esiti di un esperimento siano equiprobabili, allora la probabilità di ogni evento A è uguale alla proporzione degli esiti dello spazio campionario contenuti in A (come rapporto di casi favorevoli su casi possibili).

Esempio. Se si lanciano 2 dadi, qual è la probabilità che la somma dei valori sulla faccia superiore sia uguale a 7?

Soluzione. Assumendo che i 36 possibili esiti siano equiprobabili, poiché ci sono 6 possibili esiti che danno come somma 7,
$E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$,
la probabilità desiderata sarà uguale a 6/36 ossia 1/6.



Esempio. Se estraiamo 3 biglie a caso da un'urna che contiene 6 biglie bianche e 5 nere, qual è la probabilità che una sia bianca e le altre due nere?

Soluzione. Se teniamo conto dell'ordine di estrazione e supponiamo che ogni possibile esito dello spazio campionario sia equiprobabile, la probabilità richiesta è

$$\frac{(6\cdot 5\cdot 4)+(5\cdot 6\cdot 4)+(5\cdot 4\cdot 6)}{11\cdot 10\cdot 9}=\frac{3\cdot 120}{990}=\frac{4}{11}=0{,}3636$$

Esempio. Se estraiamo 3 biglie a caso da un'urna che contiene 6 biglie bianche e 5 nere, qual è la probabilità che una sia bianca e le altre due nere?
Soluzione. Se non teniamo conto dell'ordine di estrazione, la probabilità richiesta è

$$\frac{\binom{6}{1}\binom{5}{2}}{\binom{11}{3}} = \frac{6 \cdot 10}{165} = \frac{4}{11} = 0{,}3636$$

Esempio. Una commissione di 5 persone viene estratta da un gruppo composto di 6 uomini e 9 donne. Se la selezione avviene in modo casuale, qual è la probabilità che la commissione consti di 3 uomini e 2 donne?
Soluzione. Ognuna delle $\binom{15}{5}$ combinazioni è estratta in modo equiprobabile, quindi:

$$\frac{\binom{6}{3}\binom{9}{2}}{\binom{15}{5}} = \frac{20 \cdot 36}{3003} = \frac{240}{1001} = 0{,}2398$$

Esempio. Un'urna contiene n biglie, di cui una è speciale. Se estraiamo k biglie una alla volta, in modo tale che a ogni estrazione la probabilità di estrarre una qualunque delle biglie rimanenti sia la stessa, qual è la probabilità che la biglia speciale sia estratta?
Soluzione. Ognuno dei $\binom{n}{k}$ possibili insiemi è estratto in modo equiprobabile, quindi:

$$P(\text{si estrae la biglia speciale}) = \frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{\dfrac{(n-1)!}{(k-1)!\,(n-k)!}}{\dfrac{n!}{k!\,(n-k)!}} = \frac{k}{n}$$

In modo alternativo, sia $A_i = \{\text{la biglia speciale si estrae nell'}i\text{-esima estrazione}\}$, $i = 1, 2, \ldots, k$. Risulta $P(A_i) = 1/n$ perché ognuna delle n biglie ha uguale probabilità di essere scelta all'i-esima estrazione. Pertanto, poiché gli eventi A_i sono incompatibili si ha

$$P(\text{si estrae la biglia speciale}) = P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i) = \frac{k}{n}$$

Esempio. Supponiamo che $n+m$ biglie, di cui n siano rosse e m blu, vengano disposte in fila in modo casuale, così che ognuna delle $(n+m)!$ possibili disposizioni sia equiprobabile. Se siamo interessati solo alla successione dei colori delle biglie in fila, si provi che tutti i possibili risultati rimangono equiprobabili.

Soluzione. Consideriamo ognuna delle $(n+m)!$ possibili disposizioni ordinate delle biglie: se permutiamo tra loro le biglie rosse e facciamo lo stesso con quelle blu, la successione dei colori chiaramente non cambia. Come risultato, otteniamo che ogni successione dei colori corrisponde a $n!\,m!$ differenti disposizioni ordinate delle $n+m$ biglie, così che ogni distribuzione di colori ha probabilità di verificarsi

$$\frac{n!\,m!}{(n+m)!}$$

Ad esempio se $n=m=2$, delle $4!=24$ possibili disposizioni ordinate delle biglie, ci saranno $2!\,2!=4$ distribuzioni che danno la stessa distribuzione di colori; ad esempio

$$r_1, b_1, r_2, b_2, \qquad r_1, b_2, r_2, b_1, \qquad r_2, b_1, r_1, b_2, \qquad r_2, b_2, r_1, b_1.$$

Esempio. Nel gioco del lotto si estraggono a caso 5 numeri da un'urna contenente 90 numeri compresi tra 1 e 90. (Le cinquine sono equiprobabili). Se si fissano k numeri distinti compresi tra 1 e 90 qual è la probabilità che questi saranno tra i 5 estratti?

Soluzione. Poiché l'ordine di estrazione qui non è rilevante, lo spazio campionario è l'insieme delle $\binom{90}{5}$ combinazioni semplici di 90 oggetti in 5 classi.

L'evento $A_k = \{$i k numeri fissati sono tra i 5 estratti$\}$ è costituito dalle sequenze di S che presentano all'interno i k numeri fissati; pertanto la sua cardinalità è pari al numero di combinazioni semplici di $90-k$ oggetti in $5-k$ classi: $\binom{90-k}{5-k}$. Si ha quindi:

$P(A_k)$	$P(A_1)$	$P(A_2)$	$P(A_3)$	$P(A_4)$	$P(A_5)$
$\dfrac{\binom{90-k}{5-k}}{\binom{90}{5}}$	$\dfrac{\binom{89}{4}}{\binom{90}{5}} = \dfrac{1}{18}$	$\dfrac{\binom{88}{3}}{\binom{90}{5}} = \dfrac{2}{801}$	$\dfrac{\binom{87}{2}}{\binom{90}{5}} = \dfrac{1}{11.748}$	$\dfrac{\binom{86}{1}}{\binom{90}{5}} = \dfrac{1}{511.038}$	$\dfrac{\binom{85}{0}}{\binom{90}{5}} = \dfrac{1}{43.949.268}$

Esempio. Il problema delle concordanze. Da un'urna che contiene biglie numerate da 1 a n si estraggono casualmente in sequenza le n biglie. Diciamo che si ha una concordanza nell'i-esima estrazione se in tale estrazione viene estratta la biglia avente numero i, per $i = 1, 2, \ldots, n$. Calcolare la probabilità che non si abbia concordanza in nessuna estrazione.

Soluzione. Per $i = 1, 2, \ldots, n$ sia $E_i = \{$si ha concordanza nell'i-esima estrazione$\}$. Pertanto, la probabilità che si abbia almeno una concordanza è

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}).$$

Rappresentiamo l'esito dell'esperimento come un vettore $(\omega_1, \omega_2, \ldots, \omega_n)$, dove ω_i denota il numero estratto all'estrazione i-esima. Lo spazio campionario è quindi l'insieme delle $n!$ permutazioni

$$S = \{(\omega_1, \omega_2, \ldots, \omega_n) \colon \omega_i \in \{1, 2, \ldots, n\},\ \omega_i \neq \omega_j \text{ per } i \neq j\}.$$

L'evento $E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}$ si verifica se nelle estrazioni $i_1, i_2, \ldots, i_r$ si ha concordanza (e non è escluso che si possa avere concordanza nelle altre estrazioni). Ciò può accadere in $(n-r)!$ modi, perché nelle estrazioni $i_1, i_2, \ldots, i_r$ l'esito deve essere una concordanza, mentre nelle estrazioni rimanenti ci sono $(n-r)\,(n-r-1)\;\ldots\;2\cdot 1$ modi per individuare gli esiti dell'esperimento. Si ha pertanto

$$P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}) = \frac{(n-r)!}{n!}$$

Poiché vi sono $\binom{n}{r}$ termini in $\displaystyle\sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r})$, risulta quindi

$$\sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}) = \binom{n}{r} \frac{(n-r)!}{n!} = \frac{n!}{r!\,(n-r)!} \cdot \frac{(n-r)!}{n!} = \frac{1}{r!}$$

e dunque

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}) = \sum_{r=1}^{n} (-1)^{r+1} \frac{1}{r!}.$$

La probabilità che non si abbia concordanza in nessuna estrazione è quindi

$$p_n = P\left(\overline{\bigcup_{i=1}^{n} E_i}\right) = 1 - P\left(\bigcup_{i=1}^{n} E_i\right) = 1 - \sum_{r=1}^{n} \frac{(-1)^{r+1}}{r!} = 1 + \sum_{r=1}^{n} \frac{(-1)^{r}}{r!} = \sum_{r=0}^{n} \frac{(-1)^{r}}{r!}$$

n	p_n
1	0
2	0,5
3	0,333333
4	0,375
5	0,366667
6	0,368056
7	0,367857
8	0,367882
9	0,367879
10	0,367879

Notiamo che per n grande la probabilità

$$p_n = \sum_{r=0}^{n} \frac{(-1)^{r}}{r!}$$

può essere approssimata da

$$\mathrm{e}^{-1} = \sum_{r=0}^{+\infty} \frac{(-1)^{r}}{r!} \approx 0{,}367879$$

(Ricordiamo che $\boxed{\mathrm{e}^{x} = \sum_{r=0}^{+\infty} \frac{x^{r}}{r!}}$ per $x \in \mathbb{R}$)

Esercizio. Nell'esperimento che consiste nel lanciare n volte una moneta non truccata, indicando con t la fuoriuscita di testa e con c di croce, lo spazio campionario è

$$S = \{(\omega_1, \omega_2, \ldots, \omega_n) \colon \omega_i \in \{t, c\} \text{ per } 1 \le i \le n\},$$

e le sue 2^n sequenze sono equiprobabili. Per $1 \le k \le n$, calcolare le probabilità di

$T_k = \{\text{al } k\text{-esimo lancio esce testa}\} = \{\underline{\omega} \colon \omega_k = t; \omega_i \in \{t, c\} \text{ per } i \neq k\}$

$A_k = \{\text{nei primi } k \text{ lanci esce testa}\} = \{\underline{\omega} \colon \omega_1 = \ldots = \omega_k = t; \omega_i \in \{t, c\} \text{ per } i > k\}$

$A'_k = \{\text{nei primi } k \text{ lanci esce croce}\} = \{\underline{\omega} \colon \omega_1 = \ldots = \omega_k = c; \omega_i \in \{t, c\} \text{ per } i > k\}$

$B_k = \{\text{esce testa } k \text{ volte}\} = \{\underline{\omega} \in S \colon \sum_{i=1}^{n} \mathbf{1}_{\{\omega_i = t\}} = k\}$ ($\mathbf{1}_A = 1$ se A è vera, 0 senò).

Soluzione. Trattandosi di uno spazio campionario con esiti equiprobabili, si ha

$$P(T_k) = \frac{2^{n-1}}{2^n} = \frac{1}{2}, \qquad P(A_k) = P(A'_k) = \frac{2^{n-k}}{2^n} = \frac{1}{2^k}, \qquad P(B_k) = \frac{\binom{n}{k}}{2^n}.$$

Esercizio. Riferendosi all'esercizio precedente, mostrare che gli eventi $B_0, B_1, \ldots, B_n$ sono necessari e incompatibili, ed inoltre che

$$P(T_1 \cap A_k) = \frac{1}{2^k}, \qquad P(T_1 \cup A_k) = \frac{1}{2},$$

$$P(T_1 \cap B_k) = \frac{\binom{n-1}{k-1}}{2^n}, \qquad P(T_1 \cup B_k) = \frac{1}{2} + \frac{\binom{n-1}{k}}{2^n},$$

$$P(A_k \cap A'_k) = 0, \qquad P(A_k \cup A'_k) = \frac{1}{2^{k-1}},$$

$$P(A_k \cap B_k) = \frac{1}{2^n}, \qquad P(A_k \cup B_k) = \frac{1}{2^k} + \frac{\binom{n}{k}}{2^n} - \frac{1}{2^n},$$

$$P(A'_{k-1} \cup T_k) = \frac{1}{2} + \frac{1}{2^k}, \qquad P(A'_{k-1} \cap T_k) = \frac{1}{2^k}.$$

2.6 La probabilità come funzione di insieme continua

Una successione di eventi $\{E_n, n \geq 1\}$ è detta *successione crescente* se

$$E_1 \subset E_2 \subset \ldots \subset E_n \subset E_{n+1} \subset \ldots$$

mentre è detta *successione decrescente* se

$$E_1 \supset E_2 \supset \ldots \supset E_n \supset E_{n+1} \supset \ldots$$

Se $\{E_n, n \geq 1\}$ è una successione crescente di eventi, allora definiamo un nuovo evento, che denotiamo con $\lim\limits_{n\to\infty} E_n$, nel seguente modo

$$\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$$

Similmente, se $\{E_n, n \geq 1\}$ è una successione decrescente di eventi, definiamo $\lim\limits_{n\to\infty} E_n$ come

$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

Esempio. Nel lancio di una moneta non truccata, ripetuto indefinitamente, si consideri la successione di eventi $A_n = \{\text{nei primi } n \text{ lanci esce testa}\}$, $n = 1, 2, \ldots$. È facile vedere che

$$A_1 \supset A_2 \supset A_3 \supset \ldots ;$$

pertanto $\{A_n, n \geq 1\}$ è una successione decrescente di eventi.

Se poniamo $C_n = \{\text{esce almeno una volta testa nei primi } n \text{ lanci}\}$, si ha

$$C_1 \subset C_2 \subset C_3 \subset \ldots ;$$

segue che $\{C_n, n \geq 1\}$ è una successione crescente di eventi.

Indicando con T_n l'evento che si verifica quando esce testa al lancio n-esimo, è facile verificare che $\{T_n, n \geq 1\}$ non è una successione monotona.

Proposizione. (Continuità della probabilità) Se $\{E_n, n \geq 1\}$ è una successione di eventi, crescente o decrescente, allora

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n).$$

Esempio. Nel lancio di una moneta non truccata ripetuto indefinitamente calcolare

$$P(E^{\mathrm{p}}) = P(\{\text{esce testa per la prima volta in un lancio pari}\}),$$

$$P(E^{\mathrm{d}}) = P(\{\text{esce testa per la prima volta in un lancio dispari}\}).$$

Soluzione. Consideriamo la successione di eventi

$$F_k = \{\text{esce testa per la prima volta nel lancio } k\text{-esimo}\}, \qquad k \geq 1.$$

Gli eventi $F_1, F_2, \ldots$ sono incompatibili e, per quanto visto in precedenza, si ha

$$\begin{aligned} P(F_1) &= P(T_1) = \frac{1}{2}, \\ P(F_k) &= P(\overline{T_1} \cap \overline{T_2} \cap \cdots \cap \overline{T_{k-1}} \cap T_k) = P(A'_{k-1} \cap T_k) = \frac{1}{2^k}, \qquad k \geq 2. \end{aligned}$$

Notiamo che

$$
\begin{aligned}
E^{\mathrm{p}} &= \bigcup_{k=1}^{\infty} F_{2k} = \{\text{esce testa per la prima volta in un lancio pari}\},\\
E^{\mathrm{d}} &= \bigcup_{k=1}^{\infty} F_{2k-1} = \{\text{esce testa per la prima volta in un lancio dispari}\}.
\end{aligned}
$$

Dalla proprietà di additività numerabile e ricordando che $P\left(F_k\right) = \dfrac{1}{2^k}$ segue:

$$
P(E^{\mathrm{p}}) = P\left(\bigcup_{k=1}^{+\infty} F_{2k}\right) = \sum_{k=1}^{+\infty} P\left(F_{2k}\right) = \sum_{k=1}^{+\infty} \frac{1}{2^{2k}} = \sum_{k=1}^{+\infty} \left(\frac{1}{4}\right)^k,
$$

$$
P(E^{\mathrm{d}}) = P\left(\bigcup_{k=1}^{+\infty} F_{2k-1}\right) = \sum_{k=1}^{+\infty} P\left(F_{2k-1}\right) = \sum_{k=1}^{+\infty} \frac{1}{2^{2k-1}} = 2\sum_{k=1}^{+\infty} \left(\frac{1}{4}\right)^k.
$$

Poiché $\boxed{\displaystyle\sum_{k=1}^{\infty} x^k = \frac{x}{1-x} \qquad (-1 < x < 1)}$ si ha $P(E^{\mathrm{p}}) = \dfrac{\frac{1}{4}}{1-\frac{1}{4}} = \dfrac{1}{3}$ e $P(E^{\mathrm{d}}) = \dfrac{2}{3}$.

È utile sottolineare che sebbene sia

$$P(E^{\mathrm{p}}) + P(E^{\mathrm{d}}) = 1, \qquad E^{\mathrm{p}} \cap E^{\mathrm{d}} = \emptyset,$$

gli eventi E^{p} ed E^{d} non sono tra loro complementari. Infatti introducendo l'evento

$$E_0 = \{\text{non esce mai testa}\},$$

si ha

$$E^{\mathrm{p}} \cup E^{\mathrm{d}} \cup E_0 = S, \qquad \text{con } E^{\mathrm{p}}, E^{\mathrm{d}}, E_0 \text{ incompatibili.}$$

Inoltre, ponendo $A'_k = \{\text{nei primi } k \text{ lanci esce sempre croce}\}$ per $k \geq 1$, la successione $\{A'_1, A'_2, \ldots\}$ è decrescente ed ha come limite l'evento

$$\lim_{k\to\infty} A'_k = \bigcap_{k=1}^{\infty} A'_k = E_0.$$

Pertanto, per la continuità della probabilità, e notando che $P(A'_k) = \dfrac{1}{2^k}$, si ha

$$P(E_0) = P\left(\lim_{k\to\infty} A'_k\right) = \lim_{k\to\infty} P(A'_k) = \lim_{k\to\infty} \frac{1}{2^k} = 0.$$

2.7 La probabilità come misura della fiducia

Secondo l'impostazione soggettiva la probabilità di un evento è il *grado di fiducia che un individuo ha nel verificarsi dell'evento*.

In questa affermazione la probabilità perde la caratteristica "assoluta" di numero legato intrinsecamente all'evento per dipendere dall'opinione soggettiva dell'osservatore.

Esprimendosi in termini di scommesse la probabilità è *il prezzo che un individuo ritiene equo pagare per ricevere* 1 *se l'evento si verifica e* 0 *se l'evento non si verifica.*

ω = risultato dell'esperimento

$$\begin{cases} \omega \in A & \Rightarrow \quad \text{riceviamo } 1 \\ \omega \in \overline{A} & \Rightarrow \quad \text{riceviamo } 0 \end{cases}$$

Va inoltre imposta la seguente condizione di coerenza:

Le probabilità degli eventi vanno attribuite in modo che non sia possibile ottenere con un insieme di scommesse una vincita certa o una perdita certa.

Chiariamo il ruolo della condizione di coerenza per la probabilità soggettiva.

Sia $P(A)$ la probabilità di un evento A secondo l'impostazione soggettiva. Nel pagare $P(A)$ e nel ricevere 1 oppure 0 si guadagna $1 - P(A)$ oppure $-P(A)$, quindi almeno $-P(A)$ e al massimo $1 - P(A)$. Se $P(A)$ fosse negativa si avrebbe certamente un guadagno positivo, mentre se $P(A)$ fosse maggiore di 1 si avrebbe certamente una perdita, e nei due casi la condizione di coerenza è violata. Si ha quindi $0 \leq P(A) \leq 1$.

Se consideriamo una scommessa su S, paghiamo $P(S)$ per ricevere certamente 1; si ha quindi certamente un guadagno pari a $1 - P(S)$. Se fosse $P(S) < 1$ si avrebbe una vincita certa mentre se fosse $PS) > 1$ si avrebbe una perdita certa. Per la condizione di coerenza si ricava quindi $P(S) = 1$.

L'impostazione soggettiva alla probabilità è sottoposta a varie critiche. Lo schema di scommesse che la descrive è infatti contestabile perché la propensione o l'avversità al rischio cambia da persona a persona e per ogni individuo può variare con le circostanze connesse all'esperimento.

CAPITOLO 3 – Probabilità condizionata e indipendenza

3.1 Introduzione

3.2 Probabilità condizionata

3.3 La formula delle alternative e la formula di Bayes

3.4 Eventi indipendenti

3.5 $P(\cdot|F)$ è una probabilità

3.1 Introduzione

In questo capitolo introduciamo la

probabilità condizionata,

uno dei concetti più importanti della teoria della probabilità.

L'importanza del concetto è duplice.

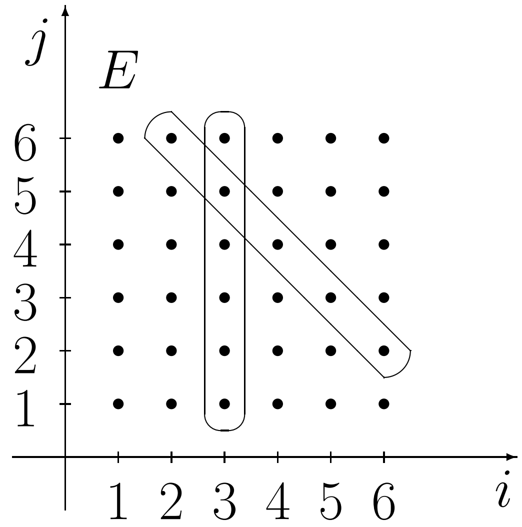
Innanzitutto si è spesso interessati a calcolare la probabilità di un evento disponendo di qualche informazione parziale.

Inoltre le probabilità condizionate sono spesso utilizzate per calcolare più facilmente le probabilità richieste.

3.1 Probabilità condizionata

Supponiamo di lanciare 2 dadi e che ognuno dei 36 possibili esiti sia *equiprobabile* con probabilità $\frac{1}{36}$. Supponiamo inoltre di osservare che l'esito del primo lancio è 3. Allora, data questa informazione, qual è la probabilità che la somma dei due dadi sia uguale a 8? Dato che il primo dado vale 3, vi sono al più 6 possibili esiti dell'esperimento: $(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)$.

Dato che all'inizio questi esiti erano equiprobabili, essi devono avere anche ora la stessa probabilità. Sapendo che il primo dado vale 3, la probabilità (condizionata) di ognuno dei 6 esiti è uguale a $\frac{1}{36}$, mentre la probabilità (condizionata) degli altri 30 punti dello spazio campionario è pari a 0. Pertanto la probabilità richiesta vale $\frac{1}{6}$, perché uno solo dei 6 risultati dà somma 8: $(3, 5)$.



In assenza d'informazione sul primo lancio la probabilità di avere somma 8 vale $\frac{5}{36}$.

Definizione. Se $P(F) > 0$, la probabilità condizionata di E dato F è data da

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Tale definizione è giustificata dalle seguenti considerazioni:

Per esperimenti dotati di spazio campionario finito e con esiti equiprobabili, abbiamo visto che per ogni evento A risulta: $P(A) = |A|/|S|$.

Pertanto, volendo esprimere in tale ambito la probabilità condizionata di E dato F, siamo condotti ad usare il rapporto di casi favorevoli al verificarsi di E (sapendo che si è verificato F) su casi possibili (gli elementi di F), cosicché:

$$P(E|F) = \frac{|E \cap F|}{|F|} = \frac{|E \cap F|/|S|}{|F|/|S|} = \frac{P(E \cap F)}{P(F)}.$$

Esempio. Nell'esperimento del lancio di un dado non truccato calcolare le probabilità condizionate di $A = \{1, 2\}$ dati gli eventi $B_1 = \{4, 5, 6\}$, $B_2 = \{1, 5, 6\}$, $B_3 = \{1, 2, 6\}$.

Soluzione. Risulta

$$P(A|B_1) = \frac{P(A \cap B_1)}{P(B_1)} = \frac{P(\emptyset)}{1/2} = 0,$$

$$P(A|B_2) = \frac{P(A \cap B_2)}{P(B_2)} = \frac{P(\{1\})}{1/2} = \frac{1/6}{1/2} = \frac{1}{3},$$

$$P(A|B_3) = \frac{P(A \cap B_3)}{P(B_3)} = \frac{P(\{1,2\})}{1/2} = \frac{1/3}{1/2} = \frac{2}{3}.$$

Pertanto, sebbene gli eventi B_1, B_2, B_3 siano equiprobabili, la probabilità condizionata di A dato B_k cambia al variare di k, ed in particolare risulta $P(A|B_2) = P(A)$.

Esempio. Uno studente sta svolgendo un test da consegnare entro un'ora. Si suppone che la probabilità che lo studente termini l'esame in meno di x ore sia uguale a $x/2$, per ogni $x \in [0,1]$. Sapendo che lo studente è ancora al lavoro dopo 3/4 d'ora, qual è la probabilità condizionata che egli usufruisca dell'intera ora?

Soluzione. Sia $L_x = \{$lo studente conclude l'esame in meno di x ore$\}$, $0 \leq x \leq 1$, e sia $F = \{$lo studente utilizza l'intera ora$\}$. Si ha $P(L_x) = x/2$, per $0 \leq x \leq 1$. Inoltre, risulta $F = \overline{L_1}$, e pertanto

$$P(F) = P(\overline{L_1}) = 1 - P(L_1) = 1 - 0{,}5 = 0{,}5.$$

Poiché $F \subset \overline{L_{3/4}}$, la probabilità cercata è

$$P(F|\overline{L_{3/4}}) = \frac{P(F \cap \overline{L_{3/4}})}{P(\overline{L_{3/4}})} = \frac{P(F)}{1 - P(L_{3/4})} = \frac{0{,}5}{1 - 0{,}375} = \frac{0{,}5}{0{,}625} = 0{,}8.$$

Esempio. Nell'esperimento del lancio di una moneta ripetuto 2 volte, supponendo che i quattro punti dello spazio campionario $S = \{cc, ct, tc, tt\}$ siano equiprobabili, qual è la probabilità condizionata che venga testa in entrambi i lanci sapendo che esce testa (a) nel primo lancio; (b) in almeno un lancio?

Soluzione. Sia $E = \{tt\} = \{\text{testa in entrambi i lanci}\}$, e $F = \{tc, tt\} = \{\text{testa al primo lancio}\}$. La probabilità cercata in (a) è allora

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(\{tt\})}{P(\{tc, tt\})} = \frac{1/4}{2/4} = \frac{1}{2}.$$

Sia $A = \{ct, tc, tt\}$; la soluzione di (b) è

$$P(E|A) = \frac{P(E \cap A)}{P(A)} = \frac{P(\{tt\})}{P(\{ct, tc, tt\})} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Notiamo che $P(E|F)$ e $P(E|A)$ sono diversi da $P(E) = \dfrac{1}{4}$.

Esempio. Nell'esperimento del lancio di una moneta ripetuto n volte, supponendo che i 2^n punti dello spazio campionario S siano equiprobabili, qual è la probabilità condizionata che venga testa in ogni lancio sapendo che esce testa (a) nel primo lancio; (b) in almeno un lancio?

Soluzione. Sia $E = \{\text{testa in ogni lancio}\}$ e $F = \{\text{testa al primo lancio}\}$. La probabilità cercata in (a) è allora

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E)}{P(F)} = \frac{1/2^n}{1/2} = \frac{1}{2^{n-1}}.$$

Sia $A = \{\text{testa in almeno un lancio}\}$. Poiché $P(A) = 1 - P(\overline{A})$, con $P(\overline{A}) = 1 - 1/2^n$, si ha

$$P(E|A) = \frac{P(E \cap A)}{P(A)} = \frac{P(E)}{1 - P(\overline{A})} = \frac{1/2^n}{1 - 1/2^n} = \frac{1}{2^n - 1}.$$

Notiamo che $P(E|F)$ e $P(E|A)$ sono diversi da $P(E) = \dfrac{1}{2^n}$.

Esempio. Nel gioco del bridge le 52 carte sono distribuite equamente a 4 giocatori (Est, Ovest, Nord e Sud). Se Nord e Sud hanno in tutto 8 picche, qual è la probabilità p_k che Est abbia k delle 5 carte di picche rimanenti? $(k = 0, 1, 2, 3, 4, 5)$

Soluzione. Si tratta di probabilità condizionata, ma è più semplice procedere con lo spazio campionario ridotto. Dato che Nord-Sud hanno in tutto 8 picche tra le loro carte, vi sono esattamente 5 picche nelle rimanenti 26 carte di Est-Ovest. Per la equiprobabilità delle distribuzioni delle carte, le probabilità cercate sono date da

$$p_0 = \frac{\binom{5}{0}\binom{21}{13}}{\binom{26}{13}} = p_5 = \frac{\binom{5}{5}\binom{21}{8}}{\binom{26}{13}} = \frac{1 \cdot 203\,490}{10\,400\,600} = \frac{9}{460} \approx 0{,}0196$$

$$p_1 = \frac{\binom{5}{1}\binom{21}{12}}{\binom{26}{13}} = p_4 = \frac{\binom{5}{4}\binom{21}{9}}{\binom{26}{13}} = \frac{5 \cdot 293\,930}{10\,400\,600} = \frac{13}{92} \approx 0{,}141$$

$$p_2 = \frac{\binom{5}{2}\binom{21}{11}}{\binom{26}{13}} = p_3 = \frac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}} = \frac{10 \cdot 352\,716}{10\,400\,600} = \frac{39}{115} \approx 0{,}339$$

Esempio. Da un'urna contenente r biglie rosse e b biglie blu si estraggono in sequenza n biglie a caso, senza reinserimento ($n \leq b+r$). Sapendo che k delle biglie estratte sono blu, qual è la probabilità condizionata che la prima biglia estratta sia blu?

Soluzione. Poniamo $E = \{$la prima biglia estratta è blu$\}$, e inoltre $B_k = \{$vengono estratte k biglie blu$\}$, $0 \leq k \leq n$. Si ha allora

$$P(E|B_k) = \frac{P(E \cap B_k)}{P(B_k)} = \frac{P(B_k|E)\,P(E)}{P(B_k)}.$$

Risulta

$$P(E) = \frac{b}{r+b}, \qquad P(B_k) = \frac{\binom{b}{k}\binom{r}{n-k}}{\binom{r+b}{n}}.$$

Infatti, $P(B_k)$ è la probabilità che in una estrazione casuale di n biglie da un'urna contenente r biglie rosse e b blu, vengano estratte k biglie blu. Notiamo inoltre che l'evento E corrisponde a B_1 quando $n=1$.

Notiamo che $P(B_k|E)$ è la probabilità che in una estrazione casuale di $n-1$ biglie da un'urna contenente r biglie rosse e $b-1$ blu, vengano estratte $k-1$ biglie blu; quindi:

$$P(B_k|E) = \frac{\binom{b-1}{k-1}\binom{r}{n-k}}{\binom{r+b-1}{n-1}}.$$

Pertanto risulta

$$P(E|B_k) = \frac{P(B_k|E)\cdot P(E)}{P(B_k)} = \frac{\binom{b-1}{k-1}\binom{r}{n-k}}{\binom{r+b-1}{n-1}} \cdot \frac{b}{r+b} \cdot \frac{\binom{r+b}{n}}{\binom{b}{k}\binom{r}{n-k}} = \frac{k}{n},$$

essendo

$$\binom{b}{k} = \frac{b\,(b-1)!}{k\,(k-1)!\,(b-k)!} = \frac{b}{k}\binom{b-1}{k-1}, \qquad \binom{r+b}{n} = \frac{n+b}{n}\binom{r+b-1}{n-1}.$$

In precedenza si è fatto uso dell'uguaglianza $P(E \cap B_k) = P(B_k|E)\,P(E)$. Questa discende immediatamente dalla definizione di probabilità condizionata. Infatti, moltiplicando ambo i membri dell'equazione $P(E|F) = \dfrac{P(E \cap F)}{P(F)}$ per $P(F)$, si ha

$$P(E \cap F) = P(F)\,P(E|F), \qquad \text{se } P(F) > 0.$$

Una generalizzazione della formula precedente è mostrata qui di seguito, ed è spesso anche nota come *legge delle probabilità composte.*

Proposizione. (Regola del prodotto) Se $P(E_1 \cap \ldots \cap E_{n-1}) > 0$, allora

$$P(E_1 \cap \ldots \cap E_n) = P(E_1)\,P(E_2|E_1)\,P(E_3|E_2 \cap E_1) \ldots P(E_n|E_1 \cap \ldots \cap E_{n-1})$$

Dimostrazione. Per la definizione di probabilità condizionata, dal 2° membro si ha

$$P(E_1)\frac{P(E_1 \cap E_2)}{P(E_1)}\frac{P(E_1 \cap E_2 \cap E_3)}{P(E_1 \cap E_2)} \ldots \frac{P(E_1 \cap E_2 \cap \ldots \cap E_n)}{P(E_1 \cap E_2 \cap \ldots \cap E_{n-1})} = P(E_1 \cap \ldots \cap E_n)$$

con le probabilità a denominatore strettamente positive perché $P(E_1 \cap \ldots \cap E_{n-1}) > 0$.

Esempio. Da un'urna contenente n biglie numerate da 1 a n si estraggono 3 biglie a caso (senza reinserimento). Assumendo che vi sia concordanza all'estrazione k-esima se in tale estrazione fuoriesce la biglia avente numero k, calcolare la probabilità
(a) di avere 3 concordanze,
(a) di avere concordanza solo nelle prime 2 estrazioni.

Soluzione. Posto $A_k = \{\text{si ha concordanza all'estrazione } k\text{-esima}\}$, dalla legge delle probabilità composte segue che la probabilità richiesta in (a) è data da

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)\, P(A_2|A_1)\, P(A_3|A_1 \cap A_2) = \frac{1}{n} \cdot \frac{1}{n-1} \cdot \frac{1}{n-2}.$$

Analogamente, la probabilità richiesta in (b) è

$$P(A_1 \cap A_2 \cap \overline{A_3}) = P(A_1)\, P(A_2|A_1)\, P(\overline{A_3}|A_1 \cap A_2) = \frac{1}{n} \cdot \frac{1}{n-1} \cdot \frac{n-3}{n-2}.$$

Esempio. Nel gioco del lotto si estraggono a caso 5 numeri da un'urna contenente 90 numeri compresi tra 1 e 90. (Le cinquine sono equiprobabili). Se si fissano k numeri distinti compresi tra 1 e 90 qual è la probabilità che questi saranno tra i 5 estratti?

Soluzione. L'evento $A_k = \{\text{i } k \text{ numeri fissati sono tra i 5 estratti}\}$ si può esprimere in termini degli eventi $B_i = \{\text{l'}i\text{-esimo numero fissato è tra i 5 estratti}\}$, $1 \leq i \leq k$:

$$A_k = B_1 \cap B_2 \cap \ldots \cap B_k, \qquad 1 \leq k \leq 5.$$

Poiché le estrazioni sono senza reinserimento, dalla regola del prodotto si ha

$$P(A_k) = P(B_1 \cap B_2 \cap \ldots \cap B_k) = P(B_1)\, P(B_2|B_1)\ \ldots\ P(B_k|B_1 \cap \ldots \cap B_{k-1}).$$

Ciò coincide con quanto visto in precedenza; infatti ad esempio per $k = 5$ risulta:

$$\begin{aligned} P(A_5) &= P(B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5) \\ &= P(B_1)\, P(B_2|B_1)\, P(B_3|B_1B_2)\, P(B_4|B_1B_2B_3)\, P(B_5|B_1B_2B_3B_4) \\ &= \frac{5}{90} \cdot \frac{4}{89} \cdot \frac{3}{88} \cdot \frac{2}{87} \cdot \frac{1}{86} = \frac{1}{43.949.268} = \frac{1}{\binom{90}{5}}. \end{aligned}$$

Esempio. Da un'urna contenente r biglie rosse e b biglie blu si estraggono in sequenza n biglie a caso, senza reinserimento ($n \leq b + r$). Sapendo che k delle biglie estratte sono blu,
(a) qual è la probabilità condizionata che la prima biglia estratta sia blu?
(b) qual è la probabilità condizionata che le prime 2 biglie estratte siano blu?

Soluzione. Poniamo $E = \{$la prima biglia estratta è blu$\}$, $F = \{$la seconda biglia estratta è blu$\}$, e inoltre $B_k = \{$vengono estratte k biglie blu$\}$, $0 \leq k \leq n$. Si ha allora

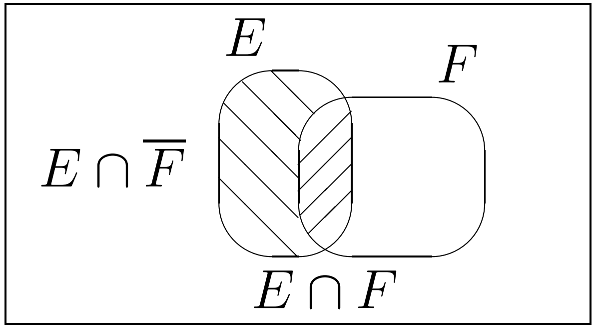
$$P(E|B_k) = \frac{k}{n}, \qquad P(E \cap F|B_k) = P(E|B_k)\, P(F|E \cap B_k) = \frac{k}{n} \cdot \frac{k-1}{n-1}.$$

Infatti, sapendo che si è verificato l'evento B_k, la sequenza delle n biglie estratte è costituita da k biglie blu e $n - k$ biglie rosse; in tal caso vi sono k casi favorevoli, su n, affinché la prima biglia sia blu, e $k - 1$ casi favorevoli, su $n - 1$, affinché la seconda biglia sia blu, sapendo che la prima biglia è blu.

3.3 La formula delle alternative e la formula di Bayes

Siano E e F due eventi, con $0 < P(F) < 1$.
L'evento E si può esprimere come

$$E = (E \cap F) \cup (E \cap \overline{F})$$



Infatti se un evento elementare appartiene all'evento E, esso inoltre appartiene o all'evento F o al suo complementare $\overline{F}$, e quindi appartiene o all'evento $E \cap F$ oppure a $E \cap \overline{F}$. Usando la proprietà di additività finita e la regola del prodotto segue:

$$\begin{aligned} P(E) &= P(E \cap F) + P(E \cap \overline{F}) \qquad (E \cap F \text{ e } E \cap \overline{F} \text{ sono incompatibili}) \\ &= P(E|F)\, P(F) + P(E|\overline{F})\, P(\overline{F}). \end{aligned}$$

Questa formula è estremamente utile in quanto permette di determinare la probabilità di un evento condizionandolo prima alla realizzazione o meno di un altro evento.

Esempio. Da un'urna contenente 5 biglie bianche e 1 biglia rossa, 6 giocatori estraggono a turno 1 biglia a caso, senza reinserimento. Qual è la probabilità che il giocatore k-esimo estragga la biglia rossa?

Soluzione. Posto $A_k = \{\text{il giocatore } k\text{-esimo estrae la biglia rossa}\}$, risulta

$$P(A_1) = \frac{1}{6}.$$

Inoltre

$$P(A_2) = P(A_2|A_1)P(A_1) + P(A_2|\overline{A_1})P(\overline{A_1}) = 0 \cdot \frac{1}{6} + \frac{1}{5} \cdot \frac{5}{6} = \frac{1}{6}.$$

Analogamente si ottiene $P(A_k) = \dfrac{1}{6}$, per $k = 1, 2, \ldots, 6$.

Notiamo che risulta $P(A_k|\overline{A_1} \cap \overline{A_2} \cap \ldots \cap \overline{A_{k-1}}) = \dfrac{1}{7-k}$, per $k = 1, 2, \ldots, 6$.

Osserviamo inoltre che gli eventi $A_1, A_2, \ldots, A_6$ sono necessari e a 2 a 2 incompatibili.

Esempio. Una compagnia assicuratrice suddivide le persone in due classi: quelle propense a incidenti (il 30%) e quelle che non lo sono (il 70%). Le statistiche mostrano che le persone propense a incidenti hanno probabilità 0,4 di avere un incidente in un anno, mentre per le altre vale 0,2.

(a) Qual è la probabilità che un nuovo assicurato abbia un incidente entro un anno?

(b) Se un nuovo assicurato ha un incidente entro un anno, qual è la probabilità che si tratti di una persona propensa agli incidenti?

Soluzione. Definiamo gli eventi $F = \{\text{una persona è propensa a incidenti}\}$ ed $E = \{\text{un nuovo assicurato ha un incidente entro un anno}\}$. Per le ipotesi fatte risulta:

$$P(F) = 0{,}3 \qquad P(\overline{F}) = 0{,}7 \qquad P(E|F) = 0{,}4 \qquad P(E|\overline{F}) = 0{,}2.$$

(a) Si ha: $P(E) = P(E|F)\,P(F) + P(E|\overline{F})\,P(\overline{F}) = 0{,}4 \cdot 0{,}3 + 0{,}2 \cdot 0{,}7 = 0{,}26.$

(b) Risulta:

$$P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{P(E|F)\,P(F)}{P(E)} = \frac{0{,}4 \cdot 0{,}3}{0{,}26} = \frac{6}{13} \approx 0{,}461.$$

Esempio. In un laboratorio di analisi l'esame del sangue è efficace al 95% nell'individuare la presenza di una certa malattia. L'esame tuttavia rileva anche dei "falsi positivi" nell'1% delle persone sane. (Con probabilità 0,01 l'esame indica erroneamente come malata una persona sana). Se lo 0,5% della popolazione è affetto dalla malattia, qual è la probabilità che una persona risultata positiva all'esame abbia la malattia?

Soluzione. Definiamo gli eventi $F = \{$la persona ha la malattia$\}$ ed $E = \{$l'esame dà esito positivo$\}$. Dalle ipotesi segue:

$$P(F) = 0{,}005 \qquad P(\overline{F}) = 0{,}995 \qquad P(E|F) = 0{,}95 \qquad P(E|\overline{F}) = 0{,}01$$

e quindi

$$P(E) = P(E|F)\,P(F) + P(E|\overline{F})\,P(\overline{F}) = 0{,}95 \cdot 0{,}005 + 0{,}01 \cdot 0{,}995 = 0{,}0147.$$

Pertanto la probabilità richiesta è

$$P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{P(E|F)\,P(F)}{P(E)} = \frac{0{,}95 \cdot 0{,}005}{0{,}0147} = \frac{0{,}00475}{0{,}0147} \approx 0{,}323.$$

Vediamo ora come si può generalizzare la seguente formula:

$$P(E) = P(E|F)\,P(F) + P(E|\overline{F})\,P(\overline{F}).$$

Supponiamo che $F_1, F_2, \ldots, F_n$ siano eventi a due a due incompatibili, ciascuno avente probabilità positiva, e necessari (ossia $\bigcup_{i=1}^{n} F_i = S$). Da queste ipotesi segue che in un esperimento si realizza uno e uno solo degli eventi $F_1, F_2, \ldots, F_n$. Scrivendo

$$E = E \cap S = E \cap \left(\bigcup_{i=1}^{n} F_i\right) = \bigcup_{i=1}^{n} (E \cap F_i) \qquad \text{(con } E \text{ evento qualsiasi)}$$

e osservando che gli eventi $E \cap F_i$, $i = 1, 2, \ldots, n$ sono a due a due incompatibili, per la proprietà di additività finita si ha

$$P(E) = P\left(\bigcup_{i=1}^{n} (E \cap F_i)\right) = \sum_{i=1}^{n} P\,(E \cap F_i) = \sum_{i=1}^{n} P(E|F_i)\,P(F_i).$$

La probabilità di E viene espressa come media ponderata delle $P(E|F_i)$, dove il peso di ciascun termine è uguale alla probabilità dell'evento rispetto al quale si condiziona.

Esempio. Un'urna contiene 3 monete; la prima è non truccata, le altre sono truccate in modo opposto: la seconda mostra testa con probabilità p, mentre la terza dà testa con probabilità $1-p$, con $0 < p < 1$. Se si sceglie una moneta a caso qual è la probabilità che lanciata mostri testa?

Soluzione. Definiamo gli eventi $F_j = \{\text{si sceglie la moneta } j\text{-esima}\}$, $j = 1,2,3$ e $T = \{\text{esce testa}\}$. Dalle ipotesi fatte segue:

$$P(F_j) = \frac{1}{3} \qquad (j = 1,2,3).$$

e inoltre

$$P(T|F_1) = 0{,}5 \qquad P(T|F_2) = p \qquad P(T|F_3) = 1-p$$

La probabilità di avere testa è quindi

$$P(T) = \sum_{j=1}^{3} P(T|F_j)\, P(F_j) = 0{,}5 \cdot \frac{1}{3} + p \cdot \frac{1}{3} + (1-p) \cdot \frac{1}{3} = \frac{1}{2}.$$

La seguente identità a volte prende il nome di **formula delle alternative** perché esprime la probabilità di un evento E a partire dalle possibili alternative $F_1, F_2, \ldots, F_n$:

$$P(E) = \sum_{i=1}^{n} P(E|F_i)\, P(F_i).$$

Supponiamo ora che E si sia verificato e di voler determinare quali degli eventi alternativi $F_1, F_2, \ldots, F_n$ si sia anch'esso verificato.

Proposizione. (Formula di Bayes) Se $F_1, F_2, \ldots, F_n$ sono eventi a due a due incompatibili, ciascuno avente probabilità positiva, e necessari, allora

$$P(F_j|E) = \frac{P(E \cap F_j)}{P(E)} = \frac{P(E|F_j)\, P(F_j)}{\sum_{i=1}^{n} P(E|F_i)\, P(F_i)} \qquad (j = 1, 2, \ldots, n).$$

Notiamo che le probabilità della formula di Bayes sommano all'unità; infatti risulta

$$\sum_{j=1}^{n} P(F_j|E) = \sum_{j=1}^{n} \frac{P(E|F_j)\, P(F_j)}{\sum_{i=1}^{n} P(E|F_i)\, P(F_i)} = 1.$$

Dalla formula di Bayes abbiamo ricavato

$$\sum_{j=1}^{n} P(F_j|E) = \sum_{j=1}^{n} \frac{P(E|F_j)\,P(F_j)}{\sum_{i=1}^{n} P(E|F_i)\,P(F_i)} = 1.$$

Notiamo che tale relazione sussiste ogniqualvolta si considerano n eventi $F_1, F_2, \ldots, F_n$ necessari e a due a due incompatibili. Infatti, per la proprietà di additività finita si ha:

$$\sum_{j=1}^{n} P(F_j) = P\left(\bigcup_{j=1}^{n} F_j\right) = P(S) = 1.$$

Esempio. In un gioco vi sono 3 carte identiche per la forma, la prima con entrambe le facce di colore rosso, la seconda con entrambe le facce di colore nero, la terza con una faccia rossa e una nera. Si sceglie a caso una carta e la si appoggia sul tavolo; se la faccia superiore della carta è rossa, qual è la probabilità che l'altra faccia sia nera?

Soluzione. Indichiamo con F_1, F_2 e F_3 gli eventi riferiti alle 3 carte, e poniamo $R = \{\text{la faccia superiore della carta scelta è rossa}\}$. Dalla formula di Bayes segue

$$\begin{aligned} P(F_3|R) &= \frac{P(R|F_3)\,P(F_3)}{P(R|F_1)\,P(F_1) + P(R|F_2)\,P(F_2) + P(R|F_3)\,P(F_3)} \\ &= \frac{\frac{1}{2}\cdot\frac{1}{3}}{1\cdot\frac{1}{3} + 0\cdot\frac{1}{3} + \frac{1}{2}\cdot\frac{1}{3}} = \frac{1}{3} \end{aligned}$$

Esempio. Si lanciano a caso n monete non truccate; per ogni moneta che mostra testa si inserisce una biglia nera in un'urna, mentre per ogni croce si inserisce una biglia bianca. Se poi si estrae a caso una biglia dall'urna, qual è la probabilità che sia nera? Se la biglia estratta è nera, qual è la probabilità che nell'urna vi erano k biglie nere?

Soluzione. Sia $E_k = \{\text{esce testa } k \text{ volte}\} = \{\text{nell'urna vi sono } k \text{ biglie nere e } n-k \text{ bianche}\}$, $k = 0, 1, \ldots, n$. Tali eventi sono a due a due incompatibili, sono necessari, e $P(E_k) = \dfrac{\binom{n}{k}}{2^n} > 0$. Sia $A = \{\text{la biglia è nera}\}$; per la formula delle alternative si ha

$$P(A) = \sum_{k=0}^{n} P(A|E_k)\, P(E_k) = \sum_{k=1}^{n} \frac{k}{n} \cdot \frac{\binom{n}{k}}{2^n} = \frac{1}{2^n} \sum_{k=1}^{n} \binom{n-1}{k-1},$$

essendo $\dfrac{k}{n} \cdot \dbinom{n}{k} = \dfrac{k}{n} \cdot \dfrac{n!}{k!\,(n-k)!} = \dfrac{(n-1)!}{(k-1)!\,(n-k)!} = \dbinom{n-1}{k-1}$. Ponendo $r = k-1$, dal teorema del binomio segue

$$P(A) = \frac{1}{2^n} \sum_{r=0}^{n-1} \binom{n-1}{r} = \frac{1}{2^n} \cdot 2^{n-1} = \frac{1}{2}.$$

Per ricavare $P(E_k|A)$ facciamo uso della formula di Bayes:

$$P(E_k|A) = \frac{P(A|E_k)\,P(E_k)}{\displaystyle\sum_{i=0}^{n} P(A|E_i)\,P(E_i)} = \frac{\dfrac{k}{n} \cdot \dfrac{\binom{n}{k}}{2^n}}{\dfrac{1}{2}} = \begin{cases} 0 & \text{se } k = 0 \\ \dfrac{\binom{n-1}{k-1}}{2^{n-1}} & \text{se } k = 1, 2, \ldots, n. \end{cases}$$

È facile verificare che la somma delle probabilità $P(E_k|A)$ è unitaria:

$$\sum_{k=0}^{n} P(E_k|A) = \sum_{k=1}^{n} \frac{\binom{n-1}{k-1}}{2^{n-1}} = \frac{1}{2^{n-1}} \sum_{r=0}^{n-1} \binom{n-1}{r} = \frac{1}{2^{n-1}} \cdot 2^{n-1} = 1.$$

3.4 Eventi indipendenti

La probabilità condizionata di E dato F non è generalmente uguale a $P(E)$. In altri termini, la conoscenza della realizzazione dell'evento F modifica in generale la possibilità del realizzarsi o meno di E.

Se $P(E|F) = P(E)$ diciamo che E è indipendente da F. Cioè, E è indipendente da F se la conoscenza della realizzazione di F non cambia la probabilità che si realizzi E.

Se $P(F) > 0$, dalla formula $P(E|F) = \dfrac{P(E \cap F)}{P(F)}$ si vede che E è indipendente da F se

$$P(E \cap F) = P(E)\, P(F).$$

Tale formula è simmetrica in E ed F, pertanto se $P(E) > 0$ e $P(F) > 0$, l'evento E è indipendente da F se F è indipendente da E e viceversa.

La seguente definizione include anche i casi in cui $P(E) = 0$ oppure $P(F) = 0$.

Definizione. Due eventi E ed F si dicono *indipendenti* se vale

$$P(E \cap F) = P(E)\, P(F).$$

Due eventi che non sono indipendenti si dicono *dipendenti*.

Esempio. Uno studente deve sottoporsi a due test. Con probabilità 0,5 supererà il primo test; con probabilità 0,4 supererà il secondo test; con probabilità 0,3 li supererà entrambi. Gli eventi relativi al superamento dei due test sono indipendenti?

Soluzione. Sia B_i l'evento che lo studente superi il test i-esimo, $i = 1, 2$. Risulta

$$P(B_1 \cap B_2) = 0{,}3 \neq 0{,}2 = 0{,}5 \cdot 0{,}4 = P(B_1)\, P(B_2),$$

quindi gli eventi B_1 e B_2 sono dipendenti.

Esercizio. Nell'esperimento che consiste nel lancio di n monete non truccate, sia $T_1 = \{$esce testa al primo lancio$\}$, $U = \{$esce lo stesso risultato negli n lanci$\}$, $A = \{$esce testa almeno 1 volta$\}$. Mostrare che T_1 e U sono indipendenti, ed inoltre che T_1 e A non sono indipendenti.

Nota. Se per gli eventi A e B risulta $A \subset B$, allora sussiste indipendenza tra i 2 eventi se e solo se $P(A) = 0$ oppure $P(B) = 1$.

Nota. Se $P(A) = 0$, allora l'evento A è indipendente da qualsiasi altro evento B.

Proposizione. Se E ed F sono eventi indipendenti, allora E ed $\overline{F}$ sono indipendenti.
Dimostrazione. Poichè risulta $E = (E \cap F) \cup (E \cap \overline{F})$, con $E \cap F$ ed $E \cap \overline{F}$ eventi incompatibili, dalla proprietà di additività finita segue

$$P(E) = P(E \cap F) + P(E \cap \overline{F}).$$

Poiché per ipotesi E ed F sono eventi indipendenti, si ha

$$P(E) = P(E)\, P(F) + P(E \cap \overline{F}),$$

ossia

$$P(E \cap \overline{F}) = P(E) - P(E)\, P(F) = P(E)\, [1 - P(F)] = P(E)\, P(\overline{F}).$$

Quindi E ed $\overline{F}$ sono indipendenti.

Notiamo pertanto che se E è indipendente da F, la probabilità che E si realizzi non è modificata dalla realizzazione o meno di F.

Inoltre, se E ed F sono indipendenti, tali sono anche $\overline{E}$ ed F, e gli eventi $\overline{E}$ ed $\overline{F}$.

Nel prossimo esempio vedremo che se E è indipendente da F e da G, allora non è detto che E sia indipendente da $F \cap G$.

Esempio. Consideriamo i seguenti eventi nel lancio di due dadi non truccati: $E = \{$la somma dei dadi è 7$\}$, $F = \{$il primo dado dà 4$\}$, $G = \{$il secondo dado dà 3$\}$. Esaminare l'indipendenza delle coppie di eventi E ed F, E e G, E ed $F \cap G$.
Soluzione. L'evento E è indipendente da F ed anche da G, in quanto

$$P(E \cap F) = P(\{(4,3)\}) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = P(E)\, P(F),$$

$$P(E \cap G) = P(\{(4,3)\}) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = P(E)\, P(G).$$

Inoltre, poiché $P(E|F \cap G) = 1$, l'evento E non è indipendente da $F \cap G$.

Ispirandoci a questo esempio, appare ragionevole definire l'indipendenza di tre eventi non limitandosi a richiedere l'indipendenza delle 3 possibili coppie, ma imponendo anche una condizione che coinvolga complessivamente i 3 eventi.

Definizione. Tre eventi E, F, G si dicono indipendenti se

$$P(E \cap F \cap G) = P(E)\, P(F)\, P(G)$$

$$P(E \cap F) = P(E)\, P(F)$$

$$P(E \cap G) = P(E)\, P(G)$$

$$P(F \cap G) = P(F)\, P(G)$$

Si noti che se E, F, G sono indipendenti, allora E è indipendente da ogni evento formato a partire da F e G. Ad esempio E è indipendente da $F \cup G$. Infatti si ha

$$\begin{aligned} P(E \cap (F \cup G)) &= P((E \cap F) \cup (E \cap G)) \\ &= P(E \cap F) + P(E \cap G) - P(E \cap F \cap G) \\ &= P(E)\, P(F) + P(E)\, P(G) - P(E)\, P(F \cap G) \\ &= P(E)\, [P(F) + P(G) - P(F \cap G)] \\ &= P(E)\, P(F \cup G). \end{aligned}$$

Esempio. Nel lancio di due monete non truccate, con $S = \{cc, ct, tc, tt\}$, stabilire se i seguenti eventi sono indipendenti:

$$T_1 = \{tc, tt\}, \qquad T_2 = \{ct, tt\}, \qquad U = \{cc, tt\}.$$

Soluzione. Si ricava facilmente che

$$\frac{1}{4} = P(\{tt\}) = P(T_1 \cap T_2) = P(T_1)\, P(T_2) = \frac{1}{2} \cdot \frac{1}{2},$$

$$\frac{1}{4} = P(\{tt\}) = P(T_1 \cap U) = P(T_1)\, P(U) = \frac{1}{2} \cdot \frac{1}{2},$$

$$\frac{1}{4} = P(\{tt\}) = P(T_2 \cap U) = P(T_2)\, P(U) = \frac{1}{2} \cdot \frac{1}{2}.$$

Gli eventi T_1, T_2, U sono dunque indipendenti a coppie. Ciononostante, essendo

$$\frac{1}{4} = P(\{tt\}) = P(T_1 \cap T_2 \cap U) \neq P(T_1)\, P(T_2)\, P(U) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8},$$

i tre eventi non sono indipendenti.

Esempio. Nell'eperimento che consiste nel generare a caso una sequenza booleana di lunghezza 3, stabilire se i seguenti eventi sono indipendenti:

$$A = \{000, 001, 010, 100\}, \quad B = \{000, 001, 100, 101\}, \quad C = \{011, 100, 110, 111\}.$$

Soluzione. Lo spazio campionario è costituito da 8 sequenze equiprobabili, quindi

$$\frac{1}{8} = P(\{100\}) = P(A \cap B \cap C) \; = \; P(A)\,P(B)\,P(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8},$$
$$\frac{3}{8} = P(\{000, 001, 100\}) = P(A \cap B) \; \neq \; P(A)\,P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4},$$
$$\frac{1}{8} = P(\{100\}) = P(A \cap C) \; \neq \; P(A)\,P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4},$$
$$\frac{1}{8} = P(\{100\}) = P(B \cap C) \; \neq \; P(B)\,P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Gli eventi A, B ,C non sono indipendenti poiché non sono indipendenti a coppie.

Definizione. Gli eventi $E_1, E_2, \ldots, E_n$ si dicono indipendenti se per ogni sottoinsieme $E_{i_1}, E_{i_2}, \ldots, E_{i_r}$, $2 \leq r \leq n$, di questi eventi si ha

$$P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}) = P(E_{i_1})\, P(E_{i_2})\, \ldots\, P(E_{i_r}).$$

Il numero di uguaglianze coinvolte nell'indipendenza di n eventi è pari al numero di sottoinsiemi di cardinalità $r = 2, 3, \ldots, n$ di un insieme di n elementi, ossia:

$$\sum_{r=2}^{n} \binom{n}{r} = 2^n - \binom{n}{1} - \binom{n}{0} = 2^n - n - 1.$$

Definizione. Gli eventi di una successione $E_1, E_2, \ldots$ si dicono indipendenti se ogni sottoinsieme finito di essi $E_{i_1}, E_{i_2}, \ldots, E_{i_r}$, $r \geq 2$, è formato da eventi indipendenti.

Esercizio. Siano A e B eventi tali che $P(A) > 0$ e $P(B) > 0$. Mostrare che le seguenti uguaglianze sono equivalenti:

$$P(A \cap B) = P(A)\, P(B), \qquad P(A|B) = P(A), \qquad P(B|A) = P(B).$$

Alcuni esperimenti talvolta consistono nell'effettuare una successione di sub-esperimenti identici, ripetuti nelle stesse condizioni, ai quali si dà il nome di *prove*.

Ad esempio, nel lancio ripetuto di una moneta, ogni lancio corrisponde ad una prova.

Talora è ragionevole supporre che gli esiti di ogni gruppo di prove non abbiano effetti sugli esiti delle altre prove. In tal caso diciamo che le prove sono indipendenti. In altri termini, la successione

$$E_1, E_2, \ldots, E_n, \ldots$$

costituisce un insieme di eventi indipendenti, dove E_i dipende esclusivamente dall'esito dell'i-esima prova.

Esempio. In una sequenza infinita di prove indipendenti ogni prova ha 2 esiti: successo con probabilità p e insuccesso con probabilità $1 - p$. Qual è la probabilità che
(a) vi sia almeno un successo nelle prime n prove;
(b) vi siano esattamente k successi nelle prime n prove;
(c) tutte le prove abbiano successo?

Soluzione. (a) Posto $E_i = \{$insuccesso alla prova i-esima$\}$, la probabilità richiesta è

$$P\left(\bigcup_{i=1}^{n} \overline{E_i}\right) = P\left(\overline{\bigcap_{i=1}^{n} E_i}\right) = 1 - P\left(\bigcap_{i=1}^{n} E_i\right) = 1 - \prod_{i=1}^{n} P(E_i) = 1 - (1-p)^n.$$

(b) Consideriamo le sequenze che nei primi n esiti hanno k successi e $n - k$ insuccessi. Ognuna di esse si realizza, per l'indipendenza, con probabilità $p^k(1-p)^{n-k}$. Ad esempio $P(\overline{E_1} \ldots \overline{E_k} E_{k+1} \ldots E_n) = P(\overline{E_1}) \ldots P(\overline{E_k}) P(E_{k+1}) \ldots P(E_n) = p^k(1-p)^{n-k}$. Vi sono $\binom{n}{k}$ sequenze di questo tipo poiché vi sono $n!/k!(n-k)!$ permutazioni distinte di k successi e $n - k$ insuccessi, quindi la probabilità richiesta è

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

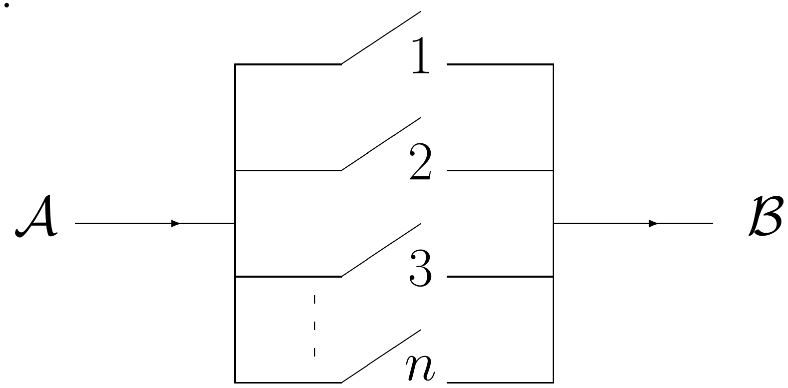
(c) La probabilità di avere n successi nelle prime n prove è, per l'indipendenza,

$$P\left(\bigcap_{i=1}^{n}\overline{E_i}\right)=\prod_{i=1}^{n}P\left(\overline{E_i}\right)=p^n.$$

Di conseguenza, per la proprietà di continuità della probabilità, la probabilità che tutte le prove abbiano successo è data da

$$\begin{aligned}
P\left(\bigcap_{i=1}^{\infty}\overline{E_i}\right) &= P\left(\lim_{n\to\infty}\bigcap_{i=1}^{n}\overline{E_i}\right)\\
&= \lim_{n\to\infty}P\left(\bigcap_{i=1}^{n}\overline{E_i}\right)\\
&= \lim_{n\to\infty}p^n=\begin{cases}0 & \text{se } p<1\\ 1 & \text{se } p=1.\end{cases}
\end{aligned}$$

Esempio. Un sistema formato da n componenti è detto sistema parallelo se funziona quando almeno uno dei suoi componenti funziona.
Il componente i-esimo, indipendentemente dagli altri, funziona con probabilità p_i, $i = 1, 2, \ldots, n$.
Qual è la probabilità che il sistema funzioni?

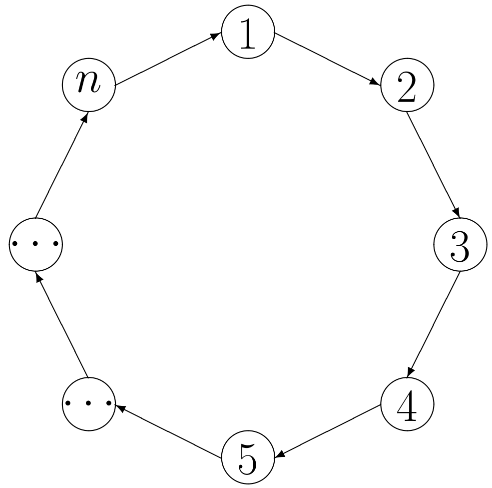


Soluzione.
Sia $E_i = \{$il componente i-esimo funziona$\}$. Allora

$$\begin{aligned} P(\text{il sistema funziona}) &= P\left(\bigcup_{i=1}^{n} E_i\right) = 1 - P\left(\bigcap_{i=1}^{n} \overline{E_i}\right) \\ &= 1 - \prod_{i=1}^{n} P\left(\overline{E_i}\right) = 1 - \prod_{i=1}^{n} (1 - p_i) \end{aligned}$$

per l'indipendenza.

Esercizio. Si consideri un sistema costituito da n unità numerate da 1 a n (vedi figura). Ciascuna unità, indipendentemente dalle altre, sceglie a caso un numero da 1 a n diverso dal proprio numero e lo invia all'unità successiva in senso orario. Calcolare la probabilità



(a) che tutte le unità ricevano dall'unità precedente un numero identico al proprio;

(b) che nessuna unità riceva dall'unità precedente un numero identico al proprio;

(c) che almeno una unità riceva dall'unità precedente un numero identico al proprio.

Valutare le probabilità per $n \to \infty$, ricordando che $\boxed{\lim_{n\to\infty}\left(1+\frac{\alpha_n}{n}\right)^n = e^{\lim_{n\to\infty}\alpha_n}}$

Soluzione.

(a) $1/(n-1)^n$ $\;(\to 0$ per $n \to \infty)$;

(b) $[(n-2)/(n-1)]^n$ $\;(\to 1/e$ per $n \to \infty)$;

(c) $1-[(n-2)/(n-1)]^n$ $\;(\to 1-1/e$ per $n \to \infty)$.

Esercizio. Nell'esperimento che consiste nel lanciare n volte una moneta truccata, indicando con t la fuoriuscita di testa e con c di croce, lo spazio campionario è

$$S = \{(\omega_1, \omega_2, \ldots, \omega_n) \colon \omega_i \in \{t, c\} \text{ per } 1 \leq i \leq n\},$$

le cui 2^n sequenze in generale non sono equiprobabili. Supponendo che per $0 < p < 1$

$$P(T_k) = P(\{\text{al } k\text{-esimo lancio esce testa}\}) = p, \qquad 1 \leq k \leq n,$$

e che i lanci siano indipendenti, calcolare le probabilità degli eventi $A_k = \{$nei primi k lanci esce testa$\}$, $A'_k = \{$nei primi k lanci esce croce$\}$, $B_k = \{$esce testa k volte$\}$.
Soluzione. Per l'indipendenza si ha

$$P(A_k) = P(T_1 \cap T_2 \cap \ldots \cap T_k) = P(T_1)\, P(T_2) \ldots P(T_k) = p^k,$$

e analogamente $P(A'_k) = (1-p)^k$. Inoltre risulta

$$P(B_k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

3.5 $P(\cdot|F)$ è una probabilità

Vedremo ora che la probabilità condizionata soddisfa i tre assiomi della probabilità.

Proposizione. Sia F un evento tale che $P(F) > 0$; allora la probabilità condizionata da F è una funzione $P(\cdot\,|F)\colon \mathcal{F} \to \mathbb{R}$ tale che

(a) $0 \leq P(A|F) \leq 1$ per ogni evento A;

(b) $P(S|F) = 1$;

(c) per ogni successione di eventi a due a due incompatibili $A_1, A_2, \ldots,$

$$P\left(\bigcup_{i=1}^{\infty} A_i \Big| F\right) = \sum_{i=1}^{\infty} P(A_i|F).$$

Dimostrazione. (a) Occorre mostrare che $0 \leq P(A \cap F)/P(F) \leq 1$. La prima disuguaglianza è ovvia, mentre la seconda discende dal fatto che $A \cap F \subset F$, da cui segue $P(A \cap F) \leq P(F)$.

(b) Si ha

$$P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1.$$

(c) Risulta

$$P\left(\bigcup_{i=1}^{\infty} A_i | F\right) = P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap F\right) \frac{1}{P(F)} = P\left(\bigcup_{i=1}^{\infty} (A_i \cap F)\right) \frac{1}{P(F)}.$$

Essendo gli eventi $A_i \cap F$ incompatibili, per la proprietà di additività numerabile si ha

$$P\left(\bigcup_{i=1}^{\infty} A_i | F\right) = \sum_{i=1}^{\infty} P(A_i \cap F) \frac{1}{P(F)} = \sum_{i=1}^{\infty} P(A_i|F),$$

e ciò conclude la dimostrazione.

Poiché la probabilità condizionata $P(\cdot\,|F)$ soddisfa gli assiomi della probabilità, ad essa si applicano le proposizioni sulla probabilità dimostrate in precedenza; ad esempio

$$P(A \cup B|F) = P(A|F) + P(B|F) - P(A \cap B|F).$$

Esercizio. Stabilire se $Q(E|F) := \dfrac{P(E \cup F)}{P(F)}$ soddisfa i 3 assiomi della probabilità.

Soluzione. $Q(E|F)$ non è una probabilità.

Proposizione. Siano F e B eventi tali che $P(F) > 0$, $P(B \cap F) > 0$ e $P(\overline{B} \cap F) > 0$; allora

$$P(A|F) = P(A|B \cap F)\, P(B|F) + P(A|\overline{B} \cap F)\, P(\overline{B}|F).$$

Dimostrazione. Dalle proprietà della probabilità condizionata segue

$$\begin{aligned} P(A|F) &= \frac{P(A \cap F)}{P(F)} = \frac{P(A \cap B \cap F)}{P(F)} + \frac{P(A \cap \overline{B} \cap F)}{P(F)} \\ &= \frac{P(A|B \cap F)\, P(B \cap F)}{P(F)} + \frac{P(A|\overline{B} \cap F)\, P(\overline{B} \cap F)}{P(F)} \\ &= P(A|B \cap F)\, P(B|F) + P(A|\overline{B} \cap F)\, P(\overline{B}|F). \end{aligned}$$

Notiamo che la formula appena dimostrata è un'estensione di quella delle alternative

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

Esempio. Un'urna contiene n biglie, di cui k sono rosse ($k \geq 3$) e le rimanenti sono blu. Se ne estraggono a caso 3 in sequenza, senza reinserimento. Se la prima estratta è rossa, qual è la probabilità che la terza estratta sia rossa?

Soluzione. Definiamo gli eventi $A_i = \{$l'i-esima biglia estratta è rossa$\}$, $i = 1, 2, 3$. Poiché si ha

$$P(A_2|A_1) = \frac{k-1}{n-1}, \qquad P(\overline{A_2}|A_1) = \frac{n-k}{n-1}$$

e

$$P(A_3|A_1 \cap A_2) = \frac{k-2}{n-2}, \qquad P(A_3|A_1 \cap \overline{A_2}) = \frac{k-1}{n-2},$$

si ottiene

$$\begin{aligned} P(A_3|A_1) &= P(A_3|A_1 \cap A_2)\, P(A_2|A_1) + P(A_3|A_1 \cap \overline{A_2})\, P(\overline{A_2}|A_1) \\ &= \frac{k-2}{n-2} \cdot \frac{k-1}{n-1} + \frac{k-1}{n-2} \cdot \frac{n-k}{n-1} = \frac{n(k-1) - 2(k-1)}{(n-2)(n-1)} \\ &= \frac{k-1}{n-1}. \end{aligned}$$

Notiamo che risulta $P(A_3|A_1) = P(A_2|A_1)$.

Un concetto importante della teoria della probabilità è quello della indipendenza condizionata di eventi.

Diciamo che due eventi A e B sono *condizionatamente indipendenti* dato F se

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

o, equivalentemente, se

$$P(A \cap B|F) = P(A|F)\, P(B|F).$$

La nozione di indipendenza condizionata si può facilmente estendere a più di 2 eventi.

Esempio. Un'urna contiene 3 monete, che danno rispettivamente testa con probabilità $\frac{1}{2}$, $\frac{1}{4}$, $\frac{3}{4}$. Se si sceglie una moneta a caso e mostra testa al primo lancio, qual è la probabilità che la stessa mostri testa anche al secondo lancio?

Soluzione. Definiamo gli eventi $F_j = \{\text{si sceglie la moneta } j\text{-esima}\}$, $j = 1, 2, 3$ e $T_i = \{\text{esce testa al lancio } i\text{-esimo}\}$, $i = 1, 2$. È ragionevole assumere che i risultati di lanci distinti della medesima moneta siano (condizionatamente) indipendenti, quindi

$$P(T_1 \cap T_2|F_1) = P(T_1|F_1)\,P(T_2|F_1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$
$$P(T_1 \cap T_2|F_2) = P(T_1|F_2)\,P(T_2|F_2) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$$
$$P(T_1 \cap T_2|F_3) = P(T_1|F_3)\,P(T_2|F_3) = \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}.$$

Pertanto si ha $P(T_1 \cap T_2) = \displaystyle\sum_{j=1}^{3} P(T_1 \cap T_2|F_j)\,P(F_j) = \frac{1}{4} \cdot \frac{1}{3} + \frac{1}{16} \cdot \frac{1}{3} + \frac{9}{16} \cdot \frac{1}{3} = \frac{7}{24}$.

Infine, essendo $P(T_1) = \dfrac{1}{2}$ si ottiene $P(T_2|T_1) = \dfrac{P(T_1 \cap T_2)}{P(T_1)} = \dfrac{7/24}{1/2} = \dfrac{7}{12} = 0{,}5833.$

CAPITOLO 4 – Variabili aleatorie

4.1 Variabili aleatorie

Problema. Un sistema è costituito da n unità numerate da 1 a n. In ogni unità si sceglie a caso un numero tra 1 e n, indipendentemente dalle altre. Diciamo che si ha concordanza nell'unità k-esima se ivi si sceglie il numero k.

Qual è la probabilità che non vi siano concordanze?
Qual è la probabilità che vi siano k concordanze?

Posto $A_k = \{\text{vi sono } k \text{ concordanze}\}$, gli eventi $A_0, A_1, \ldots, A_n$ sono necessari e a 2 a 2 incompatibili. In virtù dell'indipendenza le probabilità richieste sono:

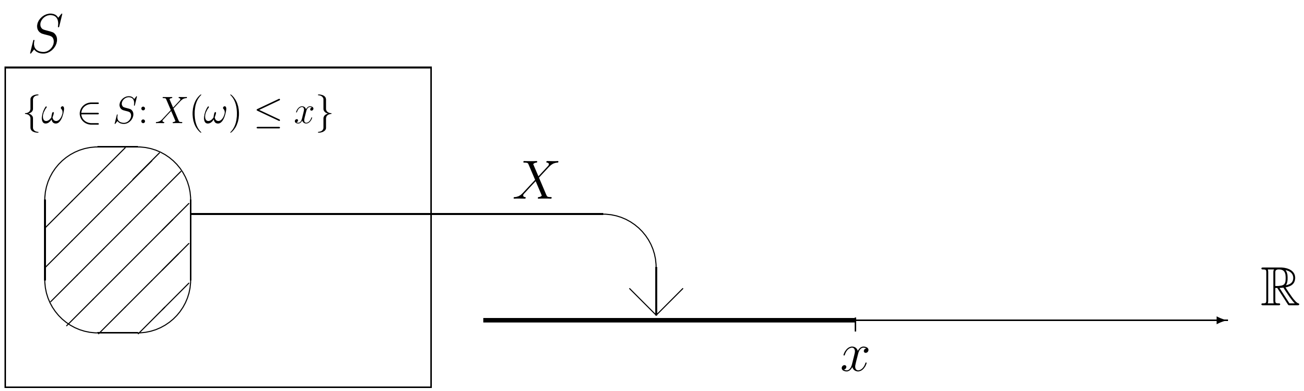
$$P(A_0) = \left(1 - \frac{1}{n}\right)^n, \qquad P(A_k) = \binom{n}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k}.$$

Notiamo che se si definisce $X =$"numero di concordanze", appare più adeguato esprimere le probabilità richieste come $P(X = 0) = P(A_0)$ e $P(X = k) = P(A_k)$.

Talora studiando un fenomeno aleatorio l'interesse va riposto su una funzione $X(\omega)$ dell'esito ω dell'esperimento. Ad esempio, lanciando due dadi siamo interessati alla somma dei due valori, oppure lanciando n volte una moneta possiamo riferirci al numero totale di teste. La quantità d'interesse, o più precisamente, queste funzioni a valori reali definite sullo spazio campionario sono note come *variabili aleatorie.*

Definizione. Dato uno spazio di probabilità $(S, \mathcal{F}, P)$, una funzione $X\colon S \to \mathbb{R}$ è detta variabile aleatoria se risulta

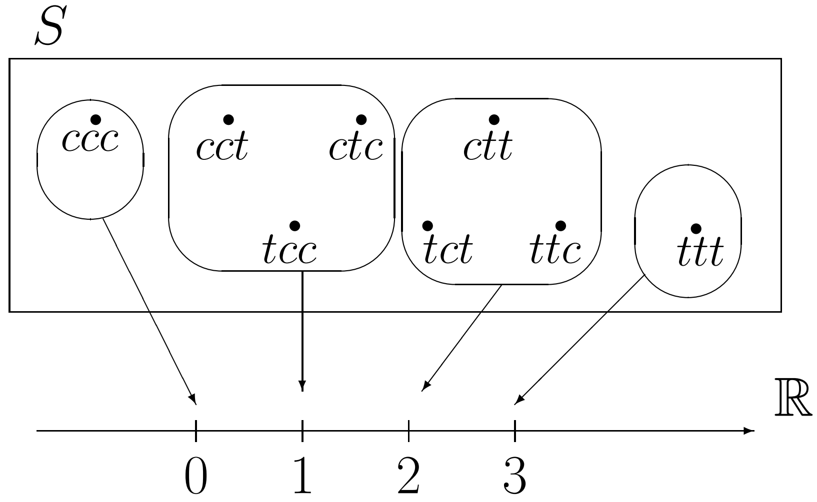
$$\{\omega \in S\colon X(\omega) \leq x\} \in \mathcal{F} \qquad \text{per ogni } x \in \mathbb{R}.$$



Poiché il valore di una variabile aleatoria è determinato dall'esito dell'esperimento, possiamo assegnare le probabilità ai possibili valori ottenuti dalla variabile aleatoria.

Esempio. Nel lancio di 3 monete non truccate sia Y il numero di teste che si ottengono; Y è una variabile aleatoria che assume i valori $0, 1, 2, 3$ con le seguenti probabilità:

$$
\begin{aligned}
P(Y=0) &= P(\{ccc\}) = \frac{1}{8} \\
P(Y=1) &= P(\{cct, ctc, tcc\}) = \frac{3}{8} \\
P(Y=2) &= P(\{ctt, tct, ttc\}) = \frac{3}{8} \\
P(Y=3) &= P(\{ttt\}) = \frac{1}{8}
\end{aligned}
$$



Poiché Y deve assumere uno tra i valori $0, 1, 2, 3$ abbiamo

$$P\left(\bigcup_{i=0}^{3}\{Y=i\}\right) = \sum_{i=0}^{3} P(Y=i) = \frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8} = 1.$$

Esempio. Si estraggono 3 biglie a caso senza reinserimento da un'urna contenente venti biglie numerate da 1 a 20. Qual è la probabilità che almeno una tra le biglie estratte abbia un numero maggiore o uguale a 17?

Soluzione. Se denotiamo con X il maggiore tra i 3 numeri estratti, X è una variabile aleatoria che assume i valori $3, 4, \dots, 20$. Inoltre, supponendo che ognuna delle $\binom{20}{3}$ possibili terne abbia uguale probabilità, si ha

$P(X = i)$	$P(X = 17)$	$P(X = 18)$	$P(X = 19)$	$P(X = 20)$
$\dfrac{\binom{i-1}{2}}{\binom{20}{3}}$	$\dfrac{\binom{16}{2}}{\binom{20}{3}} = \dfrac{2}{19}$	$\dfrac{\binom{17}{2}}{\binom{20}{3}} = \dfrac{34}{285}$	$\dfrac{\binom{18}{2}}{\binom{20}{3}} = \dfrac{51}{380}$	$\dfrac{\binom{19}{2}}{\binom{20}{3}} = \dfrac{3}{20}$

Infatti il numero di terne che compongono l'evento $\{X = i\}$ è il numero di terne per cui una biglia ha numero i e le altre due hanno numero compreso tra 1 e $i - 1$. Si ha

$$P(X \geq 17) = \sum_{i=17}^{20} P(X = i) = \frac{2}{19} + \frac{34}{285} + \frac{51}{380} + \frac{3}{20} \approx 0{,}508.$$

Esempio. Si estraggono 3 biglie a caso con reinserimento da un'urna contenente venti biglie numerate da 1 a 20. Qual è la probabilità che almeno una tra le biglie estratte abbia un numero maggiore o uguale a 17?

Soluzione. Sia Y la variabile aleatoria che descrive quante delle 3 biglie estratte abbiano un numero maggiore o uguale a 17. Tale variabile assume i valori 0, 1, 2, 3 con

$$P(Y = k) = \binom{3}{k} \left(\frac{4}{20}\right)^k \left(\frac{17}{20}\right)^{3-k}, \qquad k = 0, 1, 2, 3.$$

Pertanto la probabilità richiesta è

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \left(\frac{17}{20}\right)^3 = 1 - 0{,}614 = 0{,}386.$$

Esempio. Si lancia ripetutamente una moneta truccata, che in un singolo lancio dà testa con probabilità p, fino a che non appaia testa per la prima volta oppure si siano fatti n lanci. Se X denota il numero totale di volte che lanciamo la moneta, allora X è una variabile aleatoria che assume valori $1, 2, \ldots, n$ con probabilità

$$\begin{aligned} P(X=1) &= P(T_1) = p \\ P(X=k) &= P(C_1 \cap C_2 \cap \ldots \cap C_{k-1} \cap T_k) = (1-p)^{k-1}\, p \qquad (k = 2, 3, \ldots, n-1) \\ P(X=n) &= P(C_1 \cap C_2 \cap \ldots \cap C_n) + P(C_1 \cap C_2 \cap \ldots \cap C_{n-1} \cap T_n) = (1-p)^{n-1}, \end{aligned}$$

dove C_i e T_i rappresentano rispettivamenente la fuoriuscita di croce e testa al lancio i-esimo, e dove si è fatto uso dell'indipendenza nei lanci. Verifichiamo che

$$\begin{aligned} P\left(\bigcup_{k=1}^{n} \{X = k\} \right) &= \sum_{k=1}^{n} P(X=k) = \sum_{k=1}^{n-1} (1-p)^{k-1}\, p + (1-p)^{n-1} \\ &= p \sum_{r=0}^{n-2} (1-p)^r + (1-p)^{n-1} = p \left[\frac{1-(1-p)^{n-1}}{1-(1-p)} \right] + (1-p)^{n-1} = 1, \end{aligned}$$

avendo ricordato che $\sum_{r=0}^{m} c^r = (1-c^{m+1})/(1-c)$ per $c \neq 1$, e avendo posto $r = k-1$.

Definizione. Data una variabile aleatoria X, la funzione definita da

$$F(x) = P(X \leq x), \qquad -\infty < x < \infty$$

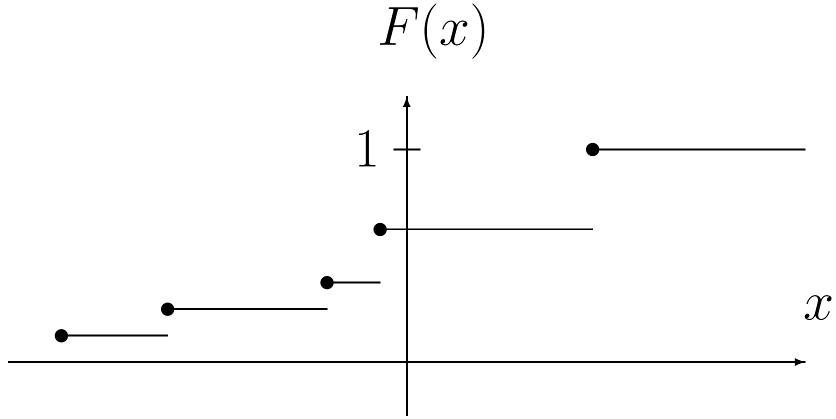
è detta *funzione di distribuzione* (o *di ripartizione*) di X.

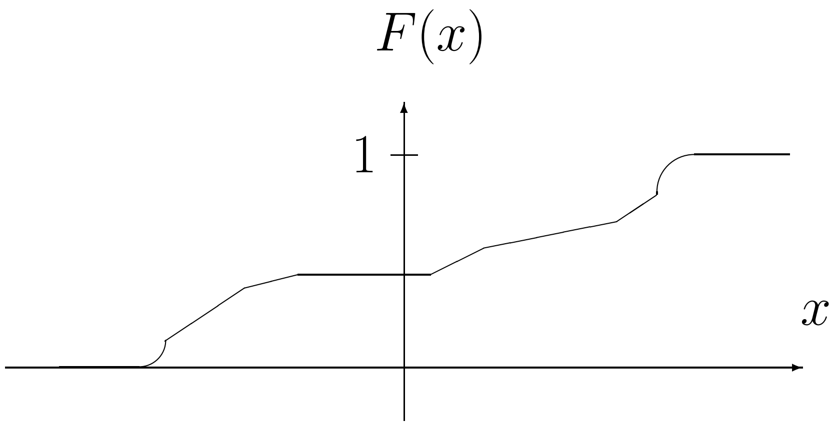
La funzione di distribuzione $F(x)$ rappresenta, per ogni $x \in \mathbb{R}$, la probabilità che la variabile aleatoria sia minore o uguale a x, ed è caratterizzata dalle seguenti proprietà:

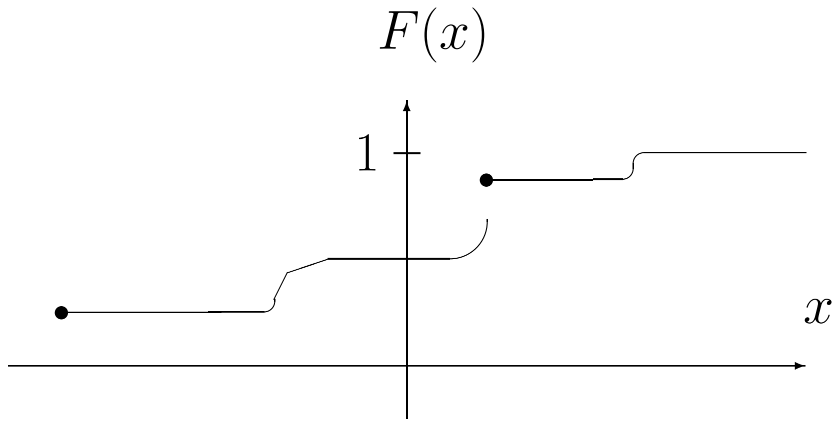
1. $F(x)$ è una funzione non decrescente, cioè risulta $F(a) \leq F(b)$ se $a < b$.
2. $\lim_{x \to \infty} F(x) = 1$.
3. $\lim_{x \to -\infty} F(x) = 0$.
4. $F(x)$ è continua a destra, ossia $\lim_{n \to \infty} F(x_n) = F(x)$ per ogni $x \in \mathbb{R}$ fissato e per ogni successione decrescente x_n, $n \geq 1$, che converge a x; equivalentemente, si ha $\lim_{h \to 0^+} F(x+h) = F(x)$ per ogni $x \in \mathbb{R}$ fissato.

Proprietà di $F(x) = P(X \le x)$:

1. $F(x)$ è non decrescente.
2. $\lim_{x\to\infty} F(x) = 1$.
3. $\lim_{x\to-\infty} F(x) = 0$.
4. $F(x)$ è continua a destra.







4.2 Variabili aleatorie discrete

Una variabile aleatoria che possa assumere al più un'infinità numerabile di valori è detta *discreta*. Per una variabile aleatoria discreta X, definiamo la *densità discreta* (o *funzione di probabilità*) $p(k)$ di X come

$$p(k) = P(X = k).$$

La densità discreta $p(k)$ è positiva al più per un'infinità numerabile di valori di k. Quindi, se X assume i valori $x_1, x_2, \ldots$, allora

$$p(x_i) \geq 0 \qquad i = 1, 2, \ldots,$$

$$p(x) = 0 \qquad \text{altrimenti.}$$

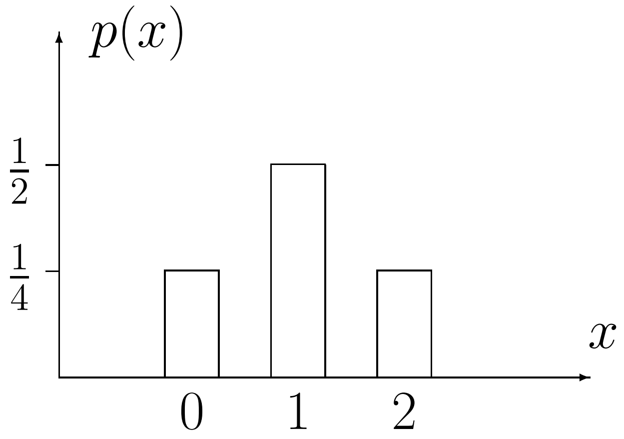
Poiché X deve assumere almeno uno dei valori x_i, abbiamo che

$$\sum_{i=1}^{\infty} p(x_i) = \sum_{i=1}^{\infty} P(X = x_i) = P\left(\bigcup_{i=1}^{\infty} \{X = x_i\}\right) = 1.$$

Può essere utile rappresentare la densità discreta in forma grafica ponendo i valori x_i in ascissa e $p(x_i)$ in ordinata. Per esempio, se la densità discreta di X è

$$p(0) = \frac{1}{4}, \qquad p(1) = \frac{1}{2}, \qquad p(2) = \frac{1}{4},$$

graficamente si ha



La densità discreta consente di calcolare la probabilità che la variabile aleatoria assuma valori in un sottoinsieme qualsiasi B di $\mathbb{R}$; ad esempio nell'intervallo $[a, b]$:

$$P(X \in B) = \sum_{k\colon x_k \in B} p(x_k), \qquad P(a \le X \le b) = \sum_{k\colon a \le x_k \le b} p(x_k).$$

Esempio. Sia X la variabile aleatoria discreta che descrive il numero di bit pari a **1** in un vettore booleano, di lunghezza n, scelto a caso.
(a) Calcolare $P(X = k)$, per $k = 0, 1, \ldots, n$.
(b) Verificare che $\sum_{k=0}^{n} P(X = k) = 1$.
(c) Calcolare $P(X \geq 1)$.

Soluzione. (a) Dividendo il numero di vettori booleani con k bit pari a **1** per il numero di vettori booleani di lunghezza n, si trae

$$P(X = k) = \frac{\binom{n}{k}}{2^n}, \qquad k = 0, 1, \ldots, n.$$

(b) Pertanto,

$$\sum_{k=0}^{n} P(X = k) = \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} = \frac{1}{2^n} \cdot 2^n = 1.$$

(c) Si ha infine

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{1}{2^n}.$$

Esempio. Sia $p(k) = c\,\dfrac{\lambda^k}{k!}$, $k = 0, 1, 2, \ldots$, dove λ è una costante positiva. Si calcoli
(a) $P(X = 0)$,
(b) $P(X > 2)$.

Soluzione. Per determinare c, imponendo che sia $\displaystyle\sum_{k=0}^{\infty} p(k) = 1$ abbiamo

$$1 = c \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = c\, e^{\lambda} \qquad \Longrightarrow \qquad c = e^{-\lambda},$$

avendo ricordato che $\boxed{\displaystyle\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x}$. Si ha quindi $p(k) = e^{-\lambda}\,\dfrac{\lambda^k}{k!}$, $k \geq 0$, e pertanto

(a) $P(X = 0) = p(0) = e^{-\lambda}\,\lambda^0/0! = e^{-\lambda}$,

(b) $P(X > 2) = 1 - P(0 \leq X \leq 2) = 1 - p(0) - p(1) - p(2) = 1 - e^{-\lambda}\left(1 + \lambda + \dfrac{\lambda^2}{2}\right)$.

Esempio. Determinare la densità della variabile aleatoria discreta X, che descrive il massimo che si ottiene lanciando 2 dadi.

Soluzione. Notiamo che $X = 1$ se esce la coppia $(1, 1)$, $X = 2$ se esce $(1, 2)$, $(2, 1)$ oppure $(2, 2)$, e così via. Posto $p(k) = P(X = k)$ risulta quindi

$$p(1) = \frac{1}{36}, \quad p(2) = \frac{3}{36}, \quad p(3) = \frac{5}{36}, \quad p(4) = \frac{7}{36}, \quad p(5) = \frac{9}{36}, \quad p(6) = \frac{11}{36},$$

ossia

$$p(k) = \frac{2k-1}{36}, \qquad k = 1, 2, \ldots, 6.$$

Notiamo che poiché

$$\boxed{\sum_{k=1}^{n} k = \frac{n(n+1)}{2}}$$

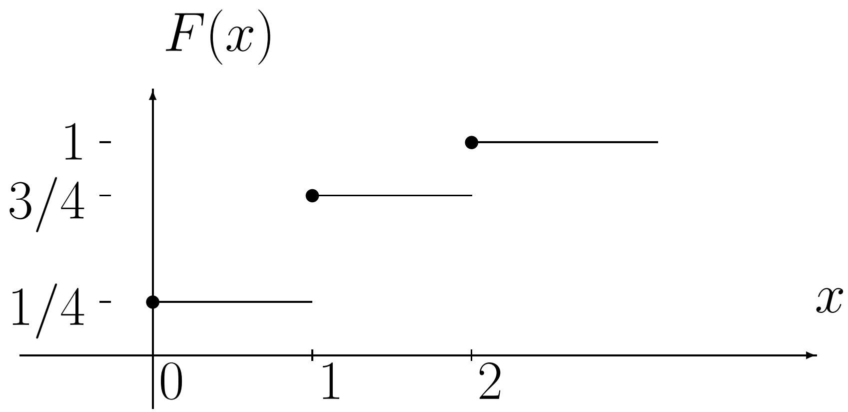
si trae

$$\sum_{k=1}^{6} p(k) = \sum_{k=1}^{6} \frac{2k-1}{36} = \frac{1}{18} \sum_{k=1}^{6} k - \frac{1}{6} = \frac{1}{18} \frac{6 \cdot 7}{2} - \frac{1}{6} = 1.$$

Se X è una variabile aleatoria discreta che assume i valori $x_1, x_2, x_3, \ldots$, dove $x_1 < x_2 < x_3 < \ldots$, allora la sua funzione di distribuzione $F(x)$ è costante negli intervalli $[x_{i-1}, x_i)$, ed in x_i ha un salto di ampiezza pari a $p(x_i)$. Quindi $F(x)$ può essere così espressa in termini della densità discreta:

$$F(x) = \sum_{k:\, x_k \le x} p(x_k) = \begin{cases} 0, & x < x_1 \\ p(x_1), & x_1 \le x < x_2 \\ p(x_1) + p(x_2), & x_2 \le x < x_3 \\ p(x_1) + p(x_2) + p(x_3), & x_3 \le x < x_4 \\ \vdots \end{cases}$$

Ad esempio, se X ha densità discreta $p(0) = \frac{1}{4}$, $p(1) = \frac{1}{2}$, $p(2) = \frac{1}{4}$, si ha



$$F(x) = \begin{cases} 0, & x < 0 \\ \dfrac{1}{4} & 0 \le x < 1 \\ \dfrac{3}{4} & 1 \le x < 2 \\ 1 & x \ge 2. \end{cases}$$

Esempio. In un gioco, in cui si lancia una moneta per 3 volte, si vincono k euro se esce testa per la prima volta al lancio k-esimo ($k = 1, 2, 3$), e si perdono c euro se non esce mai testa. Ricavare la funzione di distribuzione $F(x)$ della vincita, e determinare il valore di c affinché il gioco sia equo.

Soluzione. Indicando con X la vincita al gioco, e posto $p(k) = P(X = k)$, risulta

$$p(1) = P(T_1) = \frac{1}{2}, \qquad p(2) = P(\overline{T_1} \cap T_2) = \frac{1}{4},$$

$$p(3) = P(\overline{T_1} \cap \overline{T_2} \cap T_3) = \frac{1}{8}, \qquad p(-c) = P(\overline{T_1} \cap \overline{T_2} \cap \overline{T_3}) = \frac{1}{8},$$

Pertanto: $F(x) = 0$ per $x < -c$, $F(x) = 1/8$ per $-c \leq x < 1$, $F(x) = 5/8$ per $1 \leq x < 2$, $F(x) = 7/8$ per $2 \leq x < 3$, $F(x) = 1$ per $x \geq 3$. La vincita media è

$$\sum_k k \cdot p(k) = -c \cdot \frac{1}{8} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} = \frac{11 - c}{8}.$$

Ne segue che il gioco è equo se $c = 11$.

4.3 Valore atteso

Introduciamo uno dei più importanti concetti in calcolo delle probabilità.

Definizione. Se X è una variabile aleatoria discreta con densità discreta $p(x)$, il valore atteso (o valore medio, o speranza matematica) di X è definito da

$$E(X) = \sum_{x:\, p(x)>0} x \cdot p(x).$$

Il valore atteso di X è la media pesata di tutti i possibili valori che X può assumere, ognuno pesato con la probabilità che X lo assuma. Ad esempio, se $p(0) = p(1) = \frac{1}{2}$:

$$E(X) = 0 \cdot p(0) + 1 \cdot p(1) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

Se, invece, risulta $p(0) = \frac{1}{3}$, $p(1) = \frac{2}{3}$, allora

$$E(X) = 0 \cdot p(0) + 1 \cdot p(1) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.$$

Esempio. Calcolare il valore atteso nel lancio di un dado equilibrato.

Soluzione. Essendo $p(k) = \frac{1}{6}$ per $k = 1, 2, \ldots, 6$ otteniamo

$$E(X) = \sum_{k=1}^{6} k \cdot p(k) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = \frac{7}{2} = 3{,}5.$$

Esempio. Dato un evento A, definiamo la funzione indicatrice I_A di A come

$$I_A(\omega) = \begin{cases} 1 & \text{se } \omega \in A \\ 0 & \text{altrimenti.} \end{cases}$$

Si calcoli $E(I_A)$.

Soluzione. Poiché $p(1) = P(I_A = 1) = P(A)$ e $p(0) = P(I_A = 0) = P(\overline{A})$, abbiamo

$$E(I_A) = 0 \cdot p(0) + 1 \cdot p(1) = p(1) = P(A).$$

Pertanto, il valore atteso della variabile indicatrice di un evento è uguale alla probabilità dell'evento stesso.

Esempio. Il concorrente di un gioco a quiz deve rispondere a due domande, D_1 e D_2, ed è libero di scegliere in che ordine rispondere. Se risponde per prima alla domanda D_i gli sarà consentito di rispondere all'altra domanda solo se avrà risposto correttamente alla prima. Egli riceve V_i euro se risponde correttamente alla domanda D_i, $i = 1, 2$. Se la probabilità di $E_i = \{\text{conosce la risposta di } D_i\}$ è P_i, con E_1 e E_2 indipendenti per ipotesi, a quale domanda dovrà rispondere prima per massimizzare il guadagno atteso?

Soluzione. Sia X_i la vincita del concorrente se risponde prima a D_i $(i = 1, 2)$; si ha

$$\begin{array}{ll}
P(X_1 = 0) = 1 - P_1 & \qquad P(X_2 = 0) = 1 - P_2 \\
P(X_1 = V_1) = P_1\,(1 - P_2) & \qquad P(X_2 = V_2) = P_2\,(1 - P_1) \\
P(X_1 = V_1 + V_2) = P_1\,P_2 & \qquad P(X_2 = V_1 + V_2) = P_1\,P_2
\end{array}$$

con guadagno atteso $E[X_1] = V_1\,P_1\,(1 - P_2) + (V_1 + V_2)\,P_1\,P_2$ se risponde prima a D_1, mentre risulta $E[X_2] = V_2\,P_2\,(1 - P_1) + (V_1 + V_2)\,P_1\,P_2$ se risponde prima a D_2.

È più vantaggioso rispondere prima a D_1 se $V_1\,P_1\,(1 - P2) \geq V_2\,P_2\,(1 - P_1)$, ovvero se

$$\frac{V_1\,P_1}{1 - P_1} \geq \frac{V_2\,P_2}{1 - P_2}.$$

Esempio. Una comitiva di 120 studenti viene condotta in gita in 3 autobus. Nel primo ci sono 36 studenti, nel secondo 40, e nel terzo 44. All'arrivo si sceglie a caso uno studente tra i 120. Se X denota il numero di studenti che hanno viaggiato sull'autobus dello studente scelto a caso, calcolare $E(X)$.

Soluzione. Si ha $P(X = 36) = \frac{36}{120}$, $P(X = 40) = \frac{40}{120}$, $P(X = 44) = \frac{44}{120}$ e quindi

$$E(X) = 36 \cdot \frac{36}{120} + 40 \cdot \frac{40}{120} + 44 \cdot \frac{44}{120} = \frac{4832}{120} = 40{,}2667.$$

Il numero medio di studenti presenti su un autobus è $120/3 = 40$, che quindi è minore di $E(X)$. Questo fenomeno si verifica perché: più studenti sono presenti su un singolo autobus e più sarà probabile che lo studente scelto a caso provenga proprio da quello. Così si assegna peso maggiore agli autobus con più studenti.

Più in generale, se vi sono k autobus con $n_1, n_2, \ldots, n_k$ studenti, si ha

$$E(X) = \sum_{i=1}^{k} n_i \cdot \frac{n_i}{\sum_{j=1}^{k} n_j} = \frac{\sum_{i=1}^{k} n_i^2}{\sum_{j=1}^{k} n_j} \qquad \left(\geq \frac{\sum_{i=1}^{k} n_i}{k} \right)$$

Esempio. In seguito a trial clinici risulta che un farmaco sperimentale produce un miglioramento del 50% in pazienti gravi, un miglioramento del 5% in pazienti di media gravità e un peggioramento dell'1% in pazienti lievi. Si supponga che i pazienti affetti dalla patologia specifica siano per il 10% gravi, per il 15% medi e per il 75% lievi. Qual è il miglioramento medio prodotto dal farmaco sperimentale?

Soluzione. Descriviamo il miglioramento in percentuale prodotto dal farmaco sperimentale mediante una variabile aleatoria X discreta tale che

$$P(X = -1) = 0{,}75 \qquad P(X = 5) = 0{,}15 \qquad P(X = 50) = 0{,}10$$

Il miglioramento medio prodotto dal farmaco è del 5%, essendo

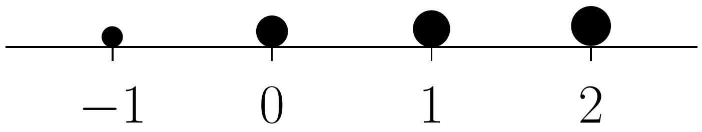
$$E(X) = \sum_x x \cdot P(X = x) = -1 \cdot 0{,}75 + 5 \cdot 0{,}15 + 50 \cdot 0{,}10 = 5.$$

Notiamo che se fosse stabilito che un farmaco si può porre in commercio se produce un miglioramento medio non minore del 5% in pazienti affetti da patologia specifica, quel farmaco sarebbe accettato sebbene causi un peggioramento nel 75% dei pazienti.

Il concetto di valore atteso è analogo al concetto fisico di *baricentro* di una distribuzione di masse. Sia X una variabile aleatoria discreta di densità discreta $p(x_i)$, $i \geq 1$. Immaginiamo una sbarra di peso trascurabile su cui sono poste delle masse

$$p(-1) = 0{,}10 \qquad p(0) = 0{,}25 \qquad p(1) = 0{,}30 \qquad p(2) = 0{,}35$$

nei punti $x_1 = -1$, $x_2 = 0$, $x_3 = 1$, $x_4 = 2$.



Allora il punto m nel quale la sbarra rimane in equilibrio è il centro di gravità, rispetto al quale è nulla la somma dei momenti dei pesi delle singole porzioni di massa, ossia

$$\sum_i [x_i - m]\, p(x_i) = 0 \qquad \Longrightarrow \qquad m = \frac{\sum_i x_i\, p(x_i)}{\sum_i p(x_i)} = \sum_i x_i\, p(x_i) = E(X).$$

Nell'esempio, $E(X) = \displaystyle\sum_{x=-1}^{2} x \cdot p(x) = -1 \cdot 0{,}10 + 0 \cdot 0{,}25 + 1 \cdot 0{,}30 + 2 \cdot 0{,}35 = 0{,}9.$

Esempio. Da un'urna contenente k biglie nere e $n-k$ biglie bianche si effettuano n estrazioni a caso senza reinserimento. Determinare $E(X)$, dove X denota l'estrazione in cui si estrae una biglia nera per la prima volta.

Soluzione. Poiché le estrazioni sono senza reinserimento l'evento $\{X=j\}$ si verifica se le prime $j-1$ biglie estratte sono bianche, la j-esima è nera, e nelle rimanenti $n-j$ estrazioni vi sono $k-1$ biglie nere. Quindi risulta

$$p(j)=P(X=j)=\frac{\binom{j-1}{0}\binom{1}{1}\binom{n-j}{k-1}}{\binom{n}{k}}=\frac{\binom{n-j}{k-1}}{\binom{n}{k}},\qquad j=1,2,\ldots,n-k+1.$$

Notiamo che

$$\sum_j p(j)=\frac{1}{\binom{n}{k}}\sum_{j=1}^{n-k+1}\binom{n-j}{k-1}=\frac{1}{\binom{n}{k}}\sum_{r=k-1}^{n-1}\binom{r}{k-1}=\frac{1}{\binom{n}{k}}\binom{n}{k}=1,$$

avendo posto $r=n-j$ e fatto uso dell'identità $\sum_{r=k}^{n}\binom{r}{k}=\binom{n+1}{k+1}$.

Si ha dunque

$$
\begin{aligned}
E(X) &= \sum_j j\,p(j) = \frac{1}{\binom{n}{k}} \sum_{j=1}^{n-k+1} j \binom{n-j}{k-1} = \frac{1}{\binom{n}{k}} \sum_{j=1}^{n-k+1} \sum_{h=1}^{j} \binom{n-j}{k-1} \\
&= \frac{1}{\binom{n}{k}} \sum_{h=1}^{n-k+1} \sum_{j=h}^{n-k+1} \binom{n-j}{k-1} = \frac{1}{\binom{n}{k}} \sum_{h=1}^{n-k+1} \sum_{r=k-1}^{n-h} \binom{r}{k-1},
\end{aligned}
$$

avendo posto $r = n - j$. Facendo uso dell'identità $\sum_{r=k}^{n} \binom{r}{k} = \binom{n+1}{k+1}$ segue

$$
E(X) = \frac{1}{\binom{n}{k}} \sum_{h=1}^{n-k+1} \binom{n-h+1}{k} = \frac{1}{\binom{n}{k}} \sum_{r=k}^{n} \binom{r}{k} = \frac{1}{\binom{n}{k}} \binom{n+1}{k+1},
$$

avendo posto $r = n-h+1$ e usato nuovamente l'identità $\sum_{r=k}^{n} \binom{r}{k} = \binom{n+1}{k+1}$. Pertanto,

$$
E(X) = \frac{n+1}{k+1}.
$$

4.4 Valore atteso di una funzione di una variabile aleatoria

Sia X una variabile aleatoria discreta di cui è nota la sua densità discreta $p(x_i)$.
Si desidera calcolare il valore atteso di una qualche funzione di X, diciamo $g(X)$.

Essendo $g(X)$ stessa una variabile aleatoria discreta, avrà una densità discreta che possiamo determinare conoscendo quella di X. Ricavata la densità discreta di $g(X)$ possiamo calcolare $E[g(X)]$ utilizzando la definizione di valore atteso.

Esempio. Sia X una variabile aleatoria che assuma i valori -1, 0, 1 con probabilità $P(X = -1) = 0{,}2 \quad P(X = 0) = 0{,}5 \quad P(X = 1) = 0{,}3$. Calcolare $E(X^2)$.

Soluzione. Sia $Y = X^2$; è immediato verificare che la densità discreta di Y è

$$P(Y = 1) = P(X^2 = 1) = P(X = \pm 1) = P(X = 1) + P(X = -1) = 0{,}5$$

$$P(Y = 0) = P(X^2 = 0) = P(X = 0) = 0{,}5$$

Quindi $E(X^2) = E(Y) = \sum_{y=0}^{1} y \cdot P(Y = y) = 0 \cdot 0{,}5 + 1 \cdot 0{,}5 = 0{,}5$.
Si noti che $0{,}5 = E(X^2) \neq [E(X)]^2 = [-1 \cdot 0{,}2 + 0 \cdot 0{,}5 + 1 \cdot 0{,}3]^2 = [0{,}1]^2 = 0{,}01$.

Se si considera che $g(X) = g(x)$ quando $X = x$, appare ragionevole che $E[g(X)]$ sia la media pesata dei valori $g(x)$, assegnando $P(X = x)$ come peso a $g(x)$.

Proposizione. Se X è una variabile aleatoria discreta che assume i valori x_i, $i \geq 1$, con probabilità $p(x_i)$, allora per ogni funzione a valori reali $g(x)$ risulta

$$E[g(X)] = \sum_i g(x_i)\, p(x_i).$$

Dimostrazione. Denotiamo con y_j, $j \geq 1$, i diversi valori di $g(x_i)$, $i \geq 1$. Allora, raggruppando tutti gli x_i che hanno stesso valore, abbiamo

$$\begin{aligned} \sum_i g(x_i)\, p(x_i) &= \sum_j \sum_{i:\, g(x_i)=y_j} g(x_i)\, p(x_i) = \sum_j y_j \sum_{i:\, g(x_i)=y_j} p(x_i) \\ &= \sum_j y_j\, P[g(X) = y_j] = E[g(X)], \end{aligned}$$

avendo fatto uso dell'identità $\boxed{P(X \in B) = \sum_{k:\, x_k \in B} p(x_k)}$ per $B = \{x_i : g(x_i) = y_j\}$.

Esempio. Sia X una variabile aleatoria che assume i valori -1, 0, 1 con probabilità $P(X = -1) = 0{,}2 \quad P(X = 0) = 0{,}5 \quad P(X = 1) = 0{,}3$. Calcolare $E(X^2)$ facendo uso di $E[g(X)] = \sum_i g(x_i)\, p(x_i)$.

Soluzione. Otteniamo $E(X^2) = \displaystyle\sum_i x_i^2\, p(x_i) = (-1)^2 \cdot 0{,}2 + 0^2 \cdot 0{,}5 + 1^2 \cdot 0{,}3 = 0{,}5$.

Esempio. Sia X una variabile aleatoria che descrive il risultato del lancio di un dado non truccato. Calcolare $E[g(X)]$, con $g(x) = \min\{x, 4\}$.

Soluzione. Si ha

$$E[g(X)] = \sum_i g(x_i)\, p(x_i) = \sum_{k=1}^{6} \min\{k, 4\}\, p(k) = \frac{1}{6}\,(1 + 2 + 3 + 4 + 4 + 4) = 3.$$

Esempio. Sia X la variabile aleatoria tale che $P(X = 0) = 1 - p$ e $P(X = 1) = p$, $0 \le p \le 1$. Calcolare il minimo di $\psi(a) := E[(X - a)^2]$, per $a \in \mathbb{R}$.

Soluzione. Risulta $\psi(a) = E[(X-a)^2] = \sum_{k=0}^{1}(k-a)^2\, p(k) = a^2\,(1-p)+(1-a)^2\, p$ e quindi $\psi'(a) = 2a(1 - p) - 2(1 - a)p = 0$ per $a = p$. Quindi il minimo di $\psi(a)$ è p.

Proposizione. (Proprietà di linearità) Sia X una variabile aleatoria discreta. Se a e b sono costanti reali, allora

$$E(a\,X+b)=a\,E(X)+b.$$

Dimostrazione. Ricordando che $E[g(X)]=\sum_i g(x_i)\,p(x_i)$, per $g(x)=a\,x+b$ è

$$E[g(X)]=\sum_i (a\,x_i+b)\,p(x_i)=a\sum_i x_i\,p(x_i)+b\sum_i p(x_i)=a\,E(X)+b,$$

avendo fatto uso dell'identità $\sum_i p(x_i)=1$.

Definizione. Sia X una variabile aleatoria discreta; la quantità $E(X^n)$, $n\geq 1$, è detta *momento di ordine* n di X. Risulta

$$E(X^n)=\sum_{x:\,p(x)>0} x^n\,p(x).$$

Segue che il valore atteso di X è anche il momento di ordine 1 di X.

Esempio. Sia X una variabile aleatoria che assuma i valori -1, 0, 1 con probabilità $P(X=-1)=0{,}2 \quad P(X=0)=0{,}5 \quad P(X=1)=0{,}3$. Calcolare $E(X^n)$.

Soluzione. Il momento di ordine n di X è dato da

$$E(X^n)=\sum_{x=-1}^{1} x^n\, P(X=x)=(-1)^n\, 0{,}2+0^n\, 0{,}5+1^n\, 0{,}3=(-1)^n\, 0{,}2+0{,}3$$

quindi $E(X^n)=0{,}5$ per n pari e $E(X^n)=0{,}1$ per n dispari.

Esempio. Determinare il momento di ordine n della variabile aleatoria $Y=aX+b$, con a e b costanti reali.

Soluzione. Facendo uso del teorema del binomio si ha

$$E[Y^n]=E[(aX+b)^n]=E\left[\sum_{k=0}^{n}\binom{n}{k}a^kX^kb^{n-k}\right]$$

e pertanto, per la proprietà di linearità,

$$E[Y^n]=\sum_{k=0}^{n}\binom{n}{k}a^kb^{n-k}E[X^k].$$

4.5 Varianza

Sebbene il valore atteso fornisca una media pesata dei possibili valori di una variabile aleatoria, esso non dà alcuna informazione riguardo alla variabilità, o dispersione, di questi valori. Per esempio, le variabili aleatorie date da

$$W = 0 \;\text{ con prob. } 1, \qquad Y = \begin{cases} -1 \text{ con prob. } 1/2 \\ +1 \text{ con prob. } 1/2, \end{cases} \qquad Z = \begin{cases} -100 \text{ con prob. } 1/2 \\ +100 \text{ con prob. } 1/2 \end{cases}$$

hanno lo stesso valore atteso, pari a 0, ma è presente maggiore dispersione nei valori di Y piuttosto che in quelli di W e nei valori di Z rispetto a Y.

Poiché ci si attende che X assuma valori disposti attorno al suo valore atteso $E(X)$, appare ragionevole misurare la variabilità di X mediante la media della distanza dal valor medio che i possibili valori di X assumono, ad esempio considerando la quantità $E(|X - \mu|)$, dove $\mu = E(X)$. Per superare alcune difficoltà di tipo matematico si preferisce invece adoperare la media delle differenze al quadrato tra X ed il suo valore atteso μ, e ciò conduce alla seguente definizione.

Definizione. Sia X una variabile aleatoria discreta di valore atteso μ; la varianza di X, denotata con $\mathrm{Var}(X)$, è così definita:

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \sum_{x:\, p(x)>0} (x-\mu)^2\, p(x).$$

Una formula alternativa per la varianza si ottiene usando la proprietà di linearità:

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

La maniera più semplice di valutare la varianza di X consiste quindi nel calcolare la differenza tra il momento del secondo ordine di X e il quadrato del suo valore atteso.

Esempio. Calcolare $\mathrm{Var}(X)$, dove X rappresenta l'esito del lancio di un dado.
Soluzione. Essendo $p(k) = \frac{1}{6}$ per $k = 1, 2, \ldots, 6$ e $E(X) = \frac{7}{2}$ otteniamo

$$\mathrm{Var}(X) = E(X^2) - \mu^2 = \sum_{k=1}^{6} k^2 \cdot p(k) - \left(\frac{7}{2}\right)^2 = \frac{1}{6}\cdot 91 - \frac{49}{4} = \frac{35}{12} \approx 2{,}9167.$$

Proposizione. (Proprietà della varianza) Sia X una variabile aleatoria discreta. Se a e b sono costanti reali, allora

$$\mathrm{Var}(a\,X+b)=a^2\,\mathrm{Var}(X).$$

Dimostrazione. Ricordando che $E(a\,X+b)=a\,E(X)+b$, si ha

$$\begin{aligned}\mathrm{Var}(a\,X+b) &= E[\{a\,X+b-E(a\,X+b)\}^2]=E[\{a\,X+b-a\,E(X)-b\}^2]\\ &= a^2\,E[\{X-E(X)\}^2]=a^2\,\mathrm{Var}(X).\end{aligned}$$

Si noti che la varianza corrisponde al momento d'inerzia per una distribuzione di masse.

Notiamo che $E(X)$ ha stesse unità di misura di X, mentre $\mathrm{Var}(X)$ ha stesse unità di misura di X^2. La seguente misura di variabilità ha invece stesse unità di X.

Definizione. Sia X una variabile aleatoria discreta; la radice quadrata di $\mathrm{Var}(X)$, che denotiamo con σ_X, è detta *deviazione standard* di X. Cioè

$$\sigma_X=\sqrt{\mathrm{Var}(X)}=\sqrt{E[(X-\mu)^2]}.$$

4.6 Le variabili aleatorie di Bernoulli e binomiali

Supponiamo di eseguire un esperimento i cui possibili esiti appartengono a due categorie, ovvero possono essere classificati come *successo* e *insuccesso*.

Poniamo $X = 1$ quando l'esito è un successo, e $X = 0$ quando l'esito è un insuccesso. La densità discreta della variabile aleatoria X è quindi

$$p(0) = P(X = 0) = 1 - p, \qquad p(1) = P(X = 1) = p \qquad (0 \leq p \leq 1)$$

dove p rappresenta la probabilità che la prova abbia avuto successo.

La variabile aleatoria X siffatta è detta variabile aleatoria di Bernoulli; la sua densità discreta può essere così espressa in forma compatta:

$$p(x) = p^x\,(1 - p)^{1-x}, \qquad x = 0, 1.$$

Notiamo che per una variabile aleatoria di Bernoulli X la funzione di distribuzione è

$$F(x) = \sum_{k \le x} p(k) = \sum_{k \le x} p^k \, (1-p)^{1-k} = \begin{cases} 0 & x < 0 \\ 1-p & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

e il momento di ordine n è dato da

$$E[X^n] = \sum_{x=0}^{1} x^n \, p(x) = \sum_{x=0}^{1} x^n \, p^x \, (1-p)^{1-x} = p.$$

Ne segue che valore atteso e varianza sono

$$E[X] = p$$

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = p(1-p).$$

Notiamo che $\mathrm{Var}(X) = 0$ se $p = 0$ e $p = 1$, mentre $\mathrm{Var}(X)$ è massima per $p = 1/2$.

Supponiamo ora di eseguire n prove in maniera indipendente, ognuna avente come possibili risultati successo (con probabilità p) e insuccesso (con probabilità $1-p$).

Sia X il numero di successi che si ottengono nelle n prove. Allora, X è detta variabile aleatoria binomiale di parametri $(n;p)$. Quindi, una variabile aleatoria di Bernoulli è semplicemente una variabile aleatoria binomiale di parametri $(1;p)$.

La densità discreta di una variabile aleatoria binomiale di parametri $(n;p)$ è data da

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n \qquad (0 \leq p \leq 1).$$

Ogni sequenza di n esiti contenenti x successi e $n-x$ insuccessi si verifica con probabilità $p^x (1-p)^{n-x}$, grazie all'ipotesi di indipendenza delle prove. La densità discreta $p(x)$ segue allora dal fatto che ci sono $\binom{n}{x}$ differenti sequenze di n esiti contenenti x successi e $n-x$ insuccessi, essendoci esattamente $\binom{n}{x}$ differenti modi di scegliere le x prove in cui si verifichino i successi. Per esempio, ci sono $\binom{4}{2} = 6$ differenti modi di avere $x = 2$ successi in $n = 4$ prove: $(ssii)$, $(sisi)$, $(siis)$, $(issi)$, $(isis)$, $(iiss)$.

Dal teorema del binomio segue facilmente che le probabilità date dalla densità discreta

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n$$

danno somma 1 per ogni $p \in [0, 1]$; infatti

$$\sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1^n = 1.$$

Esempio. Determinare la densità discreta della variabile aleatoria che descrive il numero di volte che esce testa in 5 lanci indipendenti di una moneta non truccata.

Soluzione. La variabile aleatoria in questione è binomiale di parametri $(5; \frac{1}{2})$ quindi

$p(x)$	$p(0)$	$p(1)$	$p(2)$	$p(3)$	$p(4)$	$p(5)$
$\binom{5}{x} \left(\frac{1}{2}\right)^x \left(1-\frac{1}{2}\right)^{5-x} = \binom{5}{x} \frac{1}{2^5} = \binom{5}{x} \frac{1}{32}$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

Esempio. Una fabbrica produce viti che presentano un difetto, ognuna indipendentemente dalle altre, con probabilità 0,01. Se le viti sono vendute in confezioni da 10, qual è la probabilità che un pacchetto contenga almeno due viti difettose? E se le confezioni sono da 20 viti?

Soluzione. Sia X il numero di viti difettose in una confezione da 10; essendo binomiale di parametri $(10; 0{,}01)$ la probabilità che un pacchetto abbia almeno due viti difettose è

$$\begin{aligned} P(X \geq 2) &= 1 - P(X=0) - P(X=1) \\ &= 1 - \binom{10}{0} (0{,}01)^0 \, (0{,}99)^{10} - \binom{10}{1} (0{,}01)^1 \, (0{,}99)^9 \approx 0{,}004. \end{aligned}$$

Se le confezioni sono da 20 viti il risultato è

$$P(X \geq 2) = 1 - \binom{20}{0} (0{,}01)^0 \, (0{,}99)^{20} - \binom{20}{1} (0{,}01)^1 \, (0{,}99)^{19} \approx 0{,}017.$$

Esempio. In un gioco d'azzardo un giocatore scommette su uno dei numeri compresi tra 1 e 6. Si lanciano 3 dadi. Se il numero su cui ha scommesso appare k volte, con $k = 1, 2, 3$, allora il giocatore vince vince k euro. Se invece il numero non esce, allora il giocatore perde un euro. Il gioco è equo? Ovvero, il capitale finale atteso è zero?

Soluzione. Il numero di dadi che mostra il numero su cui si è puntato è una variabile aleatoria binomiale di parametri $(3; \frac{1}{6})$. Quindi, denotando con X la vincita si ha

$$P(X=-1)=\binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3=\frac{125}{216}, \qquad P(X=1)=\binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2=\frac{75}{216},$$

$$P(X=2)=\binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1=\frac{15}{216}, \qquad P(X=3)=\binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0=\frac{1}{216},$$

da cui segue che la vincita attesa è

$$E(X)=\sum_{x:\,p(x)>0} x\,p(x)=-1\cdot\frac{125}{216}+1\cdot\frac{75}{216}+2\cdot\frac{15}{216}+3\cdot\frac{1}{216}=-\frac{17}{216}\approx -0{,}079.$$

Quindi, giocando ripetutamente, ci si attende di perdere 17 euro ogni 216 partite.

Proposizione. Se X è una variabile aleatoria binomiale di parametri $(n;p)$, allora

$$E(X)=n\,p,\qquad \mathrm{Var}(X)=n\,p\,(1-p).$$

Dimostrazione. Determiniamo il momento di X di ordine k:

$$E(X^k)=\sum_{x:\,p(x)>0}x^k\cdot p(x)=\sum_{i=0}^{n}i^k\cdot\binom{n}{i}p^i\,(1-p)^{n-i}=\sum_{i=1}^{n}i^k\cdot\binom{n}{i}p^i\,(1-p)^{n-i}.$$

Utilizzando l'identità

$$i\binom{n}{i}=i\,\frac{n!}{i!\,(n-i)!}=n\,\frac{(n-1)!}{(i-1)!\,(n-i)!}=n\binom{n-1}{i-1}$$

ricaviamo che

$$E(X^k)=np\sum_{i=1}^{n}i^{k-1}\binom{n-1}{i-1}p^{i-1}\,(1-p)^{n-i}=np\sum_{j=0}^{n-1}(j+1)^{k-1}\binom{n-1}{j}p^j\,(1-p)^{n-1-j}$$

avendo posto $j=i-1$. Si ha quindi

$$E(X^k)=n\,p\,E[(Y+1)^{k-1}]$$

dove Y è una variabile aleatoria binomiale di parametri $(n-1;p)$.

Ponendo $k = 1$ nella formula $E(X^k) = n\,p\,E[(Y+1)^{k-1}]$ ricaviamo

$$E(X) = n\,p$$

così il valore atteso di successi che si verificano in n prove indipendenti quando la probabilità di successo vale p, è pari a $n\,p$. Quindi risulta $E(Y+1) = (n-1)\,p+1$. Pertanto, ponendo $k = 2$ nella formula $E(X^k) = n\,p\,E[(Y+1)^{k-1}]$ ricaviamo

$$E(X^2) = n\,p\,E(Y+1) = n\,p\,[(n-1)\,p+1].$$

Ricordando che $E(X) = n\,p$ si ottiene infine

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = n\,p\,[(n-1)\,p+1] - n^2\,p^2 = n\,p\,(1-p),$$

così la varianza del numero di successi che si verificano in n prove indipendenti quando la probabilità di successo vale p, è pari a $n\,p\,(1-p)$.

Esempio. Se X è il numero di successi che si verificano in n prove indipendenti quando la probabilità di successo vale p, determinare valore atteso e varianza della frequenza relativa $F_n = X/n$ del numero di successi

Soluzione. Poiché X è una variabile aleatoria binomiale di parametri $(n; p)$, risulta $E(X) = n\,p$ e $\mathrm{Var}(X) = n\,p\,(1-p)$. Pertanto, ricordando la proprietà di linearità del valore atteso e la proprietà della varianza, si ha

$$E(F_n) = E\left(\frac{X}{n}\right) = \frac{1}{n}\,E(X) = \frac{1}{n}\,n\,p = p,$$

$$\mathrm{Var}(F_n) = \mathrm{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2}\,\mathrm{Var}(X) = \frac{1}{n^2}\,n\,p\,(1-p) = \frac{p\,(1-p)}{n}.$$

Notiamo, in particolare, che $E(F_n)$ è costante in n, mentre $\mathrm{Var}(F_n)$ è decrescente e tende a 0 quando n tende a $+\infty$.

Proposizione. Se X è una variabile aleatoria binomiale di parametri $(n;p)$, con $0<p<1$, allora per $k=0,1,\ldots,n$ la densità discreta sarà inizialmente strettamente crescente e successivamente strettamente decrescente, con massimo in corrispondenza del più grande intero $k \leq (n+1)p$.

Dimostrazione. Si ha

$$\frac{P(X=k)}{P(X=k-1)} = \frac{\dfrac{n!}{k!\,(n-k)!}\,p^k\,(1-p)^{n-k}}{\dfrac{n!}{(k-1)!\,(n-k+1)!}\,p^{k-1}\,(1-p)^{n-k+1}} = \frac{(n-k+1)\,p}{k\,(1-p)}.$$

Quindi $P(X=k) \geq P(X=k-1)$ se e solo se

$$(n-k+1)\,p \geq k\,(1-p)$$

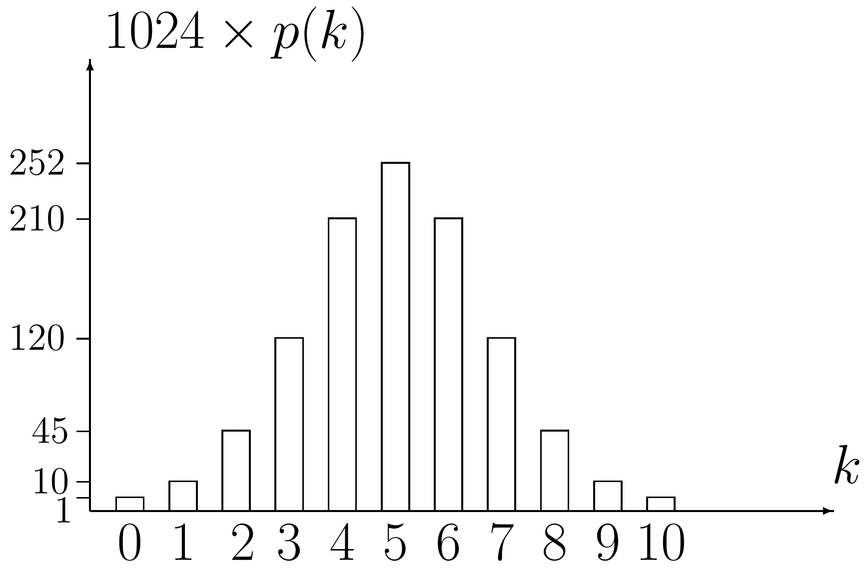
ossia

$$k \leq (n+1)p.$$

Nel caso di una variabile binomiale di parametri $(10, \frac{1}{2})$ la densità discreta

$$p(k) = \binom{10}{k} \left(\frac{1}{2}\right)^{10}$$

è strettamente crescente per $k \le 5$, ed è simmetrica: $p(k) = p(10-k) \;\; \forall k$.



Proposizione. Se X e Y sono variabili aleatorie binomiali di parametri $(n; p)$ e $(n; 1-p)$, rispettivamente, allora

$$P(X = k) = P(Y = n-k), \qquad k = 0, 1, \ldots, n.$$

Dimostrazione. Si ricava immediatamente notando che:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{n-k} (1-p)^{n-k} p^k = P(Y = n-k).$$

Durante il "Progetto Manhattan" (per la creazione della bomba nucleare), il celebre fisico Enrico Fermi chiese al Gen. Groves, direttore del progetto, quale fosse la definizione di un "grande" generale. Groves replicò che ogni generale che ha vinto 5 battaglie in sequenza può sicuramente definirsi grande. Fermi allora chiese quanti generali fossero grandi. Groves rispose circa 3 su 100. Fermi congetturò che considerando che schieramenti opposti per molte battaglie sono approssimativamente di uguali forze, la probabilità di vincere una battaglia è 1/2 e la probabilità di vincere 5 battaglie in sequenza è $(1/2)^5 = 1/32 = 0{,}03125$. Così concluse: "Quindi hai ragione Generale, circa 3 su 100. Probabilità matematica, non genio."

La soluzione di Fermi, nell'ipotesi di indipendenza tra battaglie, è la probabilità di avere 5 successi per una variabile aleatoria binomiale di parametri $(5, \frac{1}{2})$:

$$P(X = 5) = \binom{5}{5}\left(\frac{1}{2}\right)^5\left(1 - \frac{1}{2}\right)^0 = \frac{1}{32}.$$

Esempio. In un sistema elaborativo l'unità centrale prova a connettersi con n unità periferiche, ogni prova avente successo con probabilità p. Determinare media e varianza del numero totale di unità connesse, inclusa la centrale. Se una risorsa viene condivisa tra le unità connesse, determinare la frazione attesa di risorsa per ogni unità.

Soluzione. Il numero totale di unità connesse è $X+1$, con X variabile aleatoria binomiale di parametri n e p. Si ha quindi: $E(X+1) = E(X)+1 = np+1$ e $Var(X+1) = Var(X) = np(1-p)$. La frazione attesa di risorsa per ogni unità è

$$E\left(\frac{1}{X+1}\right) = \sum_{k=0}^{n} \frac{1}{k+1}\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^{n} \frac{1}{k+1}\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

$$= \frac{1}{n+1}\sum_{k=0}^{n} \frac{(n+1)!}{(k+1)!(n-k)!}p^k(1-p)^{n-k} = \frac{1}{n+1}\frac{1}{p}\sum_{j=1}^{n+1}\binom{n+1}{j}p^j(1-p)^{n+1-j}$$

$$= \frac{1-(1-p)^{n+1}}{(n+1)p} \neq \frac{1}{E(X+1)} \qquad \text{(per } j=k+1 \text{ e per la formula del binomio).}$$

4.7 La variabile aleatoria di Poisson

Una variabile aleatoria X, che assuma i valori $0, 1, 2, \ldots$, è detta variabile aleatoria di Poisson con parametro $\lambda > 0$ se

$$p(k) = P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

Notiamo che

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!} > 0 \quad \forall k;$$

inoltre si ha

$$\sum_{k=0}^{\infty} p(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1,$$

essendo

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}.$$

La variabile aleatoria di Poisson può essere utilizzata come approssimazione di una variabile aleatoria binomiale X di parametri $(n; p)$, quando n è grande e p è piccolo in modo che il prodotto $n\,p$ tenda ad un valore positivo finito. Sia $\lambda = n\,p$; allora

$$\begin{aligned} P(X=k) &= \frac{n!}{k!\,(n-k)!}\,p^k\,(1-p)^{n-k} \\ &= \frac{(n)_k}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!}\cdot\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\cdot\frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}. \end{aligned}$$

Per n grande risulta

$$\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\approx 1, \qquad \left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \qquad \left(1-\frac{\lambda}{n}\right)^k \approx 1.$$

Pertanto, quando n è grande e p è piccolo in modo che $n\,p \to \lambda > 0$ si ha

$$P(X=k)\approx \frac{\lambda^k}{k!}\,e^{-\lambda} \qquad (k=0,1,\ldots).$$

Illustriamo il significato dell'approssimazione

$$P(X=k) \approx \frac{\lambda^k}{k!}\, e^{-\lambda} \qquad (k=0,1,\ldots).$$

Se si eseguono n prove indipendenti, ognuna che dia successo con probabilità p, allora per n grande e p piccolo in modo che $n\,p$ sia un valore positivo finito, il numero totale di successi è ben approssimato da una variabile aleatoria di Poisson di parametro $\lambda = n\,p$.

Ad esempio, se $n=90$ e $p=1/18$ si ha $\lambda=5$ e quindi

$$P(X=k) = \binom{90}{k}\left(\frac{1}{18}\right)^k \left(1-\frac{1}{18}\right)^{90-k} \qquad \approx \frac{(5)^k}{k!}\, e^{-5};$$

$$P(X=0) = \left(1-\frac{1}{18}\right)^{90} = 0{,}0058 \qquad \approx e^{-5} = 0{,}0067.$$

Se $n=900$ e $p=1/180$ si ha ancora $\lambda=5$ e quindi

$$P(X=0) = \left(1-\frac{1}{180}\right)^{900} = 0{,}0066 \qquad \approx e^{-5} = 0{,}0067.$$

Esempio. Supponiamo che il numero di errori tipografici di una pagina di un libro sia descritto da una variabile aleatoria X di Poisson con parametro $\lambda = \frac{1}{2}$. Calcolare la probabilità che ci sia almeno un errore in una pagina fissata.
Soluzione. Si ha $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1/2} \approx 0{,}393$.

Esempio. Supponiamo che un pezzo prodotto da un macchinario sia difettoso con probabilità pari a 0,1. Determinare la probabilità che un lotto di 10 pezzi ne contenga al più uno difettoso.
Soluzione. La probabilità desiderata è $\binom{10}{0}(0{,}1)^0(0{,}9)^{10} + \binom{10}{1}(0{,}1)^1(0{,}9)^9 = 0{,}7361$, mentre l'approssimazione di Poisson fornisce $\frac{1^0}{0!}e^{-1} + \frac{1^1}{1!}e^{-1} = 2\,e^{-1} = 0{,}7358$.

Esempio. Il numero di richieste di stampa che giunge ad una stampante aziendale è in media 3,2 al minuto. Approssimare la probabilità che non giungano più di 2 richieste.
Soluzione. Pensiamo che le richieste di stampa giungono da un grande numero n di utenti, ognuno dei quali ha probabilità $3{,}2/n$ di fare una richiesta al minuto; allora, per l'approssimazione di Poisson, $P(X \leq 2) = e^{-3,2} + 3{,}2\,e^{-3,2} + \frac{(3{,}2)^2}{2}\,e^{-3,2} \approx 0{,}3799$.

Proposizione. Se X è una variabile aleatoria di Poisson di parametro λ, allora

$$E(X) = \lambda, \qquad \mathrm{Var}(X) = \lambda.$$

Dimostrazione. Si ha $E(X) = \displaystyle\sum_{x=0}^{\infty} x\, p(x) = \sum_{x=0}^{\infty} x\, e^{-\lambda} \frac{\lambda^x}{x!} = \lambda\, e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$.

Ponendo $k = x - 1$ e ricordando che $\displaystyle\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$ si ottiene

$$E(X) = \lambda\, e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda\, e^{-\lambda}\, e^{\lambda} = \lambda.$$

Risulta poi $E(X^2) = \displaystyle\sum_{x=0}^{\infty} x^2\, p(x) = \sum_{x=0}^{\infty} x^2\, e^{-\lambda} \frac{\lambda^x}{x!} = \lambda\, e^{-\lambda} \sum_{x=1}^{\infty} x\, \frac{\lambda^{x-1}}{(x-1)!}$.

Per $k = x - 1$ si ha: $E(X^2) = \lambda\, e^{-\lambda} \displaystyle\sum_{k=0}^{\infty} (k+1)\, \frac{\lambda^k}{k!} = \lambda\, [E(X) + E(1)] = \lambda\, (\lambda + 1)$.

Pertanto, $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \lambda\, (\lambda+1) - \lambda^2 = \lambda$.

4.8 Ulteriori distribuzioni di probabilità discrete

La variabile aleatoria geometrica

Supponiamo di ripetere in maniera indipendente una prova, che abbia probabilità di successo p, con $0 < p < 1$, fintanto che non si verifica il primo successo. Denotando con X il numero di prove necessarie per ottenere il primo successo, allora

$$P(X = n) = (1 - p)^{n-1}p, \qquad n = 1, 2, \ldots$$

Invero, si ha $X = n$ se le prime $n-1$ prove siano state un insuccesso e l'n-esima prova un successo. La formula si ricava anche per l'ipotesi di indipendenza tra le prove.

Risulta:

$$P(X = n) = (1-p)^{n-1}p > 0, \qquad \sum_{n=1}^{\infty} P(X = n) = p \sum_{n=1}^{\infty}(1-p)^{n-1} = \frac{p}{1-(1-p)} = 1,$$

avendo ricordato che $1 + c + c^2 + c^3 + \ldots = 1/(1-c)$ se $-1 < c < 1$.
La variabile aleatoria X è detta variabile aleatoria geometrica di parametro p.

Esempio. Un'urna contiene N biglie bianche e M biglie nere. Si estrae a caso una biglia alla volta, con reinserimento, fino a che non esce la prima biglia nera. Calcolare la probabilità che si debbano estrarre (a) esattamente n biglie, (b) più di k biglie.

Soluzione. Sia X il numero di biglie che si estraggono per ottenere la prima biglia nera. Allora X ha distribuzione geometrica di parametro $p = M/(M+N)$. Quindi

$$
\begin{aligned}
\text{(a)} \qquad & P(X = n) = (1-p)^{n-1}\, p; \\
\text{(b)} \qquad & P(X > k) = p \sum_{n=k+1}^{\infty} (1-p)^{n-1} = p\,(1-p)^k \sum_{j=0}^{\infty} (1-p)^j \\
& \qquad = p\,(1-p)^k \,\frac{1}{1-(1-p)} = (1-p)^k,
\end{aligned}
$$

avendo posto $j = n - k - 1$ e avendo ricordato che, per $-1 < c < 1$,

$$\sum_{j=0}^{\infty} c^j = \frac{1}{1-c}.$$

Notiamo che la probabilità $P(X > k) = (1-p)^k$ si può calcolare anche direttamente, poiché la probabilità che siano necessarie più di k prove per ottenere il primo successo è uguale alla probabilità che le prime k prove diano come esito k insuccessi.

Quindi, per una variabile aleatoria geometrica vale $P(X > k) = (1-p)^k$, $k \geq 0$.

Proposizione. (Proprietà di assenza di memoria) Se X è una variabile aleatoria geometrica di parametro p, allora per ogni $k, n \geq 0$ risulta

$$P(X > n + k | X > k) = P(X > n).$$

Dimostrazione. Per ogni $k, n \geq 0$ si ha

$$P(X > n + k | X > k) = \frac{P(\{X > n + k\} \cap \{X > k\})}{P(X > k)} = \frac{P(X > n + k)}{P(X > k)}$$

$$= \frac{(1-p)^{n+k}}{(1-p)^k} = (1-p)^n = P(X > n).$$

Esempio. Si lancia a caso una moneta. Calcolare la probabilità che esca testa per la prima volta dopo il 5º lancio sapendo che nei primi 3 lanci non esce testa.

Soluzione. Essendo $p = 1/2$, per la proprietà di assenza di memoria si ha:

$$P(X > 5 | X > 3) = P(X > 2) = \left(\frac{1}{2}\right)^2 = \frac{1}{4}.$$

Proposizione. Se X è una variabile aleatoria geometrica di parametro p, allora

$$E(X) = \frac{1}{p}, \qquad \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

Dimostrazione. Posto $q = 1 - p$ abbiamo che:

$$E(X) = \sum_{n=0}^{\infty} n\, q^{n-1} p = p \sum_{n=1}^{\infty} \frac{d}{dq}(q^n) = p \frac{d}{dq}\left(\sum_{n=0}^{\infty} q^n\right).$$

Utilizzando l'identità $\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}$ ricaviamo il valore atteso

$$E(X) = p \frac{d}{dq}\left(\frac{1}{1-q}\right) = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Calcoliamo ora il momento del secondo ordine di X:

$$E(X^2) = \sum_{n=0}^{\infty} n^2\, q^{n-1} p = p \sum_{n=1}^{\infty} \frac{d}{dq}(nq^n) = p\frac{d}{dq}\left(\sum_{n=1}^{\infty} nq^n\right) = p\frac{d}{dq}\left(\frac{q}{p}\sum_{n=1}^{\infty} nq^{n-1}p\right).$$

Ricordando che $E(X) = \sum_{n=1}^{\infty} n\, q^{n-1} p = \frac{1}{p}$, si ha

$$E(X^2) = p\frac{d}{dq}\left[\frac{q}{(1-q)^2}\right] = p\left[\frac{(1-q)^2 + 2q(1-q)}{(1-q)^4}\right] = p\left[\frac{1}{p^2} + \frac{2(1-p)}{p^3}\right] = \frac{2}{p^2} - \frac{1}{p}.$$

Infine si ottiene la varianza:

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

In particolare, se si ripete una serie di prove indipendenti, ognuna con probabilità di successo p, fino a che non si verifica il primo successo, allora il numero atteso di prove è uguale a $1/p$. Ad esempio, per un fissato valore nel lancio di un dado è $p = \frac{1}{6}$ e quindi $E(X) = 6$; per il singolo estratto nel gioco del lotto è $p = \frac{1}{18}$ e quindi $E(X) = 18$.

Esempio. Calcolare la probabilità che una variabile aleatoria geometrica X assuma valore non maggiore del suo valore atteso, se $1/p$ è intero. Valutare tale probabilità quando $p \to 1^-$ e quando $p \to 0^+$.

Soluzione. Ricordando che $E(X) = 1/p$ e che $P(X \leq k) = 1 - (1-p)^k$ si ha

$$P\{X \leq E(X)\} = 1 - (1-p)^{1/p}.$$

Pertanto

$$\lim_{p \to 1^-} P\{X \leq E(X)\} = 1 - \lim_{p \to 1^-} (1-p)^{1/p} = 1,$$

$$\lim_{p \to 0^+} P\{X \leq E(X)\} = 1 - \lim_{p \to 0^+} (1-p)^{1/p} = 1 - \lim_{x \to \infty} \left(1 - \frac{1}{x}\right)^x = 1 - e^{-1} \approx 0{,}632.$$

La variabile aleatoria ipergeometrica

Nell'estrarre n biglie senza reinserimento da un'urna che contiene N biglie, di cui m sono bianche e $N-m$ nere, sia X il numero di biglie bianche presenti tra le n estratte. Allora

$$P(X=k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \qquad k=0,1,\ldots,n.$$

Una variabile aleatoria dotata di tale densità discreta per opportuni valori n, N, m è detta variabile aleatoria *ipergeometrica*.

Ricordando che $\binom{r}{k} > 0$ quando $0 \le k \le r$, risulta $\binom{m}{k}\binom{N-m}{n-k} > 0$ per $\begin{cases} 0 \le k \le m \\ 0 \le n-k \le N-m \end{cases}$ ossia $\begin{cases} 0 \le k \le m \\ n+m-N \le k \le n \end{cases}$ da cui segue che

$$P(X=k) > 0 \quad \text{se} \quad \max(0, n+m-N) \le k \le \min(n,m).$$

Esempio. Un rivenditore acquista componenti elettriche a lotti di 10. Controlla a caso 3 componenti in ogni lotto e lo accetta solo se nessuno dei 3 pezzi controllati è difettoso. Se il 30% dei lotti ha 4 pezzi difettosi e il 70% ha 1 pezzo difettoso, qual è la percentuale dei lotti che il rivenditore rifiuterà?

Soluzione. Sia X il numero di pezzi difettosi tra i 3 controllati, e sia $B = \{$il lotto ha 4 pezzi difettosi$\}$, cosicché $\overline{B} = \{$il lotto ha 1 pezzo difettoso$\}$; si ha

$$\begin{aligned} P(X=0) &= P(X=0|B)\,P(B) + P(X=0|\overline{B})\,P(\overline{B}) \\ &= \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} \cdot \frac{3}{10} + \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}} \cdot \frac{7}{10} = \frac{6\cdot 5\cdot 4}{10\cdot 9\cdot 8} \cdot \frac{3}{10} + \frac{9\cdot 8\cdot 7}{10\cdot 9\cdot 8} \cdot \frac{7}{10} \\ &= 0{,}05 + 0{,}49 = 0{,}54 \end{aligned}$$

pertanto il rivenditore rifiuterà il 46% dei lotti. Notiamo inoltre che

$$P(B|X=0) = \frac{P(X=0|B)\,P(B)}{P(X=0)} = \frac{0{,}05}{0{,}54} \approx 0{,}0926.$$

Se si scelgono a caso n biglie senza reinserimento da un insieme di N biglie, delle quali la frazione $p = m/N$ è bianca, allora il numero di biglie bianche selezionate X ha distribuzione ipergeometrica. È ragionevole supporre che se m e N sono grandi rispetto a n, allora il fatto che si effettui o meno reinserimento ad ogni estrazione possa essere trascurabile. Non tenendo conto delle biglie già estratte, ogni altra estrazione darà una biglia bianca con probabilità approssimativamente pari a p se m e N sono grandi rispetto a n. In tal caso la legge di X è approssimata da una legge binomiale:

$$\begin{aligned}
P(X=k) &= \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} = \frac{\frac{(m)_k}{k!}\cdot\frac{(N-m)_{n-k}}{(n-k)!}}{\frac{(N)_n}{n!}} = \binom{n}{k}\frac{(m)_k\cdot(N-m)_{n-k}}{(N)_n} \\
&= \binom{n}{k}\frac{m(m-1)\ldots(m-k+1)\cdot(N-m)\ldots(N-m-n+k+1)}{N(N-1)\ldots(N-n+1)} \\
&\approx \binom{n}{k}p^k(1-p)^{n-k} \quad \text{con } p=\frac{m}{N}; \text{ e con } m \text{ e } N \text{ grandi rispetto a } n \text{ e } k.
\end{aligned}$$

Proposizione. Se X è una variabile aleatoria ipergeometrica di parametri n, N e m, allora per $p = \dfrac{m}{N}$ si ha

$$E(X) = n\,p, \qquad \mathrm{Var}(X) = n\,p\,(1-p)\left(1 - \frac{n-1}{N-1}\right).$$

Dimostrazione. Il momento di ordine k di X è dato da:

$$E(X^k) = \sum_{i=0}^{n} i^k\, P(X=i) = \sum_{i=1}^{n} i^k \binom{m}{i}\binom{N-m}{n-i}\Big/\binom{N}{n}.$$

Utilizzando le identità

$$i\binom{m}{i} = \frac{i\,m!}{i!\,(m-i)!} = \frac{m\,(m-1)!}{(i-1)!\,(m-i)!} = m\binom{m-1}{i-1}, \qquad n\binom{N}{n} = N\binom{N-1}{n-1}$$

si ha

$$E(X^k) = \frac{n\,m}{N}\sum_{i=1}^{n} i^{k-1}\binom{m-1}{i-1}\binom{N-m}{n-i}\Big/\binom{N-1}{n-1}.$$

Ponendo $j = i - 1$ in $E(X^k) = \dfrac{n\,m}{N} \displaystyle\sum_{i=1}^{n} i^{k-1} \binom{m-1}{i-1}\binom{N-m}{n-i} \Big/ \binom{N-1}{n-1}$ si ha

$$E(X^k) = \frac{n\,m}{N} \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{m-1}{j}\binom{N-m}{n-1-j} \Big/ \binom{N-1}{n-1} = \frac{n\,m}{N}\, E[(Y+1)^{k-1}]$$

con Y variabile aleatoria ipergeometrica di parametri $n-1$, $N-1$ e $m-1$. Ponendo $k = 1$ e $k = 2$ si ha rispettivamente:

$$E(X) = n\,\frac{m}{N} = n\,p, \qquad E(X^2) = \frac{n\,m}{N}\, E(Y+1) = \frac{n\,m}{N}\left[\frac{(n-1)\,(m-1)}{N-1} + 1\right].$$

Da ciò, ricordando che $p = m/N$, segue

$$\begin{aligned}
\mathrm{Var}(X) &= \frac{n\,m}{N}\left[\frac{(n-1)\,(m-1)}{N-1} + 1 - \frac{n\,m}{N}\right] = n\,p\left[\frac{(n-1)\,(Np-1)}{N-1} + 1 - n\,p\right] \\
&= n\,p\left[(n-1)\left(p - \frac{1-p}{N-1}\right) + 1 - n\,p\right] \\
&= n\,p\left[(n-1)\,p - (n-1)\,\frac{1-p}{N-1} + 1 - n\,p\right] = n\,p\,(1-p)\left(1 - \frac{n-1}{N-1}\right).
\end{aligned}$$

La variabile aleatoria uniforme discreta

Nell'estrarre una biglia da un'urna contenente n biglie numerate da 1 a N, denotiamo con X il numero estratto. Allora

$$P(X = k) = \frac{1}{N}, \qquad k = 1, 2, \ldots, N.$$

La variabile aleatoria dotata di tale densità discreta è detta *uniforme discreta*.

Proposizione. Se X è una variabile aleatoria uniforme discreta di parametro N, allora

$$E(X) = \frac{N+1}{2}, \qquad \mathrm{Var}(X) = \frac{N^2 - 1}{12}.$$

Dimostrazione. Ricordando che $\displaystyle\sum_{k=1}^{N} k = \frac{N(N+1)}{2}$, si ha

$$E(X) = \sum_{k=1}^{N} k\, P(X = k) = \frac{1}{N} \sum_{k=1}^{N} k = \frac{N+1}{2}.$$

Analogamente, poiché $\displaystyle\sum_{k=1}^{N} k^2 = \frac{N(N+1)(2N+1)}{6}$, si ha

$$E(X^2) = \sum_{k=1}^{N} k^2\, P(X = k) = \frac{1}{N}\sum_{k=1}^{N} k^2 = \frac{(N+1)(2N+1)}{6}.$$

Segue pertanto

$$\begin{aligned}
\mathrm{Var}(X) &= \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4} \\
&= \frac{N+1}{12}\,[2(2N+1) - 3(N+1)] = \frac{N+1}{12}\,(N-1) \\
&= \frac{N^2-1}{12}.
\end{aligned}$$

4.9 Proprietà delle funzioni di distribuzione

Abbiamo già visto in precedenza che la funzione di distribuzione $F(x)$ di una variabile aleatoria X rappresenta la probabilità $P(X \leq x)$, per ogni $x \in \mathbb{R}$.

Proposizione. Una funzione di distribuzione $F(x)$ è caratterizzata dalle proprietà:

1. $F(x)$ è una funzione non decrescente, cioè risulta $F(a) \leq F(b)$ se $a < b$.
2. $\lim_{x \to \infty} F(x) = 1$.
3. $\lim_{x \to -\infty} F(x) = 0$.
4. $F(x)$ è continua a destra, ossia $\lim_{n \to \infty} F(x_n) = F(x)$ per ogni $x \in \mathbb{R}$ fissato e per ogni successione decrescente x_n, $n \geq 1$, che converge a x.

Dimostrazione. La Proprietà 1 segue dal fatto che se $a < b$, allora

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\},$$

con i 2 eventi a secondo membro incompatibili. Pertanto, per la proprietà di additività finita,

$$F(b) = P(X \leq b) = P(X \leq a) + P(a < X \leq b) \geq F(a).$$

La Proprietà 2 si dimostra notando che b_n è una successione di reali che cresce verso ∞, allora gli eventi $\{X \le b_n\}$ formano una successione crescente di eventi il cui limite è $\{X < \infty\}$. Quindi, per la proprietà di continuità della probabilità, si ha

$$\lim_{n\to\infty} P(X \le b_n) = P(X < \infty) = 1.$$

La Proprietà 3 si dimostra in modo analogo alla Proprietà 2, ed è lasciato come esercizio.

Per dimostrare la Proprietà 4 notiamo che se x_n, $n \ge 1$, è una successione decrescente che converge a $x \in \mathbb{R}$, allora $\{X \le x_n\}$, $n \ge 1$, costituisce una successione di eventi decrescente il cui limite coincide con $\{X \le x\}$. Quindi, in conclusione, la proprietà di continuità fa sì che

$$\lim_{n\to\infty} P(X \le x_n) = P(X \le x).$$

Proposizione. Sia $F(x)$ la funzione di distribuzione di una variabile aleatoria X. Allora,
(i) $P(a < X \leq b) = F(b) - F(a)$ per ogni $a < b$.
(ii) $P(X < b) = \lim_{n\to\infty} F(b_n)$ per ogni $b \in \mathbb{R}$ fissato e per ogni successione crescente b_n, $n \geq 1$, che converge a b.

Dimostrazione. Nella Proposizione precedente abbiamo visto che se $a < b$, allora

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\},$$

e quindi risulta $F(b) = F(a) + P(a < X \leq b)$, da cui segue la (i).

Per dimostrare la (ii) notiamo che se b_n, $n \geq 1$, è una successione crescente che converge a b, allora dalla proprietà di continuità otteniamo:

$$P(X < b) = P\left(\lim_{n\to\infty}\{X \leq b_n\}\right) = \lim_{n\to\infty} P(X \leq b_n) = \lim_{n\to\infty} F(b_n).$$

Notiamo che una funzione di distribuzione $F(x)$ non è necessariamente continua. Infatti, limite sinistro e limite destro di $F(\cdot)$ in $x \in \mathbb{R}$ non necessariamente coincidono, essendo

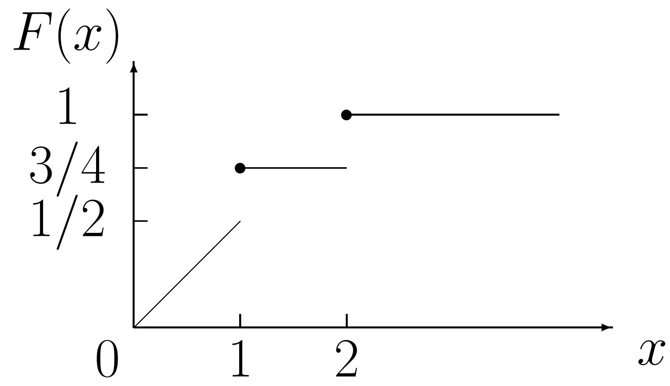
$$F(x^-) := \lim_{h \to 0^-} F(x+h) \leq \lim_{h \to 0^+} F(x+h) = F(x).$$

Se la disuguaglianza precedente è soddisfatta come uguaglianza, allora $F(\cdot)$ è continua in x e risulta $P(X = x) = F(x) - F(x^-) = 0$.

Altrimenti, $F(\cdot)$ è discontinua in x e risulta $P(X = x) = F(x) - F(x^-) > 0$.

Esempio. Sia X una variabile aleatoria avente funzione di distribuzione

$$F(x) = \begin{cases} 0, & x < 0 \\ x/2, & 0 \le x < 1 \\ 3/4, & 1 \le x < 2 \\ 1, & x \ge 2. \end{cases}$$



Calcolare (a) $P(X < 2)$, (b) $P(X = 1)$, (c) $P(X > 1/2)$, (d) $P(1 < X \le 3)$.

Soluzione. Si ha

(a) $P(X < 2) = F(2^-) = 3/4$;

(b) $P(X = 1) = F(1) - F(1^-) = 3/4 - 1/2 = 1/4$;

(c) $P(X > 1/2) = 1 - P(X \le 1/2) = 1 - F(1/2) = 1 - 1/4 = 3/4$;

(d) $P(1 < X \le 3) = F(3) - F(1) = 1 - 3/4 = 1/4$.

Appendice. Processo di Poisson

Nello studio di eventi che si verificano ad istanti (aleatori) di tempo supponiamo che, data una costante positiva λ, siano verificate le seguenti ipotesi:

1. La probabilità che un evento si verifichi in un dato intervallo di tempo di lunghezza h è uguale a $\lambda\, h + o(h)$, dove $o(h)$ indica una qualsiasi funzione $f(h)$ per la quale risulta $\lim_{h\to 0} f(h)/h = 0$. [Ad esempio, $f(h) = h^2$ è $o(h)$, mentre $f(h) = h$ non è $o(h)$.]

2. La probabilità che due o più eventi si verifichino in un intervallo di tempo di lunghezza h è uguale a $o(h)$.

3. Sia $E_i = \{$nell'intervallo $(t_{i-1}, t_i]$ si verificano k_i eventi$\}$, $i = 1, 2, \ldots, n$. Per ogni scelta degli istanti di tempo $0 \leq t_0 < t_1 < t_2 < \ldots < t_n$ e degli interi non negativi $k_1, k_2, \ldots, k_n$ gli eventi $E_1, E_2, \ldots, E_n$ sono indipendenti.

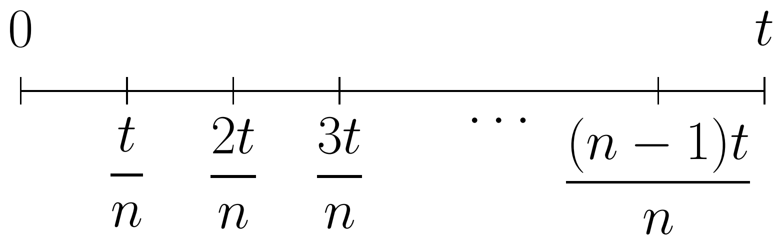
Dalle ipotesi 1 e 2 segue la seguente condizione:

4. La probabilità che non si verifichino eventi in un intervallo di tempo di lunghezza h è uguale a $1 - \lambda\, h + o(h)$.

Mostreremo che se valgono le ipotesi 1, 2 e 3, allora il numero di eventi che si verificano in ogni intervallo di lunghezza t ha distribuzione di Poisson di parametro $\lambda\, t$.

Denotiamo con $N(t)$ il numero di eventi che si verificano nell'intervallo $[0, t]$.

Suddividiamo l'intervallo $[0, t]$ in n sottointervalli disgiunti di lunghezza t/n.



Risulta

$$P\{N(t) = k\} = P(A) + P(B),$$

dove

$$A = \{k \text{ degli } n \text{ sottointervalli contengono 1 evento e gli altri } n-k \text{ ne contengono } 0\}$$

$$B = \{N(t) = k \text{ e almeno 1 sottointervallo contiene 2 o più eventi}\}$$

Mostreremo ora che $P(B) = 0$ e, al limite per $n \to \infty$, $P(A) = e^{-\lambda t}\,\dfrac{(\lambda t)^k}{k!}$, e così $P\{N(t) = k\} = e^{-\lambda t}\,\dfrac{(\lambda t)^k}{k!}$. Si ha

$$
\begin{aligned}
P(B) &= P\{N(t) = k \text{ e almeno 1 sottointervallo contiene 2 o più eventi}\} \\
&\leq P\{\text{almeno 1 sottointervallo contiene 2 o più eventi}\} \\
&= P\left(\bigcup_{i=1}^{n}\{\text{l'}i\text{-esimo sottointervallo contiene 2 o più eventi}\}\right) \\
&\leq \sum_{i=1}^{n} P\{\text{l'}i\text{-esimo sottointervallo contiene 2 o più eventi}\} \\
&= \sum_{i=1}^{n} o\left(\frac{t}{n}\right) \\
&= n\, o\left(\frac{t}{n}\right) \\
&= t\left[\frac{o(t/n)}{t/n}\right]
\end{aligned}
$$

Pertanto, $$P(B) \le t \left[\frac{o(t/n)}{t/n} \right].$$

$\forall t > 0$ si ha $\dfrac{t}{n} \to 0$ quando $n \to \infty$. Quindi, $\dfrac{o(t/n)}{t/n} \to 0$ quando $n \to \infty$ e perciò

$$\lim_{n\to\infty} P(B) = 0.$$

Si ha inoltre

$$\begin{aligned} P(A) &= P\{k \text{ degli } n \text{ sottointervalli contengono 1 evento} \\ &\qquad \text{e gli altri } n-k \text{ ne contengono } 0\} \\ &= \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned}$$

dove $p = P\{$in un intervallo di ampiezza $\dfrac{t}{n}$ si ha 1 evento$\} = \dfrac{\lambda t}{n} + o\left(\dfrac{\lambda t}{n}\right)$. Quindi,

$$P(A) = \binom{n}{k} \left[\frac{\lambda t}{n} + o\left(\frac{\lambda t}{n}\right) \right]^k \left[1 - \frac{\lambda t}{n} - o\left(\frac{\lambda t}{n}\right) \right]^{n-k}$$

Abbiamo ricavato che $P(A)$ è la probabilità di una variabile aleatoria binomiale di parametri $(n;p)$, con $p = \frac{\lambda t}{n} + o\left(\frac{\lambda t}{n}\right)$. Infatti, per $k = 0, 1, \ldots, n$:

$$P(A) = \binom{n}{k} \left[\frac{\lambda t}{n} + o\left(\frac{\lambda t}{n}\right)\right]^k \left[1 - \frac{\lambda t}{n} - o\left(\frac{\lambda t}{n}\right)\right]^{n-k}.$$

Procedendo al limite per $n \to \infty$ notiamo che $\lim_{n\to\infty} p = 0$ e che

$$\lim_{n\to\infty} n\, p = \lim_{n\to\infty} n \left[\frac{\lambda t}{n} + o\left(\frac{\lambda t}{n}\right)\right] = \lambda\, t + t \lim_{n\to\infty} \frac{o(t/n)}{t/n} = \lambda\, t.$$

Quindi, un ragionamento analogo a quello fatto in precedenza per mostrare che la distribuzione di Poisson approssima quella binomiale, ci dà

$$\lim_{n\to\infty} P(A) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

In conclusione ricaviamo, per $n \to \infty$, che

$$P\{N(t) = k\} = P(A) + P(B) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \qquad (k = 0, 1, \ldots).$$

In conclusione, nelle ipotesi 1, 2 e 3, il numero di eventi che si verificano in un intervallo fissato di lunghezza t si distribuisce secondo una variabile aleatoria di Poisson di media $\lambda\, t$, e diremo che gli eventi si verificano secondo un processo di Poisson di intensità λ.

Il valore λ è uguale al numero medio di eventi che si verificano per unità di tempo: $E[N(1)] = \lambda$; è una costante che si può determinare empiricamente.

Esempio. In un processo di Poisson, (a) ricavare la distribuzione di probabilità dell'istante in cui si verificherà il prossimo evento, (b) determinare la probabilità che avvengano almeno 3 eventi in 2 unità temporali, nel caso di intensità $\lambda = 2$.

Soluzione. (a) Sia X il tempo che trascorrerà fino al prossimo evento. Poiché $X \geq t$ se e solo se nessun evento si verificherà prima delle prossime t unità temporali, si ha

$$P(X > t) = P[N(t) = 0] = e^{-\lambda t}, \qquad t \geq 0,$$

e quindi $F(t) = P(X \leq t) = 1 - P(X > t) = 1 - e^{-\lambda t}$, $t \geq 0$.

(b) Si ha $P[N(2) \geq 3] = 1 - \sum_{k=0}^{2} P[N(2) = k] = 1 - \left(1 + 4 + \dfrac{4^2}{2}\right) e^{-4} = 1 - 13\, e^{-4}$.

CAPITOLO 5 – Variabili aleatorie continue

5.1 Introduzione

5.2 Valore atteso e varianza di una variabile aleatoria continua

5.3 La variabile aleatoria uniforme

5.4 Variabili aleatorie normali

5.5 Variabile aleatoria esponenziale

5.6 Distribuzione di una funzione di variabile aleatoria

5.1 Introduzione

Le variabili aleatorie discrete assumono un numero finito o un'infinità numerabile di valori.

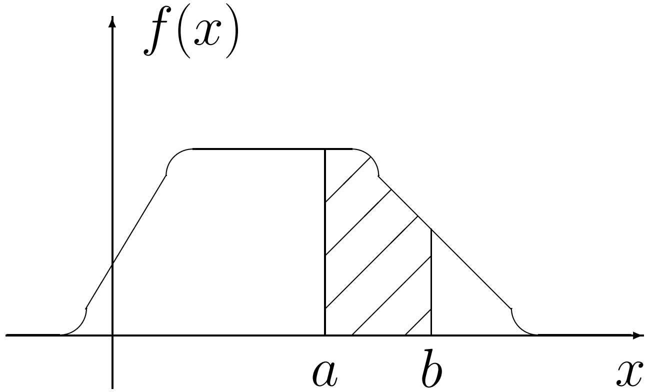
Esistono comunque variabili aleatorie il cui insieme dei valori è non numerabile, come ad esempio l'ora di arrivo di un treno o il tempo di vita di un dispositivo elettronico.

Definizione. Una variabile aleatoria X è detta continua (o, anche, assolutamente continua) se esiste una funzione $f\colon \mathbb{R} \to [0, +\infty)$ tale che per ogni sottoinsieme B di numeri reali risulta

$$P(X \in B) = \int_B f(x)\, dx.$$

La funzione f è chiamata *funzione di densità* della variabile aleatoria X, ed è di fatto caratterizzata dalle seguenti proprietà:

$$f(x) \geq 0 \quad \text{per } x \in \mathbb{R}, \qquad \int_{-\infty}^{\infty} f(x)\, dx = 1.$$



$$P(a \le X \le b) = \text{ area della regione tratteggiata } = \int_a^b f(x)\,dx.$$

Tale relazione ricalca la seguente, esaminata in passato per variabili aleatorie discrete:

$$P(a \le X \le b) = \sum_{k:\, a \le x_k \le b} p(x_k).$$

Abbiamo visto che se X è una variabile aleatoria continua, si ha

$$P(a \le X \le b) = \int_a^b f(x)\,dx.$$

Quindi, se $b = a$ si ricava

$$P(X = a) = \int_a^a f(x)\,dx = 0.$$

Pertanto, a differenza del caso discreto, la probabilità che una variabile aleatoria continua X assuma un singolo valore è uguale a zero, cosicché risulta

$$P(X < a) = P(X \le a) = F(a) = \int_{-\infty}^a f(x)\,dx$$

e quindi

$$F(b) - F(a) = P(a < X \le b) = \int_a^b f(x)\,dx.$$

Esempio. Sia X una variabile aleatoria continua di densità

$$f(x) = \begin{cases} c(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{altrimenti.} \end{cases}$$

(a) Calcolare c.

(b) Determinare $P(X > 1)$.

Soluzione. (a) Per determinare c, imponendo che sia $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$ si ha

$$1 = c\int_0^2 (4x - 2x^2)\,dx = c\left[2x^2 - \frac{2x^3}{3}\right]_{x=0}^{x=2} = c\left[8 - \frac{16}{3}\right] = c\,\frac{8}{3}$$

da cui segue $c = 3/8$.

(b) Risulta quindi

$$P(X > 1) = \int_1^{\infty} f(x)\,dx = \frac{3}{8}\int_1^2 (4x - 2x^2)\,dx = \frac{3}{8}\left[2x^2 - \frac{2x^3}{3}\right]_{x=1}^{x=2} = 1 - \frac{1}{2} = \frac{1}{2}.$$

Esempio. Il tempo (in ore) che un computer funzioni prima di bloccarsi è una variabile aleatoria continua X di densità

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

(a) Calcolare la probabilità che il computer funzioni tra 50 e 150 ore senza bloccarsi.

(b) Calcolare la probabilità che il computer funzioni per meno di 100 ore senza bloccarsi.

Soluzione. (a) Per determinare λ, imponendo che risulti $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx = 1$ otteniamo

$$1 = \lambda \int_0^\infty e^{-x/100}\, dx = \lambda \left[-100\, e^{-x/100} \right]_{x=0}^{x=\infty} = \lambda\, 100 \qquad \Longrightarrow \qquad \lambda = \frac{1}{100}$$

e pertanto

$$P(50 < X < 150) = \int_{50}^{150} \frac{1}{100}\, e^{-x/100}\, dx = \left[-e^{-x/100} \right]_{50}^{150} = e^{-1/2} - e^{-3/2} \approx 0{,}383.$$

(b) Analogamente si ha

$$P(X < 100) = \int_0^{100} \frac{1}{100}\, e^{-x/100}\, dx = \left[-e^{-x/100} \right]_0^{100} = 1 - e^{-1} \approx 0{,}632.$$

Esempio. Il tempo (in ore) di vita di certe pile per la radio è una variabile aleatoria continua X di densità

$$f(x) = \begin{cases} 0 & x \leq 100 \\ 100/x^2 & x > 100. \end{cases}$$

Determinare la probabilità che esattamente 2 pile della radio su 5 debbano essere sostituite entro le 150 ore di attività, supponendo che gli eventi $E_i = \{$l'i-esima pila va rimpiazzata entro 150 ore d'uso$\}$, $1 \leq i \leq 5$, siano indipendenti.

Soluzione. Risulta

$$P(E_i) = \int_0^{150} f(x)\,dx = 100 \int_{100}^{150} \frac{1}{x^2}\,dx = 100 \left[-\frac{1}{x}\right]_{x=100}^{x=150} = \frac{1}{3}.$$

Quindi, per l'indipendenza degli eventi E_i, la probabilità richiesta è

$$\binom{5}{2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^3 = 10 \cdot \frac{1}{9} \cdot \frac{8}{27} = \frac{80}{243} \approx 0{,}329.$$

La relazione tra la funzione di distribuzione F di una variabile aleatoria continua e la densità f è data da

$$F(x) = P\{X \in (-\infty, x]\} = \int_{-\infty}^{x} f(t)\, dt.$$

Quindi, se x è un reale in cui F è derivabile, derivando ambo i membri si ottiene

$$\frac{d}{dx} F(x) = f(x)$$

pertanto la densità è la derivata della funzione di distribuzione.

Un'interpretazione intuitiva della densità segue da:

$$P\left\{x - \frac{\varepsilon}{2} \leq X \leq x - \frac{\varepsilon}{2}\right\} = \int_{x-\varepsilon/2}^{x+\varepsilon/2} f(x)\, dx \approx \varepsilon\, f(x),$$

dove l'approssimazione sussiste quando ε è prossimo a 0 e la densità f è continua nel punto x. Si trae che la probabilità che X assuma valori in un intorno di x avente ampiezza ε è approssimativamente uguale a $\varepsilon\, f(x)$.

Esempio. Sia X una variabile aleatoria continua con funzione di distribuzione F_X e densità f_X. Determinare la densità di $Y = aX + b$, con $a \neq 0$.

Soluzione. Se $a > 0$, la funzione di distribuzione di Y è così esprimibile:

$$F_Y(x) = P(Y \leq x) = P(aX + b \leq x) = P\left(X \leq \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right),$$

mentre per $a < 0$ risulta:

$$F_Y(x) = P\left(X \geq \frac{x-b}{a}\right) = 1 - F_X\left(\frac{x-b}{a}\right).$$

Derivando rispetto ad x si ottiene infine:

$$f_Y(x) = \begin{cases} \dfrac{1}{a} f_X\left(\dfrac{x-b}{a}\right) & \text{se } a > 0 \\ -\dfrac{1}{a} f_X\left(\dfrac{x-b}{a}\right) & \text{se } a < 0 \end{cases} = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right).$$

Esempio. Sia X una variabile aleatoria continua con densità

$$f_X(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{altrimenti.} \end{cases}$$

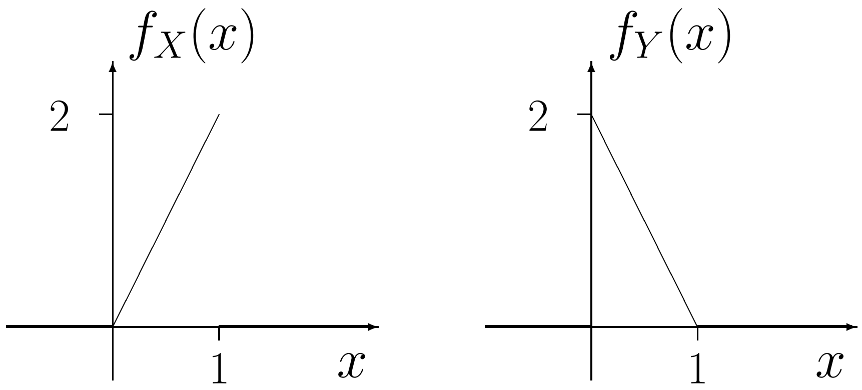
Determinare la densità di $Y = 1 - X$.

Soluzione. Ricordando che la densità di $aX + b$ è data da

$$f_{aX+b}(x) = \frac{1}{|a|}\, f_X\left(\frac{x-b}{a}\right),$$

per $a = -1$ e $b = 1$ segue che la densità di $Y = 1 - X$ è:

$$f_Y(x) = f_X(1-x) = \begin{cases} 2(1-x) & 0 < x < 1 \\ 0 & \text{altrimenti.} \end{cases}$$



5.2 Valore atteso e varianza di una variabile aleatoria continua

Come visto in precedenza, il valore atteso di una variabile aleatoria discreta X è

$$E[X] = \sum_x x\, P(X = x).$$

Analogamente, se X è una variabile aleatoria continua con densità $f(x)$, poiché

$$f(x)\, dx \approx P(x \leq X \leq x + dx) \qquad \text{per } dx \text{ piccolo},$$

si definisce il valore atteso di X come

$$E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx.$$

Esempio. Determinare $E[X]$ sapendo che la densità di X è

$$f_X(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{altrimenti.} \end{cases}$$

Soluzione. Si ha

$$E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^1 2x^2\, dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3}.$$

Esempio. Determinare $E\left[e^X\right]$ sapendo che la densità di X è

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{altrimenti.} \end{cases}$$

Soluzione. Posto $Y = e^X$, determiniamo innanzitutto la funzione di distribuzione e la densità di Y. Per $1 \le x \le e$ risulta

$$F_Y(x) = P(Y \le x) = P\left(e^X \le x\right) = P(X \le \log x) = \int_0^{\log x} f_X(y)\, dy = \log x.$$

Derivando $F_Y(x)$ si ha

$$f_Y(x) = \frac{d}{dx} F_Y(x) = \frac{d}{dx} \log x = \frac{1}{x}, \qquad 1 \le x \le e,$$

da cui segue

$$E\left[e^X\right] = E[Y] = \int_{-\infty}^{\infty} x\, f_Y(x)\, dx = \int_1^e dx = e - 1 = 1{,}71828.$$

Se X è una variabile aleatoria discreta, sappiamo che

$$E[g(X)] = \sum_i g(x_i)\, p(x_i).$$

È possibile dimostrare il seguente analogo risultato:

Proposizione. Se X è una variabile aleatoria continua con densità $f(x)$, allora per ogni funzione g a valori reali,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx.$$

Ad esempio, applicando tale Proposizione alla variabile aleatoria continua di densità

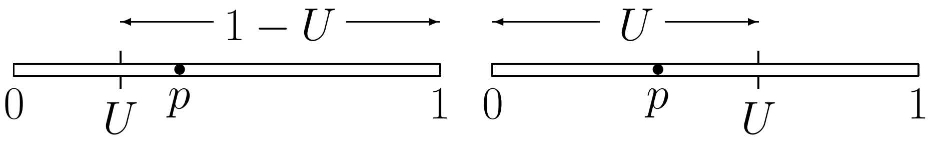
$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{altrimenti} \end{cases}$$

si ottiene

$$E\left[e^X\right] = \int_{-\infty}^{\infty} e^x\, f_X(x)\, dx = \int_0^1 e^x\, dx = [e^x]_0^1 = e - 1 = 1{,}71828.$$

Esempio. Un bastoncino di lunghezza 1 è spezzato in un punto U scelto a caso su di esso, e quindi distribuito uniformemente su $(0, 1)$. Determinare il valore atteso della lunghezza del pezzo di bastoncino che contiene il punto p, $0 \le p \le 1$.

Soluzione. Sia $L(U)$ la lunghezza del pezzo di bastoncino contenente p. Si ha



$$L(U) = \begin{cases} 1 - U & \text{se } 0 < U < p \\ U & \text{se } p < U < 1. \end{cases}$$

Quindi, essendo $f_U(x) = 1$, $0 \le x \le 1$, segue che

$$\begin{aligned} E[L(U)] &= \int_{-\infty}^{\infty} L(x)\, f_U(x)\, dx = \int_0^p (1-x)\, dx + \int_p^1 x\, dx \\ &= \left[x - \frac{x^2}{2}\right]_0^p + \left[\frac{x^2}{2}\right]_p^1 = p - \frac{p^2}{2} + \frac{1}{2} - \frac{p^2}{2} = \frac{1}{2} + p(1-p). \end{aligned}$$

Il valore atteso della lunghezza del pezzo di bastoncino contente p è massimo se p è il punto centrale, ossia se $p = 1/2$. Notiamo inoltre che $E[L(U)] \ge E[U] = 1/2$.

Corollario. Come nel caso discreto, se X è una variabile continua, con a e b costanti, si ha

$$E[aX + b] = aE[X] + b.$$

Se X è una variabile aleatoria continua di valore atteso μ, in analogia col caso discreto la varianza di X è definita da

$$\mathrm{Var}(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx.$$

La formula alternativa,

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2,$$

si dimostra imitando quanto fatto nel caso discreto. Analogamente, se a e b sono delle costanti, allora

$$\mathrm{Var}(aX + b) = a^2 Var(X).$$

Esempio. Sia X una variabile aleatoria continua avente densità

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{altrimenti.} \end{cases}$$

Determinare la funzione di distribuzione e la varianza di X.

Soluzione. La funzione di distribuzione di X è data da

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \begin{cases} 0 & x < 0 \\ \int_0^x 2t\,dt = x^2 & 0 \le x < 1 \\ 1 & x \ge 1. \end{cases}$$

Ricaviamo la varianza di X ricordando che $\mu = E[X] = 2/3$; pertanto:

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx = \int_0^1 \left(x - \frac{2}{3}\right)^2 2x\,dx$$

$$= 2\int_0^1 \left(x^3 - \frac{4}{3}x^2 + \frac{4}{9}x\right) dx = 2\left[\frac{x^4}{4} - \frac{4}{9}x^3 + \frac{2}{9}x^2\right]_0^1 = \frac{1}{18}.$$

Esempio. Sia X una variabile aleatoria continua avente funzione di distribuzione

$$F(x) = \begin{cases} 0 & x < 0 \\ \sqrt{x} & 0 \le x < 1 \\ 1 & x \ge 1. \end{cases}$$

Determinare la densità, il valore atteso e la varianza di X.

Soluzione. Derivando $F(x)$ si ottiene la densità di X:

$$f(x) = \frac{d}{dx} F(x) = \begin{cases} \dfrac{1}{2\sqrt{x}} & 0 < x < 1 \\ 0 & \text{altrimenti.} \end{cases}$$

Da questa si ricavano i momenti di X:

$$E\left[X^n\right] = \int_{-\infty}^{\infty} x^n \, f(x) \, dx = \int_0^1 \frac{x^{n-1/2}}{2} \, dx = \frac{1}{2} \left[\frac{x^{n+1/2}}{n+1/2} \right]_0^1 = \frac{1}{2n+1},$$

da cui seguono valore atteso e la varianza di X:

$$E[X] = \frac{1}{3}, \qquad \mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{5} - \frac{1}{9} = \frac{4}{45}.$$

5.3 La variabile aleatoria uniforme

Una variabile aleatoria è detta *uniforme* (o *uniformemente* distribuita) sull'intervallo $(0,1)$ se la sua densità è data da

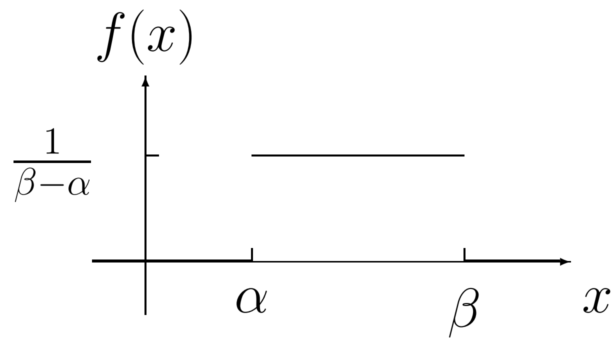
$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{altrimenti.} \end{cases}$$

Risulta $f(x) \geq 0$ e $\int_{-\infty}^{\infty} f(x)\,dx = 1$; inoltre per ogni $0 < a < b < 1$ si ha

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx = b - a.$$

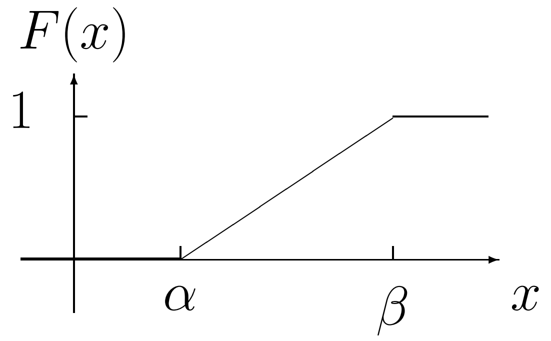
In generale, X è una variabile aleatoria uniforme sull'intervallo (α, β) se ha densità

$$f_X(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \alpha < x < \beta \\ 0 & \text{altrimenti.} \end{cases}$$



Essendo $F(x) = \int_{-\infty}^{x} f(u)\,du$, la funzione di distribuzione di una variabile aleatoria uniforme sull'intervallo (α, β) è data da

$$F(x) = \begin{cases} 0 & x < \alpha \\ \dfrac{x-\alpha}{\beta-\alpha} & \alpha \le x < \beta \\ 1 & x \ge \beta \end{cases}$$



Esempio. Sia X una variabile aleatoria continua avente distribuzione uniforme su (α, β). Determinare momenti, valore atteso e varianza di X.

Soluzione. Si ha

$$E\left[X^n\right] = \int_{-\infty}^{\infty} x^n\, f(x)\, dx = \int_{\alpha}^{\beta} \frac{x^n}{\beta-\alpha}\, dx = \frac{1}{(\beta-\alpha)} \left[\frac{x^{n+1}}{n+1}\right]_{\alpha}^{\beta} = \frac{\beta^{n+1}-\alpha^{n+1}}{(n+1)(\beta-\alpha)},$$

$$E[X] = \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} = \frac{\alpha+\beta}{2}, \qquad E\left[X^2\right] = \frac{\beta^3-\alpha^3}{3(\beta-\alpha)} = \frac{\beta^2+\alpha\beta+\alpha^2}{3},$$

$$\mathrm{Var}(X) = E\left[X^2\right] - (E[X])^2 = \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \frac{(\alpha+\beta)^2}{4} = \frac{(\beta-\alpha)^2}{12}.$$

Esempio. Sia X una variabile aleatoria avente distribuzione uniforme su $(-10, 10)$. Calcolare $P(X > 3)$, $P(X < 6)$ e $P(X > 3 \,|\, X < 6)$.

Soluzione. Dall'espressione della funzione di distribuzione

$$F(x) = \frac{x+10}{20}, \qquad -10 \leq x \leq 10$$

segue

$$P(X > 3) = 1 - F(3) = 1 - \frac{13}{20} = \frac{7}{20},$$

$$P(X < 6) = F(6) = \frac{16}{20},$$

$$P(X > 3 \,|\, X < 6) = \frac{P(3 < X < 6)}{P(X < 6)} = \frac{F(6) - F(3)}{F(6)} = \frac{3/20}{16/20} = \frac{3}{16}.$$

Esempio. Gi autobus passano ad una fermata ad intervalli di 15 minuti a partire dalle ore 7, cioè alle 7, 7:15, 7:30, 7:45, ecc. Se un passeggero arriva alla fermata in un istante uniformemente distribuito tra le 7 e le 7:30, determinare la probabilità che egli aspetti l'autobus (a) meno di 5 minuti (b) più di 10 minuti.

Soluzione. Sia X il minuto tra le 7 e le 7:30 in cui arriva il passeggero. Il passeggero aspetterà meno di 5 minuti se (e solo se) egli arriva tra le 7:10 e le 7:15 o tra le 7:25 e le 7:30. Pertanto, poiché X è uniforme sull'intervallo $(0, 30)$, si ha

$$f(x) = \begin{cases} \dfrac{1}{30} & 0 < x < 30 \\ 0 & \text{altrimenti} \end{cases}$$

e quindi la probabilità cercata in (a) è

$$P(10 < X < 15) + P(25 < X < 30) = \int_{10}^{15} \frac{1}{30}\,dx + \int_{25}^{30} \frac{1}{30}\,dx = \frac{5+5}{30} = \frac{1}{3}.$$

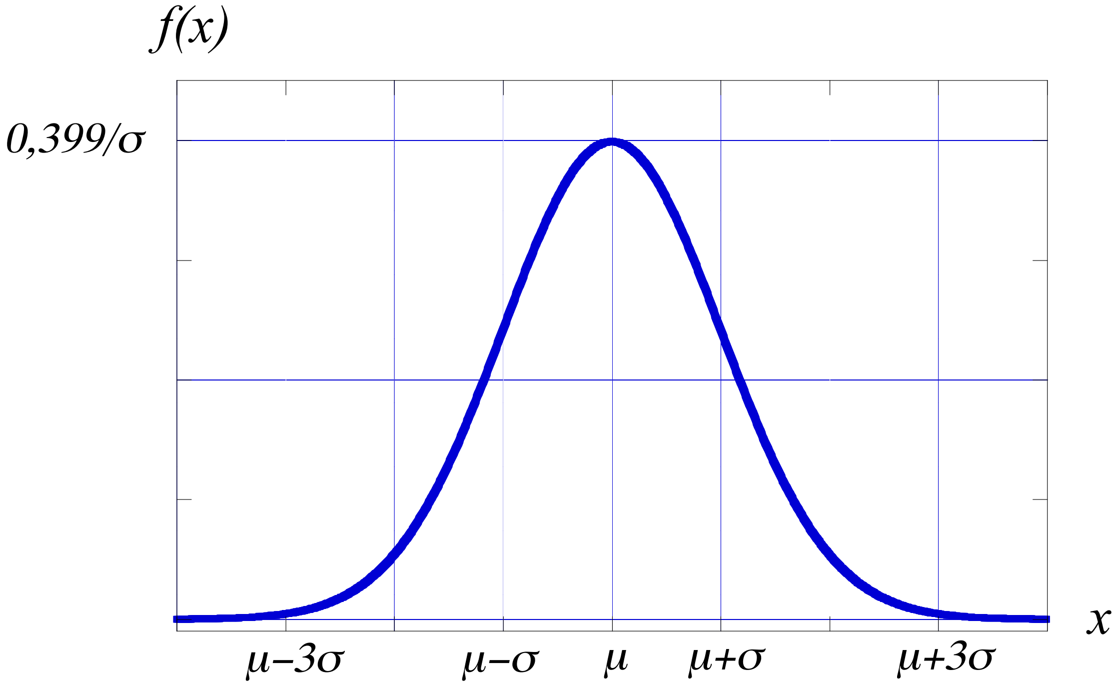
Analogamente, la probabilità cercata in (b) vale

$$P(0 < X < 5) + P(15 < X < 20) = \int_{0}^{5} \frac{1}{30}\,dx + \int_{15}^{20} \frac{1}{30}\,dx = \frac{5+5}{30} = \frac{1}{3}.$$

5.4 Variabili aleatorie normali

Una variabile aleatoria continua X è detta normale (o gaussiana) di parametri $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ se la densità di X è data da

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty.$$



La densità è una funzione con grafico a campana, è simmetrica rispetto a $x = \mu$, e possiede 2 punti di flesso: $\mu - \sigma$ e $\mu + \sigma$.

Esempio. Mostrare che se X è una variabile aleatoria normale di parametri μ e σ^2, allora valore atteso e varianza sono: $E[X] = \mu$ e $Var(X) = \sigma^2$.
Soluzione. Consideriamo la variabile $Z = (X - \mu)/\sigma$. Si ha

$$E[Z] = \int_{-\infty}^{\infty} x\, f_Z(x)\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x\, e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} \Big|_{x=-\infty}^{x=\infty} = 0$$

e quindi (integrando per parti) si ottiene

$$\mathrm{Var}(Z) = E[Z^2] - (E[Z])^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2\, e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\, dx = 1.$$

Dato che $X = \sigma Z + \mu$, si ricava

$$E[X] = \sigma\, E[Z] + \mu = \mu, \qquad \mathrm{Var}(X) = \sigma^2\, \mathrm{Var}(Z) = \sigma^2.$$

Proposizione. Se X è normale di parametri μ e σ^2, allora $Y = aX + b$ è una variabile normale di parametri $a\mu + b$ e $a^2\sigma^2$.

Dimostrazione. Supponiamo $a > 0$ (il caso $a < 0$ è analogo). La funzione di distribuzione di Y è data da

$$F_Y(x) = P(aX + b \leq x) = P\left(X \leq \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right)$$

Derivando rispetto a x si deduce che la densità di Y è

$$\begin{aligned} f_Y(x) &= \frac{d}{dx}F_Y(x) = \frac{d}{dx}F_X\left(\frac{x-b}{a}\right) = \frac{1}{a}\, f_X\left(\frac{x-b}{a}\right) \\ &= \frac{1}{a}\,\frac{1}{\sqrt{2\pi}\,\sigma}\,\exp\left\{-\frac{1}{2\sigma^2}\left(\frac{x-b}{a}-\mu\right)^2\right\} \\ &= \frac{1}{\sqrt{2\pi}\,a\sigma}\,\exp\left\{-\frac{[x-(a\mu+b)]^2}{2(a\sigma)^2}\right\}, \end{aligned}$$

che è la densità di una variabile aleatoria normale di parametri $a\mu + b$ e $a^2\sigma^2$.

Una variabile aleatoria di valore medio 0 e varianza 1 è detta *standard*.

Esempio. Se X è una variabile aleatoria di valore medio μ e varianza σ^2, determinare i valori di a e b tali che $Z = aX + b$ sia standard.

Soluzione. Affinché sia $E(Z) = 0$ e $\mathrm{Var}(Z) = 1$ deve risultare

$$E(aX + b) = a\mu + b = 0, \qquad \mathrm{Var}(aX + b) = a^2\sigma^2 = 1,$$

da cui si ottengono due diverse scelte dei valori di a e b tali che $Z = aX+b$ è standard:

$$(a,\, b) = \left(\frac{1}{\sigma},\, -\frac{\mu}{\sigma}\right), \qquad (a,\, b) = \left(-\frac{1}{\sigma},\, \frac{\mu}{\sigma}\right).$$

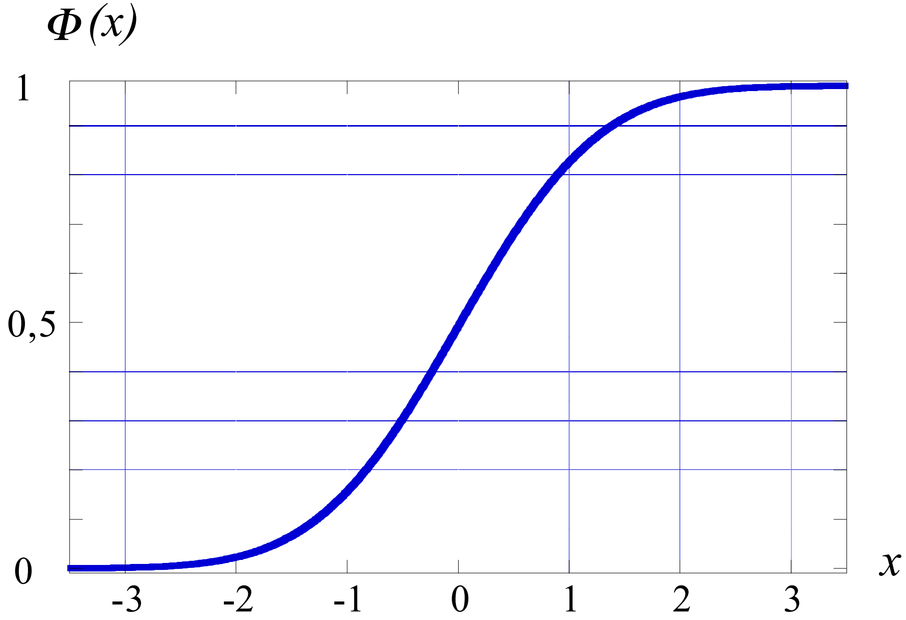
Dai risultati precedenti segue che se X è una variabile normale di parametri μ e σ^2,

$$Z = \frac{X-\mu}{\sigma} \qquad \text{e} \qquad Z' = \frac{\mu - X}{\sigma}$$

sono variabili normali di parametri 0 e 1, ossia sono *variabili normali standard*.

La funzione di distribuzione di una variabile normale standard si indica con $\Phi(x)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2}\, dz, \qquad -\infty < x < \infty.$$



Nella Tabella in Appendice sono dati i valori di $\Phi(x)$ per $x = 0;\ 0{,}01;\ 0{,}02;\ \ldots;\ 3{,}49$. Per valori negativi di x, tale funzione si può ottenere dalla seguente formula:

$$\Phi(-x) = 1 - \Phi(x), \qquad -\infty < x < \infty.$$

Dalla simmetria della densità normale standard $f(x) = \dfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$ segue la relazione di simmetria per $\Phi(x)$:

$$\Phi(-x) = 1 - \Phi(x), \qquad -\infty < x < \infty,$$

la quale afferma che se Z è una variabile normale standard allora

$$P(Z \leq -x) = P(Z > x), \qquad -\infty < x < \infty.$$

Ricordiamo che se X è una variabile normale di parametri μ e σ^2, allora $Z = (X-\mu)/\sigma$ è una variabile normale standard, e quindi la funzione di distribuzione di X è

$$F_X(x) = P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Ne segue:

$$P(a \leq X \leq b) = P(X \leq b) - P(X < a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Esempio. Sia X una variabile aleatoria normale di parametri $\mu = 3$ e $\sigma^2 = 9$. Determinare $P(2 < X < 5)$, $P(X > 0)$, $P(|X - 3| > 6)$.

Soluzione. Ponendo $Z = (X - \mu)/\sigma$ si ha:

$$\begin{aligned} P(2 < X < 5) &= P\left(\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right) = P\left(-\frac{1}{3} < Z < \frac{2}{3}\right) \\ &= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right) \approx \Phi\big(0{,}67\big) - \big[1 - \Phi\big(0{,}33\big)\big] \\ &= 0{,}7486 - (1 - 0{,}6293) = 0{,}3779 \end{aligned}$$

$$P(X > 0) = P\left(\frac{X-3}{3} > \frac{0-3}{3}\right) = P\,(Z > -1) = 1 - \Phi\big(-1\big) = \Phi\big(1\big) = 0{,}8413$$

$$\begin{aligned} P(|X - 3| > 6) &= P\big(X > 9\big) + P\big(X < -3\big) \\ &= P\left(\frac{X-3}{3} > \frac{9-3}{3}\right) + P\left(\frac{X-3}{3} < \frac{-3-3}{3}\right) \\ &= P\,(Z > 2) + P\,(Z < -2) = 1 - \Phi(2) + \Phi(-2) \\ &= 2\big[1 - \Phi\big(2\big)\big] = 2(1 - 0{,}9772) = 0{,}0456 \end{aligned}$$

Esempio. In un test d'esame in cui il punteggio assegnato è una variabile aleatoria normale X di parametri μ e σ^2 viene assegnata la lettera A a chi ha un punteggio superiore a $\mu+\sigma$, B a chi ha un punteggio tra μ e $\mu+\sigma$, C a chi ha un punteggio tra $\mu-\sigma$ e μ, D a chi ha punteggio tra $\mu-2\sigma$ e $\mu-\sigma$, ed E a chi ottiene un punteggio inferiore a $\mu-2\sigma$. Determinare le percentuali dei giudizi A, B, C, D, E.

Soluzione. Dato che

$$P(X>\mu+\sigma)=P\Big(\frac{X-\mu}{\sigma}>1\Big)=1-\Phi\big(1\big)=1-0{,}8413=0{,}1587$$

$$P(\mu<X<\mu+\sigma)=P\Big(0<\frac{X-\mu}{\sigma}<1\Big)=\Phi\big(1\big)-\Phi\big(0\big)=0{,}8413-0{,}5=0{,}3413$$

$$P(\mu-\sigma<X<\mu)=P\Big(-1<\frac{X-\mu}{\sigma}<0\Big)=\Phi\big(0\big)-\Phi\big(-1\big)=0{,}3413$$

$$P(\mu-2\sigma<X<\mu-\sigma)=P\Big(-2<\frac{X-\mu}{\sigma}<-1\Big)=\Phi\big(2\big)-\Phi\big(1\big)=0{,}1359$$

$$P(X<\mu-2\sigma)=P\Big(\frac{X-\mu}{\sigma}<-2\Big)=\Phi\big(-2\big)=1-\Phi\big(2\big)=1-0{,}9772=0{,}0228$$

segue che il 16% otterrà A, il 34% B, il 34% C, il 14% D ed il 2% E.

Esempio. Un bit deve essere trasmesso attraverso un canale soggetto a rumore. Per ridurre la possibilità di errore si invia il valore 2 quando il messaggio è **1** ed il valore -2 quando il messaggio è **0**. Se si spedisce x, allora il valore ricevuto è $R = x + Z$, per $x = \pm 2$, dove Z è una variabile aleatoria normale standard che descrive il rumore del canale. Si determini la probabilità di errore se il valore ricevuto è così decodificato:

$$\begin{aligned} &\text{se } R \geq 0{,}5 \ \text{ si decodifica } \mathbf{1}, \\ &\text{se } R < 0{,}5 \ \text{ si decodifica } \mathbf{0}. \end{aligned}$$

Soluzione. Si hanno 2 tipi di errore: uno è che il messaggio **1** sia erroneamente interpretato come **0**, l'altro è che **0** sia erroneamente interpretato come **1**. Il primo tipo di errore si realizza se il messaggio è **1** e $R = 2 + Z < 0{,}5$, il secondo si realizza se il messaggio è **0** e $R = -2 + Z \geq 0{,}5$. Posto $p_{\mathbf{k}} = P(\text{errore} \,|\, \text{il messaggio è } \mathbf{k})$, si ha

$$\begin{aligned} p_{\mathbf{1}} &= P\big(2 + Z < 0{,}5\big) = P\big(Z < -1{,}5\big) = \Phi\big(-1{,}5\big) = 1 - \Phi\big(1{,}5\big) = 0{,}0668 \\ p_{\mathbf{0}} &= P\big(-2 + Z \geq 0{,}5\big) = P\big(Z \geq 2{,}5\big) = 1 - \Phi\big(2{,}5\big) = 0{,}0062. \end{aligned}$$

5.4.1 L'approssimazione normale della distribuzione binomiale

Teorema di DeMoivre-Laplace. Sia S_n il numero di successi che si realizzano in n prove indipendenti, in ognuna delle quali il successo ha probabilità p, e sia Z una variabile normale standard. Per $a < b$ si ha:

$$\lim_{n\to+\infty} P\left\{a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right\} = P\,(a \leq Z \leq b) = \Phi\big(b\big) - \Phi\big(a\big).$$

Il teorema di DeMoivre-Laplace afferma che, se n è grande, la standardizzata di una variabile aleatoria binomiale di parametri n e p ha approssimativamente la stessa distribuzione di una variabile aleatoria normale standard.

Esempio. Sia X il numero di volte che esce testa lanciando $n = 40$ volte una moneta equa ($p = 0{,}5$). Determinare $P(X = 20)$. Ricavarne poi un'approssimazione normale e confrontare il risultato con la soluzione esatta. Cosa cambia se $p = 0{,}4$?

Soluzione. Per $p = 0{,}5$ il risultato esatto è

$$P(X = 20) = \binom{40}{20} (0{,}5)^{40} \simeq 0{,}1254.$$

Usiamo l'approssimazione normale. Mentre la variabile aleatoria binomiale è discreta a valori interi, la normale è una variabile continua ed è conveniente scrivere $P(X = i)$ come $P(i - 1/2 < X < i + 1/2)$ prima di applicare l'approssimazione normale. Si ha:

$$P(X = 20) = P\big(19{,}5 \le X < 20{,}5\big) = P\Big(\frac{19{,}5 - 20}{\sqrt{10}} < \frac{X - 20}{\sqrt{10}} < \frac{20{,}5 - 20}{\sqrt{10}}\Big)$$

$$\simeq P\Big(- 0{,}16 < \frac{X - 20}{\sqrt{10}} < 0{,}16\Big) = \Phi\big(0{,}16\big) - \Phi\big(- 0{,}16\big) = 2\,\Phi\big(0{,}16\big) - 1 = 0{,}1272.$$

Per $p = 0{,}4$ il risultato esatto e l'approssimazione normale sono:

$$P(X = 20) = \binom{40}{20} (0{,}4)^{20} (0{,}6)^{20} \simeq 0{,}0554,$$

$$P(X = 20) = P\Big(\frac{19{,}5 - 16}{\sqrt{9{,}6}} < \frac{X - 16}{\sqrt{9{,}6}} < \frac{20{,}5 - 16}{\sqrt{9{,}6}}\Big) \approx \Phi\big(1{,}45\big) - \Phi\big(1{,}13\big) = 0{,}0557.$$

5.5 Variabili aleatorie esponenziali

Una variabile aleatoria continua X è detta esponenziale (o distribuita esponenzialmente) di parametro $\lambda > 0$ se la densità di X è data da

$$f(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

La funzione di distribuzione di una variabile aleatoria esponenziale è data da

$$F(x) = P(X \leq x) = \int_0^x \lambda \mathrm{e}^{-\lambda t} dt = 1 - \mathrm{e}^{-\lambda x}, \qquad x \geq 0.$$

Si noti che $\int_{-\infty}^{\infty} f(t)dt = \int_0^{\infty} \lambda \mathrm{e}^{-\lambda t} dt = \lim_{x \to \infty} F(x) = \lim_{x \to \infty} \left(1 - \mathrm{e}^{-\lambda x}\right) = 1.$

Inoltre, per $x \geq 0$ risulta

$$P(X > x) = 1 - F(x) = \mathrm{e}^{-\lambda x}.$$

Esempio. Se X è una variabile aleatoria esponenziale di parametro λ, allora valore atteso e varianza sono: $E[X] = 1/\lambda$ e $Var(X) = 1/\lambda^2$.

Soluzione. Ricordando l'espressione della densità esponenziale, si ha

$$E[X] = \int_0^{\infty} x\, f(x)\, dx = \int_0^{\infty} x\lambda \mathrm{e}^{-\lambda x} dx$$

e quindi, integrando per parti, si ottiene

$$E[X] = -x\mathrm{e}^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} \mathrm{e}^{-\lambda x} dx = \frac{1}{\lambda} \int_0^{\infty} f(x) dx = \frac{1}{\lambda}.$$

Per calcolare la varianza di X dobbiamo prima calcolare $E[X^2]$. Si ha

$$E[X^2] = \int_0^{\infty} x^2 \lambda \mathrm{e}^{-\lambda x} \mathrm{d}x = -x^2 \mathrm{e}^{-\lambda x}\Big|_0^{+\infty} + \int_0^{\infty} 2x\mathrm{e}^{-\lambda x} \mathrm{d}x = \frac{2}{\lambda^2}.$$

Quindi

$$Var(X) = E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

Esempio. Sia X una variabile aleatoria esponenziale di parametro λ. Valutare $P\{X > E(X)\}$ e determinare il valore m tale che $P\{X > m\} = 1/2$.

Soluzione. Ricordando che $P(X > x) = \mathrm{e}^{-\lambda x}$ per $x \geq 0$, da $E(X) = \lambda^{-1}$ segue

$$P\{X > E(X)\} = P\{X > \lambda^{-1}\} = \mathrm{e}^{-1} = 0{,}3679.$$

Risulta $P\{X > m\} = 1/2$ per

$$\mathrm{e}^{-\lambda m} = \frac{1}{2}, \qquad \text{ossia per} \qquad m = \frac{1}{\lambda} \ln 2 = \frac{1}{\lambda}\, 0{,}6931.$$

Una variabile aleatoria X si dice priva di memoria se

$$P(X > s + t \,|\, X > t) = P(X > s) \qquad \text{per ogni } s, t \geq 0.$$

Se X descrive il tempo di vita di un dispositivo, la proprietà di assenza di memoria afferma che la probabilità che esso funzioni per almeno $s + t$ ore, sapendo che ha funzionato per t ore, è uguale alla probabilità iniziale che esso funzioni per almeno s ore. In altri termini, se il dispositivo ha funzionato per t ore, la distribuzione del tempo di vita residuo è uguale a quella del tempo di vita originale.

La condizione di assenza di memoria può essere scritta equivalentemente come

$$P(X > s + t) = P(X > t)\, P(X > s) \qquad \text{per ogni } s, t \geq 0.$$

La variabile aleatoria esponenziale soddisfa la proprietà di assenza di memoria, essendo

$$P(X > s + t) = e^{-\lambda(s+t)} = e^{-\lambda t}\, e^{-\lambda s} = P(X > t)\, P(X > s), \qquad s, t \geq 0,$$

ed è l'unica variabile aleatoria continua a godere di tale proprietà.

Esempio. Supponiamo che la durata di una telefonata in minuti sia una variabile aleatoria esponenziale di parametro $\lambda = \frac{1}{10}$. Calcolare la probabilità che la durata sia

(a) maggiore di 5 minuti;

(b) tra 10 e 20 minuti;

(c) tra 15 e 20 minuti sapendo che è maggiore di 10 minuti.

Soluzione. Se X rappresenta la durata della telefonata, risulta

$$P(X > 5) = 1 - F(5) = 1 - (1 - e^{-\frac{1}{10}5}) = e^{-0,5} \simeq 0{,}6065;$$

$$P(10 < X < 20) = F(20) - F(10) = 1 - e^{-1} - 1 + e^{-2} = e^{-1} - e^{-2} \simeq 0{,}2325;$$

infine

$$\begin{aligned} P(15 < X < 20 \,|\, X > 10) &= P(X > 15 \,|\, X > 10) - P(X > 20 \,|\, X > 10) \\ &= P(X > 5) - P(X > 10) = e^{-0,5} - e^{-1} \simeq 0{,}2387, \end{aligned}$$

avendo fatto uso della proprietà di assenza di memoria.

Esempio. Consideriamo un ufficio postale con due sportelli, e supponiamo che quando il Sig. Carlo entra nell'ufficio veda la Sig.ra Alice e il Sig. Bruno agli sportelli. Supponiamo che il Sig. Carlo acceda al primo sportello che si libera. Se il tempo che un impiegato dedica ad un cliente è distribuito esponenzialmente con parametro λ, qual è la probabilità che, sui tre clienti, il Sig. Carlo sia l'ultimo a lasciare l'ufficio?

Soluzione. Prendiamo in esame il tempo necessario affinché il Sig. Carlo trovi uno sportello libero. A questo punto, la Sig.ra Alice o il Sig. Bruno hanno appena lasciato lo sportello e uno dei due può trovarsi ancora allo sportello. Tuttavia, per la proprietà di assenza di memoria della variabile esponenziale, il tempo residuo di permanenza allo sportello di quest'altra persona (Alice o Bruno) è distribuito esponenzialmente con parametro λ. Per simmetria la probabilità che questa persona finisca prima del Sig. Carlo è $1/2$, in quanto anche il tempo di permanenza allo sportello del Sig. Carlo è distribuito esponenzialmente con parametro λ.

Esempio. Si supponga che il numero di chilometri che un'auto possa percorrere prima che la batteria ceda sia una variabile aleatoria esponenziale di valore atteso 10000. Se si desidera fare un viaggio di 5000 chilometri, qual è la probabilità di effettuarlo senza dover cambiare la batteria? Cosa si può dire se la distribuzione non è esponenziale?
Soluzione. Dalla proprietà di assenza di memoria della variabile esponenziale segue che il tempo di vita rimanente (in migliaia di chilometri) della batteria è una variabile esponenziale di parametro $\lambda = 1/10$. La probabilità cercata vale quindi

$$P(\text{tempo di vita rimanente} > 5) = 1 - F(5) = \mathrm{e}^{-5\lambda} = \mathrm{e}^{-1/2} \simeq 0{,}6065.$$

Se la funzione di distribuzione $F(x)$ non fosse esponenziale, la probabilità sarebbe

$$P(\text{tempo di vita} > 5 + t \,|\, \text{tempo di vita} > t) = \frac{1 - F(t+5)}{1 - F(t)},$$

dove t è il numero di chilometri già percorsi prima dell'inizio del viaggio. Quindi, se la distribuzione non è esponenziale, occorre una informazione aggiuntiva (il valore di t) per poter calcolare la probabilità richiesta.

5.6 Distribuzione di una funzione di variabile aleatoria

Supponiamo di conoscere la distribuzione di X e di voler determinare la distribuzione di $g(X)$. A tal fine è necessario esprimere l'evento $\{g(X) \leq y\}$ in termini dell'appartenenza di X ad un insieme.

Esempio. Sia X uniformemente distribuita su $(0,1)$. Determinare la funzione di distribuzione e la densità di probabilità della variabile aleatoria $Y = X^\alpha$, con $\alpha > 0$.
Soluzione. Ricordando che $F_X(x) = x$ per $0 \leq x \leq 1$, per $0 \leq y \leq 1$ si ha

$$F_Y(y) = P(Y \leq y) = P(X^\alpha \leq y) = P(X \leq y^{1/\alpha}) = F_X(y^{1/\alpha}) = y^{1/\alpha}.$$

Pertanto, la densità di probabilità di Y è data da

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{\alpha}\, y^{1/\alpha - 1}, \qquad 0 < y < 1.$$

Ad esempio, per $\alpha = 2$ si ha $f_Y(y) = \dfrac{1}{2\sqrt{y}}$, per $\alpha = 1/2$ si ha $f_Y(y) = 2y$, $0 < y < 1$.

Esempio. Sia X una variabile aleatoria continua con densità $f_X(x)$. Determinare la funzione di distribuzione e la densità di probabilità di $Y = X^2$ e $U = |X|$.

Soluzione. La funzione di distribuzione di $Y = X^2$, per $y \geq 0$, è data da

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Derivando rispetto a y si ottiene la densità

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{2\sqrt{y}} \left[f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right], \qquad y > 0.$$

Determiniamo ora la funzione di distribuzione di $U = |X|$. Per $u \geq 0$, si ha

$$F_U(u) = P(U \leq u) = P(|X| \leq u) = P(-u \leq X \leq u) = F_X(u) - F_X(-u).$$

La densità è pertanto

$$f_U(u) = \frac{d}{du} F_U(u) = f_X(u) + f_X(-u), \qquad u > 0.$$

Teorema. Sia X una variabile aleatoria continua con densità f_X, e $g(x)$ una funzione strettamente monotona (crescente o decrescente) e derivabile con continuità. Allora la variabile aleatoria Y definita da $Y = g(X)$ è continua e la sua densità è data da

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)] \left| \dfrac{d}{dy} g^{-1}(y) \right| & \text{se } y = g(x) \text{ per qualche } x \\ 0 & \text{se } y \neq g(x) \text{ per ogni } x, \end{cases}$$

dove $g^{-1}(y)$ è l'unica x tale che $y = g(x)$.

Dimostrazione. Supponiamo $g(x)$ crescente. Se y appartiene a tale intervallo è $y = g(x)$ per qualche x. Allora, se $Y = g(X)$, si ha

$$F_Y(y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

In accordo con l'enunciato del teorema, derivando si ottiene

$$f_Y(y) = f_X[g^{-1}(y)] \frac{d}{dy} g^{-1}(y),$$

dato che $g^{-1}(y)$ è non decrescente e la sua derivata è quindi non negativa.

Se y non appartiene all'insieme dei valori di g, cioè $y \neq g(x)$ per ogni x, allora $F_Y(y)$ è uguale a 0 o a 1 ed in ogni caso $f_Y(y) = 0$.

Esempio. Sia X una variabile aleatoria continua non negativa con densità f e sia $Y = X^n$. Determinare la densità $f_Y(y)$.

Soluzione. Se $g(x) = x^n$, allora

$$g^{-1}(y) = y^{1/n}, \qquad \frac{d}{dy}\, g^{-1}(y) = \frac{1}{n}\, y^{1/n-1}.$$

In virtù del teorema si ha pertanto: $f_Y(y) = \dfrac{1}{n}\, y^{1/n-1} f(y^{1/n})$.

Esempio. Sia X una variabile aleatoria uniforme in $(0, 1)$ e sia $\lambda > 0$. Mostrare che la densità di $Y = g(X) = -(\ln X)/\lambda$ è esponenziale di parametro λ.

Soluzione. Risulta $g^{-1}(y) = \mathrm{e}^{-\lambda y}$, e quindi per $y > 0$ si ha:

$$f_Y(y) = f_X[g^{-1}(y)] \left| \frac{d}{dy} g^{-1}(y) \right| = 1 \cdot \left| \frac{d}{dy} \mathrm{e}^{-\lambda y} \right| = \lambda\, \mathrm{e}^{-\lambda y}.$$

Calcolo delle Probabilità e Statistica Matematica – A.A. 2009/10

CAPITOLO 6 – Leggi congiunte di variabili aleatorie

6.1 Funzioni di distribuzione congiunte

Definizione. Date due variabili aleatorie X e Y, la funzione di distribuzione congiunta di (X, Y) è

$$F(x, y) = P\left(X \leq x, Y \leq y\right), \qquad x, y \in \mathbb{R}.$$

La funzione di distribuzione di X può essere ottenuta da quella congiunta di (X, Y) come segue:

$$\begin{aligned}\lim_{y \to +\infty} F(x, y) &= \lim_{y \to +\infty} P\left(X \leq x, Y \leq y\right) = P\left(\lim_{y \to +\infty} \{X \leq x, Y \leq y\}\right) \\ &= P\left(X \leq x, Y < +\infty\right) = P\left(X \leq x\right) = F_X(x).\end{aligned}$$

In maniera analoga la funzione di distribuzione di Y è data da

$$\lim_{x \to +\infty} F(x, y) = F_Y(y).$$

Le funzioni $F_X(x)$ e $F_Y(y)$ sono dette funzioni di distribuzione *marginali* di X e di Y.

Notiamo che sussistono i seguenti limiti:

$$\lim_{x\to+\infty}\lim_{y\to+\infty} F(x,y)=1, \qquad \lim_{x\to-\infty} F(x,y)=0, \qquad \lim_{y\to-\infty} F(x,y)=0.$$

Probabilità riferite a (X,Y) si possono esprimere in termini di $F(x,y)$.
Ad esempio, la probabilità che X sia maggiore di a e Y sia maggiore di b è data da

$$\begin{aligned} P(X>a,Y>b) &= 1-P(\{X\le a\}\cup\{Y\le b\}) \\ &= 1-[P(X\le a)+P(Y\le b)-P(X\le a,Y\le b)] \\ &= 1-F_X(a)-F_Y(b)+F(a,b). \end{aligned}$$

Inoltre, per $a_1<a_2$ e $b_1<b_2$ risulta

$$\begin{aligned} &P(a_1<X\le a_2, b_1<Y\le b_2) \\ &\quad = P(X\le a_2, b_1<Y\le b_2)-P(X\le a_1, b_1<Y\le b_2) \\ &\quad = F(a_2,b_2)-F(a_2,b_1)-F(a_1,b_2)+F(a_1,b_1). \end{aligned}$$

Procedendo al limite per $a_2\to\infty$ e $b_2\to\infty$ si ottiene la relazione precedente.

Nel caso in cui X e Y siano variabili aleatorie discrete, è conveniente definire la densità discreta congiunta di (X, Y) come segue

$$p(x, y) = P\left(X = x, Y = y\right).$$

Notiamo che

$$p(x, y) \geq 0 \quad \forall x, y, \qquad \sum_x \sum_y p(x, y) = 1.$$

Per ogni sottoinsieme $\mathcal{D}$ di $\mathbb{R}^2$ si ha

$$P\{(X, Y) \in \mathcal{D}\} = \sum_{(x,y)\in\mathcal{D}} p(x, y).$$

La densità discreta di X può essere ottenuta da $p(x, y)$ al seguente modo:

$$p_X(x) = P\left(X = x\right) = \sum_{y\,:\,p(x,y)>0} p(x, y);$$

analogamente la densità discreta di Y è data da:

$$p_Y(y) = P\left(Y = y\right) = \sum_{x\,:\,p(x,y)>0} p(x, y).$$

Esempio. Vengono scelte a caso 3 biglie da un'urna contenente 3 biglie rosse, 4 bianche e 5 blu. Se denotiamo con X e Y, rispettivamente, il numero di biglie rosse e bianche scelte, la densità congiunta di X e Y, $p(x,y) = P(X = x, Y = y)$, è data da

$$p(0,0) = \binom{5}{3}\Big/\binom{12}{3} = \frac{10}{220}, \qquad p(0,1) = \binom{4}{1}\binom{5}{2}\Big/\binom{12}{3} = \frac{40}{220},$$

$$p(0,2) = \binom{4}{2}\binom{5}{1}\Big/\binom{12}{3} = \frac{30}{220}, \qquad p(0,3) = \binom{4}{3}\Big/\binom{12}{3} = \frac{4}{220},$$

$$p(1,0) = \binom{3}{1}\binom{5}{2}\Big/\binom{12}{3} = \frac{30}{220}, \qquad p(1,1) = \binom{3}{1}\binom{4}{1}\binom{5}{1}\Big/\binom{12}{3} = \frac{60}{220},$$

$$p(1,2) = \binom{3}{1}\binom{4}{2}\Big/\binom{12}{3} = \frac{18}{220}, \qquad p(2,0) = \binom{3}{2}\binom{5}{1}\Big/\binom{12}{3} = \frac{15}{220},$$

$$p(2,1) = \binom{3}{2}\binom{4}{1}\Big/\binom{12}{3} = \frac{12}{220}, \qquad p(3,0) = \binom{3}{3}\Big/\binom{12}{3} = \frac{1}{220},$$

Le probabilità $p(x, y)$ possono essere facilmente tabulate, come di seguito mostrato.

$$p(x, y) = P(X = x, Y = y) = \frac{\binom{3}{x}\binom{4}{y}\binom{5}{3-x-y}}{\binom{12}{3}} \qquad (x, y = 0, 1, 2, 3;\ x + y \leq 3)$$

$x \setminus y$	0	1	2	3	$P(X = x)$
0	10/220	40/220	30/220	4/220	84/220
1	30/220	60/220	18/220	0	108/220
2	15/220	12/220	0	0	27/220
3	1/220	0	0	0	1/220
$P(Y = y)$	56/220	112/220	48/220	4/220	1

Le densità discrete di X e Y, dette *densità marginali*, si ottengono calcolando le somme sulle righe e sulle colonne, rispettivamente, e appaiono ai margini della tabella. Si ha:

$$P(X \geq 2) = \frac{28}{220} \approx 0{,}1273 \qquad P(Y \leq 1) = \frac{168}{220} \approx 0{,}7636$$

$$P(X \geq 2, Y \leq 1) = \frac{28}{220} \approx 0{,}1273 \qquad P(X = Y) = \frac{70}{220} \approx 0{,}3182.$$

Esempio. Supponiamo che il 15% delle famiglie in una certa comunità non abbia figli, che il 20% ne abbia 1, il 35% ne abbia 2 ed il 30% ne abbia 3. Supponiamo inoltre che in ogni famiglia ogni figlio sia con uguale probabilità maschio o femmina in maniera indipendente. Se si sceglie a caso una famiglia di questa comunità, allora il numero di maschi X e il numero di femmine Y hanno la seguente densità discreta congiunta

$x \setminus y$	0	1	2	3	$p_X(x)$
0	0,15	0,10	0,0875	0,0375	0,3750
1	0,10	0,175	0,1125	0	0,3875
2	0,0875	0,1125	0	0	0,2000
3	0,0375	0	0	0	0,0375
$p_Y(y)$	0,3750	0,3875	0,2000	0,0375	1

Ad esempio si ha:

$$
\begin{aligned}
p(0,0) &= P(\text{senza figli}) = 0{,}15 \\
p(0,1) &= P(1 \text{ figlio femmina}) = P(1 \text{ figlio})\, P(1 \text{ femmina} \,|\, 1 \text{ figlio}) = 0{,}2\,\frac{1}{2} = 0{,}1 \\
p(0,2) &= P(2 \text{ figli femmina}) = P(2 \text{ figli})\, P(2 \text{ femmine} \,|\, 2 \text{ figli}) = 0{,}35\, \left(\frac{1}{2}\right)^2 = 0{,}0875.
\end{aligned}
$$

Diciamo che due variabili aleatorie X e Y sono *congiuntamente (assolutamente) continue* se esiste una funzione $f(x,y)$ integrabile, tale che per ogni sottoinsieme $\mathcal{C}$ dello spazio delle coppie di numeri reali risulti

$$P\{(X,Y) \in \mathcal{C}\} = \iint_{(x,y)\in\mathcal{C}} f(x,y)\, dx\, dy.$$

La funzione $f(x,y)$ è chiamata densità congiunta di X e Y. Se $\mathcal{A}$ e $\mathcal{B}$ sono una qualsiasi coppia di sottoinsiemi di $\mathbb{R}$, allora definendo $\mathcal{C} = \{(x,y) : x \in \mathcal{A}, y \in \mathcal{B}\}$ otteniamo che

$$P\{X \in \mathcal{A}, Y \in \mathcal{B}\} = \int_{\mathcal{B}} \int_{\mathcal{A}} f(x,y)\, dx\, dy.$$

Poiché

$$F(a,b) = P\{X \in (-\infty, a], Y \in (-\infty, b]\,\} = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)\, dx\, dy,$$

differenziando segue che

$$f(a,b) = \frac{\partial^2}{\partial a\, \partial b} F(a,b).$$

Esempio. La distribuzione multinomiale. La distribuzione multinomiale si ottiene quando si ripete n volte un esperimento in condizioni di indipendenza, quando ogni esperimento può avere come risultato uno qualsiasi tra r possibili esiti, con probabilità $p_1, \ldots, p_r$, rispettivamente, tali che $p_1 + \ldots + p_r = 1$. Se denotiamo con X_i il numero degli n esperimenti che hanno dato come risultato l'esito i, allora si ha

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_r = n_r) = \frac{n!}{n_1! n_2! \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}, \quad \text{con} \quad \sum_{i=1}^{r} n_i = n.$$

Tale formula è verificata notando che ogni successione degli esiti degli n esperimenti che portano al fatto che l'esito i si verifichi esattamente n_i volte, per $i = 1, 2, \ldots, r$, avrà probabilità pari a $p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$, grazie all'ipotesi di indipendenza. La formula finale segue ricordando che $n!/(n_1!n_2! \ldots n_r!)$ è il numero di permutazioni di n oggetti di cui n_1 sono uguali tra loro, n_2 sono uguali tra loro, ... , n_r sono uguali tra loro.

Osserviamo che per $r = 2$ la distribuzione multinomiale si riduce a quella binomiale.

Esempio. Nel lanciare 9 volte un dado equilibrato, indicando con X_i, $1 \leq i \leq 6$, il numero di volte che il risultato è i, si ha

$$P(X_1 = 3, X_2 = 2, X_3 = 2, X_4 = 1, X_5 = 1, X_6 = 0)$$
$$= \frac{9!}{3!2!2!1!1!0!}\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^0 = \frac{9!}{3!2!2!}\left(\frac{1}{6}\right)^9 \approx 0{,}0015.$$

Esempio. Siano p, q, $1-p-q$ le probabilità che una squadra di calcio vinca, perda, pareggi una partita. Supponendo di giocare n partite indipendenti, con probabilità costanti, posto $X =$"numero di vincite" e $Y =$"numero di sconfitte", risulta

$$P(X = x, Y = y) = \frac{n!}{x!y!(n-x-y)!}\, p^x q^y (1-p-q)^{n-x-y}.$$

6.2 Variabili aleatorie indipendenti

Le variabili aleatorie X ed Y si dicono *indipendenti* se, per ogni coppia $\mathcal{A}$ e $\mathcal{B}$ di sottoinsiemi di $\mathbb{R}$, si ha

$$P\left(X \in \mathcal{A}, Y \in \mathcal{B}\right) = P\left(X \in \mathcal{A}\right) P\left(Y \in \mathcal{B}\right).$$

In altre parole, X e Y sono indipendenti se gli eventi $\{X \in \mathcal{A}\}$ e $\{Y \in \mathcal{B}\}$ sono indipendenti per ogni $\mathcal{A}$ e $\mathcal{B}$.

La condizione di indipendenza si può equivalentemente esprimere richiedendo che per ogni coppia di numeri reali x, y risulti

$$P\left(X \leq a, Y \leq b\right) = P\left(X \leq a\right)\ P\left(Y \leq b\right),$$

In termini della distribuzione congiunta $F(x,y)$ di (X,Y), si ha che X e Y sono indipendenti se e solo se

$$F(x,y) = F_X(x)\, F_Y(y), \qquad \forall x, y \in \mathbb{R}.$$

Proposizione. Le variabili aleatorie discrete X ed Y sono indipendenti se e solo se la loro densità discreta congiunta può essere espressa come

$$p(x,y) = p_X(x)\, p_Y(y), \qquad \text{per ogni}\, x, y.$$

Dimostrazione. Se vale $P(X \in \mathcal{A}, Y \in \mathcal{B}) = P(X \in \mathcal{A})\, P(Y \in \mathcal{B})$, allora la tesi segue ponendo $\mathcal{A} = \{x\}$ e $\mathcal{B} = \{y\}$. Viceversa, se vale $p(x,y) = p_X(x)\, p_Y(y)$ per ogni x, y, allora per ogni coppia di sottoinsiemi $\mathcal{A}$ e $\mathcal{B}$ di $\mathbb{R}$ risulta

$$\begin{aligned} P(X \in \mathcal{A}, Y \in \mathcal{B}) &= \sum_{x \in \mathcal{A}} \sum_{y \in \mathcal{B}} p(x,y) = \sum_{x \in \mathcal{A}} \sum_{y \in \mathcal{B}} p_X(x)\, p_Y(y) \\ &= \sum_{x \in \mathcal{A}} p_X(x) \sum_{y \in \mathcal{B}} p_Y(y) = P(X \in \mathcal{A})\; P(Y \in \mathcal{B}). \end{aligned}$$

Nel caso di variabili congiuntamente continue la condizione d'indipendenza è equivalente a richiedere che la densità congiunta possa essere fattorizzata come

$$f(x,y) = f_X(x)\, f_Y(y), \qquad \text{per ogni}\, x, y.$$

Esempio. Nell'esperimento che consiste nel lancio di una moneta ripetuto 3 volte, sia X il numero di volte che esce testa e sia Y il numero di variazioni, ossia quanti lanci danno risultati diversi dal lancio precedente. Stabilire se X e Y sono indipendenti.

Soluzione. È evidente che X e Y hanno distribuzione binomiale con parametri $(n, p) = (3, \frac{1}{2})$ e $(2, \frac{1}{2})$, rispettivamente. Ricaviamo ora la densità congiunta di (X, Y):

ω	$X(\omega)$	$Y(\omega)$
ccc	0	0
cct	1	1
ctc	1	2
ctt	2	1
tcc	1	1
tct	2	2
ttc	2	1
ttt	3	0

$x \backslash y$	0	1	2	$p_X(x)$
0	1/8	0	0	1/8
1	0	1/4	1/8	3/8
2	0	1/4	1/8	3/8
3	1/8	0	0	1/8
$p_Y(y)$	1/4	1/2	1/4	1

Risulta ad esempio $p(0,0) = \frac{1}{8} \neq p_X(0)\, p_Y(0) = \frac{1}{8} \cdot \frac{1}{4}$; ne segue che X e Y non sono indipendenti.

Esempio. Supponiamo che vengano eseguite $n+m$ prove indipendenti, ognuna delle quali abbia probabilità pari a p di risultare in un successo. Se X è il numero di successi nelle prime n prove e Y il numero di successi nelle m prove successive, allora X e Y sono indipendenti, in quanto conoscere il numero dei successi nelle prime n prove non modifica la distribuzione del numero di successi nelle ulteriori m prove, in virtù dell'ipotesi di indipendenza delle prove.

Infatti, per $x = 0, 1, \ldots, n$ e $y = 0, 1, \ldots, m$ si ha

$$P(X = x, Y = y) = P(X = x)\, P(Y = y) = \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} p^y (1-p)^{m-y}.$$

Osserviamo che se $Z = X + Y$ è il numero totale di successi nelle $n+m$ prove, allora X e Z sono dipendenti, ossia non indipendenti.

Esempio. Supponiamo che il numero di persone che entrano in un ufficio postale in un dato giorno sia una variabile di Poisson di parametro λ. Si provi che, se ogni persona che entra nell'ufficio postale è un uomo con probabilità p e donna con probabilità $1-p$, allora il numero di uomini e donne che entrano nell'ufficio postale sono variabili aleatorie di Poisson indipendenti di parametri rispettivamente pari a λp e $\lambda(1-p)$.

Soluzione. Denotiamo con X ed Y il numero di uomini e donne che entrano nell'ufficio postale. Condizionando rispetto a $X+Y$ si ottiene

$$\begin{aligned} P(X=i, Y=j) &= P(X=i, Y=j \,|\, X+Y=i+j)\, P(X+Y=i+j) \\ &\quad + P(X=i, Y=j \,|\, X+Y \neq i+j)\, P(X+Y \neq i+j). \end{aligned}$$

Ovviamente si ha $P(X=i, Y=j \,|\, X+Y \neq i+j) = 0$, e quindi

$$P(X=i, Y=j) = P(X=i, Y=j \,|\, X+Y=i+j)\, P(X+Y=i+j).$$

Essendo $X+Y$ il numero totale di persone che entrano nell'ufficio postale, risulta

$$P(X+Y=i+j) = \mathrm{e}^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}.$$

Dato che $i+j$ persone entrano nell'ufficio e con probabilità p ognuna di esse è un uomo, la probabilità che esattamente i di essi siano un uomo (e quindi j siano donne) è la densità discreta di una variabile binomiale di parametri $i+j$ e p valutata in i, ossia

$$P(X=i, Y=j \mid X+Y=i+j) = \binom{i+j}{i} p^i (1-p)^j.$$

In conclusione si perviene all'indipendenza di X e Y notando che risulta

$$\begin{aligned} P(X=i, Y=j) &= \binom{i+j}{i} p^i (1-p)^j \, \mathrm{e}^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = \mathrm{e}^{-\lambda} \frac{(\lambda p)^i \, [\lambda(1-p)]^j}{i! \, j!} \\ &= \frac{\mathrm{e}^{-\lambda p} (\lambda p)^i}{i!} \, \frac{\mathrm{e}^{-\lambda(1-p)} [\lambda(1-p)]^j}{j!} \qquad (i=0,1,\ldots;\ j=0,1,\ldots) \end{aligned}$$

e

$$P(X=i) = \frac{\mathrm{e}^{-\lambda p} (\lambda p)^i}{i!} \sum_{j=0}^{\infty} \frac{\mathrm{e}^{-\lambda(1-p)} [\lambda(1-p)]^j}{j!} = \frac{\mathrm{e}^{-\lambda p} (\lambda p)^i}{i!} \quad (i=0,1,\ldots)$$

$$P(Y=j) = \frac{\mathrm{e}^{-\lambda(1-p)} [\lambda(1-p)]^j}{j!} \sum_{i=0}^{\infty} \frac{\mathrm{e}^{-\lambda p} (\lambda p)^i}{i!} = \frac{\mathrm{e}^{-\lambda(1-p)} [\lambda(1-p)]^j}{j!} \quad (j=0,1,\ldots).$$

6.3 Somme di variabili aleatorie indipendenti

Molto spesso si è interessati a calcolare la distribuzione di $X+Y$ a partire dalle distribuzioni delle due variabili aleatorie X e Y, sotto l'ipotesi aggiuntiva che le due variabili siano indipendenti. Nel caso in cui le variabili X e Y siano assolutamente continue la variabile aleatoria $X+Y$ è a sua volta continua con densità

$$f_{X+Y}(a) = \int_{-\infty}^{+\infty} f_X(a-y) f_Y(y)\, dy.$$

In particolare si ha il seguente risultato.

Proposizione. Se X_i, $i = 1, 2, \ldots, n$, sono variabili aleatorie indipendenti con distribuzione normale di parametri, rispettivamente, μ_i e σ_i^2, allora $\displaystyle\sum_{i=1}^{n} X_i$ è una variabile aleatoria normale di parametri $\displaystyle\sum_{i=1}^{n} \mu_i$ e $\displaystyle\sum_{i=1}^{n} \sigma_i^2$.

Esempio. Somme di variabili aleatorie indipendenti di Poisson. Se X e Y sono due variabili aleatorie di Poisson indipendenti di parametri λ_1 e λ_2, rispettivamente, si calcoli la densità discreta di $X+Y$.

Soluzione. Dalle ipotesi fatte segue che per $n = 0, 1, \ldots$

$$\begin{aligned} P(X+Y=n) &= \sum_{k=0}^{n} P\left(X=k, Y=n-k\right) = \sum_{k=0}^{n} P\left(X=k\right)\, P\left(Y=n-k\right) \\ &= \sum_{k=0}^{n} \mathrm{e}^{-\lambda_1} \frac{\lambda_1^k}{k!}\, \mathrm{e}^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} = \mathrm{e}^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \frac{\lambda_1^k\, \lambda_2^{n-k}}{k!\,(n-k)!} \\ &= \frac{\mathrm{e}^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!} \lambda_1^k\, \lambda_2^{n-k} = \frac{\mathrm{e}^{-(\lambda_1+\lambda_2)}}{n!} (\lambda_1+\lambda_2)^n, \end{aligned}$$

avendo fatto uso della formula del binomio.

Si ricava che $X+Y$ ha distribuzione di Poisson di parametro $\lambda_1+\lambda_2$.

Esempio. Somme di variabili aleatorie binomiali indipendenti. Se X e Y sono due variabili aleatorie binomiali indipendenti di parametri, rispettivamente, (n, p) e (m, p), mostrare che $X + Y$ è binomiale di parametri $(n + m, p)$.

Soluzione. Dalle ipotesi fatte, per $k = 0, 1, \ldots, n + m$ segue

$$\begin{aligned} P(X+Y=k) &= \sum_{i=0}^{k} P\left(X=i, Y=k-i\right) = \sum_{i=0}^{k} P\left(X=i\right) P\left(Y=k-i\right) \\ &= \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \binom{m}{k-i} p^{k-i} (1-p)^{m-k+i} \\ &= p^k (1-p)^{n+m-k} \sum_{i=0}^{k} \binom{n}{i} \binom{m}{k-i} = \binom{n+m}{k} p^k (1-p)^{n+m-k}, \end{aligned}$$

avendo utilizzato l'identità combinatoria (formula di Vandermonde)

$$\binom{n+m}{k} = \sum_{i=0}^{k} \binom{n}{i} \binom{m}{k-i}.$$

6.4 Distribuzioni condizionate: il caso discreto

Date due variabili aleatorie discrete X e Y, si definisce la densità discreta condizionata di X dato $Y = y$, come

$$p_{X|Y}(x|y) = P\left(X = x \,|\, Y = y\right) = \frac{P\left(X = x\,, Y = y\right)}{P\left(Y = y\right)} = \frac{p(x, y)}{p_Y(y)},$$

per tutti i valori di y per i quali $p_Y(y) > 0$. In maniera analoga, per tutti gli y tali che $p_Y(y) > 0$, si definisce la funzione di distribuzione condizionata di X dato $Y = y$:

$$F_{X|Y}(x|y) = P\left(X \leq x \,|\, Y = y\right) = \sum_{a \leq x} p_{X|Y}(a|y).$$

Se X è indipendente da Y, le funzioni condizionate $p_{X|Y}(x|y)$ e $F_{X|Y}(x|y)$ coincidono con le versioni non condizionate. Infatti se X e Y sono indipendenti, si ha

$$\begin{aligned} p_{X|Y}(x|y) &= P\left(X = x \,|\, Y = y\right) = \frac{P\left(X = x, Y = y\right)}{P\left(Y = y\right)} \\ &= \frac{P\left(X = x\right)\, P\left(Y = y\right)}{P\left(Y = y\right)} = P\left(X = x\right). \end{aligned}$$

Esempio. Nell'esperimento che consiste nell'effettuare 5 estrazioni a caso da un'urna contenente 2 biglie nere e 3 biglie bianche, denotiamo con X l'estrazione in cui fuoriesce la prima biglia nera, e con Y il numero di biglie nere fuoriuscite nelle prime 3 estrazioni. Calcolare le densità discrete condizionate di X, dato $Y = y$, per $y = 0, 1, 2$.

Soluzione. Dall'esame delle $\binom{5}{2} = 10$ sequenze di estrazioni possibili segue la densità discreta congiunta di (X, Y), da cui ricava la densità discreta di Y; si perviene quindi alle densità discrete condizionate di X, dato $Y = y$, per $y = 0, 1, 2$.

$x\backslash y$	0	1	2	$p_X(x)$
1	0	2/10	2/10	4/10
2	0	2/10	1/10	3/10
3	0	2/10	0	2/10
4	1/10	0	0	1/10
$p_Y(y)$	1/10	6/10	3/10	1

x	1	2	3	4
$p_{X\|Y}(x\|0)$	0	0	0	1

x	1	2	3	4
$p_{X\|Y}(x\|1)$	1/3	1/3	1/3	0

x	1	2	3	4
$p_{X\|Y}(x\|2)$	2/3	1/3	0	0

Esempio. Supponiamo che X e Y siano due variabili aleatorie di Poisson indipendenti di parametri λ_1 e λ_2, rispettivamente. Si dimostri che la funzione di distribuzione condizionata di X dato $X+Y=n$ è binomiale di parametri n e $p=\lambda_1/(\lambda_1+\lambda_2)$.

Soluzione. Per $k=0,1,\ldots,n$ si ha che

$$\begin{aligned} P(X=k\,|\,X+Y=n) &= \frac{P(X=k, X+Y=n)}{P(X+Y=n)} = \frac{P(X=k, Y=n-k)}{P(X+Y=n)} \\ &= \frac{P(X=k)\,P(Y=n-k)}{P(X+Y=n)}. \end{aligned}$$

Ricordando che $X+Y$ ha distribuzione di Poisson di parametro $\lambda_1+\lambda_2$, e inoltre che $p=\lambda_1/(\lambda_1+\lambda_2)$, la precedente formula diventa

$$\begin{aligned} P(X=k\,|\,X+Y=n) &= \frac{\mathrm{e}^{-\lambda_1}\,\lambda_1^k}{k!}\,\frac{\mathrm{e}^{-\lambda_2}\,\lambda_2^{n-k}}{(n-k)!}\,\frac{n!}{\mathrm{e}^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n} \\ &= \frac{n!}{k!(n-k)!}\frac{\lambda_1^k\,\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n} = \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

Esempio. Da un'urna contenente k biglie nere e $n-k$ biglie bianche si effettuano n estrazioni a caso con reinserimento. Denotiamo con Y il numero di biglie nere estratte, e con X l'estrazione in cui si estrae una biglia nera per la prima volta. Determinare $p_Y(r)$, $p_{X|Y}(j|r)$ e $p_X(j)$.

Soluzione. Poiché le estrazioni sono con reinserimento, Y ha distribuzione binomiale di parametri n e $p = k/n$; quindi risulta

$$p_Y(r) = P(Y = r) = \binom{n}{r} p^r \,(1-p)^{n-r}\,, \qquad r = 0, 1, \ldots, n \qquad \left(p = \frac{k}{n}\right).$$

Notiamo che, se $r \geq 1$, $p_{X|Y}(j|r)$ è la probabilità che una sequenza casuale costituita da r biglie nere e $n-r$ biglie bianche abbia la prima biglia nera al j-esimo posto, e pertanto si ha

$$p_{X|Y}(j|r) = P(X = j|Y = r) = \binom{n-j}{r-1} \Big/ \binom{n}{r}, \qquad j = 1, 2, \ldots, n-r+1.$$

Quindi, dovendo essere $r \geq 1$ e $j \leq n - r + 1$, risulta

$$\begin{aligned} p_X(j) &= P(X = j) = \sum_{r=0}^{n} P(X = j | Y = r)\, P(Y = r) \\ &= \sum_{r=1}^{n-j+1} \frac{\binom{n-j}{r-1}}{\binom{n}{r}} \binom{n}{r} p^r\, (1-p)^{n-r} = \sum_{r=1}^{n-j+1} \binom{n-j}{r-1} p^r\, (1-p)^{n-r}. \end{aligned}$$

Ponendo $s = r - 1$ si ottiene, per $j = 1, 2, \ldots, n - r + 1$

$$\begin{aligned} p_X(j) &= \sum_{s=0}^{n-j} \binom{n-j}{s} p^{s+1}\, (1-p)^{n-s-1} \\ &= \sum_{s=0}^{n-j} \binom{n-j}{s} p^{s}\, (1-p)^{n-j-s}\, p\, (1-p)^{j-1} = p\, (1-p)^{j-1}, \end{aligned}$$

che è una probabilità di tipo geometrico.

Calcolo delle Probabilità e Statistica Matematica – A.A. 2009/10

CAPITOLO 7 – Proprietà del valore atteso

7.1 Introduzione

7.2 Valore atteso di somme di variabili aleatorie

7.2 Covarianza, Varianza di una somma e correlazioni

7.1 Introduzione

Ricordiamo che il valore atteso di una variabile aleatoria discreta X con densità $p(x)$ è

$$E[X] = \sum_x x\, p(x)$$

mentre se X è (assolutamente) continua con densità $f(x)$ si ha

$$E[X] = \int_{-\infty}^{+\infty} x\, f(x)\, dx.$$

Proposizione. Se X assume valori compresi tra a e b, allora il valore atteso è compreso tra a e b. In altre parole, se $P(a \le X \le b) = 1$ allora $a \le E[X] \le b$.

Dimostrazione. Poiché $p(x) = 0$ se x non appartiene ad $[a, b]$, nel caso discreto si ha

$$E[X] = \sum_{x:p(x)>0} x\, p(x) \ge \sum_{x:p(x)>0} a\, p(x) = a \sum_{x:p(x)>0} p(x) = a.$$

Analogamente si ricava $E[X] \le b$. La dimostrazione nel caso continuo è analoga.

7.2 Valore atteso di somme di variabili aleatorie

Siano X e Y due variabili aleatorie e g una funzione di due variabili.

Proposizione. Se X e Y hanno densità discreta congiunta $p(x, y)$, allora

$$E[g(X, Y)] = \sum_y \sum_x g(x, y)p(x, y).$$

Illustriamo un'importante applicazione di quanto su esposto. Se $E[X]$ e $E[Y]$ sono entrambe finite e poniamo $g(x, y) = x + y$, si ha

$$\begin{aligned} E[X+Y] &= \sum_x \sum_y (x+y)\, p(x, y) = \sum_x \sum_y x\, p(x, y) + \sum_x \sum_y y\, p(x, y) \\ &= \sum_x x\, p_X(x) + \sum_y y\, p_Y(y) = E[X] + E[Y]. \end{aligned}$$

Ragionando per induzione si dimostra che se $E[X_i]$ è finito per $i = 1, \ldots, n$, allora

$$E[a_1 X_1 + \cdots + a_n X_n + b] = a_1 E[X_1] + \cdots + a_n E[X_n] + b.$$

Esempio. La media campionaria. Siano $X_1, \ldots, X_n$ variabili aleatorie indipendenti e identicamente distribuite con distribuzione F e valore atteso μ. Tale sequenza costituisce un *campione casuale* della distribuzione F. La variabile aleatoria

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

è detta media campionaria di $X_1, \ldots, X_n$. Calcolare $E[\overline{X}]$.

Soluzione. Dato che $E[X_i] = \mu$, per la proprietà di linearità si ha

$$E[\overline{X}] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \mu.$$

Pertanto il valore atteso della media campionaria è μ, la media della distribuzione. La media campionaria è spesso utilizzata per dare una stima del valore atteso μ della distribuzione, se questo è sconosciuto.

Esempio. Siano X e Y due variabili aleatorie tali che $X \geq Y$. In altre parole, per ogni esito dell'esperimento, il valore assunto da X è maggiore o uguale del valore assunto da Y. Dato che $X - Y \geq 0$, si ha che $E[X - Y] \geq 0$, ossia $E[X] \geq E[Y]$.

Esempio. La disuguaglianza di Boole. Siano $A_1, \ldots, A_n$ degli eventi, e definiamo le loro variabili indicatrici

$$X_i = \begin{cases} 1 & \text{se } A_i \text{ si realizza,} \\ 0 & \text{altrimenti,} \end{cases} \qquad (i = 1, \ldots, n).$$

Poniamo $X = \sum_{i=1}^{n} X_i$, così X è il numero di eventi A_i che si realizzano. Infine sia

$$Y = \begin{cases} 1 & \text{se } X \geq 1, \\ 0 & \text{altrimenti,} \end{cases}$$

sicché Y è uguale ad 1 se si realizza almeno uno degli eventi A_i e vale 0 altrimenti.

Essendo chiaramente $X \geq Y$, dall'esempio precedente si deduce che $E[X] \geq E[Y]$. Dato che

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} P(X_i = 1) = \sum_{i=1}^{n} P(A_i)$$

e che

$$E[Y] = P\{\text{si realizza almeno uno degli } A_i\} = P\left(\bigcup_{i=1}^{n} A_i\right),$$

si ottiene infine la seguente relazione, nota come *disuguaglianza di Boole*:

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

Esempio. Valore atteso di una variabile aleatoria binomiale. Sia X una variabile aleatoria binomiale di parametri n e p. Ricordando che questa variabile rappresenta il numero di successi in n prove indipendenti, dove ogni prova ha successo con probabilità p, si ha che

$$X = X_1 + X_2 + \cdots + X_n,$$

dove

$$X_i = \begin{cases} 1 & \text{se l'}i\text{-esima prova è un successo,} \\ 0 & \text{se l'}i\text{-esima prova è un insuccesso.} \end{cases}$$

La variabile aleatoria di Bernoulli X_i ha valore atteso $E[X_i] = 0 \cdot (1 - p) + 1 \cdot p = p$. Di conseguenza risulta

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n] = n\,p.$$

Esempio. Valore atteso di una variabile aleatoria ipergeometrica. Si scelgano a caso n biglie da un'urna contenente N biglie delle quali m sono bianche. Determinare il numero atteso di biglie bianche selezionate.

Soluzione. Il numero X di biglie bianche selezionate si può rappresentare come

$$X = X_1 + X_2 + \cdots + X_m,$$

dove

$$X_i = \begin{cases} 1 & \text{se è stata scelta l'}i\text{-esima biglia,} \\ 0 & \text{altrimenti.} \end{cases}$$

Essendo

$$E[X_i] = P(X_i = 1) = P\{\text{è stata scelta l'}i\text{-esima biglia bianca }\} = \frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N},$$

si ottiene infine

$$E[X] = E[X_1] + \cdots + E[X_m] = \frac{m\,n}{N}.$$

Tale risultato può essere ottenuto anche utilizzando una rappresentazione alternativa:

$$X = Y_1 + \cdots + Y_n,$$

dove le variabili di Bernoulli

$$Y_i = \begin{cases} 1 & \text{se l'}i\text{-esima biglia selezionata è bianca,} \\ 0 & \text{altrimenti,} \end{cases} \qquad (i = 1, 2, \ldots, n).$$

non sono indipendenti, ma sono identicamente distribuite. Notiamo infatti che

$$P(Y_i = 1 \,|\, X = k) = \frac{\dbinom{n-1}{k-1}}{\dbinom{n}{k}} = \frac{\dfrac{(n-1)!}{(k-1)!\,(n-k)!}}{\dfrac{n!}{k!\,(n-k)!}} = \frac{k}{n},$$

dove a denominatore c'è il numero $\binom{n}{k}$ di sequenze booleane di lunghezza n con k elementi pari a 1 e con $n-k$ elementi pari a 0, mentre a numeratore è presente il numero $\binom{n-1}{k-1}$ di sequenze booleane di lunghezza n aventi 1 nel posto i-esimo e $k-1$ elementi pari a 1 nei rimanenti $n-1$ posti.

Notiamo che la probabilità

$$P(Y_i = 1 \,|\, X = k) = \frac{k}{n}$$

si può ottenere anche in modo diretto, tenendo presente che se si devono collocare k cifre pari a 1 e $n - k$ cifre pari a 0 in una sequenza di lunghezza n, ci sono k casi favorevoli su un totale di n affinché nel posto i-esimo della sequenza sia collocata una cifra pari a 1. Ne segue che la distribuzione di Y_i non dipende da i, essendo

$$P(Y_i = 1) = \sum_{k=0}^{n} P(Y_i = 1 \,|\, X = k)\, P(X = k) = \sum_{k=0}^{n} \frac{k}{n}\, P(X = k) = \frac{1}{n}\, E(X).$$

Si ricava pertanto

$$E(X) = n\, P(Y_i = 1) = n\, P(Y_1 = 1) = n\, \frac{m}{N}.$$

Esempio. Valore atteso del numero di accoppiamenti. N persone lanciano il loro cappello nel centro di una stanza. I cappelli vengono mescolati, e ogni persona ne prende uno a caso. Determinare il valore atteso del numero di persone che selezionano il proprio cappello.

Soluzione. Indicando con X il numero di accoppiamenti, possiamo calcolare $E[X]$ scrivendo $X = X_1 + X_2 + \cdots + X_N$, dove

$$X_i = \begin{cases} 1 & \text{se la persona } i \text{ prende il suo cappello,} \\ 0 & \text{altrimenti.} \end{cases}$$

Dato che, per ogni i, l'i-esima persona può scegliere ugualmente uno degli N cappelli, si ha

$$E[X_i] = P(X_i = 1) = \frac{1}{N},$$

da cui

$$E[X] = E[X_1] + \cdots + E[X_N] = \frac{1}{N} \cdot N = 1.$$

Quindi, in media, esattamente una persona seleziona il proprio cappello.

Esempio. Al passaggio di uno stormo di n anatre, vi sono n cacciatori che sparano all'istante, e ognuno mira ad un bersaglio a caso, indipendentemente dagli altri. Se ogni cacciatore colpisce il suo bersaglio con probabilità p, calcolare il numero atteso di anatre che non sono colpite.

Soluzione. Sia $X_i = 1$ se l'i-esima anatra non è colpita, e 0 altrimenti, per $i = 1, 2, \ldots, n$. Il numero atteso richiesto è dato da

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n].$$

Per calcolare $E[X_i] = P\{X_i = 1\}$, osserviamo che ognuno dei cacciatori colpirà, indipendentemente, l'i-esima anatra con probabilità p/n, sicché

$$P\{X_i = 1\} = \left(1 - \frac{p}{n}\right)^n$$

e quindi

$$E[X] = n\left(1 - \frac{p}{n}\right)^n.$$

Esempio. Algoritmo quick-sort. Supponiamo di disporre di un insieme di n valori distinti $x_1, x_2, \ldots, x_n$ e di volerli ordinare in modo crescente (sort). Una procedura efficace al riguardo è l'algoritmo di quick-sort, definito come segue. Se $n = 2$ l'algoritmo confronta due valori e li mette in ordine appropriato. Se $n > 2$, uno degli elementi è scelto a caso, ad esempio x_i, e si confrontano gli altri valori con x_i. I valori inferiori a x_i vengono messi tra parentesi graffe alla sinistra di x_i, mentre quelli superiori a x_i vengono messi alla destra di x_i. L'algoritmo si ripete per i valori interni alle parentesi graffe e continua finché i valori non sono tutti ordinati.
Supponiamo ad esempio di voler ordinare i seguenti 10 numeri distinti

$$5, 9, 3, 10, 11, 14, 8, 4, 17, 6.$$

Iniziamo scegliendo un numero a caso, ad esempio 10 (ogni numero può essere scelto con probabilità $1/10$). Confrontando gli altri valori con tale numero si ottiene

$$\{5, 9, 3, 8, 4, 6\}, 10, \{11, 14, 17\}$$

Rivolgiamo ora l'attenzione al sottoinsieme $\{5, 9, 3, 8, 4, 6\}$ e scegliamo a caso uno dei suoi valori, ad esempio il 6. Confrontando ognuno dei valori tra parentesi con 6 e mettendo a sinistra di esso i valori più piccoli di 6 e a destra quelli più grandi di 6 si ottiene

$$\{5, 3, 4\}, 6, \{9, 8\}, 10, \{11, 14, 17\}.$$

Considerando ora la parentesi a sinistra e scegliendo 4 per i successivi confronti, si giunge a

$$\{3\}, 4, \{5\}, 6, \{9, 8\}, 10, \{11, 14, 17\}.$$

Si continua finché non ci sono più sottoinsiemi che contengono più di un elemento.

Sia X il numero di confronti che servono all'algoritmo di quick-sort per ordinare n numeri distinti. Si ha allora che $E[X]$ è una misura dell'efficienza dell'algoritmo.
Per calcolare $E[X]$, esprimiamo X come somma di altre variabili aleatorie

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(i, j),$$

dove per $1 \le i < j \le n$, $I(i,j)$ è uguale ad 1 se i e j sono prima o poi confrontati direttamente mentre è uguale a 0 altrimenti. Quindi si ha

$$E[X] = E[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(i,j)] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[I(i,j)]$$
$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P\{i\,\mathrm{e}\,j\,\text{sono confrontati tra loro}\}.$$

Per determinare la probabilità che i e j non siano mai confrontati, osserviamo che i valori $\{i, i+1, \ldots, j-1, j\}$ sono inizialmente nella stessa parentesi e vi rimarranno se il numero scelto per il primo confronto non è compreso tra i e j. Ad esempio se il numero da confrontare con gli altri è strettamente maggiore di j, allora tutti i valori $\{i, i+1, \ldots, j-1, j\}$ andranno in una parentesi a sinistra del numero scelto, mentre se esso è strettamente inferiore a i, allora tali valori andranno messi in una parentesi a destra.

Tutti i valori $\{i, i+1, \ldots, j-1, j\}$ rimarranno quindi all'interno della stessa parentesi finché uno di essi è scelto per effettuare i confronti. Il valore scelto viene poi confrontato con gli altri compresi tra i e j. Se non è né i, né j, allora, dopo il confronto, i andrà in una parentesi alla sua sinistra e j in una parentesi alla sua destra. D'altro lato, se il valore scelto nell'insieme $\{i, i+1, \ldots, j-1, j\}$ è uguale a i o j, ci sarà un confronto diretto tra i e j.
Supponendo allora che il valore scelto sia uno dei $j-i+1$ valori tra i e j, la probabilità che si tratti di i o di j è pari a $2/(j-i+1)$. Si può quindi concludere che

$$P\{i \,\mathrm{e}\, j \,\text{sono confrontati tra loro}\} = \frac{2}{j-i+1},$$

da cui si ricava

$$E[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}.$$

Per ottenere l'ordine di grandezza di $E[X]$ per n grande, approssimiamo le somme con degli integrali. Ora

$$\sum_{j=i+1}^{n} \frac{2}{j-i+1} \approx \int_{i+1}^{n} \frac{2}{x-i+1} \mathrm{d}x = 2\log(x-i+1)\Big|_{i+1}^{n}$$
$$= 2\log(n-i+1) - 2\log(2) \approx 2\log(n-i+1).$$

Quindi

$$E[x] \approx \sum_{i=1}^{n-1} 2\log(n-i+1) \approx 2\int_{1}^{n-1} \log(n-x+1)\mathrm{d}x = 2\int_{2}^{n} \log(y)\mathrm{d}y$$
$$= 2(y\log(y) - y)\Big|_{2}^{n} \approx 2n\log(n).$$

Pertanto, per n grande, l'algoritmo di quick-sort richiede, in media, approssimativamente $2n\, log(n)$ confronti per ordinare n valori distinti.

7.3 Covarianza, Varianza di una somma e correlazioni

Proposizione. Se X e Y sono indipendenti, ed h e g sono due funzioni, allora

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Dimostrazione. Supponiamo che X ed Y siano variabili discrete con distribuzione congiunta $p(x, y)$. Allora

$$\begin{aligned} E[g(X)h(Y)] &= \sum_x \sum_y g(x)h(y)p(x,y) = \sum_x \sum_y g(x)h(y)p_X(x)p_Y(y) \\ &= \sum_x g(x)p_X(x) \sum_y h(y)p_Y(y) = E[g(X)]E[h(Y)]. \end{aligned}$$

Come il valore atteso e la varianza di una singola variabile forniscono delle informazioni sulla variabile aleatoria, così la covarianza tra due variabili aleatorie fornisce un'informazione sulla relazione tra le due variabili.

Definizione. La covarianza tra X ed Y, indicata con $Cov(X,Y)$, è definita da

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])].$$

Sviluppando il membro a destra della precedente definizione, vediamo che

$$\begin{aligned} Cov(X,Y) &= E[XY - E[X]Y - XE[Y] + E[Y]E[X]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[Y]E[X] \\ &= E[XY] - E[X]E[Y]. \end{aligned}$$

Se $Cov(X,Y) > 0$ le variabili aleatorie X ed Y si dicono *positivamente correlate*; in tal caso entrambe tendono ad assumere valore maggiore, o minore, della propria media. Al contrario, se $Cov(X,Y) < 0$ le variabili aleatorie X ed Y si dicono *negativamente correlate*; in tal caso se una tende ad assumere valore maggiore della propria media, l'altra tende ad assumere valore minore della propria media. Se $Cov(X,Y) = 0$ le variabili aleatorie X ed Y si dicono *scorrelate*, ovvero *non correlate*.

Se X e Y sono variabili indipendenti allora si ha che $Cov(X,Y)=0$, invero usando $g(x)=x$ e $h(y)=y$ nella Proposizione precedente, si ha $E[XY]=E[X]\,E[Y]$ e quindi $Cov(X,Y)=0$.

Il viceversa è falso. Un semplice esempio di due variabili dipendenti X e Y la cui covarianza è zero si ottiene supponendo che X e Y siano tali che

$$P(X=0)=P(X=1)=P(X=-1)=\frac{1}{3}, \qquad Y=\begin{cases}0 & \text{se } X\neq 0,\\ 1 & \text{se } X=0.\end{cases}$$

Essendo $XY=0$ si ha che $E[XY]=0$. Inoltre, $E[X]=0$ e quindi

$$Cov(X,Y)=E[XY]-E[X]\,E[Y]=0.$$

Tuttavia chiaramente X e Y non sono indipendenti, come si può ricavare dalla tabella seguente:

$x\backslash y$	0	1	$p_X(x)$
-1	$1/3$	0	$1/3$
0	0	$1/3$	$1/3$
1	$1/3$	0	$1/3$
$p_Y(y)$	$2/3$	$1/3$	1

Proposizione. La covarianza possiede le seguenti proprietà:

(i) $Cov(X, Y) = Cov(Y, X)$,

(ii) $Cov(X, X) = \mathrm{Var}(X)$,

(iii) $Cov(aX, Y) = a\, Cov(X, Y)$,

(iv) $Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(X_i, Y_j)$.

Dimostrazione. La (i) e la (ii) seguono immediatamente dalla definizione di covarianza. Per la proprietà di linearità del valore medio si ha la (iii):

$$Cov(aX, Y) = E[(aX - E(aX))(Y - E(Y))] = a\, Cov(X, Y).$$

Per dimostrare la (iv), ossia la proprietà di additività della covarianza, poniamo $\mu_i = E[X_i]$ e $v_j = E[Y_j]$. Allora

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mu_i, \qquad E\left[\sum_{j=1}^{m} Y_j\right] = \sum_{j=1}^{m} v_j.$$

Inoltre

$$
\begin{aligned}
Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) &= E\left[\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right)\left(\sum_{j=1}^{m} Y_j - \sum_{j=1}^{m} v_j\right)\right] \\
&= E\left[\sum_{i=1}^{n}(X_i - \mu_i)\sum_{j=1}^{m}(Y_j - v_j)\right] \\
&= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i - \mu_i)(Y_j - v_j)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} E[(X_i - \mu_i)(Y_j - v_j)],
\end{aligned}
$$

dove l'ultima uguaglianza segue in quanto il valore atteso di una somma di variabili aleatorie è uguale alla somma dei valori attesi.

Proposizione.

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} Cov(X_i, X_j).$$

Dimostrazione. Dalla precedente proposizione, posto $Y_j = X_j$, $j = 1, \ldots, n$, si ha

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i\neq j} Cov(X_i, X_j).$$

Dato che ogni coppia di indici (i, j), $i \neq j$, appare due volte nella doppia sommatoria, la proposizione segue immediatamente. (In altri termini, la *matrice di covarianza* $\{Cov(X_i, X_j)\}_{i,j}$ è simmetrica, con le varianze sulla diagonale principale).

Notiamo che dalla proposizione precedente si ha, per $a_1, \ldots, a_n$ costanti arbitrarie,

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i) + 2 \sum_{i<j} a_i a_j Cov(X_i, X_j).$$

Da ciò si trae che

$$\mathrm{Var}\,(X - Y) = \mathrm{Var}\,(X) + \mathrm{Var}\,(Y) - 2Cov(X, Y).$$

Se $X_1, \ldots, X_n$ sono a due a due scorrelate, cioè se $Cov(X_i, X_j) = 0$ per ogni $i \neq j$, risulta:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i).$$

Esempio. Siano $X_1, \ldots, X_n$ variabili indipendenti ed identicamente distribuite con valore atteso μ e varianza σ^2, e sia $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ la media campionaria. Le variabili $X_i - \overline{X}$ sono chiamate deviazioni, mentre $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$ è chiamata varianza campionaria. Determinare (a) $Var[\overline{X}]$ e (b) $E[S^2]$.

Soluzione. Per l'indipendenza delle variabili $X_1, \ldots, X_n$ si ha

$$\text{(a)} \qquad Var(\overline{X}) = \left(\frac{1}{n}\right)^2 Var\left(\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}.$$

(b) Consideriamo la seguente identità algebrica

$$
\begin{aligned}
(n-1)S^2 &= \sum_{i=1}^{n}(X_i - \mu + \mu - \overline{X})^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 + \sum_{i=1}^{n}(\overline{X} - \mu)^2 - 2(\overline{X} - \mu)\sum_{i=1}^{n}(X_i - \mu) \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 + n(\overline{X} - \mu)^2 - 2(\overline{X} - \mu)n(\overline{X} - \mu) \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\overline{X} - \mu)^2
\end{aligned}
$$

Prendendo i valori attesi di entrambi i membri dell'uguaglianza precedente si ottiene

$$
(n-1)E[S^2] = \sum_{i=1}^{n} E[(X_i - \mu)^2] - nE[(\overline{X} - \mu)^2] = n\sigma^2 - nVar(\overline{X}) = (n-1)\sigma^2,
$$

dove si è utilizzato il fatto che $E[\overline{X}] = \mu$ ed il punto (a) nell'uguaglianza finale. Dividendo per $n-1$ si ottiene che $E[S^2] = \sigma^2$.

Esempio. Calcolare la varianza di una variabile aleatoria binomiale X di parametri n e p.

Soluzione. Dato che una tale variabile rappresenta il numero di successi in n prove indipendenti quando il successo in una prova ha probabilità p, possiamo scrivere

$$X = X_1 + \cdots + X_n,$$

dove le X_i sono variabili di Bernoulli indipendenti tali che

$$X_i = \begin{cases} 1 & \text{se l'}i\text{-esima prova è un successo,} \\ 0 & \text{altrimenti.} \end{cases}$$

Si ottiene pertanto

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

Essendo

$$\mathrm{Var}(X_i) = E[X_i^2] - (E[X_i])^2 = p - p^2 = p\,(1-p),$$

si ha infine

$$\mathrm{Var}(X) = n\,p\,(1-p).$$

Esempio. Varianza del numero di accoppiamenti. Calcolare la varianza di X, il numero di persone che seleziona il proprio cappello tra N.

Soluzione Utilizzando la rappresentazione di X usata in un esempio precedente, si ha che $X = X_1 + X_2 + \cdots + X_N$, dove

$$X_i = \begin{cases} 1 & \text{se la persona } i \text{ prende il suo cappello,} \\ 0 & \text{altrimenti.} \end{cases}$$

Si ha poi

$$\text{Var}(X) = \sum_{i=1}^{N} \text{Var}(X_i) + 2 \sum_{i<j} Cov(X_i, X_j).$$

Essendo $P(X_i = 1) = 1/N$, si ha

$$E[X_i] = \frac{1}{N}, \qquad \text{Var}(X_i) = \frac{1}{N}\left(1 - \frac{1}{N}\right) = \frac{N-1}{N^2}.$$

Inoltre

$$X_i X_j = \begin{cases} 1 & \text{se l'}i\text{-esima e la } j\text{-esima persona selezionano il loro cappello,} \\ 0 & \text{altrimenti.} \end{cases}$$

Pertanto si ha

$$E[X_i X_j] = \sum_{u=0}^{1} \sum_{v=0}^{1} u\, v\, p(u,v) = p(1,1) = P(X_i = 1)\, P(X_j = 1 | X_i = 1) = \frac{1}{N}\,\frac{1}{N-1}$$

da cui segue

$$Cov(X_i, X_j) = \frac{1}{N(N-1)} - \left(\frac{1}{N}\right)^2 = \frac{N-(N-1)}{N^2(N-1)} = \frac{1}{N^2(N-1)}$$

e quindi

$$\mathrm{Var}(X) = \frac{N-1}{N} + 2\binom{N}{2}\frac{1}{N^2(N-1)} = \frac{N-1}{N} + \frac{1}{N} = 1$$

In conclusione sia media che varianza del numero di accoppiamenti sono pari a 1.

La correlazione tra due variabili aleatorie X e Y, indicata con $\rho(X, Y)$, è definita, se $Var(X)$ e $Var(Y)$ non sono nulle, dal coefficiente di correlazione

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

Proposizione. Risulta

$$-1 \leq \rho(X, Y) \leq 1.$$

Dimostrazione. Denotando con σ_X^2 e σ_Y^2 le varianze di X e Y, si ha

$$0 \leq \mathrm{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_y}\right) = \frac{\mathrm{Var}(X)}{\sigma_X^2} + \frac{\mathrm{Var}(Y)}{\sigma_Y^2} + \frac{2Cov(X, Y)}{\sigma_X\,\sigma_Y} = 2\,[1 + \rho(X, Y)]$$

da cui segue $\rho(X, Y) \geq -1$. D'altronde risulta

$$0 \leq \mathrm{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = \frac{\mathrm{Var}(X)}{\sigma_X^2} + \frac{\mathrm{Var}(Y)}{\sigma_Y^2} - \frac{2Cov(X, Y)}{\sigma_X\,\sigma_Y} = 2\,[1 - \rho(X, Y)]$$

e quindi si ha anche $\rho(X, Y) \leq 1$.

Dato che se $\text{Var}(Z) = 0$, allora Z è costante con probabilità 1, dalla precedente dimostrazione si ha che se $\rho(X,Y) = 1$ allora $Y = aX + b$, dove $a = \sigma_y/\sigma_x > 0$. Analogamente, se $\rho(X,Y) = -1$ allora $Y = aX + b$, dove $a = -\sigma_y/\sigma_x < 0$. Vale anche il viceversa, cioè se $Y = aX + b$, allora $\rho(X,Y)$ è uguale a $+1$ oppure -1 a seconda del segno di a. Infatti, poiché $Cov(X, aX+b) = a\,\text{Var}(X)$, si ha

$$\rho(X,Y) = \rho(X, aX+b) = \frac{a\,\text{Var}(X)}{\sqrt{\text{Var}(X)\,a^2\,\text{Var}(X)}} = \frac{a}{|a|} = \begin{cases} 1, & a > 0 \\ -1, & a < 0. \end{cases}$$

ll coefficiente di correlazione è una misura del grado di linearità tra X e Y. Un valore di $\rho(X,Y)$ vicino a $+1$ o a -1 indica un alto livello di linearità tra X e Y, mentre un valore vicino a 0 indica un'assenza di tale linearità. Un valore positivo di $\rho(X,Y)$ indica che Y tende a crescere quando X cresce, mentre un valore negativo indica che Y tende a decrescere quando X cresce. Se $\rho(X,Y) = 0$ allora X e Y sono scorrelate.

Dalla definizione segue che il coefficiente di correlazione $\rho(X,Y)$ è adimensionale.

Esercizio Nel lancio di un dado ripetuto n volte, sia X il numero di volte che esce 6 e Y il numero di volte che esce 5. Calcolare il coefficiente di correlazione di (X, Y).

Soluzione. Per $i, j = 1, 2, \ldots, n$ poniamo

$$X_i = \begin{cases} 1 & \text{se esce 6 al lancio } i\text{-esimo} \\ 0 & \text{altrimenti,} \end{cases} \qquad Y_j = \begin{cases} 1 & \text{se esce 5 al lancio } j\text{-esimo} \\ 0 & \text{altrimenti,} \end{cases}$$

da cui si segue che $X = \sum_{i=1}^{n} X_i$ e $Y = \sum_{j=1}^{n} Y_j$. Risulta

$$E(X_i) = E(Y_j) = \frac{1}{6}, \qquad E(X_i Y_j) = \begin{cases} 0 & \text{se } i = j, \\ \frac{1}{36} & \text{altrimenti,} \end{cases} \qquad \mathrm{Var}(X_i) = \mathrm{Var}(Y_j) = \frac{5}{36},$$

e pertanto

$$\mathrm{Cov}(X_i, Y_j) = E(X_i Y_j) - E(X_i)E(Y_j) = \begin{cases} -\frac{1}{36} & \text{se } i = j, \\ 0 & \text{altrimenti,} \end{cases}$$

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, Y_j) = -\frac{n}{36},$$

e, poiché X e Y sono variabili binomiali,

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = n\,\frac{5}{36}.$$

Infine segue:

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{-n/36}{n\,5/36} = -\frac{1}{5}.$$

Esercizio Siano X e Y tali che $p(0,0) = p(1,1) = p$ e $p(1,0) = p(0,1) = \frac{1}{2} - p$.

(i) Determinare i valori ammissibili di p.

(ii) Stabilire per quali valori di p si ha che X e Y sono indipendenti.

(iii) Calcolare $\rho(X,Y)$ e studiarlo al variare di p.

Soluzione. Dalle condizioni $p(x,y) \geq 0$ $\forall x, y$ segue $p \geq 0$ e $\frac{1}{2} - p \geq 0$, mentre $\sum_x \sum_y p(x,y) = 1$ per ogni $p \in \mathbb{R}$; quindi i valori ammissibili di p sono

$$0 \leq p \leq \frac{1}{2}.$$

(ii) Dalla seguente tabella segue che $p(x,y) = p_X(x)\, p_Y(y)$ per ogni x e y se e solo se $p = 1/4$.

$x \backslash y$	0	1	$p_X(x)$
0	p	$1/2 - p$	$1/2$
1	$1/2 - p$	p	$1/2$
$p_Y(y)$	$1/2$	$1/2$	1

Poiché X e Y sono entrambe variabili aleatorie di Bernoulli di parametro $1/2$, si ha

$$E[X] = E[Y] = \frac{1}{2}, \qquad \mathrm{Var}(X) = \mathrm{Var}(Y) = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4},$$

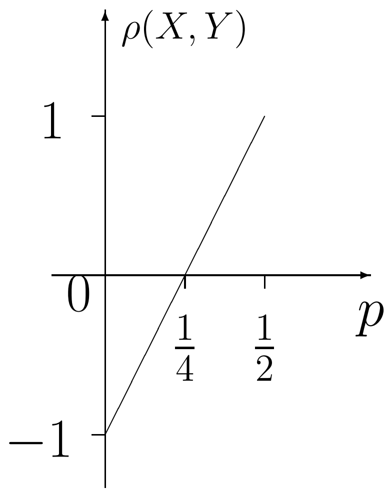
ed inoltre $E[X\,Y] = \sum_{x=0}^{1} \sum_{y=0}^{1} x\, y\, p(x,y) = p(1,1) = p$, da cui segue

$$Cov(X,Y) = E[XY] - E[X]\,E[Y] = p - \frac{1}{4}.$$

Il coefficiente di correlazione è quindi dato da

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{p - \frac{1}{4}}{\frac{1}{4}} = 4\,p - 1, \qquad 0 \le p \le \frac{1}{2}.$$

Grafico di $\rho(X,Y) = 4\,p - 1$, $0 \le p \le \frac{1}{2}$;
risulta:
$\rho(X,Y) = -1$ per $p = 0$,
$\rho(X,Y) = 0$ per $p = 1/4$,
$\rho(X,Y) = 1$ per $p = 1/2$.



Per $p = 0$ si ha:
$P(Y = 1 - X) = p(0,1) + p(1,0) = 1$
ed infatti in tal caso $\rho(X,Y) = -1$.

$x \backslash y$	0	1	$p_X(x)$
0	0	1/2	1/2
1	1/2	0	1/2
$p_Y(y)$	1/2	1/2	1

Per $p = 1/2$ si ha:
$P(Y = X) = p(0,0) + p(1,1) = 1$
ed infatti in tal caso $\rho(X,Y) = 1$.

$x \backslash y$	0	1	$p_X(x)$
0	1/2	0	1/2
1	0	1/2	1/2
$p_Y(y)$	1/2	1/2	1

Calcolo delle Probabilità e Statistica Matematica – A.A. 2009/10

CAPITOLO 8 – Teoremi Limite

8.1 Introduzione

I risultati più significativi del calcolo delle probabilità sono certamente i teoremi limite.

Di questi, i più importanti sono quelli che vengono classificati come *leggi dei grandi numeri* e come *teoremi del limite centrale*.

In generale, un teorema viene considerato una legge dei grandi numeri quando stabilisce delle condizioni sotto le quali la media di una successione di variabili aleatorie converge (in qualche senso) alla media attesa.

D'altra parte, i teoremi del limite centrale sono caratterizzati dal determinare condizioni sotto le quali la somma di un grande numero di variabili aleatorie (opportunamente standardizzata) ha la distribuzione di probabilità che tende a essere normale.

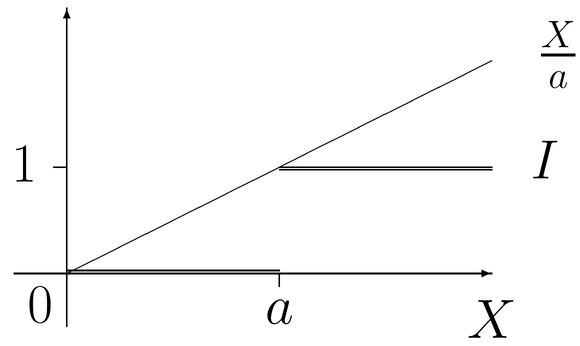
8.2 La disuguaglianza di Chebyshev e la legge debole dei grandi numeri

Proposizione. Disuguaglianza di Markov. Se X è una variabile aleatoria che assume solo valori non negativi, allora per ogni numero reale $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Dimostrazione. Per $a > 0$, sia

$$I = \mathbf{1}_{\{X \geq a\}} = \begin{cases} 1 & \text{se } X \geq a \\ 0 & \text{altrimenti.} \end{cases}$$



Ricordando che $X \geq 0$, si ha: $I \leq \dfrac{X}{a}$.

Dalle proprietà del valore atteso segue: $E[I] \leq \dfrac{E[X]}{a}$.

Ricordando che per la variabile di Bernoulli I risulta $E[I] = P(I = 1) = P(X \geq a)$, si ha

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Proposizione. Disuguaglianza di Chebyshev. Se X è una variabile aleatoria di media μ e varianza σ^2, entrambe finite, allora per ogni numero reale $k > 0$,

$$P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

Dimostrazione. Essendo $(X-\mu)^2$ una variabile aleatoria non negativa, possiamo applicare la disuguaglianza di Markov, con $a = k^2$:

$$P(X \geq a) \leq \frac{E(X)}{a} \qquad \text{diventa} \qquad P\{(X-\mu)^2 \geq k^2\} \leq \frac{E[(X-\mu)^2]}{k^2}.$$

Essendo $(X-\mu)^2 \geq k^2$ se e solo se $|X-\mu| \geq k$, si ricava infine:

$$P\{|X-\mu| \geq k\} \leq \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}.$$

Analogamente si può ottenere una limitazione inferiore per $P\{\mu - k < X < \mu + k\}$:

$$P\{|X-\mu| < k\} \geq 1 - \frac{\sigma^2}{k^2}.$$

Esempio. Si supponga che il numero di studenti che si iscrive ad un corso di lingua inglese è descritto da una variabile aletoria X di media $\mu = 50$.
(a) Cosa si può dire su $P(X \geq 75)$?
(b) Cosa si può dire su $P(40 < X < 60)$ se $\sigma^2 = 25$?
Soluzione. (a) Facendo uso della disuguaglianza di Markov, con $a = 75$, si ha

$$P(X \geq 75) \leq \frac{E(X)}{75} = \frac{50}{75} = \frac{2}{3}.$$

(b) Ricordiamo che dalla disuguaglianza di Chebyshev segue

$$P\{|X - \mu| < k\} \geq 1 - \frac{\sigma^2}{k^2}.$$

Quindi, essendo $\mu = 50$ e $\sigma^2 = 25$, per $k = 10$ si ottiene

$$P(40 < X < 60) = P\{|X - 50| < 10\} = P\{|X - \mu| < k\} \geq 1 - \frac{\sigma^2}{k^2} = 1 - \frac{25}{100} = \frac{3}{4}.$$

Esempio. Sia X uniformemente distribuita su $(0,10)$. Allora, poiché $E[X]=5$ e $\mathrm{Var}(X)=25/3$, dalla disuguaglianza di Chebyshev segue

$$P\{|X-5|>4\}\leq \frac{25}{3\cdot 16}=\frac{25}{48}\approx 0{,}52$$

mentre il valore esatto è

$$P\{|X-5|>4\}=P(X<1)+P(X>9)=\frac{1}{10}+\frac{1}{10}=0{,}2.$$

Esempio. Sia X una variabile normale di media μ e varianza σ^2. Risulta

$$P\{\mu-h\,\sigma<X<\mu+h\,\sigma\}=P\Big(-h<\frac{X-\mu}{\sigma}<h\Big)=2\,\Phi(h)-1,$$

mentre dalla disuguaglianza di Chebyshev, per $k=h\,\sigma$, si ha

$$P\{\mu-h\,\sigma<X<\mu+h\,\sigma\}=P\{|X-\mu|<h\,\sigma\}\geq 1-\frac{\sigma^2}{(h\,\sigma)^2}=1-\frac{1}{h^2}.$$

	$h=1$	$h=1{,}5$	$h=2$	$h=2{,}5$	$h=3$
$2\,\Phi(h)-1$	0,6826	0,8664	0,9544	0,9876	0,9974
$1-1/h^2$	0	0,5555	0,75	0,84	0,8888

Proposizione. Se X è una variabile aleatoria tale che $\mathrm{Var}(X) = 0$, allora

$$P\{X = E[X]\} = 1,$$

ed in tal caso X si dice *degenere*.

Teorema. La legge debole dei grandi numeri. Sia $X_1, X_2, \ldots$ una successione di variabili aleatorie i.i.d., ognuna con media finita $E[X_i] = \mu$. Allora, posto $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, per ogni $\epsilon > 0$ si ha

$$\lim_{n\to\infty} P\{|\overline{X}_n - \mu| \geq \epsilon\} = 0.$$

Dimostrazione. Assumiamo inoltre che $\mathrm{Var}(X_i) = \sigma^2$ sia finita. Poiché $E[\overline{X}_n] = \mu$ e $\mathrm{Var}[\overline{X}_n] = \sigma^2/n$, dalla disuguaglianza di Chebyshev si ha, per ogni $\epsilon > 0$

$$P\{|\overline{X}_n - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{n}\frac{1}{\epsilon^2}.$$

Procedendo al limite per $n \to \infty$ segue immediatamente l'asserto.

8.3 Il teorema del limite centrale

Vedremo ora uno dei più importanti risultati della teoria della probabilità. Esso, in breve, asserisce che la somma di un grande numero di variabili aleatorie indipendenti ha distribuzione approssimativamente normale. La dimostrazione è omessa per brevità.

Teorema. Il teorema del limite centrale. Sia $X_1, X_2, \ldots$ una successione di variabili aleatorie i.i.d., ognuna di media μ e varianza σ^2 finite. Allora, posto

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}},$$

per ogni $a \in \mathbb{R}$ si ha: $\lim_{n\to\infty} P\{Z_n \le a\} = \Phi(a) \qquad \left(= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx\right).$

In altri termini, la somma standardizzata $Z_n = \frac{X_1+\cdots+X_n-E[X_1+\cdots+X_n]}{\sqrt{\mathrm{Var}(X_1+\cdots+X_n)}}$ tende ad avere distribuzione normale standard quando $n \to \infty$. Vale quindi l'approssimazione

$$P\left(\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le a\right) \approx \Phi(a) \qquad (n \text{ grande}).$$

Esempio. Si supponga che il numero di studenti che si iscrive ad un corso universitario è descritto da una variabile alestoria di Poisson X di media $\mu = 100$. Qual è la probabilità che si iscrivano almeno 120 studenti?

Soluzione. Non è agevole determinare il valore esatto, essendo

$$P(X \geq 120) = e^{-100} \sum_{k=120}^{\infty} \frac{(100)^k}{k!} = 1 - e^{-100} \sum_{k=0}^{119} \frac{(100)^k}{k!}.$$

Ricordando che X può essere riguardata come somma di $n = 100$ variabili aleatorie di Poisson indipendenti di media 1 (e quindi di varianza 1), possiamo adoperare l'approssimazione basata sul teorema del limite centrale:

$$\begin{aligned} P(X \geq 120) &= P(X \geq 119{,}5) \qquad \text{(la correzione di continuità)} \\ &= P\left(\frac{X - 100}{\sqrt{100}} \geq \frac{119{,}5 - 100}{\sqrt{100}} \right) \\ &\approx 1 - \Phi(1{,}95) \\ &= 1 - 0{,}9744 = 0{,}0256. \end{aligned}$$

Esempio. Determinare la probabilità approssimata che lanciando a caso 10 dadi la somma X dei valori ottenuti sia compresa tra 30 e 40, estremi inclusi.

Soluzione. Indicando l'esito del lancio i-esimo con X_i, questa è una variabile uniforme discreta con $E[X_i] = 7/2$ e $\mathrm{Var}(X_i) = 35/12$. Pertanto risulta

$$E[X] = 35, \qquad \mathrm{Var}(X) = \frac{350}{12} = 29{,}1666.$$

L'approssimazione basata sul teorema del limite centrale dà quindi:

$$\begin{aligned}
P(30 \leq X \leq 40) &= P(29{,}5 \leq X \leq 40{,}5) \qquad \text{(la correzione di continuità)} \\
&= P\left(\frac{29{,}5 - 35}{\sqrt{29{,}1666}} \leq \frac{X - 35}{\sqrt{29{,}1666}} \leq \frac{40{,}5 - 35}{\sqrt{29{,}1666}}\right) \\
&\approx \Phi(1{,}02) - \Phi(-1{,}02) \\
&= 2\,\Phi(1{,}02) - 1 \\
&= 2 \cdot 0{,}8461 - 1 = 0{,}6922.
\end{aligned}$$

Esempio. Siano X_i, $i = 1, \ldots, 24$, variabili aleatorie indipendenti, con distribuzione uniforme su $(0, 1)$. Determinare in modo approssimato i valori di

$$P\left(\sum_{i=1}^{24} X_i > 14\right), \qquad P\left(8 < \sum_{i=1}^{24} X_i < 16\right).$$

Soluzione. Poiché $E[X_i] = 1/2$ e $\mathrm{Var}(X_i) = 1/12$, posto $X = \sum_{i=1}^{24} X_i$ si ha $E[X] = 12$ e $\mathrm{Var}(X) = 2$. Quindi, dal teorema del limite centrale si trae:

$$P(X > 14) = P\left(\frac{X-12}{\sqrt{2}} > \frac{14-12}{\sqrt{2}}\right) \approx 1 - \Phi(1{,}41) = 1 - 0{,}9207 = 0{,}0793$$

ed inoltre

$$\begin{aligned} P(8 < X < 16) &= P\left(\frac{8-12}{\sqrt{2}} \le \frac{X-12}{\sqrt{2}} \le \frac{16-12}{\sqrt{2}}\right) \\ &\approx \Phi(2{,}83) - \Phi(-2{,}83) \\ &= 2\,\Phi(2{,}83) - 1 \\ &= 2 \cdot 0{,}9977 - 1 = 0{,}9954. \end{aligned}$$