PERCHE' QUESTA BOZZA, E COME USARLA

(Prefazione alla vers. 0.3)

Questa versione 0.3 esce "di corsa" per correggere un errore, abbastanza grave, che era presente nell'enunciato del "teorema centrale del limite".

Continuano a mancare alcune sezioni del primo capitolo, ma ho cominciato a dare le definizioni di "insieme dei dati" (grezzi e ordinati) e di media, moda e mediana. Cercherò, nelle prossime due settimane, di continuare a riempire i buchi più evidenti nella trattazione.

Ci riuscirò? Avete solo un modo per scoprirlo, tornare sulla pagina

http://marcobar.outducks.org/SNat/Libri.html

e controllare se c'è una versione successiva. Nel frattempo, buon lavoro con questa bozza!

Firenze, 25.01.2010

Marco Barlotti

1.- STATISTICA DESCRITTIVA

1.1 - Introduzione.

Alla statistica è dedicata una celebre poesia di Carlo Alberto Salustri, detto Trilussa (Roma, 26.10.1871, Roma 21.12.1950). Poeta dialettale romanesco, nominato senatore a vita dal presidente Luigi Einaudi il giorno 1.12.1950, aveva scritto:

La Statistica

Sai ched'è la statistica? È 'na cosa
che serve pe fà un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che sposa.

Ma pè me la statistica curiosa
è dove c'entra la percentuale,
pè via che, lì, la media è sempre eguale
puro co' la persona bisognosa.

Me spiego: da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:

e, se nun entra nelle spese tue,
t'entra ne la statistica lo stesso
perch'è c'è un antro che ne magna due.

Come osserva Trilussa, e come si può rilevare dalla radice *stat* nella parola stessa, la *statistica* nasce storicamente per rilevare in modo sistematico e organizzato i dati relativi ai cittadini di uno stato. Oggi le tecniche sviluppate a tale scopo sono utilizzate nei campi più disparati.

1.2 - Dati, e loro organizzazione.

In tutto questo capitolo, considereremo un insieme finito $\mathcal{P}$ (detto *popolazione*) e una o più funzioni $\mathcal{P}\to\mathbb{R}$ dette *variabili statistiche su* $\mathcal{P}$. L'ipotesi di finitezza su $\mathcal{P}$ è coerente con la pratica; vedremo più avanti quando e come eliminare tale ipotesi.

Poiché $\mathcal{P}$ è un insieme finito, possiamo indicare i suoi elementi con $\omega_1, \omega_2, ..., \omega_n$; implicitamente assegniamo quindi agli elementi di $\mathcal{P}$ un ordine (che potrebbe essere l'ordine in cui vengono materialmente rilevati i valori delle variabili statistiche che studiamo). Se X è una variabile statistica su $\mathcal{P} = \{\omega_1, \omega_2, ..., \omega_n\}$, posto

$$x_i := \mathrm{X}(\omega_i)$$

la $n-$pla ordinata $(x_1, x_2, ..., x_n)$ si dice *insieme dei dati* (*grezzi*) (*ottenuti studiando* X *su* $\mathcal{P}$) (anche se, attenzione!, non è un insieme ma una $n-$pla ordinata); la $n-$pla ordinata $(\overline{x}_1, \overline{x}_2, ..., \overline{x}_n)$ ottenuta riordinando l'insieme dei dati in modo che sia $\overline{x}_1 \leq \overline{x}_2 \leq ... \leq \overline{x}_n$ si dice *insieme dei dati ordinati* (*ottenuti studiando* X *su* $\mathcal{P}$) (e anche in questo caso non si tratta in realtà di un insieme ma di una $n-$pla ordinata).

Se x_i è un dato, si dice *frequenza assoluta* di x_i il numero naturale che esprime la cardinalità di

$$\{j \in \{1, 2, ..., n\} \;/\; x_j = x_i\}$$

cioè il numero delle volte che x_i compare fra i dati; si dice *frequenza relativa* di x_i il rapporto fra la sua frequenza assoluta e il numero complessivo n dei dati.

Osservazione 1.2.1

I dati ottenuti sperimentalmente studiando una variabile statistica sono di fatto sempre numeri razionali. Anche quando si tratta, ad esempio, di misure di lunghezze (che quindi potrebbero in teoria assumere valori non razionali) gli strumenti che utilizziamo per le rilevazioni ci forniscono sempre approssimazioni razionali.

Osservazione 1.2.2

Potrà far comodo classificare le variabili statistiche in due tipi, come segue:

- una variabile si dice *densa in sé*, quando comunque presi due dati x_i, x_j è teoricamente possibile che, eventualmente estendendo la popolazione $\mathcal{P}$, esista un dato x_k strettamente compreso tra x_i e x_j;

- una variabile si dice *discreta* altrimenti.

1.3 - I dati raggruppati in classi.

Dall'inseme (x_1, x_2, ..., x_n) dei dati grezzi si ottengono, procedendo con opportuni criteri, certe $k_i -$ ple ordinate dette *classi* che risultano utili ai fini espositivi. L'insieme delle classi si dice *insieme dei dati raggruppati* (oppure *insieme dei dati organizzati*).

I criteri di costruzione delle classi dipendono dalla natura dei dati e da considerazioni di opportunità di vario tipo; generalmente è rilevante anche se la variabile studiata sia densa in sé oppure discreta. Devono comunque essere rispettate queste due regole:

1. – Ogni dato deve appartenere a una e una sola classe, e vi deve comparire tante volte quanta è la sua frequenza assoluta;

2. – Se a una classe appartengono i dati x_i e x_j, alla stessa classe deve appartenere ogni dato x_k compreso tra x_i e x_j.

Una prima possibilità è quella di avere tante classi quanti sono i valori distinti della variabile statistica che stiamo studiando, cioè tante classi quanti sono gli elementi dell'insieme X($\mathcal{P}$). Questa soluzione è senz'altro da escludersi nel caso di una variabile statistica densa in sé; ed è comunque proponibile, nel caso di una variabile statistica discreta, soltanto quando il numero delle classi che si ottengono è ragionevolmente "piccolo".

Per costruire le classi si procede di solito così:

1.- Si stabilisce il numero n_c delle classi. Se non abbiamo le idee chiare, un ragionevole criterio è il seguente: per n dati si stabiliscono $\sqrt{n}$ classi.

2.- Nel caso (molto frequente) che i dati grezzi siano *numeri*, si considerano il minimo m e il massimo M dell'insieme dei dati grezzi ([1])([2]). Scelto un "margine" ragionevole ε, si suddivide in n_c parti uguali l'intervallo $[m-\varepsilon,\ M+\varepsilon]$: posto

$$\delta := \frac{M-m+2\varepsilon}{n_c}$$

la prima classe sarà l'insieme dei dati che ricadono nell'intervallo $[m-\varepsilon,\ m-\varepsilon+\delta)$, la seconda classe sarà l'insieme dei dati che ricadono nell'intervallo $[m-\varepsilon+\delta, m-\varepsilon+2\delta)$, la terza classe sarà l'insieme dei dati che ricadono nell'intervallo $[m-\varepsilon+2\delta,\ m-\varepsilon+3\delta)$, e così via. È comunque possibile (e talvolta conveniente) scegliere in altro modo gli intervalli $[a_i,\ b_i)$ che definiscono le classi; in ogni caso, la differenza $b_i - a_i$ si dice *ampiezza* della $i -$ sima classe.

[1] ricordiamo che si tratta sempre di un insieme finito, quindi ha certamente massimo e minimo.

[2] la differenza $M-m$ si dice *ampiezza di variazione* (o *campo di variazione*, o *range*) dell'insieme dei dati raccolti.

3.- Distribuiti i dati nelle varie classi, si può visualizzare quanto fatto con una tabella, le cui righe forniscono le informazioni per ciascuna classe e le cui colonne invece sono così denominate: <nome della classe> (cioè l'intervallo che definisce la classe), <frequenza assoluta della classe> (cioè quanti dati appartengono a quella classe), <frequenza relativa della classe> (cioè il rapporto fra la frequenza assoluta della classe e il numero complessivo n dei dati raccolti), <frequenza percentuale della classe> (cioè l'espressione della frequenza assoluta f come percentuale del numero complessivo n dei dati raccolti, quindi semplicemente $\frac{100f}{n}$).

La colonna che esprime la frequenza relativa di ciascuna classe viene usualmente visualizzata con un istogramma, cioè con una sequenza di rettangoli nei quali la base è proporzionale all'ampiezza di ciascuna classe e l'area è proporzionale alla frequenza relativa di ciascuna classe ([3]).

La colonna che esprime la frequenza percentuale di ciascuna classe viene invece preferibilmente visualizzata con un diagramma a torta (eventualmente tridimensionale).

1.4 - Un esempio.

1.5 - Media, moda, mediana.

Sia **D** $:= (x_1, x_2, ..., x_n)$ un insieme di dati.

Si dice *media* (*aritmetica*) o *valore medio* di **D** il numero reale

$$\mathrm{M}(\mathbf{D}) := \frac{\sum_{i=1}^{n} x_i}{n} .$$

Si dice *moda* di **D** un dato che abbia la massima frequenza assoluta. Ovviamente non è detto che la moda sia unica: quando ciò avviene, l'insieme di dati **D** si dice *unimodale*; se vi sono due, tre, ..., k mode, l'insieme di dati **D** si dice (rispettivamente) *bimodale*, *trimodale*, ..., $k-$ *modale*.

Se il numero n dei dati è dispari, $n = 2k + 1$, si dice *mediana* di **D** il dato x_{k+1} ; se il numero n dei dati è pari, $n = 2k$, si dice *mediana* di **D** in numero reale $\frac{x_k + x_{k+1}}{2}$ (cioè la media fra x_k e x_{k+1}): in questo caso, la mediana non è uno dei dati (tranne il caso che sia $x_k = x_{k+1}$).

---

[3] Naturalmente, se le classi hanno tutte la stessa ampiezza allora i rettangoli che compongono l'istogramma hanno basi congruenti e quindi di fatto anche le altezze sono proporzionali alle frequenze relative!

1.6 - Altre medie.

1.7 - Percentili.

1.8 - Varianza e deviazione standard.

1.9 - Un esempio un po' elaborato.

Studiando una popolazione di rettili adulti (Iguana iguana) si sono rilevate le variabili

x = lunghezza (espressa in centimetri)

e

y = peso (espresso in grammi)

Ecco la tabella dei dati grezzi:

	lunghezza	peso
1	176	4.250
2	164	3.950
3	155	3.750
4	150	3.650
5	180	4.350
6	162	3.900
7	174	4.200
8	158	3.800
9	160	3.900
10	166	4.050

ed ecco la tabella dei dati ordinati (per la variabile "lunghezza"):

	lunghezza	peso
4	150	3.650
3	161	3.900
8	164	3.950
9	166	4.050
6	168	4.000
2	170	4.150
10	172	4.150
7	174	4.200
1	176	4.250
5	180	4.350

La mediana per la lunghezza è 169, per il peso è 4.100.

Poiché i possibili dati sono densi in sé (e presumibilmente più di dieci...) ha senso fare una tabella di classi, sia per la lunghezza che per il peso:

classe di lunghezza	frequenza assoluta	frequenza relativa	frequenza percentuale
[145, 153)	1	0,1	10%
[153, 161)	0	0	0%
[161, 169)	4	0,4	40%
[169, 177)	4	0,4	40%
[177, 185)	1	0,1	10%

classe di peso	frequenza assoluta	frequenza relativa	frequenza percentuale
[3.600, 3.760)	1	0,1	10%
[3.760, 3.920)	1	0,1	10%
[3.920, 4.080)	3	0,3	30%
[4.080, 4.240)	3	0,3	30%
[4.240, 4.400)	2	0,2	20%

con i relativi istogrammi.

Poiché i dati sono 10, è facile calcolare la media

	lunghezza	peso
4	150	3.650
3	161	3.900
8	164	3.950
9	166	4.050
6	168	4.000
2	170	4.150
10	172	4.150
7	174	4.200
1	176	4.250
5	180	4.350
totale	1.681	40.650

La lunghezza media è 168,1 e il peso medio è 4.065 grammi.

Calcoliamo ora varianza, deviazione standard, covarianza e coefficiente di correlazione.

	x_i	y_i	$(x_i - \overline{x})^2$	$(y_i - \overline{y})^2$	$(x_i - \overline{x})(y_i - \overline{y})$
4	150	3.650	$327,61$	172.225	$7.511,5$
3	161	3.900	$50,41$	27.225	$1.171,5$
8	164	3.950	$16,81$	13.225	$471,5$
9	166	4.050	$4,41$	225	$31,5$
6	168	4.000	$0,01$	4.225	$6,5$
2	170	4.150	$3,61$	7.225	$161,5$
10	172	4.150	$15,21$	7.225	$331,5$
7	174	4.200	$34,81$	18.225	$796,5$
1	176	4.250	$62,41$	34.225	$1.461,5$
5	180	4.350	$141,61$	81.225	$3.391,5$
totale	1.681	40.650	$656,9$	365.250	15.335

Varianza della lunghezza: $\frac{\sum_{i=1}^{10}(x_i-\overline{x})^2}{9} = \frac{656{,}9}{9} \approx 73$

Deviazione standard della lunghezza: $\sqrt{73} \approx 8,544$

Varianza del peso: $\frac{\sum_{i=1}^{10}(y_i-\overline{y})^2}{9} = \frac{365.250}{9} \approx 40.583$

Deviazione standard del peso: $\sqrt{40.583} \approx 201,45$

Covarianza: $\frac{\sum_{i=1}^{10}(x_i-\overline{x})(y_i-\overline{y})}{9} = \frac{15.335}{9} \approx 1.704$

Coefficiente di correlazione tra gli x_i e gli y_i:

$$\frac{1704}{\sqrt{73 \cdot 40.583}} = \frac{1704}{\sqrt{2.962.559}} \approx \frac{1704}{1721{,}2} \approx 0,99$$

La “retta di regressione lineare” è della forma

$$y = px + q$$

e (si dimostra che) i coefficienti p, q si trovano risolvendo questo sistema lineare:

$$\left(\sum_{i=1}^{n} x_i^2\right) p + \left(\sum_{i=1}^{n} x_i\right) q = \left(\sum_{i=1}^{n} x_i y_i\right)$$

$$\left(\sum_{i=1}^{n} x_i\right) p + nq = \sum_{i=1}^{n} y_i$$

ossia nel nostro caso

	x_i	y_i	x_i^2	$x_i y_i$
4	150	3.650	22.500	547.500
3	161	3.900	25.921	627.900
8	164	3.950	26.896	647.800
9	166	4.050	27.556	672.300
6	168	4.000	28.224	672.000
2	170	4.150	28.900	705.500
10	172	4.150	29.584	713.800
7	174	4.200	30.276	730.800
1	176	4.250	30.976	748.000
5	180	4.350	32.400	783.000
totale	1.681	40.650	283.233	6.848.600

$$\left(\sum_{i=1}^{n} x_i^2\right) = 283.233$$

$$\left(\sum_{i=1}^{n} x_i\right) = 1.681$$

$$\left(\sum_{i=1}^{n} x_i y_i\right) = 6.848.600$$

$$\left(\sum_{i=1}^{n} y_i\right) = 40.650$$

$$283.233\,p + 1.681q = 6.848.600$$

$$1.681\,p + 10q = 40.650$$

Consideriamo la matrice completa del sistema e riduciamola:

$$\begin{pmatrix} 283.233 & 1.681 & 6.848.600 \\ 1.681 & 10 & 40.650 \end{pmatrix}$$

Dalla seconda riga sottraiamo la prima moltiplicata per $-\frac{1.681}{283.233}$ ottenendo

$$\begin{pmatrix} 283.233 & 1.681 & 6.848.600 \\ 0 & 10 - 9,9768 & 40.650 - 40.646,734 \end{pmatrix}$$

cioè

$$\begin{pmatrix} 283.233 & 1.681 & 19.246.350 \\ 0 & 0{,}0232 & 3,266 \end{pmatrix}$$

da cui

$q \approx 140$

$p = \frac{6.848.600 - 1.681q}{283.233} = \frac{6.848.600 - 235.340}{283.233} \approx 23,35$

$y = 23,35x + 140$

1.10 - Un altro esempio.

Studiando una popolazione pesci, si sono rilevate le lunghezze e una proteina del sangue, la folchignina.

x = lunghezza (espressa in centimetri)

e

y = folchignina (espressa in U. I:, *unità immaginifiche*)

Ecco la tabella dei dati ordinati (per la variabile "lunghezza"):

	lunghezza	folchignina
1	12	0, 88
2	20	1, 3
3	24	1, 43
4	36	2, 5
5	48	1, 45
6	52	1, 31
7	60	0, 9

La mediana per la lunghezza è 36, per la folchignina è 1,31.

Calcoliamo la media

lunghezza	folchignina
12	0,88
20	1,3
24	1,43
36	2,5
48	1,45
52	1,31
60	0,9
192	9,77

La lunghezza media è circa 27,5 e la folchignina media è circa 1,4.

Calcoliamo ora varianza, deviazione standard, covarianza e coefficiente di correlazione.

	x_i	y_i	$(x_i - \overline{x})^2$	$(y_i - \overline{y})^2$	$(x_i - \overline{x})(y_i - \overline{y})$
	12	$0,88$	$240,25$	$0,27$	$8,06$
	20	$1,3$	$56,25$	$0,01$	$0,75$
	24	$1,43$	$42,875$	$0,0009$	$-0,105$
	36	$2,5$	$72,25$	$1,21$	$9,35$
	48	$1,45$	$420,25$	$0,0025$	$1,025$
	52	$1,31$	$600,25$	$0,0081$	$-2,205$
	60	$0,9$	$1.056,25$	$0,25$	$-16,25$
totale	192	$9,77$	$2.488,375$	$1,7515$	$0,625$

Varianza della lunghezza: $\frac{\sum_{i=1}^{10}(x_i-\overline{x})^2}{6} = \frac{2.488,375}{6} \approx 414,73$

Deviazione standard della lunghezza: $\sqrt{414,73} \approx 20,36$

Varianza della folchignina: $\frac{\sum_{i=1}^{10}(y_i-\overline{y})^2}{6} = \frac{1,7515}{6} \approx 0,3$

Deviazione standard della folchignina: $\sqrt{0,3} \approx 0,55$

Covarianza: $\frac{\sum_{i=1}^{10}(x_i-\overline{x})(y_i-\overline{y})}{6} = \frac{0{,}625}{6} \approx 0,104$

Coefficiente di correlazione tra gli x_i e gli y_i:

$$\frac{0{,}104}{\sqrt{414{,}73*0{,}3}} = \frac{0{,}104}{\sqrt{124{,}42}} = \frac{?}{\sqrt{?}} \approx \frac{0{,}104}{11{,}15} \approx 0,0093$$

Il coefficiente di correlazione molto vicino a zero ci dice che non c'è una correlazione lineare fra gli x_i e gli y_i. Segnando sul piano (riferito a un sistema cartesiano ortogonale monometrico positivamente orientato) i punti di coordinate (x_i, y_i) si vede (fatelo per esercizio!) qualcosa che assomiglia a una parabola. Si può allora cercare la "parabola dei minimi quadrati" che passa per quei punti (x_i, y_i).

Si dimostra che la "parabola di regressione quadratica" è della forma

$$y = ax^2 + bx + c$$

e i coefficienti a, b, c si trovano risolvendo questo sistema lineare:

$$\begin{aligned}
&\left(\sum_{i=1}^{n} x_i^4\right) a + \left(\sum_{i=1}^{n} x_i^3\right) b + \left(\sum_{i=1}^{n} x_i^2\right) c = \left(\sum_{i=1}^{n} x_i^2 y_i\right)\\
&\left(\sum_{i=1}^{n} x_i^3\right) a + \left(\sum_{i=1}^{n} x_i^2\right) b + \left(\sum_{i=1}^{n} x_i\right) c = \left(\sum_{i=1}^{n} x_i y_i\right)\\
&\left(\sum_{i=1}^{n} x_i^2\right) a + \left(\sum_{i=1}^{n} x_i\right) b + nc = \sum_{i=1}^{n} y_i
\end{aligned}$$

Trovate, per esercizio, questa parabola!

2.- ELEMENTI DI CALCOLO DELLE PROBABILITA'

2.1 - Eventi, esperimenti, risultati.

La logica ci ha insegnato ad attribuire alle *proposizioni* uno dei due possibili valori di verità 0 e 1. Spesso però si vengono a considerare enunciati attinenti a fatti che non conosciamo perché non si sono ancora verificati o perché comunque non abbiamo su di essi abbastanza informazioni. Ad esempio:

(1) nel prossimo turno del campionato di calcio, la squadra della Fiorentina pareggerà;
(2) l'attuale governo sarà ancora in carica il 20 novembre del prossimo anno;
(3) lanciando questo dado otterrò il numero 5;
(4) la somma dei numeri estratti sulla ruota di Napoli nella prossima estrazione del lotto sarà almeno 15;
(5) la prima carta di questo mazzo (che è stato mescolato mentre io non guardavo e giace coperto davanti a me) è un asso.

Ogni volta che ci si trova a considerare un enunciato attinente a fatti sconosciuti, non si può *in generale* attribuirgli "a priori" il valore 0 né il valore 1 . Come "surrogato", il calcolo delle probabilità ci consente di assegnare a un tale enunciato un numero reale nell'intervallo [0, 1] che esprime la maggiore o minore fiducia che abbiamo nel suo risultare poi vero (fiducia tanto maggiore quanto più il numero è vicino a 1).

Schematizzeremo informalmente ciascuna situazione che intendiamo studiare individuando

(a) l'*esperimento* (o *fenomeno causante*) ;
(b) una descrizione dei suoi possibili *risultati* (che si escludono reciprocamente ed esauriscono tutte le possibilità) ;
(c) un elenco di *eventi* che possono verificarsi o non verificarsi a seconda di quale particolare risultato ha avuto l'esperimento; è a ciascuno di questi eventi che vogliamo assegnare un numero reale compreso fra 0 e 1 detto "probabilità" .

Un evento si dice *certo* se qualunque risultato dell'esperimento lo rende vero. Un evento si dice *impossibile* se nessun risultato dell'esperimento lo rende vero. Due eventi si dicono *incompatibili* se nessun risultato dell'esperimento li rende veri entrambi.

Un evento che consiste nel verificarsi di un risultato si dice *evento elementare*.

Esempi

2.1.1 Esperimento: la partita della Fiorentina nel prossimo turno del campionato di calcio. Risultati: la Fiorentina vince, la Fiorentina pareggia, la Fiorentina perde. Eventi: la Fiorentina vince (evento elementare), una delle due squadre vince, la Fiorentina perde (evento elementare), nessuna delle due squadre vince, entrambe le squadre vincono, ecc. ecc.
Si noti peraltro che allo stesso esperimento si possono associare anche altri insiemi di risultati; ad esempio i possibili "punteggi" con cui si conclude la partita: 0 – 0, 1 – 0, 0 – 1, 1 – 1, 2 – 0, ecc. ecc. ; e (di conseguenza) un insieme "più ampio" di eventi: gli stessi visti sopra (ma quelli che risultavano eventi elementari ora non lo sono più), e inoltre: la partita termina a reti inviolate (evento elementare), la partita termina con una vittoria per più di tre reti di scarto, ecc. ecc.

2.1.2 Esperimento: l'evolversi della situazione politica fino al 20 novembre del prossimo anno. Risultati: tutti i diversi governi che possono teoricamente essere in carica al 20 novembre del prossimo anno. Eventi: l'attuale governo sarà ancora in carica al 20 novembre del prossimo anno (evento elementare); un diverso governo formato dalle stesse persone sarà in carica al 20 novembre; un governo con lo stesso ministro degli Esteri sarà in carica al 20 novembre del prossimo anno; un governo con tutti i ministri diversi dagli attuali sarà in carica al 20 novembre del prossimo anno; ecc., ecc., ...

2.1.3 Esperimento: lancio di un dado. Risultati: 1, 2, 3, 4, 5, 6. Eventi: un numero pari, un numero primo, il numero 3 (evento elementare), un numero maggiore di quattro, ecc. ecc.

2.1.4 Esperimento: la prossima estrazione del lotto sulla ruota di Napoli. Risultati: (1, 2, 3, 4, 5), (1, 2, 3, 4, 6), ..., (86, 87, 88, 89, 85), (86, 87, 88, 89, 90) (cioè tutte le disposizioni semplici a 5 a 5 dei numeri naturali da 1 a 90, cfr. sez. 13.3). Eventi: viene estratto il 7, il secondo estratto è 5, viene estratto l'ambo {2, 9}, viene estratto l'ambo {15, 87}, viene estratto il terno {7, 9, 16}, viene estratta la cinquina {1, 12, 19, 35, 53}, ecc. ecc.
Si noti peraltro che allo stesso esperimento si possono associare anche altri insiemi di risultati; ad esempio l'insieme delle 5 – combinazioni semplici dei numeri da 1 a 90 (cioè, per il teorema 13.4.1, i possibili insiemi di cinque numeri estraibili, senza tener conto dell'ordine di estrazione): questa scelta dell'insieme di risultati consentirebbe comunque di esprimere gli eventi "pagabili" in base alle norme sul gioco del lotto (cioè le ambate, gli ambi, i terni, le quaterne e le cinquine).

2.1.5 Esperimento: mescolare un mazzo di carte. Risultati: tutti i possibili diversi ordini in cui le carte possono presentarsi. Eventi: la prima carta è un asso, la prima carta è una carta di cuori, nelle prime dieci carte ci sono almeno due assi oppure almeno quattro carte di picche, ecc. ecc.

Esercizio 2.1.6

Fra gli eventi (1), (2), (3), (4) e (5) descritti all'inizio ce n'è qualcuno certo? Ce n'è qualcuno impossibile?

2.2 - Spazio dei risultati.

Per dare una formalizzazione matematica dei concetti introdotti nella sez. 2.1, ad ogni esperimento si associa un insieme non vuoto Ω (detto *spazio dei risultati*) i cui *elementi* "rappresentano" i possibili *risultati*; gli *eventi* ai quali vogliamo associare una *probabilità* saranno identificati con opportuni *sottoinsiemi* di Ω.

La scelta di Ω è arbitraria, come del resto (lo si è già osservato) la descrizione dei risultati dell'esperimento; vedremo tuttavia in seguito che alcune scelte sono "migliori" di altre. Ad ogni modo, fissato Ω, ogni evento risulta identificato con il sottoinsieme di Ω formato dai risultati che lo rendono vero, cosicché in particolare:

– l'evento impossibile è $\emptyset$;
– l'evento certo è Ω;
– gli eventi elementari sono quelli che hanno un solo elemento;
– due eventi sono incompatibili se e solo se sono disgiunti.

Esempio 2.2.1

Supponiamo di voler descrivere l'esperimento "lancio di due monete". Una possibile scelta di Ω è

$$\Omega_1 := \{TT, CC, TC\}$$

dove: TT è il risultato <escono due "testa">, CC è il risultato <escono due "croce">, TC è il risultato <escono una "testa" e una "croce">.

L'evento <escono due "testa"> è il sottoinsieme {TT}: si tratta dunque di un evento elementare (infatti consiste nel verificarsi di un risultato). L'evento "escono due facce uguali" è il sottoinsieme {TT, CC} ; l'evento "esce testa su almeno una moneta" è il sottoinsieme {TT, TC}.

Un'altra possibile scelta di Ω è $\Omega_2 := \{TT, CC, TC, CT\}$

dove questa volta si distingue tra TC e CT a seconda che sia uscita "testa" sull'una o sull'altra moneta.

Notiamo esplicitamente che <u>entrambe</u> le scelte per Ω sono lecite: una però, come vedremo più avanti (Esempio 2.3.6), è "più comoda" dell'altra. Notiamo anche che: è vero che certi eventi (ad esempio: "esce testa sulla prima moneta") si possono descrivere con Ω_2 (come {TT, TC}) e <u>non</u> con Ω_1 ; ma questo di per sé <u>non</u> è un motivo determinante per preferire Ω_2, almeno finché l'evento al quale siamo interessati non è uno di questi!

2.3 - Misura di probabilità.

Sia $\mathbf{\Omega}$ un insieme non vuoto. Si dice *misura di probabilità* su $\mathbf{\Omega}$ una funzione $p : \mathcal{P}(\mathbf{\Omega}) \to [0,\, 1]$ tale che:

$\boxed{\text{MP1}}$ $\mathcal{D}(p)$ è una $\sigma-$algebra, cioè:

$\boxed{\text{MP1.1}}$ $\mathbf{\Omega} \in \mathcal{D}(p)$;

$\boxed{\text{MP1.2}}$ per ogni $\mathbf{A} \in \mathcal{D}(p)$, anche $\mathbf{A}^{\mathbf{C}} \in \mathcal{D}(p)$;

$\boxed{\text{MP1.3}}$ per ogni sottoinsieme **finito o numerabile** $\boldsymbol{\mathcal{I}}$ di $\mathcal{D}(p)$, anche $\cup \boldsymbol{\mathcal{I}} \in \mathcal{D}(p)$;

$\boxed{\text{MP1.4}}$ per ogni sottoinsieme **finito o numerabile** $\boldsymbol{\mathcal{I}}$ di $\mathcal{D}(p)$, anche $\cap \boldsymbol{\mathcal{I}} \in \mathcal{D}(p)$;

$\boxed{\text{MP2}}$ $p(\mathbf{\Omega}) = 1$;

$\boxed{\text{MP3}}$ se $\mathbf{A_1}, \mathbf{A_2}, \ldots, \mathbf{A_n}, \ldots \in \mathcal{D}(p)$ e $\mathbf{A_i} \cap \mathbf{A_j} = \emptyset$ per $i \neq j$, allora

$$p(\cup \{\mathbf{A_i}\}) = \lim_{n \to +\infty} \sum_{i=1}^{n} p(\mathbf{A_i}) \qquad \text{(cioè, } p \text{ è } \sigma-\text{additiva).}$$

I sottoinsiemi di $\mathbf{\Omega}$ che appartengono al dominio di p sono gli *eventi* ai quali siamo in grado di assegnare una probabilità.

Le (i) e (ii) sono le "regole generali del gioco": da sole non bastano per determinare le probabilità dei vari eventi (ci vogliono "regole", cioè ipotesi, aggiuntive), ma da esse si possono già dedurre vari fatti.

$\boxed{\text{Teorema 2.3.1}}$

Sia $\mathbf{\Omega}$ un insieme non vuoto, e sia p una misura di probabilità su $\mathbf{\Omega}$. Allora

(a) $\emptyset \in \mathcal{D}(p)$;

(b) se $\mathbf{A_1}, \mathbf{A_2}, \ldots, \mathbf{A_s}$ sono elementi di $\mathcal{D}(p)$ a due a due disgiunti,
$$p(\mathbf{A_1} \cup \mathbf{A_2} \cup \ldots \cup \mathbf{A_s}) = p(\mathbf{A_1}) + p(\mathbf{A_2}) + \ldots + p(\mathbf{A_s});$$

(c) se $\mathbf{A} \in \mathcal{D}(p)$, si ha $p(\mathbf{A}^{\mathbf{C}}) = 1 - p(\mathbf{A})$;

(d) $p(\emptyset) = 0$;

(e) se $\mathbf{A}, \mathbf{B} \in \mathcal{D}(p)$, si ha $\mathbf{A} \backslash \mathbf{B} \in \mathcal{D}(p)$;

(f) se $\mathbf{A}, \mathbf{B} \in \mathcal{D}(p)$ e $\mathbf{A} \subseteq \mathbf{B}$, si ha $p(\mathbf{A}) \leq p(\mathbf{B})$;

(g) se $\mathbf{A}, \mathbf{B} \in \mathcal{D}(p)$, si ha $p(\mathbf{A} \cup \mathbf{B}) = p(\mathbf{A}) + p(\mathbf{B}) - p(\mathbf{A} \cap \mathbf{B})$.

Dimostrazione -

(a) Poiché $\mathbf{\Omega} \in \mathcal{D}(p)$ (per la $\boxed{\text{MP1.1}}$) e $\emptyset = \mathbf{\Omega}^{\mathbf{C}}$, si ha che $\emptyset \in \mathcal{D}(p)$ per la $\boxed{\text{MP1.2}}$.

(b) Posto $\mathbf{A_i} = \emptyset$ per $i > s$ (come possiamo fare per la (a)), l'asserto segue immediatamente dalla $\boxed{\text{MP3}}$.

(c) Sia $\mathbf{A} \in \mathcal{D}(p)$. Poiché $\mathbf{A} \cap \mathbf{A}^{\mathbf{C}} = \emptyset$, si ha
$$p(\mathbf{A}) + p(\mathbf{A}^{\mathbf{C}}) \overset{(b)}{=} p(\mathbf{A} \cup \mathbf{A}^{\mathbf{C}}) = p(\mathbf{\Omega}) \overset{\boxed{\text{MP2}}}{=} 1$$
e quindi
$$p(\mathbf{A}^{\mathbf{C}}) = 1 - p(\mathbf{A}) \qquad \text{come si voleva.}$$

(d) Poiché $\emptyset = \mathbf{\Omega}^{\mathbf{C}}$, per la (c) si trova che $p(\emptyset) = p(\mathbf{\Omega}^{\mathbf{C}}) = 1 - p(\mathbf{\Omega}) = 1 - 1 = 0$.

(e) Poiché $\mathbf{A} \backslash \mathbf{B} = \mathbf{A} \cap \mathbf{B}^{\mathbf{C}}$, per le $\boxed{\text{MP1.2}}$ e $\boxed{\text{MP1.3}}$ si ha subito l'asserto.

(f) Siano $\mathbf{A}, \mathbf{B} \in \mathcal{D}(p)$ con $\mathbf{A} \subseteq \mathbf{B}$. Allora $\mathbf{B} = \mathbf{A} \cup (\mathbf{B} \backslash \mathbf{A})$, e $\mathbf{A} \cap (\mathbf{B} \backslash \mathbf{A}) = \emptyset$. Dunque per la (b)

$$p(\mathbf{B}) = p(\mathbf{A}) + p(\mathbf{B} \backslash \mathbf{A}) \geq p(\mathbf{A})$$

ricordando che per definizione p assume solo valori non negativi.

(g) Siano $\mathbf{A}, \mathbf{B} \in \mathcal{D}(p)$.

Osserviamo in primo luogo che $\mathbf{A}$ è unione degli insiemi disgiunti $\mathbf{A} \backslash \mathbf{B}$ e $\mathbf{A} \cap \mathbf{B}$, cosicché

$$p(\mathbf{A}) = p(\mathbf{A} \backslash \mathbf{B}) + p(\mathbf{A} \cap \mathbf{B}) \qquad \text{per la } (b)\,;$$

e analogamente

$$p(\mathbf{B}) = p(\mathbf{B} \backslash \mathbf{A}) + p(\mathbf{A} \cap \mathbf{B})\,.$$

Notiamo poi che $\mathbf{A} \cup \mathbf{B}$ è unione degli insiemi a due a due disgiunti $\mathbf{A} \backslash \mathbf{B}$, $\mathbf{B} \backslash \mathbf{A}$ e $\mathbf{A} \cap \mathbf{B}$. Tenendo conto della (b) si ha quindi

$$p(\mathbf{A} \cup \mathbf{B}) = p(\mathbf{A} \backslash \mathbf{B}) + p(\mathbf{B} \backslash \mathbf{A}) + p(\mathbf{A} \cap \mathbf{B}) =$$

$$= \underbrace{p(\mathbf{A} \backslash \mathbf{B}) + p(\mathbf{A} \cap \mathbf{B})}_{p(\mathbf{A})} - p(\mathbf{A} \cap \mathbf{B}) + \underbrace{p(\mathbf{B} \backslash \mathbf{A}) + p(\mathbf{A} \cap \mathbf{B})}_{p(\mathbf{B})} = p(\mathbf{A}) + p(\mathbf{B}) - p(\mathbf{A} \cap \mathbf{B})$$

come si voleva.

Supponiamo ora che *gli eventi elementari siano tutti equiprobabili*. Attenzione: questa ipotesi non sempre è ragionevole ([4])(dipende dalla scelta di $\mathbf{\Omega}$ per rappresentare l'esperimento che stiamo considerando!) ; essa tuttavia ci consente di determinare la probabilità degli eventi elementari e, nel caso particolare che $\mathbf{\Omega}$ sia finito, di un qualsiasi evento.

$\boxed{\text{Teorema 2.3.2}}$

Sia $\mathbf{\Omega}$ un insieme non vuoto, e sia p una misura di probabilità su $\mathbf{\Omega}$ tale che

(i) ogni evento elementare appartiene a $\mathcal{D}(p)$

e

(ii) $p(\{\mathrm{x}\}) = p(\{\mathrm{y}\}) \quad \forall\, \mathrm{x}, \mathrm{y} \in \mathbf{\Omega}$.

Allora

(a) se $|\mathbf{\Omega}| = n \in \mathbb{Z}^+$,
- $(a.1)$ ogni evento elementare ha probabilità $\frac{1}{n}$;
- $(a.2)$ $\mathcal{D}(p) = \mathcal{P}(\mathbf{\Omega})$;
- $(a.3)$ per ogni $\mathbf{A} \subseteq \mathbf{\Omega}$ si ha $p(\mathbf{A}) = \frac{|\mathbf{A}|}{n}$;

(b) se $\mathbf{\Omega}$ ha infiniti elementi, ogni evento elementare ha probabilità zero .

[4] e nemmeno sempre possibile! Cfr. teorema 2.3.8.

Dimostrazione -

(a) Siano x_1, x_2, $\ldots$, x_n gli elementi di $\mathbf{\Omega}$, e supponiamo che sia $p(\{\mathrm{x}_i\}) = \varepsilon$ (per ipotesi tale valore non dipende da i).

$(a.1)$ Poiché $\mathbf{\Omega} = \bigcup_{i=1}^{n} \{\mathrm{x}_i\}$ e gli $\{\mathrm{x}_i\}$ sono a due a due disgiunti, per la (b) del teorema 2.3.1 si ha

$$1 = p(\mathbf{\Omega}) = \sum_{i=1}^{n} p\left(\{\mathrm{x}_i\}\right) = n\varepsilon$$

da cui $\varepsilon = \frac{1}{n}$.

$(a.2)$ Ogni sottoinsieme di $\mathbf{\Omega}$ appartiene a $\mathcal{D}(p)$ per la $\boxed{\text{MP1.3}}$ essendo unione di un numero finito di eventi elementari.

$(a.3)$ Se $\mathbf{A} \subseteq \mathbf{\Omega}$ con $|\mathbf{A}| = m$, siano $\mathrm{x}_{i_1}, \mathrm{x}_{i_2}, \ldots, \mathrm{x}_{i_m}$ gli m elementi di $\mathbf{A}$. Poiché $\mathbf{A}$ è unione degli m eventi elementari $\{\mathrm{x}_{i_1}\}$, $\{\mathrm{x}_{i_2}\}$, $\ldots$, $\{\mathrm{x}_{i_m}\}$ a due a due disgiunti (che hanno ciascuno probabilità $\frac{1}{n}$), applicando la (b) del teorema 2.3.1 si trova che

$$p(\mathbf{A}) = \sum_{j=1}^{m} p\left(\{\mathrm{x}_{i_j}\}\right) = m \cdot \frac{1}{n} = \frac{m}{n} = \frac{|\mathbf{A}|}{n}$$

come si voleva.

(b) Supponiamo ora che $\mathbf{\Omega}$ abbia infiniti elementi.
Sia ε la probabilità di un evento elementare, e supponiamo per assurdo che sia $\varepsilon > 0$. Per la proprietà di Archimede (teorema 9.5.2), esiste un numero naturale k tale che $k\varepsilon > 1$; scelti allora k risultati distinti $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_k \in \mathbf{\Omega}$ e posto $\mathbf{A} := \{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_k\}$, si ha (ricordando come al solito la (b) del teorema 2.3.1)

$$p(\mathbf{A}) = \sum_{j=1}^{k} p\left(\{\mathrm{x}_j\}\right) = k\varepsilon > 1$$

contro la (f) del teorema 2.3.1 e la $\boxed{\text{MP2}}$.

$\boxed{\text{Osservazione 2.3.3}}$

Il risultato espresso dalla (a) del teorema 2.3.2 è noto come *valutazione classica della probabilità.*

$\boxed{\text{Osservazione 2.3.4}}$

La (b) del teorema 2.3.2 evidenzia con chiarezza che un evento con probabilità zero non è in generale impossibile. Supponiamo ad esempio di lanciare contro un bersaglio una freccetta “a caso” (cioè senza mirare). Possiamo allora convenire che ogni punto abbia la stessa probabilità di essere colpito, e dunque (per la (b) del teorema 2.3.2) tale probabilità è zero; ciò ovviamente non significa che sia *impossibile* colpire esattamente un dato punto. Allo stesso modo (si pensi all'evento “colpire un qualunque punto diverso da $\mathbf{P}_0$” e si applichi la (c) del teorema 2.3.1) si vede che un evento con probabilità 1 non è in generale certo.

Esempio 2.3.5

Lanciamo un dado (cubico) e poniamo $\mathbf{\Omega} := \{1, 2, 3, 4, 5, 6\}$. Se il dado non è truccato, possiamo ragionevolmente supporre che gli eventi elementari siano equiprobabili; per la (a) del teorema 2.3.2, ogni evento elementare ha probabilità $\frac{1}{6}$.
Qual è la probabilità che esca un numero pari? Posto $\mathbf{A} := \{2, 4, 6\}$, per la ($a$) del teorema 2.3.2 è $p(\mathbf{A}) = \frac{3}{6} = \frac{1}{2}$.
Qual è la probabilità che esca un numero primo? Posto $\mathbf{B} := \{2, 3, 5\}$, è $p(\mathbf{B}) = \frac{3}{6} = \frac{1}{2}$.
Qual è la probabilità che esca un numero pari oppure un numero primo? Possiamo calcolarla applicando ad $\mathbf{A} \cup \mathbf{B}$ ancora la (a) del teorema 2.3.2; oppure possiamo utilizzare la (g) del teorema 2.3.1: poiché $\mathbf{A} \cap \mathbf{B} = \{2\}$, $\mathbf{A} \cap \mathbf{B}$ è un evento elementare e dunque ha probabilità $\frac{1}{6}$; ne segue che

$$p(\mathbf{A} \cup \mathbf{B}) = p(\mathbf{A}) + p(\mathbf{B}) - p(\mathbf{A} \cap \mathbf{B}) = \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \frac{5}{6} .$$

Esempio 2.3.6

Lanciamo due monete. Abbiamo già descritto (Esempio 2.2.1) due possibili spazi dei risultati:

$$\mathbf{\Omega_1} := \{\mathrm{TT, CC, TC}\} \qquad \text{e} \qquad \mathbf{\Omega_2} := \{\mathrm{TT, CC, TC, CT}\} .$$

È chiaro che non possiamo supporre <u>in</u> <u>entrambi</u> <u>i</u> <u>casi</u> che gli eventi elementari siano equiprobabili. Sarà preferibile scegliere come spazio dei risultati quello tra i due per il quale una tale ipotesi risulti adeguata alla descrizione della realtà.

Esempio 2.3.7

Esperimento: lancio di una freccetta contro un bersaglio. Che ogni punto abbia la stessa probabilità di essero colpito è supposizione un po' ardita (perché si mira al centro, quindi è più probabile colpire un punto vicino al centro che un punto vicino al bordo del bersaglio!) ([5]); e una tale ipotesi <u>non</u> è comunque sufficiente per determinare la probabilità degli eventi non elementari (tipo: viene colpito un punto del quadrante superiore destro). Occorrono ipotesi aggiuntive. Se il bersaglio è circolare, possiamo ad esempio fare la seguente ipotesi:
(•) settori circolari equiestesi hanno la stessa probabilità di essere colpiti.
Si può dimostrare che, sotto questa ipotesi, la probabilità di colpire un assegnato settore circolare è direttamente proporzionale alla sua superficie.

Teorema 2.3.8

Sia $\mathbf{\Omega}$ un insieme in cui tutti gli eventi elementari appartengono a $\mathcal{D}(p)$.
Se $|\mathbf{\Omega}| = \aleph_0$, gli eventi elementari non possono avere tutti la stessa probabilità.

Dimostrazione - Per la (b) del teorema 2.3.2, se gli eventi elementari fossero equiprobabili ciascuno di essi dovrebbe avere probabilità zero. Essendo $\mathbf{\Omega}$ l'unione degli eventi elementari, per la $\sigma -$ additività di p ($\boxed{\text{MP3}}$) dovrebbe essere $p(\mathbf{\Omega}) = 0$ contro la $\boxed{\text{MP2}}$.

---

[5] Ad ogni modo, è sufficiente supporre, ad esempio, che <per ogni circonferenza $\mathcal{C}$ con centro l'origine, ogni punto abbia la stessa probabilità $p_{\mathcal{C}}$ di essere colpito> per poter dedurre che $p_{\mathcal{C}} = 0$ per ogni $\mathcal{C}$.

2.4 - La valutazione frequentista della probabilità.

Si consideri l'esperimento "lancio di un dado".

Possiamo associargli lo spazio dei risultati $\mathbf{\Omega} := \{1, 2, 3, 4, 5, 6\}$, "decidendo" che gli eventi elementari sono tutti equiprobabili. Dunque, la probabilità che esca (ad esempio) un "5" è $\frac{1}{6}$.

Nella misura in cui il nostro ragionamento si riferisce non a una pura astrazione matematica ma a una situazione concreta, abbiamo il dovere di chiederci se la scelta di $\mathbf{\Omega}$ e la decisione di considerare equiprobabili gli eventi elementari siano *adeguate* alla rappresentazione della realtà. Ciò ovviamente dipende da fattori concreti tipo: il dado è un cubo perfetto? è costituito di materiale omogeneo? Viene lanciato con qualche particolare criterio? Ecc. ecc.

Per gli esperimenti che possono essere ripetuti più volte nelle stesse condizioni, un criterio di adeguatezza alla realtà di un modello probabilistico è il cosiddetto *criterio frequentista*.

Si consideri un esperimento che può essere ripetuto più volte nelle stesse condizioni (ad esempio: lancio di una dado, lancio di una moneta, estrazione di una pallina da un'urna; ma non: partita di calcio, evolversi della situazione politica), e sia **E** un evento fissato.

Sia n un numero intero positivo, e sia Ξ_n l'esperimento che consiste nel ripetere n volte l'esperimento considerato. Se in Ξ_n l'evento **E** risulta vero k volte, il numero

$$\frac{k}{n}$$

si dice *frequenza relativa* di **E** osservata in Ξ_n .

Si noti che la frequenza relativa è un numero reale (di fatto addirittura razionale) compreso tra zero e uno. Da una misura di probabilità "adeguata" alla rappresentazione della realtà ci si aspetta che, per ogni evento $\mathbf{E} \in \mathcal{D}(p)$, ripetendo l'esperimento un numero di volte "sufficientemente grande" la frequenza relativa di **E** risulti "vicina" a $p(\mathbf{E})$.

Di fatto, la frequenza relativa di un evento osservata in una precedente successione di esperimenti può venire scelta come stima della probabilità di quell'evento per gli esperimenti successivi; ma, come si è già osservato, questo procedimento può essere adottato solo quando si considerano esperimenti che vengono ripetuti sempre nelle stesse condizioni.

Esempio 2.4.1

Consideriamo l'esperimento che consiste nel lanciare una moneta. Scegliamo come spazio dei risultati $\mathbf{\Omega} := \{\mathrm{T},\ \mathrm{C}\}$, e supponiamo, in assenza di altre informazioni, che gli eventi elementari siano equiprobabili, cosicché $p(\mathrm{T}) := \frac{1}{2}$ e $p(\mathrm{C}) := \frac{1}{2}$.
Se ripetendo l'esperimento 1000 volte abbiamo osservato che è uscita 900 volte "Testa" e 100 volte "Croce", come possiamo valutare la probabilità che nel lancio successivo esca "Croce" ?
Se vogliamo adottare il criterio della frequenza relativa, dobbiamo rifiutare la precedente funzione p proponendone invece un'altra per la quale si abbia $p(\mathrm{T}) := \frac{9}{10}$ e $p(\mathrm{C}) := \frac{1}{10}$. In altre parole, si tratta di prendere atto che la moneta è presumibilmente truccata!

Esercizio 2.4.2

Si prenda una “moneta truccata” (ad esempio, il tappo di un barattolo di marmellata, oppure una puntina da disegno) e la si lanci 100 volte, prendendo nota del risultato di ciascun lancio. Col criterio della frequenza relativa osservata in tale successione di esperimenti, si assegni una probabilità a ciascuna delle due facce della “moneta”.
Si ripeta 50, 100, 150 volte l’esperimento prendendo ogni volta nota della frequenza relativa di ciascun risultato e confrontandola con la probabilità assegnata.

2.5 - Probabilità condizionata.

Consideriamo il seguente esperimento: da un’urna contenente i numeri 1, 2 e 3 estraiamo due numeri uno dopo l’altro.

Scegliamo come spazio dei risultati l’insieme $\mathbf{\Omega} := \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}$ delle 2 – disposizioni semplici dei numeri 1, 2 e 3 (cfr. sez. 13.3). Possiamo supporre che gli eventi elementari siano tutti equiprobabili, e che quindi ciascuno di essi abbia probabilità $\frac{1}{6}$.

L’evento <il secondo numero estratto è 3> si rappresenta col sottoinsieme $\mathbf{A} := \{(1, 3), (2, 3)\}$ di $\mathbf{\Omega}$ formato dalle coppie ordinate che hanno come secondo elemento 3, e dunque (per il teorema 2.3.2) ha *a priori* probabilità $\frac{1}{3}$. Supponiamo però di essere all’estrazione del secondo numero e di conoscere già il primo numero estratto, ad esempio: 1. Abbiamo delle informazioni in più rispetto a prima: gli eventi elementari <u>non si possono più supporre equiprobabili</u>; in effetti, tutti quelli individuati da coppie ordinate il cui primo elemento è $\neq 1$ hanno probabilità zero! Possiamo supporre equiprobabili i rimanenti due, ciascuno dei quali ha dunque adesso probabilità $\frac{1}{2}$. L’evento $\mathbf{A}$ è unione degli eventi elementari $\{(1, 3)\}$ (che ha probabilità $\frac{1}{2}$) e $\{(2, 3)\}$ (che ha probabilità 0), pertanto adesso $p(\mathbf{A}) = \frac{1}{2}$.

Abbiamo visto con questo esempio che, se sappiamo che è vero un dato evento $\mathbf{B}$, è ragionevole assegnare a tutti gli eventi una nuova misura di probabilità $p_{\mathbf{B}}$, detta *probabilità condizionata a* (oppure *da*) $\mathbf{B}$; per ogni evento $\mathbf{A}$, si scrive di solito $p(\mathbf{A}|\mathbf{B})$ anziché $p_{\mathbf{B}}(\mathbf{A})$.

L’informazione che è vero un dato evento $\mathbf{B}$ può a seconda dei casi influire profondamente oppure non influire affatto sulla probabilità di un altro evento $\mathbf{A}$. Ad esempio, tornando alla situazione esaminata sopra, sia $\mathbf{B}$ l’evento <il primo numero estratto è 3> e sia $\mathbf{A}$ l’evento <il secondo numero estratto è 3>. è chiaro che $p(\mathbf{A}|\mathbf{B}) = 0$ perché se 3 è già stato estratto come primo numero non può essere estratto come secondo! Più in generale, si ha $p(\mathbf{A}|\mathbf{B}) = 0$ ogni volta che $\mathbf{A} \cap \mathbf{B} = \emptyset$ (cioè ogni volta che $\mathbf{A}$, $\mathbf{B}$ sono incompatibili) e $p(\mathbf{A}|\mathbf{A}) = 1$.

Sia $\mathbf{\Omega}$ un insieme finito non vuoto, e sia p una misura di probabilità su $\mathbf{\Omega}$ per la quale gli eventi elementari sono tutti equiprobabili, cosicché si può applicare la (a) del teorema 2.3.2. Sia $\mathbf{B}$ un evento diverso dall’evento impossibile. Sotto queste ipotesi, per valutare la probabilità condizionata all’evento $\mathbf{B}$ è sufficiente assumere $\mathbf{B}$ come nuovo spazio dei risultati mantenendo in $\mathbf{B}$ l’ipotesi (già formulata per $\mathbf{\Omega}$) di equiprobabilità per gli eventi elementari.

Applicando la (a) del teorema 2.3.2 sia allo spazio dei risultati $\mathbf{\Omega}$ (per calcolare $p(\mathbf{A} \cap \mathbf{B})$ e $p(\mathbf{B})$) sia allo spazio dei risultati $\mathbf{B}$ (per calcolare $p(\mathbf{A}|\mathbf{B})$), si trova che

$$p(\mathbf{A} \cap \mathbf{B}) = \frac{|\mathbf{A} \cap \mathbf{B}|}{|\mathbf{\Omega}|} \ ; \qquad p(\mathbf{B}) = \frac{|\mathbf{B}|}{|\mathbf{\Omega}|} \ ; \qquad p(\mathbf{A}|\mathbf{B}) = \frac{|\mathbf{A} \cap \mathbf{B}|}{|\mathbf{B}|}$$

da cui

2.5.F1 $$p(\mathbf{A} \cap \mathbf{B}) = p(\mathbf{A}|\mathbf{B}) \cdot p(\mathbf{B})\,.$$

In generale, sia $\mathbf{\Omega}$ un qualsiasi insieme non vuoto, e sia p una misura di probabilità su $\mathbf{\Omega}$. Se $\mathbf{A}\,,\mathbf{B} \in \mathcal{D}(p)$ con $p(\mathbf{B}) \neq 0$, si definisce la probabilità di $\mathbf{A}$ condizionata a (oppure da) $\mathbf{B}$ ponendo

2.5.F2 $$p(\mathbf{A}|\mathbf{B}) = \frac{p(\mathbf{A} \cap \mathbf{B})}{p(\mathbf{B})}\,.$$

Si noti che, quando $p(\mathbf{B}) \neq 0$, la 2.5.F2 è equivalente alla 2.5.F1.

Teorema 2.5.1

Siano $\mathbf{\Omega}$ un insieme non vuoto e p una misura di probabilità su $\mathbf{\Omega}$. Si ha

(i) $\quad p(\mathbf{A}|\mathbf{A}) = 1 \quad \forall\, \mathbf{A} \in \mathcal{D}(p)$ con $p(\mathbf{A}) \neq 0$;

(ii) $\quad p(\mathbf{A}|\mathbf{\Omega}) = p(\mathbf{A}) \quad \forall\, \mathbf{A} \in \mathcal{D}(p)$;

(iii) $\quad p(\mathbf{A}|\mathbf{B}) = p(\mathbf{A} \cap \mathbf{B}|\mathbf{B}) \quad \forall\, \mathbf{A}\,,\mathbf{B} \in \mathcal{D}(p)$ con $p(\mathbf{B}) \neq 0$;

(iv) $\quad p(\mathbf{A}|\mathbf{C}) = p(\mathbf{A}|\mathbf{B}) \cdot p(\mathbf{B}|\mathbf{C}) \quad \forall\, \mathbf{A}\,,\mathbf{B}\,,\mathbf{C} \in \mathcal{D}(p)$ tali che $\mathbf{A} \subseteq \mathbf{B} \subseteq \mathbf{C}$ e $p(\mathbf{B}),\ p(\mathbf{C}) \neq 0$.

Dimostrazione - Ovvio.

Esempio 2.5.2

Esperimento: estrazione del lotto. $\mathbf{A}$:= <Il secondo numero estratto è 3> ; $\mathbf{B}$:= <Il primo numero estratto è 15>.

$$p(\mathbf{A}) = \frac{1}{90}\,, \quad p(\mathbf{B}) = \frac{1}{90}\,, \quad p(\mathbf{A} \cap \mathbf{B}) = \frac{88 \cdot 87 \cdot 86}{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86} = \frac{1}{90 \cdot 89}$$

$$p(\mathbf{A}|\mathbf{B}) = \frac{p(\mathbf{A} \cap \mathbf{B})}{p(\mathbf{B})} = \frac{1}{90 \cdot 89} \cdot 90 = \frac{1}{89}\,.$$

Esempio 2.5.3

Esperimento: lancio di un dado. $\mathbf{A}$:= <Esce 2> ; $\mathbf{B}$:= <Esce un numero pari>.

$$p(\mathbf{A}) = \frac{1}{6}\,, \quad p(\mathbf{B}) = \frac{1}{2}\,, \quad p(\mathbf{A} \cap \mathbf{B}) = \frac{1}{6} \quad (\text{essendo } \mathbf{A} \cap \mathbf{B} = \mathbf{A})$$

$$p(\mathbf{A}|\mathbf{B}) = \frac{p(\mathbf{A} \cap \mathbf{B})}{p(\mathbf{B})} = \frac{1}{6} \cdot 2 = \frac{1}{3}\,.$$

Esempio 2.5.4

Esperimento: lancio di due dadi colorati in rosso e in blu (per distinguerli).
$\mathbf{A} :=$ <Sul dado rosso esce 2> ; $\mathbf{B} :=$ <Sul dado blu esce 3> .

$$p(\mathbf{A}) = \frac{1}{6}, \quad p(\mathbf{B}) = \frac{1}{6}, \quad p(\mathbf{A} \cap \mathbf{B}) = \frac{1}{36}$$

(avendo posto $\mathbf{\Omega} := \{(i, \; j) \in \mathbb{N}^2 \;\; / \;\; 1 \leq i \leq 6, \quad 1 \leq j \leq 6\}$ ed essendo $|\mathbf{\Omega}| = 36$, $|\mathbf{A} \cap \mathbf{B}| = |\{(2, 3)\}| = 1$)
cosicché $$p(\mathbf{A}|\mathbf{B}) = \frac{p(\mathbf{A} \cap \mathbf{B})}{p(\mathbf{B})} = \frac{1}{36} \cdot 6 = \frac{1}{6} = p(\mathbf{A}).$$

Esempio 2.5.5

Esperimento: estrazione di due numeri tra i numeri 1, 2 e 3 .
$\mathbf{\Omega} := \{(1, 2), (1, 3),(2, 1),(2, 3),(3, 1),(3, 2)\}$
Supponiamo che gli eventi elementari siano equiprobabili; pertanto, ciascuno di essi ha probabilità $\frac{1}{6}$.
$\mathbf{A} :=$ <il secondo estratto è il 2> ; $\mathbf{B} :=$ <il primo estratto è il 3> .

$$p(\mathbf{A}) = \frac{2}{6} = \frac{1}{3}, \quad p(\mathbf{B}) = \frac{2}{6} = \frac{1}{3}, \quad p(\mathbf{A} \cap \mathbf{B}) = \frac{1}{6}$$

cosicché $$p(\mathbf{A}|\mathbf{B}) = \frac{p(\mathbf{A} \cap \mathbf{B})}{p(\mathbf{B})} = \frac{1}{6} \cdot 3 = \frac{1}{2}.$$

Esempio 2.5.6

Esperimento: estrazione di due numeri tra i numeri 1, 2 e 3 .
$\mathbf{\Omega} := \{(1, 2), (1, 3),(2, 1),(2, 3),(3, 1),(3, 2)\}$
Supponiamo che gli eventi elementari siano equiprobabili; pertanto, ciascuno di essi ha probabilità $\frac{1}{6}$.
$\mathbf{A} :=$ <viene estratto il 2> ; $\mathbf{B} :=$ <viene estratto il 3> .

$$p(\mathbf{A}) = \frac{4}{6} = \frac{2}{3}, \quad p(\mathbf{B}) = \frac{4}{6} = \frac{2}{3}, \quad p(\mathbf{A} \cap \mathbf{B}) = \frac{2}{6} = \frac{1}{3}$$

cosicché $$p(\mathbf{A}|\mathbf{B}) = \frac{p(\mathbf{A} \cap \mathbf{B})}{p(\mathbf{B})} = \frac{1}{3} \cdot \frac{3}{2} = \frac{1}{2}.$$

Esempio 2.5.7

Esperimento: estrazione di due numeri tra i numeri 1, 2 e 3 .
$\mathbf{\Omega} := \{(1, 2), (1, 3),(2, 1),(2, 3),(3, 1),(3, 2)\}$
Supponiamo che gli eventi elementari siano equiprobabili; pertanto, ciascuno di essi ha probabilità $\frac{1}{6}$.
$\mathbf{A} :=$ <viene estratto il 2> ; $\mathbf{B} :=$ <viene estratto l'1> .

$$p(\mathbf{A}) = \frac{4}{6} = \frac{2}{3}, \quad p(\mathbf{B}) = \frac{4}{6} = \frac{2}{3}, \quad p(\mathbf{A} \cap \mathbf{B}) = \frac{2}{6} = \frac{1}{3}$$

cosicché $$p(\mathbf{A}|\mathbf{B}) = \frac{p(\mathbf{A} \cap \mathbf{B})}{p(\mathbf{B})} = \frac{1}{3} \cdot \frac{3}{2} = \frac{1}{2}.$$

Osservazione 2.5.8

I risultati ottenuti negli esempi 2.5.6 e 2.5.7 presentano un aspetto paradossale. Essi mostrano che la probabilità di essere estratto per il numero 2 è influenzata dal fatto che venga estratto o meno il numero 3; ed è influenzata (esattamente allo stesso modo) dal fatto che venga estratto o meno il numero 1. Precisamente, sia che venga estratto il numero 3 sia che venga estratto il numero 1, la probabilità che venga estratto il numero 2 è $\frac{1}{2}$; e certamente viene estratto uno dei numeri 1 e 3 . Ma, in assenza di informazioni, la probabilità che venga estratto il numero 2 è $\frac{2}{3}$.

Esempio 2.5.9

Un altro paradosso probabilistico analogo a quello evidenziato con gli esempi 2.5.6 e 2.5.7 .
Esperimento: lancio di due monete $\mathcal{M}_1$ (da 1 Euro) e $\mathcal{M}_2$ (da 2 Euro) .
$\mathbf{\Omega} := \{$(T, T), (T, C),(C, T),(C, C)$\}$
Supponiamo che gli eventi elementari siano equiprobabili; pertanto, ciascuno di essi ha probabilità $\frac{1}{4}$.
$\mathbf{A}$:= <su entrambe le monete esce testa> ; $\mathbf{B}$:= <su almeno una moneta esce testa> ;
$\mathbf{B_1}$:= <su $\mathcal{M}_1$ esce testa> ; $\mathbf{B_2}$:= <su $\mathcal{M}_2$ esce testa> ;

$$p(\mathbf{A}) = \tfrac{1}{4}\,, \quad p(\mathbf{B}) = \tfrac{3}{4}\,, \quad p(\mathbf{B_1}) = \tfrac{1}{2}\,, \quad p(\mathbf{B_2}) = \tfrac{1}{2}\,,$$
$$p(\mathbf{A}\cap\mathbf{B}) = p(\mathbf{A}\cap\mathbf{B_1}) = p(\mathbf{A}\cap\mathbf{B_2}) = p(\mathbf{A}) = \tfrac{1}{4}\,;$$

cosicché
$$p(\mathbf{A}|\mathbf{B}) = \frac{p(\mathbf{A}\cap\mathbf{B})}{p(\mathbf{B})} = \frac{1}{4}\cdot\frac{4}{3} = \frac{1}{3}\,;$$
$$p(\mathbf{A}|\mathbf{B_1}) = \frac{p(\mathbf{A}\cap\mathbf{B_1})}{p(\mathbf{B_1})} = \frac{1}{4}\cdot 2 = \frac{1}{2}\,; \qquad p(\mathbf{A}|\mathbf{B_2}) = \frac{p(\mathbf{A}\cap\mathbf{B_2})}{p(\mathbf{B_2})} = \frac{1}{4}\cdot 2 = \frac{1}{2}\,.$$

Se sappiamo che $\mathbf{B}$ è vero, la probabilità di $\mathbf{A}$ è $\frac{1}{3}$; d'altro lato, se $\mathbf{B}$ è vero, certamente è vero almeno uno degli eventi $\mathbf{B_1}$, $\mathbf{B_2}$: in ciascuno di questi due casi, la probabilità di $\mathbf{A}$ è però $\frac{1}{2}$.

Esempio 2.5.10

In una città vivono due amici, Tizio e Caio. Tizio dice a Caio che forse andrà a trovarlo con l'autobus; ci sono 6 corse utili, e Tizio adotterà il seguente metodo: deciderà se partire o meno lanciando una moneta; stabilirà quale dei 6 autobus prendere lanciando un dado.
Caio va ad aspettare Tizio alla fermata, ma con i primi 5 autobus Tizio non arriva. Qual è la probabilità che arrivi col sesto?
Sia $\mathbf{A}_0$ l'evento "Tizio non parte", e (per $i := 1, 2, \ldots, 6$) sia $\mathbf{A}_i$ l'evento "Tizio arriva con l'i – simo autobus". Possiamo valutare che sia $p(\mathbf{A}_0) = p\left(\bigcup_{i=1}^{n} \mathbf{A}_i\right)$ (stimando che la moneta non sia truccata!) e che gli $\mathbf{A}_i$ per $i := 1, 2, \ldots, 6$ siano tutti equiprobabili (stimando che il dado non sia truccato!): pertanto $p(\mathbf{A}_0) = \frac{1}{2}$ e $p(\mathbf{A}_i) = \frac{1}{12}$ per $i > 0$.
L'evento condizionante $\mathbf{B}$ è "Tizio non arriva con nessuno dei primi 5 autobus", cioè $\mathbf{B} := (\mathbf{A}_1 \cup \mathbf{A}_2 \cup \ldots \cup \mathbf{A}_5)^{\mathbf{C}}$; per le (b) e (c) del teorema 2.3.1 ,

$$p(\mathbf{B}) = 1 - \sum_{i=1}^{5} p(\mathbf{A}_i) = 1 - \frac{5}{12} = \frac{7}{12}\,.$$

Per la 2.5.F2, $$p(\mathbf{A}_6|\mathbf{B}) = \frac{p(\mathbf{A}_6 \cap \mathbf{B})}{p(\mathbf{B})} .$$

Si noti che $\mathbf{A}_6 \subseteq \mathbf{B}$, cosicché $\mathbf{A}_6 \cap \mathbf{B} = \mathbf{A}_6$ e dunque
$$p(\mathbf{A}_6|\mathbf{B}) = \frac{p(\mathbf{A}_6)}{p(\mathbf{B})} = \frac{1}{12} \cdot \frac{12}{7} = \frac{1}{7} .$$

La conoscenza del fatto che Tizio non è arrivato con nessuno dei primi 5 autobus ha fatto aumentare la probabilità dell'evento $\mathbf{A}_6$ da $\frac{1}{12}$ a $\frac{1}{7}$. Naturalmente, la probabilità dell'evento $\mathbf{A}_0^{\mathbf{C}}$ (= "Tizio arriva") è invece *diminuita* da $\frac{1}{2}$ a $\frac{1}{7}$.

Esercizio 2.5.11

Abbiamo due dadi. Ne tiriamo uno: se è uscito un numero dispari, tiriamo anche l'altro e sommiamo i risultati ottenuti. Si dica qual è la probabilità che abbiamo tirato entrambi i dadi
– se complessivamente realizziamo un "5";
– se complessivamente realizziamo un "6";
– se complessivamente realizziamo un "7".

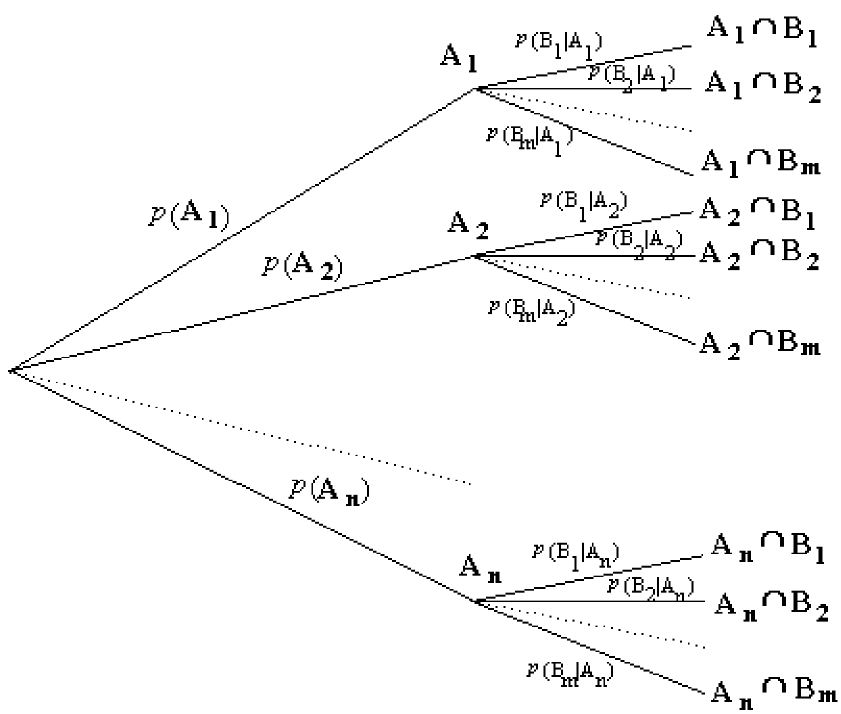
Esercizio 2.5.12

Dicei persone lanciano, successivamente, una stessa moneta non truccata. Se le prime nove ottengono "Testa", qual è la probabilità che anche la decima ottenga "Testa"?

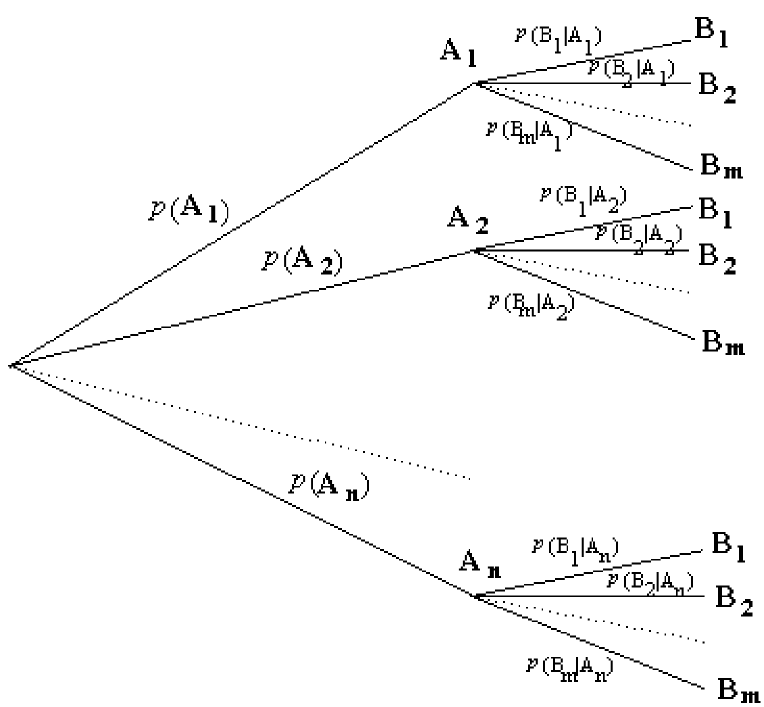
Osservazione 2.5.13

Sia $\mathbf{\Omega}$ un insieme non vuoto, e sia p una misura di probabilità su $\mathbf{\Omega}$. Siano $\mathbf{A_1}, \mathbf{A_2}, \dots, \mathbf{A_n} \in \mathcal{D}(p)$ eventi tali che $\{\mathbf{A_1}, \mathbf{A_2}, \dots, \mathbf{A_n}\}$ sia una partizione di $\mathbf{\Omega}$ (cfr. [1, 3.7]); e siano $\mathbf{B_1}, \mathbf{B_2}, \dots, \mathbf{B_m} \in \mathcal{D}(p)$ eventi con probabilità non nulla tali che $\{\mathbf{B_1}, \mathbf{B_2}, \dots, \mathbf{B_m}\}$ sia una partizione di $\mathbf{\Omega}$.
Per la formula 2.5.F1 (che si può facilmente dedurre dalla definizione di probabilità condizionata 2.5.F2) si ha
$$p(\mathbf{A_i} \cap \mathbf{B_j}) = p(\mathbf{B_j})p(\mathbf{A_i}|\mathbf{B_j}) \qquad \text{per } i := 1, 2, \dots, n .$$

Per visualizzare le $p(\mathbf{A_i} \cap \mathbf{B_j})$ è spesso comodo utilizzare un "diagramma ad albero" come il seguente

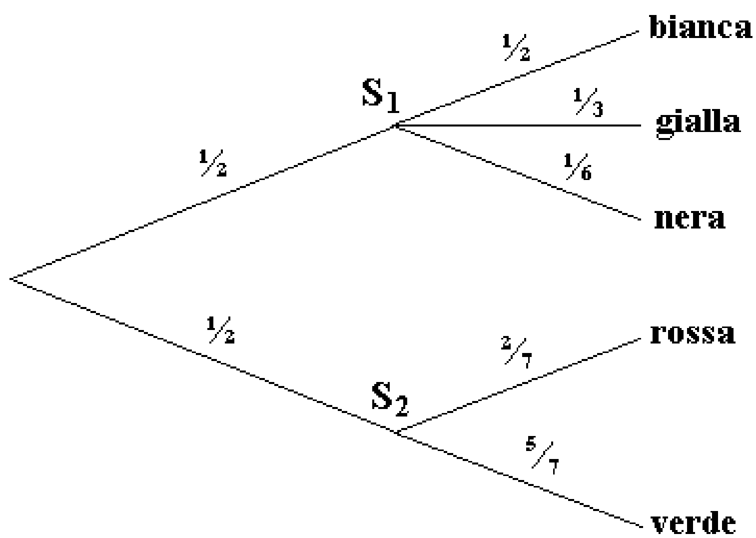


o anche, quando ciò non dia luogo ad ambiguità, come il seguente:



Esempio 2.5.14

Sono date due scatole: una contiene tre palline bianche, due gialle e una nera; l'altra contiene 2 palline rosse e 5 verdi.
Si sceglie a caso una scatola, e da essa si estrae una pallina: qual è la probabilità che la pallina estratta sia rossa?
Rappresentando la situazione con un diagramma ad albero



si trova che

$$p(\mathbf{rossa}) = \tfrac{1}{2} \cdot \tfrac{2}{7} = \tfrac{1}{7}\,.$$

Esercizio 2.5.15

Una scatola contiene 3 palline rosse e 7 palline bianche. Si estrae una pallina dalla scatola; si introduce nella scatola una pallina dell'altro colore; si estrae nuovamente una pallina dalla scatola.
Determinare la probabilità che entrambe le palline estratte siano rosse.

Esercizio 2.5.16

In un borsellino ci sono tre monete da 1 Euro. La prima è una moneta normale; la seconda ha "testa" su entrambe le facce; la terza è stata modificata in modo che, lanciandola, la probabilità di ottenere "testa" sia di $\frac{2}{3}$.
Si prende a caso una moneta dal borsellino e la si lancia. Determinare la probabilità che esca "testa".

2.6 - Indipendenza stocastica.

Sia $\mathbf{\Omega}$ un insieme non vuoto, e sia p una misura di probabilità su $\mathbf{\Omega}$; siano $\mathbf{A}$, $\mathbf{B} \in \mathcal{D}(p)$ eventi con probabilità non nulla.

Se $p(\mathbf{A}|\mathbf{B}) > p(\mathbf{A})$, si dice che l'evento $\mathbf{B}$ *favorisce* l'evento $\mathbf{A}$; se $p(\mathbf{A}|\mathbf{B}) < p(\mathbf{A})$, si dice che l'evento $\mathbf{B}$ *ostacola* l'evento $\mathbf{A}$; se $p(\mathbf{A}|\mathbf{B}) = p(\mathbf{A})$, si dice che l'evento $\mathbf{A}$ è (*stocasticamente*) *indipendente* dall'evento $\mathbf{B}$.

Teorema 2.6.1

Siano $\mathbf{\Omega}$ un insieme non vuoto e p una misura di probabilità su $\mathbf{\Omega}$; siano $\mathbf{A}, \mathbf{B} \in \mathcal{D}(p)$ eventi con probabilità non nulla. Sono fatti equivalenti:

(a) $\mathbf{A}$ è indipendente da $\mathbf{B}$;
(b) $\mathbf{B}$ è indipendente da $\mathbf{A}$;
(c) $p(\mathbf{A} \cap \mathbf{B}) = p(\mathbf{A}) \cdot p(\mathbf{B})$.

Dimostrazione - Se $p(\mathbf{A}|\mathbf{B}) = p(\mathbf{A})$, la 2.5.F1 diventa la (c). Viceversa, se vale la (c), dalla 2.5.F1 si deduce che $p(\mathbf{A}|\mathbf{B}) = p(\mathbf{A})$. Dunque $(a) \Leftrightarrow (c)$.
Ragionando allo stesso modo sulla $p(\mathbf{A} \cap \mathbf{B}) = p(\mathbf{B}|\mathbf{A}) \cdot p(\mathbf{A})$
(che si ottiene dalla 2.5.F1 scambiando $\mathbf{A}$ con $\mathbf{B}$) si ottiene che $(b) \Leftrightarrow (c)$, da cui infine $(a) \Leftrightarrow (b)$.

Teorema 2.6.2

Siano $\mathbf{\Omega}$ un insieme non vuoto e p una misura di probabilità su $\mathbf{\Omega}$; siano $\mathbf{A}, \mathbf{B} \in \mathcal{D}(p)$ eventi con probabilità non nulla. Sono fatti equivalenti:

(a) $\mathbf{A}$ favorisce $\mathbf{B}$;
(b) $\mathbf{B}$ favorisce $\mathbf{A}$;
(c) $p(\mathbf{A} \cap \mathbf{B}) > p(\mathbf{A}) \cdot p(\mathbf{B})$.

Dimostrazione - Proviamo innanzitutto che $(b) \Leftrightarrow (c)$.
$(b) \Rightarrow (c)$. Sia $p(\mathbf{A}|\mathbf{B}) > p(\mathbf{A})$. Allora

$$p(\mathbf{A}|\mathbf{B}) \cdot p(\mathbf{B}) > p(\mathbf{A}) \cdot p(\mathbf{B})$$

(essendo $p(\mathbf{B}) > 0$) e dunque (tenuto conto della 2.5.F1) la (c).
$(c) \Rightarrow (b)$. Per la 2.5.F1 si ha $\quad p(\mathbf{A}|\mathbf{B}) \cdot p(\mathbf{B}) > p(\mathbf{A}) \cdot p(\mathbf{B})$
e quindi (essendo $p(\mathbf{B}) > 0$) $\quad p(\mathbf{A}|\mathbf{B}) > p(\mathbf{A}) \quad$ ossia la (b).
Ragionando allo stesso modo con la $p(\mathbf{A} \cap \mathbf{B}) = p(\mathbf{B}|\mathbf{A}) \cdot p(\mathbf{A})$
(che si ottiene dalla 2.5.F1 scambiando $\mathbf{A}$ con $\mathbf{B}$) si ottiene che $(b) \Leftrightarrow (c)$, da cui infine $(a) \Leftrightarrow (b)$.

Osservazione 2.6.3

Il teorema 2.6.2 evidenzia il fatto che l'aumentare della probabilità di un evento al verificarsi di un altro non può essere "automaticamente" interpretata come indicazione di causalità (si veda l'esempio 2.6.8 più avanti).

Esercizio 2.6.4

Consideriamo il seguente esperimento: scegliere a caso una famiglia italiana con tre figli.
Sia **A** l'evento "ci sono figli di entrambi i sessi" e sia **B** l'evento "c'è al più una figlia femmina".
L'evento **B** favorisce od ostacola l'evento **A**? Oppure i due eventi sono stocasticamente indipendenti?
Si valuti $\frac{1}{2}$ per ciascun figlio la probabilità che sia femmina.

Esercizio 2.6.5

Per ciascuno dei seguenti esperimenti:
- scegliere a caso una famiglia italiana con due figli;
- scegliere a caso una famiglia italiana con quattro figli;

Si definiscano gli eventi **A** e **B** come nell'esercizio 2.6.4 e si stabilisca se l'evento **B** favorisce, ostacola o è stocasticamente indipendente dall'evento **A**.
Si valuti $\frac{1}{2}$ per ciascun figlio la probabilità che sia femmina.

Osservazione 2.6.6

Valutando la probabilità degli eventi col criterio frequentista, si può decidere sulla base di dati sperimentali se due eventi sono o non sono indipendenti; si possono così ad esempio ottenere informazioni sulle relazioni fra fenomeni di interesse clinico epidemiologico (fumo e tumori, colesterolo e infarto, centrali nucleari e tumori).

Esempio 2.6.7

Le percentuali di sordi e daltonici su una certa popolazione risultano essere come segue:
daltonici: 0,8%; *sordi*: 0,05%; *daltonici sordi*: 0,0004%.
Accettando la frequenza relativa come stima di probabilità, possiamo affermare che (posto **S** := *sordo* e **D** := *daltonico*)

$$p(\mathbf{S}) = \frac{5}{10.000} = \frac{5}{10^4}\,; \qquad p(\mathbf{D}) = \frac{8}{1.000} = \frac{8}{10^3}\,; \qquad p(\mathbf{S}\cap\mathbf{D}) = \frac{4}{1.000.000} = \frac{4}{10^6}\,.$$

Poiché $p(\mathbf{S}) \cdot p(\mathbf{D}) = \frac{5 \cdot 8}{10^7} = \frac{4}{10^6} = p(\mathbf{S}\cap\mathbf{D})$, possiamo abdurre che sordità e daltonismo sono patologie indipendenti.

Esempio 2.6.8

Le percentuali di fumatori e di malati di cancro su una certa popolazione risultano essere come segue:
fumatori: 55% ; *malati di cancro*: 47% ; *fumatori malati di cancro*: 38% .
Accettando la frequenza relativa come stima di probabilità, possiamo affermare che (posto $\mathbf{F} :=$ *fumatore* e $\mathbf{M} :=$ *malato di cancro*)

$$p(\mathbf{F}) = \tfrac{55}{100}\,; \qquad p(\mathbf{M}) = \tfrac{47}{100}\,; \qquad p(\mathbf{F}\cap\mathbf{M}) = \tfrac{38}{100}\,.$$

Poiché $p(\mathbf{F})\cdot p(\mathbf{M}) = \frac{55\cdot 47}{10.000} = \frac{2.585}{10^4} < p(\mathbf{F}\cap\mathbf{M})$, possiamo abdurre che i due fenomeni <u>non</u> sono indipendenti. In effetti,

$$p(\mathbf{M}|\mathbf{F}) = \tfrac{p(\mathbf{M}\cap\mathbf{F})}{p(\mathbf{F})} = \tfrac{38}{100}\cdot\tfrac{100}{55} = \tfrac{38}{55} > \tfrac{38}{100} = p(\mathbf{M})\,.$$

È, naturalmente, personale l'interpretazione di questo risultato: se si debba cioè ritenere che fumare provochi il cancro oppure che viceversa essere malati di cancro induca l'abitudine a fumare.

Esempio 2.6.9

I tre esploratori Livingston, Stanley e Davidson erano stati catturati dai feroci M'Bangi che li avevano sorpresi nel loro territorio. Condannati a morte tutti e tre, erano riusciti a strappare la seguente concessione: il capo M'Bangi avrebbe scritto su quattro pelli rozzamente conciate le lettere L, S, D e J; poi lo stregone della tribù avrebbe estratto a sorte una delle quattro pelli. La pelle contrassegnata con "L" avrebbe dato la salvezza a Livingston, quella con "S" a Stanley, quella con "D" a Davidson, quella con "J" a tutti e tre.

Consideriamo i seguenti eventi:
$\mathbf{L} :=$ <Livingston si salva> ;
$\mathbf{S} :=$ <Stanley si salva> ;
$\mathbf{D} :=$ <Davidson si salva> .
Essi sono a due a due indipendenti.

Verifichiamo ad esempio che $p(\mathbf{L}\cap\mathbf{S}) = p(\mathbf{L})\cdot p(\mathbf{S})$. Possiamo descrivere l'esperimento di estrazione a sorte della pelle mediante i quattro risultati equiprobabili
$x_{\mathrm{L}} :=$ <viene estratta la pelle contrassegnata con la lettera "L"> ;
$x_{\mathrm{S}} :=$ <viene estratta la pelle contrassegnata con la lettera "S"> ;
$x_{\mathrm{D}} :=$ <viene estratta la pelle contrassegnata con la lettera "D"> ;
$x_{\mathrm{J}} :=$ <viene estratta la pelle contrassegnata con la lettera "J"> .
Ciascuno dei corrispondenti eventi elementari ha probabilità $\frac{1}{4}$. Inoltre

$$\mathbf{L} = \{x_{\mathrm{L}}\,,x_{\mathrm{J}}\}\,; \quad \mathbf{S} = \{x_{\mathrm{S}}\,,x_{\mathrm{J}}\}\,; \quad \mathbf{L}\cap\mathbf{S} = \{x_{\mathrm{J}}\}$$

e

$$p(\mathbf{L}) = \tfrac{1}{2}\,; \quad p(\mathbf{S}) = \tfrac{1}{2}\,; \quad p(\mathbf{L}\cap\mathbf{S}) = \tfrac{1}{4}$$

cosicché

$$p(\mathbf{L}\cap\mathbf{S}) = \tfrac{1}{4} = \tfrac{1}{2}\cdot\tfrac{1}{2} = p(\mathbf{L})\cdot p(\mathbf{S})\,.$$

Dunque gli eventi **L** , **S** e **D** sono a due a due indipendenti. Si noti però che

$$p(\mathbf{L}\cap\mathbf{S}\cap\mathbf{D}) = p(\{x_1\}) = \tfrac{1}{4} \neq \tfrac{1}{8} = p(\mathbf{L})\cdot p(\mathbf{S})\cdot p(\mathbf{D})\,.$$

2.7 - Il teorema di Bayes.

Sia $\mathbf{\Omega}$ un insieme non vuoto e sia p una misura di probabilità su $\mathbf{\Omega}$; siano $\mathbf{A}, \mathbf{B} \in \mathcal{D}(p)$ eventi con probabilità non nulla.

Abbiamo visto (teorema 2.6.2) che $\mathbf{B}$ favorisce $\mathbf{A}$ se e soltanto se $\mathbf{A}$ favorisce $\mathbf{B}$. In effetti, l'influenza di $\mathbf{B}$ su $\mathbf{A}$ (intesa come $\frac{p(\mathbf{A}|\mathbf{B})}{p(\mathbf{A})}$) è proprio uguale all'influenza di $\mathbf{A}$ su $\mathbf{B}$ (intesa come $\frac{p(\mathbf{B}|\mathbf{A})}{p(\mathbf{B})}$):

Teorema 2.7.1

Sia $\mathbf{\Omega}$ un insieme non vuoto e sia p una misura di probabilità su $\mathbf{\Omega}$; siano $\mathbf{A}, \mathbf{B} \in \mathcal{D}(p)$ eventi con probabilità non nulla. Allora

$$\frac{p(\mathbf{A}|\mathbf{B})}{p(\mathbf{A})} = \frac{p(\mathbf{B}|\mathbf{A})}{p(\mathbf{B})}$$

Dimostrazione - Per la 2.5.F1 si ha che

$$p(\mathbf{A} \cap \mathbf{B}) = p(\mathbf{A}|\mathbf{B}) \cdot p(\mathbf{B})$$

ma anche (scambiando $\mathbf{A}$ con $\mathbf{B}$) che

$$p(\mathbf{A} \cap \mathbf{B}) = p(\mathbf{B}|\mathbf{A}) \cdot p(\mathbf{A}).$$

Dunque

$$p(\mathbf{A}|\mathbf{B}) \cdot p(\mathbf{B}) = p(\mathbf{B}|\mathbf{A}) \cdot p(\mathbf{A})$$

da cui (dividendo ambo i membri per $p(\mathbf{A}) \cdot p(\mathbf{B})$, che per ipotesi è diverso da zero) l'asserto.

Il teorema 2.7.1 consente di calcolare $p(\mathbf{A}|\mathbf{B})$ conoscendo $p(\mathbf{A})$, $p(\mathbf{B})$ e $p(\mathbf{B}|\mathbf{A})$. Spesso la probabilità $p(\mathbf{B})$ dell'evento condizionante non è nota; se però conosciamo oltre a $p(\mathbf{A})$ e $p(\mathbf{B}|\mathbf{A})$ anche $p(\mathbf{A}^{\mathbf{C}})$ e $p(\mathbf{B}|\mathbf{A}^{\mathbf{C}})$, possiamo calcolare $p(\mathbf{B})$. Più in generale:

Teorema 2.7.2

Sia $\mathbf{\Omega}$ un insieme non vuoto, e sia p una misura di probabilità su $\mathbf{\Omega}$. Siano $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$, $\mathbf{B} \in \mathcal{D}(p)$ eventi con probabilità non nulla tali che $\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n\}$ sia una partizione di $\mathbf{\Omega}$ (cfr. 3.7). Allora

$$p(\mathbf{B}) = \sum_{i=1}^{n} p(\mathbf{A}_i) p(\mathbf{B}|\mathbf{A}_i).$$

Dimostrazione - Applicando la 2.5.F1 alla probabilità di $\mathbf{B}$ condizionata a ciascun $\mathbf{A}_i$, si ottiene che

$$p(\mathbf{A}_i \cap \mathbf{B}) = p(\mathbf{A}_i) p(\mathbf{B}|\mathbf{A}_i)$$

e dunque, sommando membro a membro per $i := 1, 2, \ldots, n$,

$$(\bigstar) \qquad \sum_{i=1}^{n} p(\mathbf{A}_i \cap \mathbf{B}) = \sum_{i=1}^{n} p(\mathbf{A}_i) p(\mathbf{B}|\mathbf{A}_i).$$

Valutiamo il primo membro: poiché gli $\mathbf{A}_i \cap \mathbf{B}$ sono a due a due disgiunti (essendo tali gli $\mathbf{A}_i$), per la (b) del teorema 2.3.1 si ha che

$$\sum_{i=1}^{n} p(\mathbf{A}_i \cap \mathbf{B}) = p\left(\bigcup_{i=1}^{n} (\mathbf{A}_i \cap \mathbf{B})\right) = p(\left(\bigcup_{i=1}^{n} \mathbf{A}_i\right) \cap \mathbf{B}) = p(\mathbf{\Omega} \cap \mathbf{B}) = p(\mathbf{B})$$

come si voleva.

Dai teoremi 2.7.1 e 2.7.2 si ottiene immediatamente il

Teorema 2.7.3 (*di Bayes*)

Sia $\mathbf{\Omega}$ un insieme non vuoto, e sia p una misura di probabilità su $\mathbf{\Omega}$. Siano $\mathbf{A_1}, \mathbf{A_2}, \ldots, \mathbf{A_n}$, $\mathbf{B} \in \mathcal{D}(p)$ eventi con probabilità non nulla tali che $\{\mathbf{A_1}, \mathbf{A_2}, \ldots, \mathbf{A_n}\}$ sia una partizione di $\mathbf{\Omega}$ (cfr. [1, 3.7]). Allora

$$p(\mathbf{A}_i|\mathbf{B}) = \frac{p(\mathbf{A}_i)p(\mathbf{B}|\mathbf{A}_i)}{\sum_{i=1}^{n} p(\mathbf{A}_i)p(\mathbf{B}|\mathbf{A}_i)} \qquad \text{per } i := 1, 2, \ldots, n\,.$$

Dimostrazione - Dal teorema 2.7.1 si ricava che

$$p(\mathbf{A}_i|\mathbf{B}) = \frac{p(\mathbf{A}_i)p(\mathbf{B}|\mathbf{A}_i)}{p(\mathbf{B})} \qquad \text{per } i := 1, 2, \ldots, n\,.$$

Sostituendo il valore di $p(\mathbf{B})$ espresso dal teorema 2.7.2 si ha l'asserto.

Esempio 2.7.4

Abbiamo quattro scatole S_1, S_2, S_3, S_4 come segue:
– la scatola S_1 contiene: 7 palline bianche e 3 palline nere;
– la scatola S_2 contiene: 8 palline bianche e 7 palline nere;
– la scatola S_3 contiene: 1 pallina bianca e 5 palline nere;
– la scatola S_4 contiene: 5 palline bianche e 7 palline nere.

Viene scelta caso una scatola e ne è estratta a caso una pallina. Se la pallina è bianca, qual è la probabilità che sia stata scelta la prima scatola?

Sia $\mathbf{A}_i$ l'evento "è stata estratta una pallina dalla scatola S_i"; per ipotesi, gli eventi $\mathbf{A}_i$ hanno tutti la stessa probabilità (abbiamo detto: <viene scelta *a caso* una scatola...>) e dunque $p(\mathbf{A}_i) = \frac{1}{4}$ per $i := 1, 2, 3, 4$.

L'evento condizionante **B** è "è stata estratta una pallina bianca".

Per il teorema 2.7.1, $\qquad p(\mathbf{A}_1|\mathbf{B}) = \dfrac{p(\mathbf{A}_1)\cdot p(\mathbf{B}|\mathbf{A}_1)}{p(\mathbf{B})}$.

Notiamo che non è facile calcolare $p(\mathbf{B})$; mentre è immediato valutare $p(\mathbf{B}|\mathbf{A}_i)$: si ha

$$p(\mathbf{B}|\mathbf{A}_1) = \tfrac{7}{10}\,; \quad p(\mathbf{B}|\mathbf{A}_2) = \tfrac{8}{15}\,; \quad p(\mathbf{B}|\mathbf{A}_3) = \tfrac{1}{6}\,; \quad p(\mathbf{B}|\mathbf{A}_4) = \tfrac{5}{12}\,.$$

Per il teorema 2.7.2, si ha

$$p(\mathbf{B}) = \sum_{i=1}^{4} p(\mathbf{A}_i)p(\mathbf{B}|\mathbf{A}_i) = \tfrac{1}{4}\cdot\tfrac{7}{10} + \tfrac{1}{4}\cdot\tfrac{8}{15} + \tfrac{1}{4}\cdot\tfrac{1}{6} + \tfrac{1}{4}\cdot\tfrac{5}{12} = \tfrac{109}{240}$$

e dunque

$$p(\mathbf{A}_1|\mathbf{B}) = \frac{\frac{1}{4}\cdot\frac{7}{10}}{\frac{109}{240}} = \frac{42}{109} \approx 0{,}385\,.$$

Esempio 2.7.5

La produzione di wafer della Brutopia è affidata a due fabbriche F_1, F_2. La F_1 copre il 60% della produzione nazionale, la F_2 solo il 40%; si sa inoltre che i due diversi tipi di wafer sono così distribuiti:
– F_1 : 80% al cacao, 20% alla nocciola;
– F_2 : 60% al cacao, 40% alla nocciola.
Ci viene offerto un wafer al cacao prodotto in Brutopia. Qual è la probabilità che sia stato prodotto nella fabbrica F_1 ?
Sia $\mathbf{A}_i$ l'evento "ci viene offerto un wafer prodotto in F_i"; per ipotesi, $p(\mathbf{A}_1) = \frac{60}{100} = \frac{3}{5}$ e $p(\mathbf{A}_2) = \frac{40}{100} = \frac{2}{5}$.
L'evento condizionante **B** è "ci viene offerto un wafer al cacao" .

Per il teorema 2.7.1 , $$p(\mathbf{A}_1|\mathbf{B}) = \frac{p(\mathbf{A}_1)\cdot p(\mathbf{B}|\mathbf{A}_1)}{p(\mathbf{B})} .$$

Notiamo che non è facile calcolare $p(\mathbf{B})$; mentre è immediato valutare $p(\mathbf{B}|\mathbf{A}_1)$ e $p(\mathbf{B}|\mathbf{A}_2)$: si ha
$$p(\mathbf{B}|\mathbf{A}_1) = \frac{4}{5} ; \quad p(\mathbf{B}|\mathbf{A}_2) = \frac{3}{5} .$$
Per il teorema 2.7.2, si ha
$$p(\mathbf{B}) = p(\mathbf{A}_1)p(\mathbf{B}|\mathbf{A}_1) + p(\mathbf{A}_2)p(\mathbf{B}|\mathbf{A}_2) = \frac{3}{5} \cdot \frac{4}{5} + \frac{2}{5} \cdot \frac{3}{5} = \frac{18}{25}$$
e dunque
$$p(\mathbf{A}_1|\mathbf{B}) = \frac{\frac{3}{5} \cdot \frac{4}{5}}{\frac{18}{25}} = \frac{2}{3} .$$

Esercizio 2.7.6

Sono date due monete: una non è truccata, l'altra ha due "Testa" . Prendiamo a caso una moneta e la lanciamo due volte; entrambe le volte viene "Testa". Qual è la probabilità che la moneta sia quella truccata?

Esercizio 2.7.7

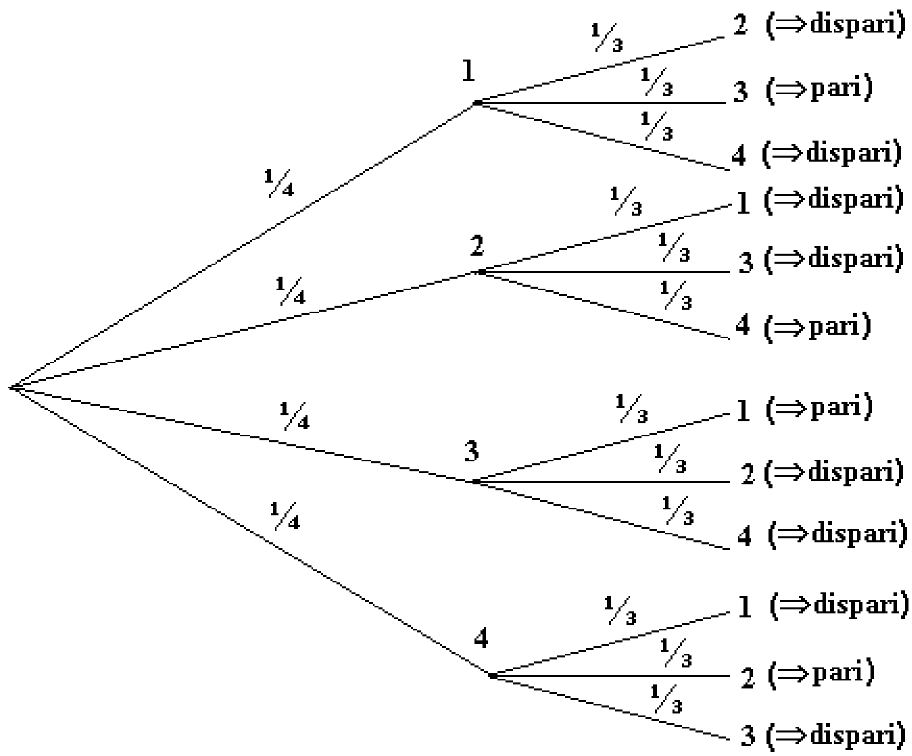
Sappiamo che in una partita di cento dadi ce n'è uno truccato, che presenta il "sei" su due o più facce (ma ignoriamo su quante). Scegliamo a caso un dado fra quei cento e lo lanciamo tre volte; esce tre volte "sei". Qual è la probabilità che sia il dado truccato? Qual è la probabilità che il dado abbia il "sei" su due facce? E su tutte le facce?

Osservazione 2.7.8

Le situazioni in cui trova applicazione il teorema 2.7.2 possono essere visualizzate con un "diagramma ad albero" (come si è già visto nell'Osservazione 2.5.13). La probabilità dell'evento **B** si ottiene come somma delle probabilità calcolate (con la "regola del prodotto", cioè con la 2.5.F1) lungo tutti i "cammini" che terminano in **B** (più correttamente: che terminano nei vari $\mathbf{A}_i \cap \mathbf{B}$) .

Esempio 2.7.9

Da un'urna contenente i numeri 1, 2, 3, 4 se ne estraggono successivamente due. Qual è la probabilità che la somma dei due numeri estratti sia pari?



I vari "cammini" che terminano con "**(⇒ pari)**" forniscono le probabilità che dobbiamo sommare per ottenere il risultato desiderato. La probabilità cercata è dunque

$$\frac{1}{4}\cdot\frac{1}{3}+\frac{1}{4}\cdot\frac{1}{3}+\frac{1}{4}\cdot\frac{1}{3}+\frac{1}{4}\cdot\frac{1}{3}=\frac{1}{3}\,.$$

<u>2.8 - Variabili aleatorie.</u>

Abbiamo finora scelto di descrivere gli effetti di un "esperimento" mediante un insieme di "risultati" i cui sottoinsiemi rappresentano gli "eventi". Molto spesso ci fa comodo associare un numero a ciascun risultato; ciò consente di evidenziare certi eventi particolarmente significativi per il nostro modo di considerare l'esperimento.

Esempio 2.8.1

Esperimento: lancio di due dadi. Spazio dei risultati: $\boldsymbol{\Omega} := \{(1,\ 1), (1,\ 2), (1,\ 3), \ldots, (6,\ 4), (6,\ 5), (6,\ 6)\}$. A ciascun risultato possiamo associare la somma dei punti espressi dai dadi (questo, in molti casi, è l'unico dato significativo).

Esempio 2.8.2

Stesso esperimento, stesso spazio dei risultati dell'esempio 2.8.1 . Questa volta però stiamo giocando a "Monòpoli" e siamo "finiti in prigione": il nostro interesse è ottenere "numeri doppi" per poter uscire di prigione. Associeremo ai risultati (1, 1), (2, 2), (3, 3), (4, 4), (5, 5) e (6, 6) il numero 1 (per dire: successo!) e a tutti gli altri risultati il numero 0 (per dire: insuccesso!).

Se all'esperimento che stiamo considerando è collegata una scommessa, ci può interessare associare a ogni risultato un numero che quantifichi il guadagno (o la perdita) stabilito per il verificarsi di tale risultato.

Esempio 2.8.3

Esperimento: lancio di un dado; spazio dei risultati $\mathbf{\Omega} := \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$. Scommessa associata: se viene "1", vinco 1 Euro; se viene "2", vinco 2 Euro; se viene "3", vinco 3 Euro; se viene "4", vinco 4 Euro; se viene "5", **perdo** 5 Euro; se viene "6", **perdo** 6 Euro.
Associeremo al risultato "1" il numero 1; al risultato "2" il numero 2; al risultato "3" il numero 3; al risultato "4" il numero 4; al risultato "5" il numero -5; al risultato "6" il numero -6 .

Sia $\mathbf{\Omega}$ un insieme non vuoto, e sia p una misura di probabilità su $\mathbf{\Omega}$. Si dice *variabile aleatoria* su $\mathbf{\Omega}$ ogni applicazione $\mathbf{X}$: $\mathbf{\Omega} \to \mathbb{R}$ tale che la retroimmagine di ogni intervallo appartenga a $\mathcal{D}(p)$.

La scelta di una variabile aleatoria determina un "certo modo" di guardare ai possibili risultati del fenomeno che studiamo; assumono quindi particolare significato quegli eventi che sono formati da tutti e soli i risultati a cui la variabile aleatoria associa uno stesso valore.

Sia $\mathbf{\Omega}$ un insieme non vuoto, sia p una misura di probabilità su $\mathbf{\Omega}$ e sia $\mathbf{X}$ una variabile aleatoria su $\mathbf{\Omega}$. Per ogni $x_0 \in \mathbb{R}$, si pone

$$\mathbf{E}_{x_0} := \{\omega \in \mathbf{\Omega} \,/\, \mathbf{X}(\omega) = x_0\} \,.$$

Poiché $\mathbf{X}$ è una variabile aleatoria, $\mathbf{E}_{x_0} \in \mathcal{D}(p)$ per ogni $x_0 \in \mathbb{R}$. Inoltre, gli $\mathbf{E}_{x_0}$ non vuoti costituiscono una partizione di $\mathbf{\Omega}$; si potrebbe scegliere come spazio dei risultati l'insieme degli $\mathbf{E}_{x_0}$ non vuoti, o addirittura l'insieme dei numeri reali x_0 per i quali $\mathbf{E}_{x_0} \neq \emptyset$.

La probabilità $p(\mathbf{E}_{x_0})$ dell'insieme $\mathbf{E}_{x_0}$ si indica di solito con la notazione $p(\mathbf{X} = x_0)$.

Per ogni x_0 , x_1 , $x_2 \in \mathbb{R}$, si considerano anche gli insiemi
$\{\omega \in \mathbf{\Omega} \,/\, \mathbf{X}(\omega) \le x_0\}$ la cui probabilità si indica con la notazione $p(\mathbf{X} \le x_0)$
e $\{\omega \in \mathbf{\Omega} \,/\, x_1 \le \mathbf{X}(\omega) \le x_2\}$ la cui probabilità si indica con la notazione $p(x_1 \le \mathbf{X} \le x_2)$.

Si dice *funzione di ripartizione* della variabile aleatoria $\mathbf{X}$ la funzione $\mathbb{R} \to [0,\ 1]$ che associa a ogni numero reale x la probabilità $p(\mathbf{X} \le x)$.

Osservazione 2.8.4

Se la variabile aleatoria **X** assume soltanto un numero finito di valori x_1, x_2, $\ldots$, x_s oppure un'infinità numerabile di valori x_1, x_2, $\ldots$, x_n, $\ldots$, il valore della funzione di ripartizione di **X** resta completamente determinato dalle probabilità $p(\mathbf{X} = x_i)$.

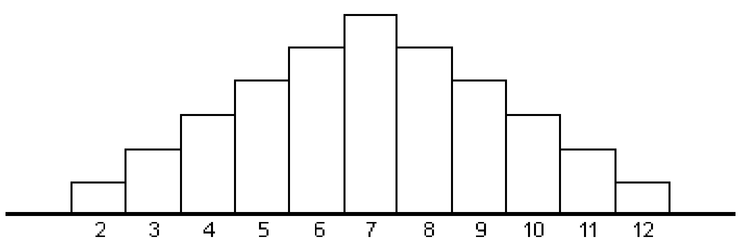
Vediamo come si può rappresentare la situazione negli esempi considerati in precedenza.

Esempio 2.8.1

Possiamo tabulare come segue le probabilità degli eventi $\mathbf{E}_{x_0}$:

x_0	2	3	4	5	6	7	8	9	10	11	12
$p(\mathbf{X} = x_0)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Possiamo anche disegnare un istogramma



nel quale l'area della colonna corrispondente a ciascun valore x_0 ([6]) di **X** è la probabilità $p(\mathbf{X} = x_0)$, mentre la funzione di ripartizione si può calcolare in x_0 misurando l'area complessiva delle colonne "fino a x_0 incluso".

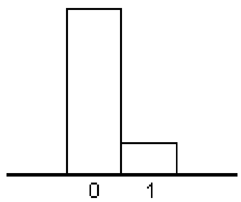
Esempio 2.8.2

Possiamo tabulare come segue le probabilità degli eventi $\mathbf{E}_{x_0}$:

x_0	0	1
$p(\mathbf{X} = x_0)$	$\frac{5}{6}$	$\frac{1}{6}$

[6] in questo caso x_0 assume tutti i valori interi da 2 a 12 .

Possiamo anche disegnare un istogramma



con le stesse convenzioni adottate per l'esempio 2.8.1 .

Esempio 2.8.3

Possiamo tabulare come segue le probabilità degli eventi $\mathbf{E}_{x_0}$:

x_0	1	2	3	4	-5	-6
$p(\mathbf{X} = x_0)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Possiamo anche disegnare un istogramma

con le stesse convenzioni adottate per gli esempi 2.8.1 e 2.8.2 .

2.9 - Speranza matematica.

Siano dati un esperimento, uno spazio dei risultati $\mathbf{\Omega}$ relativo ad esso, una misura di probabilità p su $\mathbf{\Omega}$ e una variabile aleatoria $\mathbf{X}$ su $\mathbf{\Omega}$.

Se ripetiamo n volte l'esperimento dato, la media dei valori che otteniamo per $\mathbf{X}$ è

$$m = \sum_{i=1}^{s} \frac{v_i x_i}{n}$$

dove v_i è il numero delle volte che si è verificato l'evento

$$\mathbf{E}_{x_i} := \{\omega \in \mathbf{\Omega} \,/\, \mathbf{X}(\omega) = x_i\}.$$

In altri termini,

$$m = \sum_{i=1}^{s} \frac{v_i}{n} x_i \qquad \text{dove } \frac{v_i}{n} \text{ è la frequenza relativa dell'evento } \mathbf{E}_{x_i} \text{ (cfr. 2.3)}.$$

Se vogliamo una stima "a priori" di tale media, possiamo sostituire alla frequenza relativa di $\mathbf{E}_{x_i}$ la probabilità $p(\mathbf{X} = x_i)$. Si giunge così alla seguente definizione:

Se **X** assume soltanto un numero finito di valori x_1 , x_2 , $\dots$, x_s , si dice *valore medio* o *speranza matematica* della variabile aleatoria **X** il numero

$$\mu_{\mathbf{X}} := p(\mathbf{X} = x_1) \cdot x_1 + p(\mathbf{X} = x_2) \cdot x_2 + \dots + p(\mathbf{X} = x_s) \cdot x_s = \sum_{i=1}^{s} p(\mathbf{X} = x_i) \cdot x_i \,.$$

Questa definizione può essere estesa al caso in cui **X** assume un'infinità numerabile di valori x_1 , x_2 , $\dots$, x_n , $\dots$, ponendo

$$\mu_{\mathbf{X}} := \lim_{n \to +\infty} \sum_{i=1}^{n} p(\mathbf{X} = x_i) \cdot x_i$$

ma solo se esiste ed è finito il

$$\lim_{n \to +\infty} \sum_{i=1}^{n} p(\mathbf{X} = x_i) \cdot |x_i| \,.$$

Tale condizione garantisce che il valore medio non dipenda dall'ordine con cui si considerano gli x_i.

Vediamo ora il valore medio della variabile aleatoria negli esempi visti in 2.8.

Esempio 2.9.1

Il valore medio della variabile aleatoria considerata nell'esempio 2.8.1 è

$$\mu_{\mathbf{X}} := \tfrac{1}{36} \cdot 2 + \tfrac{1}{18} \cdot 3 + \tfrac{1}{12} \cdot 4 + \tfrac{1}{9} \cdot 5 + \tfrac{5}{36} \cdot 6 + \tfrac{1}{6} \cdot 7 +$$
$$+ \tfrac{5}{36} \cdot 8 + \tfrac{1}{9} \cdot 9 + \tfrac{1}{12} \cdot 10 + \tfrac{1}{18} \cdot 11 + \tfrac{1}{36} \cdot 12 = \tfrac{252}{36} = 7$$

quindi (effettuando un "alto" numero di lanci) ci si può aspettare di avanzare "in media" di 7 caselle per ogni lancio dei dadi.

Esempio 2.9.2

Il valore medio della variabile aleatoria considerata nell'esempio 2.8.2 è

$$\mu_{\mathbf{X}} := \tfrac{1}{6} \cdot 1 + \tfrac{5}{6} \cdot 0 = \tfrac{1}{6}$$

quindi (effettuando un "alto" numero di lanci) ci si può aspettare "in media" <un sesto di successo> per ogni lancio dei dadi. Naturalmente, <un sesto di successo> non ha senso; intuitivamente, si potrebbe pensare che ciò equivalga a dire: otterremo "in media" un successo ogni sei lanci di dadi. Si veda l'esempio 2.13.3.

Esempio 2.9.3

Il valore medio della variabile aleatoria considerata nell'esempio 2.8.3 è

$$\mu_{\mathbf{X}} := \frac{1}{6}\cdot 1+\frac{1}{6}\cdot 2+\frac{1}{6}\cdot 3+\frac{1}{6}\cdot 4+\frac{1}{6}\cdot(-5)+\frac{1}{6}\cdot(-6) = -\frac{1}{6} = -0{,}1(6)$$

quindi (effettuando un "alto" numero di lanci) ci si può aspettare di perdere poco più di 0,166 Euro (circa Lit 300) a lancio.

Teorema 2.9.4

Siano $\mathbf{\Omega}$ un insieme non vuoto, p una misura di probabilità su $\mathbf{\Omega}$ e $\mathbf{X}$ una variabile aleatoria su $\mathbf{\Omega}$. Sia $\mu_{\mathbf{X}}$ il valore medio di $\mathbf{X}$. Siano a, b numeri reali con $b \neq 0$.
Posto $$\mathbf{Y}(\omega) := a\mathbf{X}(\omega)+b \qquad \forall \omega \in \mathbf{\Omega}$$
$\mathbf{Y}$ è una variabile aleatoria su $\mathbf{\Omega}$ con valore medio $a\mu_{\mathbf{X}}+b$.

Dimostrazione - Dimostriamo il teorema nel caso particolare in cui $\mathbf{X}$ assume un numero finito di valori x_1, x_2, …, x_s (e quindi $\mathbf{Y}$ assume i valori ax_1+b, ax_2+b, …, ax_s+b). Si ha $p(\mathbf{Y}=ax_i+b)=p(\mathbf{X}=x_i)$; il valore medio di $\mathbf{Y}$ è dunque

$$\sum_{i=1}^{s} p(\mathbf{Y}=ax_i+b)\cdot(ax_i+b) = \sum_{i=1}^{s} p(\mathbf{X}=x_i)\cdot(ax_i+b) =$$

$$= \sum_{i=1}^{s} p(\mathbf{X}=x_i)\cdot ax_i + \sum_{i=1}^{s} p(\mathbf{X}=x_i)\cdot b = a\cdot \underbrace{\left(\sum_{i=1}^{s} p(\mathbf{X}=x_i)\cdot x_i\right)}_{=\mu_{\mathbf{X}}} - b\cdot \underbrace{\left(\sum_{i=1}^{s} p(\mathbf{X}=x_i)\right)}_{=1} = a\mu_{\mathbf{X}}+b\,.$$

Una variabile aleatoria si dice *centrata* se il suo valore medio è zero.

Abbiamo visto (esempio 2.8.3) come una variabile aleatoria può quantificare l'esito di una scommessa. Intuitivamente, una scommessa è equa se la corrispondente variabile aleatoria è centrata: infatti in tal caso ("a lungo andare") non si vince né si perde, in media, niente.

Corollario 2.9.5

Siano $\mathbf{\Omega}$ un insieme non vuoto, p una misura di probabilità su $\mathbf{\Omega}$ e $\mathbf{X}$ una variabile aleatoria su $\mathbf{\Omega}$. Sia $\mu_{\mathbf{X}}$ il valore medio di $\mathbf{X}$.
Posto $$\mathbf{Y}(\omega) := \mathbf{X}(\omega)-\mu_{\mathbf{X}} \qquad \forall \omega \in \mathbf{\Omega}$$
$\mathbf{Y}$ è una variabile aleatoria centrata su $\mathbf{\Omega}$.

Dimostrazione - Si applichi il teorema 2.9.4 con $a := 1$ e $b := -\mu_{\mathbf{X}}$.

2.10 - Varianza e deviazione standard.

Siano dati un esperimento, uno spazio dei risultati $\mathbf{\Omega}$ relativo ad esso ed una misura di probabilità p su $\mathbf{\Omega}$.

Ad ogni variabile aleatoria $\mathbf{X}$ su $\mathbf{\Omega}$ con valore medio $\mu_{\mathbf{X}}$, possiamo associare un'altra variabile aleatoria $(\mathbf{X}-\mu_{\mathbf{X}})^2$ detta *deviazione quadratica* di $\mathbf{X}$: ad ogni risultato $\omega \in \mathbf{\Omega}$, la deviazione quadratica associa il quadrato dello "scostamento dal valore medio" di $\mathbf{X}(\omega)$.

Sia $\mathbf{X}$ una variabile aleatoria su $\mathbf{\Omega}$. Il valore medio della deviazione quadratica di $\mathbf{X}$ si indica con $\mathrm{Var}(\mathbf{X})$ e si dice *varianza* di $\mathbf{X}$; la radice quadrata della varianza di $\mathbf{X}$ si indica con $\sigma_{\mathbf{X}}$ e si dice *deviazione standard* (o *scarto quadratico medio*) di $\mathbf{X}$. Dunque, se $\mathbf{X}$ assume soltanto i valori $x_1, x_2, \ldots, x_s$:

$$\mathrm{Var}(\mathbf{X}) := \sum_{i=1}^{s} p(\mathbf{X}=x_i)\cdot(x_i-\mu_{\mathbf{X}})^2 \,; \qquad \sigma_{\mathbf{X}} := \sqrt{\sum_{i=1}^{s} p(\mathbf{X}=x_i)\cdot(x_i-\mu_{\mathbf{X}})^2} \,.$$

$\boxed{\text{Osservazione 2.10.1}}$

Per valutare di quanto una variabile aleatoria si discosta dal proprio valore medio, potremmo pensare di associare ad ogni evento $\mathbf{E}_x$ lo "scostamento dalla media" (il termine tecnico è "deviazione") $x_i - x$, e calcolare poi la media di tali scostamenti; ma ciò equivarrebbe a considerare il valore medio della variabile aleatoria $\mathbf{X}-\mu_{\mathbf{X}}$, che (teorema 2.9.5) è sempre zero! Infatti gli scostamenti in più e in meno dalla media, se presi col loro segno, si compensano.
Bisogna dunque considerare gli scostamenti dalla media indipendentemente dal segno; potremmo scegliere la cosiddetta "deviazione assoluta" $|x_i-\mu_{\mathbf{X}}|$ ma di fatto si preferisce considerare la deviazione quadratica $(x_i-\mu_{\mathbf{X}})^2$.
Nelle applicazioni alla realtà fisica, molto spesso i numeri che una variabile aleatoria associa ai risultati dell'esperimento esprimono la misura di qualche grandezza (l'altezza delle persone in un campione di popolazione; l'energia sviluppata in una interazione fra particelle; ecc. ecc.) e quindi tali numeri sono riferiti a una unità di misura (l'altezza ad esempio si può misurare in metri, centimetri, piedi, pollici, ecc. ecc.). Anche il valore medio è riferito alla stessa unità di misura. Nel calcolo della deviazione quadratica e della varianza si introduce però un elevamento al quadrato; considerando anziché la varianza la deviazione standard possiamo continuare a riferirci alla stessa unità di misura.

Vediamo ora la varianza e la deviazione standard della variabile aleatoria negli esempi visti in 2.8.

Esempio 2.10.2

Abbiamo visto nell'esempio 2.9.1 che il valore medio della variabile aleatoria **X** considerata nell'esempio 2.8.1 è $\mu_{\mathbf{X}} = 7$. Si ha dunque

$$\mathrm{Var}(\mathbf{X}) := \tfrac{1}{36} \cdot (7-2)^2 + \tfrac{1}{18} \cdot (7-3)^2 + \tfrac{1}{12} \cdot (7-4)^2 + \tfrac{1}{9} \cdot (7-5)^2 + \tfrac{5}{36} \cdot (7-6)^2 +$$

$$+ \tfrac{1}{6} \cdot (7-7)^2 + \tfrac{5}{36} \cdot (7-8)^2 + \tfrac{1}{9} \cdot (7-9)^2 + \tfrac{1}{12} \cdot (7-10)^2 + \tfrac{1}{18} \cdot (7-11)^2 + \tfrac{1}{36} \cdot (7-12)^2 =$$

$$= \tfrac{25+32+27+16+5+0+5+16+27+32+25}{36} = \tfrac{210}{36} = \tfrac{35}{6}\,.$$

e $$\sigma_{\boldsymbol{X}} := \sqrt{\tfrac{35}{6}} \approx 2{,}415\,.$$

Esempio 2.10.3

Abbiamo visto nell'esempio 2.9.2 che il valore medio della variabile aleatoria **X** considerata nell'esempio 2.8.2 è $\mu_{\mathbf{X}} = \frac{1}{6}$. Si ha dunque

$$\mathrm{Var}(\mathbf{X}) := \tfrac{1}{6} \cdot (1 - \tfrac{1}{6})^2 + \tfrac{5}{6} \cdot (0 - \tfrac{1}{6})^2 = \tfrac{25}{216} + \tfrac{5}{216} = \tfrac{30}{216} = \tfrac{5}{36}$$

e $$\sigma_{\boldsymbol{X}} := \sqrt{\tfrac{5}{36}} = \tfrac{\sqrt{5}}{6} \approx 0{,}373\,.$$

Esempio 2.10.4

Abbiamo visto nell'esempio 2.9.3 che il valore medio della variabile aleatoria **X** considerata nell'esempio 2.8.3 è $\mu_{\mathbf{X}} = -\frac{1}{6}$. Si ha dunque

$$\mathrm{Var}(\mathbf{X}) := \tfrac{1}{6} \cdot (1 + \tfrac{1}{6})^2 + \tfrac{1}{6} \cdot (2 + \tfrac{1}{6})^2 + \tfrac{1}{6} \cdot (3 + \tfrac{1}{6})^2 +$$

$$+ \tfrac{1}{6} \cdot (4 + \tfrac{1}{6})^2 + \tfrac{1}{6} \cdot (-5 + \tfrac{1}{6})^2 + \tfrac{1}{6} \cdot (-6 + \tfrac{1}{6})^2 =$$

$$= \tfrac{1}{6} \cdot \left(\tfrac{49+169+361+625+841+1225}{36}\right) = \tfrac{545}{36}$$

e $$\sigma_{\boldsymbol{X}} := \sqrt{\tfrac{545}{36}} = \tfrac{\sqrt{545}}{6} \approx 3{,}89\,.$$

Teorema 2.10.5

Siano $\mathbf{\Omega}$ un insieme non vuoto, p una misura di probabilità su $\mathbf{\Omega}$ e $\mathbf{X}$ una variabile aleatoria su $\mathbf{\Omega}$. Siano a, b numeri reali con $a \neq 0$.
Posto (come nel teorema 2.9.4) $\mathbf{Y}(\omega) := a\mathbf{X}(\omega) + b \quad \forall \omega \in \mathbf{\Omega}$
si ha
$$\mathrm{Var}(\mathbf{Y}) = a^2 \cdot \mathrm{Var}(\mathbf{X})$$
e
$$\sigma_{\mathbf{Y}} = |a| \cdot \sigma_{\mathbf{X}}.$$

Dimostrazione - Sia $\mu_{\mathbf{X}}$ il valore medio di $\mathbf{X}$; allora, per il teorema 2.9.4, il valore medio di $\mathbf{Y}$ è $a\mu_{\mathbf{X}} + b$. Per definizione, $\mathrm{Var}(\mathbf{Y})$ è il valore medio di
$$(\mathbf{Y} - (a\mu_{\mathbf{X}} + b))^2 = (a\mathbf{X} + b - a\mu_{\mathbf{X}} - b)^2 = (a \cdot (\mathbf{X} - \mu_{\mathbf{X}}))^2 = a^2 \cdot (\mathbf{X} - \mu_{\mathbf{X}})^2$$
e dunque, applicando ancora il teorema 2.9.4,
$$\mathrm{Var}(\mathbf{Y}) = a^2 \cdot \mathrm{Var}(\mathbf{X}).$$
Dalla definizione di $\sigma_{\mathbf{Y}}$ si ha infine che
$$\sigma_{\mathbf{Y}} = \sqrt{\mathrm{Var}(\mathbf{Y})} = \sqrt{a^2 \cdot \mathrm{Var}(\mathbf{X})} = \sqrt{a^2} \cdot \sqrt{\mathrm{Var}(\mathbf{X})} = |a| \cdot \sigma_{\mathbf{X}}.$$

Sia $\mathbf{X}$ una variabile aleatoria con valore medio $\mu_{\mathbf{X}}$ e deviazione standard $\sigma_{\mathbf{X}}$ non nulla. Si dice *variabile aleatoria standardizzata* associata a $\mathbf{X}$ la $\frac{\mathbf{X} - \mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}$.

Teorema 2.10.6

Sia $\mathbf{X}$ una variabile aleatoria con deviazione standard non nulla. La variabile aleatoria standardizzata associata a $\mathbf{X}$ ha valore medio 0 e deviazione standard 1.

Dimostrazione - Si applichino i teoremi 2.9.5 e 2.10.5.

2.11 - La disuguaglianza di Chebyshev.

Il valore medio $\mu_{\mathbf{X}}$ di una variabile aleatoria è in un certo senso il "centro" attorno al quale vengono a cadere i valori assunti dalla variabile aleatoria stessa; ci si aspetta che, ripetendo l'esperimento, solo "poche" volte $\mathbf{X}$ assuma valori "molto distanti" da $\mu_{\mathbf{X}}$. Sperimentalmente, si trova che i valori assunti dalla variabile aleatoria non differiscono "quasi mai" dal valor medio per più di **tre deviazioni standard**. Una stima (molto più cauta) della situazione è espressa dal

Teorema 2.11.1 ("disuguaglianza di Chebyshev")

Siano $\mathbf{\Omega}$ un insieme non vuoto, p una misura di probabilità su $\mathbf{\Omega}$ e $\mathbf{X}$ una variabile aleatoria su $\mathbf{\Omega}$ con valore medio $\mu_{\mathbf{X}}$ e deviazione standard σ. Sia h un numero reale positivo.

Posto $$\mathbf{E}_h^* := \{\omega \in \mathbf{\Omega} \,/\, |\mathbf{X}(\omega) - \mu_{\mathbf{X}}| > h\sigma\} \quad \left(= \bigcup_{|x-\mu| > h\sigma} \mathbf{E}_x \right),$$

si ha $\quad p(\mathbf{E}_h^*) < \frac{1}{h^2}$.

Ciò si esprime anche scrivendo $\quad p(|\mathbf{X} - \mu_{\mathbf{X}}| > h\sigma) < \frac{1}{h^2}$.

Dimostrazione - Omettiamo la dimostrazione di questo teorema.

Osservazione 2.11.2

La probabilità che la variabile aleatoria differisca dal valore medio per più di tre deviazioni standard si può stimare mediante la disuguaglianza di Chebyshev ponendo $h := 3$.
Si ottiene $$p(|\mathbf{X} - \mu_{\mathbf{X}}| > 3\sigma) < \tfrac{1}{9}\ .$$
Come si è già osservato, questa stima è fin troppo cauta.

Esempio 2.11.3

Applichiamo la disuguaglianza di Chebyshev con $h := 2$ alla variabile aleatoria $\mathbf{X}$ considerata nell'esempio 2.8.1 .
Abbiamo visto che $\mu_{\mathbf{X}} = 7$ (esempio 2.9.1) e $\sigma_{\mathbf{X}} \approx 2{,}415$ (esempio 2.10.2) . Per il teorema di Chebyshev si ha
$$p(\mathbf{E}_2^*) < \tfrac{1}{4}\ .$$
In realtà, essendo $\mathbf{E}_2^* = \mathbf{E}_2 \cup \mathbf{E}_3 \cup \mathbf{E}_{11} \cup \mathbf{E}_{12}$ (con gli $\mathbf{E}_i$, come sappiamo, a due a due disgiunti), si ha
$$p(\mathbf{E}_2^*) = p(\mathbf{E}_2) + p(\mathbf{E}_3) + p(\mathbf{E}_{11}) + p(\mathbf{E}_{12}) = \tfrac{6}{36} = \tfrac{1}{6}\ .$$

2.12 - Esperimenti binomiali.

Un esperimento si dice *di Bernoulli* se lo descriviamo mediante uno spazio dei risultati che ha esattamente due elementi ([7]) ; tali due risultati vengono convenzionalmente indicati come "successo" e "insuccesso" .

Esempi

2.12.1 Lancio di una moneta.
2.12.2 Lancio di una moneta truccata (tappo di barattolo; puntina da disegno).
2.12.3 Lancio di un dado (successo := "esce un numero pari") .
2.12.4 Lancio di un dado (successo := "esce 3") .
2.12.5 Lancio di due dadi (successo := "escono numeri doppi") .
2.12.6 Lancio di una freccetta contro un bersaglio (successo := "si colpisce il settore centrale") .
2.12.7 Estrazione del lotto sulla ruota di Napoli (successo := "è uscito il terno che ho giocato") .

[7] In sostanza, un esperimento è di Bernoulli quando nel descriverne i possibili risultati ci interessa soltanto se una certa situazione si verifica oppure no. Non è dunque l' esperimento in sé che merita un nome particolare, ma il nostro modo di esaminarne i risultati.

Sia $\boldsymbol{\mathcal{B}}$ un esperimento di Bernoulli. Un esperimento $\boldsymbol{\mathcal{E}}$ si dice *esperimento binomiale* (*associato a* $\boldsymbol{\mathcal{B}}$) se:
(i) $\boldsymbol{\mathcal{E}}$ consiste nel ripetere più volte $\boldsymbol{\mathcal{B}}$; ogni ripetizione di $\boldsymbol{\mathcal{B}}$ si dice un *tentativo*;
(ii) la probabilità di successo non varia al variare dei tentativi;
(iii) gli eventi <al j – simo tentativo si verifica un successo> sono a due a due indipendenti al variare di j.

Sia $\boldsymbol{\mathcal{B}}$ un esperimento di Bernoulli, e sia $\boldsymbol{\Omega}$ lo spazio dei risultati associato.

Sia n un numero intero positivo. All'esperimento binomiale $\boldsymbol{\mathcal{E}_n}$ consistente nel ripetere n volte $\boldsymbol{\mathcal{B}}$ si può associare come spazio dei risultati il prodotto cartesiano

$$\underbrace{\boldsymbol{\Omega} \times \boldsymbol{\Omega} \times \ldots \times \boldsymbol{\Omega}}_{n \text{ volte}} = \boldsymbol{\Omega}^n .$$

Sia $\mathbf{X} : \boldsymbol{\Omega}^n \to \{0, 1, \ldots, n\}$ l'applicazione che associa a ogni n – pla ordinata di elementi di $\boldsymbol{\Omega}$ il numero dei successi che vi compaiono. Questa $\mathbf{X}$ è una variabile aleatoria, detta *numero dei successi* (*su* n *tentativi*).

Teorema 2.12.8

Sia $\boldsymbol{\mathcal{B}}$ un esperimento di Bernoulli, con probabilità di successo a e probabilità di insuccesso b (cosicché $b = 1 - a$).
Sia $\boldsymbol{\mathcal{E}_n}$ l'esperimento binomiale che consiste nel ripetere n volte $\boldsymbol{\mathcal{B}}$, e sia $\mathbf{X}$ la variabile aleatoria "numero dei successi" ad esso associata. Per $k := 0, 1, \ldots, n$, si ha

$$p(\mathbf{X} = k) \;=\; \binom{n}{k} a^k b^{n-k}.$$

Inoltre, il valore medio di $\mathbf{X}$ è na e si ha

$$\mathrm{Var}(\mathbf{X}) = nab , \qquad \sigma_{\mathbf{X}} = \sqrt{nab} .$$

Dimostrazione - Calcoliamo $p(\mathbf{X} = k)$ per $k := 0, 1, \ldots, n$.
Tale probabilità è la somma delle probabilità dei diversi eventi elementari nei quali si verificano k successi; tali eventi elementari (cioè n – ple ordinate di elementi di $\boldsymbol{\Omega}$) sono in numero di $\binom{n}{k}$ perché corrispondono ai diversi modi di scegliere k indici tra 1 e n; inoltre, la probabilità di ciascuno di essi è $a^k b^{n-k}$ (poiché le probabilità di successo e insuccesso non variano al ripetersi di $\boldsymbol{\mathcal{B}}$, e gli esiti dei tentativi ripetuti sono eventi indipendenti).
Dunque

$$p(\mathbf{X} = k) = \binom{n}{k} a^k b^{n-k}$$

come si voleva dimostrare. Notiamo che

$$\sum_{i=1}^{n} p(\mathbf{X} = k) = \sum_{i=1}^{n} \binom{n}{k} a^k b^{n-k} = (a+b)^n = 1^n = 1$$

come ci si aspettava.

Calcoliamo ora il valor medio di **X**:

$$\mu_{\mathbf{X}} := \sum_{k=0}^{n} p(\mathbf{X}=k)\cdot k = \sum_{k=0}^{n} k\cdot\binom{n}{k} a^k b^{n-k} =$$

$$= \sum_{k=1}^{n} k\cdot\binom{n}{k} a^k b^{n-k} \overset{\text{Teorema 13.4.5 }(c)}{=} \sum_{k=1}^{n} n\cdot\binom{n-1}{k-1} a^k b^{n-k} = na \sum_{k=1}^{n}\binom{n-1}{k-1} a^{k-1} b^{n-k} =$$

$$\overset{i:=k-1}{=} na \sum_{i=0}^{n-1}\binom{n-1}{i} a^i b^{(n-1)-i} = na(a+b)^{n-1} = na\cdot 1^{n-1} = na$$

Dunque ci aspettiamo “in media” ([8]) na successi eseguendo n esperimenti di Bernoulli.

Calcoliamo ora la varianza di **X**:

$$\mathrm{Var}(\mathbf{X}) := \sum_{k=0}^{n} p(\mathbf{X}=k)\cdot(k-\mu_{\mathbf{X}})^2 = \sum_{k=0}^{n} p(\mathbf{X}=k)\cdot(k-na)^2 = \sum_{k=0}^{n} p(\mathbf{X}=k)\cdot(n^2a^2-2nak+k^2) =$$

$$= n^2a^2\sum_{k=0}^{n} p(\mathbf{X}=k) - 2na\sum_{k=0}^{n} p(\mathbf{X}=k)\cdot k + \sum_{k=0}^{n} p(\mathbf{X}=k)\cdot k^2 =$$

$$= n^2a^2\cdot 1 - 2na\mu_{\mathbf{X}} + \sum_{k=0}^{n} p(\mathbf{X}=k)\cdot k^2 = n^2a^2 - 2n^2a^2 + \sum_{k=0}^{n} p(\mathbf{X}=k)\cdot k^2 .$$

Si ha

$$\sum_{k=0}^{n} p(\mathbf{X}=k)\cdot k^2 = \sum_{k=0}^{n} k^2\binom{n}{k} a^k b^{n-k} =$$

$$= \sum_{k=1}^{n} k^2\binom{n}{k} a^k b^{n-k} \overset{\text{Teorema 13.4.5 }(c)}{=} \sum_{k=1}^{n} kn\binom{n-1}{k-1} a^k b^{n-k} =$$

$$= na\sum_{k=1}^{n} k\binom{n-1}{k-1} a^{k-1} b^{n-k} \overset{\substack{[i:=k-1]\\ [m:=n-1]}}{=} na\sum_{i=0}^{m}(i+1)\binom{m}{i} a^i b^{m-i} =$$

$$= na\left(\sum_{i=0}^{m} i\binom{m}{i} a^i b^{m-i} + \sum_{i=0}^{m}\binom{m}{i} a^i b^{m-i}\right) = na\left(ma+(a+b)^m\right) = na\left((n-1)a+1^m\right) =$$

$$= na(na-a+1) = na(na+b) = n^2a^2 + nab$$

cosicché infine si è trovato che

$$\mathrm{Var}(\mathbf{X}) = n^2a^2 - 2n^2a^2 + \sum_{k=0}^{n} p(\mathbf{X}=k)\cdot k^2 = n^2a^2 - 2n^2a^2 + n^2a^2 + nab = nab$$

e quindi possiamo infine calcolare la deviazione standard

$$\sigma_{\mathbf{X}} := \sqrt{\mathrm{Var}(\mathbf{X})} = \sqrt{nab} .$$

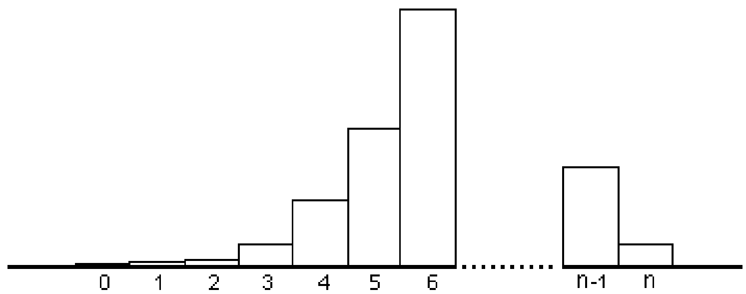
[8] la media è riferita a più ripetizioni dell' esperimento binomiale $\mathcal{E}_n$.

Osservazione 2.12.9

Possiamo tabulare come segue le probabilità $p(\mathbf{X}=k)$:

k	0	1	2	$\ldots$	$n-1$	n
$p(\mathbf{X}=k)$	b^n	nab^{n-1}	$\binom{n}{2}a^2b^{n-2}$	$\ldots$	nab^{n-1}	b^n

Possiamo anche disegnare un istogramma



nel quale l'area della colonna corrispondente a ciascun valore k ([9]) di $\mathbf{X}$ è la probabilità $p(\mathbf{X}=k)$, mentre la funzione di ripartizione si può calcolare in k misurando l'area complessiva delle colonne "fino a k incluso". Si tenga presente che valori diversi di a, b e n danno naturalmente luogo a istogrammi diversi.

Osservazione 2.12.10

Sia $\mathcal{E}_n$ l'esperimento binomiale che consiste nel ripetere n volte un fissato esperimento di Bernoulli $\mathcal{B}$, e sia $\mathbf{X}$ la variabile aleatoria "numero dei successi" ad esso associata.
Raddoppiando n raddoppia il valor medio di $\mathbf{X}$ (e raddoppia la varianza di $\mathbf{X}$) ma la deviazione standard viene moltiplicata per $\sqrt{2}$ (e $\sqrt{2}$ è sensibilmente più piccolo di 2).
Passando da n a $100n$ ripetizioni dell'esperimento di Bernoulli, il valor medio di $\mathbf{X}$ è moltiplicato per 100 ma la deviazione standard (che misura il margine di variazione rispetto alla media) viene moltiplicato solo per 10 : il margine percentuale di errore, quindi, diminuisce!

Esempio 2.12.11

Esperimento: lancio di un dado. Successo: Esce un quadrato perfetto (cioè 1 oppure 4). Probabilità di successo: $a=\frac{1}{3}$; probabilità di insuccesso: $b=1-a=\frac{2}{3}$.

Su 300 lanci: valor medio 100; deviazione standard $\sqrt{\frac{2\cdot 300}{9}}=\sqrt{66{,}6}\simeq 8{,}16$. Il numero di successi sarà quasi certamente compreso tra 75 e 125 (tre deviazioni standard attorno al valore medio); il rapporto $\frac{\text{scarto}}{\text{successi}}$ è $\frac{25}{100}$, ossia il 25% di margine in più o meno.

[9] in questo caso k assume tutti i valori interi da 0 a n .

Su 30.000 lanci: valor medio 10.000; deviazione standard $\sqrt{\frac{2\cdot 30.000}{9}} = \sqrt{6666} \simeq 81{,}6$. Il numero di successi sarà quasi certamente compreso tra 9.750 e 10.250 (tre deviazioni standard attorno al valore medio); il rapporto $\frac{\text{scarto}}{\text{successi}}$ è $\frac{250}{10.000} = \frac{25}{1.000}$, ossia il 2,5% di margine in più o meno.

Esempio 2.12.12

Sappiamo che in un certo libro di 500 pagine ci sono 300 errori "distribuiti casualmente". Qual è la probabilità che a pag. 37 ci siano esattamente 3 errori? Possiamo ragionare così: l'ipotesi che la distribuzione degli errori sia casuale esprime il fatto che per ciascun errore ogni pagina ha uguale probabilità che vi compaia. Dunque: per ogni errore, la probabilità che esso sia a pag. 37 ("successo") è $\frac{1}{500}$; la probabilità che esso sia in una qualsiasi altra pagina ("insuccesso") è $b = 1 - a = \frac{499}{500}$. Ogni errore è un esperimento di Bernoulli; l'insieme dei 300 errori è un esperimento binomiale costituito da 300 esperimenti di Bernoulli ($n = 300$). La probabilità che a pag. 37 ci siano esattamente 3 errori è dunque

$$\binom{n}{3} a^3 b^{n-3} = \binom{300}{3} \frac{1}{500^3} \cdot \frac{499^{497}}{500^{497}} .$$

Si noti che non è immediato calcolare questo numero!

2.13 - Fino a raggiungere il successo.

Consideriamo l'esperimento $\mathcal{E}$ che consiste nel ripetere un dato esperimento di Bernoulli $\mathcal{B}$ finché non si ottiene un successo. Possiamo scegliere $\mathbb{Z}^+$ come spazio dei risultati: la variabile aleatoria che si considera in questo caso (detta *tempo di attesa*) è l'identità su $\mathbb{Z}^+$ (essa associa n al risultato di un esperimento che ha richiesto n ripetizioni di $\mathcal{B}$ per ottenere un successo).

Teorema 2.13.1

Sia $\mathcal{B}$ un esperimento di Bernoulli con probabilità di successo a e probabilità di insuccesso b (cosicché $b = 1 - a$).
Sia $\mathcal{E}$ l'esperimento che consiste nel ripetere $\mathcal{B}$ finché non si ottiene un successo, e sia $\mathbf{X}$ la variabile aleatoria "tempo di attesa" ad esso associata. Si ha $p(\mathbf{X} = k) = ab^{k-1}$ $\forall k \in \mathbb{Z}^+$ e il valore medio di $\mathbf{X}$ è $\frac{1}{a}$.

Dimostrazione - L'unico risultato di $\mathcal{E}$ al quale $\mathbf{X}$ associa il valore k è quello per il quale nelle prime $k - 1$ ripetizioni di $\mathcal{B}$ si ha un insuccesso e nella $k -$ sima ripetizione di $\mathcal{B}$ si ha un successo. Poiché i risultati nelle ripetizioni di $\mathcal{B}$ sono indipendenti, la probabilità di quanto sopra è ab^{k-1}, come si voleva.
Omettiamo il calcolo del valore medio di $\mathbf{X}$.

Osservazione 2.13.2

Nella situazione vista sopra, le $p(\mathbf{X} = i)$ formano una successione geometrica con termine iniziale a e ragione b.

Esempio 2.13.3

Giocando a “Monopoli”, per “uscire di prigione” bisogna fare “numeri doppi”. Quanti turni ci si deve aspettare di rimanere, in media, in “prigione”?
L’esperimento di Bernoulli qui considerato è quello visto all’esempio 2.12.5, con “successo” := “numeri doppi”. La probabilità a di successo è $\frac{1}{6}$ (perché?) e il valore medio della variabile aleatoria “tempo di attesa” è $\frac{1}{a}$ cioè 6. Ci si deve dunque aspettare di rimanere “in prigione”, in media, 6 turni.

Esempio 2.13.4

Al “Gioco delle Carriere”, si resta alla “Panchina al Parco” finché tirando i dadi non si realizza 7, 11 o “numeri doppi”. Quanti turni ci si deve aspettare di rimanere, in media, alla “Panchina al Parco”?
L’esperimento di Bernoulli qui considerato consiste nel lancio di due dadi, con “successo” := “7, 11 o numeri doppi”. La probabilità a di successo è $\frac{7}{18}$ (perché?) e il valore medio della variabile aleatoria “tempo di attesa” è $\frac{1}{a}$ cioè $\frac{18}{7}$. Ci si deve dunque aspettare di rimanere alla “Panchina al Parco”, in media, poco più di 2 turni e mezzo.

Esercizio 2.13.5

Al “Gioco delle Carriere”, si resta all’“Ospedale” finché tirando (entrambi) i dadi non si realizza 5 o meno di 5. Quanti turni ci si deve aspettare di rimanere, in media, all’“Ospedale”?

Esercizio 2.13.6

Nel gioco “La grande abbuffata” di Alex Randolph, si avanza con la pedina su un percorso di ventuno caselle tirando i dadi con la seguente regola. Si lancia un dado e si osserva il punteggio x_1 ottenuto; si può scegliere se avanzare di x_1 caselle oppure lanciare un altro dado; in quest’ultimo caso, detto x_2 il punteggio ottenuto col secondo dado, se $x_1 + x_2 > 7$ si torna alla partenza, se invece $x_1 + x_2 \leq 7$ si avanza di $2 \cdot (x_1 + x_2)$ caselle.
Stabilire una strategia per decidere (in funzione di x_1 e della casella su cui ci si trova) se conviene lanciare il secondo dado.
Suggerimento: si valuti in funzione di x_1 e della casella su cui ci si trova la speranza matematica di avanzamento quando non si tira il secondo dado e quando invece lo si tira.

2.14 - Variabili aleatorie con immagine non numerabile.

Siano $\mathbf{\Omega}$ un insieme non vuoto, p una misura di probabilità su $\mathbf{\Omega}$ e $\mathbf{X}$ una variabile aleatoria su $\mathbf{\Omega}$. Se $\mathbf{X}$ assume un'infinità non numerabile di valori, le $p(\mathbf{X} = x_0)$ sono spesso tutte uguali a zero, e comunque non consentono in generale di determinare la probabilità degli eventi "interessanti" individuabili mediante $\mathbf{X}$: si ricorre quindi ad un approccio completamente diverso, che possiamo descrivere qui soltanto per grandi linee e in un caso particolare.

Sia $\mathbf{f}$ una funzione $\mathbb{R} \to \mathbb{R}$ con la proprietà che: detta $\mathbf{F}$ la funzione di ripartizione della variabile aleatoria $\mathbf{X}$ (cfr. sez. 2.8), si ha

$$(\bigstar) \qquad \mathbf{F}(x_0) = \int_{-\infty}^{x_0} \mathbf{f}(\mathrm{t})\, \mathrm{dt}\,.$$

Non è detto in generale che una tale $\mathbf{f}$ esista; se essa esiste, si dice che $\mathbf{f}$ è la *densità di probabilità* di $\mathbf{X}$ e (comunque presi x_1, $x_2 \in \mathbb{R}$ con $x_1 \leq x_2$) si ha

$$p(x_1 \leq \mathbf{X} \leq x_2) = \int_{x_1}^{x_2} \mathbf{f}(\mathrm{t})\, \mathrm{dt}\,.$$

Si definiscono per $\mathbf{X}$ il valore medio $\mu_{\mathbf{X}}$, la varianza $\mathrm{Var}(\mathbf{X})$ e la deviazione standard $\sigma_{\mathbf{X}}$ ponendo

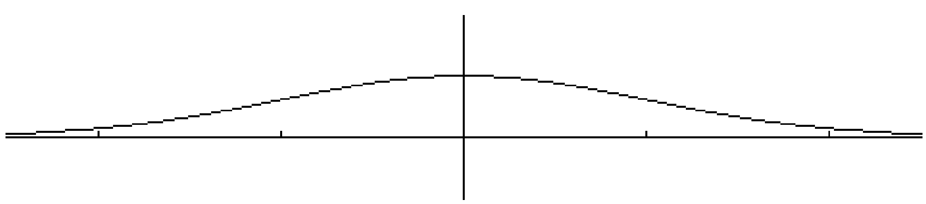
$$\mu_{\mathbf{X}} := \int_{-\infty}^{+\infty} \mathrm{x} \cdot \mathbf{f}(\mathrm{x})\, \mathrm{dx}$$

$$\mathrm{Var}(\mathbf{X}) := \int_{-\infty}^{+\infty} (\mathrm{x} - \mu_{\mathbf{X}})^2 \mathbf{f}(\mathrm{x})\, \mathrm{dx} \qquad \text{e} \qquad \sigma_{\mathbf{X}} := \sqrt{\mathrm{Var}(\mathbf{X})}\,.$$

La funzione

$$\mathbf{f}_0(\mathrm{x}) := \frac{e^{-\frac{\mathrm{x}^2}{2}}}{\sqrt{2\pi}}$$

si dice *densità normale standard*. Il suo grafico è approssimativamente il seguente:



$\boxed{\text{Esercizio 2.14.1}}$

Si studi la funzione

$$\mathbf{f}_0\,(\mathrm{x}) := \frac{e^{-\frac{\mathrm{x}^2}{2}}}{\sqrt{2\pi}}$$

e se ne tracci approssivamente il grafico.

Si dice che una variabile aleatoria **X** *ha distribuzione di probabilità normale standard* se la densità di probabilità ad essa associata è la densità normale standard. Si dimostra che una variabile aleatoria con distribuzione di probabilità normale standard ha valore medio 0 e deviazione standard 1 .

Per calcolare approssimativamente le probabilità di una variabile aleatoria con distribuzione di probabilità normale standard, si può usare la seguente tabulazione dei valori di $\int_0^x \mathbf{f}_0(\mathrm{t})\,\mathrm{dt}$ per $0 \leq x \leq 4$:

Tabella 2.14.T1

x	$\int_0^x \mathbf{f}_0(\mathrm{t})\,\mathrm{dt}$	x	$\int_0^x \mathbf{f}_0(\mathrm{t})\,\mathrm{dt}$	x	$\int_0^x \mathbf{f}_0(\mathrm{t})\,\mathrm{dt}$	x	$\int_0^x \mathbf{f}_0(\mathrm{t})\,\mathrm{dt}$
0,00	0,0000	1,00	0,3413	2,00	0,4772	3,00	0,4987
0,05	0,0199	1,05	0,3531	2,05	0,4798	3,05	0,4989
0,10	0,0398	1,10	0,3643	2,10	0,4821	3,10	0,4990
0,15	0,0596	1,15	0,3749	2,15	0,4842	3,15	0,4992
0,20	0,0793	1,20	0,3849	2,20	0,4861	3,20	0,4993
0,25	0,0987	1,25	0,3944	2,25	0,4878	3,25	0,4994
0,30	0,1179	1,30	0,4032	2,30	0,4893	3,30	0,4995
0,35	0,1368	1,35	0,4115	2,35	0,4906	3,35	0,4996
0,40	0,1554	1,40	0,4192	2,40	0,4918	3,40	0,4997
0,45	0,1736	1,45	0,4265	2,45	0,4929	3,45	0,4997
0,50	0,1915	1,50	0,4332	2,50	0,4938	3,50	0,4998
0,55	0,2088	1,55	0,4394	2,55	0,4946	3,55	0,4998
0,60	0,2257	1,60	0,4452	2,60	0,4953	3,60	0,4998
0,65	0,2422	1,65	0,4505	2,65	0,4960	3,65	0,4998
0,70	0,2580	1,70	0,4554	2,70	0,4965	3,70	0,4999
0,75	0,2734	1,75	0,4599	2,75	0,4970	3,75	0,4999
0,80	0,2881	1,80	0,4641	2,80	0,4974	3,80	0,4999
0,85	0,3023	1,85	0,4678	2,85	0,4978	3,85	0,4999
0,90	0,3159	1,90	0,4713	2,90	0,4981	3,90	0,5000
0,95	0,3289	1,95	0,4744	2,95	0,4984	3,95	0,5000

Poiché la $\mathbf{f}_0$ è una funzione pari (cfr. sez. 21.4) e $\mathbf{f}_0(x) \approx 0$ se $x \geq 4$, la tabella 2.14.T1 consente di stimare $p(x_1 \leq \mathbf{X} \leq x_2)$ quando **X** ha distribuzione di probabilità normale standard.

Esempio 2.14.2

Sia $\mathbf{X}$ una variabile aleatoria con distribuzione di probabilità normale standard. Utilizziamo la Tabella 2.14.T1 per stimare approssimativamente $p(-1 \leq \mathbf{X} \leq 2)$.
Per considerazioni di simmetria, $p(-1 \leq \mathbf{X} \leq 2) = p(0 \leq \mathbf{X} \leq 1) + p(0 \leq \mathbf{X} \leq 2)$. Dalla tabella si ricava che $p(0 \leq \mathbf{X} \leq 1) \approx 0{,}3413$ e $p(0 \leq \mathbf{X} \leq 2) \approx 0{,}4772$. Pertanto $p(-1 \leq \mathbf{X} \leq 2) \approx 0{,}8185$.

Esempio 2.14.3

Sia $\mathbf{X}$ una variabile aleatoria con distribuzione di probabilità normale standard. Utilizziamo la Tabella 2.14.T1 per stimare approssimativamente $p(\mathbf{X} \leq -1)$.
Per considerazioni di simmetria, $p(\mathbf{X} \leq -1) = p(1 \leq \mathbf{X}) = \frac{1}{2} - p(0 \leq \mathbf{X} \leq 1)$. Dalla tabella si ricava che $p(0 \leq \mathbf{X} \leq 1) \approx 0{,}3413$. Pertanto $p(\mathbf{X} \leq -1) \approx 0{,}1587$.

Si dice che una variabile aleatoria $\mathbf{X}$ *ha distribuzione di probabilità normale* se la variabile aleatoria standardizzata ottenuta da essa (cfr. sez. 2.10) ha distribuzione di probabilità normale standard.

Se conosciamo il valore medio e la deviazione standard di una variabile aleatoria $\mathbf{X}$ con distribuzione di probabilità normale, possiamo ancora utilizzare la Tabella 2.14.T1 per stimare approssimativamente $p(x_1 \leq \mathbf{X} \leq x_2)$ per $x_1, x_2 \in \mathbb{R}$ con $x_1 \leq x_2$.

Esempio 2.14.4

Il signor Rossi deve essere in ufficio ogni giorno alle 9:00. Egli fa in modo di arrivare attorno alle 8:55: la differenza, espressa in minuti, tra l'ora di effettivo arrivo del signor Rossi e le 8:55 può essere descritta con una variabile aleatoria $\mathbf{X}$ con distribuzione di probabilità normale che ha valor medio 0 e deviazione standard 2,5.
Qual è la probabilità che il sig. Rossi arrivi al lavoro in ritardo (cioè dopo le 9:00)?
Dobbiamo calcolare $p(5 \leq \mathbf{X})$, ossia (passando alla variabile aleatoria standardizzata $\mathbf{Y} := \frac{\mathbf{X}-0}{2{,}5} = \frac{2}{5}\mathbf{X}$), dobbiamo calcolare $p(2 \leq \mathbf{Y})$. La Tabella 2.14.T1 ci fornisce $p(0 \leq \mathbf{Y} \leq 2) \approx 0{,}4772$. Ricordando che $p(0 \leq \mathbf{Y}) = \frac{1}{2}$, si trova che $p(5 \leq \mathbf{X}) \approx 0{,}0228$.
Pertanto il sig. Rossi arriva in ufficio in ritardo meno del 2,3% delle volte.

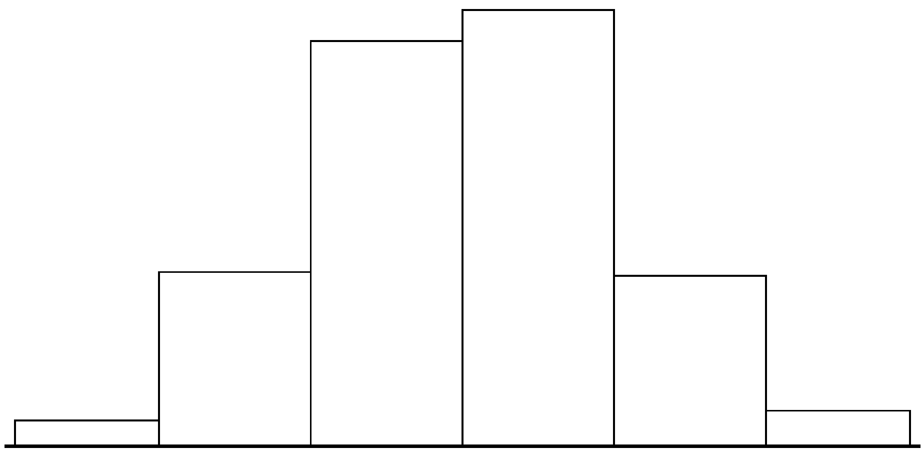
Esempio 2.14.5

Nella gelida Brutopia, la temperatura nel mese di Dicembre può essere descritta mediante una variabile aleatoria $\mathbf{X}$ con distribuzione di probabilità normale che ha valor medio -8^0 C e deviazione standard 5^0 C. Qual è la probabilità che in un dato istante la temperatura sia superiore a 0^0 C?
Dobbiamo calcolare $p(0 \leq \mathbf{X})$, ossia (passando alla variabile aleatoria standardizzata $\mathbf{Y} := \frac{\mathbf{X}+8}{5}$), dobbiamo calcolare $p(1{,}6 \leq \mathbf{Y})$. La Tabella 2.14.T1 ci fornisce $p(0 \leq \mathbf{Y} \leq 1{,}6) \approx 0{,}4452$. Ricordando che $p(0 \leq \mathbf{Y}) = \frac{1}{2}$, si trova che $p(0 \leq \mathbf{X}) \approx 0{,}0548$. Pertanto nel mese di Dicembre in Brutopia la temperatura supera gli 0 gradi Celsius solo il 5,48% delle volte.

2.15 - Approssimazione mediante variabili aleatorie con distribuzione di probabilità normale.

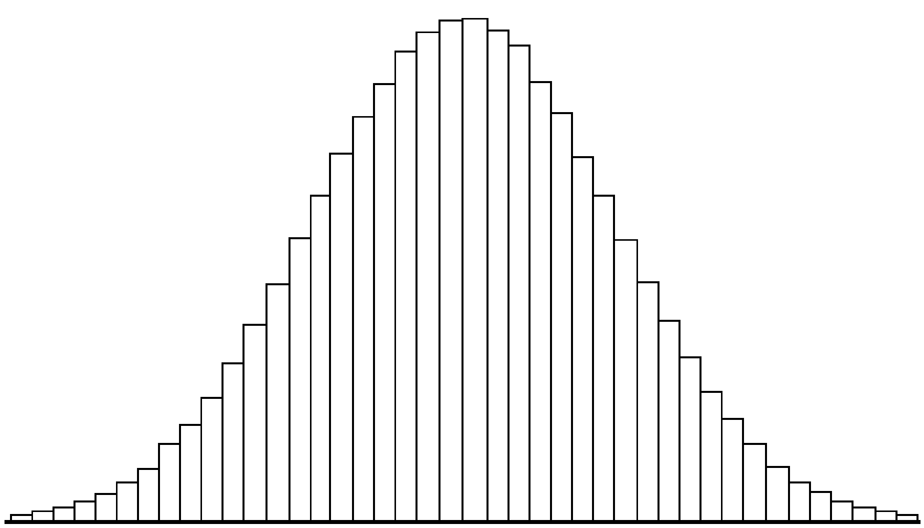
Sia dato un esperimento, e sia **X** una variabile aleatoria ad esso associata. Se **X** assume "molti valori" in un certo intervallo aperto (a, b) di numeri reali, può essere "matematicamente conveniente" descrivere l'esperimento mediante una variabile aleatoria che assume tutti i valori di quell'intervallo.

Consideriamo un caso concreto. L'esperimento consista nello scegliere a caso un cittadino italiano maggiorenne; e la variabile aleatoria **X** associ a ogni risultato (cioè a ogni cittadino italiano maggiorenne) il suo peso espresso in chilogrammi. è chiaro che **X** ha immagine finita, perché è definita su un insieme finito! La distribuzione di **X** può essere tabulata mediante un istogramma, avendo scelto un certo numero di ampiezze di pesi:

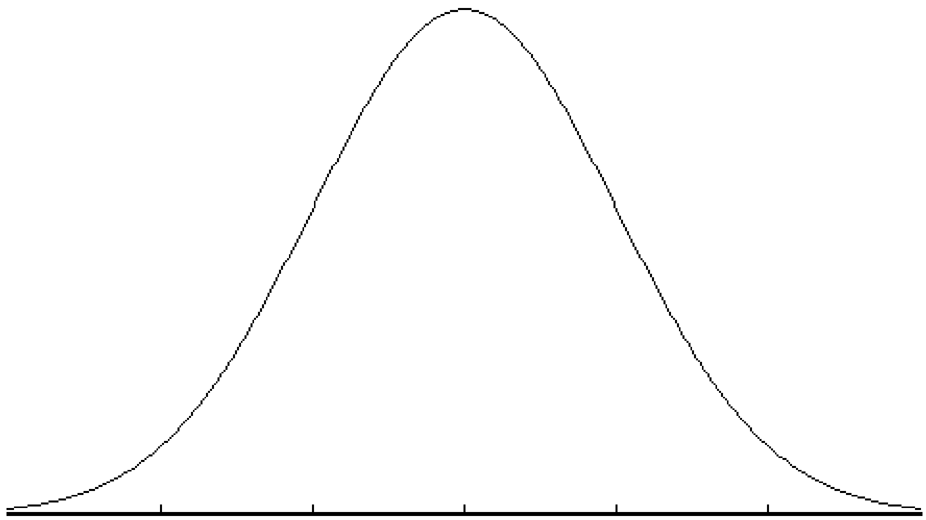


L'istogramma è fatto in modo che l'area di ogni blocco sia proporzionale al numero di individui che hanno il peso compreso nell'intervallo indicato alla base del blocco. Se l'area totale è uguale a 1, possiamo interpretare l'area di ciascun blocco come la probabilità che un individuo abbia peso compreso nell'intervallo indicato alla base.

Per migliorare la rappresentazione fornita dall'istogramma possiamo aumentare il numero degli intervalli, diminuendone di conseguenza le ampiezze per mantenere uguale a 1 l'area totale :



Continuando ad aumentare il numero degli intervalli diminuendone le ampiezze (in modo che l'area totale resti uguale a 1), l'istogramma si "avvicina" sempre più al grafico della densità di probabilità di una variabile aleatoria con distribuzione di probabilità normale, dalla caratteristica forma "a campana":



Stimando valore medio e deviazione standard della variabile aleatoria, si possono applicare le tecniche viste nella sezione 2.14 per calcolare la probabilità che la variabile aleatoria assuma valori in un assegnato intervallo.

Si dimostra che la variabile aleatoria "numero dei successi su n tentativi" (cfr. sez. 2.12) può essere approssivamente descritta, per n "sufficientemente grande", mediante una variabile aleatoria con distribuzione di probabilità normale.

Esempio 2.15.1

Si consideri nuovamente il problema visto nell'esempio 2.12.12.
Il valore $p(\mathbf{X} = 3)$ può essere stimato sostituendo alla variabile aleatoria $\mathbf{X}$ "numero dei successi su 300 tentativi" la variabile aleatoria $\mathbf{X_1}$ con distribuzione di probabilità normale che ha lo stesso valore medio $\frac{3}{5}$ e la stessa deviazione standard $\sqrt{\frac{3 \cdot 499}{5 \cdot 500}} \approx 0{,}77$, e calcolando $p(2{,}5 \leq \mathbf{X_1} \leq 3{,}5)$.
Per effettuare il calcolo, dobbiamo ricondurci alla variabile aleatoria standardizzata $\mathbf{Y}$ ottenuta da $\mathbf{X_1}$; si ha

$$\mathbf{Y} := \frac{\mathbf{X_1} - \frac{3}{5}}{\frac{77}{100}} = \frac{100 \cdot \mathbf{X_1} - 60}{77}$$

e dobbiamo calcolare $p(\frac{190}{77} \leq \mathbf{Y} \leq \frac{290}{77})$ ossia $p(2{,}5 \leq \mathbf{Y} \leq 3{,}8)$. Utilizzando la Tabella 2.14.T1 si trova che $p(0 \leq \mathbf{Y} \leq 2{,}5) \approx 0{,}4938$ e $p(0 \leq \mathbf{Y} \leq 3{,}8) \approx 0{,}4999$ per cui la probabilità cercata è circa 0,0061 .

3.- STATISTICA INFERENZIALE

3.1 - Il "teorema centrale del limite".

Sia Ω un insieme, sia p una misura di probabilità su Ω e siano X_1, X_2 due variabili aleatorie su Ω. Si dice che X_1 e X_2 sono *indipendenti* se comunque presi due intervalli (anche illimitati) I_1, I_2 contenuti in $\mathbb{R}$ gli eventi

$$E_1 := \{\omega \in \mathbf{\Omega} \,/\, \mathbf{X}_1(\omega) \in I_1\} \quad \text{e} \quad E_2 := \{\omega \in \mathbf{\Omega} \,/\, \mathbf{X}_2(\omega) \in I_2\}$$

sono indipendenti (ossia $p(E_1 \cap E_2) = p(E_1) \cdot p(E_2)$).

Teorema 3.1.1 ("Teorema centrale del limite")

Sia data una successione X_1, X_2, ..., X_n, ... di variabili aleatorie, e siano μ_i, σ_i^2 rispettivamente il valor medio e la varianza di ciascuna variabile aleatoria X_i. Supponiamo inoltre che

(C1) le funzioni X_i siano tutte limitate;

(C2) le X_i siano tutte a due a due indipendenti;

(C3) si abbia $\lim\limits_{n \to +\infty} \sum\limits_{i=1}^{n} \sigma_i^2 = +\infty$.

Posto, per ogni $n \in \mathbb{N}$,

$$S_n := \sum_{i=1}^{n} X_i, \qquad M_n := \frac{S_n}{n} = \frac{\sum_{i=0}^{n} X_i}{n},$$

per n che tende a $+\infty$ le variabili aleatorie S_n e M_n tendono a una variabile aleatoria normale, nel senso che (dette F_{S_n} e F_{M_n} le densità di probabilità delle variabili aleatorie standardizzate associate a S_n e M_n) per ogni $t \in \mathbb{R}$ si ha

$$\lim_{n \to +\infty} F_{S_n}(t) = \lim_{n \to +\infty} F_{M_n}(t) = \mathbf{f}_0(x) = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \text{ (cfr. sez. 2.14)}.$$

Dimostrazione - Omettiamo la dimostrazione di questo teorema.

Osservazione 3.1.2

La condizione C3 esprime formalmente la condizione intuitiva che le varianze σ_i^2 delle X_i "non devono essere troppo piccole". Si può dimostrare che una condizione sufficiente (ma non necessaria!) affinché valga la C3 è che sia

$$\lim_{n \to +\infty} \sigma_i^2 \neq 0 .$$

Osservazione 3.1.3

La prima parte del Teorema Centrale del Limite (quella che riguarda le S_n) spiega perché così tante distribuzioni delle variabili osservate in natura siano (approssimativamente) normali. Infatti i fenomeni descritti da tali variabili sono spesso la somma di moltissimi fattori indipendenti, che possono essere rappresentati con variabili aleatorie che verificano le condizioni C1, C2 e C3: ad esempio, caratteristiche come l'altezza e il peso di un individuo adulto dipendono dall'interazione di molti fattori genetici fra loro indipendenti.

3.2 - Campioni.

Sia $n \in \mathbb{N}$. Un sottoinsieme di n elementi di una popolazione $\mathcal{P}$ si dice un *campione* di $\mathcal{P}$ di ordine n (o anche un $n-$ *campione* di $\mathcal{P}$).
Se X è una variabile statistica su $\mathcal{P}$ (e quindi anche su ogni campione di $\mathcal{P}$), per ogni $n-$ campione C_n di $\mathcal{P}$ indichiamo

- con $X(C_n)$ l'insieme dei valori assunti da X sugli elementi di C_n;
- con $M(X(C_n))$ la media di $X(C_n)$;
- con $V(X(C_n))$ la varianza di $X(C_n)$.

Teorema 3.2.1

Siano μ la media e σ^2 la varianza di una variabile statistica X su una popolazione $\mathcal{P}$. L'insieme delle medie $M(X(C_n))$ al variare di C_n nell'insieme di tutti gli $n-$ campioni di $\mathcal{P}$ ha media μ e varianza $\frac{\sigma^2}{n}$.

Dimostrazione - Omettiamo la dimostrazione di questo teorema.

Sia Ω_n l'insieme di tutti gli $n-$ campioni di $\mathcal{P}$, e sia $X_i : \Omega_n \to \mathbb{R}$ la funzione che ad ogni $C_n \in \Omega_n$ associa il valore $X(\omega_i)$ assunto da X sull'$i-$ simo elemento di C_n; X_i è una variabile aleatoria, e le X_i verificano le ipotesi C1, C2 e C3 del teorema centrale del limite, dunque al crescere di n la variabile aleatoria

$$M_n := \frac{\sum_{i=0}^{n} X_i}{n} = M(X(C_n))$$

si comporta sempre più come una variabile aleatoria con distribuzione normale.

Ne segue che, lavorando su un particolare $n-$campione C_n di $\mathcal{P}$, purché n sia "sufficientemente grande" (e vedremo più avanti che cosa si deve intendere con ciò) la probabilità che $M(X(C_n))$ differisca dalla media μ di $\mathcal{P}$ per più di due (o tre, o...) deviazioni standard può essere calcolata utilizzando la $\boxed{\text{Tabella 2.14.T1}}$. Purché, però, si conosca la deviazione standard a cui tende quella di M_n, ossia (per il teorema 3.2.1) $\frac{\sigma}{\sqrt{n}}$. Generalmente, pr non confondere la deviazione standard σ di X su $\mathcal{P}$ con la deviazione standard di M_n, quest'ultima viene detta *errore standard della media*.

Se in generale $M(X(C_n))$ è una "buona" stima della media μ di X su $\mathcal{P}$ (nel senso che tende a μ al crescere di n), possiamo dire la stessa cosa di $V(X(C_n))$ come stima della varianza σ^2 di X su $\mathcal{P}$? Non proprio.

$\boxed{\text{Teorema 3.2.2}}$

Sia σ^2 la varianza di una variabile statistica X su una popolazione $\mathcal{P}$. L'insieme delle varianze $V(X(C_n))$ al variare di C_n nell'insieme di tutti gli $n-$campioni di $\mathcal{P}$ ha media $\frac{n-1}{n}\sigma^2$ e varianza che tende a zero al tendere di n a $+\infty$.

Dimostrazione - Omettiamo la dimostrazione di questo teorema.

Il teorema 3.2.2 suggerisce di definire la varianza (e quindi la deviazione standard) in modo leggermente diverso per i campioni.

Sia C_n un $n-$campione di $\mathcal{P}$. Si dice *varianza campionaria* dell'insieme

$$X(C_n) = \{x_1, x_2, ..., x_n\}$$

dei valori assunti dalla variabile statistica X sul campione C_n di ordine n il numero

$$s^2(X(C_n)) := \frac{\sum_{i=1}^{n} (M(X(C_n)) - x_i)^2}{n-1} = \frac{n}{n-1} V(X(C_n)).$$

Alla luce di questa definizione, il teorema 3.2.2 può essere così riformulato:

$\boxed{\text{Teorema 3.2.3}}$

Sia σ^2 la varianza di una variabile statistica X su una popolazione $\mathcal{P}$. L'insieme delle varianze campionarie $s^2(X(C_n))$ al variare di C_n nell'insieme di tutti gli $n-$campioni di $\mathcal{P}$ ha media σ^2 e varianza che tende a zero al tendere di n a $+\infty$.

Possiamo dunque utilizzare la varianza campionaria di una variabile statistica per stimare la varianza (e quindi la deviazione standard) della stessa variabile statistica su tutta la popolazione.

Tutto quanto si è detto fin qui vale, come si è osservato, purché n sia “sufficientemente grande”. Gli statistici concordano generalmente che la distribuzione della media campionaria si può considerare normale, e la varianza campionaria si può assumere come valida stima della varianza sull’intera popolazione, **tutte le volte che $\boldsymbol{n \geq 30}$**.

Esempio 3.2.4

Lavorando con un campione di 30 cavie (su una popolazione $\mathcal{P}$ assai più ampia) abbiamo rilevato i seguenti pesi:

$$28, 32, 37, 29, 31, 30, 32, 26, 32, 27, 29, 30, 28, 31, 31,$$

$$30, 34, 39, 31, 33, 32, 32, 28, 30, 25, 27, 28, 26, 29, 29\,.$$

Che cosa possiamo dire circa la media della variabile statistica “peso” X sull’intera popolazione $\mathcal{P}$?

Poiché il nostro campione è formato da 30 elementi, possiamo supporre che la media campionaria abbia distribuzione normale e possiamo stimare la deviazione standard di X su $\mathcal{P}$ mediante la varianza campionaria.

La media calcolata sul campione è $\frac{906}{30} = 30,2$; la varianza campionaria è

$$\frac{\sum_{i=1}^{30} (30{,}2 - x_i)^2}{29} = \frac{2728}{290} \approx 9,4$$

cosicchè $\sigma \approx 3{,}07$ e l’errore standard della media (cioè la deviazione standard della media campionaria) è (per il teorema 3.2.1)

$$\frac{\sigma}{\sqrt{n}} \approx \frac{3{,}07}{5{,}477} \approx 0{,}56\,.$$

La Tabella 2.14.T1 ci dice che è inferiore a $0{,}0456$ la probabilità che la media calcolata su tutta la popolazione differisca da quella del nostro campione per più di 2 errori standard; in altri termini il peso medio calcolato su tutta la popolazione rientra (al 95, 44 %) nell’intervallo

$$[29,08;\ 31,32]\,.$$

Non è un gran risultato: l’errore commesso (al 95, 44 %) è fino a $1,12$ cioè (espresso in percentuale del valore che cerchiamo) fino al 3,7 %. Se vogliamo dati più precisi, però, non abbiamo altra scelta che operare su un campione più numeroso.

3.3 - Test di ipotesi.

Vogliamo adesso mostrare come si possono utilizzare le informazioni su una variabile statistica che abbiamo ricavato da un campione della popolazione per accettare o rifiutare ipotesi su affermazioni che riguardano proprietà di tale variabile su tutta la popolazione.

Ad esempio: vogliamo capire, esaminando un campione, se le confezioni di caffè prodotte da una certa ditta contengono davvero il peso dichiarato in etichetta; oppure: vogliamo decidere se il tasso alcoolico di una certa marca di vino corrisponde davvero a quello dichiarato in etichetta. E così via.

Poiché le medie campionarie (su campioni di almeno 30 elementi) seguono la distribuzione normale, la tabella solita ci permette di calcolare la probabilità che la media del nostro campione si discosti dalla media vera (che è il valor medio di tutte le medie campionarie) per più di un tot prefissato, oppure la probabilità che la media del nostro campione superi la media vera per più di un tot prefissato, et similia.

Nei test di ipotesi si segue il procedimento inverso: prima si fissa la probabilità α, poi ci si chiede se (ad esempio) il valore per cui la media del nostro campione si discosta dalla media vera rientra nell'intervallo in cui sarebbe a priori caduto (o non sarebbe a priori caduto, a seconda di quel che cerchiamo) con probabilità α.

In pratica, si procede come segue.

1. – Si fissa l'ipotesi da sottoporre a verifica (detta *ipotesi nulla*).

2. – Se ne deduce la negazione (detta *ipotesi alternativa*).

3. – Si sceglie un *livello di significatività* α per il test: α indica la probabilità di rifiutare l'ipotesi nulla quando invece essa è vera. La probabilità invece di non rifiutare una ipotesi nulla che in realtà è falsa si dice *rischio del consumatore*. Si può dimostrare che, fissata l'ampiezza del campione, al diminuire del livello di significatività cresce il rischio del consumatore, e viceversa: questo va tenuto presente nel fissare il valore di α.

3. – Si determina (anche in funzione di α) un intervallo (o due intervalli) di valori per la nostra variabile statistica che non si dovrebbero ottenere qualora l'ipotesi nulla fosse vera (*regione di rifiuto*) e il complementare (*regione di accettazione*). Se la regione di rifiuto è costituita da un intervallo, il test si dice *a una coda*; se invece è costituita da due intervalli, il test si dice *a due code*.

4. – Si esegue il test sui dati del campione a nostra disposizione, e si traggono le dovute conseguenze.

Esempio 3.3.1

Una ditta produttrice di lampadine sostiene che la durata media di ogni lampadina prodotta è 1600 ore.

Estraendo un campione di 100 lampadine, si è trovata una durata media di 1570 ore e una varianza campionaria di 14400 (ore^2). Possiamo fidarci del produttore?

Se le lampadine durano più di quanto dichiarato, noi siamo contenti. Dunque porremo:

ipotesi nulla: $\mu \geq 1600$.

ipotesi alternativa: $\mu < 1600$.

Scegliamo un livello di significatività di 0,05. Questo significa che cerchiamo un intervallo di valori $(-\infty, y_0)$ (per la variabile aleatoria normale standard Y) tale che $p(\mathrm{Y} < y_0) < 0,05$. Il valore di y_0 lo troviamo leggendo "alla rovescia" (rispetto a come ci eravamo abituati a fare) la Tabella 2.14.T1 : esso è (circa) $-1,65$. Dunque rifiuteremo l'ipotesi nulla (con livello di significatività $0,05$) se il valore standardizzato della media campionaria da noi rilevato non appartiene all'intervallo di accettazione $(-1,65;\ +\infty)$. Tale valore standardizzato si calcola ricordando che la media delle medie campionarie è la media μ dell'intera popolazione (cioè 1600 in base all'ipotesi nulla) mentre la deviazione standard delle medie campionarie (cioè l'errore standard della media) è $\frac{\sigma}{\sqrt{n}}$. A sua volta, σ si stima attraverso la varianza campionaria, dunque $\sigma = 120$. Il valore standardizzato della media da noi rilevata è dunque

$$\frac{\text{(Media rilevata)}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{1570-1600}{\frac{120}{10}} = -\frac{30}{12} = -2,5\,.$$

Poiché tale valore non appartiene all'intervallo di accettazione $(-1,65;\ +\infty)$, rifiutiamo l'ipotesi nulla al livello di significatività di $0,05$.

Decidiamo dunque di intraprendere una *class action* contro la ditta produttrice delle lampadine, ma il giudice competente ci chiede per scrupolo se l'ipotesi nulla verrebbe rifiutata anche a un livello di significatività di $0,01$. Livello di significatività più basso, ricordiamo, significa minore probabilità di rifiutare erroneamente l'ipotesi nulla (dunque la richiesta del giudice è a garanzia degli interessi della ditta produttrice!).

Per rispondere alla domanda del giudice, i calcoli sono esattamente gli stessi: cambia soltanto l'intervallo di accettazione (cioè il valore y_0 che lo individua). Sempre dalla Tabella 2.14.T1 ricaviamo il nuovo valore di y_0, che è "grosso modo" a metà strada fra $-2,30$ e $-2,35$; quindi possiamo porre $y_0 := -2,325$. Ancora una volta il valore standardizzato della media da noi rilevata non appartiene all'intervallo di accettazione $(-2,325;\ +\infty)$, quindi possiamo rifiutare l'ipotesi nulla anche al livello di significatività $0,01$.

Si arriva così al processo. Ma l'avvocato che difende la ditta produttrice delle lampadine presenta un'obiezione: l'ipotesi nulla non può essere $\mu \geq 1600$ ma piuttosto, come dichiarato sulle confezioni, $\mu = 1600$. Che cosa cambia, se vogliamo rifiutarla (come stabilito dal giudice) al livello di significatività di $0,01$?

Questa volta la regione di rifiuto è formata da due intervalli. Questo significa che cerchiamo due intervalli di valori $(-\infty,\ -y_0)$ e $(y_0,\ +\infty)$ (per la variabile aleatoria normale standard Y) tale che $p(\mathrm{Y} < -y_0) + p(y_0 < \mathrm{Y}) < 0,01$. Con le usuali condsiderazioni di simmetria sulla distribuzione normale standard, cerchiamo nella Tabella 2.14.T1 un y_0 tale che $p(y_0 < \mathrm{Y}) < 0,005$, ossia tale che $p(0 \leq \mathrm{Y} \leq y_0) = 0,495$. Si trova $y_0 \approx 2,58$. L'intervallo di accettazione è dunque $(-2,58;\ 2,58)$. Il valore standardizzato della media da noi rilevato è $-2,5$, quindi appartiene a tale intervallo: non possiamo rifiutare l'ipotesi nulla ($\mu = 1600$) al livello di significatività di $0,01$ e il giudice manda assolta la ditta produttrice delle lampadine. Si noti che il risultato del test *non conferma* l'ipotesi nulla: semplicemente, smentisce (al livello di significatività di $0,01$) l'ipotesi alternativa; i dati sono compatibili con l'ipotesi nulla, ma potrebbero essere compatibili anche con altre, diverse ipotesi.

3.4 - Teoria del test Z.

Il test che abbiamo applicato nell'esempio 3.3.1 si chiama test Z, e trova applicazione tutte le volte che la varianza σ^2 della popolazione $\mathcal{P}$ è nota, oppure può essere stimata con la varianza campionaria; quest'ultima condizione si traduce operativamente nel fatto che il campione deve consistere di almeno 30 esemplari.

Sotto queste ipotesi, possiamo formalizzare l'algoritmo del test distinguendo sostanzialmente tre casi.

Primo Caso: L'ipotesi nulla è $\mu \leq \mu_0$, l'ipotesi alternativa è $\mu > \mu_0$.

Il test è "a una coda". Sia $\mu_{\mathbf{X}}$ la media della variabile statistica X ricavata dal nostro campione, e sia σ^2 la varianza di X sulla popolazione $\mathcal{P}$ (nota per qualche motivo, oppure stimata mediante la varianza campionaria calcolata sul nostro campione). Sia α il livello di significatività scelto per il test.

Si determina dalla $\boxed{\text{Tabella 2.14.T1}}$ un valore y_0 tale che (indicata con Y la variabile aleatoria normale standard)

$$p(y_0 < \mathrm{Y}) < \alpha$$

ossia tale che

$$p(0 \leq \mathrm{Y} \leq y_0) = 0,5 - \alpha\,.$$

L'intervallo di rifiuto è $(y_0,\ +\infty)$. Se

$$\frac{\mu_{\mathbf{X}} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > y_0$$

(cioè se il valore standardizzato di $\mu_{\mathbf{X}}$ cade nell'intervallo di rifiuto) rifiutiamo l'ipotesi nulla al livello di significatività α.

ATTENZIONE! Nel caso opposto, cioè se il valore standardizzato di $\mu_{\mathbf{X}}$ cade nell'intervallo di accettazione, *non si dice* che "accettiamo l'ipotesi nulla" ma piuttosto che "non rifiutiamo l'ipotesi nulla".

Secondo Caso: L'ipotesi nulla è $\mu \geq \mu_0$, l'ipotesi alternativa è $\mu < \mu_0$.

Questo secondo caso è del tutto analogo al primo. Il test è ancora "a una coda". Sia $\mu_{\mathbf{X}}$ la media della variabile statistica X ricavata dal nostro campione, e sia σ^2 la varianza di X sulla popolazione $\mathcal{P}$ (nota per qualche motivo, oppure stimata mediante la varianza campionaria calcolata sul nostro campione). Sia α il livello di significatività scelto per il test.

Si determina dalla $\boxed{\text{Tabella 2.14.T1}}$ un valore y_0 tale che (indicata con Y la variabile aleatoria normale standard)

$$p(y_0 < \mathrm{Y}) < \alpha$$

ossia tale che

$$p(0 \leq \mathrm{Y} \leq y_0) = 0,5 - \alpha\,.$$

L'intervallo di rifiuto è $(-\infty, -y_0)$. Se

$$\frac{\mu_{\mathbf{X}}-\mu_0}{\frac{\sigma}{\sqrt{n}}} < -y_0$$

(cioè se il valore standardizzato di $\mu_{\mathbf{X}}$ cade nell'intervallo di rifiuto) rifiutiamo l'ipotesi nulla al livello di significatività α.

ATTENZIONE! Nel caso opposto, cioè se il valore standardizzato di $\mu_{\mathbf{X}}$ cade nell'intervallo di accettazione, *non si dice* che "accettiamo l'ipotesi nulla" ma piuttosto che "non rifiutiamo l'ipotesi nulla".

Terzo Caso: L'ipotesi nulla è $\mu = \mu_0$, l'ipotesi alternativa è $\mu \neq \mu_0$.

Questa volta il test è "a due code". Come al solito, sia $\mu_{\mathbf{X}}$ la media della variabile statistica X ricavata dal nostro campione, e sia σ^2 la varianza di X sulla popolazione $\mathcal{P}$ (nota per qualche motivo, oppure stimata mediante la varianza campionaria calcolata sul nostro campione). Sia α il livello di significatività scelto per il test.

Si determina dalla Tabella 2.14.T1 un valore y_0 tale che (indicata con Y la variabile aleatoria normale standard)

$$p(y_0 < \mathrm{Y}) < \frac{\alpha}{2}$$

ossia tale che

$$p(0 \leq \mathrm{Y} \leq y_0) = 0,5 - \frac{\alpha}{2} .$$

La regione di rifiuto è $(-\infty, -y_0) \cup (y_0, +\infty)$. Se

$$\frac{\mu_{\mathbf{X}}-\mu_0}{\frac{\sigma}{\sqrt{n}}} < -y_0 \quad \text{oppure} \quad \frac{\mu_{\mathbf{X}}-\mu_0}{\frac{\sigma}{\sqrt{n}}} > y_0$$

(cioè se il valore standardizzato di $\mu_{\mathbf{X}}$ cade nella regione di rifiuto) rifiutiamo l'ipotesi nulla al livello di significatività α.

ATTENZIONE! Nel caso opposto, cioè se il valore standardizzato di $\mu_{\mathbf{X}}$ cade nell'intervallo di accettazione, *non si dice* che "accettiamo l'ipotesi nulla" ma piuttosto che "non rifiutiamo l'ipotesi nulla".

3.5 - Teoria del test t ("di Student").

Come possiamo procedere quando il numero di elementi del campione è troppo piccolo perché possiamo utilizzare la varianza campionaria come stima per la varianza della variabile statistica sull'intera popolazione?

Il problema fu affrontato da William S. Gosset (1876-1937), un dipendente della fabbrica di birra irlandese *Guinness*. Egli riuscì a determinare la densità di probabilità associata alla variabile aleatoria "media campionaria" sui campioni di n elementi (con $n < 30$) ***nell'ipotesi che la variabile statistica di riferimento sulla popolazione $\mathcal{P}$ abbia distribuzione normale*** (si ricordi che, grazie al teorema centrale del limite, nel caso del test Z una siffatta ipotesi non era affatto necessaria).

Sia $\nu \in \mathbb{R}$. Una variabile aleatoria *ha la distribuzione* t *di Student con* ν *gradi di libertà* se la sua densità di probabilità è

$$\mathbf{f}(y) := \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

dove Γ è la funzione $\mathbb{R}\to\mathbb{R}$ definita da

$$\Gamma(x) := \int_0^{+\infty} t^{x-1}e^{-t}dt$$

La funzione $\Gamma(x)$ che compare nella definizione di **f** è nota come "funzione gamma di Eulero". Non è difficile osservare (ragionando "per parti" sulle primitive di $t^{x-1}e^{-t}$) che

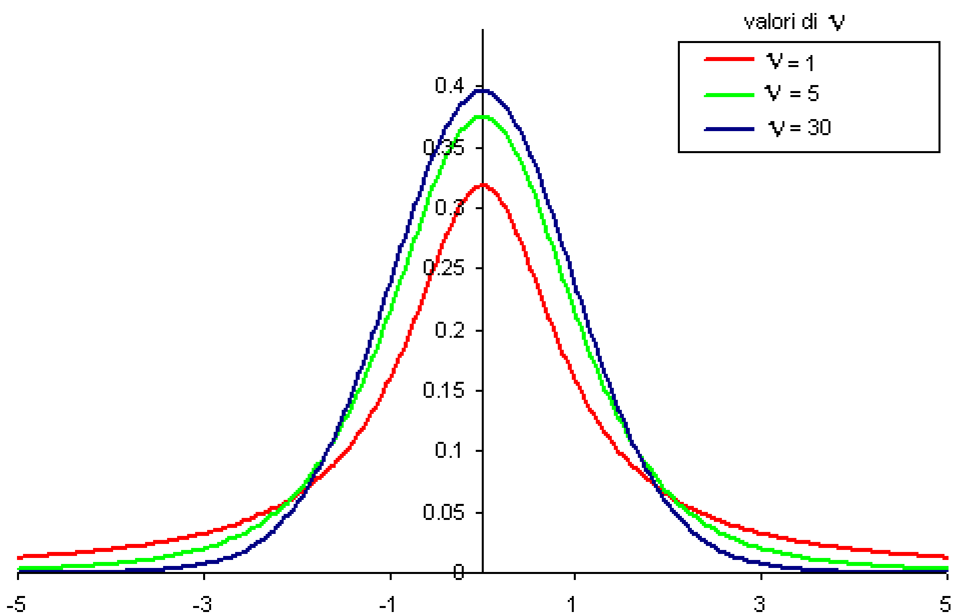
$$\Gamma(x) = (x-1)\Gamma(x-1)\,;$$

inoltre, è facile verificare che $\Gamma(1) = 1$. Per induzione, si prova agevolmente che per ogni $n \in \mathbb{N}$ si ha

$$\Gamma(n) = (n-1)!\,.$$

Si dimostra che una variabile aleatoria che abbia la distribuzione t di Student con ν gradi di libertà ha valor medio 0 e varianza $\frac{\nu}{\nu-2}$. Al tendere di ν a $+\infty$, tale variabile aleatoria tende a una variabile aleatoria con distribuzione normale standard.

Questo è il grafico della densità di probabilità per le distribuzioni t di Student con 1, 5 e 30 gradi di libertà:



Ai fini pratici, si fa riferimento a tabelle analoghe alla Tabella 2.14.T1, una per ogni valore di ν $(1 \leq \nu \leq 29)$. Addirittura, per l'applicazione del test t (formalmente del tutto analoga a quella del test Z) è sufficiente un'unica tabella che fornisca in funzione del grado di libertà ν i "valori critici" y_0 corrispondenti ai livelli di significatività standard ($\alpha := 0,05$ e $\alpha := 0,01$).

Tabella 3.5.T1

ν	$\alpha = 0,05$		$\alpha = 0,01$	
	test "a una coda" (y_0 tale che $p(y>y_0)=\alpha$)	test "a due code" (y_0 tale che $p(y>y_0)=\frac{\alpha}{2}$)	test "a una coda" (y_0 tale che $p(y>y_0)=\alpha$)	test "a due code" (y_0 tale che $p(y>y_0)=\frac{\alpha}{2}$)
1	6,314	12,706	31,821	63,657
2	2,920	4,303	6,965	9,925
3	2,353	3,182	4,541	5,841
4	2,132	2,776	3,747	4,604
5	2,015	2,571	3,365	4,032
6	1,943	2,447	3,143	3,707
7	1,895	2,365	2,998	3,499
8	1,860	2,306	2,896	3,355
9	1,833	2,262	2,821	3,250
10	1,812	2,228	2,764	3,169
11	1,796	2,201	2,718	3,106
12	1,782	2,179	2,681	3,055
13	1,771	2,160	2,650	3,012
14	1,761	2,145	2,624	2,977
15	1,753	2,131	2,602	2,947
16	1,746	2,120	2,583	2,921
17	1,740	2,110	2,567	2,898
18	1,734	2,101	2,552	2,878
19	1,729	2,093	2,539	2,861
20	1,725	2,086	2,528	2,845
21	1,721	2,080	2,518	2,831
22	1,717	2,074	2,508	2,819
23	1,714	2,069	2,500	2,807
24	1,711	2,064	2,492	2,797
25	1,708	2,060	2,485	2,787
26	1,706	2,056	2,479	2,779
27	1,703	2,052	2,473	2,771
28	1,701	2,048	2,467	2,763
29	1,699	2,045	2,462	2,756

Esempio 3.5.1

Una ditta produttrice di vino mette il commercio il suo prodotto in bottiglie del contenuto dichiarato di 750 ml.

Dal controllo effettuato su un campione di 6 bottiglie risultano i seguenti valori (in ml.):

751,0 746,5 746,0 750,5 751,0 749,0 .

Vogliamo stabilire, con un livello di significatività del 5%, se il contenuto medio delle bottiglie è quello dichiarato.

Poniamo:

ipotesi nulla: $\mu = 750,0$.

ipotesi alternativa: $\mu \neq 750,0$.

La media campionaria da noi rilevata è $\mu_{\mathbf{X}} = \frac{4494}{6} = 749,0$, la varianza campionaria ad essa relativa è $s_{\mathbf{X}} = \frac{25{,}5}{5} = 5,1$.

La regione di rifiuto, ricavata dalla tabella Tabella 2.14.T1 per $\nu := 5$, $\alpha := 0,05$ e test a due code, è $(-\infty, -2,571) \cup (2,571, +\infty)$. Il valore standardizzato della media da noi rilevata è

$$\frac{\mu_{\mathbf{X}} - \mu}{\frac{s_{\mathbf{X}}}{\sqrt{n}}} = \frac{749 - 750}{\frac{\sqrt{5{,}1}}{\sqrt{6}}} \approx -\frac{1}{0{,}922} \approx -1,08 \, .$$

Poiché tale valore appartiene all'intervallo di accettazione $(-2,571; +2,571)$, non rifiutiamo l'ipotesi nulla al livello di significatività di $0,05$: non c'è evidenza significativa che il contenuto medio differisca da quello dichiarato in etichetta.