Indice

Introduzione

Uno dei concetti più importanti della Statistica è quello di "popolazione". Con questo vocabolo si intende l'insieme delle unità (statistiche) che si vogliono esaminare. Il concetto di popolazione assume un ruolo importante nella partizione della statistica in *statistica descrittiva* e *statistica induttiva*.

Come è già stato ampiamente precisato nel primo corso di statistica, la statistica descrittiva si occupa di descrivere con opportuni metodi (procedure) le caratteristiche più salienti dei fenomeni (caratteri), oggetto di indagine, rilevabili sulle singole unità statistiche. Queste descrizioni non sono altro che processi di sintesi dei "dati" ricavabili dalle unità statistiche a disposizione. Se si suppone di avere a disposizione sei unità statistiche in corrispondenza delle quali il carattere quantitativo X assume i valori 3, 4, 6, 7, 9 e 10, il calcolo della loro media aritmetica, che è pari a

$$\frac{1}{6}(3+4+6+7+9+10)=\frac{39}{6}=6,5$$

è un chiaro esempio di sintesi statistica.

Le unità statistiche a disposizione possono essere tutte le unità della popolazione di interesse o una parte (un campione) delle stesse. Il valore della sintesi che si effettua sui dati di un campione raramente coincide con il valore della sintesi relativa all'intera popolazione. La statistica induttiva (o inferenza statistica) predispone i metodi per estendere (generalizzare) la sintesi effettuata sui dati campionari alla sintesi relativa alla totalità della popolazione di interesse. Questa estensione implica un margine di incertezza. Il ricorso alla rilevazione parziale (campionaria) fa diminuire i costi ed anche il tempo per lo svolgimento della rilevazione. Per poter effettuare la rilevazione campionaria bisogna prima scegliere la modalità di estrazione delle unità dalla popolazione e successivamente bisogna determinare il numero n di unità da estrarre. Questi due aspetti hanno notevole influenza sul margine di incertezza che si ha nella generalizzazione dalla sintesi campionaria alla sintesi dell'intera popolazione. È opportuno precisare subito che le generalizzazioni sono possibili solo se le estrazioni, delle n unità statistiche dalla popolazione avvengono in modo casuale, ovvero se tutte le unità della popolazione hanno la "stessa possibilità" di essere estratte e far così parte del campione.

Per meglio chiarire quanto sinora esposto si riprenda l'esempio precedente: la popolazione di interesse è costituita da $N = 6$ unità statistiche $U_1, U_2, \ldots, U_6$ ed il carattere X assume in corrispondenza delle sei unità statistiche rispettivamente i valori: 3, 4, 6, 7, 9, 10. Nella popolazione la media aritmetica μ del carattere X è pari a

$$\frac{1}{6}(3 + 4 + 6 + 7 + 9 + 10) = \frac{39}{6} = 6,5\ .$$

Si supponga ora di estrarre a caso un campione di $n = 2$ unità statistiche. Si supponga altresì che le estrazioni avvengano con le due seguenti modalità:

i) *con riposizione*. Si estrae innanzitutto, a caso, una unità fra le sei, si annota il valore di X ottenuto e si ripone l'unità estratta nella popolazione che così viene ricostituita. Quindi si estrae, a caso, dalla popolazione una seconda unità e si annota il valore di X così ottenuto.

ii) *senza riposizione*. Si estrae innanzitutto, a caso, una unità fra le sei e si annota il valore di X ottenuto. Quindi dalle restanti $N - 1 = 5$ unità della popolazione si estrae a caso una seconda unità e si annota il valore di X così ottenuto.

Si riportano, rispettivamente nella Tabella 1.1 e nella Tabella 1.2, le due liste dei campioni possibili che si ottengono con le due modalità di estrazione. Con x_1 si

$x_1 \backslash x_2$	3	4	6	7	9	10
3	$(3,3)$	$(3,4)$	$(3,6)$	$(3,7)$	$(3,9)$	$(3,10)$
4	$(4,3)$	$(4,4)$	$(4,6)$	$(4,7)$	$(4,9)$	$(4,10)$
6	$(6,3)$	$(6,4)$	$(6,6)$	$(6,7)$	$(6,9)$	$(6,10)$
7	$(7,3)$	$(7,4)$	$(7,6)$	$(7,7)$	$(7,9)$	$(7,10)$
9	$(9,3)$	$(9,4)$	$(9,6)$	$(9,7)$	$(9,9)$	$(9,10)$
10	$(10,3)$	$(10,4)$	$(10,6)$	$(10,7)$	$(10,9)$	$(10,10)$

Tabella 1.1: *Estrazioni con riposizione (lista campioni possibili).*

è indicato il primo valore estratto e con x_2 il secondo valore estratto.

Nella Tabella 1.1 la coppia $(x_1 = 6; x_2 = 7)$ indica uno degli $N^2 = 36$ possibili campioni ottenibili con questa modalità di estrazione. Una casella del prospetto indica un campione possibile.

$x_1 \backslash x_2$	3	4	6	7	9	10
3	/	$(3,4)$	$(3,6)$	$(3,7)$	$(3,9)$	$(3,10)$
4	$(4,3)$	/	$(4,6)$	$(4,7)$	$(4,9)$	$(4,10)$
6	$(6,3)$	$(6,4)$	/	$(6,7)$	$(6,9)$	$(6,10)$
7	$(7,3)$	$(7,4)$	$(7,6)$	/	$(7,9)$	$(7,10)$
9	$(9,3)$	$(9,4)$	$(9,6)$	$(9,7)$	/	$(9,10)$
10	$(10,3)$	$(10,4)$	$(10,6)$	$(10,7)$	$(10,9)$	/

Tabella 1.2: *Estrazioni senza riposizione (lista campioni possibili).*

Nel caso delle estrazioni senza riposizione (Tabella 1.2) i campioni possibili risultano $N^2 - N = 36 - 6 = 30$, in quanto l'unità estratta nella prima estrazione non può più presentarsi sulla seconda estrazione e quindi, rispetto alle estrazioni con riposizione, vengono a mancare gli N campioni che si trovano nelle caselle barrate.

Si può calcolare ora, per ogni campione estraibile, la media campionaria $\bar{x} = \frac{1}{2}(x_1 + x_2)$. Si hanno così 36 medie campionarie nel caso di estrazioni con riposizione e 30 medie campionarie nel caso di estrazioni senza riposizione.

Campione (x_1, x_2)	Media campionaria $\bar{x}$	Campione (x_1, x_2)	Media campionaria $\bar{x}$	Campione (x_1, x_2)	Media campionaria $\bar{x}$
$(3,3)$	3	$(6,3)$	$4,5$	$(9,3)$	6
$(3,4)$	$3,5$	$(6,4)$	5	$(9,4)$	$6,5$
$(3,6)$	$4,5$	$(6,6)$	6	$(9,6)$	$7,5$
$(3,7)$	5	$(6,7)$	$6,5$	$(9,7)$	8
$(3,9)$	6	$(6,9)$	$7,5$	$(9,9)$	9
$(3,10)$	$6,5$	$(6,10)$	8	$(9,10)$	$9,5$
$(4,3)$	$3,5$	$(7,3)$	5	$(10,3)$	$6,5$
$(4,4)$	4	$(7,4)$	$5,5$	$(10,4)$	7
$(4,6)$	5	$(7,6)$	$6,5$	$(10,6)$	8
$(4,7)$	$5,5$	$(7,7)$	7	$(10,7)$	$8,5$
$(4,9)$	$6,5$	$(7,9)$	8	$(10,9)$	$9,5$
$(4,10)$	7	$(7,10)$	$8,5$	$(10,10)$	10

Tabella 1.3: *Estrazioni con riposizione (medie campionarie).*

Raggruppando le medie campionarie che hanno lo stesso valore si ottiene la seguente distribuzione delle medie campionarie (Tabella 1.4).

Media campionaria $\bar{x}$	Numero campioni $P(\bar{x})$
3	1
$3,5$	2
4	1
$4,5$	2
5	4
$5,5$	2
6	3
$6,5$	6
7	3
$7,5$	2
8	4
$8,5$	2
9	1
$9,5$	2
10	1
Numero totale campioni	36

Tabella 1.4: *Estrazioni con riposizione (distribuzione delle medie campionarie).*

Nel caso delle estrazioni senza riposizione la distribuzione delle 30 medie campionarie risulta:

Media campionaria $\bar{x}$	Numero campioni $P(\bar{x})$
$3,5$	2
$4,5$	2
5	4
$5,5$	2
6	2
$6,5$	6
7	2
$7,5$	2
8	4
$8,5$	2
$9,5$	2
Numero totale campioni	30

Tabella 1.5: *Estrazioni senza riposizione (distribuzione delle medie campionarie).*

Le due distribuzioni delle medie campionarie sono differenti.

È opportuno far presente ora che sia nel caso di estrazioni con riposizione che nel caso di estrazioni senza riposizione, mentre è possibile ricavare, con procedimento deduttivo, la lista dei campioni possibili e quindi la distribuzione delle medie campionarie, non è invece possibile prevedere quale campione della lista si otterrà a seguito di una reale estrazione campionaria. In altre parole l'operazione di estrazione di un campione di n unità appartiene ai cosiddetti "esperimenti casuali". Ovvero a quegli esperimenti per i quali il risultato di una reale sperimentazione può variare nell'ambito di una lista di risultati possibili.

Per rendere più chiara l'esemplificazione si supporrà che con l'estrazione del campione si vogliano *avere informazioni sul valore della media μ della popolazione*. Si osservi che il valore di μ non deve essere noto, altrimenti non avrebbe senso estrarre il campione. Forse la cosa più ragionevole che si può fare è quella di assegnare all'ignoto μ il valore della media aritmetica del campione estratto. Come si è mostrato in precedenza, però, il valore della media campionaria solitamente varia da un campione all'altro. Inoltre, i valori delle medie campionarie raramente coincidono con il valore vero (ma ignoto di $\mu = 6,5$). Il campione estratto, potendo essere uno dei 36 campioni possibili nel caso di estrazioni con riposizione (ovvero uno qualsiasi dei 30 campioni nel caso di estrazioni senza riposizione), può dare informazioni su μ molto difformi dal vero valore di quest'ultimo. Ciò non di meno, il metodo statistico permette, se il campione è estratto a caso, di valutare anche lo *scostamento che "mediamente"* si ha tra il valore della media campionaria ed il valore di μ. È altresì possibile ricavare, utilizzando il valore $\bar{x}$ della media del campione (realmente estratto) e lo scostamento medio delle medie campionarie, un intervallo nel quale "molto verosimilmente" è compreso l'ignoto valore di μ. Questa procedura è basata, tra l'altro, sulla conoscenza delle caratteristiche della distribuzione delle medie campionarie.

Per approfondire alcune delle affermazioni sino ad ora fatte si può calcolare la media $M_1\left(\bar{X}\right)$ di tutte le medie campionarie nonché la loro varianza $Var\left(\bar{X}\right)$ sia per le estrazioni con riposizione che per quelle senza riposizione. Dal prospetto in Tabella 1.6 si ottiene: $M_1\left(\bar{X}\right) = \dfrac{234}{36} = 6,5$; $Var\left(\bar{X}\right) = \dfrac{112,5}{36} = 3,125$. Si constata così che $M_1\left(\bar{X}\right)$, ovvero la media di tutte le 36 medie campionarie, coincide con il valore della media che il carattere X assume nella popolazione ($\mu = M_1(X) = 6,5$). La varianza $Var(X)$ del carattere X nella popolazione è pari a $6,25$. Si riscontra che $Var\left(\bar{X}\right) = \dfrac{Var(X)}{2}$. In seguito si dimostrerà che, nel caso di campionamento con riposizione:

$$\begin{cases} M_1\left(\bar{X}\right) = \mu \\ Var\left(\bar{X}\right) = \dfrac{Var(X)}{n} \end{cases},$$

essendo n la numerosità del campione.

$\bar{x}$	$P(\bar{x})$	$\bar{x} \cdot P(\bar{x})$	$\left[\bar{x} - M_1\left(\bar{X}\right)\right]$	$\left[\bar{x} - M_1\left(\bar{X}\right)\right]^2 \cdot P(\bar{x})$
3	1	3	$-3,5$	$12,25$
$3,5$	2	7	$-3,0$	$18,00$
4	1	4	$-2,5$	$6,25$
$4,5$	2	9	$-2,0$	$8,00$
5	4	20	$-1,5$	$9,00$
$5,5$	2	11	$-1,0$	$2,00$
6	3	18	$-0,5$	$0,75$
$6,5$	6	39	$0,0$	$0,00$
7	3	21	$+0,5$	$0,75$
$7,5$	2	15	$+1,0$	$2,00$
8	4	32	$+1,5$	$9,00$
$8,5$	2	17	$+2,0$	$8,00$
9	1	9	$+2,5$	$6,25$
$9,5$	2	19	$+3,0$	$18,00$
10	1	10	$+3,5$	$12,25$
Totali	36	234		$112,50$

Tabella 1.6: *Calcolo di $M_1\left(\bar{X}\right)$ e $Var\left(\bar{X}\right)$ nel caso di estrazioni con riposizione.*

Ecco ora il calcolo di $M_1\left(\bar{X}\right)$ e $Var\left(\bar{X}\right)$ nel campionamento senza riposizione.

$\bar{x}$	$P(\bar{x})$	$\bar{x} \cdot P(\bar{x})$	$\left[\bar{x} - M_1\left(\bar{X}\right)\right]$	$\left[\bar{x} - M_1\left(\bar{X}\right)\right]^2 \cdot P(\bar{x})$
$3,5$	2	7	$-3,0$	$18,00$
$4,5$	2	9	$-2,0$	$8,00$
5	4	20	$-1,5$	$9,00$
$5,5$	2	11	$-1,0$	$2,00$
6	2	12	$-0,5$	$0,50$
$6,5$	6	39	$0,0$	$0,00$
7	2	14	$+0,5$	$0,50$
$7,5$	2	15	$+1,0$	$2,00$
8	4	32	$+1,5$	$9,00$
$8,5$	2	17	$+2,0$	$8,00$
$9,5$	2	19	$+3,0$	$18,00$
Totali	30	195		$75,00$

Tabella 1.7: *Calcolo di $M_1\left(\bar{X}\right)$ e $Var\left(\bar{X}\right)$ nel caso di estrazioni senza riposizione.*

Dal prospetto in Tabella 1.7 si ottiene: $M_1\left(\bar{X}\right) = \dfrac{195}{30} = 6,5$; $Var\left(\bar{X}\right) = \dfrac{75}{30} = 2,5$. Anche con questo tipo di estrazioni la media dell'insieme delle 30 medie campionarie coincide con $\mu = M_1\left(X\right)$. Inoltre la variabilità delle medie campionarie risulta inferiore a quella che si ottiene con il campionamento con riposizione. In effetti si dimostra che nel campionamento senza riposizione:

$$\begin{cases} M_1\left(\bar{X}\right) = \mu \\ Var\left(\bar{X}\right) = \dfrac{Var(X)}{n} \cdot \dfrac{N-n}{N-1} \end{cases},$$

essendo N la numerosità della popolazione ed n quella del campione.

L'esempio mostra che le due procedure campionarie, mentre forniscono lo stesso valor medio delle $\bar{X}$, differiscono invece per la variabilità delle medie campionarie. In particolare, la varianza delle $\bar{X}$ risulta più piccola nel campionamento senza riposizione

$$\frac{Var(X)}{n} \cdot \frac{N-n}{N-1} \quad < \quad \frac{Var(X)}{n} \qquad\qquad (n \geq 2) \ .$$

Si possono ora trarre le prime conclusioni. Innanzitutto si è sottolineato che l'estrazione di un campione "casuale" dà luogo a risultati che variano da un campione all'altro e che l'estensione dai dati del campione a quelli della popolazione di interesse si basa anche sulla conoscenza delle distribuzioni campionarie. Le distribuzioni campionarie – nell'esempio si è considerato la distribuzione delle medie campionarie – sono generate dallo schema di campionamento. Lo schema di campionamento non è altro che un esperimento casuale, cioè un esperimento del quale si può conoscere l'insieme dei risultati possibili ma non si può prevedere quale si verificherà. Prima di poter fare le generalizzazioni di cui sopra bisogna familiarizzare con quella parte della matematica "applicata" che studia gli esperimenti casuali, cioè con il calcolo delle probabilità, in quanto, si perdoni la ripetizione, l'estrazione di un campione non è altro che un esperimento casuale.

Con il calcolo delle probabilità si imparerà a ricavare la distribuzione campionaria delle medie campionarie e poi tramite la stessa si effettueranno le generalizzazioni dal campione alla popolazione. In altre parole con la "singola media campionaria" non si è in grado di effettuare la generalizzazione, ma con la conoscenza del valore della media campionaria unita alla conoscenza delle "caratteristiche" della distribuzione delle medie campionarie è possibile ricavare utili informazioni sulla media μ della popolazione. Quanto detto per la media si può generalizzare alla varianza, alla frequenza relativa ecc..

CAPITOLO PRIMO

Elementi di calcolo delle probabilità

In questo capitolo si esamineranno gli elementi di calcolo delle probabilità indispensabili per una introduzione all'*inferenza statistica*.

1 Modelli probabilistici

Nell'introduzione si è precisato che il "calcolo delle probabilità" si occupa dello studio degli esperimenti casuali cui appartengono le estrazioni casuali campionarie. Gli esperimenti casuali sono rappresentati da particolari modelli matematici che sono denominati modelli probabilistici.

1.1 Esperimenti casuali

Ai fini della presente trattazione è possibile dividere gli esperimenti in deterministici quando la replicazione dello stesso "sotto le medesime condizioni" dà luogo allo stesso risultato. Si ha invece un esperimento casuale quando, pur facendo il possibile per mantenere la sperimentazione sotto le stesse condizioni, i risultati possono essere diversi da una replicazione all'altra.

Ecco ora alcuni esperimenti casuali:

a) *lancio di una moneta*. Come risultato del lancio (dell'esperimento) si considera la figura stampigliata sulla faccia superiore: Testa, Croce.

b) *lancio di un dado* sulle cui sei facce sono stati stampigliati i numeri: 1, 2, 3, 4, 5, 6. Come risultato di un lancio si considera il numero stampigliato sulla faccia superiore: 1, 2, 3, 4, 5, 6.

c) *Sesso di un nascituro*. In una sperimentazione – una nascita – si considerano due risultati: maschio, femmina.

d) *Estrazione con riposizione* di un campione casuale di $n = 2$ unità da una popolazione di $N = 6$ unità. I risultati di questo campionamento sono tutti gli $N^2 = 36$ possibili campioni che sono stati elencati nell'introduzione.

In ciascuno degli esperimenti sopra riportati è possibile predisporre la lista dei risultati possibili *ma non è possibile prevedere con esattezza quale sarà il* *risultato in una replicazione dell'esperimento.*

——————— . . ——————— . . ———————

Se invece si replica un numero di volte sufficientemente elevato lo stesso esperimento si osservano nei risultati "regolarità" la cui conoscenza può essere utile nel trattare gli esperimenti casuali.

——————— . . ——————— . . ———————

È ora necessario precisare che nel predisporre la lista dei risultati possibili (o elementari) bisogna che:

i) i singoli risultati siano fra loro incompatibili;

ii) la lista sia esaustiva.

In altre parole, la lista dei risultati di un esperimento casuale, deve soddisfare le stesse due proprietà che si hanno nella lista delle modalità di un carattere qualitativo o quantitativo, a suo tempo studiate in statistica descrittiva.

——————— . . ——————— . . ———————

Per comprendere cosa siano le "regolarità" cui si è accennato in precedenza si sono eseguite, con una moneta di 50 centesimi di euro, 200 lanci. I lanci sono stati suddivisi in $k = 20$ serie: ciascuna serie si compone di $n = 10$ lanci. I risultati dei singoli lanci, riportati in Tabella 1.8, sono:

- Testa $=$ {Faccia con la stampigliatura 50 centesimi};
- Croce $=$ {l'altra faccia}.

1^a	2^a	3^a	4^a	5^a	6^a	7^a	8^a	9^a	10^a	11^a	12^a	13^a	14^a	15^a	16^a	17^a	18^a	19^a	20^a
C	T	C	T	C	C	T	C	C	C	T	T	T	C	T	T	C	C	T	C
C	T	T	T	C	C	T	C	C	T	T	C	C	T	C	C	T	C	C	C
T	C	C	C	T	T	T	C	T	T	T	T	C	C	T	T	T	T	C	T
C	T	T	C	T	C	C	C	T	T	T	T	T	T	T	C	C	C	T	C
C	T	T	T	C	C	C	C	T	C	C	C	C	T	T	T	C	C	C	C
C	C	T	T	T	T	T	T	T	C	T	C	T	C	T	T	T	T	C	T
C	C	C	C	C	C	C	T	C	C	T	C	T	C	T	C	C	T	C	C
C	C	C	C	C	T	C	C	T	T	C	T	C	T	C	T	C	T	T	C
T	T	T	C	T	T	T	T	T	C	T	T	C	T	C	C	C	C	C	C
T	C	T	C	T	T	C	T	T	C	T	T	T	T	C	T	T	C	T	T
3T	5T	6T	4T	5T	5T	5T	4T	7T	4T	8T	6T	5T	6T	6T	6T	4T	4T	4T	3T

Tabella 1.8: *Risultati delle* 20 *serie di* 10 *lanci.*

Si supponga di essere interessati al risultato T= $\{$Testa$\}$. Il numero di Teste $n\,(\mathrm{T})$ ottenute nelle singole serie di $n = 10$ lanci ha assunto i valori: 3, 5, 6, 4, 5, 5, 5, 4, 7, 4, 8, 6, 5, 6, 6, 6, 4, 4, 4, 3. La frequenza relativa di T, $fr\,(\mathrm{T}) = \dfrac{n\,(\mathrm{T})}{n}$, nelle $k = 20$ serie, è variata da un minimo di $0,3$ ad un massimo di $0,8$. Si potrà così "congetturare" che verosimilmente anche *nella prossima serie di* $n = 10$ *lanci la frequenza relativa sarà compresa fra* $0,3$ *e* $0,8$. Questa congettura non è sicura in quanto potrebbe accadere che nei prossimi $n = 10$ lanci si abbiano, ad esempio, per $n\,(\mathrm{T})$ i valori: 0, 1, 2, 9, 10. La congettura si fonda sull'esperienza di $k = 20$ serie di $n = 10$ lanci e si basa anche sulla "speranza" che nella prossima serie le frequenze oscillino come nelle prime $k = 20$ serie di lanci.

Se si allungano le singole $k = 20$ serie eseguendo $n = 50$ lanci per ciascuna serie, l'esperienza mostra che, nel passare da una serie all'altra, le variazioni delle $fr\,(\mathrm{T})$ sono molto più limitate; la frequenza relativa minima potrebbe essere pari a $0,36$ e quella massima a $0,66$. L'esperienza, in altre parole, mostra che aumentando n da 10 a 50 la frequenza relativa di T è meno variabile; conseguentemente è possibile congetturare che nei prossimi $n = 50$ lanci la $fr\,(\mathrm{T})$ sia compresa nell'intervallo $[0,36; 0,66]$.

Se si eseguono $n = 250$ lanci per ciascuna serie l'esperienza mostra una ulteriore riduzione delle variazioni fra le $fr\,(\mathrm{T})$. Se si allungassero ancora le serie le oscillazioni di $fr\,(\mathrm{T})$ tenderebbero a ridursi ulteriormente. In conclusione, tenendo fisso $k = 20$, l'esperienza mostra che:

- qualunque sia il valore di n (10, 50, 250, ecc.) la $fr\,(\mathrm{T})$ oscilla intorno al valore $0,5$;
- all'aumentare di n le oscillazioni di $fr\,(\mathrm{T})$ si riducono sempre di più e tendono a concentrarsi intorno a $0,5$. Questa proprietà è nota come "stabilizzazione delle frequenze relative col crescere delle prove".

Oltre che nel campo dei giochi di sorte, del lancio di monete, del lancio di dadi, la regolarità sopra descritta si presenta anche in altri campi per i quali a prima vista sembrerebbe che le prove non si svolgano sotto le stesse condizioni.

Si pensi a questo proposito alla frequenza relativa di neonati di sesso maschile ($fr\,(\mathrm{M}) = \dfrac{n(\mathrm{M})}{n}$) osservabile in $k = 20$ gruppi di nascite ciascuna caratterizzata da n nascite. Si può organizzare questa sperimentazione raggruppando le nascite, che si avranno presso la clinica Mangiagalli di Milano, in gruppi di n nascite successive, prendendo per n gli stessi valori considerati per il lancio della moneta da 50 centesimi di euro. La frequenza relativa $fr\,(\mathrm{M})$ dei maschi mostrerà la stessa stabilizzazione vista con i lanci della moneta con una sola differenza: il valore intorno al quale si stabilizzerà $fr\,(\mathrm{M})$ è un po' più grande di $0,5$. In effetti questo valore è un po' più grande di $0,5$. È noto, infatti, dalla demografia che, nel genere umano le nascite maschili sono un po' più frequenti di quelle femminili. Nel 2007 in Italia si sono avuti $n = 563.933$ nati. I maschi sono risultati $n(\text{M}) = 290.330$ e le femmine sono risultate $n(\text{F}) = 273.603$. La frequenza relativa dei maschi fra i nati in in Italia è così pari a: $fr\,(\text{M}) = 0,51483$. Si sono svolte molte sperimentazioni in campi differenti e sì è sempre notato che sotto condizioni sperimentali "costanti" la frequenza relativa di un determinato risultato tende, all'aumentare di n, a stabilizzarsi intorno ad un valore che non necessariamente è uguale a $0,5$. È opportuno sottolineare che questa proprietà è stata evidenziata da esperienze basate su milioni di osservazioni. In altre parole questa proprietà non è stata dedotta "a tavolino" con regole della matematica.

Si può a questo punto affermare che il valore intorno al quale si "stabilizza" la frequenza relativa di un risultato di un esperimento casuale è una *caratteristica* di quell'esperimento. Con ciò si intende affermare che ripetendo le serie di prove la frequenza relativa si stabilizzerà sempre intorno allo stesso valore. Trattasi di una caratteristica utile nei casi in cui bisogna trattare con esperimenti per i quali si possa parlare di ripetitività della sperimentazione "sotto le stesse condizioni".

La proprietà della stabilizzazione della frequenza relativa col crescere del numero delle prove costituirà la base empirica per il concetto di probabilità. Il concetto (l'interpretazione) di probabilità legato alla proprietà di stabilizzazione delle frequenze relative è particolarmente utile nell'inferenza in quanto l'estrazione di un campione da una popolazione è una operazione ripetibile sotto le stesse condizioni.

1.2 La costruzione di modelli probabilistici

Nel prospetto che segue sono riportate le frequenze ottenute replicando n volte un esperimento casuale che si compone dei risultati elementari $e_1, e_2, \ldots, e_i, \ldots, e_s$.

Risultati elementari	Frequenze dei risultati	Frequenze relative
e_1	$n\,(e_1)$	$fr\,(e_1)$
e_2	$n\,(e_2)$	$fr\,(e_2)$
$\vdots$	$\vdots$	$\vdots$
e_i	$n\,(e_i)$	$fr\,(e_i)$
$\vdots$	$\vdots$	$\vdots$
e_s	$n\,(e_s)$	$fr\,(e_s)$
Totale	n	1

Si tenga presente che: $fr\,(e_i) = \dfrac{n\,(e_i)}{n} \geq 0$; $\displaystyle\sum_{i=1}^{s} fr\,(e_i) = 1$.

Si può sintetizzare la sperimentazione fornendo la lista dei risultati con accanto i corrispettivi valori delle frequenze relative.

Si è ora pronti per individuare gli elementi che costituiscono un modello probabilistico. Si è in precedenza precisato che un modello probabilistico deve rappresentare un esperimento casuale. Pertanto nel modello si ha l'insieme dei risultati dell'esperimento $e_1, e_2, \ldots, e_i, \ldots, e_s$ cui vanno associati dei numeri positivi $P(e_1), P(e_2), \ldots, P(e_i), \ldots, P(e_s)$ che hanno le stesse proprietà formali delle frequenze relative. In particolare: $P(e_i) > 0$, $i = 1, 2, \ldots, s$; $\sum_{i=1}^{s} P(e_i) = 1$. Questi numeri positivi, che hanno le stesse proprietà formali delle frequenze relative, vengono denominati probabilità dei risultati elementari.

La struttura matematica del modello probabilistico è così costituita dall'insieme dei risultati elementari $e_1, e_2, \ldots, e_i, \ldots, e_s$ ai quali sono associate le corrispettive probabilità $P(e_i)$.

Le probabilità $P(e_i)$ rappresentano in questi modelli le frequenze relative $fr(e_i)$ di lungo periodo. Il "calcolo delle probabilità" insegna, partendo da questa struttura del modello probabilistico, a ricavare la probabilità da associare a risultati più complessi di quelli previsti dalla lista dei risultati elementari.

Il modello probabilistico per rappresentare i risultati nel lancio della moneta in precedenza presentati è il seguente.

Risultati possibili	Probabilità
Testa	$0,5$
Croce	$0,5$
Totale	1

Le probabilità $P(\mathrm{T}) = P(\mathrm{C}) = 0,5$ nel modello probabilistico ora presentato rappresentano le frequenze relative di lungo periodo dei due risultati possibili.

Il modello probabilistico per rappresentare il sesso di un nascituro è il seguente.

Sesso	Probabilità
Maschile	$0,51483$
Femminile	$0,48517$
Totale	1

Anche in questo caso la $P(\mathrm{M}) = 0,51483$ e la $P(\mathrm{F}) = 0,48517$ rappresentano le corrispettive frequenze relative di lungo periodo.

Prima di passare oltre conviene soffermarsi su due questioni. *La prima* riguarda la possibilità di costruire modelli probabilistici aventi la stessa struttura formale di quelli visti sino ad ora ma che non rappresentano esperimenti casuali di tipo ripetitivo. La struttura matematica è la stessa (lista risultati cui sono associati numeri positivi tali che la loro somma sia uguale ad uno) ma i numeri positivi rappresentano cose diverse: frequenze relative di lungo periodo per gli esperimenti casuali di tipo ripetitivo, valutazioni soggettive per gli esperimenti non di tipo ripetitivo. È bene precisare che il calcolo delle probabilità, partendo da una struttura matematica astratta, conduce a conseguenze (teoremi) che riguardano questa struttura astratta. Questi teoremi sono ovviamente validi sia se le probabilità rappresentano una frequenza di lungo periodo sia se rappresentano qualcosa d'altro.

La seconda questione che preme far presente riguarda la generalità dei risultati (teoremi, leggi) che vengono ricavati nel senso che questi valgono per un qualsiasi modello probabilistico.

———— . . ———— . . ————

Conviene ora occuparsi del come vengono assegnate le probabilità nei modelli probabilistici che rappresentano esperimenti casuali di tipo ripetitivo.

Le strade che si possono seguire sono basate:

1. su eventuali simmetrie presenti fra i risultati possibili;

2. sulle frequenze relative calcolate in un gran numero di prove.

Spesso in molti modelli probabilistici legati ai giochi di sorte, ai lanci di dadi, di monete, ecc., si assegna una probabilità uguale a tutti i risultati possibili. Così ai risultati testa e croce del modello riguardante il lancio di una moneta si assegnano le probabilità $0,5$ e $0,5$. Analogamente si assegna probabilità $\frac{1}{6}$ a ciascuna delle sei facce di un dado nel modello che rappresenta i risultati del lancio dello stesso. In tutti questi casi si suppone che dal punto di vista fisico, geometrico, ecc., i risultati possibili si trovino, per quanto riguarda il loro verificarsi in una prova, sullo stesso piano, quindi sembra "ragionevole" assegnare ad essi la stessa probabilità. Se i risultati possibili sono s si assegna probabilità $\frac{1}{s}$ a ciascuno di essi.

Questo approccio si basa sulla convinzione che, ripetendo un numero di volte sufficientemente elevato l'esperimento, le frequenze relative dei risultati possibili siano pressoché uguali fra loro e quindi nel modello probabilistico devono essere pari a $\frac{1}{s}$.

I modelli probabilistici che assegnano uguale probabilità a tutti gli eventi possibili sono denominati modelli "uniformi". L'estrazione dei campioni casuali – del tipo visti in precedenza – da una popolazione può essere rappresentata da un modello probabilistico uniforme.

Esempio 1.2.1 *Modello probabilistico per $n = 2$ estrazioni con riposizione dalla popolazione con $N = 6$ unità.*

I risultati possibili sono costituiti dagli $N^2 = 36$ campioni possibili riportati sul prospetto a doppia entrata presentato nell'introduzione. A ciascuno dei 36 campioni possibili si associa la probabilità $\frac{1}{36} = \frac{1}{N^2}$. I singoli risultati possibili possono essere indicati dalle coppie (x_1, x_2) dei valori estraibili con le due osservazioni.

Esempio 1.2.2 *Modello probabilistico per $n = 2$ estrazioni senza riposizione.*

I risultati possibili sono gli $N \cdot (N - 1) = 30$ campioni possibili riportati nell'introduzione. A ciascuno dei 30 campioni possibili si assegna probabilità $\frac{1}{30} = \frac{1}{N} \cdot \frac{1}{N-1}$. I singoli risultati possibili possono essere indicati dalle coppie (x_1, x_2) dei valori estraibili con le due osservazioni.

————— . . ————— . . —————

Se si ha motivo di ritenere che manchi la simmetria fra i risultati possibili allora non si può usare questo procedimento per assegnare le probabilità.

Si pensi ad esempio ad un dado di materiale plastico nel quale sia stata introdotta una pallina di piombo in posizione eccentrica rispetto al baricentro geometrico del dado. Questa "eccentricità" tenderà a favorire alcune facce a scapito delle altre e quindi non è più ragionevole assegnare la stessa probabilità a tutte le facce.

In altre parole, se non è possibile intravedere simmetria fra i risultati, viene a mancare il convincimento che le frequenze relative dei risultati siano le stesse e quindi non si può assegnare la stessa probabilità a tutti gli eventi. In questi casi non resta che porre le probabilità uguali alle frequenze relative che si otterranno replicando molte volte l'esperimento. Questo è il procedimento adottato dalle aziende assicuratrici per i rischi che vengono assicurati. In effetti, sulla scorta di molte rilevazioni si calcolano le frequenze relative di alcuni rischi e quindi si costruiscono i modelli probabilistici che sono alla base, assieme ad altri elementi, del calcolo del premio che il cliente deve corrispondere per assicurare un certo rischio.

Esempio 1.2.3 *Ecco un altro esempio per il quale non è possibile far riferimento ai modelli uniformi.*

Sia X la durata della vita in anni di un neonato di sesso maschile in Italia. Relativamente alla durata della vita si possono considerare i seguenti risultati possibili:
$e_1 = \{X \leq 5\}$; $e_2 = \{5 < X \leq 10\}$; $e_3 = \{10 < X \leq 15\}$; $e_4 = \{15 < X \leq 20\}$; $e_5 = \{20 < X \leq 25\}$; $e_6 = \{25 < X \leq 30\}$; $e_7 = \{30 < X \leq 35\}$; $e_8 = \{35 < X \leq 40\}$; $e_9 = \{40 < X \leq 45\}$; $e_{10} = \{45 < X \leq 50\}$; $e_{11} = \{50 < X \leq 55\}$; .....; $e_{19} = \{90 < X \leq 95\}$; $e_{20} = \{95 < X\}$.
Per assegnare le probabilità ai risultati sopra indicati si procede (concettualmente) come segue. Si considerano $N = 100.000$ neonati di sesso maschile e si segue questa popolazione fino al suo esaurimento (per morte). Sia n_1 il numero dei neonati che muoiono entro 5 anni dalla nascita, sia n_2 il numero di neonati che muoiono dopo un periodo compreso fra i 5 e i 10 anni dalla nascita ($5 < X \leq 10$), sia .... Ovviamente $N = 100.000 = n_1 + n_2 + \ldots + n_{20}$. La probabilità che la durata sia quella indicata da e_i è allora fornita da $P(e_i) = \dfrac{n_i}{100.000}$. Ovviamente $\sum_{i=1}^{20} \dfrac{n_i}{100.000} = 1$. Il modello probabilistico per la durata della vita di un neonato di sesso maschile in Italia è riportato in Tabella 1.9.

Età X in anni	Decessi n_i	Probabilità di morte $\times 1.000$
$0 < X \leq 5$	483	$4,83$
$5 < X \leq 10$	60	$0,6$
$10 < X \leq 15$	77	$0,77$
$15 < X \leq 20$	270	$2,70$
$20 < X \leq 25$	388	$3,88$
$25 < X \leq 30$	405	$4,05$
$30 < X \leq 35$	406	$4,06$
$35 < X \leq 40$	523	$5,23$
$40 < X \leq 45$	748	$7,48$
$45 < X \leq 50$	1.166	$11,66$
$50 < X \leq 55$	1.884	$18,84$
$55 < X \leq 60$	2.980	$29,8$
$60 < X \leq 65$	4.849	$48,49$
$65 < X \leq 70$	7.139	$71,39$
$70 < X \leq 75$	10.839	$108,39$
$75 < X \leq 80$	15.102	$151,02$
$80 < X \leq 85$	17.768	$177,68$
$85 < X \leq 90$	18.482	$184,82$
$90 < X \leq 95$	11.107	$111,07$
$95 < X$	5.324	$53,24$
Totale neonati considerati	100.000	1.000

Tabella 1.9: *Modello probabilistico per la durata della vita di un neonato di sesso maschile in Italia (dati: Istat, Roma, 2004)*

Dalla tavola di mortalità indicata si desume che la probabilità che ha un neonato di morire fra i 20 ed i 25 anni è pari al $3,88‰$ ($3,88$ per 1.000: per ogni 1.000 neonati ne muoiono $3,88$ fra i 20 ed i 25 anni). Si può anche dire che per i neonati la probabilità di avere una durata della vita compresa fra 20 e 25 anni è pari al $3,88‰$. Analogamente, la tavola mostra che la probabilità che la durata di vita di un neonato sia compresa fra 70 e 75 anni è pari a $108,33‰$.

1.3 Probabilità di un risultato che si compone di alcuni risultati elementari

Si supponga che nell'esperimento del lancio di un dado si sia interessati alla valutazione della probabilità di ottenere un numero dispari. Il risultato E (punteggio dispari) si ottiene sia se si ha e_1 (faccia con il numero 1) sia se si ha E_3 (faccia con il numero 3), sia se si ha e_5 (faccia con il numero 5). Per costruzione e_1, e_3 ed e_5 (come anche e_2, e_4 ed e_6) sono fra loro incompatibili. Pertanto replicando n volte l'esperimento si ha $n(E) = n(e_1) + n(e_3) + n(e_5)$. Conseguentemente

$$\begin{aligned} fr(E) &= \frac{n(E)}{n} \\ &= \frac{n(e_1) + n(e_3) + n(e_5)}{n} \\ &= fr(e_1) + fr(e_3) + fr(e_5) \ , \end{aligned}$$

ovvero la frequenza relativa di un numero dispari – fra i sei numeri possibili – è pari alla somma delle frequenze relative dei tre risultati elementari e_1, e_3 ed e_5.

Trattasi di un risultato generale. In effetti, se il risultato A si compone di alcuni risultati elementari allora la sua frequenza – in n prove – è data dalla somma delle frequenze dei risultati elementari di cui si compone A. Consegue che la frequenza relativa di A è pari alla somma delle frequenze relative dei risultati elementari di cui si compone A.

Esempio 1.3.1 *Frequenza relativa di ottenere un numero almeno uguale a 5 nel lancio del dado.*

Il risultato A=(numero almeno uguale a 5) si ottiene sia se si realizza e_5 (numero uguale a 5) sia se si realizza e_6 (numero uguale a 6). Pertanto $n(A) = n(e_5) + n(e_6)$ e quindi

$$\begin{aligned} fr(A) &= \frac{n(A)}{n} = \\ &= \frac{n(e_5)}{n} + \frac{n(e_6)}{n} \\ &= fr(e_5) + fr(e_6) \ . \end{aligned}$$

Dato che nei modelli probabilistici la probabilità di un risultato rappresenta una frequenza relativa (almeno negli esperimenti ripetibili), sembra del tutto ragionevole *definire la probabilità di un risultato A come somma delle probabilità dei risultati elementari di cui si compone*.

Esempio 1.3.2 *Determinazione della probabilità di ottenere un numero dispari nel lancio di un dado (onesto).*

Il risultato R (numero dispari) si compone dei risultati semplici e_1, e_3, e_5 a ciascuno dei quali compete probabilità $\frac{1}{6}$. Pertanto:

$$\begin{aligned} P(R) &= P(e_1) + P(e_3) + P(e_5) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 0,5 \; . \end{aligned}$$

Esempio 1.3.3 *Determinazione della probabilità di ottenere un numero almeno uguale a* 5 *nel lancio di un dado (onesto).*

Il risultato A={numero ≥ 5} si compone dei risultati semplici e_5 ed e_6. Pertanto

$$\begin{aligned} P(A) &= P(e_5) + P(e_6) \\ &= \frac{1}{6} + \frac{1}{6} = 0,3333 \; . \end{aligned}$$

Esempio 1.3.4 *Determinare la probabilità che un neonato italiano di sesso maschile muoia entro i 10 anni dalla nascita.*

Dal modello probabilistico costruito per la durata della vita X in anni dei neonati italiani di sesso maschile si desume innanzitutto che:

a) $P(e_1) = P(X \leq 5) = 0,00483$

b) $P(e_2) = P(5 < X \leq 10) = 0,00060$.

Inoltre, A= $\{X \leq 10\}$ si compone dei risultati semplici $e_1 = \{X \leq 5\}$ ed $e_2 = \{5 < X \leq 10\}$. Pertanto

$$\begin{aligned} P(X \leq 10) &= P(X \leq 5) + P(5 < X \leq 10) \\ &= 0,00483 + 0,00060 = 0,00543 \; . \end{aligned}$$

Per apprezzare il valore di questa probabilità si moltiplica, solitamente, per 1000 e si ottiene $5,43‰$. Questo valore informa che su 1000 neonati ne muoiono entro i 10 anni dalla nascita $5,43$.

——————— . . ——————— . . ———————

L'applicazione della definizione di probabilità di un risultato A che si compone di un certo numero di risultati semplici conduce ad una celebre regola nel caso dei modelli uniformi.

Si supponga che un modello uniforme sia costituito da s risultati semplici, a ciascuno dei quali compete probabilità $\frac{1}{s}$. Si consideri ora un risultato A costituito da k risultati elementari ($k \leq s$).

Dalla definizione della probabilità di A consegue che P(A) è la somma di k termini ciascuno uguale a $\frac{1}{s}$, di modo che

$$P(A) = \overbrace{\frac{1}{s} + \frac{1}{s} + \ldots + \frac{1}{s}}^{\text{k volte}} = \frac{k}{s} \ .$$

In generale in un modello uniforme la probabilità di un risultato A è fornita dal rapporto fra il numero k dei risultati semplici di cui si compone A (risultati favorevoli ad A) ed il numero s dei risultati possibili.

Esempio 1.3.5 *Probabilità di ottenere nella seconda estrazione il numero $x_2 = 7$ nel caso delle $n = 2$ estrazioni con riposizione dalla popolazione di $N = 6$ unità, presentata nell'introduzione.*

È stato già evidenziato che questo campionamento si può rappresentare con un modello uniforme. Pertanto a ciascuno degli $s = 36$ campioni possibili compete probabilità $\frac{1}{36}$. I campioni con il secondo estratto uguale a 7 sono i seguenti sei $(3, 7)$; $(4, 7)$; $(6, 7)$; $(7, 7)$; $(9, 7)$; $(10, 7)$. Pertanto, la probabilità cercata è pari a

$$P\left(x_2 = 7\right) = \frac{6}{36} = \frac{1}{6} = 0,1666 \ .$$

Esercizio 1 *Ricavare nel caso delle $n = 2$ estrazioni con riposizione la probabilità che x_1 (il primo valore estratto) sia uguale a 7.*

Esempio 1.3.6 *Probabilità di ottenere nella prima estrazione il numero $x_1 = 7$ nel caso delle $n = 2$ estrazioni senza riposizione dalla popolazione di $N = 6$ unità presentata nell'introduzione.*

In questo campionamento si hanno $s = 30$ campioni possibili. Di questi i seguenti $k = 5$ sono quelli con il primo estratto uguale a 7: $(7, 3)$; $(7, 4)$; $(7, 6)$; $(7, 9)$; $(7, 10)$. Conseguentemente

$$P\left(x_1 = 7\right) = \frac{5}{30} = \frac{1}{6} = 0,1616 \ .$$

Esercizio 2 *Ricavare nel caso delle $n = 2$ estrazioni senza riposizione la probabilità che x_2 (il secondo estratto) sia uguale a 7.*

1.4 Algebra degli eventi.

Nel linguaggio del calcolo delle probabilità i risultati degli esperimenti casuali vengono anche denominati eventi.

Nel paragrafo 1.3 si è evidenziato che la probabilità di un risultato (evento) è fornita dalla somma delle probabilità dei risultati (eventi) elementari che lo compongono.

Non sempre però è semplice impiegare questo approccio anche nel caso dei modelli uniformi. Si consideri a questo proposito il caso di un campionamento con riposizione con $n = 4$ da una popolazione con $N = 100$. I campioni possibili sono ora $100^4 = 100.000.000$. Pertanto può diventare complicato valutare la probabilità di eventi effettuando il rapporto fra casi favorevoli e casi possibili.

Il calcolo delle probabilità, partendo dalla definizione di probabilità degli eventi vista in 1.3, ricava alcune leggi (di probabilità) che sono molto utili nel determinare la probabilità di eventi che appaiono complessi. La tecnica consiste nell'esprimere la probabilità di eventi complessi in termini delle probabilità di eventi più semplici per i quali è più immediata la determinazione delle probabilità stesse.

In questo contesto, nel descrivere i risultati (gli eventi), si impiegherà il linguaggio degli insiemi e la loro algebra.

In seguito sarà considerato sia l'insieme costituito da tutti gli eventi elementari del modello, che sarà indicato con Ω, sia l'insieme che non contiene alcun evento semplice che sarà indicato con $\emptyset$ (insieme vuoto).

Definizione (Complemento di un evento) Il complemento di un evento A, indicato con $\bar{\text{A}}$, è l'evento che si compone di tutti gli eventi elementari di Ω che non appartengono ad A.

Esempio 1.4.1 *Si consideri il modello probabilistico associato al lancio di un dado. Gli eventi elementari sono e_1, e_2, e_3, e_4, e_5, e_6; pertanto, $\Omega = \{e_1, e_2, e_3, e_4, e_5, e_6\}$. Sia R l'evento "numero dispari": $R = \{e_1, e_3, e_5\}$.*

Il complemento di R è l'evento $\overline{\text{R}}$ "numero pari". Ovviamente per individuare di quanti eventi elementari si compone $\overline{\text{R}}$ si considerano tutti gli eventi di Ω e si tolgono quelli che appartengono ad R. Si ha così $\overline{\text{R}} = \{e_2, e_4, e_6\}$.

Esempio 1.4.2 *Si consideri il modello probabilistico associato al lancio di due monete.*

I risultati elementari possibili sono: TT, TC, CT, CC; TT indica testa con la prima e testa con la seconda moneta, TC indica testa con la prima e croce con la seconda, ecc.. Pertanto Ω={TT, TC, CT, CC}. Sia A l'evento almeno una testa; ovviamente A= {TT, TC, CT}. Il complemento di A è l'evento nessuna testa. Si ha così $\overline{\text{A}}$={CC}.

Esempio 1.4.3 *Si consideri il modello probabilistico presentato in 1.2 relativo alla durata della vita X in anni dei neonati di sesso maschile in Italia.*

Gli eventi elementari sono:
$e_1 = \{X \leq 5\}$; $e_2 = \{5 < X \leq 10\}$; $e_3 = \{10 < X \leq 15\}$;......; $e_{19} = \{90 < X \leq 95\}$; $e_{20} = \{95 < X\}$. Si consideri l'evento B morte di un neonato entro i 10 anni: B= $\{X \leq 10\} = \{e_1, e_2\}$. Si consideri l'evento $\overline{\text{B}} = \{10 < X\} = \{e_3, e_4\, e_5, e_6, \ldots, e_{19}, e_{20}\}$.

Esempio 1.4.4 *Si consideri l'evento* $\Omega = \{e_1, e_2, \ldots, e_s\}$.

Si consideri il suo complemento $\overline{\Omega}$. Per individuare gli eventi elementari di cui si compone $\overline{\Omega}$ bisogna considerare tutti gli eventi di Ω e togliere quelli che appartengono ad Ω. Ovviamente dopo questa operazione non resta nessun evento elementare. Pertanto $\overline{\Omega}$ non contiene alcun evento elementare, ovvero $\overline{\Omega} = \emptyset$ (insieme vuoto).

Definizione (Intersezione fra due eventi) L'intersezione degli eventi A e B, indicata con (A∩B) è l'evento che si compone di tutti gli eventi elementari che appartengono contemporaneamente sia ad A sia a B.

Esempio 1.4.5 *Si consideri il modello probabilistico associato al lancio di un dado. Si considerino i seguenti quattro eventi:*

- *E={il punteggio eccede 3};*
- *F={il punteggio è un numero pari};*
- *G={il punteggio non è divisibile per 3};*
- *H={il punteggio non è nè 3 nè 4}.*

Gli eventi E, F, G ed H sono composti dai seguenti eventi elementari:

E=$\{e_4, e_5, e_6\}$; *F*=$\{e_2, e_4, e_6\}$; *G*=$\{e_1, e_2, e_4, e_5\}$; *H*=$\{e_1, e_2, e_5, e_6\}$.

Si ricavino le intersezioni: (E∩H); (F∩G); ($\overline{F}$∩G).
Si ottengono immediatamente dalla definizione le intersezioni: (E∩H)=$\{e_5, e_6\}$; (F∩G)=$\{e_2, e_4\}$. Per l'intersezione ($\overline{\text{F}}$∩G) si ricava innanzitutto $\overline{\text{F}}$=$\{e_1, e_3, e_5\}$. Quindi ($\overline{\text{F}}$∩G)=$\{e_1, e_5\}$.

Esempio 1.4.6 *Si consideri ancora il modello associato al lancio di un dado e si considerino gli eventi: A=* $\{e_1, e_2\}$*; B=* $\{e_4, e_5, e_6\}$*. Si determinino le intersezioni (A∩B); (A∩$\overline{A}$); (B∩$\overline{B}$); (A∩$\overline{B}$).*

Si ricava immediatamente che $\overline{\text{A}} = \{e_3, e_4, e_5, e_6\}$ e che $\overline{\text{B}} = \{e_1, e_2, e_3\}$. Pertanto (A∩B)= $\emptyset$; (A∩$\overline{\text{A}}$)= $\emptyset$; (B∩$\overline{\text{B}}$)= $\emptyset$; (A∩$\overline{\text{B}}$)= $\{e_1, e_2\}$.

L'esempio ha mostrato che l'intersezione fra due eventi può coincidere con l'insieme vuoto.

Definizione (Eventi incompatibili) Due eventi si dicono incompatibili o escludentisi se non hanno eventi elementari in comune. Questo significa che l'intersezione fra due eventi incompatibili coincide con l'evento vuoto $\emptyset$.

Si ricordi che tutti gli eventi elementari di un modello probabilistico $(e_1, e_2, \ldots, e_s)$ sono per definizione incompatibili. Si osservi che sono anche incompatibili gli eventi A e $\overline{\text{A}}$, essendo A un evento qualsiasi.

Definizione (Intersezione fra tre o più eventi) L'intersezione fra gli eventi A, B e C indicata con (A e B e C) o anche (A∩B∩C) è l'evento che si compone di tutti gli eventi elementari che appartengono contemporaneamente sia ad A, sia a B, sia a C.

In modo analogo si definisce l'intersezione fra quattro o più eventi.

Esempio 1.4.7 *Relativamente al modello associato al lancio di un dado si determini l'intersezione (A∩B∩C), essendo:*

$$A = \{e_1, e_2, e_3\};\ B = \{e_2, e_4, e_6\};\ C = \{e_2, e_3, e_4, e_5, e_6\}.$$

È immediato verificare che (A∩B∩C)= $\{e_2\}$.

Riveste un ruolo importante nel linguaggio degli insiemi anche il concetto di unione.

Definizione (Unione di eventi) L'unione degli eventi A e B, indicata con (A∪B), è l'evento che si compone degli eventi elementari che appartengono ad almeno uno dei due.

L'unione di k eventi A_1, A_2,...,A_k, indicata con ($A_1 \cup A_2 \cup \ldots \cup A_k$) si compone degli eventi elementari che appartengono ad almeno uno di essi.

Esempio 1.4.8 *Gli eventi elementari del modello associato al lancio di un dado sono:* e_1, e_2, e_3, e_4, e_5 *ed* e_6*. Si ricavi l'evento "punteggio divisibile per due o per tre".*

Si indichi con A l'evento "punteggio divisibile per due": A= $\{e_2, e_4, e_6\}$. Si indichi con B l'evento "punteggio divisibile per tre": B= $\{e_3, e_6\}$. Consegue che $(\text{A} \cup \text{B}) = \{e_2, e_3, e_4, e_6\}$.

Esempio 1.4.9 *Gli eventi elementari del modello probabilistico associato al lancio di tre monete sono i seguenti otto TTT, TTC, TCT, CTT, TCC, CTC, CCT, CCC. L'evento TTC indica che con la prima moneta si è avuto testa, con la seconda testa e con la terza croce. Analogo è il significato degli altri eventi elementari.*

Si considerino ora gli eventi

$$\begin{aligned} \mathrm{E} &= \{\text{nessuna testa}\} = \{\text{CCC}\}\ ; \\ \mathrm{F} &= \{\text{una testa}\} = \{\text{TCC, CTC, CCT}\}\ ; \\ \mathrm{G} &= \{\text{due teste}\} = \{\text{TTC, TCT, CTT}\}; \\ \mathrm{H} &= \{\text{tre teste}\} = \{\text{TTT}\}\ . \end{aligned}$$

Si determini ora l'evento "Almeno due teste". Almeno due teste significa $\{$o due teste o tre teste$\}$=(G$\cup$H)=$\{$TTC, TCT, CTT, TTT$\}$.

Si determini altresì l'evento "Al più una testa". Al più una testa significa zero teste o una testa. Pertanto, $\{$Al più una testa$\}$=(E$\cup$F)=$\{$CCC, CCT, CTC, TCC$\}$.

Definizione (Evento differenza tra due eventi) L'evento differenza fra A e B, si indica con (A-B), è costituito da tutti gli eventi elementari di A che non appartengono a B.

In altre parole per individuare gli eventi elementari di (A-B) si *devono togliere ad A gli eventi della intersezione fra A e B.*

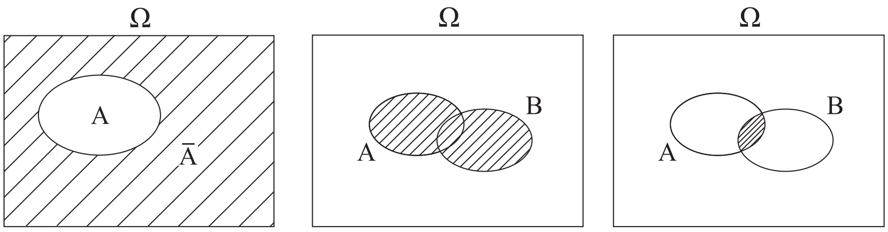
Esempio 1.4.10 *Si considerino gli eventi A=* $\{e_1, e_3, e_4, e_8\}$ *e B=* $\{e_1, e_2, e_5, , e_8, e_9\}$. *Si determini l'evento differenza (A-B).*

A tal fine si ricava l'intersezione (A$\cap$B)=$\emptyset$, cioè A e B sono eventi incompatibili. In questo caso per ottenere (A-B) bisogna togliere ad A l'insieme vuoto, cioè (A-B)=A.

Definizione (Eventi incompatibili) Gli eventi A_1, A_2,..., A_k si dicono incompatibili se nessun evento elementare appartiene a più di uno di essi, o ciò che è lo stesso se l'intersezione fra due qualsiasi di essi è uguale all'evento vuoto $\emptyset$.

——— . . ——— . . ———

Nello studio dell'algebra degli eventi si fa largo uso dei diagrammi di Venn. Con questi diagrammi si rappresenta l'evento Ω, cioè l'evento costituito da tutti gli eventi elementari del modello probabilistico, con un rettangolo. Un generico evento è rappresentato solitamente da una figura chiusa (cerchio, ellisse, quadrato, ecc..) incluso in Ω. Nel grafico (1.1) si sono visualizzati (con tratteggio) gli eventi: $\overline{\mathrm{A}}$, (A$\cup$B), (A$\cap$B).

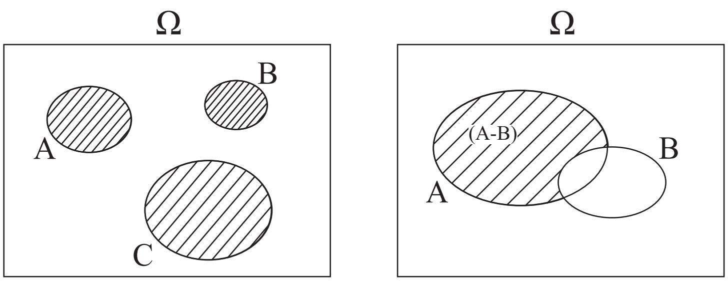


(a) Evento $\overline{\text{A}}$ (b) Unione fra due eventi compatibili (c) Intersezione fra due eventi compatibili

Figura 1.1: *Operazioni su eventi compatibili.*

Le parti tratteggiate indicano rispettivamente gli eventi: $\overline{\text{A}}$, (A∪B) ed (A∩B).

Nei due rettangoli del grafico (1.2) si sono riportate rispettivamente l'unione di tre eventi incompatibili e l'evento differenza (A-B).



(a) Unione di tre eventi incompatibili. (b) Evento (A-B).

Figura 1.2: *Operazioni su eventi compatibili ed incompatibili.*

1.5 Probabilità dell'evento Ω, dell'evento unione, dell'evento complementare e dell'evento $\emptyset$

In 1.3 si è assunto che la probabilità di un evento A è fornita dalla somma delle probabilità degli eventi elementari (del modello probabilistico) di cui si compone.

Si ricaveranno ora, partendo da tale assunzione, le probabilità dell'evento Ω, dell'evento unione e dell'evento complementare di un prefissato evento.

Come è ben noto, in ogni modello probabilistico, la somma delle probabilità di tutti gli eventi semplici $e_1, e_2, \ldots, e_s$, di cui si compone, è pari ad 1. Dato che in ogni modello probabilistico $\Omega = \{e_1, e_2, \ldots, e_s\}$ e dato che la probabilità di un evento è pari alla somma delle probabilità degli eventi elementari di cui si compone, si ha

$$P(\Omega) = \sum_{i=1}^{s} P(e_i) = 1 \ . \tag{1.1}$$

L'evento Ω è detto evento certo in quanto in ogni replicazione dell'esperimento casuale si deve verificare uno degli eventi possibili $e_1, e_2, \ldots, e_s$ dell'esperimento e quindi in ogni replicazione si verifica Ω.

———— . . ———— . . ————

Siano A e B due eventi incompatibili. Per semplicità si supponga che $\mathrm{A} = \{e_1, e_3\}$ e che $\mathrm{B} = \{e_2, e_5, e_6\}$. Visto che A e B sono incompatibili l'unione A∪B si compone di tutti gli eventi semplici di A e di tutti gli eventi semplici di B, ovvero $(\mathrm{A} \cup \mathrm{B}) = \{e_1, e_3, e_2, e_5, e_6\}$. Pertanto, nell'esempio considerato,

$$\begin{aligned} P(\mathrm{A} \cup \mathrm{B}) &= P(e_1) + P(e_3) + P(e_2) + P(e_5) + P(e_6) \\ &= \{P(e_1) + P(e_3)\} + \{P(e_2) + P(e_5) + P(e_6)\} \ . \end{aligned}$$

Tenendo ora presente che $P(e_1) + P(e_3) = P(\mathrm{A})$ e che $P(e_2) + P(e_5) + P(e_6) = = P(\mathrm{B})$ si ha $P(\mathrm{A} \cup \mathrm{B}) = P(\mathrm{A}) + P(\mathrm{B})$. La deduzione ora trovata vale qualunque sia l'evento A e qualunque sia l'evento B purché essi siano fra loro incompatibili.

Si può pertanto affermare che se due eventi A e B sono incompatibili – ovvero $(\mathrm{A} \cap \mathrm{B}) = \emptyset$ – allora

$$P(\mathrm{A} \cup \mathrm{B}) = P(\mathrm{A}) + P(\mathrm{B}) \ . \tag{1.2}$$

Esempio 1.5.1 Si consideri il modello probabilistico associato all'estrazione con riposizione $n = 2$ unità dalla popolazione di $N = 6$ unità presentata nell'introduzione (si vedano gli esempi di estrazioni campionarie di questo tipo considerati in 1.3). Trattasi di un modello probabilistico uniforme i cui eventi elementari sono gli $s = N^2 = 36$ campioni possibili, a ciascuno dei quali compete probabilità $\frac{1}{36}$. Si determini la probabilità che la media campionaria $\overline{X}$ sia uguale a $3,5$ o sia uguale a 8.

Gli eventi $\{\overline{X} = 3,5\}$ e $\{\overline{X} = 8\}$ sono fra loro incompatibili. L'evento $\{\overline{X} = 3,5\}$ si compone degli eventi semplici (campioni): $(3,4)$ e $(4,3)$. L'evento $\{\overline{X} = 8\}$ si compone degli eventi semplici (campioni): $(6,10)$, $(7,9)$, $(9,7)$ e $(10,6)$. Pertanto:

$$P\left(\overline{X} = 3,5\right) = \frac{\text{numero campioni favorevoli}}{\text{numero campioni possibili}} = \frac{2}{36} \ ;$$

$$P\left(\overline{X} = 8\right) = \frac{\text{numero campioni favorevoli}}{\text{numero campioni possibili}} = \frac{4}{36} \ .$$

Pertanto, essendo gli eventi $\left(\overline{X}=3,5\right)$ e $\left(\overline{X}=8\right)$ incompatibili, si può applicare la (1.2) e quindi

$$P\left[\left(\overline{X}=3,5\right)\cup\left(\overline{X}=8\right)\right] = \frac{2}{36}+\frac{4}{36}=\frac{1}{6}=0,16\overline{6}$$

———— . . ———— . . ————

Si dimostrerà ora che:

$$P\left(\mathrm{A}\right)=1-P\left(\overline{\mathrm{A}}\right) \ . \tag{1.3}$$

Dimostrazione. L'evento $\overline{\mathrm{A}}$ si compone di tutti gli eventi elementari di Ω che non appartengono ad A. In altre parole:

1. gli eventi A e $\overline{\mathrm{A}}$ sono incompatibili;
2. l'unione $\left(\mathrm{A}\cup\overline{\mathrm{A}}\right)=\Omega$.

Consegue che $P\left(\mathrm{A}\cup\overline{\mathrm{A}}\right)=P\left(\mathrm{A}\right)+P\left(\overline{\mathrm{A}}\right)$ (perché A e $\overline{\mathrm{A}}$ sono incompatibili). Inoltre $\left(\mathrm{A}\cup\overline{\mathrm{A}}\right)=\Omega$, pertanto $P\left(\Omega\right)=P\left(\mathrm{A}\right)+P\left(\overline{\mathrm{A}}\right)$. Dalla (1.1) si sa che $P\left(\Omega\right)=1$. Pertanto $1=P\left(\mathrm{A}\right)+P\left(\overline{\mathrm{A}}\right)$, da cui si ottiene il risultato $P\left(\mathrm{A}\right)=1-P\left(\overline{\mathrm{A}}\right)$.

Esempio 1.5.2 Si consideri, ancora una volta, l'estrazione con riposizione di $n=2$ unità dalla popolazione di $N=6$ unità presentata nell'introduzione e si ricavi la probabilità che la media campionaria $\overline{X}$ sia almeno uguale a 4. L'evento $\overline{X}$ almeno uguale a 4 si compone, considerando la lista dei valori che effettivamente la media campionaria può assumere, dei valori: 4; $4,5$; 5; $5,5$; 6; $6,5$; 7; $7,5$; 8; $8,5$; 9; $9,5$; 10. Pertanto

$$\begin{aligned} P\left(\overline{X}\geq 4\right) &= P\left[\left(\overline{X}=4\right)\cup\left(\overline{X}=4,5\right)\cup\left(\overline{X}=5\right)\cup\left(\overline{X}=5,5\right)\cup\left(\overline{X}=6\right)\cup\right.\\ &\quad \cup\left(\overline{X}=6,5\right)\cup\left(\overline{X}=7\right)\cup\left(\overline{X}=7,5\right)\cup\left(\overline{X}=8\right)\cup\left(\overline{X}=8,5\right)\cup\\ &\quad \left.\cup\left(\overline{X}=9\right)\cup\left(\overline{X}=9,5\right)\cup\left(\overline{X}=10\right)\right] \ . \end{aligned}$$

Risulta più semplice considerare l'evento complementare di $\overline{X}\geq 4$ che è l'evento $\overline{X}<4$. L'evento $\overline{X}<4$ si realizza sia se $\overline{X}=3,5$, sia se $\overline{X}=3$. È indubbiamente più agevole calcolare $P\left(\overline{X}<4\right)$ che $P\left(\overline{X}\geq 4\right)$. L'evento $\left(\overline{X}<4\right)=\left[\left(\overline{X}=3\right)\cup\left(\overline{X}=3,5\right)\right]$. Inoltre gli eventi $\left(\overline{X}=3\right)$ e $\left(\overline{X}=3,5\right)$ sono incompatibili. Pertanto

$$\begin{aligned} P\left(\overline{X}<4\right) &= P\left[\left(\overline{X}=3\right)\cup\left(\overline{X}=3,5\right)\right]\\ &= P\left(\overline{X}=3\right)+P\left(\overline{X}=3,5\right) \ . \end{aligned}$$

Dalla distribuzione delle medie campionarie – riportata nell'introduzione – si ricava che vi è un campione con $\overline{X} = 3$ e due campioni con $\overline{X} = 3,5$. Pertanto $P\left(\overline{X} = 3\right) = \frac{1}{36}$ e $P\left(\overline{X} = 3,5\right) = \frac{2}{36}$. Conseguentemente

$$P\left(\overline{X} < 4\right) = \frac{1}{36} + \frac{2}{36} = \frac{3}{36} = \frac{1}{12} = 0,083\overline{3} \ .$$

Applicando ora la (1.3) si ricava

$$\begin{aligned} P\left(\overline{X} \geq 4\right) &= 1 - P\left(\overline{X} < 4\right) \\ &= 1 - \frac{3}{36} = \frac{33}{36} = 0,916\overline{6} \ . \end{aligned}$$

———— . . ———— . . ————

Con $\emptyset$ si è indicato l'evento che non contiene nessun evento semplice. Consegue che replicando continuamente un esperimento l'evento $\emptyset$ non si realizza mai. Per questo motivo $\emptyset$ è anche detto evento impossibile.

Si dimostrerà ora che:

$$P\left(\emptyset\right) = 0 \ . \tag{1.4}$$

Dimostrazione. Si è mostrato in precedenza che $\emptyset = \overline{\Omega}$. Pertanto, applicando la (1.3) e ricordando che $P\left(\Omega\right) = 1$ si ha

$$\begin{aligned} P\left(\emptyset\right) &= P\left(\overline{\Omega}\right) \\ &= 1 - P\left(\Omega\right) \\ &= 1 - 1 = 0 \ . \end{aligned}$$

———— . . ———— . . ————

Nel caso in cui gli eventi $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_k$ siano fra loro incompatibili è possibile dimostrare facilmente che la probabilità dell'unione è data dalla somma delle probabilità dei singoli eventi, in formula

$$P\left(\mathrm{A}_1 \cup \mathrm{A}_2 \cup \cdots \cup \mathrm{A}_k\right) = \sum_{i=1}^{k} P\left(\mathrm{A}_i\right) \ . \tag{1.5}$$

Dimostrazione. Si supponga $k = 3$. È noto dall'algebra degli insiemi che $\left(\mathrm{A}_1 \cup \mathrm{A}_2 \cup \mathrm{A}_3\right) = \left[\mathrm{A}_1 \cup \left(\mathrm{A}_2 \cup \mathrm{A}_3\right)\right]$. Si ponga $\left(\mathrm{A}_2 \cup \mathrm{A}_3\right) = \mathrm{B}$. Pertanto

$$\left(\mathrm{A}_1 \cup \mathrm{A}_2 \cup \mathrm{A}_3\right) = \left(\mathrm{A}_1 \cup \mathrm{B}\right) \ .$$

Conseguentemente $P\left(\mathrm{A}_1 \cup \mathrm{A}_2 \cup \mathrm{A}_3\right) = P\left(\mathrm{A}_1 \cup \mathrm{B}\right)$.

Per ipotesi di eventi $\mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3$ sono incompatibili, cioè non hanno eventi elementari in comune. Consegue che anche A_1 e B sono incompatibili. Pertanto per la (1.2) si ha

$$P(\mathrm{A}_1 \cup \mathrm{A}_2 \cup \mathrm{A}_3) = P(\mathrm{A}_1) + P(\mathrm{B}) \ .$$

Ovviamente $P(\mathrm{B}) = P(\mathrm{A}_2 \cup \mathrm{A}_3) = P(\mathrm{A}_2) + P(\mathrm{A}_3)$. Sostituendo questa relazione nella precedente si ha così

$$P(\mathrm{A}_1 \cup \mathrm{A}_2 \cup \mathrm{A}_3) = P(\mathrm{A}_1) + P(\mathrm{A}_2) + P(\mathrm{A}_3) \ .$$

È immediato generalizzare il risultato anche per $k > 3$.

Esempio 1.5.3 *Si determini, tramite la (1.5), la probabilità che la media campionaria $\overline{X}$ sia almeno uguale a 4.*

Si consideri l'estrazione con riposizione di $n = 2$ unità dalla popolazione di $N = 6$ unità presentata nell'introduzione.

Si è visto nell'esempio precedente che

$$\begin{aligned}\left(\overline{X} \geq 4\right) \quad = \quad & \left[\left(\overline{X} = 4\right) \cup \left(\overline{X} = 4,5\right) \cup \left(\overline{X} = 5\right) \cup \left(\overline{X} = 5,5\right) \cup \left(\overline{X} = 6\right) \cup \right. \\ & \cup \left(\overline{X} = 6,5\right) \cup \left(\overline{X} = 7\right) \cup \left(\overline{X} = 7,5\right) \cup \left(\overline{X} = 8\right) \cup \left(\overline{X} = 8,5\right) \cup \\ & \left. \cup \left(\overline{X} = 9\right) \cup \left(\overline{X} = 9,5\right) \cup \left(\overline{X} = 10\right)\right]\end{aligned}$$

Gli eventi che compongono l'unione sono fra loro incompatibili, pertanto si può impiegare la (1.5).

Dalla distribuzione delle medie campionarie riportata nell'introduzione si ricava il prospetto in Tabella 1.10 la cui terza colonna riporta le probabilità che la media campionaria assuma rispettivamente i valori: 4; $4,5$;. . . .

$\overline{X}$	Numero dei campioni	Probabilità
4	1	$1/36$
$4,5$	2	$2/36$
5	4	$4/36$
$5,5$	2	$2/36$
6	3	$3/36$
$6,5$	6	$6/36$
7	3	$3/36$
$7,5$	2	$2/36$
8	4	$4/36$
$8,5$	2	$2/36$
9	1	$1/36$
$9,5$	2	$2/36$
10	1	$1/36$
Campioni favorevoli $\left(\overline{X} \geq 4\right)$	33	$33/36$

Tabella 1.10: *Probabilità che la media campionaria $\overline{X}$ sia almeno uguale a 4 (estrazioni con riposizione).*

In conclusione

$$\begin{aligned} P\left(\overline{X} \geq 4\right) &= P\left(\overline{X} = 4\right) + P\left(\overline{X} = 4,5\right) + \cdots + P\left(\overline{X} = 10\right) \\ &= \frac{1}{36} + \frac{2}{36} + \cdots + \frac{1}{36} = \frac{33}{36} \ . \end{aligned}$$

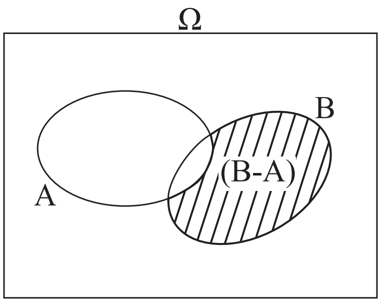
Si ottiene ovviamente lo stesso valore ottenuto con la relazione $P\left(\overline{X} \geq 4\right) = 1 - P\left(\overline{X} < 4\right)$.

———— . . ———— . . ————

Si dimostrerà ora che nel caso in cui gli eventi A e B sono compatibili

$$P\left(\mathrm{A} \cup \mathrm{B}\right) = P\left(\mathrm{A}\right) + P\left(\mathrm{B}\right) - P\left(\mathrm{A} \cap \mathrm{B}\right) \ . \tag{1.6}$$

Dimostrazione. L'unione $(\mathrm{A} \cup \mathrm{B})$ si compone degli eventi elementari che appartengono ad almeno uno di essi. Si possono elencare gli eventi elementari di $(\mathrm{A} \cup \mathrm{B})$ dividendo gli stessi in due gruppi: in un gruppo si considerano tutti gli eventi che appartengono ad A e nell'altro gruppo si considerano gli eventi elementari di B che non appartengono ad A, ovvero nell'altro gruppo si hanno gli eventi elementari appartenenti all'evento differenza $(\mathrm{B} - \mathrm{A})$. In questo modo gli eventi dell'unione sono ripartiti in *due eventi fra loro incompatibili* A *e* $(\mathrm{B} - \mathrm{A})$. Il seguente diagramma di Venn illustra la scomposizione.



In altre parole valgono le uguaglianze:

i) $(\mathrm{A} \cup \mathrm{B}) = [\mathrm{A} \cup (\mathrm{B} - \mathrm{A})]$;

ii) $P\left(\mathrm{A} \cup \mathrm{B}\right) = P\left[\mathrm{A} \cup (\mathrm{B} - \mathrm{A})\right]$.

Essendo A e $(\mathrm{B} - \mathrm{A})$ incompatibili la probabilità della loro unione si calcola con la (1.2) e quindi

$$P\left(\mathrm{A} \cup \mathrm{B}\right) = P\left(\mathrm{A}\right) + P\left(\mathrm{B} - \mathrm{A}\right) \ .$$

L'evento B è costituito dagli eventi di B che appartengono anche ad A e dagli eventi di B che non appartengono ad A.

$$\mathrm{B} \;=\; \underbrace{\left(\begin{array}{c}\text{Eventi di B che}\\ \text{appartengono}\\ \text{ad A}\end{array}\right)}_{(\mathrm{B} \cap \mathrm{A})} \cup \underbrace{\left(\begin{array}{c}\text{Eventi di B che}\\ \text{non appartengono}\\ \text{ad A}\end{array}\right)}_{\left(\mathrm{B} \cap \overline{\mathrm{A}}\right)} \ .$$

Si osservi che gli eventi $(\mathrm{B} \cap \mathrm{A})$ e $\left(\mathrm{B} \cap \overline{\mathrm{A}}\right)$ sono fra loro incompatibili. Si osservi, altresì, che $\left(\mathrm{B} \cap \overline{\mathrm{A}}\right) = (\mathrm{B} - \mathrm{A})$. Pertanto

$$B = (\mathrm{B} \cap \mathrm{A}) \cup (\mathrm{B} - \mathrm{A}) \quad \Rightarrow \quad P(B) = P\left[(\mathrm{B} \cap \mathrm{A}) \cup (\mathrm{B} - \mathrm{A})\right] \ .$$

Applicando la (1.2) si ha

$$P(B) = P(\mathrm{B} \cap \mathrm{A}) + P(\mathrm{B} - \mathrm{A}) \quad \Rightarrow \quad P(\mathrm{B} - \mathrm{A}) = P(B) - P(\mathrm{B} \cap \mathrm{A}) \ .$$

Sostituendo quest'ultima relazione nell'ultima formula ricavata per $P(\mathrm{A} \cup \mathrm{B})$ si ha

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

Esempio 1.5.4 *Si consideri il lancio di un dado (onesto). Si ricavi la probabilità di ottenere un punteggio divisibile per due o per tre.*

Sia A l'evento punteggio divisibile per due: $\mathrm{A} = \{e_2, e_4, e_6\}$. Sia B l'evento punteggio divisibile per tre: $\mathrm{B} = \{e_3, e_6\}$. L'evento intersezione $(\mathrm{A} \cap \mathrm{B}) = \{e_6\}$. L'evento unione $(\mathrm{A} \cup \mathrm{B}) = \{e_2, e_3, e_4, e_6\}$, pertanto

$$\begin{aligned} P(\mathrm{A} \cup \mathrm{B}) &= P(e_2) + P(e_3) + P(e_4) + P(e_6) \\ &= \frac{4}{6} \ . \end{aligned}$$

Si può ricavare $P(\mathrm{A} \cup \mathrm{B})$ anche impiegando la (1.6). A tal fine si ricava, innanzi tutto:

$$\begin{aligned} P(\mathrm{A}) &= P(e_2) + P(e_4) + P(e_6) = \frac{3}{6} \ ; \\ P(\mathrm{B}) &= P(e_3) + P(e_6) = \frac{2}{6} \ ; \\ P(\mathrm{A} \cap \mathrm{B}) &= P(e_6) = \frac{1}{6} \ . \end{aligned}$$

In conclusione

$$\begin{aligned} P(\mathrm{A} \cup \mathrm{B}) &= P(\mathrm{A}) + P(\mathrm{B}) - P(\mathrm{A} \cap \mathrm{B}) \\ &= \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6} \ . \end{aligned}$$

Si osservi che

$$P(\mathrm{A} \cup \mathrm{B}) = \left[P(e_2) + P(e_4) + P(e_6)\right] + \left[P(e_3) + P(e_6)\right] - P(e_6) \ .$$

Quest'ultima rappresentazione fa capire che nel caso di eventi compatibili la somma $P(\mathrm{A}) + P(\mathrm{B})$ conteggia due volte la probabilità degli eventi che appartengono sia ad A sia a B. Per "pareggiare i conti" bisogna allora sottrarre a $P(\mathrm{A}) + P(\mathrm{B})$ la probabilità della loro intersezione.

Si possono richiamare ora alcuni risultati conseguiti in questo paragrafo:

1. $P(\Omega) = 1;$ (1.1)
2. $P(\mathrm{A} \cup \mathrm{B}) = \begin{cases} P(\mathrm{A}) + P(\mathrm{B}) & \text{se } (\mathrm{A} \cap \mathrm{B}) = \emptyset \quad (1.2) \\ P(\mathrm{A}) + P(\mathrm{B}) - P(\mathrm{A} \cap \mathrm{B}) & \text{se } (\mathrm{A} \cap \mathrm{B}) \neq \emptyset; \quad (1.6) \end{cases}$
3. $P(\mathrm{A}) = 1 - P\left(\overline{\mathrm{A}}\right)$, o anche $P\left(\overline{\mathrm{A}}\right) = 1 - P(\mathrm{A})$; (1.3)
4. $P(\emptyset) = 0;$ (1.4)
5. se $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_k$ sono eventi incompatibili, allora

$$P(\mathrm{A}_1 \cup \mathrm{A}_2 \cup \cdots \cup \mathrm{A}_k) = \sum_{i=1}^{k} P(\mathrm{A}_i) \ . \qquad (1.5)$$

1.6 Probabilità e modelli probabilistici di eventi rappresentabili in tabella 2×2

In questo paragrafo si svolgono esercizi che riguardano le probabilità di eventi rappresentabili in tabelle 2×2.

Esempio 1.6.1 *Probabilità di superare esami universitari.*

Si indichi con S l'esame di Statistica, con D l'esame di Diritto Pubblico e con $(\mathrm{S} \cap \mathrm{D})$ entrambi gli esami. Le matricole di una Facoltà di economia superano entro il primo anno i sopra indicati esami con le seguenti probabilità: $P(\mathrm{S}) = 0,572$; $P(\mathrm{D}) = 0,658$; $P(\mathrm{D} \cap \mathrm{S}) = 0,486$. In una tabella 2×2 è possibile riportare non solo queste probabilità ma anche quelle degli eventi complementari $\overline{\mathrm{S}}$, $\overline{\mathrm{D}}$, nonché le probabilità delle altre tre intersezioni: $(\mathrm{S} \cap \overline{\mathrm{D}})$, $(\overline{\mathrm{S}} \cap \mathrm{D})$ e $(\overline{\mathrm{S}} \cap \overline{\mathrm{D}})$. Ovviamente $(\mathrm{S} \cap \overline{\mathrm{D}})$ significa che nel primo anno si supera Statistica ma non Diritto Pubblico, $(\overline{\mathrm{S}} \cap \mathrm{D})$ significa che si supera Diritto Pubblico ma non Statistica, mentre $(\overline{\mathrm{S}} \cap \overline{\mathrm{D}})$ significa che non si supera né Statistica né Diritto Pubblico.

	S	$\overline{\mathrm{S}}$	
D	$0,486$	$0,172$	$0,658$
$\overline{\mathrm{D}}$	$0,086$	$0,256$	$0,342$
	$0,572$	$0,428$	$1,000$

Tabella 1.11: *Probabilità di superare gli esami di Statistica S e di Diritto Pubblico D.*

Nella riga marginale sono riportate le probabilità $P(\mathrm{S})$ e $P(\overline{\mathrm{S}}) = 1 - P(\mathrm{S})$. Analogamente nella colonna marginale si riportano le probabilità

$$P(\mathrm{D}) \quad \text{e} \quad P(\overline{\mathrm{D}}) = 1 - P(\mathrm{D})\ .$$

Nella casella (S, D) si pone la probabilità $P(\mathrm{S} \cap \mathrm{D})$. Nelle altre caselle si sono riportate le probabilità delle altre tre intersezioni che si ricavano per differenza. In effetti:

i) $\mathrm{D} = (\mathrm{D} \cap \mathrm{S}) \cup (\mathrm{D} \cap \overline{\mathrm{S}})$, inoltre $(\mathrm{D} \cap \mathrm{S})$ è incompatibile con $(\mathrm{D} \cap \overline{\mathrm{S}})$. Pertanto $P(\mathrm{D}) = P(\mathrm{D} \cap \mathrm{S}) + P(\mathrm{D} \cap \overline{\mathrm{S}}) \rightarrow P(\mathrm{D} \cap \overline{\mathrm{S}}) = P(\mathrm{D}) - P(\mathrm{D} \cap \mathrm{S}) \rightarrow P(\mathrm{D} \cap \overline{\mathrm{S}}) = 0,658 - 0,486 = 0,172$.

ii) $\mathrm{S} = (\mathrm{S} \cap \mathrm{D}) \cup (\mathrm{S} \cap \overline{\mathrm{D}})$, inoltre $(\mathrm{S} \cap \mathrm{D})$ è incompatibile con $(\mathrm{S} \cap \overline{\mathrm{D}})$ per cui, dopo qualche passaggio si ottiene:

$$P(\mathrm{S} \cap \overline{\mathrm{D}}) = P(\mathrm{S}) - P(\mathrm{S} \cap \mathrm{D}) = 0,572 - 0,486 = 0,086\ .$$

iii) $\overline{\mathrm{S}} = (\overline{\mathrm{S}} \cap \mathrm{D}) \cup (\overline{\mathrm{S}} \cap \overline{\mathrm{D}})$, inoltre $(\overline{\mathrm{S}} \cap \mathrm{D})$ è incompatibile con $(\overline{\mathrm{S}} \cap \overline{\mathrm{D}})$ per cui, dopo qualche passaggio si ottiene:

$$P(\overline{\mathrm{S}} \cap \overline{\mathrm{D}}) = P(\overline{\mathrm{S}}) - P(\overline{\mathrm{S}} \cap \mathrm{D}) = 0,428 - 0,172 = 0,256\ .$$

Prima di procedere allo svolgimento di qualche esercizio è opportuno notare che oltre alle relazioni sopra indicate si ha anche:

$$(\mathrm{S} \cup \overline{\mathrm{S}}) = \Omega;\ (\mathrm{D} \cup \overline{\mathrm{D}}) = \Omega;\ \left\{(\mathrm{S} \cap \mathrm{D}) \cup (\mathrm{S} \cap \overline{\mathrm{D}}) \cup (\overline{\mathrm{S}} \cap \mathrm{D}) \cup (\overline{\mathrm{S}} \cap \overline{\mathrm{D}})\right\} = \Omega\ .$$

Queste relazioni indicano che le componenti degli eventi delle unioni sopra indicate costituiscono liste di eventi di tre modelli probabilistici.

Esempio 1.6.2 *Ricavare la probabilità che una matricola ha di superare "almeno uno dei due esami".*

{Almeno uno dei due esami} $= (\mathrm{S} \cup \mathrm{D})$. Gli eventi S e D sono fra loro compatibili per cui

$$\begin{aligned} P(\text{superare almeno uno dei due esami}) &= P(\mathrm{S}) + P(\mathrm{D}) - P(\mathrm{S} \cap \mathrm{D}) \\ &= 0,572 + 0,658 - 0,486 \\ &= 0,744\ . \end{aligned}$$

Si tenga altresì presente che l'evento {almeno uno dei due esami} si compone degli eventi incompatibili: $(\mathrm{S} \cap \mathrm{D})$, $(\overline{\mathrm{S}} \cap \mathrm{D})$ e $(\mathrm{S} \cap \overline{\mathrm{D}})$. In altre parole

$$\begin{aligned} P(\mathrm{S} \cup \mathrm{D}) &= P\left[(\mathrm{S} \cap \mathrm{D}) \cup (\overline{\mathrm{S}} \cap \mathrm{D}) \cup (\mathrm{S} \cap \overline{\mathrm{D}})\right] \\ &= P(\mathrm{S} \cap \mathrm{D}) + P(\overline{\mathrm{S}} \cap \mathrm{D}) + P(\mathrm{S} \cap \overline{\mathrm{D}}) \\ &= 0,486 + 0,172 + 0,086 = 0,744\ . \end{aligned}$$

Esempio 1.6.3 *Ricavare la probabilità che una matricola, nel primo anno, non superi né Statistica né Diritto Pubblico.*

Trattasi della $P(\overline{\mathrm{S}} \cap \overline{\mathrm{D}})$ che era stata già ricavata nella relazione $P(\overline{\mathrm{S}} \cap \overline{\mathrm{D}}) = P(\overline{\mathrm{S}}) - P(\overline{\mathrm{S}} \cap \mathrm{D}) = 0,256$. Questa probabilità può essere ricavata anche in altro modo. L'evento {nessuno dei due esami} è il complementare di {almeno uno dei due esami}. Pertanto

$$\begin{array}{rcl} P(\overline{\mathrm{S}} \cap \overline{\mathrm{D}}) = & P(\overline{\mathrm{S} \cup \mathrm{D}}) & = 1 - P(\mathrm{S} \cup \mathrm{D}) \\ & \downarrow & \\ P(\overline{\mathrm{S}} \cap \overline{\mathrm{D}}) = & 1 - 0,744 & = 0,256 \ . \end{array}$$

Esempio 1.6.4 *Ricavare la probabilità che lo studente non superi almeno uno dei due esami.*

L'evento {non superare almeno uno dei due esami} si indica con l'unione $(\overline{\mathrm{S}} \cup \overline{\mathrm{D}})$. Pertanto si può applicare la relazione

$$\begin{aligned} P(\overline{\mathrm{S}} \cup \overline{\mathrm{D}}) &= P(\overline{\mathrm{S}}) + P(\overline{\mathrm{D}}) - P(\overline{\mathrm{S}} \cap \overline{\mathrm{D}}) \\ &= 0,428 + 0,342 - 0,256 = 0,514 \ . \end{aligned}$$

A questa probabilità si può pervenire in altro modo. L'evento

$$(\overline{\mathrm{S}} \cup \overline{\mathrm{D}}) = \left[(\overline{\mathrm{S}} \cap \mathrm{D}) \cup (\mathrm{S} \cap \overline{\mathrm{D}}) \cup (\overline{\mathrm{S}} \cap \overline{\mathrm{D}})\right] \ .$$

Come è già stato più volte precisato, le tre intersezioni che compongono l'unione $(\overline{\mathrm{S}} \cup \overline{\mathrm{D}})$ sono tre eventi incompatibili. Pertanto

$$\begin{aligned} P(\overline{\mathrm{S}} \cup \overline{\mathrm{D}}) &= P(\overline{\mathrm{S}} \cap \mathrm{D}) + P(\mathrm{S} \cap \overline{\mathrm{D}}) + P(\overline{\mathrm{S}} \cap \overline{\mathrm{D}}) \\ &= 0,172 + 0,086 + 0,256 = 0,514 \ . \end{aligned}$$

1.7 Probabilità dell'intersezione

Siano A e B due risultati relativi allo stesso esperimento casuale. I risultati A e B si compongono di un certo numero di risultati elementari. Si vuole trovare una regola per l'assegnazione della probabilità all'evento intersezione $(\mathrm{A} \cap \mathrm{B})$. A tal fine si eseguono n replicazioni dell'esperimento. Le frequenze dei risultati A, B, $\overline{\mathrm{A}}$ e $\overline{\mathrm{B}}$ nonché delle loro intersezioni sono riportate nella Tabella 1.12.

	A	$\overline{\mathrm{A}}$	
B	$n\,(\mathrm{A} \cap \mathrm{B})$	$n\,(\overline{\mathrm{A}} \cap \mathrm{B})$	$n(\mathrm{B})$
$\overline{\mathrm{B}}$	$n\,(\mathrm{A} \cap \overline{\mathrm{B}})$	$n\,(\overline{\mathrm{A}} \cap \overline{\mathrm{B}})$	$n(\overline{\mathrm{B}})$
	$n(\mathrm{A})$	$n(\overline{\mathrm{A}})$	n

Tabella 1.12: *Frequenze dei risultati A, B, $\overline{A}$, $\overline{B}$ e delle loro intersezioni.*

Si possono calcolare così le *frequenze relative* dei due risultati A, B e della loro intersezione: $fr(\mathrm{A}) = \dfrac{n(\mathrm{A})}{n}$; $fr(\mathrm{B}) = \dfrac{n(\mathrm{B})}{n}$ e $fr(\mathrm{A} \cap \mathrm{B}) = \dfrac{n(\mathrm{A} \cap \mathrm{B})}{n}$. Le frequenze relative sopra indicate fanno riferimento alla totalità delle n prove.

È possibile, sempre sulla scorta delle frequenze riportate nella Tabella 2×2, ricavare la frequenza relativa di B nell'ambito delle $n(\mathrm{A})$ prove in cui si è verificato A. Si ha così: $fr(\mathrm{B}|\mathrm{A}) = \dfrac{n(\mathrm{A} \cap \mathrm{B})}{n(\mathrm{A})}$, frequenza relativa di B nell'ambito delle prove in cui si è verificato A. Questa frequenza è denominata *frequenza relativa condizionata* (frequenza relativa di B condizionata ad A). In modo analogo si può calcolare la $fr(\mathrm{A}|\mathrm{B}) = \dfrac{n(\mathrm{A} \cap \mathrm{B})}{n(\mathrm{B})}$.

Solitamente $fr(\mathrm{B}|\mathrm{A}) \neq fr(\mathrm{B})$. Fra le seguenti frequenze relative $fr(\mathrm{A} \cap \mathrm{B})$, $fr(\mathrm{A})$ e $fr(\mathrm{B}|\mathrm{A})$ vale la relazione $fr(\mathrm{A} \cap \mathrm{B}) = fr(\mathrm{A}) \cdot fr(\mathrm{B}|\mathrm{A})$.

Dimostrazione.

$$fr(\mathrm{A} \cap \mathrm{B}) = \frac{n(\mathrm{A} \cap \mathrm{B})}{n} ,$$

moltiplicando numeratore e denominatore per $n(\mathrm{A})$ si ottiene

$$fr(\mathrm{A} \cap \mathrm{B}) = \frac{n(\mathrm{A} \cap \mathrm{B})}{n} \cdot \frac{n(\mathrm{A})}{n(\mathrm{A})} \rightarrow fr(\mathrm{A} \cap \mathrm{B}) = \frac{n(\mathrm{A})}{n} \cdot \frac{n(\mathrm{A} \cap \mathrm{B})}{n(\mathrm{A})} .$$

Ricordando che $\dfrac{n(\mathrm{A})}{n} = fr(\mathrm{A})$ e che $\dfrac{n(\mathrm{A} \cap \mathrm{B})}{n(\mathrm{A})} = fr(\mathrm{B}|\mathrm{A})$ si ottiene il risultato. In modo analogo si dimostra che $fr(\mathrm{A} \cap \mathrm{B}) = fr(\mathrm{B}) \cdot fr(\mathrm{A}|\mathrm{B})$.

All'aumentare di n queste frequenze relative tendono a stabilizzarsi. Sembra allora ragionevole istituire fra le probabilità del modello probabilistico la stessa relazione moltiplicativa esistente fra le frequenze relative. In altre parole anche nel modello probabilistico valgono le relazioni

$$P(\mathrm{A} \cap \mathrm{B}) = P(\mathrm{A}) \cdot P(\mathrm{B}|\mathrm{A}) \tag{1.7}$$

$$P(\mathrm{A} \cap \mathrm{B}) = P(\mathrm{B}) \cdot P(\mathrm{A}|\mathrm{B}) . \tag{1.8}$$

Il simbolo $P(\mathrm{B}|\mathrm{A})$ indica la probabilità di B dato A e rappresenta, nel modello, la frequenza relativa $fr(\mathrm{B}|\mathrm{A})$. Significato analogo ha il simbolo $P(\mathrm{A}|\mathrm{B})$.

Dalla (1.7) e dalla (1.8) si ricava:

$$P(\mathrm{B}|\mathrm{A}) = \frac{P(\mathrm{A} \cap \mathrm{B})}{P(\mathrm{A})} , \quad \text{purché } P(\mathrm{A}) > 0 ; \tag{1.9}$$

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{A} \cap \mathrm{B})}{P(\mathrm{B})} , \quad \text{purché } P(\mathrm{B}) > 0 . \tag{1.10}$$

Esempio 1.7.1 *Estrazioni senza riposizione.*

Si abbia una classe di $N = 10$ studenti di cui 4 maschi e 6 femmine. Si estraggono senza riposizione $n = 2$ studenti. Siano: M_1 l'evento estrazione di un maschio alla prima estrazione, M_2 l'evento estrazione di un maschio alla seconda estrazione, F_1 l'evento estrazione di una femmina alla prima estrazione ed F_2 l'evento estrazione di una femmina alla seconda estrazione. Si calcoli la probabilità di avere due maschi.

L'evento {due maschi} coincide con l'evento intersezione $(\mathrm{M}_1 \cap \mathrm{M}_2)$. Pertanto si può applicare la relazione $P(\mathrm{M}_1 \cap \mathrm{M}_2) = P(\mathrm{M}_1) \cdot P(\mathrm{M}_2|\mathrm{M}_1)$.
Il simbolo $P(\mathrm{M}_1)$ indica la probabilità di ottenere un maschio alla prima estrazione. Si calcola questa probabilità facendo riferimento ad una estrazione da una popolazione con $N = 10$ studenti di cui 4 maschi.
Il simbolo $P(\mathrm{M}_2|\mathrm{M}_1)$ indica la probabilità di ottenere maschio alla seconda estrazione sapendo che si è ottenuto maschio nella prima estrazione. Si calcola questa probabilità facendo riferimento ad una estrazione da una popolazione con 9 studenti di cui 3 maschi.
Applicando il modello uniforme ad entrambe le estrazioni si ha:

$$\begin{aligned} P(\mathrm{M}_1) &= \frac{\text{N° casi favorevoli}}{\text{N° casi possibili}} = \frac{\text{4 maschi}}{\text{10 studenti}} = 0,4 \\ P(\mathrm{M}_2|\mathrm{M}_1) &= \frac{\text{N° casi favorevoli}}{\text{N° casi possibili}} = \frac{\text{3 maschi}}{\text{9 studenti}} = 0,3\bar{3}\ . \end{aligned}$$

In conclusione

$$P(\mathrm{M}_1 \cap \mathrm{M}_2) = \frac{4}{10} \cdot \frac{3}{9} = \frac{12}{90} = 0,13\bar{3}\ .$$

L'esperimento dell'estrazione di due studenti è stato suddiviso in due sub-esperimenti (prima estrazione, seconda estrazione), dei quali il secondo è condizionato dal risultato del primo. Senza questa suddivisione si sarebbe dovuto ricavare un unico modello costituito da tutti i 10×9 campioni possibili a ciascuno dei quali si sarebbe dovuto assegnare probabilità $\frac{1}{90}$.

Si consideri l'esempio delle $n = 2$ estrazioni senza riposizione dalla popolazione di $N = 6$ unità presentato nell'introduzione ed in altri paragrafi.
Conseguentemente

$$P(\mathrm{M}_1 \cap \mathrm{M}_2) = \frac{\text{N° campioni con 2 maschi}}{\text{N° campioni possibili}}\ .$$

Nell'esempio considerato quest'ultimo approccio è più complicato di quello basato sulla regola moltiplicativa $P(\mathrm{A} \cap \mathrm{B}) = P(\mathrm{A}) \cdot P(\mathrm{B}|\mathrm{A})$.

——— . . ——— . . ———

Si è ora pronti per introdurre il concetto di indipendenza in probabilità.

Definizione (Indipendenza in probabilità) L'evento B è indipendente (in probabilità) dall'evento A se

$$P\left(\mathrm{B}|\mathrm{A}\right) = P\left(\mathrm{B}\right)\ .$$

Esempio 1.7.2 Si abbia una classe con $N = 10$ studenti di cui 4 maschi e 6 femmine. Si estraggono con riposizione $n = 2$ studenti. Con questa modalità di estrazioni la composizione della popolazione resta immutata, conseguentemente il risultato della prima estrazione non influenza il risultato della seconda estrazione. Quindi $P\left(\mathrm{M}_2|\mathrm{M}_1\right) = \frac{4}{10}$.
Inoltre $P\left(\mathrm{M}_1\right) = \frac{4}{10}$ da cui

$$P\left(\mathrm{M}_1 \cap \mathrm{M}_2\right) = \frac{4}{10} \cdot \frac{4}{10} = \frac{16}{100} = 0,16\ .$$

Prima di procedere si tenga ben presente che M_1 e $\mathrm{M}_2|\mathrm{M}_1$ indicano eventi differenti. Come si può calcolare $P\left(\mathrm{M}_2\right)$ senza ricorrere all'insieme di tutti i campioni possibili? Si mostrerà che ciò è possibile ricorrendo allo schema della tabella 2×2. In effetti $\mathrm{M}_2 = \left(\mathrm{M}_1 \cap \mathrm{M}_2\right) \cup \left(\mathrm{F}_1 \cap \mathrm{M}_2\right)$.
Inoltre $\left(\mathrm{M}_1 \cap \mathrm{M}_2\right) \cap \left(\mathrm{F}_1 \cap \mathrm{M}_2\right) = \emptyset$ (si noti che $\overline{\mathrm{M}}_1 = \mathrm{F}_1$).
Pertanto

$$\begin{aligned} P\left(\mathrm{M}_2\right) &= P\left(\mathrm{M}_1 \cap \mathrm{M}_2\right) + P\left(\mathrm{F}_1 \cap \mathrm{M}_2\right) \\ &= P\left(\mathrm{M}_1\right) \cdot P\left(\mathrm{M}_2|\mathrm{M}_1\right) + P\left(\mathrm{F}_1\right) \cdot P\left(\mathrm{M}_2|\mathrm{F}_1\right)\ . \end{aligned} \tag{1.11}$$

La (1.11) vale sia per le estrazioni con riposizione che con quelle senza riposizione.
Nel caso di *estrazioni con riposizione* si ha:

$$P\left(\mathrm{M}_1\right) = \frac{4}{10}\ ,\ \ P\left(\mathrm{F}_1\right) = \frac{6}{10}\ ,\ \ P\left(\mathrm{M}_2|\mathrm{M}_1\right) = \frac{4}{10}\ ,\ \ P\left(\mathrm{M}_2|\mathrm{F}_1\right) = \frac{4}{10}\ .$$

Pertanto

$$P\left(\mathrm{M}_2\right) = \frac{4}{10} \cdot \frac{4}{10} + \frac{6}{10} \cdot \frac{4}{10} = \frac{4}{10}\left(\frac{4}{10} + \frac{6}{10}\right) = \frac{4}{10}\ .$$

Dunque dato che $P\left(\mathrm{M}_2|\mathrm{M}_1\right) = P\left(\mathrm{M}_2\right)$, si può affermare che nello schema delle estrazioni con riposizione M_2 è indipendente in probabilità da M_1.
Si noti che $P\left(\mathrm{M}_2|\mathrm{F}_1\right) = P\left(\mathrm{M}_2\right)$ e che quindi M_2 è indipendente in probabilità da $\mathrm{F}_1 = \overline{\mathrm{M}}_1$. Si noti altresì che $P\left(\mathrm{M}_1\right) = P\left(\mathrm{M}_2\right)$.

Nel caso di *estrazioni senza riposizione* si ha:

$$P\left(\mathrm{M}_1\right)=\frac{4}{10}\ ,\ \ P\left(\mathrm{F}_1\right)=\frac{6}{10}\ ,\ \ P\left(\mathrm{M}_2|\mathrm{M}_1\right)=\frac{3}{9}\ ,\ \ P\left(\mathrm{M}_2|\mathrm{F}_1\right)=\frac{4}{9}\ .$$

Pertanto

$$P\left(\mathrm{M}_2\right)=\frac{4}{10}\cdot\frac{3}{9}+\frac{6}{10}\cdot\frac{4}{9}=\frac{36}{90}=0,4\ .$$

Essendo $P\left(\mathrm{M}_2|\mathrm{M}_1\right)=\dfrac{3}{9}=0,33\bar{3}\neq P\left(\mathrm{M}_2\right)=0,4$ non si può affermare che M_2 è indipendente in probabilità da M_1.
Si noti che anche $P\left(\mathrm{M}_2|\mathrm{F}_1\right)=\dfrac{4}{9}=0,44\bar{4}\neq P\left(\mathrm{M}_2\right)=0,4$. Non si può infine affermare che M_2 è indipendente in probabilità da F_1. Si noti che anche nel campionamento senza riposizione $P\left(\mathrm{M}_1\right)=P\left(\mathrm{M}_2\right)$.

———— . . ———— . . ————

È tempo di ritornare alla teoria. Si è visto in precedenza che la probabilità dell'intersezione $\left(\mathrm{A}\cap\mathrm{B}\right)$ è sempre fornita dalla regola moltiplicativa

$$P\left(\mathrm{A}\cap\mathrm{B}\right)=P\left(\mathrm{A}\right)\cdot P\left(\mathrm{B}|\mathrm{A}\right)\ .$$

Nel caso in cui B sia indipendente da A, ovvero nel caso in cui $P\left(\mathrm{B}|\mathrm{A}\right)=P\left(\mathrm{B}\right)$, la probabilità dell'intersezione diventa

$$P\left(\mathrm{A}\cap\mathrm{B}\right)=P\left(\mathrm{A}\right)\cdot P\left(\mathrm{B}\right)\ . \tag{1.12}$$

Si può allora concludere affermando che:

$$\begin{cases} P\left(\mathrm{A}\cap\mathrm{B}\right)=P\left(\mathrm{A}\right)\cdot P\left(\mathrm{B}|\mathrm{A}\right)\ , & \text{sempre} \\ P\left(\mathrm{A}\cap\mathrm{B}\right)=P\left(\mathrm{A}\right)\cdot P\left(\mathrm{B}\right)\ , & \text{se B è indipendente in probabilità da A}\ . \end{cases}$$

L'indipendenza (in probabilità) è una relazione simmetrica, cioè se B è indipendente da A allora anche A è indipendente da B.
Dimostrazione. Per ipotesi $P\left(\mathrm{B}|\mathrm{A}\right)=P\left(\mathrm{B}\right)$, conseguentemente vale la (1.12) ovvero $P\left(\mathrm{A}\cap\mathrm{B}\right)=P\left(\mathrm{A}\right)\cdot P\left(\mathrm{B}\right)$. Pertanto

$$P\left(\mathrm{A}|\mathrm{B}\right)=\frac{P\left(\mathrm{A}\cap\mathrm{B}\right)}{P\left(\mathrm{B}\right)}\quad\text{risulta pari a}\quad P\left(\mathrm{A}|\mathrm{B}\right)=\frac{P\left(\mathrm{A}\right)P\left(\mathrm{B}\right)}{P\left(\mathrm{B}\right)}=P\left(\mathrm{A}\right)\ .$$

Si è così dimostrato che l'indipendenza di B da A implica l'indipendenza di A da B. In modo analogo si dimostra che l'indipendenza di A da B implica l'indipendenza di B da A.

Definizione (Indipendenza fra due eventi) Gli eventi A e B si dicono indipendenti se

$$P(\mathrm{A} \cap \mathrm{B}) = P(\mathrm{A}) \cdot P(\mathrm{B}) \ .$$

Da questa relazione si ricava agevolmente che:

$$\textbf{i)} \quad P(\mathrm{B}|\mathrm{A}) = P(\mathrm{B}) \ ; \qquad \textbf{ii)} \quad P(\mathrm{A}|\mathrm{B}) = P(\mathrm{A}) \ .$$

In conclusione per il calcolo delle probabilità dell'intersezione si possono usare le formule:

$$\begin{aligned} P(\mathrm{A} \cap \mathrm{B}) &= \begin{cases} P(\mathrm{A}) \cdot P(\mathrm{B}|\mathrm{A}) \\ P(\mathrm{B}) \cdot P(\mathrm{A}|\mathrm{B}) \end{cases} , \quad \text{sempre} \\ P(\mathrm{A} \cap \mathrm{B}) &= P(\mathrm{A}) \cdot P(\mathrm{B}) \ , \quad \text{solo se A e B sono indipendenti.} \end{aligned}$$

———— . . ———— . . ————

Il problema più difficile che si presenta nel valutare la probabilità dell'intersezione è "sapere" se è realistico o meno assumere che i due eventi siano indipendenti. Non esistono regole generali e spesso è solo l'esperienza che può dire se due eventi possono ritenersi indipendenti. Dovrebbe cioè accadere che in una lunga serie di prove

$$fr(\mathrm{A}|\mathrm{B}) = fr(\mathrm{A}) \ , \quad \text{ovvero} \quad \frac{n(\mathrm{A} \cap \mathrm{B})}{n(\mathrm{B})} = \frac{n(\mathrm{A})}{n} \ .$$

Esiste però un caso generale molto importante che dà luogo all'indipendenza. Questo accade quando gli eventi A e B fanno riferimento a parti fisicamente separate.

Esempio 1.7.3 Nella scatola S_A vi sono 3 palline rosse e 7 palline nere. Nella scatola S_B vi sono 5 palline rosse e 5 palline nere. Si estragga una pallina da S_A ed una pallina da S_B. Sia A l'evento estrazione di pallina rossa da S_A e B l'evento estrazione pallina rossa da S_B. Calcolare $P(\mathrm{A} \cap \mathrm{B})$.

Dato che gli eventi A e B fanno riferimento a parti fisicamente separate allora si possono ritenere indipendenti. Conseguentemente

$$\begin{aligned} P(\mathrm{A} \cap \mathrm{B}) &= P(\mathrm{A}) \cdot P(\mathrm{B}) \\ &= \frac{3}{10} \cdot \frac{5}{10} = 0,15 \ . \end{aligned}$$

———— . . ———— . . ————

Estensione della regola moltiplicativa al caso di tre eventi

Si considerino i tre eventi A, B e C. Si vuole calcolare la probabilità della loro intersezione. Si può, a tal proposito, estendere la regola moltiplicativa:

$$P(\mathrm{A} \cap \mathrm{B} \cap \mathrm{C}) = P(\mathrm{A}) \cdot P(\mathrm{B}|\mathrm{A}) \cdot P[\mathrm{C}|(\mathrm{A} \cap \mathrm{B})] \tag{1.13}$$

Esempio 1.7.4 Si abbia una scatola con 6 palline rosse e 4 palline nere. Si estraggono senza riposizione tre palline. Si determinino le probabilità di ottenere:

a) tre palline nere;

b) pallina nera alla prima estrazione, rossa alla seconda estrazione e nera alla terza estrazione;

c) una pallina nera e due rosse.

Conviene dividere l'esperimento in tre parti (ognuna per ogni estrazione) che sono – in questo esempio – fra loro dipendenti.
Sia R_i l'evento pallina rossa alla i-ma estrazione: $i = 1, 2, 3$.
Sia N_i l'evento pallina nera alla i-ma estrazione: $i = 1, 2, 3$.

a)
$$\begin{aligned} P(\mathrm{N}_1 \cap \mathrm{N}_2 \cap \mathrm{N}_3) &= P(\mathrm{N}_1) \cdot P(\mathrm{N}_2|\mathrm{N}_1) \cdot P[\mathrm{N}_3|(\mathrm{N}_1 \cap \mathrm{N}_2)] \\ &= \frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} = \frac{24}{720} \,. \end{aligned}$$

b)
$$\begin{aligned} P(\mathrm{N}_1 \cap \mathrm{R}_2 \cap \mathrm{N}_3) &= P(\mathrm{N}_1) \cdot P(\mathrm{R}_2|\mathrm{N}_1) \cdot P[\mathrm{N}_3|(\mathrm{N}_1 \cap \mathrm{R}_2)] \\ &= \frac{4}{10} \cdot \frac{6}{9} \cdot \frac{3}{8} = \frac{72}{720} \,. \end{aligned}$$

c)
$$P(\text{una pallina nera e due rosse}) = \; ?$$

$$\left\{ \begin{array}{c} \text{Una pallina nera} \\ \text{e due rosse} \end{array} \right\} = (\mathrm{N}_1 \cap \mathrm{R}_2 \cap \mathrm{R}_3) \cup (\mathrm{R}_1 \cap \mathrm{N}_2 \cap \mathrm{R}_3) \cup (\mathrm{R}_1 \cap \mathrm{R}_2 \cap \mathrm{N}_3) \;.$$

Le tre successioni di eventi sono fra loro incompatibili. Pertanto

$$P\left(\begin{array}{c} \text{una pallina nera} \\ \text{e due rosse} \end{array} \right) = P(\mathrm{N}_1 \cap \mathrm{R}_2 \cap \mathrm{R}_3) + P(\mathrm{R}_1 \cap \mathrm{N}_2 \cap \mathrm{R}_3) + P(\mathrm{R}_1 \cap \mathrm{R}_2 \cap \mathrm{N}_3) \;.$$

$$\begin{aligned} P(\mathrm{N}_1 \cap \mathrm{R}_2 \cap \mathrm{R}_3) &= \frac{4}{10} \cdot \frac{6}{9} \cdot \frac{5}{8} = \frac{120}{720} \,, \\ P(\mathrm{R}_1 \cap \mathrm{N}_2 \cap \mathrm{R}_3) &= \frac{6}{10} \cdot \frac{4}{9} \cdot \frac{5}{8} = \frac{120}{720} \,, \\ P(\mathrm{R}_1 \cap \mathrm{R}_2 \cap \mathrm{N}_3) &= \frac{6}{10} \cdot \frac{5}{9} \cdot \frac{4}{8} = \frac{120}{720} \,. \end{aligned}$$

In conclusione

$$P\begin{pmatrix}\text{una pallina nera}\\ \text{e due rosse}\end{pmatrix} = 3 \cdot \frac{120}{720}\,.$$

Se tre eventi fanno riferimento a parti fisicamente separate allora si ritengono indipendenti ed in questo caso

$$P\left(\mathrm{A} \cap \mathrm{B} \cap \mathrm{C}\right) = P\left(\mathrm{A}\right) \cdot P\left(\mathrm{B}\right) \cdot P(\mathrm{C})\,. \tag{1.14}$$

L'ultima regola moltiplicativa si può estendere al caso di k eventi. Se i k eventi $A_1, A_2, \ldots, A_k$ fanno riferimento a parti fisicamente separate allora si ritengono indipendenti ed in questo caso:

$$P\left(\mathrm{A}_1 \cap \mathrm{A}_2 \cap \cdots \cap \mathrm{A}_k\right) = P\left(\mathrm{A}_1\right) \cdot P\left(\mathrm{A}_2\right) \cdot \cdots \cdot P\left(\mathrm{A}_k\right) \tag{1.15}$$

OSSERVAZIONE IMPORTANTE. Se si estrae da una popolazione un campione di n unità con la tecnica della riposizione, si ritiene che le estrazioni siano indipendenti in quanto *con la riposizione* non si modifica la composizione della popolazione pertanto ciò che si verifica in una estrazione non influenza le prove successive.

2 *Variabili casuali*

Spesso quando si esegue un esperimento casuale si è interessati solo al valore assunto da una quantità associata ad ogni risultato elementare. Analogamente, partendo dal modello probabilistico è possibile associare con qualche "regola" ad ogni evento elementare – $e_1, e_2, \ldots, e_s$ – un valore numerico $X\left(e_i\right)$. Dato che nel modello probabilistico ad ogni evento e_i ($i = 1, \ldots, s$) è associata la probabilità $P\left(e_i\right)$, questa stessa probabilità viene associata anche a $X\left(e_i\right)$. Il prospetto sotto riportato illustra il processo logico che "trasforma" gli eventi e_i nei valori $X\left(e_i\right)$.

Modello Probabilistico		Prospetto Valori $X\left(e_i\right)$, $P\left(e_i\right)$	
e_1	$P\left(e_1\right)$	$X\left(e_1\right)$	$P\left(e_1\right)$
e_2	$P\left(e_2\right)$	$X\left(e_2\right)$	$P\left(e_2\right)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
e_i	$P\left(e_i\right)$	$X\left(e_i\right)$	$P\left(e_i\right)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
e_s	$P\left(e_s\right)$	$X\left(e_s\right)$	$P\left(e_s\right)$

Questa trasformazione dà luogo ad una variabile casuale X. La variabile casuale X è una entità complessa che assegna un valore $X(e_i)$ ad ogni evento elementare $e_i \in \Omega$. Da un punto di vista matematico $X : \Omega \to \mathcal{R}^1$.

————— . . ————— . . —————

Esempio 2.0.5 Si lanci una moneta equilibrata 5 volte e si consideri il numero X di teste che si ottengono. Per ricavare la variabile casuale (v.c.) X bisogna innanzitutto costruire il modello probabilistico che rappresenti l'esperimento. In ogni lancio si può avere T (testa) oppure C (croce). Per individuare gli eventi elementari associati ai cinque lanci basta indicare le possibili $2^5 = 32$ sequenze di T e C che si possono verificare nei cinque lanci.

Sequenze	
$\mathrm{TTTTT} = e_1$	$\mathrm{TTCCC} = e_{17}$
$\mathrm{TTTTC} = e_2$	$\mathrm{TCTCC} = e_{18}$
$\mathrm{TTTCT} = e_3$	$\mathrm{TCCTC} = e_{19}$
$\mathrm{TTCTT} = e_4$	$\mathrm{TCCCT} = e_{20}$
$\mathrm{TCTTT} = e_5$	$\mathrm{CTTCC} = e_{21}$
$\mathrm{CTTTT} = e_6$	$\mathrm{CTCTC} = e_{22}$
$\mathrm{TTTCC} = e_7$	$\mathrm{CTCCT} = e_{23}$
$\mathrm{TTCTC} = e_8$	$\mathrm{CCTTC} = e_{24}$
$\mathrm{TCTTC} = e_9$	$\mathrm{CCTCT} = e_{25}$
$\mathrm{CTTTC} = e_{10}$	$\mathrm{CCCTT} = e_{26}$
$\mathrm{TTCCT} = e_{11}$	$\mathrm{TCCCC} = e_{27}$
$\mathrm{TCTCT} = e_{12}$	$\mathrm{CTCCC} = e_{28}$
$\mathrm{CTTCT} = e_{13}$	$\mathrm{CCTCC} = e_{29}$
$\mathrm{TCCTT} = e_{14}$	$\mathrm{CCCTC} = e_{30}$
$\mathrm{CTCTT} = e_{15}$	$\mathrm{CCCCT} = e_{31}$
$\mathrm{CCTTT} = e_{16}$	$\mathrm{CCCCC} = e_{32}$

La sequenza TCTTC indica: testa al primo lancio, croce al secondo lancio, testa al terzo lancio, testa al quarto lancio e croce al quinto lancio. Significato analogo hanno le altre sequenze. Un modo per individuare tutte le sequenze è quello di suddividerle in gruppi dove all'interno di ogni gruppo si hanno le sequenze con lo stesso numero di T. Il numero di teste può assumere solo i valori: $0, 1, 2, 3, 4, 5$. Si considerino ora, come esempio, le sequenze con una testa. Per individuare le sequenze con una testa bisogna tenere presente che quest'ultima può presentarsi alla prima estrazione (nelle altre quattro estrazioni si ha C), alla seconda estrazione (nelle altre quattro estrazioni si ha C), alla terza, alla quarta o alla quinta estrazione. Si hanno così 5 casi con una testa.

Analogamente, per individuare le sequenze con due teste, bisogna individuare le coppie di estrazioni in cui si ha T (nelle altre tre si ha C). Queste coppie sono: $(1^a, 2^a)$, $(1^a, 3^a)$, $(1^a, 4^a)$, $(1^a, 5^a)$, $(2^a, 3^a)$, $(2^a, 4^a)$, $(2^a, 5^a)$, $(3^a, 4^a)$, $(3^a, 5^a)$, $(4^a, 5^a)$. Si hanno così dieci sequenze con due teste (e ovviamente tre croci). Il seguente prospetto riporta per ogni numero di teste (su cinque lanci) il numero delle sequenze.

Numero teste su cinque lanci	Numero di sequenze
0	1
1	5
2	10
3	10
4	5
5	1
Totale	32

In seguito si vedrà che il numero di sequenze aventi lo stesso numero di teste x su $n = 5$ lanci è pari alle combinazioni di 5 oggetti (diversi, le singole estrazioni) presi x alla volta. Individuati gli $s = 32$ eventi elementari si può assegnare – perché la moneta è equilibrata – ad ogni evento elementare la stessa probabilità $\dfrac{1}{32} = 0,03125$. Si tenga presente che i cinque lanci possono ritenersi indipendenti per cui la probabilità di una successione può essere ottenuta con la regola moltiplicativa, utilizzando cioè le probabilità che fanno riferimento alle singole prove $P(\mathrm{T}) = 0,5$; $P(\mathrm{C}) = 0,5$. In formula:

$$\begin{aligned} P(\mathrm{TTCCT}) &= P(\mathrm{T} \cap \mathrm{T} \cap \mathrm{C} \cap \mathrm{C} \cap \mathrm{T}) \\ &= P(\mathrm{T}) \cdot P(\mathrm{T}) \cdot P(\mathrm{C}) \cdot P(\mathrm{C}) \cdot P(\mathrm{T}) \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{32} = 0,03125 \; . \end{aligned}$$

In definitiva, la variabile casuale numero di teste in cinque lanci di una moneta equilibrata, è rappresentata nel prospetto che segue.

Modello probabilistico		Variabile casuale X		Modello probabilistico		Variabile casuale X	
Eventi elementari e_i	$P(e_i)$	$X(e_i)$	$P(e_i)$	Eventi elementari e_i	$P(e_i)$	$X(e_i)$	$P(e_i)$
TTTTT $= e_1$	$0,03125$	5	$0,03125$	TTCCC $= e_{17}$	$0,03125$	2	$0,03125$
TTTTC $= e_2$	$0,03125$	4	$0,03125$	TCTCC $= e_{18}$	$0,03125$	2	$0,03125$
TTTCT $= e_3$	$0,03125$	4	$0,03125$	TCCTC $= e_{19}$	$0,03125$	2	$0,03125$
TTCTT $= e_4$	$0,03125$	4	$0,03125$	TCCCT $= e_{20}$	$0,03125$	2	$0,03125$
TCTTT $= e_5$	$0,03125$	4	$0,03125$	CTTCC $= e_{21}$	$0,03125$	2	$0,03125$
CTTTT $= e_6$	$0,03125$	4	$0,03125$	CTCTC $= e_{22}$	$0,03125$	2	$0,03125$
TTTCC $= e_7$	$0,03125$	3	$0,03125$	CTCCT $= e_{23}$	$0,03125$	2	$0,03125$
TTCTC $= e_8$	$0,03125$	3	$0,03125$	CCTTC $= e_{24}$	$0,03125$	2	$0,03125$
TCTTC $= e_9$	$0,03125$	3	$0,03125$	CCTCT $= e_{25}$	$0,03125$	2	$0,03125$
CTTTC $= e_{10}$	$0,03125$	3	$0,03125$	CCCTT $= e_{26}$	$0,03125$	2	$0,03125$
TTCCT $= e_{11}$	$0,03125$	3	$0,03125$	TCCCC $= e_{27}$	$0,03125$	1	$0,03125$
TCTCT $= e_{12}$	$0,03125$	3	$0,03125$	CTCCC $= e_{28}$	$0,03125$	1	$0,03125$
CTTCT $= e_{13}$	$0,03125$	3	$0,03125$	CCTCC $= e_{29}$	$0,03125$	1	$0,03125$
TCCTT $= e_{14}$	$0,03125$	3	$0,03125$	CCCTC $= e_{30}$	$0,03125$	1	$0,03125$
CTCTT $= e_{15}$	$0,03125$	3	$0,03125$	CCCCT $= e_{31}$	$0,03125$	1	$0,03125$
CCTTT $= e_{16}$	$0,03125$	3	$0,03125$	CCCCC $= e_{32}$	$0,03125$	0	$0,03125$

Si è così avuta una prima rappresentazione della variabile che consiste nell'indicare le coppie di valori $X(e_i)$ e $P(e_i)$.

In molti contesti è utile rappresentare la v.c. X indicando sia i valori k diversi $x_1 < x_2 < \cdots < x_k$ che la stessa assume sia le corrispettive probabilità $p(x_1), p(x_2), \ldots, p(x_k)$, essendo $p(x_j) = P[X = x_j]$, $j = 1, 2, \ldots, k$. L'evento $(X = x_j)$ si compone di tutti gli eventi elementari in corrispondenza dei quali $X = x_j$. Pertanto

$$p(x_j) = P[X = x_j] \quad = \quad \begin{array}{c}\text{Somma delle probabilità di tutti gli eventi} \\ \text{elementari in corrispondenza dei quali } X = x_j\end{array} .$$

Ovviamente $\sum_{j=1}^{k} p(x_j) = 1$.

Numero delle teste in 5 lanci x_j	Numero sequenze con $X = x_j$	Probabilità di una sequenza	$p(x_j)$
0	1	$\frac{1}{32}$	$1 \cdot \frac{1}{32} = 0,03125$
1	5	$\frac{1}{32}$	$5 \cdot \frac{1}{32} = 0,15625$
2	10	$\frac{1}{32}$	$10 \cdot \frac{1}{32} = 0,3125$
3	10	$\frac{1}{32}$	$10 \cdot \frac{1}{32} = 0,3125$
4	5	$\frac{1}{32}$	$5 \cdot \frac{1}{32} = 0,15625$
5	1	$\frac{1}{32}$	$1 \cdot \frac{1}{32} = 0,03125$
Totali	32		$1,00000$

2.1 Variabili casuali indipendenti

Sullo stesso modello probabilistico si può definire più di una variabile casuale. Ad esempio in corrispondenza di ogni evento elementare e_i si possono avere due v.c. $X(e_i)$ e $Y(e_i)$. Dalle triplette $(X(e_i), Y(e_i), P(e_i); i = 1, 2, \ldots, s)$ si possono ricavare le due distribuzioni di probabilità delle variabili casuali X e Y:

$$\begin{array}{cc} \overbrace{\begin{array}{cr} x_1 & p(x_1) \\ \vdots & \vdots \\ x_j & p(x_j) \\ \vdots & \vdots \\ x_k & p(x_k) \\ \hline & 1,00 \end{array}}^{X} & \qquad \overbrace{\begin{array}{cr} y_1 & p(y_1) \\ \vdots & \vdots \\ y_j & p(y_j) \\ \vdots & \vdots \\ y_r & p(y_r) \\ \hline & 1,00 \end{array}}^{Y} \end{array}$$

nonché la distribuzione congiunta delle due variabili casuali che solitamente viene rappresentata in una tabella a doppia entrata (Tabella 1.13).

	x_1	x_2	$\cdots$	x_j	$\cdots$	x_k	$P(Y=y_t)$
y_1	$p(x_1,y_1)$	$p(x_2,y_1)$	$\cdots$	$p(x_j,y_1)$	$\cdots$	$p(x_k,y_1)$	$p(y_1)$
y_2	$p(x_1,y_2)$	$p(x_2,y_2)$	$\cdots$	$p(x_j,y_2)$	$\cdots$	$p(x_k,y_2)$	$p(y_2)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
y_t	$p(x_1,y_t)$	$p(x_2,y_t)$	$\cdots$	$p(x_j,y_t)$	$\cdots$	$p(x_k,y_t)$	$p(y_t)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
y_r	$p(x_1,y_r)$	$p(x_2,y_r)$	$\cdots$	$p(x_j,y_r)$	$\cdots$	$p(x_k,y_r)$	$p(y_r)$
$P(X=x_j)$	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_j)$	$\cdots$	$p(x_k)$	1

Tabella 1.13: *Distribuzioni di probabilità delle v.c. X ed Y e distribuzione congiunta delle due v.c..*

Il simbolo $p(x_j, y_t) = P[(X = x_j) \cap (Y = y_t)]$ indica una probabilità congiunta. Sommando per riga si hanno le probabilità marginali $p(y_t)$. Sommando per colonna si hanno le probabilità marginali $p(x_j)$.

Esempio 2.1.1 *Si lanci una moneta ben equilibrata $n = 5$ volte. Sia X la v.c. numero di teste totali e sia Y la v.c. numero di teste nei primi tre lanci. Ricavare la distribuzione congiunta di (X, Y).*

Conviene riportare la lista dei 32 eventi elementari con accanto i valori delle variabili X e Y (Tabella 1.14).

Eventi elementari e_i	$X(e_i)$	$Y(e_i)$	$Z(e_i)$	Eventi elementari e_i	$X(e_i)$	$Y(e_i)$	$Z(e_i)$
TTTTT $= e_1$	5	3	2	TTCCC $= e_{17}$	2	2	0
TTTTC $= e_2$	4	3	1	TCTCC $= e_{18}$	2	2	0
TTTCT $= e_3$	4	3	1	TCCTC $= e_{19}$	2	1	1
TTCTT $= e_4$	4	2	2	TCCCT $= e_{20}$	2	1	1
TCTTT $= e_5$	4	2	2	CTTCC $= e_{21}$	2	2	0
CTTTT $= e_6$	4	2	2	CTCTC $= e_{22}$	2	1	1
TTTCC $= e_7$	3	3	0	CTCCT $= e_{23}$	2	1	1
TTCTC $= e_8$	3	2	1	CCTTC $= e_{24}$	2	1	1
TCTTC $= e_9$	3	2	1	CCTCT $= e_{25}$	2	1	1
CTTTC $= e_{10}$	3	2	1	CCCTT $= e_{26}$	2	0	2
TTCCT $= e_{11}$	3	2	1	TCCCC $= e_{27}$	1	1	0
TCTCT $= e_{12}$	3	2	1	CTCCC $= e_{28}$	1	1	0
CTTCT $= e_{13}$	3	2	1	CCTCC $= e_{29}$	1	1	0
TCCTT $= e_{14}$	3	1	2	CCCTC $= e_{30}$	1	0	1
CTCTT $= e_{15}$	3	1	2	CCCCT $= e_{31}$	1	0	1
CCTTT $= e_{16}$	3	1	2	CCCCC $= e_{32}$	0	0	0

Tabella 1.14: $n = 5$ *lanci di una moneta: eventi elementari e valori assunti dalle v.c.* X, Y *e* $Z = X - Y$.

Nel prospetto che segue si è costruita la tabella a doppia entrata nelle cui caselle si è riportato il numero degli eventi elementari in corrispondenza dei quali le due variabili assumono i valori indicati nella riga e nella colonna madre.

$Y\backslash X$	0	1	2	3	4	5	
0	1	2	1	0	0	0	4
1	0	3	6	3	0	0	12
2	0	0	3	6	3	0	12
3	0	0	0	1	2	1	4
	1	5	10	10	5	1	32

Moltiplicando le frequenze di ogni casella per $\frac{1}{32}$ si ottiene la seguente distribuzione delle probabilità di (X, Y).

$Y\backslash X$	0	1	2	3	4	5	
0	$0,03125$	$0,0625$	$0,03125$	0	0	0	$0,125$
1	0	$0,09375$	$0,1875$	$0,09375$	0	0	$0,375$
2	0	0	$0,09375$	$0,1875$	$0,09375$	0	$0,375$
3	0	0	0	$0,03125$	$0,0625$	$0,03125$	$0,125$
	$0,03125$	$0,15625$	$0,3125$	$0,3125$	$0,15625$	$0,03125$	$1,000$

—————— . . —————— . . ——————

Si è ora pronti per dare la seguente

Definizione (Indipendenza in probabilità fra due v.c.) Le v.c. X ed Y sono indipendenti in probabilità se

$$p(x_j, y_t) = p(x_j) \cdot p(y_t) \ , \tag{2.1}$$

per ogni coppia (x_j, y_t), con $j = 1, 2, \ldots, k$ e $t = 1, 2, \ldots, r$.

Esempio 2.1.2 Le v.c. (X, Y) definite nell'Esempio 2.1.1 non sono indipendenti. Infatti, vi sono molte probabilità congiunte che essendo uguali a zero non possono essere uguali al prodotto delle probabilità marginali.

Esempio 2.1.3 Si lanci una moneta ben equilibrata 5 volte e si considerino la v.c. Y numero di teste sui primi tre lanci e la v.c. Z numero di teste negli ultimi due lanci. Impiegando la solita lista delle 32 sequenze (Tabella 1.14) si è ottenuto il prospetto indicante il numero degli eventi elementari e quindi la distribuzione delle probabilità congiunte delle v.c. (Y, Z).

Numero delle sequenze

$Y \backslash Z$	0	1	2	
0	1	2	1	4
1	3	6	3	12
2	3	6	3	12
3	1	2	1	4
	8	16	8	32

Distribuzione di probabilità

$Y \backslash Z$	0	1	2	
0	$0,03125$	$0,0625$	$0,03125$	$0,125$
1	$0,09375$	$0,1875$	$0,09375$	$0,375$
2	$0,09375$	$0,1875$	$0,09375$	$0,375$
3	$0,03125$	$0,0625$	$0,03125$	$0,125$
	$0,25$	$0,5$	$0,25$	$1,000$

Si controlla immediatamente che le v.c. Y ed Z sono fra loro indipendenti. Infatti, ogni probabilità congiunta è uguale al prodotto delle probabilità marginali.

—————— . . —————— . . ——————

Se due v.c. sono definite su parti fisicamente separate di un esperimento le stesse si ritengono indipendenti. Questo è proprio il caso delle v.c. Y e Z dell'Esempio 2.1.3. In effetti, Y è definita nei primi tre lanci e Z invece nel quarto e quinto. I primi tre lanci sono fisicamente separati dagli ultimi due. In altre parole l'esperimento (cinque lanci) è stato diviso in due parti fisicamente separate per cui i risultati di una delle due parti non può influenzare l'altra.

Quanto sino ad ora precisato per due v.c. si estende facilmente al caso di tre, quattro,..., variabili casuali. In particolare nel caso di tre v.c.: X (che assume i valori $x_1, \ldots, x_j, \ldots, x_k$), Y (che assume i valori $y_1, \ldots, y_t, \ldots, y_r$) e

Z (che assume i valori $z_1, \ldots, z_l, \ldots, z_c$) si *ha l'indipendenza in probabilità se le loro probabilità congiunte si fattorizzano nel prodotto delle corrispettive tre probabilità marginali*:

$$P\left[(X = x_j) \cap (Y = y_t) \cap (Z = z_l)\right] = P\left(X = x_j\right) \cdot P\left(Y = y_t\right) \cdot P\left(Z = z_l\right). \quad (2.2)$$

per ogni $j = 1, \ldots, k$, $t = 1, \ldots, r$ ed $l = 1, \ldots, c$. Se le tre v.c. fanno riferimento a parti fisicamente separate di un esperimento si ritengono indipendenti.

2.2 Aspettativa e valore atteso

Siano $e_1, \ldots, e_i, \ldots, e_s$ i risultati elementari di un esperimento casuale. Dopo aver replicato n volte l'esperimento casuale i risultati si sono presentati con le frequenze: $n\left(e_1\right), \ldots, n\left(e_i\right), \ldots, n\left(e_s\right)$; $\sum_{i=1}^{s} n\left(e_i\right) = n$. Siano $X\left(e_1\right), \ldots, X\left(e_i\right), \ldots, X\left(e_s\right)$ i valori che assume la v.c. X in corrispondenza dei risultati elementari $e_1, \ldots, e_i, \ldots, e_s$. Si calcoli la media aritmetica di questi valori.

$$\begin{aligned} M_1(X) = \frac{1}{n}\sum_{i=1}^{s} X\left(e_i\right) \cdot n\left(e_i\right) &= \sum_{i=1}^{s} X\left(e_i\right) \cdot \frac{n\left(e_i\right)}{n} \\ &= \sum_{i=1}^{s} X\left(e_i\right) \cdot fr\left(e_i\right) . \end{aligned} \quad (2.3)$$

All'aumentare di n le frequenze relative $fr\left(e_i\right)$ si stabilizzano intorno alle probabilità. Se si sostituiscono nella (2.3) le $fr\left(e_i\right)$ con le $p\left(e_i\right)$ si ottiene l'aspettativa della v.c. X fornita da

$$E(X) = \sum_{i=1}^{s} X\left(e_i\right) \cdot p\left(e_i\right) . \quad (2.4)$$

L'aspettativa $E(X)$ rappresenta il valore cui "tende" $M_1(X)$ all'aumentare di n.

Esempio 2.2.1 *Si lanci due volte un dado ben equilibrato. Si consideri la v.c. somma S dei punteggi dei due lanci. Ricavare l'aspettativa di S.*

Per poter utilizzare la (2.4) bisogna innanzitutto ricavare il modello probabilistico relativo al lancio dei due dadi e determinare $S\left(e_i\right)$.

Eventi elementari e_i	$S(e_i)$	$P(e_i)$	$S(e_i) \cdot P(e_i)$	Eventi elementari e_i	$S(e_i)$	$P(e_i)$	$S(e_i) \cdot P(e_i)$
$(1,1)=e_1$	2	1/36	$2 \cdot 1/36$	$(4,1)=e_{19}$	5	1/36	$5 \cdot 1/36$
$(1,2)=e_2$	3	$''$	$3 \cdot 1/36$	$(4,2)=e_{20}$	6	$''$	$6 \cdot 1/36$
$(1,3)=e_3$	4	$''$	$4 \cdot 1/36$	$(4,3)=e_{21}$	7	$''$	$7 \cdot 1/36$
$(1,4)=e_4$	5	$''$	$5 \cdot 1/36$	$(4,4)=e_{22}$	8	$''$	$8 \cdot 1/36$
$(1,5)=e_5$	6	$''$	$6 \cdot 1/36$	$(4,5)=e_{23}$	9	$''$	$9 \cdot 1/36$
$(1,6)=e_6$	7	$''$	$7 \cdot 1/36$	$(4,6)=e_{24}$	10	$''$	$10 \cdot 1/36$
$(2,1)=e_7$	3	$''$	$3 \cdot 1/36$	$(5,1)=e_{25}$	6	$''$	$6 \cdot 1/36$
$(2,2)=e_8$	4	$''$	$4 \cdot 1/36$	$(5,2)=e_{26}$	7	$''$	$7 \cdot 1/36$
$(2,3)=e_9$	5	$''$	$5 \cdot 1/36$	$(5,3)=e_{27}$	8	$''$	$8 \cdot 1/36$
$(2,4)=e_{10}$	6	$''$	$6 \cdot 1/36$	$(5,4)=e_{28}$	9	$''$	$9 \cdot 1/36$
$(2,5)=e_{11}$	7	$''$	$7 \cdot 1/36$	$(5,5)=e_{29}$	10	$''$	$10 \cdot 1/36$
$(2,6)=e_{12}$	8	$''$	$8 \cdot 1/36$	$(5,6)=e_{30}$	11	$''$	$11 \cdot 1/36$
$(3,1)=e_{13}$	4	$''$	$4 \cdot 1/36$	$(6,1)=e_{31}$	7	$''$	$7 \cdot 1/36$
$(3,2)=e_{14}$	5	$''$	$5 \cdot 1/36$	$(6,2)=e_{32}$	8	$''$	$8 \cdot 1/36$
$(3,3)=e_{15}$	6	$''$	$6 \cdot 1/36$	$(6,3)=e_{33}$	9	$''$	$9 \cdot 1/36$
$(3,4)=e_{16}$	7	$''$	$7 \cdot 1/36$	$(6,4)=e_{34}$	10	$''$	$10 \cdot 1/36$
$(3,5)=e_{17}$	8	$''$	$8 \cdot 1/36$	$(6,5)=e_{35}$	11	$''$	$11 \cdot 1/36$
$(3,6)=e_{18}$	9	$''$	$9 \cdot 1/36$	$(6,6)=e_{36}$	12	$''$	$12 \cdot 1/36$
							$252/36 = 7$

Dal prospetto si ricava

$$\begin{aligned} E(S) &= \sum_{i=1}^{36} S(e_i) \cdot P(e_i) = \sum_{i=1}^{36} S(e_i) \cdot \frac{1}{36} \\ &= 252 \cdot \frac{1}{36} = 7 \,. \end{aligned}$$

Il risultato ottenuto informa che effettuando un numero n elevato di coppie di lanci la media aritmetica della somma dei due punteggi è prossima a 7.

———— . . ———— . . ————

Alcune volte può essere più conveniente ricavare $E(X)$ facendo riferimento alla distribuzione di probabilità di X. In effetti partendo dalla (2.4)

$$E(X) = \sum_{i=1}^{s} X(e_i) \cdot p(e_i) \ ,$$

è possibile raggruppare gli eventi elementari in k gruppi, $1 \leq k \leq s$. Nel gruppo j-mo ($j = 1, 2, \ldots, k$) si hanno gli eventi elementari con $X(e_i) = x_j$. Pertanto

$$\begin{aligned}
E(S) &= \sum_{j=1}^{k} \left\{ \sum_{\{i:X(e_i)=x_j\}} X(e_i) \cdot p(e_i) \right\} \\
&= \sum_{j=1}^{k} \left\{ \sum_{\{i:X(e_i)=x_j\}} x_j \cdot p(e_i) \right\} = \sum_{i=1}^{k} x_j \cdot \left\{ \sum_{\{i:X(e_i)=x_j\}} p(e_i) \right\} \\
&= \sum_{j=1}^{k} x_j \cdot p(x_j) \ .
\end{aligned}$$

In conclusione è possibile calcolare l'aspettativa di una v.c. anche con la formula

$$E(X) = \sum_{j=1}^{k} x_j \cdot p(x_j) \ . \tag{2.5}$$

Esempio 2.2.2 *Calcolare l'aspettativa della v.c. S dell'esempio precedente con la (2.5).*

Per poter applicare la (2.5) bisogna innanzitutto ricavare la distribuzione della v.c. S che si ottiene con il prospetto in Tabella 1.15 che riporta nell'ultima colonna i prodotti $s_j \cdot p(s_j)$.

s_j	Eventi elementari	Numero eventi	$p(s_j)$	$s_j \cdot p(s_j)$
2	$(1,1)$	1	$1/36$	$2/36$
3	$(1,2)$; $(2,1)$	2	$2/36$	$6/36$
4	$(1,3)$; $(2,2)$; $(3,1)$	3	$3/36$	$12/36$
5	$(1,4)$; $(2,3)$; $(3,2)$; $(4,1)$	4	$4/36$	$20/36$
6	$(1,5)$; $(2,4)$; $(3,3)$; $(4,2)$; $(5,1)$	5	$5/36$	$30/36$
7	$(1,6)$; $(2,5)$; $(3,4)$; $(4,3)$; $(5,2)$; $(6,1)$	6	$6/36$	$42/36$
8	$(2,6)$; $(3,5)$; $(4,4)$; $(5,3)$; $(6,2)$	5	$5/36$	$40/36$
9	$(3,6)$; $(4,5)$; $(5,4)$; $(6,3)$	4	$4/36$	$36/36$
10	$(4,6)$; $(5,5)$; $(6,4)$	3	$3/36$	$30/36$
11	$(5,6)$; $(6,5)$	2	$2/36$	$22/36$
12	$(6,6)$	1	$1/36$	$12/36$
		36	$1,00$	$252/36 = 7$

Tabella 1.15: *Distribuzione di probabilità della v.c. S.*

—————— . . —————— . . ——————

In conclusione, per il calcolo dell'aspettativa di una v.c., si dispone delle due formule

$$E(X) = \begin{cases} \displaystyle\sum_{i=1}^{s} X(e_i) \cdot p(e_i) & (2.4) \\ \displaystyle\sum_{j=1}^{k} x_j \cdot p(x_j) \ . & (2.5) \end{cases}$$

Dato che $E(X)$ ha la stessa struttura matematica di $M_1(X)$, deriva che $E(X)$ ha le stesse proprietà di $M_1(X)$ studiate in statistica descrittiva. In particolare sono utili le seguenti proprietà di $E(X)$.

$$\textbf{i)} \qquad E(a + bX) = a + b \cdot E(X) \ . \tag{2.6}$$

$$\textbf{ii)} \qquad E(X + Y) = E(X) + E(Y) \ . \tag{2.7}$$

Dimostrazioni.

i) Sia $Y = a + bX$, allora $Y(e_i) = a + b \cdot X(e_i)$. Pertanto

$$\begin{aligned} E(Y = a + bX) &= \sum_{i=1}^{s} [a + b \cdot X(e_i)] \cdot p(e_i) \\ &= \sum_{i=1}^{s} a \cdot p(e_i) + \sum_{i=1}^{s} b \cdot X(e_i) \cdot p(e_i) \\ &= a \cdot \sum_{i=1}^{s} p(e_i) + b \cdot \sum_{i=1}^{s} X(e_i) \cdot p(e_i) \\ &= a + b \cdot E(X) \end{aligned}$$

La relazione ora trovata vale per qualsiasi valore di a e di b. In particolare:

- per $b = 0 \rightarrow E(a + 0 \cdot X) = a + 0 \cdot E(X) = a$;
- per $b = 1 \rightarrow E(a + 1 \cdot X) = a + 1 \cdot E(X) = a + E(X)$;
- per $a = 0 \rightarrow E(0 + b \cdot X) = 0 + b \cdot E(X) = b \cdot E(X)$.

ii) $E(X + Y) = E(X) + E(Y)$.

Sia $Z = X + Y$, ovvero $Z(e_i) = X(e_i) + Y(e_i)$. Pertanto

$$\begin{aligned} E(X+Y) &= E(Z) = \sum_{i=1}^{s} Z(e_i) \cdot p(e_i) \\ &= \sum_{i=1}^{s} [X(e_i) + Y(e_i)] \cdot p(e_i) \\ &= \sum_{i=1}^{s} X(e_i)\, p(e_i) + \sum_{i=1}^{s} Y(e_i) \cdot p(e_i) \\ &= E(X) + E(Y)\ . \end{aligned}$$

Questa proprietà è molto importante perché permette di calcolare l'aspettativa di una somma in base alle aspettative delle sue componenti il cui calcolo è in molti casi più immediato.

Esempio 2.2.3 *Si lanci due volte un dado bene equilibrato. Sia X il punteggio che si consegue con il primo lancio e sia Y il punteggio che si consegue con il secondo lancio. Sia $S = X + Y$. Ricavare l'aspettativa di $S = X + Y$ con la (2.7).*

Il modello probabilistico associato al primo lancio dà luogo alla distribuzione di probabilità per la v.c. X riportata in Tabella (1.16).

Eventi elementari $e_i = x_i$	$p(x_i)$	$x_i \cdot p(x_i)$
1	$1/6$	$1 \cdot 1/6$
2	$1/6$	$2 \cdot 1/6$
3	$1/6$	$3 \cdot 1/6$
4	$1/6$	$4 \cdot 1/6$
5	$1/6$	$5 \cdot 1/6$
6	$1/6$	$5 \cdot 1/6$
	$1,00$	$E(X) = 3,5$

Tabella 1.16: *Distribuzione di probabilità della v.c. X.*

La v.c. Y ha la stessa distribuzione di probabilità della v.c. X. Pertanto $E(Y) = 3,5$. In conclusione

$$E(S) = E(X) + E(Y) = 3,5 + 3,5 = 7\ .$$

———— . . ———— . . ————

La (2.7) si generalizza agevolmente al caso della somma di c variabili casuali

$$E\left(X_1 + X_2 + \cdots + X_c\right) = E\left(X_1\right) + E\left(X_2\right) + \cdots + E\left(X_c\right) \ . \tag{2.8}$$

——— . . ——— . . ———

L'aspettativa del prodotto fra due v.c. si può calcolare con le formule

$$E(X \cdot Y) = \sum_{i=1}^{s} X\left(e_i\right) \cdot Y\left(e_i\right) \cdot p\left(e_i\right) \tag{2.9}$$

$$= \sum_{j=1}^{k} \sum_{t=1}^{r} x_j \cdot y_t \cdot p\left(x_j, y_t\right) \ . \tag{2.10}$$

Nel caso del prodotto fra variabili casuali generalmente si ha:

$$E(X \cdot Y) \neq E(X) \cdot E(Y) \ .$$

Tuttavia vale il seguente *Teorema*.

Teorema (Valore atteso del prodotto di v.c. indipendenti) Se X ed Y sono indipendenti allora

$$E(X \cdot Y) = E(X) \cdot E(Y) \ . \tag{2.11}$$

Dimostrazione. La v.c. X assume i valori $x_1, \ldots, x_j, \ldots, x_k$. La v.c. Y assume i valori $y_1, \ldots, y_t, \ldots, y_r$. Per l'ipotesi di indipendenza $p\left(x_j, y_t\right) = p\left(x_j\right) \cdot p\left(y_t\right)$, per ogni coppia (j, t). Pertanto, in forza della (2.10)

$$\begin{aligned} E(X \cdot Y) &= \sum_{j=1}^{k} \sum_{t=1}^{r} x_j \cdot y_t \cdot p\left(x_j, y_t\right) \\ &= \sum_{j=1}^{k} \sum_{t=1}^{r} x_j \cdot y_t \cdot p\left(x_j\right) \cdot p\left(y_t\right) \\ &= \sum_{j=1}^{k} x_j \cdot p\left(x_j\right) \cdot \sum_{t=1}^{r} y_t \cdot p\left(y_t\right) = E(X) \cdot E(Y) \ . \end{aligned}$$

Esempio 2.2.4 *Si lanci una moneta equilibrata $n = 5$ volte. Sia X la v.c. numero di teste totali e sia Y la v.c. numero di teste nei primi tre lanci. Ricavare l'aspettativa del loro prodotto con la (2.10).*

Per poter calcolare l'aspettativa del prodotto bisogna ricavare, per ogni coppia (x_j, y_t), il valore del prodotto ponderato $x_j \cdot y_t \cdot p\left(x_j, y_t\right)$. A tal fine si può impiegare il prospetto che segue ove le probabilità congiunte sono quelle ricavate in precedenza.

$y \backslash x$	0	1	2	3	4	5
0	$0,00$	$0,00$	$0,00$	$0,00$	$0,00$	$0,00$
1	$0,00$	$1 \cdot 1 \cdot 0,09375$	$1 \cdot 2 \cdot 0,1875$	$1 \cdot 3 \cdot 0,09375$	$0,00$	$0,00$
2	$0,00$	$0,00$	$2 \cdot 2 \cdot 0,09375$	$2 \cdot 3 \cdot 0,1875$	$2 \cdot 4 \cdot 0,09375$	$0,00$
3	$0,00$	$0,00$	$0,00$	$3 \cdot 3 \cdot 0,03125$	$3 \cdot 4 \cdot 0,0625$	$3 \cdot 5 \cdot 0,03125$

L'aspettativa $E(X \cdot Y)$, fornita dalla sommatoria dei $6 \times 4 = 24$ prodotti $x_j \cdot y_t \cdot p\left(x_j, y_t\right)$, risulta $7,875$.
L'aspettativa di X risulta

$$\begin{aligned}
E(X) &= \sum_{j=1}^{6} x_j \cdot p\left(x_j\right) \\
&= 0 \cdot 0,03125 + 1 \cdot 0,15625 + 2 \cdot 0,3125 + 3 \cdot 0,3125 + \\
&\quad +4 \cdot 0,15625 + 5 \cdot 0,03125 \\
&= 2,5\ .
\end{aligned}$$

L'aspettativa di Y risulta

$$\begin{aligned}
E(Y) &= \sum_{t=1}^{4} y_t \cdot p\left(y_t\right) \\
&= 0 \cdot 0,125 + 1 \cdot 0,375 + 2 \cdot 0,375 + 3 \cdot 0,125 \\
&= 1,5\ .
\end{aligned}$$

Si rileva che per le variabili (X, Y) risulta

$$\begin{aligned}
E(X \cdot Y) &\neq E(X) \cdot E(Y) \\
7,875 &\neq 3,75\ .
\end{aligned}$$

Esempio 2.2.5 *Si lanci una moneta equilibrata $n = 5$ volte. Sia Y la v.c. numero di teste nei primi tre lanci e Z la v.c. numero di teste negli ultimi due lanci. Si è visto in precedenza che Y e Z sono v.c. indipendenti. Calcolare $E(Y \cdot Z)$ con la formula*

$$E(Y \cdot Z) = \sum_{t=1}^{4} \sum_{l=1}^{3} y_t \cdot z_l \cdot p\left(y_t, z_l\right)\ , \tag{2.10}$$

e verificare che $E(Y \cdot Z) = E(Y) \cdot E(Z)$.

$y \backslash z$	0	1	2
0	$0,00$	$0,00$	$0,00$
1	$0,00$	$1 \cdot 1 \cdot 0,1875$	$1 \cdot 2 \cdot 0,09375$
2	$0,00$	$2 \cdot 1 \cdot 0,1875$	$2 \cdot 2 \cdot 0,09375$
3	$0,00$	$3 \cdot 1 \cdot 0,0625$	$3 \cdot 2 \cdot 0,03125$

$\boxed{E(Y \cdot Z) = 1,5}$

Si ricava immediatamente che $E(Y) = 1,5$ e che $E(Z) = 1$. Si verifica effettivamente quanto previsto dalla (2.11) nel caso di v.c. indipendenti.

2.3 Varianza

Oltre all'aspettativa riveste una notevole importanza nell'inferenza anche la varianza di una v.c.. Siano $e_1, \ldots, e_i, \ldots, e_s$ i risultati elementari di un esperimento casuale. Siano $fr(e_1), \ldots, fr(e_i), \ldots, fr(e_s)$ le frequenze relative degli eventi elementari ottenute dopo aver replicato l'esperimento n volte. Siano $X(e_1), \ldots, X(e_i), \ldots, X(e_s)$ i valori che assume la v.c. X in corrispondenza degli eventi elementari.

La varianza riscontrata dai valori $X(e_i)$ nelle n prove è fornita da

$$\begin{matrix}\text{Varianza} \\ \text{sperimentale}\end{matrix} = \sum_{i=1}^{s} \left[X(e_i) - M_1(X)\right]^2 \cdot fr(e_i) \ .$$

All'aumentare del numero di prove le frequenze relative "tendono" alle probabilità $p(e_i)$ e la media aritmetica sperimentale $M_1(X)$ tende all'aspettativa $E(X)$. Sostituendo nella espressione precedente le $fr(e_i)$ con le $p(e_i)$ e la $M_1(X)$ con $E(X)$ si ottiene la varianza della v.c. X

$$Var(X) = \sigma^2(X) = \sum_{i=1}^{s} \left[X(e_i) - E(X)\right]^2 \cdot p(e_i) \ . \tag{2.12}$$

La (2.12) informa che la varianza della v.c. X altro non è che l'aspettativa degli scarti al quadrato $[X(e_i) - E(X)]^2$. Si può allora rappresentare la varianza come segue

$$\sigma^2(X) = Var(X) = E\left\{(X - \mu)^2\right\} \ , \tag{2.13}$$

ove si è posto $\mu = E(X)$.

Se si dispone della distribuzione di probabilità della v.c. X, la varianza si può calcolare anche con la formula

$$Var(X) = \sum_{j=1}^{k} (x_j - \mu)^2 \cdot p(x_j) \ . \tag{2.14}$$

——————— . . ——————— . . ———————

Dalla (2.13) si ricava – utilizzando le proprietà del valore atteso (2.6) e (2.7) – il procedimento indiretto per il calcolo della varianza

$$\begin{aligned} E\left\{(X-\mu)^2\right\} &= E\left\{X^2+\mu^2-2\mu X\right\} \\ &= E\left(X^2\right)+E\left(\mu^2\right)-2\mu E\left(X\right) \\ &= E\left(X^2\right)+\mu^2-2\mu^2 \\ &= E\left(X^2\right)-\mu^2 \end{aligned} \tag{2.15}$$

——————— . . ——————— . . ———————

Saranno molto utili in seguito le seguenti proprietà della varianza.

i) $Var(a+bX)=b^2\cdot Var(X)\ .$ (2.16)

ii) $Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)\ .$ (2.17)

Dimostrazioni.

i) Sia $Y=a+bX$, allora $Y\left(e_i\right)=a+bX\left(e_i\right)$. Pertanto

$$\begin{aligned} Var(Y=a+bX) &= \sum_{i=1}^{s}\left\{\left[a+bX\left(e_i\right)\right]-\left[a+b\mu\right]\right\}^2\cdot p\left(e_i\right) \\ &= \sum_{i=1}^{s} b^2\cdot\left[X\left(e_i\right)-\mu\right]^2\cdot p\left(e_i\right) \\ &= b^2\cdot\sum_{i=1}^{s}\left[X\left(e_i\right)-\mu\right]^2\cdot p\left(e_i\right) \\ &= b^2\cdot Var(X) \end{aligned}$$

La relazione ora trovata vale per qualsiasi valore di a e di b. In particolare:

- per $b=0\rightarrow Var(a+0\cdot X)=Var(a)=0^2\cdot Var(X)=0$; (la varianza di una v.c. costante $X=a$ è uguale a 0);
- per $b=1\rightarrow Var(a+1\cdot X)=1^2\cdot Var(X)=Var(X)$;
- per $a=0\rightarrow E(0+b\cdot X)=b^2\cdot Var(X)$.

ii) $Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$.

Per comodità si pone: $Z = X + Y$, $E(Y) = \eta$, $E(X) = \mu$. È noto che $E(Z) = E(X + Y) = \mu + \eta$. Dalla (2.13) si ottiene

$$\begin{aligned} Var(X+Y) &= E\left\{[(X+Y)-(\mu+\eta)]^2\right\} \\ &= E\left\{[(X-\mu)+(Y-\eta)]^2\right\} \\ &= E\left\{(X-\mu)^2\right\} + E\left\{(Y-\eta)^2\right\} + 2E\left\{(X-\mu)(Y-\eta)\right\} \end{aligned}$$

L'aspettativa del prodotto fra gli scarti $(X - \mu)$ ed $(Y - \eta)$ è denominata covarianza fra X ed Y. In conclusione

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y) \ .$$

La $Cov(X, Y)$ si può calcolare con le formule

$$\begin{aligned} Cov(X,Y) &= \sum_{i=1}^{s} [X(e_i) - \mu][Y(e_i) - \eta] \cdot p(e_i) && (2.18) \\ &= \sum_{j=1}^{k} \sum_{t=1}^{r} (x_j - \mu)(y_t - \eta) \cdot p(x_j, y_t) \ . && (2.19) \end{aligned}$$

È immediato verificare che

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y) \ .$$

———— . . ———— . . ————

Anche per la covarianza vi è un procedimento indiretto

$$\begin{aligned} Cov(X,Y) &= E\{(X-\mu)(Y-\eta)\} \\ &= E\{X \cdot Y - X \cdot \eta - \mu \cdot Y + \mu \cdot \eta\} \\ &= E(X \cdot Y) - \eta \cdot E(X) - \mu \cdot E(Y) + E(\mu \cdot \eta) \\ &= E(X \cdot Y) - \eta \cdot \mu - \mu \cdot \eta + \mu \cdot \eta \\ &= E(X \cdot Y) - \mu \cdot \eta \ . && (2.20) \end{aligned}$$

Teorema (Covarianza fra due v.c. indipendenti) Se X ed Y sono indipendenti allora

$$Cov(X, Y) = 0 \ .$$

Dimostrazione. Per ipotesi X ed Y sono indipendenti. Consegue dalla (2.11) che $E(X \cdot Y) = \mu \cdot \eta$. Sostituendo questa relazione nella (2.20) si ha

$$Cov(X, Y) = \mu \cdot \eta - \mu \cdot \eta = 0 \ .$$

———— . . ———— . . ————

Conseguentemente, *se X ed Y sono v.c. indipendenti allora*

$$Var(X+Y) = Var(X) + Var(Y) \ . \tag{2.21}$$

La (2.21) si generalizza al caso di c v.c. indipendenti.

Teorema (Varianza della somma di v.c. indipendenti) Se $X_1, X_2, \ldots, X_c$ sono fra loro indipendenti allora si può dimostrare che

$$Var\left(X_1 + X_2 + \cdots + X_c\right) = Var\left(X_1\right) + Var\left(X_2\right) + \cdots + Var\left(X_c\right) \ . \tag{2.22}$$

———— . . ———— . . ————

Si svolgeranno ora alcuni esercizi sulla varianza.

Esempio 2.3.1 *Sia S la v.c. somma dei punteggi ottenibili in due lanci di un dado ben equilibrato. Ricavare la varianza di S con la (2.14).*

In precedenza si è determinato il valore dell'aspettativa $E(S) = 7$. In Tabella 1.17 si sono riportati i calcoli necessari per l'applicazione della (2.14).

s_j	$(s_j - 7)$	$(s_j - 7)^2$	$p\left(s_j\right)$	$(s_j - 7)^2 \cdot p\left(s_j\right)$
2	-5	25	$1/36$	$25/36$
3	-4	16	$2/36$	$32/36$
4	-3	9	$3/36$	$27/36$
5	-2	4	$4/36$	$16/36$
6	-1	1	$5/36$	$5/36$
7	0	0	$6/36$	0
8	1	1	$5/36$	$5/36$
9	2	4	$4/36$	$16/36$
10	3	9	$3/36$	$27/36$
11	4	16	$2/36$	$32/36$
12	5	25	$1/36$	$25/36$
			$1,00$	$210/36 = Var(S)$

Tabella 1.17: *Calcoli necessari per la determinazione della varianza di S tramite la (2.14).*

Esempio 2.3.2 *Sia S la v.c. somma dei punteggi ottenibili in due lanci di un dado ben equilibrato. Ricavare la $Var(S)$ con la (2.21).*

La v.c. X punteggio che si consegue con il primo lancio e la v.c. Y punteggio che si consegue con il secondo lancio sono fra loro indipendenti. Pertanto

$$Var(S) = Var(X+Y) = Var(X) + Var(Y) \ .$$

Inoltre la v.c. Y ha la stessa distribuzione di probabilità della v.c. X, per cui

$$Var(S) = 2 \cdot Var(X) \ .$$

La Tabella 1.18 riporta i calcoli per la determinazione della $Var(X)$. In precedenza si era calcolata l'aspettativa di X: $E(X) = 3,5$.

x_j	$(x_j - 3,5)$	$(x_j - 3,5)^2$	$p(x_j)$	$(x_j - 3,5)^2 \cdot p(x_j)$
1	$-2,5$	$6,25$	$1/6$	$6,25/6$
2	$-1,5$	$2,25$	$1/6$	$2,25/6$
3	$-0,5$	$0,25$	$1/6$	$0,25/6$
4	$+0,5$	$0,25$	$1/6$	$0,25/6$
5	$+1,5$	$2,25$	$1/6$	$2,25/6$
6	$+2,5$	$6,25$	$1/6$	$6,25/6$
			$1,00$	$17,5/6 = Var(X)$

Tabella 1.18: *Calcoli necessari per la determinazione della varianza di X.*

Si ha così $Var(X) = \dfrac{17,5}{6}$ e quindi

$$Var(S) = 2 \cdot Var(X) = 2 \cdot \frac{17,5}{6} = \frac{35}{6} = \frac{210}{36}$$

————— . . ————— . . —————

Come è noto per misurare la variabilità di una v.c. X si usano gli scarti medi assoluti. Fra questi il più impiegato è lo scarto quadratico medio che coincide con la media quadratica del modulo degli scarti.

$$\begin{aligned} \text{Scarto quadratico medio} &= M_2\{|X - \mu|\} = \sqrt{E(X-\mu)^2} \\ &= \sqrt{Var(X)} \ . \end{aligned}$$

Lo scarto quadratico medio di una v.c. si indica con $\sigma(X)$. Si ricordi che $\sigma(X)$ è espresso nella stessa unità di misura di X, mentre la varianza è espressa con il quadrato dell'unità di misura di X.

Esempio 2.3.3 *Si ricavi $\sigma(X)$, essendo S il punteggio ottenuto lanciando due volte un dado ben equilibrato.*

Si ricava $\sigma(X) = \sqrt{\dfrac{210}{36}} = 2,4152$. Si può allora affermare che lanciando due volte un dado (ben equilibrato) il punteggio complessivo S assume valori mediamente pari a 7. Inoltre gli scarti $|S - \mu|$ valgono mediamente (in media quadratica) $2,4152$.

3 *Elementi di calcolo combinatorio*

Nella specificazione delle leggi di probabilità di alcune variabili casuali vengono impiegate le disposizioni, le permutazioni e le combinazioni.

3.1 Le disposizioni

Il simbolo $(N)_n$ indica le disposizioni di N oggetti (diversi) presi n alla volta, $n \leq N$. Le disposizioni sono tutti i gruppi che si possono formare prendendo n degli N oggetti dati. Due gruppi qualsiasi differiscono: per almeno un elemento o per l'ordine in cui gli stessi sono disposti. Si mostrerà che

$$(N)_n = N \cdot (N-1) \cdot \cdots \cdot (N-n+1) \; . \tag{3.1}$$

Dimostrazione. Si considerino N oggetti diversi $a_1, a_2, \ldots, a_N$. Con $n = 1$ si possono formare i seguenti N gruppi ($(N)_1 = N$ disposizioni):

$$(a_1)\,, (a_2)\,, \ldots, (a_N) \; .$$

Con $n = 2$ si possono formare $(N)_2 = N \cdot (N-1)$ gruppi in quanto a ciascun gruppo con un oggetto si associa (nella seconda posizione) uno fra i restanti $(N-1)$ oggetti. Con $n = 3$ si possono formare $(N)_3 = N \cdot (N-1) \cdot (N-2)$ gruppi in quanto ad ogni gruppo possibile di numerosità due si può associare (nella terza posizione) uno qualsiasi fra i restanti $(N-2)$ oggetti. Procedendo in questo modo si riscontra che il numero delle disposizioni $(N)_n$ è fornito da:

$$(N)_n = N \cdot (N-1) \cdot \cdots \cdot (N-n+1) \; .$$

Esempio 3.1.1 *Siano $(a_1, a_2, a_3, a_4, a_5, a_6)$ le unità di una popolazione. Formare i campioni di numerosità $n = 2$ nel caso di estrazioni senza riposizione.*

Il generico campione viene indicato con la coppia

$$(a_j, a_i) \quad \text{con} \quad \begin{cases} j = 1,2,3,4,5,6 \\ i = 1,2,3,4,5,6 \quad (i \neq j) \end{cases} ,$$

dove a_j ed a_i indicano rispettivamente il primo ed il secondo elemento estratto. Il numero dei campioni possibili è

$$(6)_2 = 6 \cdot 5 = 30 \; .$$

Si osservi che due campioni differiscono per:

- l'ordine di estrazione: $(a_1, a_3) \neq (a_3, a_1)$;

- la composizione: $(a_1, a_3) \neq (a_1, a_4)$.

Esempio 3.1.2 *Considerando la popolazione dell'esempio precedente, formare (sempre nel caso di estrazioni senza riposizione) i campioni di numerosità* $n = 3$*.*

Il numero di campioni possibili è

$$(6)_3 = 6 \cdot 5 \cdot 4 = 120 \; .$$

Il generico campione viene indicato con la tripletta

$$(a_j, a_i, a_t) \quad \text{con} \quad \begin{cases} j = 1, 2, 3, 4, 5, 6 \\ i = 1, 2, 3, 4, 5, 6 \quad (i \neq j) \\ t = 1, 2, 3, 4, 5, 6 \quad (t \neq j; t \neq i) \end{cases} \; .$$

I seguenti due campioni sono diversi per la posizione delle unità: $(a_2, a_1, a_5) \neq (a_1, a_2, a_5)$. I seguenti due campioni sono diversi per la composizione: $(a_4, a_2, a_1) \neq (a_5, a_2, a_3)$.

Esempio 3.1.3 *Si determini il numero delle parole ottenibili utilizzando* $n = 4$ *delle seguenti lettere:* A, C, E, I, N, O*.*

Il numero delle parole di quattro lettere coincide con $(6)_4 = 6 \cdot 5 \cdot 4 \cdot 3 = 360$ parole. Ecco alcune di queste parole: $(CENA)$, $(CINA)$, $(CAIO)$, $(CONI)$, $(NOCE)$, $(NOCI)$, $(ACNE)$.

3.2 Le permutazioni

Si considerino le disposizioni

$$(N)_n = N \cdot (N - 1) \cdot \dots \cdot (N - n + 1) \; ,$$

e si ponga $n = N$. Si ha

$$\begin{aligned} (N)_N &= N \cdot (N - 1) \cdot \dots \cdot (N - N + 1) \\ &= N \cdot (N - 1) \cdot \dots \cdot 2 \cdot 1 = N! \; . \end{aligned} \tag{3.2}$$

Il prodotto $N \cdot (N-1) \cdot \dots \cdot 2 \cdot 1$ viene indicato con $N!$. Il fattoriale di N fornisce il numero delle permutazioni di N oggetti diversi. Si osservi che ogni permutazione comprende tutti gli N oggetti. Pertanto una permutazione differisce dall'altra solo per la posizione degli N oggetti.

Esempio 3.2.1 *Si abbia una popolazione con* $N = 3$*. Si elenchino i campioni possibili di numerosità* $n = 3$*, nel caso di campionamento senza riposizione.*

Il numero di campioni possibili è dato dalle permutazioni di tre oggetti: $3! = 3 \cdot 2 \cdot 1 = 6$. La lista dei campioni possibili è la seguente: (a_1, a_2, a_3), (a_1, a_3, a_2), (a_2, a_1, a_3), (a_2, a_3, a_1), (a_3, a_1, a_2), (a_3, a_2, a_1).

Esempio 3.2.2 *Elencare le parole che si possono formulare con le lettere: D, E, U (si osservi che alcune parole potrebbero non avere nessun significato nella lingua italiana).*

La lista delle parole è la seguente: (DEU), (DUE), (EDU), (EUD), (UDE), (UED).

3.3 Le combinazioni

Il simbolo

$$\binom{N}{n} = \frac{(N)_n}{n!}\,, \qquad (n \leq N) \tag{3.3}$$

indica le combinazioni di N oggetti presi n alla volta. Due combinazioni differiscono per almeno un elemento.

Esempio 3.3.1 *Si abbiano $N = 5$ oggetti $(a_1, a_2, a_3, a_4, a_5)$. Si elenchino le combinazioni con $n = 3$.*

Bisogna, in altre parole, individuare tutte le triplette che differiscono per la composizione: $\{a_1, a_2, a_3\}$, $\{a_1, a_2, a_4\}$, $\{a_1, a_2, a_5\}$, $\{a_1, a_3, a_4\}$, $\{a_1, a_3, a_5\}$, $\{a_1, a_4, a_5\}$, $\{a_2, a_3, a_4\}$, $\{a_2, a_3, a_5\}$, $\{a_2, a_4, a_5\}$, $\{a_3, a_4, a_5\}$.
Applicando la (3.3) si ha $\displaystyle\binom{5}{3} = \frac{(5)_3}{3!} = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10$.

Nel caso di estrazioni "in blocco" di n unità da una popolazione di N unità, il numero dei campioni possibili è fornito dalle combinazioni $\displaystyle\binom{N}{n}$ di N oggetti presi n alla volta. Con la tecnica delle estrazioni in blocco le n unità vengono estratte contemporaneamente, viene cioè a mancare l'ordine di estrazione. Conseguentemente due campioni estratti in blocco possono differire solo per la composizione.

———— . . ———— . . ————

Si cercherà di dare ora una giustificazione alla (3.3). Per comodità si farà riferimento all'Esempio 3.3.1. Da ogni combinazione (tripletta non ordinata) è possibile ricavare $3!$ triplette ordinate (queste triplette differiscono fra loro solo per l'ordine di presentazione). In particolare dalla tripletta non ordinata $\{a_1, a_2, a_3\}$ si ottengono le seguenti $3! = 6$ triplette ordinate: (a_1, a_2, a_3), (a_1, a_3, a_2), (a_2, a_1, a_3), (a_2, a_3, a_1), (a_3, a_1, a_2), (a_3, a_2, a_1). Consegue che il numero delle triplette ordinate ottenibili dalle $\binom{5}{3} = 10$ triplette non ordinate sono pari a

$$\binom{5}{3} \cdot 3! = 10 \cdot 6 = 60\ .$$

In generale il prodotto

$$\underbrace{\binom{N}{n}}_{\substack{\text{Combinazioni}\\ \text{(differiscono}\\ \text{per almeno}\\ \text{un elemento)}}} \times \underbrace{n!}_{\substack{\text{Permutazioni}\\ \text{(differiscono}\\ \text{per}\\ \text{l'ordinamento)}}} ,$$

fornisce il numero dei gruppi di n oggetti ottenibili da N oggetti diversi. Questi gruppi differiscono per la loro composizione o per il loro ordinamento. Il numero di questi gruppi coincide con le disposizioni di N oggetti (diversi) presi n alla volta. In definitiva $(N)_n = \binom{N}{n} \cdot n!$. Da cui si ricava $\binom{N}{n} = \frac{(N)_n}{n!}$.

Esempio 3.3.2 *Si abbia una popolazione con* $N = 4$ *unità statistiche* (a_1, a_2, a_3, a_4). *Si elenchino i campioni senza riposizione e quelli in blocco nel caso* $n = 2$.

Campioni in blocco	Campioni senza riposizione
$\{a_1, a_2\}$	$(a_1, a_2); (a_2, a_1)$
$\{a_1, a_3\}$	$(a_1, a_3); (a_3, a_1)$
$\{a_1, a_4\}$	$(a_1, a_4); (a_4, a_1)$
$\{a_2, a_3\}$	$(a_2, a_3); (a_3, a_2)$
$\{a_2, a_4\}$	$(a_2, a_4); (a_4, a_2)$
$\{a_3, a_4\}$	$(a_3, a_4); (a_4, a_3)$
$\binom{4}{2} = 6$	$(4)_2 = 12$

——— · · ——— · · ———

Si consideri la (3.3) che fornisce il numero delle combinazioni di N oggetti presi n alla volta:

$$\binom{N}{n} = \frac{(N)_n}{n!} = \frac{N \cdot (N-1) \cdot \dots \cdot (N-n+1)}{n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1}\,, \tag{3.3}$$

e si moltiplichi numeratore e denominatore per

$$(N-n)! = (N-n) \cdot (N-n-1) \cdot \dots \cdot 2 \cdot 1\,.$$

Si ottiene

$$\begin{aligned}\binom{N}{n} &= \frac{N \cdot (N-1) \cdot \dots \cdot (N-n+1)}{n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1} \cdot \frac{(N-n) \cdot (N-n-1) \cdot \dots \cdot 2 \cdot 1}{(N-n) \cdot (N-n-1) \cdot \dots \cdot 2 \cdot 1} \\ &= \frac{N!}{n!(N-n)!} \; . \end{aligned} \tag{3.4}$$

È immediato verificare che il numero delle combinazioni di N oggetti presi n alla volta è uguale al numero delle combinazioni di N oggetti presi $(N-n)$ alla volta. In formula

$$\binom{N}{n} = \binom{N}{N-n} \; . \tag{3.5}$$

Questo risultato è del tutto plausibile ove si pensi che ad ogni combinazione di n oggetti è possibile associare una combinazione di $(N-n)$ oggetti che si compone di tutti gli $(N-n)$ oggetti non compresi nella prima combinazione.

Esempio 3.3.3 *Si abbia una popolazione con* $N = 6$ *unità statistiche* $(a_1, a_2, a_3, a_4, a_5, a_6)$. *Si elenchino i campioni in blocco di numerosità* $n = 2$ *ed* $n = 4$.

Il numero dei campioni possibili coincide (Tabella 1.19) in quanto ad ogni campione di due unità corrisponde (e viceversa) un campione di quattro unità:

$$\begin{aligned}\binom{6}{2} &= \frac{6!}{2! \cdot 4!} = \frac{720}{2 \cdot 24} = 15 \\ \binom{6}{4} &= \frac{6!}{4! \cdot 2!} = 15 \; .\end{aligned}$$

Campioni in blocco con $n = 2$	Campioni in blocco con $n = 4$
$\{a_1, a_2\}$	$\{a_3, a_4, a_5, a_6\}$
$\{a_1, a_3\}$	$\{a_2, a_4, a_5, a_6\}$
$\{a_1, a_4\}$	$\{a_2, a_3, a_5, a_6\}$
$\{a_1, a_5\}$	$\{a_2, a_3, a_4, a_6\}$
$\{a_1, a_6\}$	$\{a_2, a_3, a_4, a_5\}$
$\{a_2, a_3\}$	$\{a_1, a_4, a_5, a_6\}$
$\{a_2, a_4\}$	$\{a_1, a_3, a_5, a_6\}$
$\{a_2, a_5\}$	$\{a_1, a_2, a_3, a_6\}$
$\{a_2, a_6\}$	$\{a_1, a_3, a_4, a_5\}$
$\{a_3, a_4\}$	$\{a_1, a_2, a_5, a_6\}$
$\{a_3, a_5\}$	$\{a_1, a_2, a_4, a_6\}$
$\{a_3, a_6\}$	$\{a_1, a_2, a_4, a_5\}$
$\{a_4, a_5\}$	$\{a_1, a_2, a_3, a_6\}$
$\{a_4, a_6\}$	$\{a_1, a_2, a_3, a_5\}$
$\{a_5, a_6\}$	$\{a_1, a_2, a_3, a_4\}$

Tabella 1.19: *Elenco dei campioni in blocco di numerosità $n = 2$ ed $n = 4$.*

———— . . ———— . . ————

Conviene fare cenno, ora, a due convenzioni. Il simbolo $\binom{N}{N}$ indica il numero di combinazioni di N oggetti presi N alla volta. Vi è una sola combinazione di questo tipo ed è quella che comprende tutti gli N oggetti. In altre parole $\binom{N}{N} = 1$. Applicando la (3.4) per $n = N$ si ottiene

$$\binom{N}{N} = \frac{N!}{N!(N-N)!} = \frac{1}{0!} .$$

Ma essendo $\binom{N}{N} = 1$ bisogna porre $0! = 1$ per convenzione. Anche se il simbolo $\binom{N}{0}$ è privo di significato ($\binom{N}{0}$ indica le combinazioni di N oggetti presi 0 alla volta), si conviene di porre $\binom{N}{0} = 1$ in modo da preservare la relazione (3.5) $\binom{N}{n} = \binom{N}{N-n}$ anche per $n = N$. In effetti per $n = N$ si ha

$$1 = \binom{N}{N} = \binom{N}{N-N} \to 1 = \binom{N}{N} = \binom{N}{0} .$$

In conclusione, sembra del tutto ragionevole porre: $0! = 1$; $\binom{N}{0} = 1$.

4 *Variabili casuali più comuni*

In questa sezione si presenteranno alcune variabili casuali che trovano vasto impiego.

4.1 Variabile casuale indicatore

La variabile casuale indicatore I assume solo i valori 0 ed 1, rispettivamente con probabilità $(1-p)$ e p, con $0 < p < 1$:

$$I = \begin{array}{cc} \text{Valori} & \text{Probabilità} \\ \left\{ \begin{array}{c} 0 \\ 1 \end{array} \right. & \begin{array}{c} 1-p \\ p \\ \hline 1,00 \end{array} \end{array} . \tag{4.1}$$

È immediato valutare l'aspettativa e la varianza della variabile casuale indicatore.

$$E(I) = 0 \cdot (1-p) + 1 \cdot p = p \,, \tag{4.2}$$

$$\begin{aligned} Var(I) &= (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p = p^2 \cdot (1-p) + (1-p)^2 \cdot p \\ &= p \cdot (1-p) \cdot [p + (1-p)] = p \cdot (1-p) \,. \end{aligned} \tag{4.3}$$

La v.c. indicatore è spesso associata al verificarsi o meno di un evento A in una prova. In particolare: se si verifica A la v.c. I assume valore 1, se non si verifica A (si verifica $\overline{\text{A}}$) la v.c. I assume valore 0. Sia $p = P(\text{A})$ e $(1-p) = P(\overline{\text{A}})$. In conclusione

$$I = \begin{cases} 0 & \text{con probabilità} \quad (1-p) = P(\overline{\text{A}}) \\ 1 & \text{con probabilità} \quad p = P(\text{A}) \,. \end{cases}$$

La variabile casuale indicatore è anche detta variabile "contatore" in quanto tramite essa è possibile rappresentare il numero di volte in cui si verifica l'evento A in n prove ripetute.

Esempio 4.1.1 Si lanci una moneta equilibrata $n = 3$ volte. Sia X la v.c. numero di teste ottenibili nelle tre prove. La v.c. X può assumere i valori $x = 0, 1, 2, 3$. È possibile rappresentare la v.c. X come somma di tre v.c. indicatori: I_1, I_2, I_3. L'indicatore I_1 fa riferimento al primo lancio. Se nel primo lancio si ha testa $I_1 = 1$, se nel primo lancio si ha croce $I_1 = 0$. Dato che la moneta è equilibrata

$P(I_1 = 1) = P(I_1 = 0) = 0,5$. La distribuzione di I_1 risulta

$$I_1 \quad = \quad \begin{cases} 0 & \text{con probabilità} \quad 0,5 \\ 1 & \text{con probabilità} \quad 0,5 \ . \end{cases}$$

L'indicatore I_2 fa riferimento al secondo lancio ed assume i valori 0 ed 1 a seconda che nel secondo lancio si ottenga croce o testa. La distribuzione di I_2 risulta

$$I_2 \quad = \quad \begin{cases} 0 & \text{con probabilità} \quad 0,5 \\ 1 & \text{con probabilità} \quad 0,5 \ . \end{cases}$$

Analogamente I_3 fa riferimento al terzo lancio ed assume valore 0 ed 1 a seconda che nel terzo lancio si ottenga croce o testa. La distribuzione di I_3 risulta

$$I_3 \quad = \quad \begin{cases} 0 & \text{con probabilità} \quad 0,5 \\ 1 & \text{con probabilità} \quad 0,5 \ . \end{cases}$$

È così evidente che il valore che assume X coincide con il numero degli indicatori pari ad 1, di modo che sommando i tre indicatori si ottiene il valore di X. In formula

$$X = I_1 + I_2 + I_3 \ . \tag{4.4}$$

Ad esempio se nei tre lanci si ha la sequenza TCT, allora $X = 2$. I tre indicatori assumono i valori: $I_1 = 1$, $I_2 = 0$ ed $I_3 = 1$. Si constata così che effettivamente $X = I_1 + I_2 + I_3 = 2$. Si osservi infine che i tre indicatori sono v.c. indipendenti aventi la stessa distribuzione. Conseguentemente, dato che i tre indicatori hanno la distribuzione

$$I_j \quad = \quad \begin{cases} 0 & \quad 0,5 \\ 1 & \quad 0,5 \end{cases} \qquad (j = 1, 2, 3) \ ,$$

si ha in forza della (4.2) e (4.3)

$$\begin{cases} E(I_j) = 0,5 \\ Var(I_j) = 0,5 \cdot 0,5 = 0,25 \end{cases} \qquad (j = 1, 2, 3) \ .$$

Infine dalla (2.8) si ottiene

$$\begin{aligned} E(X) \quad &= \quad E(I_1 + I_2 + I_3) = E(I_1) + E(I_2) + E(I_3) \\ &= \quad 0,5 + 0,5 + 0,5 = 3 \cdot 0,5 = 1,5 \ . \end{aligned}$$

Inoltre, tenuto conto che i tre indicatori sono v.c. indipendenti, si può impiegare la (2.22) ricavando così

$$\begin{aligned} Var(X) &= Var(I_1 + I_2 + I_3) = Var(I_1) + Var(I_2) + Var(I_3) \\ &= 0,25 + 0,25 + 0,25 = 3 \cdot 0,25 = 0,75 \ . \end{aligned}$$

Si osservi che si è riusciti – con l'ausilio degli indicatori – a valutare l'aspettativa e la varianza di X senza conoscere la legge di probabilità di X.

Si osservi, infine, che la procedura impiegata in questo esempio si può generalizzare agevolmente. Sia A un evento che ha probabilità $p = P(\mathrm{A})$ di verificarsi in una prova. Si supponga di effettuare n prove in condizioni di indipendenza. Sia X il numero di volte in cui si presenta A nelle n prove. Sia I_j l'indicatore associato alla j-ma prova ($j = 1, 2, \ldots, n$) che assume valore 1 o valore 0 se nella j-ma prova si ha rispettivamente l'evento A o l'evento "non A" ($\overline{\mathrm{A}}$). Ovviamente $X = I_1 + \cdots + I_j \cdots + I_n$. Si ricava così che

$$E(X) = n \cdot p; \quad Var(X) = n \cdot p \cdot q \ ,$$

dove $q = 1 - p$.

4.2 Variabile casuale binomiale

Prima di studiare la v.c. binomiale conviene presentare lo sviluppo del binomio di Newton. Come è noto:

i) $$(a+b)^2 = a^2 + 2ab + b^2 = \binom{2}{0} a^0 b^{2-0} + \binom{2}{1} a^1 b^{2-1} + \binom{2}{2} a^2 b^{2-2} \ .$$

ii) $$\begin{aligned} (a+b)^3 &= a^3 + 3ab^2 + 3a^2b + b^3 \\ &= \binom{3}{0} a^0 b^{3-0} + \binom{3}{1} a^1 b^{3-1} + \binom{3}{2} a^2 b^{3-2} + \binom{3}{3} a^3 b^{3-3} \ . \end{aligned}$$

In generale vale la seguente regola che si dà senza dimostrazione

$$(a+b)^n = \binom{n}{0} a^0 b^{n-0} + \binom{n}{1} a^1 b^{n-1} + \cdots + \binom{n}{x} a^x b^{n-x} + \cdots + \binom{n}{n} a^n b^{n-n} \ , \tag{4.5}$$

essendo n un numero intero positivo. Il simbolo $\binom{n}{x}$ presente nella (4.5) indica le combinazioni di n elementi presi x alla volta. Proprio perché $\binom{n}{x}$ è presente nello sviluppo del binomio di Newton lo stesso è denominato anche "coefficiente binomiale".

È possibile ora definire la v.c. binomiale di parametri n e p, essendo n un numero intero e $0 < p < 1$.

Definizione (Variabile casuale binomiale) La variabile casuale binomiale X è una v.c. che assume i valori $x = 0, 1, \ldots, n$ con probabilità

$$P(X = x) = p(x) = \binom{n}{x} p^x (1-p)^{n-x} \ , \quad x = 0, 1, \ldots, n \ . \tag{4.6}$$

Per poter accettare la (4.6) come legge di probabilità di una variabile casuale bisogna dimostrare che:

i) $p(x) = \binom{n}{x} p^x (1-p)^{n-x} > 0$ per ogni $x = 0, 1, \ldots, n$;

ii) $\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = 1$.

La verifica della prima condizione è immediata. Per quanto concerne la seconda condizione si consideri lo sviluppo del binomio

$$\begin{aligned}[p + (1-p)]^n \quad &= \quad \binom{n}{0} p^0 (1-p)^{n-0} + \cdots + \binom{n}{x} p^x (1-p)^{n-x} + \\ &\quad + \cdots + \binom{n}{n} p^n (1-p)^{n-n} \ . \end{aligned} \tag{4.7}$$

Il generico termine della sommatoria $\binom{n}{x} p^x (1-p)^{n-x}$ coincide con la (4.6). La (4.7) è chiaramente uguale ad 1 in quanto $[p+(1-p)]^n = 1^n = 1$. In conclusione la (4.6) è la legge di probabilità di una variabile casuale, che per ovvi motivi è denominata variabile casuale binomiale.

———— . . ———— . . ————

Chiarito che la formula

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \tag{4.6}$$

fornisce la legge di probabilità di una v.c. si cercherà ora di precisare in quali contesti è applicabile la variabile casuale binomiale. Il contesto è quello delle prove ripetute n volte “sotto le stesse” condizioni in modo che:

i) la probabilità $p = P(\mathrm{A})$ di un evento A resti costante nelle n prove;

ii) le n prove siano indipendenti.

Si dimostrerà allora che la v.c. X, che indica il numero delle volte che A si presenta nelle n prove, ha legge di probabilità fornita proprio da

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \ , \quad x = 0, 1, \ldots, n \ . \tag{4.6}$$

Dimostrazione. Per individuare i possibili risultati elementari conseguibili nelle n prove basta elencare tutte le sequenze di A e $\overline{\text{A}}$ che possono verificarsi nelle n prove. Il loro numero è pari a 2^n (si veda l'Esempio 2.0.5 della Sezione 2: cinque lanci di una moneta equilibrata). Si consideri una successione con $X = x$. Significa che vi sono x prove con A ed $(n-x)$ prove con $\overline{\text{A}}$. Per chiarire il concetto si supponga $n = 5$ e $x = 2$. Una possibile successione di questo tipo è: $\overline{\text{A}}\ \overline{\text{A}}$ A $\overline{\text{A}}$ A. Deriva, per l'indipendenza fra le prove, che

$$\begin{aligned} P\left(\overline{\text{A}}\,\overline{\text{A}}\,\text{A}\,\overline{\text{A}}\,\text{A}\right) &= P\left(\overline{\text{A}}\right) \cdot P\left(\overline{\text{A}}\right) \cdot P(\text{A}) \cdot P\left(\overline{\text{A}}\right) \cdot P(\text{A}) \\ &= \left\{P\left(\overline{\text{A}}\right)\right\}^3 \cdot \{P(\text{A})\}^2 = p^2 \cdot (1-p)^{5-2} \ . \end{aligned}$$

Si osservi che $p^2 \cdot (1-p)^{5-2}$ è il valore della probabilità di ogni successione con $x = 2$. Per calcolare la probabilità che $X = 2$ bisogna allora sommare la probabilità di tutte le successioni con due A e tre $\overline{\text{A}}$. In formula

$$P(X = 2) = \left(\begin{array}{c} \text{Numero delle} \\ \text{successioni} \\ \text{con due A e tre } \overline{\text{A}} \end{array} \right) \cdot p^2 \cdot (1-p)^{5-2} \ .$$

Il numero delle successioni con $x = 2$ è dato dal numero delle coppie di prove con A ottenibili da cinque prove. Le coppie di prove con A sono: $(1^a, 2^a)$; $(1^a, 3^a)$; $(1^a, 4^a)$; $(1^a, 5^a)$; $(2^a, 3^a)$; $(2^a, 4^a)$; $(2^a, 5^a)$; $(3^a, 4^a)$; $(3^a, 5^a)$; $(4^a, 5^a)$. Si hanno 10 coppie che non sono altro che le combinazioni $\binom{5}{2} = 10$ di cinque elementi (le prove) presi due alla volta (le due prove con A). Pertanto

$$P(X = 2) = \binom{5}{2} \cdot p^2 \cdot (1-p)^{5-2} \ .$$

Il ragionamento si può ovviamente ripetere per un qualsiasi valore di n e di x. In definitiva la probabilità che X assuma valore x nelle n prove è pari a

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \ , \quad x = 0, 1, \ldots, n \ .$$

Esempio 4.2.1 *Sia $p = 0,6$ la probabilità che un farmaco sia efficace. Il farmaco viene somministrato a $n = 5$ pazienti omogenei. Sia X il numero di pazienti che guariscono. Ricavare la distribuzione di probabilità della v.c. X.*

Per poter rispondere al quesito bisogna controllare se è ragionevole ritenere che $p = 0,6$ per ogni paziente e se è ragionevole che vi sia indipendenza nel processo di guarigione fra i cinque pazienti.

i) Si può ritenere che p sia costante perché i pazienti sono omogenei.

ii) Si può supporre che i pazienti siano fisicamente separati in modo da evitare vicendevoli influenze.

Conseguentemente per calcolare le probabilità della v.c. X si può applicare la formula

$$p(x) = \binom{5}{x} \cdot 0,6^x \cdot 0,4^{5-x} \ , \quad x = 0,1,2,3,4,5,6 \ .$$

In particolare:

$$\begin{aligned}
p(0) &= \binom{5}{0} \cdot 0,6^0 \cdot 0,4^5 = 1 \cdot 1 \cdot 0,01024 &&= 0,01024 \ ; \\
p(1) &= \binom{5}{1} \cdot 0,6^1 \cdot 0,4^4 = 5 \cdot 0,6 \cdot 0,0256 &&= 0,0768 \ ; \\
p(2) &= \binom{5}{2} \cdot 0,6^2 \cdot 0,4^3 = 10 \cdot 0,36 \cdot 0,064 &&= 0,2304 \ ; \\
p(3) &= \binom{5}{3} \cdot 0,6^3 \cdot 0,4^2 = 10 \cdot 0,216 \cdot 0,16 &&= 0,3456 \ ; \\
p(4) &= \binom{5}{4} \cdot 0,6^4 \cdot 0,4^1 = 5 \cdot 0,1296 \cdot 0,4 &&= 0,2592 \ ; \\
p(5) &= \binom{5}{5} \cdot 0,6^5 \cdot 0,4^0 = 1 \cdot 0,07776 \cdot 1 &&= 0,07776. \\
& && \overline{1,00000}
\end{aligned}$$

Esempio 4.2.2 *Si lanci una moneta ben equilibrata $n = 5$ volte. Si ricavi la legge di probabilità di X che conta il numero di teste ottenute negli $n = 5$ lanci.*

La probabilità di ottenere T in una prova è $p = 0,5$. I lanci sono fra loro indipendenti per cui X è una v.c. binomiale con $n = 5$ e $p = 0,5$. Pertanto

$$\begin{aligned}
p(x) &= \binom{5}{x} \cdot 0,5^x \cdot 0,5^{5-x} \\
&= \binom{5}{x} \cdot 0,5^5 = \binom{5}{x} \cdot 0,03125 \ .
\end{aligned}$$

Si ha così:

$$\begin{aligned}
&p(0) = 1 \cdot 0,03125 = 0,03125 \ ; \quad &&p(1) = 5 \cdot 0,03125 = 0,15625 \ ; \\
&p(2) = 10 \cdot 0,03125 = 0,3125 \ ; \quad &&p(3) = 10 \cdot 0,03125 = 0,3125 \ ; \\
&p(4) = 5 \cdot 0,03125 = 0,15625 \ ; \quad &&p(5) = 1 \cdot 0,03125 = 0,03125 \ .
\end{aligned}$$

Si ricaveranno ora l'aspettativa e la varianza della v.c. binomiale.

L'aspettativa della v.c. binomiale è fornita da

$$E(X) = \sum_{x=0}^{n} x \cdot p(x) = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x} \; .$$

Lo svolgimento di questo calcolo presenta qualche difficoltà. È possibile, ricorrendo alla tecnica degli indicatori, ricavare agevolmente il valore di $E(X)$ e di $Var(X)$ di una binomiale.

Sia I_j la v.c. indicatore che assume valore 1 se nella j-ma prova si verifica l'evento A e valore 0 se nella j-ma prova si verifica l'evento $\overline{\text{A}}$. La v.c. X si può rappresentare come somma degli n indicatori I_j ($j = 1, 2, \ldots, n$)

$$X = I_1 + \cdots + I_j + \cdots + I_n \; .$$

Gli indicatori hanno stessa distribuzione

$$I_j \quad = \quad \begin{cases} 1 & \text{con probabilità} \quad p = P(\text{A}) \\ 0 & \text{con probabilità} \quad (1-p) = P(\overline{\text{A}}) \; . \end{cases}$$

Si sa che: $E\left(I_j\right) = p$; $Var\left(I_j\right) = p \cdot q$, con $q = 1 - p$. Pertanto

$$E(X) = E\left(I_1\right) + \cdots + E\left(I_j\right) + \cdots + E\left(I_n\right) = np \tag{4.8}$$

Dato che gli indicatori sono indipendenti si ha

$$Var(X) = Var\left(I_1\right) + \cdots + Var\left(I_j\right) + \cdots + Var\left(I_n\right) = npq \tag{4.9}$$

Si fa notare che la (4.8) e la (4.9) erano state già ricavate nel Paragrafo 4.1 (variabile casuale indicatore).

Esempio 4.2.3 *Ricavare $E(X)$ e $Var(X)$ nel caso degli $n = 5$ pazienti a cui è stato somministrato un farmaco. X è la v.c. numero di pazienti che guariscono.*

Essendo $p = 0,6$ si ha

$$E(X) = n \cdot p = 5 \cdot 0,6 = 3 \; .$$

Ciò significa che degli $n = 5$ pazienti ne guariscono in media 3.

$$Var(X) \quad = \quad n \cdot p \cdot (1-p) = 5 \cdot 0,6 \cdot 0,4 = 1,2 \; .$$

$\sigma(X) = \sqrt{1,2} = 1,0954$. Questo valore informa che il numero dei guariti differisce da 3 per un valore che mediamente è pari a $1,0954$.

4.3 Variabile casuale ipergeometrica

Prima di studiare la v.c. ipergeometrica conviene presentare (senza dimostrazione) una importante relazione fra coefficienti binomiali.

$$\binom{r}{0}\cdot\binom{N-r}{n-0}+\binom{r}{1}\cdot\binom{N-r}{n-1}+\cdots+\binom{r}{x}\cdot\binom{N-r}{n-x}+\cdots+\binom{r}{n}\cdot\binom{N-r}{n-n}=\binom{N}{n}\ . \tag{4.10}$$

Esempio 4.3.1 Sia $N = 9$; $r = 4$; $n = 3$. Verificare numericamente la (4.10).

$$\binom{N}{n}=\frac{9!}{3!\cdot 6!}=\frac{9\cdot 8\cdot 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1\cdot}{3\cdot 2\cdot 1\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}=84\ .$$

Nell'esempio che si sta considerando

$$\binom{r}{x}\cdot\binom{N-r}{n-x}=\binom{4}{x}\cdot\binom{5}{3-x}\ .$$

Pertanto

$$\begin{aligned}\binom{4}{0}\binom{5}{3}+\binom{4}{1}\binom{5}{2}+\binom{4}{2}\binom{5}{1}+\binom{4}{3}\binom{5}{0} &= 1\cdot 10+4\cdot 10+6\cdot 5+4\cdot 1\\ &= 84\end{aligned}$$

——— . . ——— . . ———

Definizione (Variabile casuale ipergeometrica) La v.c. ipergeometrica X di parametri $n \leq N$; $r < N$ (n, r, ed N sono interi positivi) assume i valori $x = 0, 1, 2, \ldots, n$ con probabilità

$$p(x)=\frac{\binom{r}{x}\cdot\binom{N-r}{n-x}}{\binom{N}{n}}\ ,\quad x=0,1,2,\ldots,n\ . \tag{4.11}$$

Per poter accettare la (4.11) come legge di probabilità di una variabile casuale bisogna dimostrare che

i) $\dfrac{\binom{r}{x}\cdot\binom{N-r}{n-x}}{\binom{N}{n}} > 0$ per ogni $x = 0, 1, \ldots, n$;

ii) $\displaystyle\sum_{x=0}^{n} \frac{\binom{r}{x}\cdot\binom{N-r}{n-x}}{\binom{N}{n}} = 1.$

La prima condizione è immediatamente verificata perché: $\binom{N}{n} > 0$, $\binom{r}{x} > 0$ per ogni $x = 0, 1, 2, \ldots, n$. Inoltre $\binom{N-r}{n-x} \geq 0$. Il coefficiente binomiale $\binom{a}{b} = 0$ se $b > a$. La seconda condizione è immediata conseguenza della (4.10). Infatti

$$\begin{aligned}
\sum_{x=0}^{n} \frac{\binom{r}{x}\cdot\binom{N-r}{n-x}}{\binom{N}{n}} &= \frac{1}{\binom{N}{n}} \cdot \sum_{x=0}^{n} \binom{r}{x}\cdot\binom{N-r}{n-x} \\
&= \frac{1}{\binom{N}{n}} \cdot \binom{N}{n} = 1\ .
\end{aligned}$$

——— . . ——— . . ———

Chiarito che la formula (4.11) fornisce la legge di probabilità di una variabile casuale si precisa ora che il contesto nel quale è impiegabile la v.c. ipergeometrica è quello delle estrazioni in blocco o delle estrazioni (una alla volta) senza riposizioni nel caso in cui l'ordine di estrazione non sia rilevante.

——— . . ——— . . ———

(Estrazioni in blocco)

Si ha una popolazione di N unità (palline per comodità espositiva) di cui r rosse ed $(N - r)$ bianche. Si estraggono in blocco n palline. Per semplicità si suppone $n \leq r$. Con X si indica la v.c. numero di palline rosse presenti nel campione. I valori che può assumere X sono: $x = 0, 1, 2, \ldots, n$.

Il numero di campioni possibili nel caso del campionamento in blocco è dato dalle combinazioni delle N palline prese n alla volta:

$$\begin{array}{c}\text{Numero dei campioni possibili}\\ \text{con estrazioni in blocco}\end{array} = \binom{N}{n}\ .$$

Supponendo che i campioni siano "alla pari", per quanto concerne la possibilità di estrazione si può attribuire a ciascuno di essi la stessa probabilità $1/\binom{N}{n}$. In altre parole si può impiegare un modello probabilistico uniforme per il campionamento in blocco. Consegue che

$$P[X=x] = \frac{\text{N° Campioni favorevoli}}{\text{N° Campioni possibili}} = \frac{\begin{array}{c}\text{Numero dei}\\ \text{campioni con } x\\ \text{palline rosse}\end{array}}{\binom{N}{n}}$$

Ecco come conteggiare il numero dei campioni con x palline rosse. In ogni campione "favorevole" devono esservi x palline rosse ed $(n-x)$ palline bianche. Le x palline rosse devono provenire dalle r palline rosse della popolazione. Si possono avere $\binom{r}{x}$ gruppi di x palline rosse. Le $(n-x)$ palline bianche del campione provengono dalle $(N-r)$ palline bianche della popolazione. Si possono così avere $\binom{N-r}{n-x}$ gruppi di $(n-r)$ palline bianche.

Associando ogni gruppo di x palline rosse con ogni gruppo di $(n-x)$ palline bianche si ottengono $\binom{r}{x} \cdot \binom{N-r}{n-x}$ campioni con n palline di cui x sono rosse. In definitiva il numero dei campioni favorevoli è $\binom{r}{x} \cdot \binom{N-r}{n-x}$. Conseguentemente

$$P[X=x] = p(x) = \frac{\binom{r}{x} \cdot \binom{N-r}{n-x}}{\binom{N}{n}} \tag{4.11}$$

Esempio 4.3.2 *Si ha una popolazione con $N=11$ palline di cui $r=6$ rosse e 5 bianche. Si estraggono in blocco $n=5$ palline. Elencare il numero dei campioni con $x=2$ palline rosse.*

Il numero dei campioni con $x=2$ palline rosse è dato da

$$\binom{6}{2}\binom{5}{3} = \frac{6!}{2!\cdot 4!} \cdot \frac{5!}{3!\cdot 2!} = \frac{720}{2\cdot 24} \cdot \frac{120}{6\cdot 2} = 15\cdot 10 = 150\ .$$

Si indichino con $(r_1, r_2, r_3, r_4, r_5, r_6)$ le sei palline rosse e con $(b_1, b_2, b_3, b_4, b_5)$ le cinque palline bianche. In Tabella 1.20 si riportano nella colonna madre i 15 gruppi di due palline rosse e nella riga madre i 10 gruppi con 3 palline bianche.

	$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$	$\begin{pmatrix} b_1 \\ b_2 \\ b_4 \end{pmatrix}$	$\begin{pmatrix} b_1 \\ b_2 \\ b_5 \end{pmatrix}$	$\begin{pmatrix} b_1 \\ b_3 \\ b_4 \end{pmatrix}$	$\begin{pmatrix} b_1 \\ b_3 \\ b_5 \end{pmatrix}$	$\begin{pmatrix} b_1 \\ b_4 \\ b_5 \end{pmatrix}$	$\begin{pmatrix} b_2 \\ b_3 \\ b_4 \end{pmatrix}$	$\begin{pmatrix} b_2 \\ b_3 \\ b_5 \end{pmatrix}$	$\begin{pmatrix} b_2 \\ b_4 \\ b_5 \end{pmatrix}$	$\begin{pmatrix} b_3 \\ b_4 \\ b_5 \end{pmatrix}$
(r_1, r_2)										
(r_1, r_3)										
(r_1, r_4)										
(r_1, r_5)										
(r_1, r_6)										
(r_2, r_3)										
(r_2, r_4)										
(r_2, r_5)										
(r_2, r_6)										
(r_3, r_4)										
(r_3, r_5)										
(r_3, r_6)										
(r_4, r_5)										
(r_4, r_6)										
(r_5, r_6)										

Tabella 1.20: *Gruppi di due palline rosse (colonna madre) e gruppi di tre palline bianche (riga madre).*

Ciascuna casella rappresenta un campione con due palline rosse indicate nella colonna madre e tre palline bianche indicate nella riga madre. Il numero delle caselle è $\binom{6}{2} \cdot \binom{5}{3} = 15 \cdot 10 = 150$.

Esempio 4.3.3 *Relativamente al campionamento dell'Esempio 4.3.2 calcolare la probabilità di ottenere $x = 2$ palline rosse.*

Si può applicare la (4.11). Si ha così

$$p(2) = \frac{\binom{6}{2} \cdot \binom{5}{3}}{\binom{11}{5}} = \frac{150}{\frac{11!}{5!6!}} = \frac{150}{\frac{11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}} = \frac{150}{462} = 0,324675\ .$$

Esempio 4.3.4 *Un lotto di dieci articoli ne contiene due difettosi. Si supponga di estrarre in blocco $n = 4$. Ricavare la probabilità che il numero dei pezzi difettosi X presenti nel campione sia: uguale a 0; uguale a 1; uguale a 2; uguale a 3.*

Il numero X di pezzi difettosi presenti nel campione segue la legge ipergeometrica con $N = 10$, $r = 2$ ed $n = 4$ (si noti che in questo esempio $n > r$). Per ricavare le probabilità richieste si proverà ad impiegare la (4.11). Pertanto

$$\begin{aligned} P[X=0] &= \frac{\binom{2}{0}\cdot\binom{8}{4}}{\binom{10}{4}} = \frac{1\cdot 70}{210} = 0,33\overline{3}\ ; \\ P[X=1] &= \frac{\binom{2}{1}\cdot\binom{8}{3}}{\binom{10}{4}} = \frac{2\cdot 56}{210} = 0,53\overline{3}\ ; \\ P[X=2] &= \frac{\binom{2}{2}\cdot\binom{8}{2}}{\binom{10}{4}} = \frac{1\cdot 28}{210} = 0,13\overline{3}\ . \end{aligned}$$

$P[X = 3] = 0$ in quanto X non può essere superiore ad r; ovvero, se vi sono $r = 2$ pezzi difettosi nel lotto non è possibile, con il tipo di campionamento considerato, trovarne $x = 3$ nel campione. L'evento $(X = 3)$ è un evento impossibile. Si provi, ad ogni modo, ad impiegare la (4.11) anche per $x = 3$. Si ha

$$P[X=3] = \frac{\binom{2}{3}\cdot\binom{8}{1}}{\binom{10}{4}} = \frac{0\cdot 8}{210} = 0\ .$$

In precedenza si era precisato che se $b < a$ allora $\binom{b}{a} = 0$. Il simbolo $\binom{2}{3}$ indicando il numero delle combinazioni di due oggetti presi tre alla volta non può che essere uguale a zero.

———— . . ———— . . ————

L'Esempio 4.3.4 ha mostrato che i valori che può assumere X nel caso della v.c. ipergeometrica non necessariamente sono tutti i valori $x = 0, 1, 2, \ldots, n$. In effetti, nel definire una v.c. ipergeometrica, si pongono solo le condizioni: $r < N$ ed $n \leq N$. Si dimostra allora che i valori che X può assumere sono i seguenti

$$\max(0; n + r - N) \leq x \leq \min(r; n)\ .$$

Si calcolino i limiti di x per

i) $N = 8$, $n = 4$, $r = 5$;

ii) $N = 10$, $r = 4$, $n = 5$;

Nel caso **i)** si ha: $\max(0; 9 - 8) \leq x \leq \min(5; 4) \rightarrow (1 \leq x \leq 4)$.
Nel caso **ii)** si ha: $\max(0; 9 - 10) \leq x \leq \min(4; 5) \rightarrow (0 \leq x \leq 4)$.

——— . . ——— . . ———

(Estrazioni senza riposizione)

Come si è già accennato in precedenza, alla distribuzione ipergeometrica è possibile pervenire anche per altra via. Si abbia una popolazione costituita da N palline di cui r rosse e $N - r$ bianche. Si estraggono una dietro l'altra n palline senza riposizione. Si indichi con X il numero di palline rosse presenti nel campione.

Come è stato ben evidenziato in 3.1 il numero dei campioni possibili, nel caso delle estrazioni senza riposizione, è fornito dalle disposizioni $(N)_n$ di N oggetti presi n alla volta.

Si ricordi che nel caso delle disposizioni due campioni possono differire per la composizione o anche solo per l'ordine di presentazione. I campioni in blocco differiscono per la composizione e sono pari alle combinazioni $\binom{N}{n}$ di N oggetti presi n alla volta. Da ogni campione in blocco si possono formare $n!$ campioni che differiscono per l'ordine di presentazione. Consegue che i campioni possibili nel caso delle estrazioni senza riposizione sono $\binom{N}{n} n! = (N)_n$. Sempre nel caso del campionamento in blocco il numero dei campioni con x palline rosse è fornito da $\binom{r}{x}\binom{N-r}{n-x}$.

Consegue che, nel caso del campionamento senza riposizione il numero dei campioni con x palline rosse è fornito da $\binom{r}{x}\binom{N-r}{n-x} n!$.

In conclusione, nel caso del campionamento senza riposizione, la probabilità di ottenere x palline rosse è data da

$$P[X = x] = \frac{\text{Numero di campioni favorevoli}}{\text{Numero di campioni possibili}} = \frac{\binom{r}{x}\binom{N-r}{n-x} n!}{\binom{N}{n} n!}$$

$$= \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}.$$

Si riottiene la stessa formula (4.11) ottenuta con il campionamento in blocco.

———— . . ———— . . ————

Si può dimostrare che l'aspettativa e la varianza della v.c. ipergeometrica di parametri n, r e N sono pari a

$$E(X) = \sum_{x=0}^{n} x \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = n\frac{r}{N}; \tag{4.12}$$

$$Var(x) = \sum_{x=0}^{n} \left(x - n\frac{r}{N}\right)^2 \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = n\frac{r}{N}\left(1-\frac{r}{N}\right)\frac{N-n}{N-1}. \tag{4.13}$$

Esempio 4.3.5 *(Confronto fra l'aspettativa e la varianza delle v.c. binomiale ed ipergeometrica)*

Una popolazione è costituita da N palline di cui r rosse e $N-r$ bianche.

a) *Si estraggono con riposizione n palline.*

Sia X il numero di palline rosse del campione. Si ha uno schema di prove ripetute in cui la probabilità dell'evento A (pallina rossa in una estrazione) è costantemente uguale a $\frac{r}{N} = p$ (questo valore indica anche la frequenza relativa delle palline rosse nella popolazione). Inoltre, data la riposizione, le prove sono fra loro indipendenti. Conseguentemente, la legge di probabilità di X è

$$p(x) = \binom{n}{x} p^x(1-p)^{n-x}. \tag{4.14}$$

In precedenza è stato dimostrato che per la v.c. binomiale

$$E(X) = np = n\frac{r}{N};$$

$$Var(X) = np(1-p) = n\frac{r}{N}\left(1-\frac{r}{N}\right).$$

b) *Si estraggono senza riposizione n palline.*

Sia X il numero di palline rosse del campione. Ora la legge di probabilità di X è quella di una v.c. ipergeometrica di parametri n, r, N. Pertanto

$$p(x) = \frac{\begin{pmatrix} r \\ x \end{pmatrix}\begin{pmatrix} N-r \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}. \tag{4.15}$$

Inoltre

$$E(X) = n\frac{r}{N} \tag{4.16}$$

$$Var(X) = n\frac{r}{N}\left(1 - \frac{r}{N}\right)\frac{N-n}{N-1}. \tag{4.17}$$

In conclusione i due schemi di estrazioni portano a due v.c. aventi medie uguali e varianze differenti.

La varianza di X che si ottiene nel cado di estrazioni senza riposizione è minore di quella della v.c. che si ha in caso di estrazioni con riposizione in quanto $\frac{N-n}{N-1} < 1$ (tranne il caso banale $n = 1$).

Il rapporto $\frac{N-n}{N-1}$ è detto fattore di correzione (rispetto alla binomiale).

Esempio 4.3.6 *Si abbia un lotto con $N = 12$ pezzi di cui $r = 4$ difettosi. Si estraggono $n = 3$ pezzi. Sia X il numero di pezzi difettosi del campione. Calcolare la distribuzione di probabilità di X e+ l'aspettativa e la varianza di X sia nel caso di estrazioni con riposizione che nel caso di estrazioni senza riposizione. (Estrazioni senza riposizione)*

$$p(x) = \frac{\begin{pmatrix} 4 \\ x \end{pmatrix}\begin{pmatrix} 8 \\ 3-x \end{pmatrix}}{\begin{pmatrix} 12 \\ 3 \end{pmatrix}} = \frac{\begin{pmatrix} 4 \\ x \end{pmatrix}\begin{pmatrix} 8 \\ 3-x \end{pmatrix}}{220} \qquad (x = 0, 1, 2, 3)$$

x	$\begin{pmatrix} 4 \\ x \end{pmatrix}$	$\begin{pmatrix} 8 \\ 3-x \end{pmatrix}$	$\begin{pmatrix} 4 \\ x \end{pmatrix}\begin{pmatrix} 8 \\ 3-x \end{pmatrix}$	p(x)
0	1	56	56	$0{,}2545\overline{45}$
1	4	28	112	$0{,}5090\overline{90}$
2	6	8	48	$0{,}2181\overline{81}$
3	4	1	4	$0{,}018\overline{18}$
			220	1,000

$$E(X) = n\frac{r}{N} = 3\frac{4}{12} = 1$$
$$Var(X) = n\frac{r}{N}\left(1 - \frac{r}{N}\right)\frac{12-3}{12-1} = 0.545454.$$

(Estrazioni con riposizione)

$$p(x) = \left(\begin{array}{c} 3 \\ x \end{array}\right)\left(\frac{4}{12}\right)^x\left(1 - \frac{4}{12}\right)^{3-x} \qquad (x = 0, 1, 2, 3)$$

x	$\binom{3}{x}$	$(0,333)^x$	$(0,666)^{3-x}$	p(x)
0	1	1,000	0,296296	0,296296
1	3	0,333$\bar{3}$	0,44444	0,44444
2	3	0,11111	0,6666	0,22222
3	1	0,03703	1,000	0,03703
				1,000

$$E(X) = n\frac{r}{N} = 3\frac{4}{12} = 1$$
$$Var(X) = n\frac{r}{N}\left(1 - \frac{r}{N}\right) = 0.666\bar{6}.$$

Si osservi che i valori estremi della v.c. X (x=0 e x=3) hanno più probabilità nella distribuzione binomiale che nella distribuzione ipergeometrica.

Si osservi, infine, che il fattore di correzione $\frac{N-n}{N-1}$ era già stato impiegato nella introduzione dove si è confrontata la distribuzione delle medie campionarie nel caso delle estrazioni con riposizione con la distribuzione delle medie campionarie nel caso delle estrazioni senza riposizione.

4.4 Variabile casuale normale

Nella trattazione sinora fatta sulle variabili casuali si è supposto che le stesse potessero assumere solo un numero limitato di valori. Esistono tuttavia molti esperimenti casuali che danno luogo a variabili che possono assumere, ipoteticamente, tutti gli infiniti valori compresi in un intervallo. Si pensi alla durata delle gomme per le automobili, alla durata delle lampade, al diametro delle sferette prodotte in serie da una macchina, ecc..

Si può allora convenire di rappresentare i risultati di alcuni esperimenti casuali con una variabile continua X. Si ipotizzi di aver replicato $n = 70$ volte uno di tali esperimenti. I 70 valori della variabile X sono compendiati nella distribuzione di frequenze che segue, ove, per comodità, sono riportate le frequenze relative e le frequenze relative specifiche.

X	frequenze	frequenze relative	Ampiezza classi	Frequenze relative specifiche
2-4	3	0,043	2	0,0215
4-6	7	0,100	2	0,0500
6-8	15	0,214	2	0,1070
8-10	20	0,286	2	0,1430
10-12	14	0,200	2	0,1000
12-14	6	0,086	2	0,0430
14-16	5	0,071	2	0,0360
Tot	70	1,000		

Si ricordi che la frequenza relativa specifica (frs) di una classe è fornita dal rapporto fra la frequenza relativa e l'ampiezza della classe:

$$frs = \frac{\text{frequenza relativa}}{\text{ampiezza}}. \tag{4.18}$$

La rappresentazione grafica della distribuzione delle frequenze relative si effettua riportando in ascissa le classi ed in ordinata le frequenze relative specifiche delle classi stesse. Si ottengono così tanti rettangoli quante sono le classi. L'area di ogni rettangolo è uguale alla frequenza relativa della classe.

In effetti, dalla (4.18) si ricava:

$$\text{frequenza relativa} = \text{ampiezza} \times frs;$$

quest'ultimo prodotto coincide proprio con l'area del rettangolo.

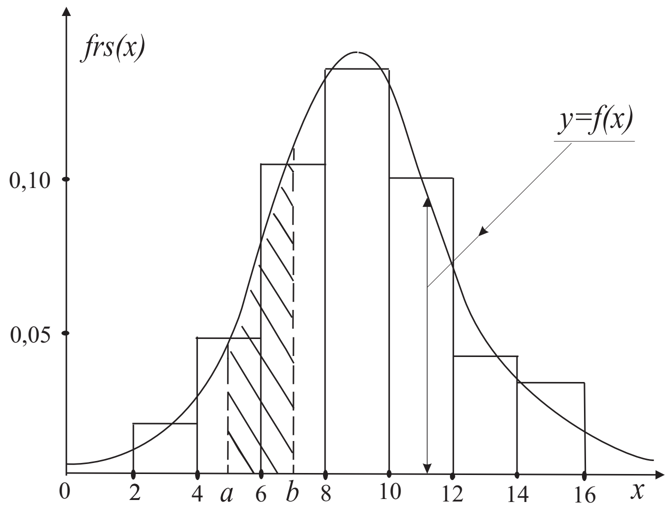


Figura 1.3: *Rappresentazione grafica delle frequenze relative.*

Nel passare da un rettangolino a quello successivo la $frs(x)$ sperimenta variazioni (salti) di notevole entità. Ciò dipende dal fatto che con sole 70 replicazioni dell'esperimento si sono dovute fare poche classi con ampiezza pari a 2. Se si facessero $n = 500$ replicazioni dell'esperimento allora si potrebbero fare più raggruppamenti con classi di minor ampiezza. Si noterebbero, allora, variazioni più limitate nel passare da un rettangolino a quello successivo.

Per descrivere l'andamento delle frequenze relative specifiche si sostituisce l'andamento reale (a scalini) con una curva continua $y = f(x)$ denominata curva delle frequenze relative specifiche.

La curva $y = f(x)$, che rappresenta le frequenze relative specifiche deve essere tale che:

i) $f(x) \geq 0$ per quei valori x in cui si manifesta la variabile X;

ii) l'area sottesa alla curva deve essere uguale alla somma delle frequenze relative, cioè deve essere uguale a 1.

Disponendo di una curva delle frequenze relative specifiche è possibile ottenere la frequenza relativa che fa riferimento all'intervallo (a, b) valutando l'area della curva compresa fra a e b (area tratteggiata della curva precedente).

Le curve più frequentemente impiegate per descrivere le distribuzioni dei caratteri continui sono la curva di Pareto, la curva log-normale e la curva normale. Le equazioni delle ordinate di queste curve si trovano in molti manuali di statistica [Zenga (1998), Zenga (2007), Frosini, Magagnoli].

Bisogna rappresentare ora la variabile casuale continua nel contesto probabilistico delle variabili casuali. Si hanno così le variabili casuali continue. Queste variabili casuali sono caratterizzate da una funzione $y = f(x)$, detta funzione di *densità* di probabilità che rappresenta nel modello probabilistico le frequenze relative specifiche.

La funzione di densità di probabilità deve soddisfare le stesse proprietà della curva delle frequenze relative specifiche. In altre parole:

i) $f(x) \geq 0$ per ogni $a < x < b$, a e b sono gli estremi entro i quali si manifesta la variabile casuale;

ii) $\displaystyle\int_a^b f(x)dx = 1$ (coincide con l'area sottesa alla curva).

In particolare la probabilità cumulata della variabile casuale X è fornita dall'integrale

$$P[X \leq x] = \int_a^x f(t)dt \qquad (a < x < b)$$

e coincide con l'area sottesa a $f(t)$ (funzione di densità di probabilità) per t che va da a a x.

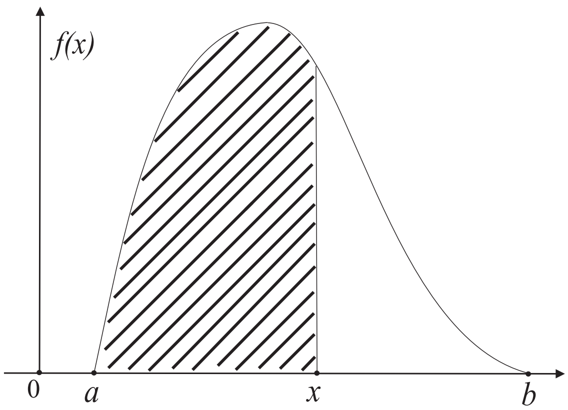


Figura 1.4: *Rappresentazione grafica della funzione di densità di probabilità e della probabilità cumulata.*

———— . . ———— . . ————

La variabile casuale più impiegata nell'inferenza statistica è la variabile casuale normale.

La funzione di densità di probabilità della v.c. normale è fornita da

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \qquad , -\infty < x < +\infty, \qquad (4.19)$$

essendo i parametri μ e σ^2 rispettivamente l'aspettativa e la varianza della v.c. normale.

Come è noto la curva normale è simmetrica intorno a μ ed ha un caratteristico andamento campanulare. Per valutare le probabilità riguardanti questa distribuzione è molto utile far riferimento alla funzione cumulata delle probabilità o funzione di ripartizione $F(x) = P[X \leq x]$ che è data dall'area a sinistra di x della curva di densità $f(x)$. Ad esempio, si ricava immediatamente che

$$P[x_1 < X \leq x_2] = F(x_2) - F(x_1).$$

Sfortunatamente la $F(x)$ non è esprimibile con una funzione matematica semplice.

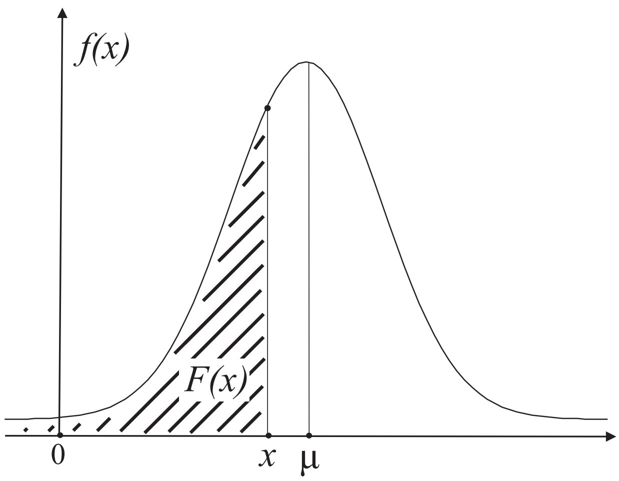


Figura 1.5: *Grafico della curva normale.*

E' stata così tabulata la funzione $F(x)$ per un cospicuo numero di valori x per il caso particolare $\mu = 0$ e $\sigma^2 = 1$.
In effetti si riescono a valutare i valori $F(x)$ di una qualsiasi variabile casuale normale, ($-\infty < \mu < +\infty; \sigma^2 > 0$), ricorrendo alla tabulazione della variabile casuale normale con $\mu = 0$ e $\sigma^2 = 1$.

——————— . . ——————— . . ———————

La variabile casuale normale con $\mu = 0$ e $\sigma^2 = 1$ è denominata variabile casuale normale standard ed è solitamente indicata con Z. La funzione di densità di Z è fornita da

$$f(z) = \frac{1}{1\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{z-0}{1}\right)^2\right\} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\}. \tag{4.20}$$

Si osservi che la v.c. normale standard è simmetrica intorno allo zero.

———— . . ———— . . ————

In seguito saranno utili le seguenti proprietà.
Proprietà 1 (Questa proprietà viene data senza dimostrazione)

Sia X una v.c. normale. Sia $Y = a + bX$. Allora anche la v.c. Y è una variabile casuale normale con $E(Y) = a + bE(X)$ e $Var(Y) = b^2 Var(X)$.

Proprietà 2

Sia X una v.c. con aspettativa μ e varianza σ^2. Sia $Z = \frac{X-\mu}{\sigma} = -\frac{\mu}{\sigma} + \frac{1}{\sigma}X$. Allora

$$E(Z) = -\frac{\mu}{\sigma} + \frac{1}{\sigma}E(X) = -\frac{\mu}{\sigma} + \frac{\mu}{\sigma} = 0$$

$$Var(Z) = \left(\frac{1}{\sigma}\right)^2 Var(x) = 1.$$

Si noti che Z è una trasformazione lineare di X.

Tutte le v.c con aspettativa 0 e varianza 1 sono denominate v.c. standard. Questa proprietà vale per qualsiasi v.c. e non solo per la v.c. normale.

———— . . ———— . . ————

Corollario Sia X una v.c. normale con $E(X) = \mu$ e $Var(X) = \sigma^2$. Allora, la v.c. $Z = \frac{X-\mu}{\sigma} = -\frac{\mu}{\sigma} + \frac{1}{\sigma}X$ è una v.c. normale standard.

Dimostrazione. Per la proprietà 1 la v.c. Z è una v.c. normale in quanto Z è una trasformazione lineare di X con $a = -\frac{\mu}{\sigma}$ e $b = \frac{1}{\sigma}$. Inoltre dalla proprietà 2 deriva che Z è una v.c. standard. In conclusione la v.c. $Z = \frac{X-\mu}{\sigma}$ è una v.c. normale standard di cui è disponibile la tabulazione della funzione cumulata delle probabilità.

———— . . ———— . . ————

Ecco come ricavare le probabilità cumulate di una normale qualsiasi tramite quelle della normale standardizzata.

Si abbia una v.c. normale X di parametri μ e σ (qualsiasi) e si voglia calcolare la $P[X \leq x]$.

L'evento $[X \leq x]$ è equivalente all'evento $[X - \mu \leq x - \mu]$, nonchè all'evento $\left[\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right]$. In formula

$$[X \leq x] \longleftrightarrow \left[\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right].$$

L'equivalenza significa che se si verifica l'evento $[X \leq x]$ allora si verifica l'evento $\left[\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right]$ e viceversa. Consegue che questi due eventi hanno la stessa probabilità di verificarsi. Quindi

$$F(x) = P[X \leq x] = P\left[\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right]. \tag{4.21}$$

La variabile $\frac{X-\mu}{\sigma} = Z$ è una variabile normale standard. La probabilità cumulata $P[Z \leq z]$ della v.c. normale standard Z è universalmente indicata con il simbolo $\Phi(z)$: $P[Z \leq z] = \Phi(z)$.

Tenuto conto di ciò la (4.21) diventa

$$F(x) = P[X \leq x] = P\left[Z \leq \frac{x-\mu}{\sigma}\right] = \Phi\left(z = \frac{x-\mu}{\sigma}\right). \tag{4.22}$$

La tabulazione di $\Phi(z)$ è riportata in fondo al libro. Essendo la normale standard simmetrica intorno allo zero è stato allora sufficiente tabulare la $\Phi(z)$ solo per $z \geq 0$.

Dal grafico che segue si osserva che

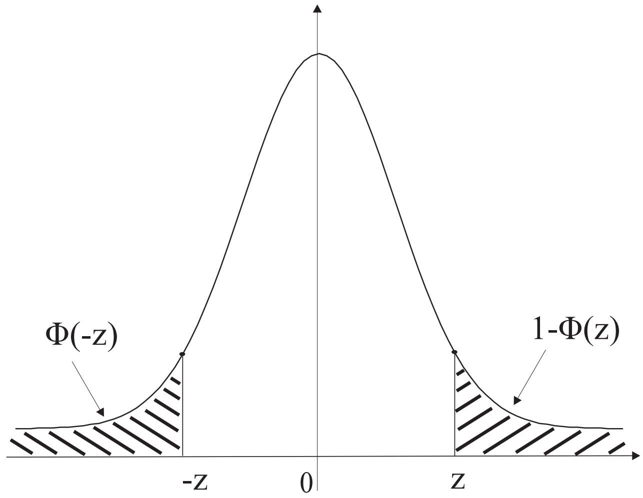
$$P(Z \leq -z) = P[Z > z],$$

ovvero, ricordando che $P(Z > z) = 1 - P[Z \leq z]$ si ha

$$P[Z \leq -z] = 1 - P[Z \leq z],$$

ovvero

$$\Phi(-z) = 1 - \Phi(z), \quad (z \geq 0). \tag{4.23}$$



Spesso si desidera valutare la probabilità che Z sia compresa in un intervallo simmetrico intorno allo zero.

$$\begin{aligned} P[-z < Z \leq z] &= P[Z \leq z] - P[Z \leq -z] \\ &= \Phi(z) - [1 - \Phi(z)] \\ &= 2\Phi(z) - 1, \quad (z \geq 0). \end{aligned} \tag{4.24}$$

Esempio 4.4.1 *Ricavare le seguenti probabilità:*

i) $P[-1 \leq Z \leq +1]$

ii) $P[-2 \leq Z \leq +2]$

iii) $P[-3 \leq Z \leq +3]$.

Utilizzando la (4.24) e le tavole della funzione $\Phi(z)$ si ottiene:

i) $P[-1 \leq Z \leq +1] = 2\Phi(1) - 1 = 2 \cdot 0,8413 - 1 = 0,6826$

ii) $P[-2 \leq Z \leq +2] = 2\Phi(2) - 1 = 2 \cdot 0,9772 - 1 = 0,9544$

iii) $P[-3 \leq Z \leq +3] = 2\Phi(3) - 1 = 2 \cdot 0,9987 - 1 = 0,9974$.

Esempio 4.4.2 *Una variabile casuale è distribuita normalmente con* $\mu = 50$ *e* $\sigma = 5$. *Ricavare:*

i) $P[X > 62]$

ii) $P[|X - 50| < 8]$

iii) $P[|X-40|>5]$.

i)
$$\begin{aligned} P[X>62] &= 1-P[X\leq 62] \\ &= 1-\Phi\left(\frac{62-50}{5}\right) \\ &= 1-\Phi\left(\frac{12}{5}\right) \\ &= 1-\Phi(2,4) \\ &= 1-0,9918 = 0,0082 \end{aligned}$$

ii) $P[|X-50|<8] = P[50-8<X<50+8] = P[42\leq X\leq 58]$.
Standardizzando la v.c. X si ha

$$\begin{aligned} P[|X-50|<8] \quad &= P\left[\tfrac{42-50}{5}\leq \tfrac{X-50}{5}\leq \tfrac{58-50}{5}\right] \\ &= P[-1,6\leq Z\leq 1,6] \\ &= 2\Phi(1,6)-1 \\ &= 2\cdot 0,9452-1 = 0,8904 \end{aligned}$$

iii)
$$\begin{aligned} P[|X-40|>5] &= 1-P[|X-40|\leq 5] \\ &= 1-P[40-5\leq X\leq 40+5] \\ &= 1-P[35\leq X\leq 45] \\ &= 1-P\left[\frac{35-50}{5}\leq\frac{X-50}{5}\leq\frac{45-50}{5}\right] \\ &= 1-P[-3\leq Z\leq -1] \\ &= 1-[\Phi(-1)-\Phi(-3)] \\ &= 1-[(1-\Phi(1))-(1-\Phi(3))] \\ &= 1-\Phi(3)+\Phi(1) \\ &= 1-0,9987+0,8413 = 0,8426. \end{aligned}$$

————— . . ————— . . —————

Si presenta ora la *proprietà degli intervalli tipici* della distribuzione normale.

Sia X una v.c. normale con $E(X)=\mu$ e $Var(X)=\sigma^2$. Si può dimostrare che per ogni $\gamma>0$ si ha

$$P[\mu-\gamma\sigma\leq X\leq \mu+\gamma\sigma] = 2\Phi(\gamma)-1. \tag{4.25}$$

Dimostrazione $P[\mu - \gamma\sigma \leq X \leq \mu + \gamma\sigma] = P[X \leq \mu + \gamma\sigma]+$ $-P[X \leq \mu - \gamma\sigma]$. Impiegando la (4.22) si ha

$$\begin{aligned} P[\mu - \gamma\sigma \leq X \leq \mu + \gamma\sigma] &= \Phi\left(\frac{\mu + \gamma\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - \gamma\sigma - \mu}{\sigma}\right) \\ &= \Phi(\gamma) - \Phi(-\gamma) = 2\Phi(\gamma) - 1. \end{aligned} \tag{4.26}$$

Per $\gamma = 1$, 2 e 3 si hanno i cosiddetti intervalli tipici. (Si riveda l'esempio (4.4.1)).

a) $P[\mu - \sigma \leq X \leq \mu + \sigma] = 2\Phi(1) - 1 = 0,6826$;

b) $P[\mu - 2\sigma \leq X \leq \mu + 2\sigma] = 2\Phi(2) - 1 = 0,9544$;

c) $P[\mu - 3\sigma \leq X \leq \mu + 3\sigma] = 2\Phi(3) - 1 = 0,9974$;

5 *Alcune leggi di calcolo delle probabilità*

In questo capitolo si esamineranno alcune leggi probabilistiche che rivestono particolare importanza nell'inferenza statistica.

5.1 La disuguaglianza di Cebiceff

Sia X una v.c. con $E(X) = \mu$ e $Var(X) = \sigma^2$. Allora

$$P[|X - \mu| \geq c] \leq \frac{Var(X)}{c^2}, \; (c > 0). \tag{5.1}$$

Si farà la dimostrazione nel caso X sia una v.c. discreta che assume i valori x_j con probabilità $p(x_j)$, $j = 1, 2, \ldots, s$.
Dimostrazione. Assegnato un numero positivo c si costruiscano due gruppi di valori della v.c. X:

i) Primo gruppo $|x_j - \mu| < c$;

ii) Secondo gruppo $|x_j - \mu| \geq c$.

La varianza di X è fornita da

$$\begin{aligned} Var(X) &= \sum_{j=1}^{s} (x_j - \mu)^2 p(x_j) \\ &= \sum_{j:|x_j - \mu| < c} (x_j - \mu)^2 p(x_j) + \sum_{j:|x_j - \mu| \geq c} (x_j - \mu)^2 p(x_j) \end{aligned} \tag{5.2}$$

Essendo le due ultime sommatorie quantità non negative si ha

$$Var(X) \geq \sum_{j:|x_j-\mu|\geq c} (x_j-\mu)^2 p(x_j). \tag{5.3}$$

Ciascun termine dell'ultima sommatoria è tale che

$$|x_j-\mu| \geq c \longrightarrow (x_j-\mu)^2 \geq c^2 \longrightarrow$$
$$\longrightarrow \sum_{j:|x_j-\mu|\geq c} (x_j-\mu)^2 p(x_j) \geq \sum_{j:|x_j-\mu|\geq c} c^2 p(x_j).$$

Tenuto conto di quest'ultima disuguaglianza la (5.3) diventa

$$Var(X) \geq c^2 \sum_{j:|x_j-\mu|\geq c} p(x_j). \tag{5.4}$$

La sommatoria $\sum_{j:|x_j-\mu|\geq c} p(x_j)$ non è altro che la probabilità dell'evento $|X-\mu| \geq c$ che si compone di tutti i valori x_j che hanno uno scarto assoluto (in modulo) da μ maggiore o uguale a c. Pertanto la (5.4) si può anche scrivere così

$$Var(X) \geq c^2 P(|X-\mu| \geq c).$$

Da quest'ultima diseguaglianza si ottiene la (5.1). ■

——— . . ——— . . ———

La disuguaglianza (5.1) fornisce una valutazione per eccesso della probabilità che X abbia uno scarto (in valore assoluto) da μ almeno uguale a c.

Alcune volte la diseguaglianza può fornire informazioni banali come mostra il seguente

Esempio 5.1.1 *Sia X una v.c. con $\mu = 10$ e con $\sigma = 2$. Sia $c = 2$.*Allora dalla (5.1) si ricava

$$P[|X-10| \geq 2] \leq \frac{2^2}{2^2} = 1.$$

Trattasi di un risultato banale in quanto per ogni evento A la sua probabilità è compresa fra zero e 1: $0 \leq P(A) \leq 1$.

Si ponga ora $c = 2,5$. La (5.1) informa che

$$P[|X-10| \geq 2,5] \leq \frac{2^2}{2,5^2} = \frac{4}{6,25} = 0,64.$$

Quest'ultimo risultato informa che per ogni v.c. X con $\sigma = 2$, la probabilità che lo scarto $|X-\mu|$ sia maggiore o uguale a 2,5 è al più uguale a 0,64.

La diseguaglianza informa altresì che, a parità di c la probabilità che lo scarto sia $|X-\mu| \geq c$ "tende " ad aumentare con σ^2.

Il valore (per eccesso) fornito dalla disuguaglianza, alcune volte, può essere molto più elevato del valore vero, come mostra l'esempio che segue.

Esempio 5.1.2 *Sia X una v.c. normale con $\mu = 10$ e $\sigma = 2$. Sia $c = 3,5$. Ricavare il valore vero della $P[|X - 10| \geq 3,5]$ ed il valore fornito dalla disuguaglianza di Cebiceff.*

Dalla (5.1) si ricava

$$P[|X - 10| \geq 3,5] \leq \frac{2^2}{3,5^2} = \frac{4}{12,25} = 0,3265.$$

Ecco ora il valore vero di tale probabilità

$$\begin{aligned}
P[|X - 10| \geq 3,5] &= 1 - P[|X - 10| < 3,5]. \\
P[|X - 10| < 3,5] &= P[6,5 \leq X \leq 13,5] \\
&= \Phi\left(\frac{13,5 - 10}{2}\right) - \Phi\left(\frac{6,5 - 10}{2}\right) \\
&= \Phi(1,75) - \Phi(-1,75) = 2\Phi(1,75) - 1 \\
&= 2 \cdot 0,9599 - 1 = 0,9198.
\end{aligned}$$

Pertanto

$$P[|X - 10| \geq 3,5] = 1 - 0,9198 = 0,0802.$$

Si nota così una notevole divergenza fra il valore vero (0,0802) ed il valore in eccesso (0,3265) fornito dalla diseguaglianza di Cebiceff.

Quanto evidenziato dagli esempi 5.1.1 e 5.1.2 non deve far credere che la diseguaglianza di Cebiceff non trovi nessuna applicazione. Innanzitutto va tenuto presente che per la sua validità si prescinde dalla conoscenza della legge di probabilità della v.c.X,e che quindi può impiegarsi in qualsiasi situazione purchè della variabile casuale si conosca il valore di μ e di σ.

Inoltre, la diseguglianza di Cebiceff è molto utile per esaminare alcuni aspetti del comportamento di variabili casuali associate al replicarsi del numero delle prove di un esperimento, nel caso in cui le prove crescano indefinitivamente.

5.2 Legge (debole) dei grandi numeri

La diseguaglianza di Cebiceff è valida per qualsiasi v.c.Z con aspettativa $E(Z) < \infty$ e varianza $Var(Z) < \infty$. Pertanto è valida anche per la v.c. frequenza relativa campionaria $Y = \dfrac{X}{n}$, essendo X una v.c. binomiale di parametri

n e p. In effetti:

$$E(Y) = E\left(\frac{1}{n}X\right) = \frac{1}{n}E(X) = \frac{np}{n} = p$$
$$Var(Y) = Var\left(\frac{1}{n}X\right) = \left(\frac{1}{n}\right)^2 Var(X) = \frac{1}{n^2}npq = \frac{pq}{n}.$$

Si osservi che la v.c.Y assume i valori: $\frac{0}{n}, \frac{1}{n}, \ldots, \frac{x}{n}, \ldots, \frac{n-1}{n}, \frac{n}{n}$.

Si ricordi che X indica il numero delle volte in cui il risultato A si presenta in n prove e che $p = P(A)$ è la probabilità di A in una prova.

Applicando la diseguaglianza di Cebiceff alla v.c. Y si ha:

$$0 \leq P\left\{|\frac{X}{n} - E\left(\frac{X}{n}\right)| \geq c\right\} \leq \frac{Var\left(\frac{X}{n}\right)}{c^2} \longrightarrow$$
$$\longrightarrow 0 \leq P\left\{\left|\frac{X}{n} - p\right| \geq c\right\} \leq \frac{pq}{c^2 n}.$$

Questa diseguaglianza è valida per qualsiasi valore di n. All'aumentare di n il rapporto $\frac{pq}{c^2}\frac{1}{n}$ tende a zero. Conseguentemente, tende a zero anche la $P\left\{\left|\frac{X}{n} - p\right| \geq c\right\}$, in quanto questa probabilità è compresa fra lo zero e $\frac{pq}{c^2}\frac{1}{n}$.

Si è così dimostrata la cosiddetta

Legge (debole) dei grandi numeri

$$\lim_{n\to\infty} P\left\{\left|\frac{X}{n} - p\right| \geq c\right\} = 0, \tag{5.5}$$

essendo X una v.c. binomiale con parametri n e p.

La legge dei grandi numeri afferma che per n "grande" la divergenza fra la v.c. frequenza relativa $\frac{X}{n}$ e la sua aspettativa p (probabilità di A in una prova) è molto limitata. Inoltre all'aumentare di n la divergenza fra $\frac{X}{n}$ e $p = P(A)$ tende a ridursi. Questo significa che la v.c. $\frac{X}{n}$ tende a stabilizzarsi intorno a p. Infatti fissato un valore di c piccolo a piacere, all'aumentare di n tende a zero la probabilità di avere uno scarto fra $\frac{X}{n}$ e p, maggiore di c.

Questa legge che riguarda relazioni esistenti fra entità astratte (entità previste nel modello probabilistico) ha nel mondo reale il suo corrispettivo costituito dalla “proprietà della stabilizzazione delle frequenze relative nel lungo periodo”. Questa “concordanza” fra *mondo astratto* delle variabili casuali e *mondo reale* delle frequenze relative sperimentali fa comprendere che i “modelli probabilistici” sono idonei a rappresentare gli esperimenti casuali. Si badi però che la legge dei grandi numeri non dimostra la “proprietà della stabilizzazione delle frequenze relative col crescere del numero delle prove”, essendo quest’ultima ottenuta solo tramite la constatazione empirica del comportamento delle frequenze relative (sperimentali).

5.3 Teorema del “limite” centrale

Il teorema del “limite” centrale riveste un ruolo molto importante nei casi in cui bisogna ricavare la funzione cumulata delle probabilità di una v.c. che si può rappresentare come somma di un gran numero di variabili casuali. Si darà la versione semplice di detto teorema, senza la dimostrazione, che richiederebbe tecniche di analisi matematica non elementari.

Teorema (del limite centrale) *Si abbia la successione* $X_1, X_2, \ldots, X_i, \ldots \ldots, X_n$ *di v.c. Per ipotesi le v.c.* X_i*:*

i) hanno la stessa distribuzione di probabilità;

ii) sono indipendenti in probabilità;

iii) hanno aspettativa $\mu = E(X_i)$ *e varianza* $\sigma^2 = Var(X_i)$ *finite.*

E’ possibile, allora, dimostrare che

$$P\left\{\frac{(X_1 + X_2 + \ldots + X_i + \ldots + X_n) - n\mu}{\sigma\sqrt{n}} \leq z\right\}$$

è approssimabile bene, all’aumentare di n*, da* $\Phi(z)$*, per ogni* z*, essendo* $\Phi(z)$ *la funzione cumulata delle probabilità della v.c. normale standard.*

——————— . . ——————— . . ———————

Ecco una illustrazione del teorema del limite centrale. Si consideri la v.c. $Y_n = X_1 + X_2 + \ldots + X_i + \ldots + X_n$.
Ovviamente

$$\begin{aligned} E(Y_n) &= E(X_1 + X_2 + \ldots + X_i + \ldots + X_n) \\ &= E(X_1) + E(X_2) + \ldots + E(X_i) + \ldots + E(X_n). \end{aligned}$$

Per ipotesi

$$E(X_i) = \mu \quad (i = 1, 2, \ldots, n).$$

Pertanto

$$E(Y_n) = \mu + \ldots + \mu + \ldots + \mu = n\mu.$$

Per ipotesi le v.c. X_i hanno varianza σ^2 e sono fra loro indipendenti. Pertanto

$$\begin{aligned}
Var(Y_n) &= Var(X_1 + X_2 + \ldots + X_i + \ldots + X_n) \\
&= Var(X_1) + Var(X_2) + \ldots + Var(X_i) + \ldots + Var(X_n) \\
&= \sigma^2 + \ldots + \sigma^2 + \ldots + \sigma^2 = n\sigma^2.
\end{aligned}$$

Ovviamente

$$\sigma(Y_n) = \sqrt{n}\sigma.$$

Standardizzando la v.c. Y_n si ha

$$Z_n = \frac{Y_n - E(Y_n)}{\sigma(Y_n)} = \frac{Y_n - n\mu}{\sqrt{n}\sigma}.$$

Si ricordi che: $E(Z_n) = 0$, $Var(Z_n) = 1$.
Si vuole valutare

$$P[Y_n \leq y] = P\left[\frac{Y_n - n\mu}{\sqrt{n}\sigma} \leq \frac{y - n\mu}{\sqrt{n}\sigma}\right] = P\left[Z_n \leq \frac{y - n\mu}{\sqrt{n}\sigma}\right].$$

Il teorema afferma che al crescere di n

$$\begin{aligned}
P[Y_n \leq y] &= P\left[\frac{Y_n - n\mu}{\sqrt{n}\sigma} \leq \frac{y - n\mu}{\sqrt{n}\sigma}\right] \\
&= P\left[Z_n \leq \frac{y - n\mu}{\sqrt{n}\sigma}\right] \simeq \Phi\left(\frac{y - n\mu}{\sqrt{n}\sigma}\right).
\end{aligned} \tag{5.6}$$

L'approssimazione della $P[Y_n \leq y]$ con la probabilità cumulata $\Phi\left(\frac{y - n\mu}{\sqrt{n}\sigma}\right)$ cresce al crescere di n.

Questo teorema deve la sua importanza al fatto che esso vale quale che sia la forma della distribuzione di X_i. Tuttavia, quanto più le v.c. X_i sono prossime alla v.c. normale tanto più velocemente $\frac{Y_n - n\mu}{\sqrt{n}\sigma}$ tende ad avere una funzione cumulata prossima a quella di una v.c. normale standard.

In molti casi, l'approssimazione è già buona per $n = 30$. Nel paragrafo che segue si applicherà il teorema del limite centrale per ricavare la probabilità cumulata di una binomiale con n elevato. In seguito, il teorema del limite centrale verrà impiegato per ricavare la distribuzione delle medie campionarie.

5.4 Approssimazione della distribuzione binomiale con quella normale

Sia X una variabile casuale binomiale di parametri n e p. Nel capitolo precedente si è rappresentata la v.c. binomiale come somma di n v.c. indicatori I_i fra loro indipendenti ed aventi la stessa probabilità di successo:

$$X = I_1 + I_2 + \ldots + I_n.$$

Inoltre $E(I_i) = p$ e $Var(I_i) = pq$. Pertanto in forza del teorema del limite centrale

$$P[X \leq x] = P\left[\frac{X - np}{\sqrt{npq}} \leq \frac{x - np}{\sqrt{npq}}\right] \simeq \Phi\left(\frac{x - np}{\sqrt{npq}}\right),$$

purchè n sia sufficientemente elevato.

Esempio 5.4.1 *Si lancia una moneta, ben equilibrata, $n = 100$ volte. Sia X il numero di teste ottenibili in 100 lanci. Ricavare*

$$P[X \leq 54] = \sum_{i=0}^{54} \binom{100}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{100-i}.$$

Questa probabilità si può approssimare, essendo n abbastanza grande, con

$$\Phi\left(\frac{54 - 100 \cdot 0,5}{\sqrt{100 \cdot 0,5 \cdot 0,5}}\right) = \Phi\left(\frac{4}{5}\right) = \Phi(0,8) = 0,7881.$$

L'approssimazione della binomiale con la normale si ritiene soddisfacente per $npq \geq 5$. Attualmente molti programmi di calcolo automatico forniscono i valori esatti delle probabilità cumulate della v.c. binomiale.

——————— . . ——————— . . ———————

Si tenga presente che nel caso in cui p è molto piccolo ($p = 0,02$) l'approssimazione alla normale è molto lenta. Ciò dipende dalla notevole asimmetria della v.c. indicatore I_i che assume i valori 0 e 1 rispettivamente con probabilità $(1 - p) = 0,98$ e $p = 0,02$. Si noti che l'aspettativa $E(I_i) = 0,02$.
Se si ha una v.c. binomiale X con $n = 60$ e $p = 0,02$ si ha $E(X) = 60 \times 0,02 = 1,2$ e $Var(X) = 1,176$. In questo caso, per approssimare la probabilità della binomiale non si impiega la v.c. normale bensì una v.c. discreta denominata v.c. di Poisson che ha la seguente legge di probabilità

$$p(x) = \frac{e^{-\mu}\mu^x}{x!}, \quad x = 0, 1, 2, ...$$

essendo $\mu > 0$ l'unico parametro della distribuzione che coincide con l'aspettativa della v.c.

Esempio 5.4.2 *Si supponga che un istituto bancario internazionale abbia concesso $n = 100$ prestiti a piccoli imprenditori dislocati in differenti aree geografiche. Da pluriennale esperienza si sa che la probabilità di insolvenza (entro un anno) è pari a $p = 0,02$. Si supponga che le v.c. indicatori I_i associate all'insolvenza $(I_i = 1)$ o meno $(I_i = 0)$ dell'i-mo imprenditore $(i = 1, 2, ..., 100)$ siano fra loro indipendenti. Si supponga altresì che la probabilità di insolvenza sia costantemente pari a $p = 0,02$. Sia X il numero di insolvenze in un anno. Ricavare*

i) $E(X)$;

ii) $P(X \leq 2)$.

Con le ipotesi indicate X ha un distribuzione binomiale con $n = 100$ e $p = 0,02$.

i) Pertanto

$$E(X) = 100 \cdot 0,02 = 2$$

ii)

$$\begin{aligned} P(X \leq 2) &= P\left[(X = 0) \cup (X = 1) \cup (X = 2)\right] \\ &= P(X = 0) + P(X = 1) + P(X = 2). \end{aligned}$$

Nel caso della v.c. binomiale in considerazione

$$p(x) = \left(\begin{array}{c} 100 \\ x \end{array} \right) (0,02)^x \, (0,98)^{100-x} = \frac{(100)_x}{x!} (0,02)^x (0,98)^{100-x}.$$

Si ricordi che $\left(\begin{array}{c} N \\ h \end{array} \right) = \dfrac{(N)_h}{h!}$, essendo $(N)_n$ le dispozioni di N oggetti presi h alla volta:

$$(N)_h = N(N-1)\ldots(N-h+1).$$

Si ha così

$$p(0) = \left(\begin{array}{c} 100 \\ 0 \end{array} \right) (0,02)^0 \, (0,98)^{100} = 1 \cdot 1 \cdot 0,13262.$$

$$p(1) = \left(\begin{array}{c} 100 \\ 1 \end{array} \right) (0,02)^1 \, (0,98)^{99} = 100 \cdot 0,02 \cdot 0,98^{99} = 0,27065$$

$$p(2) = \left(\begin{array}{c} 100 \\ 2 \end{array} \right) (0,02)^2 \, (0,98)^{98} = \frac{100 \cdot 99}{2} \cdot 0,0004 \cdot 0,13809 = 0,27341.$$

In conclusione

$$P[X \leq 2] = 0,13262 + 0,27065 + 0,27341 = 0,67668.$$

Si proverà ad approssimare le sopra calcolate probabilità con la legge di probabilità di una Poisson con $\mu = E(X) = 2$

$$\begin{aligned}
p(x) &= \frac{e^{-\mu}\mu^x}{x!} = \frac{e^{-2}2^x}{x!} = \frac{2^x}{e^2 x!} \\
p(0) &= \frac{2^0}{e^2 0!} = \frac{1}{7,38905 \cdot 1} = 0,13533 \\
p(1) &= \frac{2^1}{7,38905 \cdot 1!} = \frac{2}{7,38905 \cdot 1} = 0,27067 \\
p(2) &= \frac{2^2}{7,38905 \cdot 2!} = \frac{4}{7,38905 \cdot 2} = 0,27067
\end{aligned}$$

Si osservi che queste tre ultime probabilità sono molto prossime alle probabilità esatte della binomiale. I calcoli svolti in questo esempio sono stati eseguiti con una normale “vecchia” calcolatrice tascabile.

CAPITOLO SECONDO

Elementi di inferenza

Come è stato già precisato nell'introduzione l'inferenza statistica predispone i metodi per estendere (generalizzare) le sintesi effettuate sui dati campionari alla totalità della popolazione di interesse.

L'inferenza predispone metodi che si possono dividere in due gruppi:

- metodi per la stima;
- metodi per le verifiche di ipotesi.

————— . . ————— . . —————

Con la stima si vuole valutare numericamente qualche caratteristica della popolazione (media aritmetica, varianza, coefficiente di correlazione, ecc..) sulla base dei dati del campione.

Esempio 5.4.3 *Si vogliono avere informazioni sul valore dell'incognita percentuale delle matricole, di una facoltà universitaria alquanto affollata, che fumano almeno cinque sigarette al giorno.*

La cosa più naturale che si può fare, per avere informazioni sulla predetta percentuale, consiste nello scegliere un campione casuale dalla popolazione di tutte le matricole e calcolare la percentuale di matricole, del campione, che fumano almeno cinque sigarette al giorno. Si supponga che detta percentuale sia risultata 30%. Si riterrà, allora, che la percentuale incognita riferentesi a tutte le matricole è pari ad un valore "prossimo" al 30%.

————— . . ————— . . —————

Per ipotesi statistica si intende una congettura circa: qualche caratteristica della popolazione (il valore di una percentuale, la media aritmetica di un carattere quantitativo, ecc..), il tipo di distribuzione di una variabile (normale, di Pareto, ecc..), il valore del coefficiente di correlazione fra due variabili X e Y della popolazione, ecc..

La problematica che si ha nelle verifiche di ipotesi consiste nel dover scegliere fra due ipotesi statistiche messe a confronto, sulla base di dati campionari. Per test statistico si intende, appunto, una procedura che sulla base dei dati campionari permette di scegliere una delle due ipotesi messe a confronto.

Esempio 5.4.4 *In un reparto un determinato tipo di macchina ha una percentuale di pezzi di scarto pari al* 3%*. Si intende acquistare un nuovo tipo di macchina per la quale i venditori garantiscono una percentuale di pezzi di scarto pari all'*1%*. Il compratore dubita che la nuova macchina possa operare con una percentuale di scarti inferiore al* 3% *ed intende avere una dimostrazione sulla qualità di quest'ultima. Su* $n = 1000$ *pezzi prodotti con la nuova macchina si sono avuti* $x = 12$ *pezzi di scarto.*

Sulla base di questa informazione campionaria è possibile ricavare una procedura (test statistico) che permette di optare per l'ipotesi del venditore o per l'ipotesi del compratore. Secondo il venditore la percentuale di pezzi di scarto con la nuova macchina è $p_1 = 1\%$. Secondo il compratore la percentuale di pezzi di scarto con la nuova macchina è $p_0 = 3\%$.

——— . . ——— . . ———

In questo volumetto saranno considerati solo gli aspetti elementari delle problematiche della stima.

6 *Stima*

Si introdurrà ora la simbologia che verrà impiegata nella trattazione della stima facendo riferimento al *campionamento con riposizione* di n unità da una popolazione di N unità.

In corrispondenza delle N unità $u_1, u_2, \ldots, u_N$ della popolazione il carattere quantitativo X assume rispettivamente i valori $a_1, a_2, \ldots, a_N$.

Si indichi con μ la media aritmetica del carattere X nella popolazione:

$$\mu = \frac{1}{N} \sum_{j=1}^{N} a_j. \tag{6.1}$$

Si indichi con σ^2 la varianza del carattere X nella popolazione:

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} (a_j - \mu)^2. \tag{6.2}$$

Con il campionamento si vuole stimare l'ignota media μ della popolazione.

Si indichi con X_1 la v.c. che rappresenta il valore di X ottenibile con la prima estrazione. La v.c. X_1 può assumere i valori $a_1, a_2, \ldots, a_N$ in quanto questi sono i valori di X associati alle N unità estraibili. Se si suppone che le N unità della popolazione hanno la stessa probabilità $\dfrac{1}{N}$ di essere estratte si può allora affermare che la v.c. X_1 assume i valori $a_1, a_2, \ldots, a_N$ rispettivamente con probabilità $\dfrac{1}{N}, \ldots, \dfrac{1}{N}, \ldots, \dfrac{1}{N}$. La distribuzione di X_1 può essere sintetizzata come segue

$$X_1 \equiv \left\{ \begin{array}{ccccc} a_1, & \ldots, & a_j, & \ldots, & a_N \\ \frac{1}{N}, & \ldots, & \frac{1}{N}, & \ldots, & \frac{1}{N} \end{array} \right\}.$$

Si indichi con X_2 la v.c. che rappresenta il valore di X ottenibile con la seconda estrazione. Dato che il campionamento è con riposizione, dopo ogni estrazione e successiva riposizione, si ricompone la popolazione di partenza. Pertanto la distribuzione di X_2 coincide con quella di X_1:

$$X_2 \equiv \left\{ \begin{array}{ccccc} a_1, & \ldots, & a_j, & \ldots, & a_N \\ \frac{1}{N}, & \ldots, & \frac{1}{N}, & \ldots, & \frac{1}{N} \end{array} \right\}.$$

In generale la v.c. X_i, che indica il valore di X ottenibile alla i-ma estrazione, ha la seguente distribuzione di probabilità

$$X_i \equiv \left\{ \begin{array}{ccccc} a_1, & \ldots, & a_j, & \ldots, & a_N \\ \frac{1}{N}, & \ldots, & \frac{1}{N}, & \ldots, & \frac{1}{N} \end{array} \right\}, \quad (i = 1, 2, ..., n). \qquad (6.3)$$

In conclusione le v.c. X_i hanno la stessa distribuzione (quella indicata nella (6.3)). Inoltre, le v.c. X_i, data la riposizione di ogni unità estratta, sono fra loro indipendenti in probabilità. Si è ora pronti per definire il *campionamento casuale*.

Per campione casuale si intende la successione delle n variabili casuali campionarie indipendenti

$$(X_1, \ldots, X_i, \ldots, X_n)$$

aventi la stessa distribuzione (6.3).

Con x_i si indica invece il valore effettivamente ottenuto alla i-ma estrazione ($i = 1, 2, \ldots, n$). In altre parole x_i è una determinazione della v.c. X_i ($i = 1, 2, \ldots, n$). Pertanto $(x_1, \ldots, x_i, \ldots, x_n)$ sono i valori di X effettivamente ottenuti con il campione.

Si tenga allora ben presente che $(X_1, \ldots, X_i, \ldots, X_n)$ indica una successione di n v.c. indipendenti ed identicamente distribuite, mentre $(x_1, \ldots, x_i, \ldots, x_n)$ indica una successione di n valori (n numeri).

——— . . ——— . . ———

Si è in precedenza affermato che con il campionamento si intende stimare l'ignota media μ della popolazione.

Dopo aver estratto il campione il metodo più semplice che si può impiegare per valutare μ è quello di stimarlo con la media aritmetica dei valori $(x_1, \ldots, x_i, \ldots, x_n)$. Si ha così

$$\bar{x} = \frac{1}{n}(x_1 + \ldots + x_i + \ldots + x_n).$$

Ovviamente, essendo x_i una determinazione della v.c. X_i ($i = 1, 2, \ldots, n$), deriva che $\bar{x}$ è una determinazione della v.c. media campionaria

$$\bar{X} = \frac{1}{n}(X_1 + \ldots + X_i + \ldots + X_n). \tag{6.4}$$

Come è stato ampiamente mostrato nell'introduzione nel passare da un campione all'altro varia il valore che effettivamente assume $\bar{X}$. Essendo $\bar{X}$ una v.c. ciò significa che ha una distribuzione di probabilità, una aspettativa $E(\bar{X})$, una varianza $Var(\bar{X})$, ecc..

La v.c. $\bar{X}$ è uno stimatore di μ, mentre $\bar{x}$ è una stima di μ.

Ecco come determinare $E(\bar{X})$ e $Var(\bar{X})$.

Conviene ricavare innanzitutto $E(\bar{X}_i)$ e $Var(\bar{X}_i)$. Ricordando che la v.c. X_i ha la distribuzione (6.3) si ha

$$E(X_i) = \sum_{j=1}^{N} a_j \frac{1}{N} = \frac{1}{N}\sum_{j=1}^{N} a_j = \mu \quad (i = 1, 2, \ldots, n). \tag{6.5}$$

In altre parole, alla i-ma estrazione si ha un valore X_i mediamente uguale alla media aritmetica che nella popolazione ha il carattere X.

$$Var(X_i) = \sum_{j=1}^{N}(a_j - \mu)^2 \frac{1}{N} = \frac{1}{N}\sum_{j=1}^{N}(a_j - \mu)^2 = \sigma^2 \quad (i = 1, 2, \ldots, n). \tag{6.6}$$

Anche la varianza di X_i coincide con la varianza che nella popolazione ha il carattere X.

Si è ora pronti per ricavare $E(\bar{X})$ e $Var(\bar{X})$.

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)$$

Per la (6.5) la $E(X_i) = \mu$, per cui

$$E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}n\mu = \mu.$$

In altre parole l'aspettativa della v.c. media campionaria coincide con la media aritmetica μ della popolazione.

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}Var\left(\sum_{i=1}^{n} X_i\right).$$

Per ipotesi le v.c. X_i sono fra loro indipendenti per cui

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var\left(X_i\right).$$

Inoltre per la (6.6) $Var(X_i) = \sigma^2$. Pertanto

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \sigma^2 = n\sigma^2.$$

Tenuto conto di quest'ultimo risultato la $Var(\bar{X})$ diventa

$$Var(\bar{X}) = \frac{1}{n^2}n\sigma^2 = \frac{1}{n}\sigma^2. \tag{6.7}$$

In altre parole la varianza della media campionaria $\bar{X}$ è inversamente proporzionale alla numerosità del campione.

Nell'introduzione si era considerata l'estrazione da una popolazione di $N = 6$ unità, di tutti i campioni con riposizione di numerosità $n = 2$ e si era constatato che la media di tutte le medie campionarie era proprio uguale a μ e la varianza di tutte le medie campionarie era proprio uguale a $\frac{\sigma^2}{2} = \frac{\sigma^2}{n}$.

———— . . ———— . . ————

Naturalmente la media campionaria $\bar{X}$, non è l'unico, anche se il più naturale, stimatore per μ. In effetti si potrebbe anche stimare μ con la mediana degli n valori $(x_1, x_2, \ldots, x_i, \ldots, x_n)$.

Avendo la possibilità di stimare μ in modi diversi – con diversi stimatori – si pone il problema di scegliere fra stimatori alternativi. La scelta si baserà sulle caratteristiche (proprietà) degli stimatori.

In generale uno stimatore è tanto più buono, quanto più fornisce valori prossimi all'ignota quantità da stimare.

Nel caso in esame uno stimatore è tanto più buono quanto più fornisce valori (stime) prossimi a μ.

Si è ora pronti per generalizzare i concetti esposti.

6.1 Proprietà degli stimatori

i) Si supporrà di voler stimare la caratteristica θ della popolazione: θ può essere una media, un indice di variabilità, una frequenza relativa, ecc..

ii) Lo stimatore per θ si otterrà dal campione casuale $(X_1, X_2, \ldots, X_i, \ldots, X_n)$.

iii) Sia $T = g(X_1, X_2, \ldots, X_i, \ldots, X_n)$ uno stimatore per θ.

T si configura come funzione $g(X_1, X_2, \ldots, X_i, \ldots, X_n)$ delle n v.c. campionarie. Conseguentemente T è una v.c. campionaria.

——————— . . ——————— . . ———————

(Proprietà della non distorsione)
Una delle proprietà più apprezzate per uno stimatore T per θ è la non distorsione.

Definizione (Stimatore "non distorto" o "corretto") Uno stimatore T per θ si dice "non distorto" o "corretto" se

$$E(T) = \theta, \text{ per qualsiasi valore di } \theta.$$

Il seguente grafico aiuta a comprendere il significato della non distorsione.

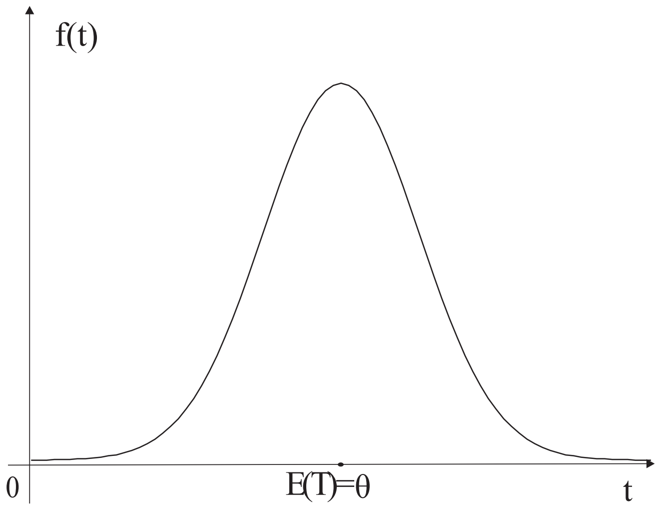


Figura 1.6: *Distribuzione di probabilità di uno stimatore non distorto.*

In ascissa sono riportati i valori t di T ed in ordinata la funzione di densità $f(t)$ di T.

Dato che T è uno stimatore corretto per θ ciò significa che il valore "centrale di T", cioè $E(T)$, coincide con θ.

Per uno stimatore corretto gli scarti positivi $(T - \theta)$, per $(T > \theta)$, bilanciano gli scarti negativi $(T - \theta)$, per $(T < \theta)$. Il bilanciamento significa che $E(T - \theta) = 0$. In effetti $E(T - \theta) = E(T) - \theta$: la differenza $E(T) - \theta$ è ovviamente pari a 0 per uno stimatore corretto. Se $E(T) \neq \theta$ lo stimatore è distorto. Per *distorsione* di uno stimatore T per θ si intende la differenza $E(T) - \theta$.

Per apprezzare la proprietà della non distorsione si supponga che $E(T) - \theta > 0$, si suppone cioè che vi sia una distorsione positiva. Ciò significa, come mostra il grafico, che lo stimatore T fornisce, mediamente, valori t maggiori di θ.

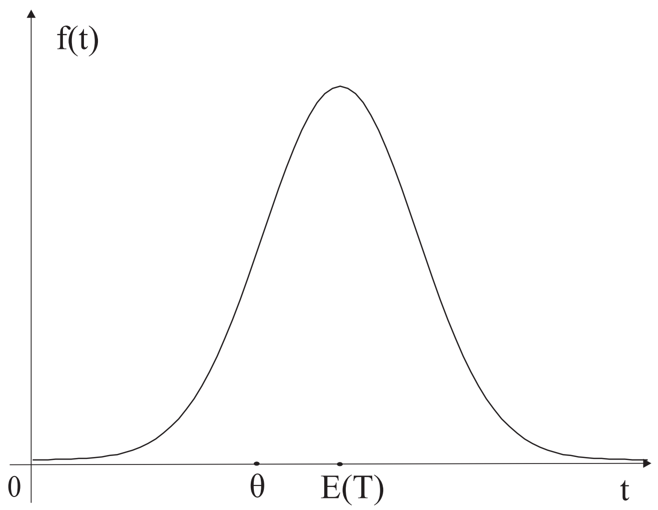


Figura 1.7: *Distribuzione di probabilità di uno stimatore con distorsione positiva.*

Se $E(T) \neq \theta$ si ha distorsione negativa. Ciò significa, come mostra il grafico, che lo stimatore T fornisce valori t che sono mediamente minori di θ.

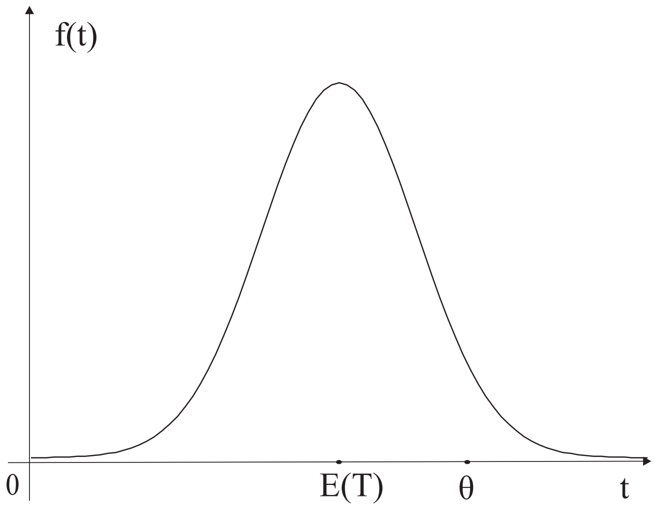


Figura 1.8: *Distribuzione di probabilità di uno stimatore con distorsione negativa.*

Esempio 6.1.1 In precedenza si è dimostrato che $E(\bar{X}) = \mu$. Si può allora affermare che la media campionaria $\bar{X}$ è uno stimatore corretto (non distorto) per μ.

——————— . . ——————— . . ———————

(Varianza di stimatori non distorti)
Può accadere che per stimare θ si disponga di due stimatori non distorti T_1 e T_2. In altre parole

$$E(T_1) = E(T_2) = \theta \text{ (per ogni } \theta).$$

Sorge allora la necessità di introdurre un criterio per poter scegliere fra due stimatori non distorti. Il criterio dovrà essere tale da far preferire quello stimatore che fornisce valori più prossimi all'ignoto θ. In altre parole si dovrà optare per quello stimatore i cui valori siano più concentrati intorno a θ. Il criterio può allora essere basato sull'ordine di grandezza degli scarti $|T - \theta|$, ovvero può essere basato sulla variabilità degli stimatori nel senso che fra due stimatori non distorti si preferirà quello con la varianza (scarto quadratico medio) più piccola.

In conclusione per valutare uno stimatore non distorto si considera la sua varianza.

Il grafico 1.9 si riferisce alle distribuzioni di probabilità di due stimatori non distorti (T_1 e T_2) per θ con $Var(T_1) < Var(T_2)$.

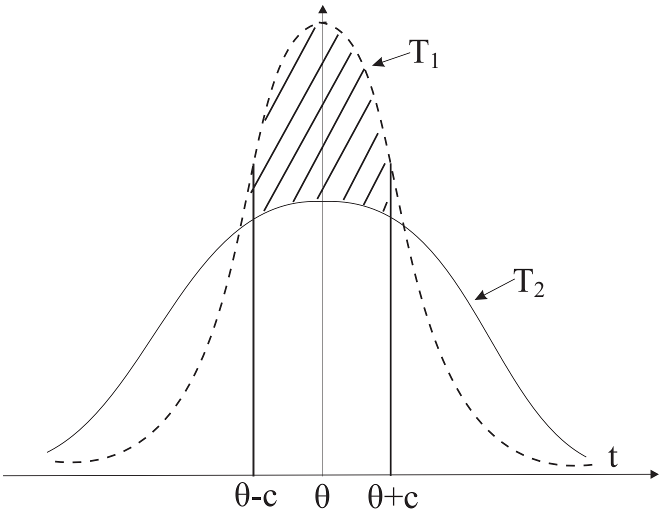


Figura 1.9: *Distribuzioni di probabilità di due stimatori non distorti per* θ *con* $Var(T_1) < Var(T_2)$.

Si noti che essendo $\sigma(T_1) < \sigma(T_2)$, verosimilmente

$$P\{\theta - c \leq T_1 \leq \theta + c\} > P\{\theta - c \leq T_2 \leq \theta + c\}.$$

La differenza tra le due probabilità coincide con l'area tratteggiata nel grafico.

————— . . ————— . . —————

(Proprietà della consistenza)

Un'altra proprietà desiderabile, per uno stimatore T per θ, è che deve fornire, all'aumentare di n, valori T sempre prossimi a θ. Si ha così la proprietà della consistenza. La proprietà della consistenza riguarda il comportamento dello stimatore all'aumentare di n. Per illustrare questa proprietà conviene indicare lo stimatore con T_n proprio per evidenziare che T_n è calcolato su un campione di ampiezza n.

Definizione (Consistenza semplice) Lo stimatore T_n per θ è consistente se, per ogni $c > 0$, si ha

$$\lim_{n\to\infty} P\{|T_n - \theta| \geq c\} = 0.$$

All'aumentare di n, la probabilità che T_n differisca da θ per più di un prefissato valore $c \geq 0$, tende a zero.

Esempio 6.1.2 *Consistenza dello stimatore* $\bar{X}_n$.
Si sa che

$$E(\bar{X}_n) = \mu;\ Var(\bar{X}_n) = \frac{1}{n}\sigma^2.$$

Applicando la disuguaglianza di Cebiceff alla v.c. $\bar{X}_n$ si ha

$$P\{|\bar{X}_n - \mu| \geq c\} \leq \frac{Var(\bar{X}_n)}{c^2}$$

$$\lim_{n\to 0} P\{|\bar{X}_n - \mu| \geq c\} \leq \lim_{n\to 0} \frac{1}{c^2}\frac{1}{n}\sigma^2.$$

Al divergere di n la parte di destra della disuguaglianza tende a zero. Conseguentemente tende a zero anche $P\{|\bar{X}_n - \mu| \geq c\}$.

Esempio 6.1.3 *Consistenza della frequenza relativa campionaria* $\hat{p}_n = \dfrac{X}{n}$.

Si vuole stimare la frequenza relativa p di una caratteristica A della popolazione con la frequenza relativa campionaria $\hat{p}_n = \dfrac{X}{n}$, essendo X il numero delle volte in cui A si è presentato nel campione con riposizione di numerosità n.

Si vuole dimostrare che

$$\lim_{n\to 0} P\left\{|\frac{X}{n} - p| \geq c\right\} = 0.$$

Questa verifica coincide con la legge (debole) dei grandi numeri in precedenza dimostrata nel paragrafo (5.2).

6.2 Stima intervallare

Si ipotizzi che T sia uno stimatore per θ.

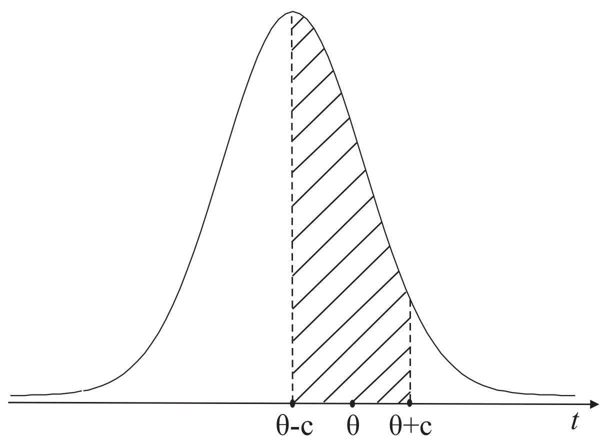


Figura 1.10: *Distribuzione di probabilità dello stimatore T per θ.*

In Fig. 1.10 si è riportato il grafico della funzione di densità $f(t)$ della v.c. campionaria T. (Si è ipotizzato che T sia una v.c. continua).

La probabilità che T sia compreso nell'intervallo $\theta \pm c$ è rappresentata dall'area tratteggiata. Al diminuire di c la probabilità prima evidenziata tende a zero. E' praticamente certo che il valore di T ottenuto dal campione non sia uguale al valore θ da stimare. Si preferisce, allora, fornire per θ non un solo valore (il valore di T che effettivamente si ha nel campione), ma un intervallo di valori nel quale dovrebbe essere compreso, con un certo grado di probabilità, l'ignoto valore di θ. Si passa così dalla stima puntuale (che fornisce un solo valore per θ) alla stima intervallare (che fornisce, sempre per θ, un intervallo di valori).

Si passa da una informazione su θ costituita da un solo valore di T (valore che ha però bassissima probabilità di essere uguale a θ) ad una informazione su θ costituita da un intervallo "casuale" di valori (H, V) che ha una probabilità elevata $(1-\alpha)$ di contenere al suo interno l'ignoto θ. H e V sono variabili casuali campionarie che delimitano gli estremi casuali dell'intervallo. In formula

$$P\{H \leq \theta \leq V\} = 1 - \alpha. \tag{6.8}$$

La relazione (6.8) si legge così: è pari a $(1-\alpha)$ la probabilità che l'intervallo casuale (H, V) contenga al suo interno l'ignoto valore θ.

Gli estremi H e V sono variabili casuali campionarie (basate cioè su $(X_1, X_2, \ldots, X_i, \ldots, X_n)$).

La probabilità $(1-\alpha)$ è denominata livello di confidenza e solitamente assume i seguenti valori: $0,90$; $0,95$; $0,99$.

Cosa significa che $(1-\alpha)$ è ad esempio uguale a $0,95$? Significa che se si scelgono $B = 1000$ campioni – ciascuno di numerosità n – e per ogni campione si ricava l'intervallo casuale (t, v), allora di questi 1000 intervalli 950 comprendono θ e 50 non lo comprendono. Vi è cioè un rischio probabilistico α che l'intervallo casuale non comprenda al suo interno il valore di θ.

Ovviamente, siccome il valore di θ è ignoto, non è possibile sapere se il particolare intervallo (t, v) ricavato con l'unico campione effettivamente estratto, contenga o meno al suo interno il valore ignoto θ. Si può solo "confidare" che il campione sia uno dei 950 intervalli buoni sugli ipotetici 1000 campioni ricavabili con la stessa procedura. Si può però essere sfortunati ed essere incappati in uno dei 50 (su 1000) intervalli non buoni.

6.3 Sintesi sulla stima puntuale di una media μ e di una frequenza relativa p e stima puntuale della varianza σ^2

Nelle pagine precedenti si sono considerate diverse proprietà degli stimatori $\bar{X}$ per μ e $\hat{p}_n$ per p. Ecco una sintesi di tali risultati.

i) Sia $(X_1, X_2, \ldots, X_i, \ldots, X_n)$ un campione casuale estratto con riposizione da una popolazione in cui il carattere X ha media aritmetica μ e varianza σ^2. Sia
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
la v.c. media campionaria. Si è dimostrato che

(a) $E(\bar{X}) = \mu$, pertanto $\bar{X}$ è uno stimatore corretto per μ;

(b) $Var(\bar{X}) = \frac{1}{n}\sigma^2$;

(c) $\bar{X}$ è uno stimatore consistente per μ;

(d) al divergere di n la media campionaria (teorema del limite centrale) tende alla distribuzione normale; quest'ultimo risultato permetterà di ricavare l'intervallo di confidenza per μ.

ii) Sia $p = \frac{N(A)}{N}$ la frequenza relativa della caratteristica A in una popolazione di N unità. $N(A)$ indica il numero delle unità della popolazione con la caratteristica A.

Sia X il numero di unità con la caratteristica A che si è riscontrato in un campione con riposizione di n unità.

Sia $\hat{p} = \frac{X}{n}$ la v.c. campionaria frequenza relativa. Si è dimostrato che

(a) $E(\hat{p}) = p$, pertanto $\hat{p}$ è uno stimatore corretto per p;

(b) $Var(\hat{p}) = \frac{1}{n}p(1-p)$;

(c) $\hat{p}$ è uno stimatore consistente per p;

(d) la v.c. X ha una distribuzione asintoticamente normale (per il teorema del limite centrale). Quest'ultimo risultato permetterà di ricavare l'intervallo di confidenza per p.

——————— . . ——————— . . ———————

Si mostrerà ora come ricavare uno stimatore non distorto per la varianza σ^2 del carattere X della popolazione, sempre nel caso di campionamento con riposizione.

Si considerino gli n valori campionari $(x_1, x_2, \ldots, x_i, \ldots, x_n)$ e la loro media aritmetica $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Sia A un numero reale qualsiasi. Per la proprietà di "minimo" della media aritmetica, vale la relazione:
$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - A)^2 - n(\bar{x} - A)^2.$$

Dato che A può essere un numero reale qualsiasi, si può porre $A = \mu$. La relazione precedente diventa

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \mu)^2 - n(\bar{x} - \mu)^2.$$

La relazione precedente vale se si sostituiscono i valori x_i e $\bar{x}$ con le corrispettive variabili casuali campionarie X_i e $\bar{X}$. Si ha così

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2. \tag{6.9}$$

Si calcolino ora le aspettative delle due parti della (6.9).

$$\begin{aligned} E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] &= E\left[\sum_{i=1}^{n}(X_i - \mu)^2\right] - nE(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^{n}\left[E(X_i - \mu)^2\right] - nVar(\bar{X}) \\ &= \sum_{i=1}^{n} Var(X_i) - nVar(\bar{X}). \end{aligned}$$

Per la (6.6) e per la (6.7) $Var(X_i) = \sigma^2$ e $Var(\bar{X}) = \frac{1}{n}\sigma^2$.

Pertanto la relazione precedente diventa

$$E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \sum_{i=1}^{n}\sigma^2 - n\frac{1}{n}\sigma^2 = n\sigma^2 - \sigma^2 = (n-1)\sigma^2$$

Dividendo per $(n-1)$ prima e seconda parte si ha

$$\begin{aligned} \frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] &= \sigma^2 \\ E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] &= \sigma^2 \end{aligned} \tag{6.10}$$

Si è così dimostrato che dividendo la devianza campionaria $\sum_{i=1}^{n}(X_i - \bar{X})^2$ per $(n-1)$ si ottiene la varianza campionaria corretta

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

S^2 è così uno stimatore corretto (non distorto) della varianza σ^2 della popolazione. Per ottenere uno stimatore corretto per σ^2 bisogna quindi dividere la devianza campionaria $\sum_{i=1}^{n}(X_i - \bar{X})^2$ per $(n-1)$ e non per n come potrebbe sembrare. In effetti, se si dividesse per n si otterrebbe:

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = E\left[\frac{(n-1)}{n}\frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]$$
$$= \frac{(n-1)}{n}E\left[\frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{n-1}{n}\sigma^2.$$

L'ultima relazione mostra che per n grande non vi è, praticamente, differenza fra la varianza campionaria $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ e la varianza campionaria corretta $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

6.4 Intervalli di confidenza per μ

Come è noto per lo stimatore puntuale $\bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i$ si ha: $E(\bar{X}) = \mu$ e $Var(\bar{X}) = \frac{1}{n}\sigma^2$. Inoltre, dato che le n v.c. del campione casuale $(X_1, X_2, \ldots, X_i, \ldots, X_n)$ sono indipendenti, identicamente distribuite e con comune aspettativa μ e varianza σ^2 (per ipotesi μ e σ^2 sono valori finiti), allora in forza del teorema del limite centrale la loro somma $Y_n = X_1 + X_2 + \ldots + X_i + \ldots + X_n$ tende al crescere di n a distribuirsi come una v.c. normale con $E(Y_n) = n\mu$ e $Var(Y_n) = n\sigma^2$.

Conseguentemente anche la v.c. $\bar{X}_n = \frac{1}{n}Y_n$, tende a distribuirsi, per la prima proprietà della v.c. normale, come una v.c. normale con $E(\bar{X}_n) = \mu$ e $Var(\bar{X}_n) = \frac{1}{n}\sigma^2$. Consegue altresì che anche la v.c. standardizzata $Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ tende al crescere di n a distribuirsi come una v.c. Z normale standard. Pertanto (si veda il grafico 1.11) deriva che

$$P\left\{-z_{\left(1-\frac{\alpha}{2}\right)} \leq Z_n \leq z_{\left(1-\frac{\alpha}{2}\right)}\right\} = (1-\alpha) \tag{6.11}$$

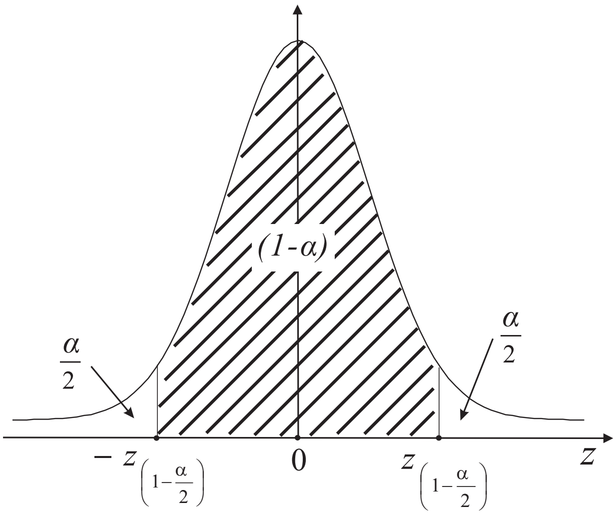


Figura 1.11: *Grafico della densità di una v.c. Z normale standard.*

La (6.11) informa che è $(1-\alpha)$ la probabilità che la media campionaria standardizzata $Z_n = \dfrac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ è compresa nell'intervallo $\left(-z_{\left(1-\frac{\alpha}{2}\right)}, z_{\left(1-\frac{\alpha}{2}\right)}\right)$. Il valore $z_{\left(1-\frac{\alpha}{2}\right)}$ si trova sulla tavola $\Phi(z)$ ed è quel valore di z tale che $\Phi(z) = \left(1-\frac{\alpha}{2}\right)$. Si consideri ora la diseguaglianza indicata in (6.11)

$$\left\{-z_{\left(1-\frac{\alpha}{2}\right)} \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\left(1-\frac{\alpha}{2}\right)}\right\} \tag{6.12}$$

e si moltiplichino i suoi termini per $\dfrac{\sigma}{\sqrt{n}}$. Si ottiene così la seguente diseguaglianza (equivalente alla prima)

$$\left\{-z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}\right\}.$$

Si sottragga ai tre termini il valore $\bar{X}_n$. Si ottiene la seguente diseguaglianza equivalente alle altre

$$\left\{-\bar{X}_n - z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X}_n + z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}\right\}.$$

Moltiplicando i termini di quest'ultima per -1 si ottiene

$$\left\{\bar{X}_n + z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{X}_n - z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}\right\},$$

che può ovviamente essere scritta come segue

$$\left\{\bar{X}_n - z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}\right\}, \tag{6.13}$$

La (6.13) è equivalente alla (6.12) il che significa che se si verifica l'una si verifica anche l'altra e viceversa. Ciò significa anche che i due eventi casuali

$$\left\{-z_{\left(1-\frac{\alpha}{2}\right)} \leq Z_n \leq z_{\left(1-\frac{\alpha}{2}\right)}\right\} \text{ e } \left\{\bar{X}_n - z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}\right\}$$

hanno la stessa probabilità $(1-\alpha)$ di verificarsi. In conclusione

$$P\left\{\bar{X}_n - z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}\right\} = (1-\alpha). \tag{6.14}$$

La (6.14) si legge come segue: è $(1-\alpha)$ la probabilità che l'intervallo casuale $\left(\bar{X}_n - z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}};\ \bar{X}_n + z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}\right)$ contenga al suo interno l'ignoto valore di μ. Nel caso considerato l'estremo inferiore dell'intervallo coincide con la v.c. $H = \bar{X}_n - z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}$ mentre l'estremo superiore coincide con la v.c. $V = \bar{X}_n + z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}$.
L'intervallo si ottiene quindi aggiungendo (per l'estremo superiore) e togliendo (per l'estremo inferiore) allo stimatore puntuale $\bar{X}_n$ il valore $z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}$.

Esempio 6.4.1 *Ricavare l'intervallo di confidenza per μ con livello di confidenza* $(1-\alpha) = 0,954$.

Bisogna innanzitutto ricavare $z_{\left(1-\frac{\alpha}{2}\right)}$.

$$\alpha = (1-0,954) = 0,046; \qquad \frac{\alpha}{2} = 0,023; \qquad \left(1-\frac{\alpha}{2}\right) = 0,977.$$

Dalle tavole della normale standard si ottiene $z_{(0,977)} = 2$.
L'intervallo di confidenza con $(1-\alpha) = 0,954$ è così fornito da: $\bar{X} \pm 2\frac{\sigma}{\sqrt{n}}$.

Osservazione L'impiego della (6.14) si basa, tra l'altro, sull'impiego del teorema del limite centrale che per n grande permette di approssimare la distribuzione di probabilità di $\bar{X}_n$ con quella della v.c. normale. Si ritiene che l'approssimazione si possa ritenere buona per $n \geq 50$.

Esempio 6.4.2 *La varianza del carattere X nella popolazione è pari a $\sigma^2 = 16$. Si è estratto un campione casuale con $n = 64$ e si è ottenuto $\sum_{i=1}^{64} x_i = 400$.*

Ricavare gli intervalli di confidenza per μ *sia con* $(1-\alpha) = 0,94$ *che con* $(1-\alpha) = 0,98$.

Si ricava innanzitutto la media campionaria

$$\bar{x} = \frac{1}{64}400 = 6,25.$$

Per $(1-\alpha) = 0,95$ si ha $\left(1-\frac{\alpha}{2}\right) = 0,975$, da cui $\Phi(z) = 0,975$ fornisce (tramite le tavole) $z_{\left(1-\frac{\alpha}{2}\right)} = 1,96$. Per $(1-\alpha) = 0,98$ si ha $\left(1-\frac{\alpha}{2}\right) = 0,99$ e tramite le tavole $\Phi(z)$ si ha $z_{(0,99)} = 2,33$.
I due intervalli sono allora forniti da

i) per $(1-\alpha) = 0,95$ si ha $\bar{x} \pm 1,96\frac{\sqrt{16}}{\sqrt{64}} \rightarrow$
$6,25 \pm 1,96 \cdot 0,5 \rightarrow 6,25 \pm 0,98 \rightarrow (5,27; 7,23)$;

ii) per $(1-\alpha) = 0,98$ si ha $6,25 \pm 2,33 \cdot 0,5 \rightarrow 6,25 \pm 1,65 \rightarrow (5,085; 7,415)$.

L'ampiezza dell'intervallo di confidenza è fornito da

$$\left(\bar{x} + z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}\right) - \left(\bar{x} - z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}\right) = 2z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}.$$

Si osservi che aumentando $(1-\alpha)$ aumenta anche l'ampiezza dell'intervallo.

Esempio 6.4.3 *La varianza del carattere* X *è ancora* $\sigma^2 = 16$. *Si è estratto un campione casuale di* $n = 256$ *unità e si è ottenuto* $\sum_{i=1}^{256} x_i = 1600$. *Ricavare gli intervalli di confidenza con* $(1-\alpha) = 0,95$ *e con* $(1-\alpha) = 0,98$.

$$\bar{x} = \frac{1}{256}1600 = 6,25$$

$$\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{16}}{\sqrt{256}} = \frac{4}{16} = 0,25.$$

I due intervalli risultano:

i) per $(1-\alpha) = 0,95$ si ha $\bar{x} \pm 1,96 \cdot 0,25 \rightarrow 6,25 \pm 0,49 \rightarrow (5,76; 6,74)$;

ii) per $(1-\alpha) = 0,98$ si ha $\bar{x} \pm 2,33 \cdot 0,25 \rightarrow 6,25 \pm 0,5825 \rightarrow (5,6675; 6,8325)$.

Si noti che a parità di $(1-\alpha)$ nel passare da $n = 64$ a $n = 4 \cdot 64 = 256$ la lunghezza degli intervalli diminuisce.

————— . . ————— . . —————

Per poter impiegare gli intervalli di confidenza

$$\bar{X} \pm z_{\left(1-\frac{\alpha}{2}\right)} \frac{\sigma}{\sqrt{n}}$$

bisogna conoscere il valore della varianza σ^2 della popolazione (è quanto è stato ipotizzato negli esempi 2 e 3).

Nel caso in cui σ^2 non è noto si stima il valore di quest'ultimo con la varianza campionaria corretta

$$S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Quindi si impiega il seguente intervallo di confidenza

$$\bar{X} \pm z_{\left(1-\frac{\alpha}{2}\right)} \frac{\sqrt{S^2}}{\sqrt{n}}. \tag{6.15}$$

Per poter impiegare la (6.15) è necessario avere una numerosità campionaria con $n \geq 150$.

Esempio 6.4.4 *Si vuole ricavare un intervallo di confidenza per l'ignota media μ del carattere X. Si desidera $(1-\alpha) = 0,95$. Si è estratto un campione casuale con $n = 200$ ed è risultato:* $\sum_{i=1}^{200} x_i = 30.000$; $\sum_{i=1}^{200} x_i^2 = 4.545.000$.

La media $\bar{x}$ risulta 150.

La devianza campionaria si può calcolare con la relazione seguente (procedimento indiretto per il calcolo della devianza)

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2. \tag{6.16}$$

Si ha così

$$\sum_{i=1}^{200} (x_i - \bar{x})^2 = 4.545.000 - 200 \cdot 150^2 \qquad \rightarrow$$

$$\sum_{i=1}^{200} (x_i - 150)^2 = 4.545.000 - 200 \cdot 22.500 = 45.000.$$

Pertanto per S^2 otteniamo il valore

$$s^2 = \frac{1}{200-1} \cdot 45.000 = 226,13065.$$

L'intervallo di confidenza risulta così

$$\bar{x} \pm 1,95 \cdot \frac{\sqrt{226,13}}{\sqrt{200}} \to \bar{x} \pm 1,95 \cdot 1,0633 \to 150 \pm 2,0735 \to (147,9265; 152,0735).$$

6.5 Intervalli di confidenza per la frequenza relativa p

In una popolazione di N unità statistiche la frequenza relativa di una caratteristica A è pari a $p = \dfrac{N(A)}{N}$, essendo $N(A)$ il numero di unità statistiche con la caratteristica A. Si estrae con riposizione un campione di numerosità n. Sia X il numero delle estrazioni in cui si è presentata A. La v.c. X ha una distribuzione binomiale di parametri n e p. Sia $\hat{p} = \dfrac{X}{n}$ lo stimatore puntuale per p. Come è noto: $E(\hat{p}) = p$ e $Var(\hat{p}) = \dfrac{pq}{n}$. Inoltre, per n elevato la v.c. X tende a distribuirsi secondo una v.c. normale. Conseguentemente anche $\hat{p} = \dfrac{1}{n}X$ tende a distribuirsi secondo una v.c. normale con aspettativa p e varianza $\dfrac{pq}{n}$. Standardizzando $\hat{p}$ si ottiene la v.c. $\dfrac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$ che al crescere di n tende a distribuirsi come una v.c. normale standard. Pertanto, sempre per n elevato si ha

$$P\left\{-z_{\left(1-\frac{\alpha}{2}\right)} \leq \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \leq z_{\left(1-\frac{\alpha}{2}\right)}\right\} = (1-\alpha). \tag{6.17}$$

Dalla disuguaglianza indicata nella (6.17)

$$\left\{-z_{\left(1-\frac{\alpha}{2}\right)} \leq \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \leq z_{\left(1-\frac{\alpha}{2}\right)}\right\} \tag{6.18}$$

si ottiene, con la stessa procedura impiegata per la disuguaglianza (6.12) riguardante $\bar{X}_n$, la relazione

$$\left\{\hat{p} - z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{pq}{n}} \leq p \leq \hat{p} + z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{pq}{n}}\right\} \tag{6.19}$$

che è equivalente alla (6.18), pertanto

$$P\left\{\hat{p} - z_{(1-\frac{\alpha}{2})}\sqrt{\frac{pq}{n}} \le p \le \hat{p} + z_{(1-\frac{\alpha}{2})}\sqrt{\frac{pq}{n}}\right\} = (1-\alpha). \tag{6.20}$$

In conclusione l'intervallo di confidenza per p con "grado di confidenza" $(1-\alpha)$ è fornito da

$$\hat{p} \pm z_{(1-\frac{\alpha}{2})}\frac{\sqrt{pq}}{\sqrt{n}}. \tag{6.21}$$

L'intervallo (6.21) non si può utilizzare perchè non si conosce il valore di p necessario per calcolare $\sqrt{p(1-p)}$. Si possono allora seguire due strade: ricavare un intervallo conservativo, oppure sostituire $\sqrt{p(1-p)}$ con $\sqrt{\hat{p}(1-\hat{p})}$.
(Intervallo di confidenza conservativo)

Si sostituisce $p(1-p)$ con il valore massimo che lo stesso può assumere. La funzione $p(1-p)$ assume l'andamento indicato in Fig.1.12. Il valore massimo di $p(1-p)$ si ha per $p = 0,5$. Inoltre

$$\max p(1-p) = 0,25$$

$$\max \sqrt{p(1-p)} = 0,50.$$

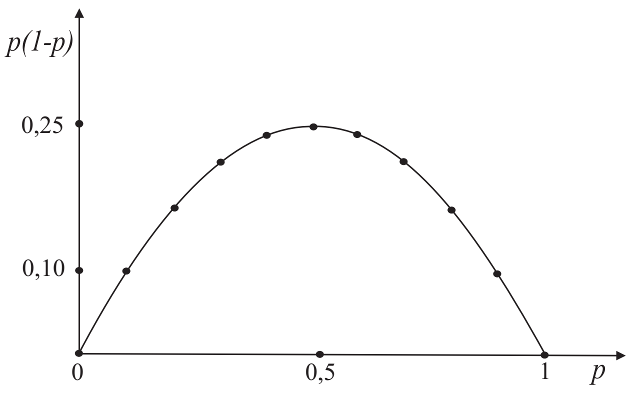


Figura 1.12: *Grafico della funzione* $p(1-p)$.

Con la sopra indicata sostituzione si ha l'intervallo

$$\hat{p} \pm z_{(1-\frac{\alpha}{2})}\frac{0,50}{\sqrt{n}}. \tag{6.22}$$

L'intervallo (6.22) contiene l'intervallo (6.21) in quanto

$$\hat{p} - z_{\left(1-\frac{\alpha}{2}\right)}\frac{0,50}{\sqrt{n}} \leq \hat{p} - z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sqrt{pq}}{\sqrt{n}}$$

ed in quanto

$$\hat{p} + z_{\left(1-\frac{\alpha}{2}\right)}\frac{0,50}{\sqrt{n}} \geq \hat{p} + z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sqrt{pq}}{\sqrt{n}}.$$

In altre parole i due intervalli sono centrati entrambi in $\hat{p}$, ma quello conservativo ha una maggiore ampiezza. Conseguentemente la sua probabilità di contenere p al suo interno è $\geq (1-\alpha)$.

Esempio 6.5.1 *Si vuole ricavare un intervallo di confidenza con* $(1-\alpha) = 0,95$ *dell'ignota frequenza relativa* p *che la caratteristica* A *ha in una popolazione. Si estrae un campione con riposizione di numerosità* $n = 40$ *e si è ottenuto* $x = 25$.

Si ha così $\hat{p} = \dfrac{25}{40} = 0,625$. L'intervallo "conservativo" per p risulta $0,625 \pm 1,96\frac{0,50}{\sqrt{40}} \rightarrow 0,625 \pm 1,96\frac{0,50}{6,3245} \rightarrow 0,625 \pm 0,155 \rightarrow (0,47; 0,78)$. Il risultato ottenuto ci permette di "confidare" che l'ignota frequenza relativa sia compresa fra il 47% ed il 78%.

(Intervallo di confidenza ottenuto sostituendo $p(1-p)$ con $\hat{p}(1-\hat{p})$)

Per poter sostituire $p(1-p)$ con $\hat{p}(1-\hat{p})$ bisogna che la numerosità campionaria sia alquanto elevata.

Esempio 6.5.2 *Ricavare l'intervallo di confidenza con* $(1-\alpha) = 0,954$ *di una ignota frequenza relativa sulla base di un campione con riposizione di numerosità* $n = 400$ *nel quale si è riscontrato* $x = 80$.

Per $(1-\alpha) = 0,954$ si ha $z_{\left(1-\frac{\alpha}{2}\right)} = 2$. La stima puntuale è pari a $\hat{p} = \dfrac{80}{400} = 0,20$. Pertanto $\sqrt{\hat{p}(1-\hat{p})} = \sqrt{0,2 \cdot 0,8} = 0,40$. L'intervallo di confidenza risulta così:
$0,20 \pm 2\frac{0,40}{\sqrt{400}} \rightarrow 0,20 \pm 2\frac{0,4}{20} \rightarrow 0,20 \pm 0,04 \rightarrow (0,16; 0,24)$.
Se si fosse proceduto con l'intervallo conservativo si sarebbero ottenuti i seguenti estremi (dell'intervallo): $0,20 \pm 2\frac{0,50}{20} \rightarrow 0,20 \pm 0,05 \rightarrow (0,15; 0,25)$.

6.6 Determinazione della numerosità campionaria per la stima della media μ

Sino ad ora la numerosità n del campione è stata pre-assegnata. E' arrivato il momento di esaminare come viene determinato il valore di n. In questo paragrafo si esaminerà come determinare n nel caso della stima di una aspettativa μ.

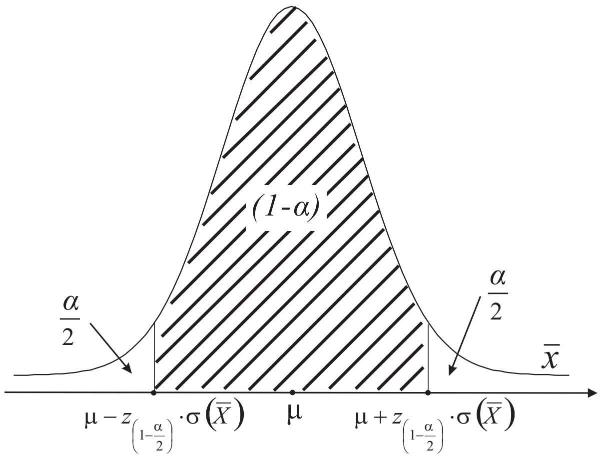
——— · · ——— · · ———

Per n grande la v.c. media campionaria $\bar{X}$ ha una distribuzione che si approssima con quella di una v.c. normale avente la stessa aspettativa e lo stesso scarto quadratico medio di $\bar{X}$ come è noto:

$$\begin{cases} E(\bar{X}) = \mu & \text{(media aritmetica della popolazione)}, \\ \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}} & \text{essendo } \sigma \text{ lo scarto quadratico medio della popolazione}. \end{cases}$$

Pertanto, per le note proprietà della normale

$$P\left\{\mu - z_{\left(1-\frac{\alpha}{2}\right)}\sigma(\bar{X}) \leq \bar{X} \leq \mu + z_{\left(1-\frac{\alpha}{2}\right)}\sigma(\bar{X})\right\} = (1-\alpha).$$



I valori di $\bar{X}$ compresi nell'intervallo sopra indicato hanno uno scarto in modulo da $\bar{X}$ che è minore o uguale a $z_{\left(1-\frac{\alpha}{2}\right)}\sigma(\bar{X})$. Pertanto, la relazione probabilistica precedente può anche scriversi così

$$P\left\{|\bar{X} - \mu| \leq z_{\left(1-\frac{\alpha}{2}\right)}\sigma(\bar{X})\right\} = (1-\alpha). \tag{6.23}$$

Quest'ultima relazione può essere utilizzata per ricavare la numerosità campionaria.

Si può procedere come segue.

i) Il ricercatore fissa lo scarto massimo "tollerato" fra la media campionaria $\bar{X}$ e l'ignota media μ. In altre parole, indicando questo scarto con d, si vuole che:

$$|\bar{X} - \mu| \leq d.$$

ii) Dato che si opera con campioni casuali non si può pretendere che la condizione $|\bar{X} - \mu| \leq d$ sia soddisfatta per ogni campione. In effetti, il ricercatore pretende che la stessa diseguaglianza si realizzi con probabilità

$(1-\alpha)$. Con ciò significando che vi è un rischio probabilistico pari ad α che si abbia $|\bar{X}-\mu| > d$.

iii) I due requisiti precisati in $i)$ e $ii)$ sono compendiati nella relazione

$$P\{|\bar{X}-\mu| \le d\} = (1-\alpha). \tag{6.24}$$

Tenuto conto della (6.22) si desume che per soddisfare la (6.24) bisogna che

$$z_{\left(1-\frac{\alpha}{2}\right)}\sigma(\bar{X}) = d. \tag{6.25}$$

Ricordando che $\sigma(\bar{X}) = \dfrac{\sigma}{\sqrt{n}}$ la (6.25) diventa
$z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}} = d \to z_{\left(1-\frac{\alpha}{2}\right)}\sigma = \sqrt{n}d \to \sqrt{n} = \frac{z_{\left(1-\frac{\alpha}{2}\right)}\sigma}{d}$.
In definitiva, passando al quadrato, si ha

$$n = \frac{z^2_{\left(1-\frac{\alpha}{2}\right)}\sigma^2}{d^2}. \tag{6.26}$$

La (6.26) fornisce la numerosità campionaria necessaria, per garantire con probabilità $(1-\alpha)$, che vi sia uno scarto in modulo fra stimatore $\bar{X}$ e aspettativa μ minore o uguale a d.

Dalla (6.26) si desume che n aumenta:

i) al crescere della varianza σ^2 del carattere X della popolazione;

ii) al crescere della probabilità $(1-\alpha)$, o meglio al crescere di $z^2_{\left(1-\frac{\alpha}{2}\right)}$;

iii) al diminuire del valore d dello scarto tollerato.

Esempio 6.6.1 *Si sa che nella popolazione la varianza σ^2 del carattere X è uguale a 200. Determinare la numerosità n del campione (con riposizione) che garantisce*

$$P\{|\bar{X}-\mu| \le 4\} = 0,954.$$

Per ricavare n si deve applicare la (6.26). Da $(1-\alpha) = 0,954$ deriva $z_{\left(1-\frac{\alpha}{2}\right)} = 2$. In definitiva

$$n = 2^2\frac{200}{4^2} = 50.$$

Dato che $n \ge 30$ si può ritenere che effettivamente la distribuzione di $\bar{X}$ si approssimi bene – per il teorema del limite centrale – ad una distribuzione normale.

Esempio 6.6.2 *Si ipotizzi che nella popolazione la varianza* σ^2 *del carattere* X *sia pari a* 100*. Si vuole* $d = 2$ *e* $(1-\alpha) = 0,94$*. Ricavare la numerosità* n *del campione che garantisce*

$$P\{|\bar{X} - \mu| \leq 2\} = 0,94.$$

Da $(1-\alpha) = 0,94$ si ottiene $\alpha = 0,06$ ed $\dfrac{\alpha}{2} = 0,03$. Dalle tavole di $\Phi(z)$ si trova $z_{(1-0,03)} = z_{(0,97)} = 1,881$. Applicando la (6.26) si ottiene

$$n = (1,881)^2 \frac{100}{2^2} = 89.$$

Per poter applicare la (6.26)

$$n = \frac{z^2_{\left(1-\frac{\alpha}{2}\right)} \sigma^2}{d^2}$$

bisogna conoscere il valore di σ^2.

Ecco come bisogna procedere se non si conosce il valore di σ^2.

a) Si può far riferimento – se possibile – ad esperienze analoghe in grado di fornire un valore σ^{*2} in sostituzione dell'ignoto valore σ^2.

b) Si realizza preliminarmente un campione pilota di numerosità n_1. Con questo campione si stima σ^2 con la varianza campionaria corretta

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2, \tag{6.27}$$

essendo $\bar{x}_1$ la media campionaria di questo campione pilota.

Quindi si determina n con la formula

$$n = \frac{z^2_{\left(1-\frac{\alpha}{2}\right)} s_1^2}{d^2}. \tag{6.28}$$

Si ricava la numerosità campionaria aggiuntiva $n_2 = (n - n_1)$ e si estraggono conseguentemente altre n_2 unità dalla popolazione.

Ovviamente il campione totale è fornito dalle unità del primo campione di numerosità n_1 e dalle unità statistiche del secondo campione di numerosità n_2. La numerosità totale è $n_1 + n_2 = n$.

———— . . ———— . . ————

Ecco ora una procedura per la determinazione di n utile nel caso in cui il carattere X assume valori non negativi ed abbia $\mu > 0$.

i) Il ricercatore fissa lo scarto *relativo* massimo “tollerato” fra la media campionaria $\bar{X}$ e μ. Lo scarto relativo è fornito da $\dfrac{\bar{X}-\mu}{\mu}$. Se $\dfrac{\bar{X}-\mu}{\mu}=0,05$ significa che lo scarto fra $\bar{X}$ e μ è uguale al 5% del valore di μ. In definitiva, indicando il valore dello scarto relativo con $\delta>0$ si fissa la condizione che
$$\frac{|\bar{X}-\mu|}{\mu}\leq\delta\rightarrow|\bar{X}-\mu|\leq\delta\mu.$$
Ad esempio se $\delta=0,10$ significa che si tollera che lo scarto in modulo $|\bar{X}-\mu|$ sia inferiore o uguale a $0,10\mu$ (al 10% del valore di μ).

ii) Il ricercatore stabilisce che la condizione precisata in $i)$ abbia probabilità di verificarsi pari ad $(1-\alpha)$.

iii) I due requisiti precedenti sono compendiati nella relazione
$$P\{\frac{|\bar{X}-\mu|}{\mu}\leq\delta\}=(1-\alpha),$$
ovvero
$$P\{|\bar{X}-\mu|\leq\delta\mu\}=(1-\alpha). \tag{6.29}$$

Ponendo $\delta\mu=d$ risulta evidente che la (6.29) coincide con la (6.24). Pertanto per trovare il valore di n che soddisfa la (6.29) basta applicare la (6.26)
$$n=\frac{z^2_{\left(1-\frac{\alpha}{2}\right)}\sigma^2}{d^2}$$
e sostituire d^2 con $\delta^2\mu^2$. Si ha così
$$\begin{aligned} n&=z^2_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma^2}{\delta^2\mu^2}\\ n&=z^2_{\left(1-\frac{\alpha}{2}\right)}\frac{1}{\delta^2}\left(\frac{\sigma}{\mu}\right)^2. \end{aligned} \tag{6.30}$$

La (6.30) differisce dalla (6.26) perchè fa intervenire il coefficiente di variazione $\dfrac{\sigma}{\mu}$ del carattere X che risulta più stabile nel tempo dello scarto quadratico medio. Ciò deriva dal fatto che il coefficiente di variazione CV è invariante alle trasformazioni di scala. Infatti se $Y=aX$ $(a>0)$, allora:

i) $\sigma(Y)=a\sigma(X)$, $M_1(Y)=aM_1(X)$

ii) $CV(Y)=\dfrac{\sigma(Y)}{M_1(Y)}=\dfrac{a\sigma(X)}{aM_1(X)}=\dfrac{\sigma(X)}{M_1(X)}=CV(X)$

Se X indica il reddito individuale che si riscontra in un anno e Y il reddito individuale dell'anno successivo e si suppone che i redditi individuali siano aumentati del 5% (si ipotizza cioè $Y = 1,05X$) allora il coefficiente di variazione non si è modificato. Conseguentemente si può impiegare il coefficiente di variazione di X per valutare la numerosità campionaria relativa ai redditi Y.

Esempio 6.6.3 *Da molte indagini sul reddito personale si è constatato che il coefficiente di variazione dello stesso sia pari a* $0,80$*. Ricavare la numerosità campionaria che garantisca con probabilità* $(1-\alpha) = 0,97$ *uno scarto relativo (in modulo) fra* $\bar{X}$ *e* μ *inferiore o uguale a* $\delta = 0,05$*.*

Per poter applicare la (6.30) è necessario innanzitutto ricavare il valore di $z_{\left(1-\frac{0,03}{2}\right)} = z_{(0,985)}$. Dalla tavola $\Phi(z)$ si ottiene $z_{(0,985)} = 2,17$. Pertanto

$$n = (2,17)^2 \frac{1}{(0,05)^2}(0,80)^2 = 1205.$$

6.7 Determinazione della numerosità campionaria per la stima della frequenza relativa p

Sia $\hat{p} = \dfrac{X}{n}$ la frequenza relativa campionaria. Come è noto: $E(\hat{p}) = p, Var(\hat{p}) = \dfrac{1}{n}pq$. Per n elevato $\hat{p}$ ha una distribuzione che si può approssimare con la corrispettiva v.c. normale. Pertanto

$$P\{|\hat{p}-p| \le z_{\left(1-\frac{\alpha}{2}\right)}\sigma(\hat{p})\} = (1-\alpha). \tag{6.31}$$

Quest'ultima relazione può essere impiegata per ricavare n seguendo la stessa procedura vista per μ. Si pone, pertanto, la condizione che con probabilità $(1-\alpha)$ si abbia $|\hat{p}-p| \le d$. Il valore di $(1-\alpha)$ ed il valore di d vengono fissati dal ricercatore. Si vuole che

$$P\{|\hat{p}-p| \le d\} = (1-\alpha). \tag{6.32}$$

Pertanto, tenuto conto della (6.31), la relazione richiesta implica che

$$z_{\left(1-\frac{\alpha}{2}\right)}\sigma(\hat{p}) = d.$$

Dato che $\sigma(\hat{p}) = \dfrac{\sqrt{pq}}{\sqrt{n}}$ la relazione precedente diventa

$z_{\left(1-\frac{\alpha}{2}\right)}\dfrac{\sqrt{pq}}{\sqrt{n}} = d \rightarrow \sqrt{n} = z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sqrt{pq}}{d} \rightarrow$

$$n = z^2_{\left(1-\frac{\alpha}{2}\right)}\frac{pq}{d^2}. \tag{6.33}$$

Per poter applicare la (6.33) bisognerebbe conoscere il valore di p (che è oggetto di stima). A parità di $z^2_{\left(1-\frac{\alpha}{2}\right)}$ e di d^2 il valore di n cresce al crescere di $p(1-p)$. Allora si può scegliere il valore massimo di $p(1-p)$che è uguale a $0,25$ e si realizza per $p = 0,5$. Si considera, quindi, ai fini del valore da assegnare a n, la condizione più sfavorevole. Ciò significa che per valori di $p \neq 0,5$ ci vorrebbe una numerosità inferiore per garantire il valore di $(1-\alpha)$ ed il valore di d. In conclusione, mettendosi nella condizione più sfavorevole, il valore di n è fornito da

$$n = z^2_{\left(1-\frac{\alpha}{2}\right)} \frac{0,25}{d^2}. \tag{6.34}$$

Esempio 6.7.1 *Si determini n in modo che: $d = 0,02$ e $(1-\alpha) = 0,954$.*

Come è noto per $(1-\alpha) = 0,954$ si ha $z_{\left(1-\frac{\alpha}{2}\right)} = 2$. Pertanto la numerosità risulta

$$n = 2^2 \frac{0,25}{(0,02)^2} = 2500. \tag{6.35}$$

La condizione $d = 0,02$ significa, ipotizzando ad esempio che $p = 0,4$, che si tollera uno scarto di $0,02$ fra $\hat{p}$ e p, ovvero $\hat{p}$ deve trovarsi nell'intervallo $0,38 \leq \hat{p} \leq 0,42$ con probabilità $0,954$.

————— . . ————— . . —————

Porsi nella condizione più sfavorevole è abbastanza ragionevole se l'ignota frequenza relativa p è compresa nell'intervallo $(0,3; 0,7)$. In effetti in questo intervallo il valore di $p(1-p)$ oscilla da $0,21$ a $0,25$ ed il conseguente valore di n secondo la (6.33) oscilla da 2100 (per $p = 0,3$ e $p = 0,7$) a 2500 per $p = 0,5$. Se invece l'ignoto p è situato verso gli estremi del suo campo di variazione $(0 < p < 1)$, il ricorso all'ipotesi più sfavorevole conduce a numerosità campionarie decisamente più elevate di quelle necessarie (se si conoscesse il valore di p). Se si ipotizza, ad esempio, $p = 0,25$ la (6.33) fornisce

$$n = 2^2 \left(\frac{1}{0,02}\right)^2 (0,15 \cdot 0,85) = 1275.$$

Si ha così un valore di n decisamente più piccolo di quello della ipotesi più sfavorevole.

————— . . ————— . . —————

Per ovviare a questo inconveniente si può effettuare un primo campionamento pilota di numerosità n_1 per avere informazioni su p e poi usare la stima $\hat{p}_1$, ottenuta con questo primo campione, per ottenere n con la (6.33) ove p viene sostituito con $\hat{p}_1$. Quindi si fa un secondo campionamento di numerosità $n_2 = n - n_1$.

Ecco il dettaglio di questa procedura

i) Si effettua un primo campionamento pilota di numerosità n_1. Si calcola $\hat{p}_1 = \dfrac{X_1}{n_1}$ essendo X_1 il numero di unità portatrici della caratteristica di interesse che si sono riscontrate nel primo campione.

ii) Si ricava n con la formula

$$n = z^2_{\left(1-\frac{\alpha}{2}\right)} \frac{\hat{p}_1(1-\hat{p}_1)}{d^2}.$$

iii) Si effettua un secondo campionamento di numerosità $n_2 = n - n_1$. Si calcola $\hat{p}_2 = \dfrac{X_2}{n_2}$, essendo X_2 il numero delle presenze della caratteristica di interesse che si riscontrano nel secondo campione.

iv) Lo stimatore di p è fornito da

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n}.$$

Esempio 6.7.2 *Ricavare n con il campionamento in due parti con $(1-\alpha) = 0,954$ e $d = 0,02$.*

i) Si può prendere $n_1 = 500$. Si può supporre che $X_1 = 100$ di modo che $p_1 = \frac{100}{500} = 0,2$.

ii) La numerosità campionaria totale diventa

$$n = 2^2 \frac{1}{(0,02)^2 0,2 \cdot 0,8} = 1600.$$

iii) La numerosità $n_2 = 1600 - 500 = 1100$. Si può ora supporre $X_2 = 300$ di modo che

$$\hat{p}_2 = \frac{300}{1100} = 0,2727.$$

iv) La stima di p con il campione totale diventa

$$\hat{p} = \frac{100 + 300}{1600} = \frac{400}{1600} = 0,25.$$