ELEMENTI DI STATISTICA [1]

Alberta Albertella [(1)] – Luigi Mussio [(2)]
[(1)] Politecnico di Milano – DICA – Piazza Leonardo da Vinci, 32 – 20133 Milano
Tel. 02-2399-6282 – Fax 02-2399-6530 – e-mail alberta.albertella@polimi.it
[(2)] Politecnico di Milano – DICA – Piazza Leonardo da Vinci, 32 – 20133 Milano
Tel. 02-2399-6501 – Fax 02-2399-6602 – e-mail luigi.mussio@polimi.it

RIASSUNTO

Le variabili statistiche, le variabili casuali (con gli elementi di calcolo delle probabilità, cui aggiungere i complementi di calcolo delle probabilità, oggigiorno ormai parti integranti di qualsiasi trattazione sulle variabili casuali stesse) e le distribuzioni campionarie costituiscono sicuramente gli elementi della statistica e le basi per qualsiasi studio di essa.

PARTE I – VARIABILI STATISTICHE

1.1 Introduzione

Il termine *statistica* risale al XVII secolo. Esso è usato, per la prima volta, da professori tedeschi di quelle discipline che oggi si chiamerebbero *scienze politiche*, per indicare una scienza che deve essere studiata dai futuri uomini di Stato, per fornire tutte le informazioni riguardanti lo Stato. In origine cioè la statistica non si basa sui dati numerici, ma solo su dati qualitativi. Tuttavia sono convogliate in essa nozioni fondamentali provenienti da altri campi di indagine che ne hanno, a poco a poco, trasformato la struttura. In tal senso, si fa riferimento principalmente:

- alla Dottrina della Sorte, antico nome del calcolo delle probabilità;
- all'Aritmetica Politica, sorta a Londra sotto l'influenza dell'empirismo inglese che, sviluppando il lato numerico della statistica, propone un metodo classificatorio dei fenomeni naturalistici relativi alle popolazioni (nascite, morti, censimenti, ecc.);
- alle discipline che per secoli si erano dedicate all'Arte del misurare, come l'Astronomia, la Geodesia e la Topografia.

Storicamente la statistica comincia con l'essere descrittiva. Infatti è necessario dapprima accumulare informazioni, criticarle, analizzarle, sintetizzarle, e solo in seguito, anche per l'apporto del calcolo delle probabilità, si constatano analogie e permanenze, e si individuano forme generali di legami strutturali che permettono di formulare teorie relative ai fenomeni osservati.

Gli stessi due momenti che caratterizzano lo sviluppo storico si ritrovano anche nelle applicazioni pratiche della statistica. Tuttavia accanto ai due importanti filoni della *statistica descrittiva* e dello *studio dei modelli probabilistici*, è presente oggi, in pieno sviluppo, tutta la problematica che concerne l'*inferenza statistica*, ossia quell'insieme di metodologie che permettono di decidere, in base ai dati ricavabili da campioni di ampiezza, anche limitata, sulla correttezza o meno di ipotesi precedentemente fatte sul comportamento dei fenomeni sotto esame.

[1] Questo lavoro riporta, pressoché integralmente, nello stile degli autori, quanto esposto nei capitoli 1, 2, 3 ed 8 del libro: Fondamenti di statistica, di Giovanna Togliatti (Hoepli, Milano, 1976), dove le note, scritte dagli autori del presente lavoro, servono a colmare i quasi quaranta anni passati dall'epoca di edizione del libro suddetto, così da rendere questi quattro capitoli ancora pienamente attuali.

Nel seguito, sono esposti elementi di statistica, essenzialmente finalizzati all'analisi di risultati di esperimenti di tipo tecnico e scientifico, ed al controllo delle caratteristiche della produzione di massa. Nonostante si possa notare una certa prevalenza di applicazioni tratte dall'ingegneria civile, il tipo di esposizione e la scelta degli esempi, sono impostati in modo da non privilegiare nessun tipo di utilizzatore a danno di altro, in modo tale che le metodologie esposte possano essere ugualmente accessibili a tutti.
Il livello delle conoscenze matematiche necessarie per la comprensione del seguito è molto modesto. Inoltre parecchi concetti, appena accennati all'inizio, sono poi ripresi e approfonditi. Questa impostazione offre il vantaggio di avvicinare in modo graduale, sia come linguaggio che come esempi, a concetti non facili, che richiedono maturazione e ripensamento.

1.2 La varabile statistica ad una dimensione

Data una popolazione composta da N individui, di qualsiasi tipo essi siano e qualunque sia il processo generativo comune che li ha formati, la statistica insegna come classificare questa popolazione in relazione ad un unico suo aspetto, quello che interessa l'indagine, e come descriverla o rappresentarla sinteticamente, in tutte le sue peculiarità somatiche, attraverso una serie di indici significativi.
E' importante innanzitutto, prima di qualsiasi indagine statistica, delimitare con precisione la popolazione che si vuole esaminare. Questo si ottiene mediante la definizione di una caratteristica comune, in maniera indifferenziata, a tutti i componenti della popolazione, tale che non ammetta ambiguità o indeterminatezza nel tracciare la separazione fra gli individui interni alla popolazione e quelli esterni.
Ad esempio, la popolazione dei cittadini italiani, nati in un certo giorno, non è ancora una popolazione ben definita, fino a quando non si è precisato che si includono coloro che, pur nati all'estero, in quel giorno, hanno diritto alla cittadinanza italiana e si escludono invece quelli nati sul territorio nazionale a cui non compete, alla nascita, la cittadinanza italiana. Altra popolazione completamente definita potrebbe essere quella delle autovetture targate MI, oppure quella dei risultati delle misure di grandezze ottenute con un certo strumento e con certi criteri operativi.
Una volta definita la popolazione si può esaminare la struttura, sotto il punto di vista di altre caratteristiche che non si presentano in maniera uniforme per tutti gli individui. Anzi gli individui nella popolazione si distinguono proprio in base alle differenze tra alcune delle caratteristiche possedute da ciascuno di essi. Si chiama attributo /argomento di una popolazione una particolare caratteristica degli individui della stessa che si manifesta in essi con forme differenti. Esempi di attributi /argomenti delle popolazioni sopra citate sono:

- per la popolazione dei cittadini italiani, nati in un certo giorno, il Comune di appartenenza, il numero i figli già presenti nel nucleo familiare in cui essi si inseriscono, il peso, la statura, l'esistenza o meno di una certa malformazione, ecc.
- per la popolazione delle autovetture targate MI, l'anno di fabbricazione, la cilindrata, la marca, ecc.
- per la popolazione dei risultati delle misure di una grandezza l'attributo è unico: il valore numerico delle misure ottenute.

Studiare una popolazione dal punto di vista di un suo attributo /argomento significa esaminare come si distribuiscono le varie forme dell'attributo /argomento sugli individui della popolazione, cioè classificare gli individui secondo le varie forme dell'attributo /argomento.

Per poter fare questo, occorre che l'attributo/argomento X, in relazione al quale si vuole effettuare l'organizzazione della popolazione, possegga i seguenti requisiti:

- può assumere forme distinte x_i tra loro incompatibili, ovvero ogni individuo deve possedere solo una forma dell'attributo;
- ogni forma x_i rappresenta un concetto di classe, ovvero deve essere possibile che, almeno in linea di principio, parecchi individui posseggano la stessa forma dell'attributo;
- in ogni individuo della popolazione è presente una forma x_i, ovvero l'attributo deve essere presente, pur sotto diverse forme, in ciascun individuo della popolazione;
- esistono almeno due individui, nella popolazione, in possesso di forme diverse dell'attributo, ovvero la popolazione stessa non si presenta totalmente omogenea.

Classificare la popolazione C, secondo l'attributo/argomento X, significa individuare la forma che spetta a ciascun individuo di questa popolazione. Pertanto preso in considerazione un individuo si deve sempre poter decidere quale forma dell'attributo/argomento è presente nell'individuo. La successione delle diverse forme x_i non sempre può rappresentarsi con numeri in maniera immediata; tuttavia questa possibilità esiste sempre nei casi, relativi alle misure, che interessano principalmente questa trattazione.
Inoltre l'attributo/argomento X può assumere:

- un numero finito di forme e, in tal caso, è detto discreto;
- un numero infinito di forme, tutte contenute con continuità in un intervallo limitato: $a-b$, ed allora è detto di tipo continuo.

Tuttavia nella pratica, si ha sempre a che fare con attributi/argomenti di tipo discreto, poiché ogni risultato di misurazioni, per sensibile che sia lo strumento con cui esse si effettuano, è sempre una serie discreta, anche se numerosa, di valori numerici che differiscono tra loro di quantità finite. Di conseguenza, solo il caso di attributi/argomenti discreti è preso in considerazione, cosicché la successione dei valori argomentali di X può rappresentarsi con una successione di x_i $(i=1,2,...,n)$.

Per mezzo dell'operazione di classificazione, secondo l'attributo/argomento X, degli N individui che formano, nel loro insieme, la popolazione C, si sono raggruppati assieme quegli individui che posseggono un'uguale forma dell'attributo/argomento. In questo modo, si sono contati gli individui in ciascun gruppo, ottenendo i valori delle *frequenze assolute*: $f_1, f_2, ..., f_n$. Esse rappresentano il numero degli individui compresi nelle classi definite dai valori argomentali corrispondenti: $x_1, x_2, ... x_n$. Con queste due serie in corrispondenza biunivoca si costruisce l'espressione:

$$X\begin{cases} x_1 & x_2 & x_3 & ... & x_n \\ f_1 & f_2 & f_3 & ... & f_n \end{cases} \qquad \text{tale che:} \qquad \sum_{i=1}^{n} f_i = N \tag{2.1}$$

nella quale la prima riga rappresenta la successione dei valori dell'attributo/argomento, mentre nella seconda riga sono riportate le frequenze assolute f_i, cioè i numeri di individui della popolazione che posseggono l'attributo/argomento X, rispettivamente sotto le rispettive forme: $x_1, x_2, ..., x_n$. L'espressione (2.1) può scriversi anche sotto la forma:

$$X \begin{cases} x_1 & x_2 & \dots & x_n \\ \dfrac{f_1}{N} & \dfrac{f_2}{N} & \dots & \dfrac{f_n}{N} \end{cases} \qquad \text{tale che:} \qquad \sum_{i=1}^{n} \frac{f_i}{N} = 1 \qquad (2.2)$$

che differisce dalla precedente, perché in essa, alla successione delle frequenze assolute f_i, è stata sostituita la successione delle *frequenze relative* f_i / N. Le uguaglianze:

$$\sum_{i=1}^{n} f_i = N \qquad \text{e} \qquad \sum_{i=1}^{n} \frac{f_i}{N} = 1$$

non sono inutili aggiunte, ma servono ad esprimere la garanzia che la popolazione è analizzata per intero e che tutti i suoi individui sono sinteticamente rappresentati in X. Le frequenze, assolute e relative, sono additive, infatti se si considerano indifferenziati due valori argomentali: $x_i = x_j = x_0$, la frequenza assoluta di x_0 è: $f_0 = f_i + f_j$, e quella relativa $f_0/N = f_i/N + f_j/N$. La variabile X, costruita come specificato e rappresentata due espressioni (2.1) e (2.2) [2], è detta *variabile statistica semplice*, cioè *ad una dimensione*.

Esempio 1.2.1

Popolazione: gli allievi che frequentano un determinato corso. Argomento/attributo sotto cui è analizzata la popolazione: la statura. Forme dell'attributo: le misure delle stature espresse in centimetri. Dall'analisi dei singoli individui è risultata la seguente variabile statistica semplice (ossia ad una dimensione).

$$X \begin{cases} 164 & 165 & 166 & 167 & 168 & 169 & 170 & 171 & 172 & 173 & 174 & 175 & 176 \\ 2 & 1 & 0 & 0 & 2 & 2 & 10 & 4 & 6 & 2 & 4 & 7 & 1 \end{cases}$$

$$\begin{cases} 177 & 178 & 179 & 180 & 181 & 182 & 183 & 184 & 185 & 186 & 187 & 188 & 189 \\ 3 & 4 & 6 & 11 & 4 & 1 & 0 & 6 & 5 & 3 & 1 & 0 & 1 \end{cases} \qquad \sum_{i=1}^{26} f_i = 86$$

In alcuni casi, la classificazione di una popolazione, rispetto ad un determinato valore argomentale, può essere fatta usando non più le singole forme dell'attributo in questione, ma le *classi* di opportuna ampiezza. Ad esempio, si può determinare il numero di individui il cui valore argomentale è compreso fra x_i e x_j, estremi inclusi, ed associare alla classe: $x_i - x_j$, questo numero che è chiamato frequenza della classe.

[2] Coloro che scrivono sono soliti non numerare le formule ed altre espressioni, richiamandole invece dettagliatamente nel testo. Invece in questo caso, per rimanere più aderenti all'originale del libro, si è scelto di mantenere la numerazione, ivi presente.

Inoltre ogni classe è contraddistinta da un valore medio o punto medio della classe: $\left(x_i + x_j\right)/2$, e da due limiti di classe, generalmente situati a metà fra: x_{i-1} e x_i, e fra x_j e x_{j+1}, che non coincidono così con i dati osservati.

<u>Esempio 1.2.2</u>
Si voglia costruire la distribuzione di frequenze relativa ai pesi di 60 studenti maschi, misurati di chilo in chilo. I valori ottenuti sono i seguenti:

63	74	68	75	60	65	62	57	68	79	71	66	72	72	64	58	67	62	69	67
69	71	65	76	57	74	63	80	63	74	54	69	70	75	72	66	78	60	64	67
63	61	69	64	74	61	73	66	82	61	64	58	68	71	60	66	58	70	84	68

Il peso maggiore è di $84\,kg$, quello minore è di $54\,kg$ e l'intervallo fra questi è di: $84 - 54 = 30\,kg$. Una scelta conveniente dell'ampiezza della classe è di $2\,kg$ e, di conseguenza, i punti medi delle classi sono 54.5, 55.5, $ecc.$, mentre gli intervalli di classe sono: $54 - 55$, $56 - 57$, $58 - 59$, $ecc.$
Con questa scelta, i limiti delle classi sono: 55.5, 57.5, 59.5, $ecc.$, e la variabile statistica risulta ordinata, come nella prima delle seguenti tabelle. Ovviamente si può stabilire un'altra distribuzione di frequenze, scegliendo una diversa ampiezza delle classi, ad esempio, di $4\,kg$; in tal caso, si ottiene la seconda delle stesse tabelle.

Peso (kg)	*Frequenza*
54 – 55	1
56 – 57	2
58 – 59	3
60 – 61	6
62 – 63	6
64 – 65	6
66 – 67	7
68 – 69	8
70 – 71	5
72 – 73	4
74 – 75	6
76 – 77	1
78 – 79	2
80 – 81	1
82 – 83	1
84 – 85	1
Totale	60

Peso (kg)	*Frequenza*
54 – 57	3
58 – 61	9
62 – 65	12
66 – 69	15
70 – 73	9
74 – 77	7
78 – 81	3
82 – 85	2
Totale	60

1.3 Rappresentazione grafica di una variabile statistica semplice

Della variabile X si possono dare alcuni tipi di rappresentazioni grafiche. Si riportino, in scala opportuna, su una semiretta, i valori numerici dell'attributo/argomento X, dopo averli disposti in ordine crescente, in cosicché risultino distribuiti sul segmento $x_1 - x_n$. In corrispondenza a ciascun valore x_i si riporti, in direzione normale alla semiretta, un segmento di lunghezza proporzionale alla frequenza totale f_i di quel valore argomentale. Si avrà così un *diagramma di frequenza* (in Fig. 1.3.1, è rappresentato il diagramma di frequenza relativo al precedente esempio 1.2.1).

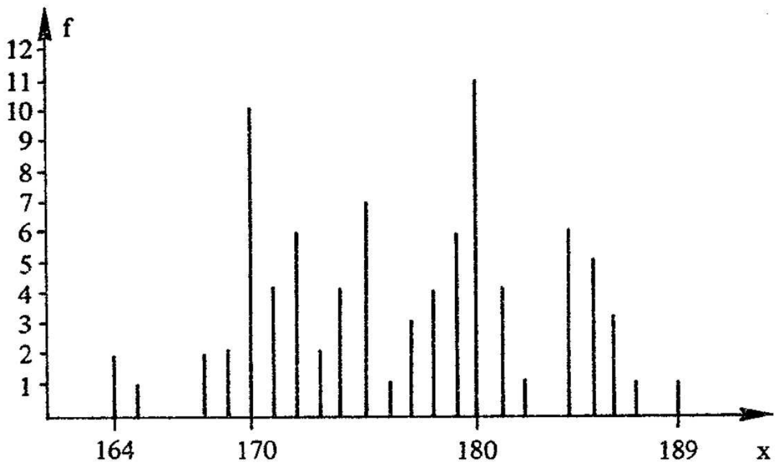


Fig. 1.3.1 – Diagramma di frequenza relativo all'Esempio 1.2.1

Tuttavia quando la serie dei valori argomentali è molto numerosa e compresa in un intervallo $x_1 - x_n$ assai limitato, può essere conveniente un'altra rappresentazione grafica della variabile statistica. La serie dei valori argomentali, riportata graficamente sulla semiretta, è suddivisa in tanti intervalli o classi, ciascuno contenente un certo gruppo di valori argomentali.

Dati i punti di separazione: x'_1, x'_2, ..., x'_m, o limiti delle classi che, per evitare ambiguità, è meglio non coincidano con i valori argomentali propri della variabile statistica, si costruisca, sopra ogni intervallo costituito dal segmento: $x'_i - x'_{i+1}$, un rettangolo di area proporzionale al numero degli individui della popolazione in possesso di valore argomentale compreso entro i limiti: x'_i e x'_{i+1}. La figura risultante si chiama *istogramma.* Talvolta l'ampiezza delle classi coincide con l'unità di misura delle x ed è mantenuta costante. Tuttavia in generale, questo non accade ed anzi spesso è opportuno modificare l'ampiezza delle classi in quella zona in cui le frequenze diminuiscono sensibilmente.

Pertanto le altezza dei singoli rettangoli sono calcolate in modo da rispettare la definizione data di istogramma. Indicando con m_0 l'unità di misura delle x (che può anche essere l'intervallo di classe più frequentemente utilizzato) e con: $k_i = m_i / m_0$, il numero di unità di misura contenuto nella classe di ampiezza m_i, l'ordinata y_i da assegnare al rettangolo di base questa classe, risulta:

$$y_i = f_i / k_i$$

dove f_i rappresenta la frequenza assoluta degli individui presenti nella classe di ampiezza m_i (e l'area totale, sottesa dall'istogramma, è misurata dalla relazione: $\sum f_i = N$). Molto spesso, con un cambiamento di unità nella scala delle ordinate, le altezze dei rettangoli sono rappresentate con i valori:

$$y'_i = \frac{f_i / N}{k_i}$$

e di conseguenza, l'area totale, sottesa dall'istogramma, diventa un'area *normalizzata*, cioè uguale ad 1.

<u>Esempio 1.3.1</u>

I seguenti dati rappresentano la distribuzione dei comuni italiani secondo classi di popolazione al 1° gennaio 1960 e l'istogramma corrispondente ha unità di misura: $m_0 = 1000$ abitanti (essendo: $N = \sum f_i = 7698$).

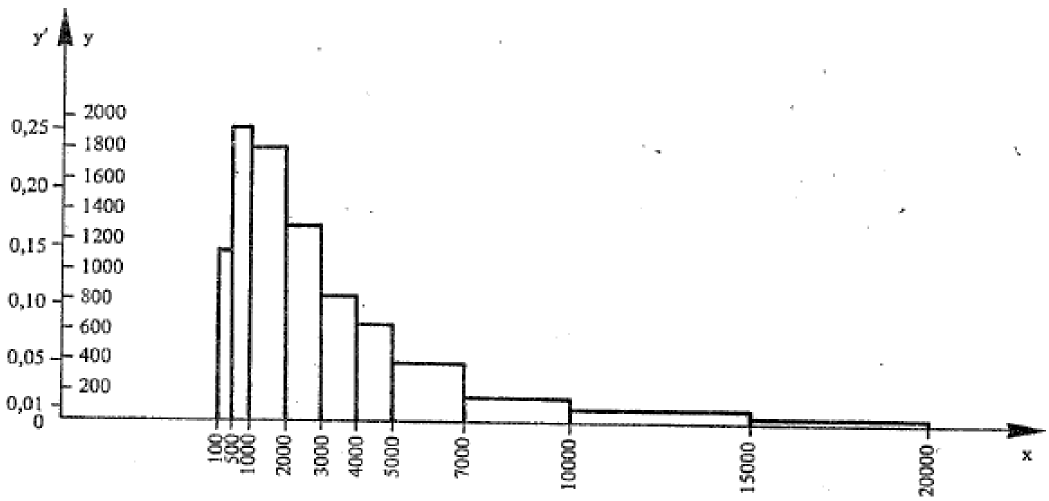


Fig. 1.3.2 – Istogramma relativo alla distribuzione dei comuni italiani secondo il numero degli abitanti

Classi di popolazione	Ampiezza delle classi m_i	Numero dei comuni f_i	k_i	$y_i = \frac{f_i}{k_i}$	$y'_i = \frac{f_i/N}{k_i}$
$100 - 500$	400	449	0.4	1.123	0.146
$501 - 1.000$	500	969	0.5	1.938	0.252
$1.001 - 2.000$	1.000	1.793	1	1.793	0.233
$2.001 - 3.000$	1.000	1.280	1	1.280	0.166
$3.001 - 4.000$	1.000	824	1	824	0.107
$4.001 - 5.000$	1.000	630	1	630	0.082
$5.001 - 7.000$	2.000	765	2	382	0.050
$7.001 - 10.000$	3.000	472	3	156	0.020
$10.001 - 15.000$	5.000	369	5	74	0.010
$15.001 - 20.000$	5.000	147	5	29	0.004

Nella costruzione degli istogrammi, ha notevole importanza la scelta dell'ampiezza dell'intervallo di ciascuna classe. Infatti da questa scelta, a volte in modo assai macroscopico, dipende la formazione dell'istogramma. L'ampiezza dell'intervallo deve essere scelta, pur con qualche arbitrarietà, tenendo conto di questi elementi:

- estensione del campo di variabilità dell'attributo/argomento della variabile;
- numero complessivo di individui componenti la popolazione;
- grado di attendibilità delle rilevazioni dell'attributo/argomento stesso.

Nel caso che il valore dell'attributo/argomento, proprio di ciascun individuo, è ottenuto mediante misure, è inutile ripartire l'istogramma in classi il cui intervallo è inferiore alla sensibilità della misura stessa.
La rappresentazione mediante istogramma, con intervallo di ampiezza uniforme (della variabile statistica già ordinata in classi dell'Esempio 1.2.2), mostra chiaramente come, al variare della ampiezza dell'intervallo, cambi sensibilmente la forma dell'istogramma corrispondente (a riguardo, si veda la Fig. 1.3.3).

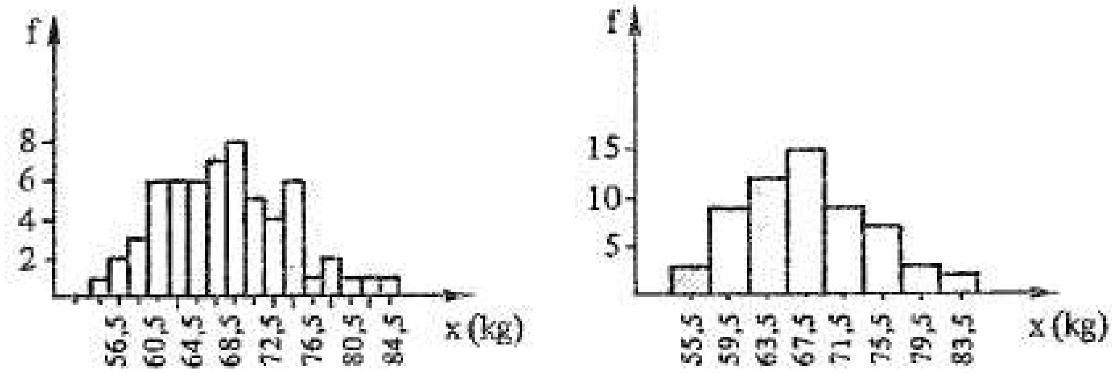


Fig. 1.3.3 – Istogrammi relativi alla stessa popolazione, con diverso intervallo di classe

Il problema della scelta dell'intervallo (in cui suddividere la scala dei valori argomentali per la costruzione dell'istogramma) perde la sua importanza, qualora si passi dalla rappresentazione della variabile statistica a quella della variabile statistica cumulata.
Si chiama *funzione cumulativa di frequenza o funzione di distribuzione* la seguente doppia successione in corrispondenza biunivoca:

$$X\begin{cases} x_1 & x_2 & \dots & x_n \\ F_1 = \dfrac{f_1}{N} & F_2 = \dfrac{f_1 + f_2}{N} & \dots & F_n = \dfrac{f_1 + f_2 + f_3 \dots + f_n}{N} \end{cases}$$

cioè quella nella quale in corrispondenza di ogni valore x_i dell'argomento, è posta la somma delle frequenze di tutti gli elementi che hanno valori inferiori od uguali a x_i.
Si chiama *diagramma cumulativo di frequenza o diagramma di distribuzione* la rappresentazione grafica della funzione cumulativa di frequenza che si costruisce facendo corrispondere, in un diagramma cartesiano, all'ascissa x_i, l'ordinata proporzionale a:

$$F_i = \frac{f_1 + f_2 + ... + f_i}{N} = \sum_{r=1}^{i} \frac{f_r}{N} \qquad \text{essendo, per definizione:} \qquad F_n = \sum_{r=1}^{n} \frac{f_r}{N} = 1$$

Frequentemente la variabile statistica cumulata è rappresentata mediante istogramma, con intervalli di ampiezza uniforme in cui (come mostra la Fig. 1.4.1) sulla classe i-esima è costruito un rettangolo di area proporzionale a F_i.

1.4 Moda, mediana, quantili

Nelle serie dei valori argomentali, alcuni sono di particolare importanza, per la descrizione della variabile statistica cui si riferiscono. Quando questa variabile è rappresentata nella forma (2.1), si possono definire alcune caratteristiche [3] della variabile stessa.

- La *moda* è quel valore argomentale cui corrisponde la massima frequenza, rispetto ai valori argomentali che lo precedono e lo seguono. Il valore modale non può mai occupare il primo o l'ultimo posto nella successione dei valori argomentali x_i e, qualora il valore di massima frequenza coincida con il primo o l'ultimo termine, la distribuzione si dice zero-modale. Una distribuzione è unimodale o plurimodale a seconda che esista una sola frequenza che supera tutte le altre, oppure che esistano, nella serie delle frequenze, altri massimi relativi.
- La *mediana* è quel valore argomentale la cui frequenza f_m soddisfa le seguenti condizioni. Avendo disposto in ordine crescente i valori argomentali: $x_1, x_2, ... x_n$, cui corrispondono le frequenze $f_1, f_2, ..., f_n$, si ha che:

  - la somma delle frequenze che precedono f_m è minore della metà della frequenza totale;
  - la somma di f_m e delle frequenze che la precedono è uguale o maggiore della metà della frequenza totale.

  Qualora la variabile statistica sia rappresentata con un istogramma, la mediana è l'ascissa del punto la cui ordinata divide in due parti uguali l'area coperta dall'istogramma.
- I *quantili* sono quei valori argomentali per cui sono soddisfatte le condizioni, già poste per la mediana, sostituendo solo un quarto o tre quarti, al posto di metà, nella loro definizione [4].

Quando invece la serie dei valori argomentali è particolarmente numerosa e la variabile statistica è ordinata per classi, dal grafico, si può individuare la classe modale. Tuttavia in questi casi, è più comodo considerare non già l'istogramma rettangolare, ma un nuovo diagramma ottenuto interpolandolo, nel senso intuitivo e

[3] Tutte le caratteristiche (di questo paragrafo) sono utili anche per lo studio di dati qualitativi, con la sola condizione che siano ordinabili.

[4] I quantili possono essere altresì definiti a partire dalla mediana, prendendo in considerazione gli scarti in valore assoluto, rispetto ad essa. In questo caso, il secondo quantile segna ancora la metà dei dati (proprio con le stesse regole di definizione della mediana); esso raggruppa la parte centrale dei dati e lascia la parte restante nelle code. Come la mediana è una caratteristica robusta, in quanto poco sensibile alla presenza di dati anomali e, in particolare, di errori grossolani; il suo nome è derivato dalla letteratura specialistica inglese/americana: *mean absolute value* (*mav*) e potrebbe essere tradotto, con il termine, peraltro poco usato: *valore assoluto mediano*.

corrente della parola. Questa curva, detta *curva di frequenza*, è tracciata in modo tale che l'area totale, racchiusa fra di essa e l'asse x, sia uguale all'area totale dei rettangoli dell'istogramma (come mostra la Fig. 1.4.1).

La ragione fondamentale, per la sostituzione della curva di frequenza all'istogramma, sta nel fatto che il contorno rettangolare equivale all'ipotesi di una distribuzione uniforme degli elementi delle singole classi, mentre è più attendibile che la distribuzione delle frequenze dei valori argomentali, interni ad ogni intervallo, si uniformi all'andamento generale dell'istogramma.

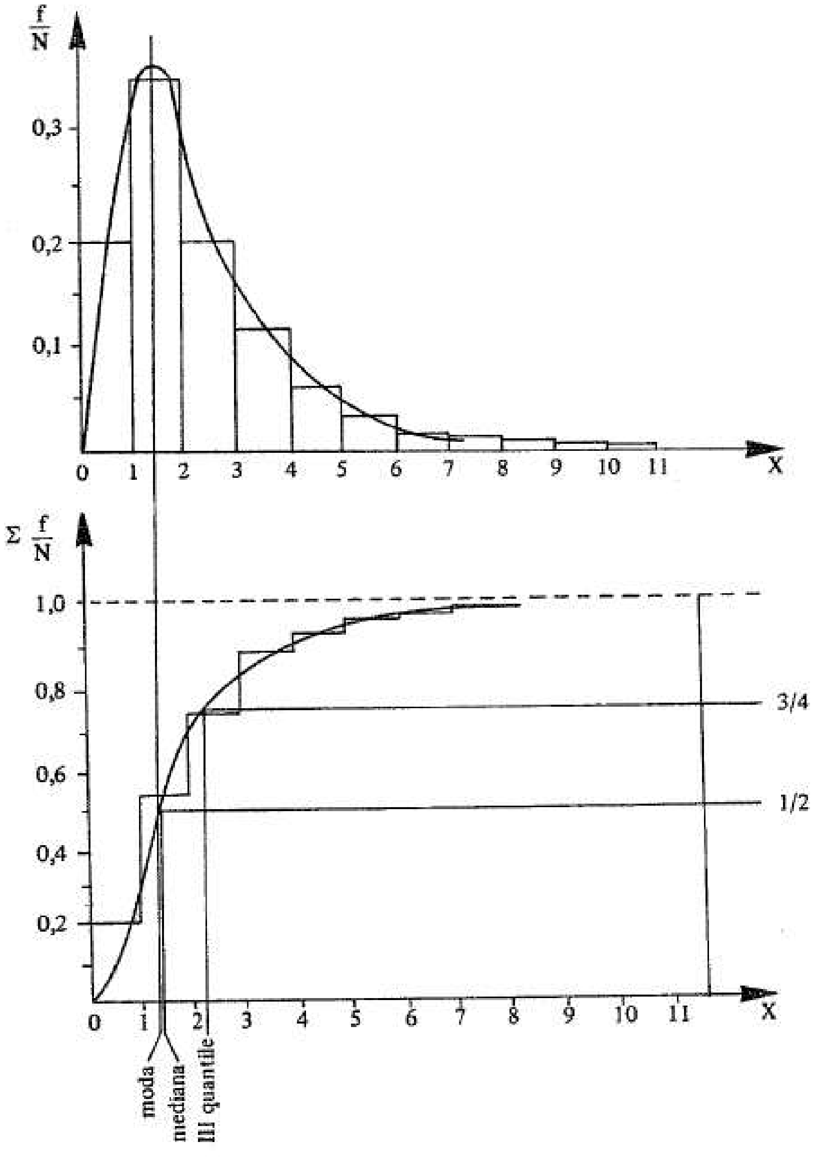


Fig. 1.4.1 – Determinazione di moda, mediana e quantili per una v.s. ordinata sotto forma di istogramma

Disegnata la curva di frequenza, risulta individuato il massimo la cui ascissa corrisponde al valore modale della distribuzione.

La determinazione grafica della mediana e dei quantili si esegue utilizzando l'istogramma delle frequenze cumulate in cui si è sostituito il contorno curvilineo al corrispondente contorno a gradini. Si interseca la curva interpolatrice con le rette: $y = 1/2$, $y = 1/4$ e $y = 3/4$, e si determinano le ascisse dei punti intersezione, le quali rappresentano rispettivamente il valore della mediana, del primo e del terzo quantile [5] (come illustra la Fig. 1.4.1).

1.5 Rappresentazione sintetica di una variabile statistica ad una dimensione

La rappresentazione grafica della variabile statistica nelle forme sopra viste, pur essendo spesso assai espressiva, da un punto di vista qualitativo, permette di ricavare solo scarse indicazioni quantitative sulla distribuzione, ed è così necessaria una descrizione matematica più accurata. A questo scopo, sono stati introdotti alcuni indici che, ricavati usando tutti i dati componenti la variabile statistica, precisano alcune sue caratteristiche essenziali.
Questi indici descrittivi sono, nel loro insieme, chiamati *statistiche*, ed è comunque necessario premettere alcune definizioni (per poter essere più sintetici nel seguito).

Si chiama *momento k-esimo rispetto ad un polo* θ, di una distribuzione di frequenza o variabile statistica ad una dimensione, la seguente espressione:

$$m_{k,\theta} = \sum_{i=1}^{n} (x_i - \theta)^k \frac{f_i}{N} \quad . \tag{5.1}$$

La serie dei momenti di una variabile statistica, rispetto ad un particolare valore del polo, ad esempio lo zero, basta a rappresentare tutte le caratteristiche della sua distribuzione.

I momenti che nella pratica sono più significativi per rappresentare la funzione sono i tre seguenti:

- Il momento di primo ordine $(k = 1)$ della variabile rispetto al polo $\theta = 0$, che è detto *media* della variabile statistica:

$$m_{1,0} = \sum_{i=1}^{n} x_i \frac{f_i}{N} = M \tag{5.2}$$

- Il momento di secondo ordine $(k = 2)$ della variabile rispetto al polo $\theta = 0$, che è detto *valore quadratico medio*:

$$m_{2,0} = \sum_{i=1}^{n} x_i^2 \frac{f_i}{N} = M_2 \tag{5.3}$$

- Il momento di secondo ordine $(k = 2)$ della variabile rispetto al polo $\theta = M$, che è detto *varianza* o *momento centrale del secondo ordine*:

[5] Il valore assoluto mediano (mav) può essere ricavato approssimativamente, sempre facendo uso della curva di frequenza, calcolando la differenza tra le ascisse, corrispondenti al terzo ed al primo quantile, e facendone la metà.

$$m_{2,M} = \sum_{i=1}^{n} (x_i - M)^2 \frac{f_i}{N} = \sigma^2 \tag{5.4}$$

La differenza: $x_i - M$, è detto *scarto*. La radice quadrata della varianza σ si chiama *scarto quadratico medio* (*sqm*) o *deviazione standard*, ed è espressa nella stessa unità di misura dei valori argomentali x e degli scarti. Si badi tuttavia che le due denominazioni: varianza e sqm., sono frequentemente confuse fra loro. Dalle definizioni precedenti si derivano alcune proprietà.

- La media M è il valore del polo per cui è minimo il momento centrale del secondo ordine [6]. Infatti volendo ricercare il valore di θ per il quale è minima la funzione:

$$M_2(\theta) = \sum_{i=1}^{n} (x_i - \theta)^2 \frac{f_i}{N}$$

occorre derivare la funzione stessa rispetto a θ ed uguagliare a zero la derivata:

$$\frac{d\sum_{i=1}^{n} (x_i - \theta)^2 \frac{f_i}{N}}{d\theta} = -2\sum_{i=1}^{n} (x_i - \theta)\frac{f_i}{N} = 0 \qquad \text{da cui} \qquad \theta = \sum_{i=1}^{n} x_i \frac{f_i}{N} = M$$

- La variabile statistica formata dagli scarti ha valore medio nullo. Infatti per la definizione di media:

$$\sum_{i=1}^{n} (x_i - M)\frac{f_i}{N} = \sum_{i=1}^{N} x_i \frac{f_i}{N} - M = 0$$

- Tra i tre indici sopra definiti esiste la seguente relazione:

$$\sigma^2 = M_2 - M^2 \tag{5.5}$$

Infatti dalla (5.4) sviluppando il quadrato si ha:

$$\sigma^2 = \sum_{i=1}^{n} x_i^2 \frac{f_i}{N} - 2M\sum_{i=1}^{n} x_i \frac{f_i}{N} + M^2 \sum_{i=1}^{n} \frac{f_i}{N}$$

cosicché il primo termine a secondo membro è M_2, per la (5.3), il secondo termine $-2M^2$, per la (5.2), e la sommatoria del terzo termine è uguale alla unità, per la (2.2); da qui si deduce facilmente la (5.5).

[6] La moda è il valore del polo per cui è minimo il momento centrale di ordine zero ed il *valore assoluto medio* (ovvero la media dei valori assoluti degli scarti, rispetto alla media: *MAV*) è il valore del polo per cui è minimo il momento centrale di ordine uno.

Questa relazione risulta spesso utile per il calcolo di σ^2, perché permette di evitare l'esecuzione delle n sottrazioni: $x_i - M$, che costituiscono una mole di calcoli non trascurabile, per n molto grande, e che facilmente introducono errori di approssimazione.

Quando i valori argomentali sono suddivisi in classi, come valori x_i da introdurre nella (5.1) si usano i punti medi delle classi. Può esistere una certa ambiguità là dove la prima classe comprende quegli individui il cui valore argomentale è *inferiore a* … e l'ultima classe quella il cui valore argomentale è *superiore a* … Tuttavia poiché ogni variabile statistica si riferisce ad un fenomeno noto, non si hanno, di solito, difficoltà a stabilire l'ordine di grandezza del valore più piccolo e del più grande, ed a determinare così, sia pure in via approssimata, il punto medio della prima e dell'ultima classe.
Ai due momenti: media e varianza, è associato un importante significato descrittivo della variabile statistica.

- La media M è un indice di posizione, cioè determina la posizione della popolazione cui la variabile statistica si riferisce, sulla scala dei valori argomentali. Con un'analogia meccanica, i valori della frequenza sono interpretati come masse, disposte lungo un asse, ed aventi come coordinata il valore dell'argomento corrispondente, cosicché la media coincide con il baricentro di queste masse.
- La varianza σ^2 è un indice di variabilità, perché misura la dispersione dei valori dell'argomento degli individui intorno al valore medio. Il concetto di varianza è fondamentale in statistica, perché la statistica è un metodo per studiare la variabilità di una caratteristica in una popolazione. Storicamente si è scelto questo indice per la necessità di misurare la variabilità esistente tra i dati, mettendo a confronto le distanze fra ciascun individuo della popolazione ed un termine di confronto globalmente. L'analogia meccanica che meglio si presta per comprendere il significato della varianza è il momento di inerzia I, purché si interpretino ancora le frequenze relative come masse e le differenze $(x_i - M)$, cioè gli scarti, come le distanze delle masse dal centro di rotazione:

$$I = \sum_{i=1}^{n} \frac{f_i}{N} (x_i - M)^2 = \sigma^2$$

Inoltre tenendo conto della (2.2). si può interpretare σ^2 come rotore, cioè quella distanza, dall'asse di rotazione, in cui si può concentrare tutta la massa ($\sum f_i / N$). per conservare invariato il momento di inerzia. Dalla definizione meccanica di momento di inerzia, risulta che I è tanto maggiore, a parità di massa, quanto maggiori sono le distanze dall'asse (e tornando al significato statistico, la varianza σ^2 è una misura della dispersione dei vari individui, intorno al loro valore medio).

Riassumendo la moda esprime una proprietà della *curva* interpolatrice della distribuzione, la mediana [7] una proprietà dell'*area* individuata dalla curva stessa, mentre la media e la varianza rappresentano le proprietà *meccaniche* della distribuzione.

[7] Anche il valore assoluto mediano (mav) è una proprietà dell'area individuata dalla curva stessa.

Tra la media, la mediana e la moda sussiste una relazione per cui il valore della mediana è sempre compreso fra quello della media e quello della moda. Una distribuzione si dice *asimmetrica*, quando i tre valori predetti sono diversi fra loro, ed una misura della asimmetria è data dall'indice di *skewness*:

$$S = \frac{M - Moda}{\sigma} \qquad (5.6)$$

Un secondo modo per misurare l'asimmetria di una distribuzione è dato dal coefficiente γ che fa uso del momento centrale del terzo ordine:

$$\gamma = \frac{\sum (x_i - M)^3 \frac{f_i}{N}}{\sigma^3} = \sum \left(\frac{x_i - M}{\sigma} \right)^3 \frac{f_i}{N} \qquad (5.7)$$

che risulta nullo, nel caso essa sia simmetrica, mentre è positivo, se il diagramma di frequenza presenta una *coda* verso destra, e negativo, in caso contrario.

Tuttavia gli indici di asimmetria devono sempre essere usati con discernimento, perché possono verificarsi casi in cui, ad esempio, si annullano il momento del terzo ordine e questi indici, senza che la variabile statistica sia simmetrica. Inoltre occorre rilevare che in entrambe, le (5.6) e (5.7), le entità $(M - Moda)$ o $(x_i - M)$ sono misurate in termini di sqm σ. Questa procedura, costantemente applicata in seguito, riflette il fatto che, in una distribuzione, il valore della differenza, fra due valori argomentali qualunque, non ha significato se non lo si confronta con l'entità della dispersione che caratterizza la distribuzione stessa.

Se si desidera una più dettagliata descrizione matematica della variabile occorre effettuare il calcolo di momenti di ordine superiore [8]. Tuttavia è più arduo dare significati descrittivi soddisfacenti agli altri momenti. Infatti un'interpretazione intuitiva generale rischia di cadere in difetto, anche per tipi di distribuzione non eccessivamente sofisticati.

1.6 La disuguaglianza di Tchebycheff

Il valore descrittivo dei due indici: media e varianza, è messo in risalto ottimamente dalla disuguaglianza di Tchebycheff che permette di assegnare il limite inferiore al numero o frequenza degli scarti, relativi ad una variabile statistica, in valore assoluto minori di una costante positiva assegnata L.

In altre parole, questa disuguaglianza determina un limite inferiore al numero o frequenza degli individui i cui valori dell'argomento siano contenuti nell'intervallo: $M - L$ e $M + L$. Infatti indicando con v gli scarti, per definizione di varianza, si ha:

---

[8] Il primo momento successivo è il momento centrale del quarto ordine da cui si ottiene il coefficiente β (che mostra la curtosi di una distribuzione). La parola deriva dal greco antico, significa becco e, in generale, serve a spiegare il comportamento delle code (laddove la loro presenza o rimozione modificano il coefficiente β), anche se molti possono essere i contro-esempi. Una curiosità precisa che becchi arrotati servivano come forbici (un'invenzione solo medioevale), per tagliare i capelli (da cui una sostituzione di parola). Invece più importante è precisare che $\beta = 3$ indica la distribuzione normale (centrale in statistica e presentata nel seguito), mentre in generale (e sempre con molti contro-esempi), $\beta > 3$ è una spia di code piuttosto piene (e/o della presenza di dati anomali) e $\beta < 3$ di code parecchio vuote.

$$\sigma^2 = \nu_1^2 \frac{f_1}{N} + \nu_2^2 \frac{f_2}{N} + \ldots + \nu_n^2 \frac{f_n}{N} \qquad (6.1)$$

Si suppone che, nella (6.1), le ν_i siano state ordinate con valore assoluto crescente ed i primi s valori dell'argomento siano inferiori, in valore assoluto, alla costante L, mentre i rimanenti valori: $\nu_{s+1}, \nu_{s+2}, \ldots, \nu_n$, siano maggiori (od almeno uguali) a L. Dopodiché nella (6.1), si pongono i primi s valori di ν uguali a zero ed i rimanenti $n-s$ uguali a L. Con questa sostituzione, si diminuisce (od al massimo non si aumenta) il secondo membro, cosicché:

$$\sigma^2 \geq L^2 \left(\frac{f_{s+1}}{N} + \frac{f_{s+2}}{N} + \ldots + \frac{f_n}{N} \right) \qquad (6.2)$$

Nella (6.2), la somma di frequenze tra parentesi rappresenta la frequenza totale F' degli scarti il cui valore assoluto è superiore o uguale a L. Se con F si indica la frequenza totale degli scarti in valore assoluto inferiori a L, si avrà $F/N + F'/N = 1$. Allora la (6.2) si può scrivere nella forma:

$$\sigma^2 \geq L^2 \left(1 - \frac{F}{N} \right) \qquad (6.3)$$

Dalla (6.3), si ricava F/N e, posto $L = \lambda\sigma$, dove λ misura l'intervallo L, prendendo come unità di misura degli intervalli la quantità σ, si ottiene così la disuguaglianza di Tchebycheff.

$$\frac{F}{N} \geq 1 - \frac{1}{\lambda^2} \qquad (6.4)$$

Il valore limite della frequenza F/N, ricavabile dalla (6.4), è significativo solo nel caso in cui esso sia maggiore di zero e pertanto λ deve essere maggiore dell'unità.

In Fig. 1.6.1, è riportarta a tratto continuo la curva rappresentativa della funzione $F^* = F/N = 1 - 1/\lambda^2$ per valori di λ compresi fra 1 e 5.

Da questa curva, si ricava: più del 75% degli individui della popolazione è racchiuso in un intorno della media dell'ampiezza di $\pm 2\sigma$; più del 90% degli individui della popolazione è contenuto in un intorno delle media dell'ampiezza di $\pm 3\sigma$ [9].

La disuguaglianza di Tchebycheff è di notevole importanza in statistica, giustificando perché i due valori: di media e varianza, nella maggior parte dei casi, bastano a descrivere sufficientemente bene una variabile statistica ed una popolazione. Infatti noti questi due valori, è possibile localizzare e limitare la zona di maggiore addensamento della popolazione, sulla scala degli attributi/argomenti.

---

[9] Dalla stessa disuguaglianza, si ricava invece che il primo intorno della media dell'ampiezza di $\pm\sigma$ potrebbe anche essere vuoto.

Tuttavia occorre notare che l'uguaglianza nella relazione (6.4) si verifica solo se tutti gli scarti, inferiori (in modulo) ad L, sono proprio nulli (perché coincidenti con la media) e tutti gli altri sono esattamente uguali a $\pm L$ (perché distanti L dalla media), eventualità che, in pratica, non si presenta mai. Da queste condizioni, molto restrittive, segue che la frequenza, relativa di un certo scarto, nei casi che interessano le applicazioni, è notevolmente superiore a: $1-1/\lambda^2$.

Nella stessa figura, è tracciata, con tratto discontinuo, la curva $f(\lambda)$ corrispondente ad una particolare distribuzione, molto frequente nella pratica, detta distribuzione normale, dalla quale risulta l'andamento della frequenza degli scarti compresi nell'intervallo: $M \pm \lambda\sigma$. L'importanza della disuguaglianza di Tchebycheff risiede soprattutto nel fatto che, per derivarla, non si è fatta alcun'ipotesi sul tipo di distribuzione di frequenza ed inoltre che essa dimostra qualitativamente, come la frequenza degli scarti molto elevati (rispetto alla varianza), tenda rapidamente a zero.

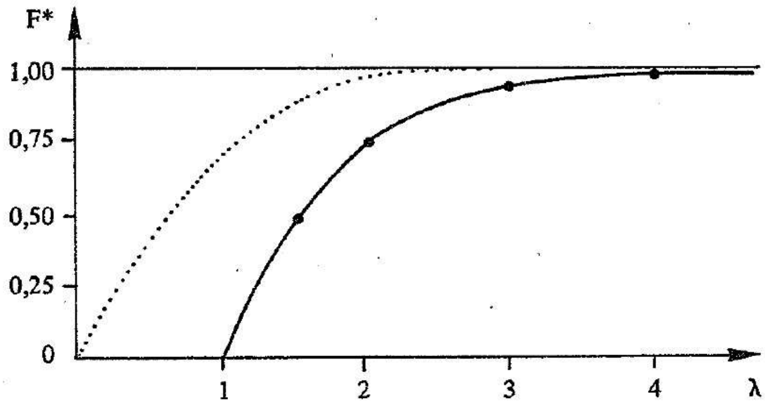


Fig. 1.6.1 – La funzione $F^* = 1 - 1/\lambda^2$ (rappresentata a tratto continuo) con la funzione $f(\lambda)$, di analogo significato, calcolata per la distribuzione normale (rappresentata tratteggiata)

1.7 Alcuni significati collaterali del termine media

Da un punto di vista strettamente metodologico, la media (o momento di ordine 1 rispetto allo zero) ha un valore puramente descrittivo e, in molti casi, se presa isolatamente da altri momenti successivi, ha significato molto limitato, cioè dice assai poco sulla struttura interna della variabile statistica stessa. Tuttavia per moltissimi tipi di variabili statistiche, tra le quali sono quelle che più interessano questo studio, il significato che è attribuito al termine *media* trascende il puro e semplice senso descrittivo, per significare qualcosa di interno ed intrinseco al fenomeno generale della variabile.

In questi casi, gli individui della popolazione si addensano attorno a questo valore caratteristico, quasi ad esprimere una tendenza della popolazione intera a voler assumere quel valore, senza differenziazione tra gli individui. Allora la media coincide con una componente costante della struttura intima della popolazione che agenti perturbatori hanno mascherato, nei singoli individui, dietro una variabilità contingente. Pertanto il significato della media è, per le variabili statistiche di questo tipo, preponderante su quello degli altri indici, giacché palesa qualcosa di intimo nella sostanza della variabile statistica.

Altre volte poi, quando si parla di media o se ne calcola il valore, si insegue un terzo significato, altrettanto preciso, che tuttavia occorre aver presente, per non incorrere in confusione ed ambiguità. Infatti in alcuni casi speciali, ma non rari, si cerca di semplificare una questione che coinvolge un insieme di più individui, sintetizzando con un unico valore i differenti valori degli attributi/argomenti, posseduti dagli individui. A questo valore unico, utilizzato come il valore dell'attributo/argomento, posseduto identicamente da tutti gli individui della popolazione, è dato il nome di media. La sostituzione dei singoli valori argomentali, posseduti dagli individui, con il valore della media, non altera la visione e la trattazione d'insieme del fenomeno nel quale agisce quella popolazione. Tuttavia più correttamente dovrebbe essere chiamata *media relativa*, poiché il suo contenuto di utilità è giustificato solo da applicazioni ad un fenomeno particolare, estraneo alla variabile statistica stessa ed alla sua descrizione.

In questo caso, non ha senso parlare di media di due o più quantità, ma ha senso parlare di media di esse allo scopo della valutazione sintetica di un'altra grandezza, da essa dipendente. Infatti date le grandezze omogenee: $a_1, a_2, ..., a_m$, delle quali interessa considerare la funzione $f(a_1, a_2, ..., a_m)$, se per un dato valore a, si ha: $f(a, a, ..., a) = f(a_1, a_2, ..., a_m)$, questo significa che, agli effetti del calcolo della funzione f, di interesse specifico, è come se tutte le variabili avessero un unico valore a. Allora a è solo la media relativa di: $a_1, a_2, ..., a_m$, agli effetti del calcolo di f.

Il significato relativo di media è di uso frequentissimo anche nel parlare comune: ogni volta, ad esempio, che si parla di consumo medio per abitante di un certo prodotto. Infatti questo valore medio è calcolato in vista non già del consumatore che non si sente affatto rappresentato, in quel numero, e neppure di un'indagine sulla distribuzione statistica dei consumi, ma solo dal punto di vista delle necessità dell'approvvigionamento del prodotto cui non interessa come il prodotto sia suddiviso, nel consumo tra i diversi individui, ma solo il consumo medio di ogni individuo. In base a questo consumo medio ed al numero degli individui, può essere così predisposto il quantitativo da approvvigionare.

Pertanto ogni qualvolta si sente parlare o si vuole parlare di media [10], è importante avere presente i differenti significati e saper applicare giustamente l'uno o l'altro, con discernimento [11].

---

[10] Esistono inoltre altri tipi di media (utilizzati talvolta per motivi specifici ed alcuni impiegati anche nel seguito); ad esempio:

- radice della media quadratica: $x_2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$
- media geometrica: $x_0 = \sqrt[n]{\prod_{i=1}^{n} x_i}$
- media armonica: $x_{-1} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$
- media ponderata (potata): $\bar{x}_p = \frac{\sum_{i=1}^{n} p_i x_i}{\sum_{i=1}^{n} p_i}$ con: $p_i = peso$ (se: $p_i = 0$, per qualche i).

[11] In ambito metrologico, media e sqm sono altresì usati per indicare un valore nominale di una certa grandezza ed un intervallo, dove i valori possibili di questa grandezza sono generalmente raccolti.
Inoltre all'sqm (radice quadrata della varianza, a sua volta, dipendente dalla media), possono essere sostituiti indici di mutua variabilità, facenti uso dei valori argomentali, oppure delle sole frequenze relative (essendo il primo solo quantitativo e non robusto, ed il secondo anche qualitativo e robusto):

- delta di Gini: $\Delta = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} |x_i - x_j|}{n \cdot (n-1)} \quad (\forall\ j \neq i)$

Una perfetta identità formale si ha fra le variabili statistiche e le variabili casuali (presentate nel seguito); tuttavia occorre segnalare che il loro contenuto concettuale è completamente diverso. Infatti le variabili statistiche sono il risultato di classificazioni effettive, già eseguite su individui concreti, mentre le variabili casuali sono classificazioni ipotetiche, cioè semplicemente pensate, che non riguardano la composizione delle popolazioni che sono o si suppongono note, ma piuttosto le operazioni di estrazione a caso, da fare sulle popolazioni stesse [12].

1.8 La variabile statistica a due dimensioni

Data una popolazione di individui ì classificabili secondo due caratteristiche possedute, se ciascuna di queste ha tutti i quattro requisiti necessari (già elencati per la variabile semplice), perché sia considerabile oggetto di classificazione statistica, la variabile costruita su questa popolazione, tenendo conto della duplice modalità degli individui, è detta variabile statistica doppia od a due dimensioni.

Il risultato dell'analisi, dove ogni individuo della popolazione è esaminato e classificato, in base ad una duplice successione di valori argomentali (X, Y), è costituita da una tabella a doppia entrata (come mostra la Fig. 1.8.1) che è la rappresentazione numerica della variabile statistica a due dimensioni.

X / Y	x_1	x_2	...	x_{m-1}	x_m	*Totali per riga* q_i
y_1	f_{11}	f_{12}	...	$f_{1(m-1)}$	f_{1m}	q_1
y_2	f_{21}	f_{22}	...	$f_{2(m-1)}$	f_{2m}	q_2
...	...	...	...	...	...	...
y_{n-1}	$f_{(n-1)1}$	$f_{(n-1)2}$	...	$f_{(n-1)(m-1)}$	$f_{(n-1)n}$	q_{n-1}
y_n	f_{n1}	f_{n2}	...	$f_{n(m-1)}$	f_{nm}	q_n
Totali per colonna p_j	p_1	p_2	...	p_{m-1}	p_m	N

Fig. 1.8.1 – Tabella a doppia entrata rappresentante una variabile statistica a due dimensioni

Le due serie di valori $(x_1, x_2, ..., x_m)$ e $(y_1, y_2, ..., y_m)$, ovvero i valori assunti dagli attributi/argomenti di X e Y, costituiscono l'elemento metrico della tabella, mentre gli $(m\,n)$ valori f_{ij}, cioè gli $(m\,n)$ valori delle caselle interne della tabella, costituiscono l'elemento connettivo.

- rapporto di concentrazione: *nulla* $0 \leq R = \dfrac{\Delta}{2\mu} \leq 1$ *max*
- entropia di Shannon: $H = -\sum_{i=1}^{n} f_i \, log_c \, f_i$
- rapporto di concentrazione: *nulla* $0 \leq \dfrac{H}{log_c \, n} \leq 1$ *max* $c = 2$ (teoria dell'informazione)
  $c = e$ (variabili statistiche continue).

[12] Una concezione, più moderna, interpreta addirittura le variabili casuali, come puri modelli matematici (stocastici), in perfetta analogia alle forme geometriche (astratte) di cui i campi discreti di punti sono le rappresentazioni (concrete).

Ogni valore f_{ij} rappresenta la frequenza assoluta o numero degli individui della popolazione, in possesso della coppia $(x_i \; y_i)$ di valori degli argomenti X e Y, cosicché se tutti gli individui della popolazione sono esaminati, si ha:

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{f_{ij}}{N}=1 \tag{8.1}$$

essendo N il numero totale di individui della popolazione. L'ultima riga e l'ultima colonna, della stessa tabella, contengono rispettivamente

$p_j = \sum_{i=1}^{n} f_{ij} =$ frequenza assoluta con cui X assume il valore x_j, qualunque sia il valore assunto di Y

$q_i = \sum_{j=1}^{m} f_{ij} =$ frequenza assoluta con cui Y assume il valore y_i, qualunque sia il valore assunto di X,

cosicché si ha anche, per la (8.1): $\sum_{j=1}^{m} p_j = \sum_{i=1}^{n} q_i = N$.

Di conseguenza, q_i sono le frequenze assolute della variabile statistica semplice Y, ottenuta analizzando la popolazione solo per la caratteristica Y, mentre p_j sono le frequenze assolute della variabile statistica semplice X, ottenuta analizzando la popolazione solo per la caratteristica X.
La media e la varianza generale della popolazione, per il solo attributo/argomento Y (indipendentemente da X), sono:

$$M_y = \sum_{i=1}^{n} y_i \frac{q_i}{N} \qquad \sigma_y^2 = \sum_{i=1}^{n}\left(y_i - M_y\right)^2 \frac{q_i}{N} \tag{8.2}$$

e le corrispondenti statistiche, per il solo attributo/argomento X, sono:

$$M_x = \sum_{j=1}^{m} x_j \frac{p_j}{N} \qquad \sigma_y^2 = \sum_{j=1}^{m}\left(x_j - M_x\right)^2 \frac{p_j}{N} \tag{8.3}$$

Per illustrare la variabile statistica doppia, si riprende la popolazione dell'Esempio 1.2.1, esaminando sia la statura che il peso della popolazione. La tabella di Fig. 1.8.2 riporta i risultati ottenuti dall'analisi dei singoli individui, a ciascuno dei quali corrisponde una particolare coppia di valori argomentali: la statura in centimetri e il peso in chilogrammi. Con i dati di questa tabella, si costruisce la tabella di Fig. 1.8.3, rappresentativa della variabile statistica a due dimensioni, legata a quella popolazione. L'attributo/argomento X della tabella rappresenta la statura degli individui e l'attributo/argomento Y il loro peso.

X (cm)	Y (kg)	X (cm)	Y (kg)	X (cm)	Y (kg)	X (cm)	Y (kg)	X (cm)	Y (kg)	X (cm)	Y (kg)
172	60	185	71	171	64	170	68	174	68	170	70
181	80	175	65	173	65	185	76	179	68	168	64
176	73	170	70	179	75	178	67	170	60	172	75
175	66	170	66	171	71	180	70	180	67	185	80
169	55	184	73	173	75	178	78	184	73	185	72
164	56	170	60	186	75	172	70	165	56	177	69
184	87	186	78	180	72	185	75	172	63	180	74
174	65	175	73	171	68	180	70	177	67	178	65
170	64	180	73	175	66	181	80	169	69	164	60
170	61	178	70	189	72	172	76	179	65	175	70
180	72	182	75	174	74	177	63	181	65	181	82
186	83	174	68	172	65	175	64	187	87		
179	73	184	82	175	67	179	75	180	73		
170	57	180	73	171	71	170	56	168	65		
179	67	184	78	180	67	184	82	180	80		

Fig. 1.8.2 – Popolazione caratterizzata da due valori argomentali: peso ed altezza (cm)

1.9 Curve di regressione e di variabilità

Uno dei modi più comodi per la rappresentazione grafica della variabile statistica doppia è il diagramma di dispersione, ottenuto rappresentando simbolicamente le frequenze f_{ij}, delle coppie di valori (x_j, y_i), con cerchi di superficie proporzionale a f_{ij}, in corrispondenza dell'ascissa x_j e dell'ordinata y_i.

Tuttavia qualsiasi tipo di rappresentazione grafica di una variabile statistica doppia permette solo valutazioni grossolanamente qualitative, della struttura della popolazione. Invece più espressiva è la rappresentazione mediante due particolari curve, ottenibili dopo una prima rielaborazione sintetica dei dati. Come la variabile statistica semplice si può sinteticamente riassumere in due indici: media e varianza, identicamente si può riassumere la tabella a doppia entrata in due sole colonne che, in generale, bastano a caratterizzare la variabile doppia.

A riguardo, innanzitutto è necessario stabilire in base a considerazioni teoriche od all'intuizione, oppure per semplice comodità, quale attributo/argomento sia da considerare indipendente e quale invece dipendente dall'altro. Il primo è indicato con X e il secondo con Y. In molti casi, questa distinzione è semplice; ad esempio, in tutti i casi in cui un attributo/argomento è una caratteristica ambientale (ora, temperatura, macchina operatrice, ecc.) che condiziona il risultato di una certa operazione, senza esserne condizionata. In altri casi, la scelta può invece risultare dubbia, cosicché possono essere convenienze pratiche a suggerire la scelta dell'attributo/argomento indipendente X e di quello dipendente Y.

Dopodichè considerando l'accoppiamento della prima colonna della tabella, contenente i valori argomentali Y, con ciascuna colonna successiva, si ottengono m variabili statistiche semplici, ciascuna caratterizzata da un diverso valore dell'argomento X, posseduto dai propri individui.

Y \ X	164	5	6	7	8	9	170	1	2	3	4	5	6	7	8	9	180	1	2	3	4	5	6	7	8	189	q_i
55						1																					1
6	1	1					1																				3
7							1																				1
8																											
9																											
60							2		1																		4
1							1																				1
2																											
3									1					1													2
4					1		1	1				1															4
5					1				1	1	1	1			1	1		1									8
6												2															2
7												1		1	1	1	2										6
8							1	1			2					1											5
9						1								1													2
70							2		1			1			1		2										7
1								2														1					3
2																	2					1				1	4
3												1	1			1	3				2						8
4											1						1										2
5							1		1	1						2			1			1	1				8
6									1													1					2
7																											
8															1						1		1				3
9																											
80																	1	2				1					4
1																		1			2						
2																							1				3
3																											1
4																											
5																											
6																											
87																					1					1	2
p_j	2	1			2	2	10	4	6	2	4	7	1	3	4	6	11	4	1		6	5	3	1		1	86

Fig. 1.8.3 – Variabile statistica bidimensionale rappresentata nella forma di tabella a doppia entrata

Per ciascuna di queste variabili semplici, con formule, derivate dalle (5.2) e (5.4), si possono calcolare i due indici fondamentali: media e varianza:

$$M_{yj} = \frac{y_1 f_{1j} + y_2 f_{2j} + \ldots + y_n f_{nj}}{p_j} = \frac{\sum_{i=1}^{n} y_i f_{ij}}{p_j} \qquad (9.1)$$

$$\sigma_{yj}^2 = \frac{\sum_{i=1}^{n} \left(y_i - M_{yj}\right)^2 f_{ij}}{p_j} \tag{9.2}$$

ottenendo così un tabella sintetica, ridotta a sole quattro colonne:

X	M_{y_j}	$\sigma_{y_j}^2$	p_j
x_1	M_{y_1}	$\sigma_{y_1}^2$	p_1
x_2	M_{y_2}	$\sigma_{y_2}^2$	p_2
.	.	.	.
.	.	.	.
.	.	.	.
x_m	M_{y_m}	$\sigma_{y_m}^2$	p_m

La (9.1) prende il nome di *media parziale per colonna* e la (9.2) di *varianza parziale per colonna*. Questa tabella sintetica si può facilmente rappresentare per mezzo di due curve: la *curva di regressione* e la *curva di variabilità* nelle quali rispettivamente le medie M_{y_j} e le varianze $\sigma_{y_j}^2$ sono riportate sulla scala dei valori dell'attributo/argomento indipendente X.

A mo' di esempo, le curve di regressione e di variabilità forniscono la rappresentazione grafica della variabile statistica doppia, presente nella tabella di Fig. 1.8.2, scegliendo l'attributo/argomento: statura degli individui X, a buon diritto, come indipendente, poiché si presume che sia la statura a condizionare, in parte, il peso Y (attributo/argomento dipendente) e non viceversa.

x_j	M_{y_j}	$\sigma_{y_j}^2$	p_j	x_j	M_{y_j}	$\sigma_{y_j}^2$	p_j
164	58.0	4.0	2	7	66.3	6.2	3
5	56.0	0	1	8	70.0	24.5	4
6				9	70.5	15.9	6
7				180	71.9	11.7	11
8	64.5	0.3	2	1	76.7	46.7	4
9	62.0	49.0	2	2	75.0	0	1
170	64.1	36.3	10	3			
1	68.5	8.3	4	4	79.2	25.7	6
2	68.2	35.8	6	5	74.8	10.2	5
3	70.0	25.0	2	6	78.6	10.9	3
4	68.8	10.7	4	7	87.0	0	1
5	67.3	8.5	7	8			
6	73.0	0	1	9	72.0	0	1

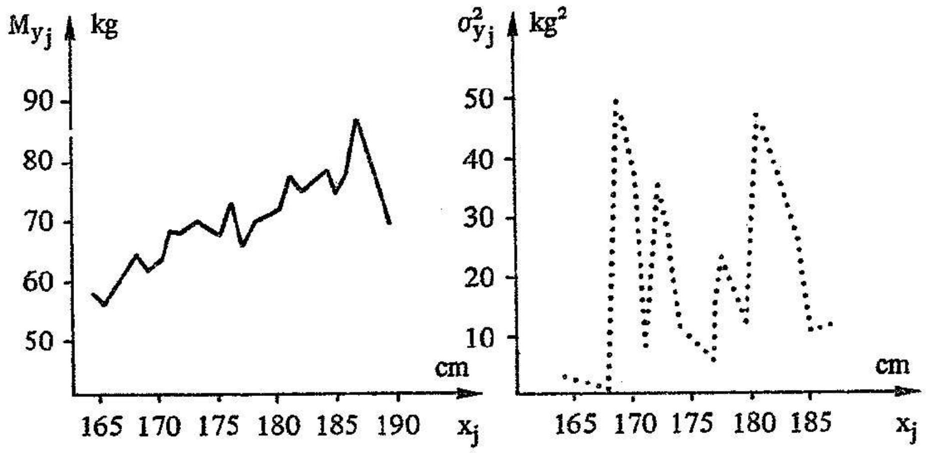


Fig. 1.9.1 – Curva di regressione e curva di variabilità

Mediante la (9.1) e la (9.2) si calcolano le medie e le varianze parziali (nel calcolo è conveniente mantenere una sola cifra decimale) ed il risultato è mostrato nella tabella di Fig. 1.9.1 con la quale si tracciano le curve di regressione e di varibilità corrispondenti.

La curva di regressione presenta un andamento abbastanza regolare, mentre molto più irregolare risulta l'andamento della curva di variabilità, in gran parte, dovuto all'esiguo numero di individui, contenuto in alcune colonne parziali, soprattutto per i valori estremi del campo.

La costruzione delle due curve: di regressione e di variabilità, è molto utile, per seguire l'andamento di un processo produttivo. Infatti dato un processo produttivo in serie ed il prodotto costituito da un pezzo il quale deve possedere particolari requisiti, ad esempio, una dimensione prestabilita, ogni giorno si esamina tutta (o parte) della produzione giornaliera e, in funzione della data (variabile indipendente), si riporta in grafico la media delle misure ottenute, dalla verifica delle dimensioni dei pezzi, e la varianza delle misure stesse. La prima curva (curva di regressione) permette di controllare la presenza di eventuali fattori che perturbano sistematicamente la produzione, spostando progressivamente la dimensione media del pezzo, mentre la seconda curva (curva di variabilità) permette di controllare l'eventuale insorgere nella produzione di cause perturbatrici di carattere accidentale.

1.10 Misura della dipendenza fra gli argomenti di una variabile statistica doppia

Lo studio, non già delle due variabili X e Y indipendentemente (analizzandole come variabili statistiche semplici marginali, oppure in due famiglie condizionate), ma sintetizzate con una misura, richiede un nuovo indice, il più possibile significativo, capace di stabilire quali relazioni esistono fra le due variabili, presenti simultaneamente in un gruppo di individui.

In questo caso, si vuole un criterio che giudichi se e quanto le due variabili sono reciprocamente dipendenti. Questa interdipendenza è facile da mettere in luce e da studiare in fenomeni semplici e ripetibili, per mezzo di esperienze condotte secondo schemi opportunamente studiati. Tuttavia diventa assai problematico rivelarla, quando il fenomeno si presenta complesso ed inoltre la dipendenza è così tenue da essere facilmente mascherata da altre cause perturbatrici.

Questa indagine mira a detereminare, se un dato individuo della popolazione, studiata sotto due aspetti, in possesso di un certo valore x_j dell'argomento X, possa indifferentemente possedere qualsiasi valore

dell'argomento Y , oppure se invece la frequenza con cui assume i diversi valori di Y è dipendente dal valore dell'argomento X posseduto.
Come in ogni problema statistico, si vuole dare un indice di questa che è detta dipendenza fra gli argomenti, indice che deve, in un certo senso, rappresentare la misura della dipendenza stessa. Innanzitutto occorre definire i due casi limite, mai verificati nei dati di un'esperienza (ma mentalmente concepibili): i casi della *perfetta dipendenza* e della *perfetta indipendenza* fra le due variabili x e y .

- Si ha perfetta dipendenza, quando il fatto che un individuo possegga un determinato valore x_j di X determina univocamente quale valore y_i di Y l'individuo stesso possiede. In una tabella a doppia entrata, tipo quella di Fig. 1.8.1, si ha, in ogni colonna x_j una sola frequenza f_{ij} diversa da 0, in quanto tutti gli individui, aventi il valore argomentale x_j , hanno tutti lo stesso valore argomentale y_i . Si può allora pensare cioè che esista una funzione: $y' = f(x)$, che lega i due argomenti della variabile statistica doppia, detta curva di dipendenza. L'equazione della curva di dipendenza $y' = f(x)$ può essere determinata:

  - empiricamente cercando funzioni arbitrarie ritenute adatte a descrivere bene la relazione fra x e y ;
  - razionalmente creando schermi logici costituenti una possibile spiegazione della relazione fra x e y e non una semplice descrizione.

  Nella maggioranza dei casi, non si hanno a disposizione curve di dipendenza di questo tipo. Allora si assume, come curva di dipendenza, la curva di regressione (precedentemente descritta), perché questa costituisce, in realtà, la rappresentazione più fedele dell'andamento di Y in funzione di X , ricavabile dai soli dati sperimentali.

- Si ha una perfetta indipendenza, quando il comportamento della variabile dipendente Y può essere ricavato, dall'esperimento stesso, osservando come si comporta Y , in assenza di X .
  Ad esempio, se la variabile statistica doppia è costruita sulla popolazione *campi coltivati a grano in una certa regione*, ed i due argomenti di classificazione sono: $X =$ *quantità di fertilizzante somministrato a ciascun campo*, $Y =$ *quantità di grano prodotto da ciascun campo*, il comportamento di Y può essere empiricamente determinato, *indipendentemente* da X , osservando il rendimento di quei campi cui non è stato somministrato alcun fertilizzante.
  Tuttavia nella maggioranza dei casi, non è facile o non è possibile, analizzare Y , in assenza di X , e si deve pertanto ricorrere ad ipotesi logicamente ammissibili. Precisamente la statistica propone, come ipotesi di *indipendenza in media*, o *correlativa*, quella che, in caso di indipendenza, le medie dei valori di Y (variabile dipendente) rimangono invariate, per ogni valore x_j di X . Allora le medie parziali per la colonna M_{y_j} devono essere tutte uguali, tra loro, ed uguali alla media generale M_y , cosicché ogni colonna risulti provvista di una certa stabilità del valore medio.

Infine con una definizione più esauriente, si ha perfetta indipendenza quando, separando la popolazione in tanti gruppi, in base ai valori di X, ciascuno può considerarsi come un gruppo di estrazioni, a caso, dalla popolazione, presa in toto. Ogni gruppo ha media oscillante, intorno alla media generale, con oscillazioni tanto minori, quanto è maggiore il suo numero di individui. Poiché aver scelto gli individui in base a x, non provoca una corrispondente scelta degli individui rispetto a y, essi si comportano come estratti a caso e non si ha dipendenza fra X e Y. Ad esempio, nella seguente tabella Y è, in media, indipendente da X, perché la media di y, in corrispondenza a diversi valori x_j, è sempre: $M_{y_j} = 3$.

X / *Y*	*1*	*2*	*5*	*7*	*10*	*16*	*Totali*
1	1			1	2	1	5
2	3		1	1	4	5	14
3	8	15	3	1	6	7	40
4	3		1	1	4	5	14
5	1			1	2	1	5
Totali	16	15	5	5	18	19	78
Medie parziali y_j	3	3	3	3	3	3	3

Nei casi intermedi, occorre interpretare il valore argomentale y_i, presente in un individuo qualunque, tra quelli aventi valore argomentale x_j. La media M_{y_j} è il valore dominante, in questo gruppo, poiché le eventuali differenze: $y_i - M_{y_j}$, sono dovute a fattori accidentali o comunque indipendenti dalla variabile X, presente con lo stesso valore, in tutti gli individui del gruppo. Per questo motivo, le differenze: $y_i - M_{y_j}$, sono dette *scarti indipendenti* e rappresentano le dispersioni dei singoli valori y_i, intorno alla loro media, dovute a cause che nulla hanno a che fare con l'influenza di X. Nel caso si sia precedentemente fissata una relazione: $y' = f(x)$, assunta come curva di dipendenza, in corrispondenza di ogni x_j, si ha una ben determinata y'_j ed ancora le differenze: $y_i - y'_j$, sono dovute a motivi accidentali (od indipendenti da x_j comunque). Invece analizzando il significato di una singola media per colonna M_{y_j}, questa è scomponibile in due parti: una indipendente da X (che esiste anche in assenza di X), ed una causata da quel valore di x_j particolare, posseduto da tutti gli individui di cui M_{y_j} è la media.

Nell'esempio precedente, il valore medio del frumento raccolto nei campi cui è stata somministrata una certa quantità x_j di fertilizzante si compone di: una parte, raccolta comunque, anche senza fertilizzante, più una parte, dovuta invece al fertilizzante, in misura x_j.

Dato che si è convenuto di considerare la media generale M_y come il valore argomentale corrispondente all'ipotesi di *indipendenza correlativa* o *in media*, si può indicare la parte di M_{y_j}, dipendente da x_j, che prende il nome di *scarto dipendente*, come: $M_{y_j} - M_y$. L'attendibilità, o peso, di tale scarto: $M_{y_j} - M_y$, dipende ovviamente dal numero p_j di individui, a disposizione, per calcolare la media per colonna M_{y_j}.

Dopodiché definiti gli scarti dipendenti ed indipendenti, si osserva che, se si ha, fra X e Y:

- una perfetta dipendenza, tutti gli individui con valore argomentale x_j hanno la stessa y_i, coincidente con la media M_{y_j} e pertanto tutti gli scarti indipendenti: $y_i - M_{yj}$, sono nulli;
- una perfetta indipendenza, tutte le medie perziali per colonne M_{y_j} sono uguali fra loro ed uguali alla media generale M_y ed allora tutti gli scarti dipendenti: $M_{y_j} - M_y$, sono nulli.

Un indice, semplice ed intuitivo, per rappresentare il grado di dipendenza, fra X e Y, è il rapporto:

$$\frac{\sum |scarti\ \ dipendenti\ \ |}{\sum |scarti\ \ indipenden\ \ ti\ |}$$

Tuttavia questo rapporto ha il difetto di essere infinito, nel caso di perfetta dipendenza, mentre:

$$I = \frac{\sum |scarti\ \ dipendenti\ \ |}{\sum |scarti\ \ \ indipenden\ \ ti| + \sum |scarti\ \ dipendenti\ \ |}$$

si annulla, nel caso di perfetta indipendenza, e vale uno, nel caso di perfetta dipendenza (perché numeratore e denominatore diventano uguali). I valori intermedi (ad esempio: $I = 0{,}2$ ed $I = 0{,}7$) misurano il minore o maggior grado di indipendenza fra X e Y.

In pratica, si preferisce usare un indice quadratico di dipendenza I_2 che si può scrivere sotto due forme diverse, a seconda che si utilizzi, come curva di dipendenza, una particolare funzione: $y' = f(x)$, oppure la curva di regressione;

$$I_2 = \frac{\sum_j (y'_j - M_y)^2 p_j}{\sum_i \sum_j (y_i - y'_j)^2 f_{ij} + \sum_j (y'_j - M_y)^2 p_j}$$

$$I_2 = \frac{\sum_j (M_{y_j} - M_y)^2 p_j}{\sum_i \sum_j (y_i - M_{y_j})^2 f_{ij} + \sum_j (M_{y_j} - M_y)^2 p_j} = \eta_x^2$$

Nella seconda forma, l'indice quadratico di dipendenza prende il nome di *Indice di Pearson* ed è quello più frequentemente usato. Fondamentale è comunque la differenza di significato esistente fra η_x^2 e I_2. Infatti il primo misura il grado di dipendenza di Y da X, mentre il secondo misura quanto bene la particolare funzione: $y' = f(x)$, usata per il calcolo di I_2, può rappresentare la dipendenza (potendosi presentare il caso in cui, utilizzata erroneamente una y', per costituire un modello adeguato della variabilità indotta in Y dal variare di X, si ricava un valore di I_2 prossimo a zero, mentre contemporaneamente il valore di η_x^2, che non richiede, per essere calcolato, nessuna ipotesi di modello, risulta invece elevato).

Se si dividono numeratore e denominatore di η_x^2 per N, si ha al numeratore: $\frac{1}{N}\sum_j \left(M_{y_j} - M_y\right)^2 p_j$, che si può interpretare come varianza della variabile statistica semplice, costituita dalle medie per colonna M_{y_j} (posto che, a ciascun M_{y_j}, sia assegnata una frequenza relativa uguale a p_j / N), ed al denominatore:

$$\frac{1}{N}\sum_{j=1}^{m}\sum_{i=1}^{n}\left(y_i - M_{y_j}\right)^2 f_{ij} + \frac{1}{N}\sum_{j=1}^{m}\left(M_{y_j} - M_y\right)^2 p_j = \bar{\sigma}_y^2 + \sigma_{\bar{y}}^2 =$$

$$= \frac{1}{N}\sum_j\sum_i\left(y_i^2 - 2M_{y_j}y_i + M_{y_j}^2\right)f_{ij} + \frac{1}{N}\sum_j\left(M_{y_j}^2 - 2M_{y_j}M_y + M_y^2\right)p_j =$$

$$= \frac{1}{N}\sum_j\left(M_{y_j^2} - 2M_{y_j}^2 + M_{y_j}^2\right)p_j + \frac{1}{N}\sum_j\left(M_{y_j}^2 - 2M_{y_j}M_y + M_y^2\right)p_j =$$

$$= \frac{1}{N}\sum_j\left(M_{y_j^2} - M_{y_j}^2\right)p_j + \frac{1}{N}\sum_j\left(M_{y_j}^2 - 2M_{y_j}M_y + M_y^2\right)p_j = \frac{1}{N}\sum_j\left(M_{y_j^2} - 2M_{y_j}M_y + M_y^2\right)p_j =$$

$$= M_{y^2} - 2M_y^2 + M_y^2 = M_{y^2} - M_y^2 = \sigma_y^2 = \frac{1}{N}\sum_j\sum_i\left(y_i - M_y\right)^2 f_{ij} = \sum_i\left(y_i - M_y\right)^2\frac{q_i}{N}$$ [13]

cosicché l'indice di Pearson si trasforma nel rapporto della varianze fra le medie per colonna con la varianza generale della popolazione. Invece un'analoga traformazione si può fare su I_2, solo se vale la relazione di ortogonalità fra gli scarti dipendenti e gli scarti indipendenti: $\sum\sum\left(y_i - y'_j\right)\left(y'_j - M_y\right) = 0$ (dove la doppia sommatoria si esegue su entrambi gli indici: i e j), che si verifica, ad esempio, quando l'equazione della curva di dipendenza: $y' = f(x)$, è ricavata con un procedimento ai *minimi quadrati*:

$$\eta_x^2 = \frac{\sum_j\left(M_{y_j} - M_y\right)^2 p_j}{N\cdot\sigma_y^2} \qquad\qquad I_2 = \frac{\sum_j\left(y'_j - M_y\right)^2 p_j}{N\cdot\sigma_y^2} \qquad (10.1)$$

[13] La suddetta identità è detta teorema di decomposizione ortogonale della varianza e stabilisce che la somma della varianza spiegata (cioè della varianza tra le medie per colonna, con la media generale) e della varianza residua (cioè della media delle varianze per colonna) è uguale alla varianza generale. A riguardo, particolarmente importante è proprio l'ortogonalità che esclude una covarianza tra scarti spiegati e scarti residui, come tra medie e varianze (semplici), mentre si ha una dipendenza lineare tra medie e dati.

Anche I_2 e η_x^2 sono uguali a 0, quando si ha perfetta indipendenza, ed hanno valore compreso fra 0 ed 1, in tutti gli altri casi. La dipendenza fra X e Y è da considerarsi nulla quando: $\eta_x^2 < 0.2$ [14]; è presente, ma debole, quando: $0.2 < \eta_x^2 < 0.4$, sensibile quando: $0.4 < \eta_x^2 < 0.6$, forte quando: $0.6 < \eta_x^2 < 0.8$, e fortissima quando: $\eta_x^2 > 0.8$. Tenendo presente la (10.1), si può calcolare l'indice di Pearson per la popolazione dell'esempio del paragrafo 1.8 (la cui variabile statistica doppia è contenuta nella tabella di Fig. 1.8.2). A riguardo, si calcola dapprima la media: $M_y = 70.08$. Dopodiché usando questo valore medio ed i dati della tabella di Fig. 1.9.1, si ricava il numeratore:

$$\sum_{j=1}^{m}\left(M_{y_j} - M_y\right)^2 p_j = 2450$$

Infine, con la seconda delle (8.2), si calcola la varianza generale σ_y^2 rispetto ad Y: $\sigma_y^2 = 48{,}59$, cosicché introducendo questo risultato nella (10.1), si ha:

$$\eta_x^2 = \frac{2450}{86 \times 48.59} = 0{,}61$$

potendo dedurre che la dipendenza fra X e Y è abbastanza sensibile.

La Fig. 1.10.1 illustra graficamente i concetti esposti. La retta: $y = M_y = cost$, rappresenta il luogo in cui dovrebbero cadere tutti i punti che rappresentano gli individui della popolazione, se fra x e y si ha perfetta indipendenza. La curva può rappresentare la curva di regressione, oppure una curva di dipendenza scelta:

---

[14] Per dipendenze piuttosto basse, è altresì utile rifarsi al concetto di indipendenza stocastica, come decomposizione di una frequenza doppia nel prodotto delle due frequenze marginali, e definire le contingenze come differenza fra la frequenza doppia ed il prodotto delle frequenze marginali:

$c_{ij} = f_{ij} - p_i q_j$ con: $-1 \le c_{ij} \le 1$ dove: $c_{ij} = 0$ è la condizione di indipendenza.

Dopodiché calcolata la semicontingenza media: $C_0 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left|c_{ij}\right|$, si definiscono gli indici di Bonferroni (monolaterali e bilaterali):

$$\begin{cases} \beta_X = \dfrac{c_0}{1 - \sum_{i=1}^{n} p_i^2} & ;\ \textit{indipendenza}\ \ 0 \le \beta_X \le 1\ \ \textit{perfetta dipendenza}\ \ x = h(y) \\ \beta_Y = \dfrac{c_0}{1 - \sum_{j=1}^{m} q_j^2} & ;\ \textit{indipendenza}\ \ 0 \le \beta_Y \le 1\ \ \textit{perfetta dipendenza}\ \ y = g(x) \end{cases}$$

$$\begin{cases} \beta_0 = \sqrt{\beta_X \beta_Y} & ;\ \textit{indipendenza}\ \ 0 \le \beta_0 \le 1\ \ \textit{perfetta dipendenza bilaterale} \\ \beta_{-1} = \dfrac{2\beta_X \beta_Y}{\beta_X + \beta_Y} & ;\ \textit{indipendenza}\ \ 0 \le \beta_{-1} \le 1\ \ \textit{perfetta dipendenza bilaterale} \end{cases}$$

$y' = f(x)$, e su questa devono cadere tutti i punti, se si ha perfetta dipedenza. In realtà, i punti cadono sulle n rette di equazione: $x = x_j$, e sono p_j su ciascuna di esse. L'indice di dipendenza esprime il rapporto fra lo scostamento della curva dalla retta: $y = M_y = cost$, e la dispersione dei punti rispetto alla stessa retta $y = M_y = cost$.

Se come equazione: $y' = f(x)$, relativa ad una certa variabile statistica doppia, si sceglie una retta od una parabola del secondo ordine, oppure una parabola di k-esimo ordine, variano i valori degli indici, calcolati in corrispondenza: $I_2^{(1)}, I_2^{(2)}, ..., I_2^{(K)}$,secondo la relazione: $I_2^{(1)} \leq I_2^{(2)} \leq ... \leq I_2^{(K)} \leq \eta_x^2$. Infatti esiste un limite superiore alla successione I_2 che è rappresentato dall'indice di dipendenza di Pearson η_x^2, relazione che sottolinea maggiormente l'importanza della curva di regressione, come curva rappresentativa della dipendenza fra i due argomenti della varibile statistica doppia [15].

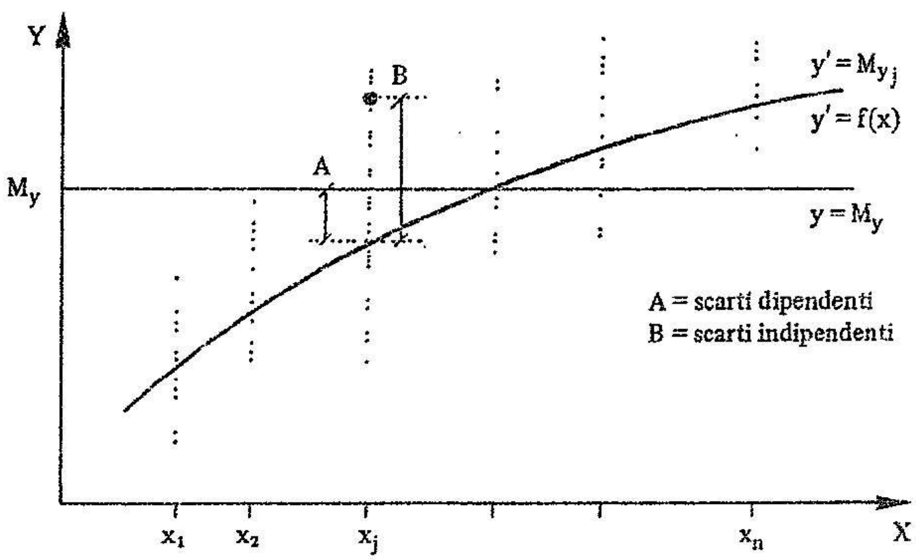


Fig. 1.10.1 – Rappresentazione grafica degli scarti dipendenti e indipendenti

1.11 Coefficienti di regressione

Nel caso esista una dipendenza, fra i due attributi/argomenti X e Y, di una variabile statistica doppia, si può proporre, con diversi criteri, una relazione analitica fra le due variabili. Tuttavia quando la dipendenza non è perfetta, non è possibile trovare una funzione, per far corrispondere esattamente ai valori argomentali x quelli corrispondenti y, ed il problema diventa un problema di interpolazione, da risolvere con opportune metodologie. Allora particolare semplicità presenta il caso in cui la funzione interpolante: $y' = f(x)$, può essere considerata una retta, assumendo la forma: $y' = ax + b$.

Se la dipendenza fra X e Y non è perfetta, i valori di y' differiscono dai corrispondenti y, qualunque sia la scelta di coefficienti a e b; cosicché è ben evidente, avendo la differenza: $y' - y$, il significato di uno scarto,

[15] Un'alternativa alla regressione classica è data dalla regressione robusta, dove le mediane sono al posto delle medie ed i valori assoluti mediani (mav) al posto delle varianze. In questo caso non vale ovviamente il teorema di decomposizione ortogonale della varianza: comunque il mav spiegato è dato dal mav delle mediane condizionate ed il mav residuo dalla mediana dei mav condizionati.

che occorra cercare valori dei parametri che rendano gli scarti, nel loro insieme, i più piccoli possibili. Si tratta di un problema di stima ed il metodo più usato è quello dei *minimi quadrati*, consistente nell'imporre una scelta dei parametri, fatta in modo che la somma dei quadrati degli scarti sia minima.

Pertanto detti y_k i valori argomentali di Y e y'_k i valori: $y'_k = ax_k + b \ (k = 1, 2, ..., N)$, si ha:

$$\sum_{k=1}^{N} (y_k - y'_k)^2 = \sum_{k=1}^{N} (y_k - (ax_k + b))^2 = min \tag{11.1}$$

La (11.1) è una espressione quadratica nelle quantità: a e b, considerate variabili. Condizione necessaria e sufficiente, affinchè la (11.1) sia soddisfatta, è che le derivate parziali del primo membro, rispetto ad a e b siano uguali a zero, cosicché si ottengono due equazioni lineari nelle due incognite a e b, effettuate le derivazioni:

$$\begin{cases} \sum_{k=1}^{N} x_k (y_k - (ax_k + b)) = 0 \\ \sum_{k=1}^{N} (y_k - (ax_k + b)) = 0 \end{cases} \qquad \begin{cases} a\sum x^2 + b\sum x = \sum xy \\ a\sum x + Nb = \sum y \end{cases} \qquad \text{(11.2) ed (11.3)}$$

Risolvendo il sistema, si ottengono i valori di a e b, e si può scrivere l'equazione della retta interpolatrice: $y' = ax + b$ (dove, per brevità, si omette l'apice su y, per indicare la retta di regressione).

Il coefficiente a è il coefficiente direttivo della retta, è chiamato *coefficiente di regressione lineare* e ha grande importanza, mentre il coefficiente b dipende solo dalla scelta dell'origine degli assi e non ha alcun significato statistico. Ad esempio, si ha: $a = 0$, Y è linearmente indipendente da X (a riguardo, non è inutile osservare che spesso X e Y sono grandezze diverse ed a non è un numero fisso, né un numero puro, in generale, ma una grandezza il cui valore numerico dipende dalle unità di misura prescelte).

Se il valore di b non interessa, può essere eliminato, riferendo i valori: x_k e y_k, alle loro medie $M(X)$ e $M(Y)$, considerando le differenze: $v_{x_k} = x_k - M(X)$ e $v_{y_k} = y_k - M(Y)$. In questo caso, si ha: $\sum v_{x_k} = \sum v_k = 0$, e si ricava semplicemente: $b = 0$, dalla seconda delle (11.3), mentre la prima equazione fornisce il valore di a:

$$a = \frac{\sum v_{x_k} v_{y_k}}{\sum v_{x_k}^2} \tag{11.4}$$

fondamentale, per il suo significato e perché permette il calcolo immediato del coefficiente di regressione. Dopodiché dividendo numeratore e denominatore della (11.4), per il numero N delle coppie di valori a disposizione; il denominatore diventa:

$$\sigma_x^2 = \frac{1}{N} \sum_j \left(x_j - M_x\right)^2 p_j$$

cioè la varianza della variabile indipendente. Analogamente il numeratore ha l'espressione (dove per il terzo termine dell'uguaglianza, si procede in analogia allo sviluppo della (5.5)):

$$\sigma_{xy} = \frac{1}{N} \sum \sum \left(x_j - M_x\right)\left(y_i - M_y\right) f_{ij} = \frac{1}{N} \sum \sum x_j y_i f_{ij} - M_x M_y$$

che, detta *covarianza* ed è indicata con il simbolo σ_{xy}, può essere interpretata come una varianza congiunta delle due variabili X e Y. Con le posizioni fatte, la (11.4) diventa allora:

$$a = \frac{\sigma_{xy}}{\sigma_x^2} \tag{11.5}$$

dove il coefficiente di regressione lineare è dato dal rapporto tra la covarianza e la varianza della variabile indipendente.

Analogamente per la dipendenza di X e Y, si pone: $x = a' y + b'$, trovando i due parametri, sempre con il metodo dei minimi quadrati, con equazioni formalmente identiche alle (11.3). Se la dipendenza tra le due varibili è perfetta, risulta evidentemente: $a' = 1/a$, e la retta coincide con la precedente. Tuttavia in generale, questo non accade, minimizzando: $\sum (y' - y)^2$, nel primo caso e: $\sum (x' - x)^2$, nel secondo (cioè le somme dei quadrati delle distanze dei punti osservati dalla retta interpolante, misurate rispettivamente nel senso di: y e x). Pertanto riferite ancora: x e y, alle loro medie, l'espressione di a' diventa:

$$a' = \frac{\sigma_{xy}}{\sigma_y^2} \tag{11.6}$$

dove σ_y^2 è la varianza della Y, considerata come variabile indipendente, mentre il numeratore non varia (non variando evidentemente la covarianza, con lo scambio fra X e Y).

D'altra parte, il problema della regressione lineare è molto più complesso, non bastando determinare la retta di regressione, ma dovendosi anche valutare, quanto bene rappresenti l'andamento dei dati sperimentali.

1.12 Coefficiente di correlazione lineare

L'indice quadratico di dipendenza (presentato nel paragrafo 1.10) permette di misurare l' adattamento di una variabile statistica doppia ad una curva di dipendenza, di forma e tipo qualsiasi, purchè ben determinata e definita, anche numericamente (cioè essendo una funzione calcolabile, per ogni valore di x).

Spessissimo invece si presenta il quesito di misurare il grado di dipendenza fra gli attributi/argomenti X e Y, di una variabile statistica doppia, rappresentata da una curva della quale si stabilisce solo la famiglia, la forma ed il tipo, senza prefissarne i parametri, cosicché i valori $f(x)$ non possono essere calcolati.

Particolare interesse riveste il caso in cui la famiglia delle curve di dipendenza è costituita da rette ed il legame fra X e Y è lineare, senza ulteriori precisazioni, sui parametri della relazione stessa. In questo caso particolare, un indice, chiamato *coefficiente di correlazione lineare*, misura il grado di adattamento dei dati sperimentali all'ipotesi che, fra gli attributi/argomenti X e Y della variabile statistica, esista un legame di tipo lineare.

La correlazione è un aspetto più generale della regressione, in quanto considera la relazione, intercedente fra le due variabili, senza porre la condizione di dipendenza di natura causale tra le variabili stesse. Infatti nella correlazione si considera l'interdipendenza fra le variabili, nel senso che sia una che l'altra possono fungere da variabile indipendente.

Un metodo rigoroso, per introdurre il coefficiente di correlazione lineare, è suggerito da Bonferroni che facilmente si può estendere a curve di tipo diverso (ad esempio, di secondo o terzo grado).

Dati N individui della popolazione esaminata (cui corrispondono altrettante coppie di valori: x e y), si rappresenta ogni individuo con un punto, nel piano XY. Come noto, tre punti qualsiasi, rappresentanti tre individui qualsiasi, rispettivamente di coordinate: (x_1, y_1), (x_2, y_2) e (x_3, y_3), sono allineati, quando si annulla il determinate costruito sulle coordinate:

$$det\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = 0 \qquad (12.1)$$

Con N coppie di valori rappresentanti ciascun individuo si può costruire la matrice:

$$X = \begin{vmatrix} 1 & 1 & 1 & ... & 1 \\ x_1 & x_2 & x_3 & ... & x_N \\ y_1 & y_2 & y_3 & ... & y_N \end{vmatrix} . \qquad (12.2)$$

L'indice numerico progressivo con cui sono contrassegnati i valori: x e y non serve ad individuare il valore argomentale posseduto dall'individuo corrispondente, ma solo ad indicare che: x_i ed y_i, sono due valori argomentali, posseduti dall'individuo al posto i-esimo nella successione di individui della popolazione (supposti ordinati, secondo una regola qualsiasi). In generale, le N colonne della matrice (12.2) non sono tutte diverse, una dall'altra, ma ogni coppia di valori degli argomenti è ripetuta, tante volte, quanto la frequenza di quella coppia di valori della variabile statistica doppia.

Innanzitutto è opportuno ricordare che la matrice AA', ottenuta moltiplicando A, per la sua trasposta A', è simmetrica ed ha la seguente forma:

$$AA' = \begin{vmatrix} \sum a_{1i}^2 & \sum a_{1i}a_{2i} & \dots & \sum a_{1i}a_{mi} \\ \sum a_{2i}a_{1i} & \sum a_{2i}^2 & \dots & \sum a_{2i}a_{mi} \\ \dots & \dots & \dots & \dots \\ \sum a_{mi}a_{1i} & \sum a_{mi}a_{2i} & \dots & \sum a_{mi}^2 \end{vmatrix} \qquad (12.3)$$

essendo:

$$A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{vmatrix} \qquad (12.4)$$

Il determinate della matrice AA' possiede le seguenti proprietà:

- se: $n < m \Rightarrow det(AA') = 0$
- se: $n > m \Rightarrow det(AA') = \sum \delta_m^2(A)$

La seconda proprietà è fondamentale, perché stabilisce che il determinante della matrice AA' è uguale alla somma dei quadrati dei determinanti $\delta_m(A)$ di tutte le matrici quadrate di ordine m estraibili da A.
Ritornando alla matrice (12.2) e ricordando la condizione di allineamento di tre punti qualsiasi, espressa dall'annullarsi dei determinanti del terzo ordine (12.1) (dove (x_1, y_1), (x_2, y_2) e (x_3, y_3), indicano tre coppie di valori degli attributi/argomenti, posseduti da tre qualsiasi individui della popolazione), quando il diagramma è rettilineo, tutti i punti si trovano allineati su una retta, tutti i determinanti di terzo ordine formabili con i dati sperimentali risultano nulli ed è nullo anche il determinante della matrice $XX^|$, data da:

$$XX^| = \begin{vmatrix} N & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \\ \sum y_i & \sum x_i y_i & \sum y_i^2 \end{vmatrix} \qquad (i = 1, 2, \dots, N) \qquad (12.5)$$

Viceversa l'annullarsi del determinante della matrice (12.5) porta all'annullarsi di tutti i determinanti di terzo ordine estraibili dalla matrice (12.2). Infatti sempre per la suddetta seconda proprietà, essendo $det(XX^|)$ uguale ad una somma di termini tutti non negativi (perché dei quadrati), il suo valore è nullo, solo se sono nulli tutti i termini della somma (che sono tutti determinanti di tipo (12.1)).
Pertanto il verificarsi della uguaglianza: $det(XX^|) = 0$, è condizione necessaria e sufficiente, affinchè il diagramma dei punti sia lineare e lineare sia anche il legame fra X e Y. Invece il determinante della matrice (12.5), qualora non si verifichi la perfetta dipendenza lineare fra X e Y, non risulta nullo e tuttavia si

ha sempre: $det(XX^{|}) > 0$, essendo una somma di termini tutti positivi. Allora dividendo le singole righe della matrice (12.5) per N, il segno di uguaglianza o disuguaglianza nella relazione: $det(XX^{|}) \geq 0$, non muta, ma i termini di questo determinante assumono significati particolari. Infatti la sommatoria di x_i (con $i = 1,2,...,N$), divisa per N, è la media M_x dei valori argomentali X dalla popolazione; identicamente la sommatoria di y_i, divisa per N, è la media M_y dei valori argomentali Y dalla popolazione. Procedendo poi analogamente per tutti gli altri termini, anch'essi assumono il significato di medie:

$$\frac{1}{N^{3}} det(XX^{|}) = det \begin{vmatrix} 1 & M_x & M_y \\ M_x & M_{x^2} & M_{xy} \\ M_y & M_{xy} & M_{y^2} \end{vmatrix} \geq 0 \qquad (12.6)$$

Sviluppando il determinante ed aggiungendo e sottraendo $M_x^2 M_y^2$ si ottiene:

$$\left(M_{x^2} - M_x^2\right)\left(M_{y^2} - M_y^2\right) - \left(M_{xy} - M_x M_y\right)^2 \geq 0 \qquad (12.7)$$

$$\left(M_{xy} - M_x M_y\right)^2 \leq \left(M_{x^2} - M_x^2\right)\left(M_{y^2} - M_y^2\right) \qquad (12.8)$$

da cui si ricava la seguente disuguaglianza:

$$\frac{\left(M_{xy} - M_x M_y\right)^2}{\left(M_{x^2} - M_x^2\right)\left(M_{y^2} - M_y^2\right)} \leq 1 \qquad (12.9)$$

La radice quadrata del suo primo membro è, per definizione, il coefficiente di correlazione lineare:

$$r = \frac{M_{xy} - M_x M_y}{\sqrt{\left(M_{x^2} - M_x^2\right)\left(M_{y^2} - M_y^2\right)}} \qquad (12.10)$$

La (12.9) risulta soddisfatta se: $|r| \leq 1$, e conseguentemente il coefficiente di correlazione lineare ha un valore sempre compreso tra -1 e $+1$. E' uguale a ± 1, solo quando si ha una perfetta dipendenza lineare, è uguale a zero, in caso di perfetta indipendenza lineare fra X e Y, ed è compreso fra $+1$ e 0, oppure fra 0 e -1, in tutti gli altri casi. Ricordando poi la (5.5), il coefficiente di correlazione lineare ha espressione:

$$r = \frac{M_{xy} - M_x M_y}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \qquad (12.11)$$

Il calcolo del valore M_{xy} è eseguito, usando la seguente espressione:

$$M_{xy} = \frac{1}{N}\sum_{i=1}^{n}\ \sum_{j=1}^{m} x_j y_i f_{ij} = \frac{1}{N}\sum_{j=1}^{m} x_j M_{y_j} p_j \qquad (12.12)$$

A differenza dell'indice di Pearson che è una misura assolutamente generale della dipendenza, fra i due attributi/argomenti della variabile statistica doppia, il *coefficiente di correlazione lineare* ha significato solo quando la dipendenza ha un andamento pressochè rettilineo. Infatti esso è un indice che mette in luce la maggiore o minore approssimazione della retta come rappresentante della correlazione fra gli argomenti della variabile statistica stessa.

In Fig. 1.12.1, sono rappresentati sul piano X, Y casi di variabili statistiche a due dimensioni; i punti rappresentano ciascuno un individuo della popolazione in esame. Per ognuno dei quattro casi illustrati è dato il valore del coefficiente di correlazione r che varia da 0, nel primo caso, ad 1, per l'ultimo. Degno di nota è il fatto che, nel secondo caso, pur essendo evidente la dipendenza fra X e Y, r è nullo, poiché il legame fra i due attributi/argomenti è lungi dall'essere rappresentabile con una retta.

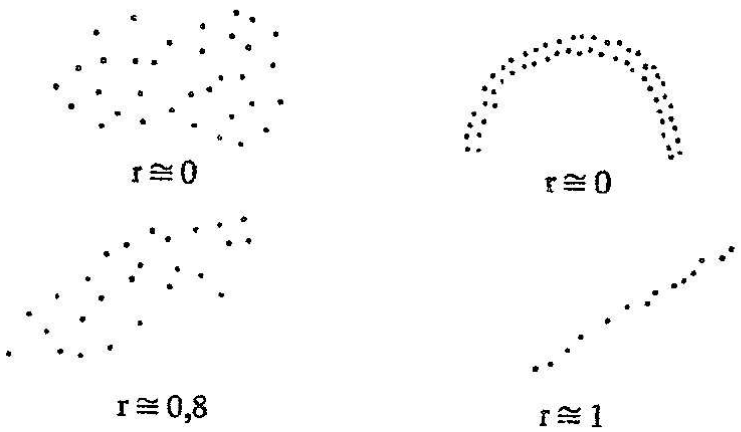


Fig. 1.12.1 – Valori del coefficiente di correlazione r per vari insiemi di punti

Introducendo nella (12.11) la definizione di scarto, si deduce un'altra espressione di r, molto utilizzata:

$$r = \frac{M_{v_x v_y}}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \qquad (12.13)$$

che permette anche di evidenziare il legame intercorrente tra il coefficiente di correlazione lineare ed i coefficienti di regressione lineare. Infatti confrontando la (12.13) con le (11.5) ed (11.6), si osserva che r è la media geometrica di a ed a', i quali sono così uno l'inverso dell'altro, solo quando si ha una perfetta dipendenza lineare $\left(r^2 = 1\right)$.

L'esempio della popolazione di individui, analizzati dal punto di vista della statura X e del peso Y (costituito dai dati contenuti nella tabella di Fig. 1.8.2), si presta anche per il calcolo del coefficiente di correlazione lineare r e dei coefficienti di regressione lineare a ed a'. Infatti può essere accettabile l'ipotesi che, fra statura e peso degli individui, esista un legame da ritenersi lineare, in prima approssimazione. Dati i valori:

$$M_x = 176{,}65 \qquad M_y = 70{,}08$$
$$M_{xy} = 12408{,}68 \qquad \sigma_{xy} = M_{v_x v_y} = 28{,}73$$
$$\sigma_x^2 = 34{,}41 \qquad \sigma_y^2 = 48{,}59$$

da questi, si ricava

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{28{,}73}{34{,}41} = 0{,}83 \qquad a' = \frac{\sigma_{xy}}{\sigma_y^2} = \frac{28{,}73}{48{,}59} = 0{,}59$$
$$r = \frac{M_{xy} - M_x M_y}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{28{,}73}{\sqrt{34{,}41 \times 48{,}59}} = 0{,}70 = \sqrt{a \cdot a'}$$

L'intensità della correlazione lineare esistente fra statura e peso è discreta ed è più forte l'influenza della statura sul peso che viceversa. Infatti il coefficiente direttivo a della retta $v_y = a v_x$ è $0{,}83$, mentre quello a' della $v_x = a' v_y$ è solo $0{,}59$; così le equazioni, in forma completa, delle rette di regressione sono:

$$y - M_y = a(x - M_x) \qquad x - M_x = a'(y - M_y)$$
$$y = ax + (M_y - aM_x) \qquad x = a' y + (M_x - a' M_y)$$
$$y = 0{,}83\, x - 76{,}54 \qquad x = 0{,}59\, y + 135{,}22$$

1.13 Regressione lineare multipla

La regressione semplice esprime la dipendenza lineare di una variabile Y da una variabile X che si considera indipendente. Tuttavia molti fenomeni dipendono da più cause e non possono essere ricondotti allo schema sopra indicato. Ad esempio, se la posizione planimetrica di un punto del coronamento di una diga varia nel tempo, gli spostamenti dipendono certamente dal livello dell'acqua nel bacino artificiale, dalla temperatura ambiente e da altri fattori legati al tipo di struttura, al suo ancoraggio alla roccia dei fianchi, ecc. In questi casi, occorre ipotizzare un legame funzionale tra alcune variabili primarie: $X_1, X_2, \dots, X_n$, che possono anche essere correlate tra loro, ed una variabile dipendente Y, espresso dalla relazione, dove il caso, più semplice e più comune, è quello lineare (sottostante):

$$y = f(x_{.1}, x_2, \dots, x_n) \qquad (13.1)$$

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n \qquad (13.2)$$

Il problema richiede la stima di n parametri a, assegnati m gruppi di valori (con $m > n$), di Y e X_k $(k = 1, 2, ..., N)$. Come nel caso della regressione semplice, è sempre impossibile, a meno di una perfetta dipendenza lineare e dell'assenza di errori di misura, trovare un insieme di valori a_i, tali da soddisfare tutte le m relazioni (13.2), per cui occorre accettare l'esistenza di residui v_i, ottenendo un sistema di equazioni di misura od agli errori:

$$y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + ...a_n x_{ni}) = v_i \qquad (13.3)$$

Tra le infinite soluzioni di questo sistema, si sceglie, come nel caso della regressione semplice, quella che rende minima la somma dei quadrati degli scarti v_i, cioè:

$$\sum v_i^2 = \sum [y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + ...a_n x_{ni})]^2 = min \quad . \qquad (13.4)$$

Anche in questo caso, il minimo si trova annullando tutte le derivate, rispetto ai paramentri a, ottenendo un sistema lineare di n equazioni in n incognite a:

$$\begin{cases} a_0 m + a_1 \sum x_{1i} + ...a_n \sum x_{ni} = \sum y_i \\ a_0 \sum x_{1i} + a_1 \sum x_{1i}^2 + ...a_n \sum x_{1i} x_{ni} = \sum x_{1i} y_i \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_0 \sum x_{ni} + a_1 \sum x_{ni} x_{1i} + ...a_n \sum x_{ni}^2 = \sum x_{ni} y_i \end{cases} \qquad (13.5)$$

Ancora se non interessa il termine additivo a_0, si sostituiscono alle osservazioni originali, gli scarti rispetto alle loro medie: $v_y = y - M(Y)$ e $v_{x_k} = x_k - M(X_k)$, cosicché effettuata, nel sistema (13.5), una traslazione di assi, si elimina la prima equazione e si tolgono i coefficienti di a_0, per cui il sistema diventa:

$$\begin{cases} a_1 \sum v_{x_1}^2 + a_2 \sum v_{x_1} v_{x_2} + ...a_n \sum v_{x_1} v_{x_n} = \sum v_{x_1} v_y \\ a_1 \sum v_{x_2} v_{x_1} + a_2 \sum v_{x_2}^2 + ...a_n \sum v_{x_2} v_{x_n} = \sum v_{x_2} v_y \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_1 \sum v_{x_n} v_{x_1} + a_2 \sum v_{x_n} v_{x_2} + ...a_n \sum v_{x_n}^2 = \sum v_{x_n} v_y \end{cases}$$

Anche il problema della regressione lineare multipla è molto vasto e complicato. Infatti è necessario valutare, oltre alla validità globale dell'equazione di regressione, tra tutte le X, quali maggiormente influenzano la Y e quali, pur ritenute importanti, sono invece superflue, nella costruzione di un modello che descriva il comportamento della Y, in funzione delle variabili X.

PARTE II – ELEMENTI DI CALCOLO DELLE PROBABILITA'

2.1. Legge empirica del caso e teorema di Bernoulli

Le variabili statistiche, dove ad ogni valore numerico dell'argomento x_i, oppure ad ogni coppia di valori x_j e y_i è fatta corrispondere la frequenza con cui compaiono, nella popolazione esaminata, rappresentano un metodo di classificazione di dati, ottenuti tramite un esperimento, ormai concluso. Infatti i dati da classificare provengono dall'aver analizzato, secondo un certo argomento, o più argomenti insieme, tutti gli individui, costituenti una certa popolazione, chiaramente definita. L'esperimento è concluso nel senso che, una volta delimitata la popolazione, il risultato della classificazione, cioè la variabile statistica costruita, è ormai univocamente determinato e, se la classificazione è rifatta, salvo errori, deve portare alla stessa serie di frequenze relative, precedentemente ottenute.

Negli esempi ed in alcuni altri casi, la popolazione su cui si costruisce una variabile statistica ad una o più dimensioni è chiaramente scelta e definita (ad esempio, i pezzi prodotti da una certa macchina in un dato periodo di tempo oppure gli iscritti sui registri dello stato civile di un Comune). Tuttavia molto frequentemente la popolazione è costituita dai risultati di *esperimenti ripetitivi*, come la successione di *teste e croci* ottenute tirando una moneta, il numero di pezzi difettosi contati in scatole di n pezzi prodotti da una fabbrica, le misure ripetute della capacità di un certo condensatore. Un esperimento in cui alcuni animali sono alimentati con diversi prodotti, per determinare le capacità nutritive di questi ultimi, può essere fatto una sola volta sugli stessi animali (tuttavia l'esperimento può essere pensato come il primo di una serie illimitata di esperimenti analoghi su analoghi animali e così anch'esso può considerarsi ripetitivo, cioè eseguito con identiche modalità operative).

I risultati di esperimenti ripetitivi possono a volte essere conosciuti a priori con esattezza, ma ciò solo quando sono sufficientemente note le leggi che li regolano. Ad esempio, sarà possibile sapere a priori il numero di eclissi che un osservatorio astronomico registrerà anno dopo anno, perché si ritiene di conoscere le leggi della meccanica celeste (ed inoltre queste ultime sono sufficientemente semplici per essere utilizzate in calcoli pratici).

La grande maggioranza degli esperimenti ripetitivi porta invece a risultati che non è possibile predire con esattezza. Anche se si cerca, con grandissima cura di tenere, sotto controllo tutti i fattori che possono influenzare tale risultato e di far sì che essi non cambino, da un esperimento all'altro, il risultato può variare da un'osservazione all'altra, in modo irregolare, escludendo ogni possibilità di predizione. Se questo sia dovuto a piccole ed incontrollabili variazioni, nelle condizioni iniziali dell'esperimento, alla complessità delle leggi che lo regolano od all'azione di tanti e piccoli fattori disturbanti, è cosa difficile da stabilire. Tuttavia è certo che i risultati sono soggetti a fluttuazioni, a caso, che non possono essere calcolate a priori.

In questo caso, si parla di esperimenti *casuali* o *stocastici* ed una sistematica registrazione dei risultati di esperimenti stocastici (popolazione di individui) forma il materiale (valori argomentali) per una classificazione statistica dove si associa, ad ogni valore argomentale, la frequenza relativa con cui tale valore si presenta, nella serie di esperimenti fatti.

Il risultato di un esperimento casuale può essere descritto come estrazione, a caso, di un individuo da una popolazione, più vasta, costituita da tutti i risultati ottenibili ripetendo l'esperimento, infinite volte, sempre nelle stesse condizioni. Per *estrazione a caso* si intende un'operazione che abbia tutte le caratteristiche opposte a quelle della scelta. Dell'individuo estratto, non si deve sapere assolutamente nulla prima della sua estrazione, salvo che è uno degli individui della popolazione; non deve esserci nessun elemento di scelta (neppure di natura inconscia), ma deve essere possibile, coi medesimi atti e con le medesime operazioni di estrazione, prendere l'uno o l'altro degli individui, in maniera assolutamente indipendente dalla caratteristica posseduta. In altre parole, in rapporto all'operazione dell'estrazione, gli individui della popolazione devono comportarsi come se tutti esattamente identici. Inoltre l'operazione dell'estrarre, a caso, deve risultare assolutamente indipendente dal valore argomentale, posseduto dall'individuo estratto.

Se non è possibile prevedere i singoli risultati di una sequenza di esperimenti casuali, si può tuttavia notare che l'insieme di tutti i risultati, nonostante il comportamento irregolare di ciascuno di essi, presenta una notevole uniformità.

Per approfondire un po' questa uniformità, sia A l'insieme (o popolazione) di tutti i possibili risultati di un certo esperimento casuale, da ripetersi un gran numero di volte, nelle stesse condizioni operative. Sia B un sottoinsieme di A: quando un particolare risultato (o individuo) è compreso in B, si dice che si è verificato l'evento E, in caso contrario, si dice che l'evento E non si è verificato. Ad esempio, se l'esperimento casuale consiste nel lancio di un dado, la popolazione A dei risultati possibili è costituita dai numeri: $1,2,3,4,5,6$. Una sottopopolazione B può essere quella dei numeri pari: $2,4,6$, e l'evento E (che si può verificare o no) è dato da un valore argomentale pari dell'individuo estratto, a caso, cioè appartenente a B. Ripetendo l'esperimento un numero di volte n, si può determinare il numero f di volte in cui E si verifica e la sua frequenza relativa f/n. Se si determina la frequenza relativa dell'evento E, per valori crescenti di n, in generale, si osserva che essa presenta una marcata tendenza a diventare più o meno costante, al crescere di n, tendendo ad un limite, con l'esperimento ripetuto all'infinito.

Questa affermazione non può essere dimostrata né matematicamente, in quanto si riferisce a fenomeni d'osservazione, né praticamente, perché un esperimento non può essere ripetuto infinite volte. Tuttavia ad ogni evento E, dipendente da un esperimento casuale, è possibile associare un numero P tale che, in una lunga serie di ripetizioni dell'esperimento, la frequenza relativa di E sia praticamente uguale a P.

Questa tipica forma di regolarità statistica, genericamente chiamata *legge empirica del caso*, costituisce la base empirica di ogni teoria statistica e di ogni modello matematico dei fenomeni che presentano questo tipo di regolarità. Pertanto la legge empirica del caso costituisce l'anello di passaggio fra il campo della sperimentazione e quello del ragionamento matematico, di per sé, incapace di provare qualcosa nel dominio dei fatti fisici, ma essenziale allo scopo di descriverli, analizzarli e fare predizioni su questi.

Per formulare una teoria, utile come modello matematico, per i fenomeni dove compare regolarità statistica, insieme ad ogni evento E, si introduce una grandezza P, detta probabilità dell'evento E in quel particolare esperimento casuale. Allora ogniqualvolta la probabilità di un evento E, in un esperimento casuale, è P, si intende che, in una lunga serie di ripetizioni dell'esperimento, è praticamente certo che la frequenza relativa di E sia circa uguale a P.

Tuttavia volendo costruire una teoria, per l'elaborazione logica di modelli matematici di vari fenomeni casuali, una definizione della probabilità, intuitiva e sperimentale, non è conveniente, mentre è necessario definire la probabilità in base alle sue caratteristiche operative, stabilite in modo assiomatico. D'altra parte, affinché una teoria possa funzionare, quale modello di fenomeni casuali, è necessario che gli assiomi di definizione della probabilità siano compatibili anche con la derivazione, fatta come limite della frequenza relativa.

Dati gli eventi: $E_1, E_2, \cdots, E_n$, relativi ad un esperimento (supponendo due eventi E_i ed E_j con $i \neq j$ in mutua esclusione), si ha un nuovo evento: $E_i + E_j$ (con E_i ed E_j eventi indifferenziati), quando l'evento: $S = E_1 + E_2 + \cdots + E_n$, è un evento certo (cioè un evento verificato ad ogni ripetizione dell'esperimento). In questo caso, si dice che P è una funzione di probabilità e $P(E_i)$ la probabilità dell'evento E_i, se valgono i seguenti assiomi.

- $P(E_i) \geq 0$
- $P(E_i + E_j) = P(E_i) + P(E_j)$
- $P(S) = 1$

In pratica, le probabilità degli eventi sono assegnate come:

- frequenze relative ricavate da ripetizioni dell'esperimento;
- rapporto tra il numero di possibili risultati, corrispondenti al verificarsi di E, e quello totale dei risultati;
- misura della fiducia che si ha nel verificarsi di un evento.

Se le probabilità sono ottenute in uno di questi modi, soddisfano i tre assiomi precedenti. Tuttavia lo sviluppo formale della teoria probabilistica non richiede l'interpretazione fisica dei numeri $P(E_i)$, affinché questi numeri soddisfino i tre assiomi.
La natura del concetto di probabilità è ulteriormente chiarita dal teorema di Bernoulli (1713), o legge dei grandi numeri, di cui si dà solo l'enunciato.

In una serie di prove, *un evento che abbia probabilità costante* P *ed in ciascuna di esse si presenti* f *volte*, *la probabilità della disuguaglianza*:

$$\left| P - \frac{f}{n} \right| < \varepsilon \qquad (1.1)$$

con ε *positivo e comunque assegnato, tende a* 1, *al crescere del numero* n *delle prove*.

Il teorema di Bernoulli non è una dimostrazione della legge empirica del caso. Quest'ultima enuncia solo un'osservazione sul comportamento della frequenza relativa nei fenomeni ripetitivi, mentre il teorema di Bernoulli è una relazione analitica fra entità matematiche: la probabilità di un evento e la probabilità di una disuguaglianza. Tuttavia il teorema di Bernoulli aiuta a chiarire il senso in cui, nell'enunciato della legge empirica del caso, è usata la locuzione: *la frequenza relativa* f/n *per* $n \to \infty$ *tende ad un limite*. Infatti non è un limite, nel senso dell'analisi matematica, perché non è possibile, fissato un ε a piacere, determinare un n^* tale che per $n > n^*$ sia verificata la disuguaglianza (1.1). Invece il teorema di Bernoulli dimostra che è possibile, aumentando n, far sì che la probabilità della (1.1) si avvicini ad 1 quanto si vuole: è un nuovo tipo di convergenza verso un limite, regolata dal teorema di Bernoulli, che è detta *convergenza in probabilità*, proprio per questo.

2.2 Probabilità totali e composte

Secondo un'impostazione frequentemente adottata si possono rappresentare tutti i possibili risultati, di un esperimento ripetitivo, con punti in un insieme S, detti *punti campione*, associando a ciascuno la probabilità di essere estratto a caso, cioè di verificarsi in una qualunque prova. Un evento E può essere rappresentato da uno solo o da un insieme di punti di S, come in Fig. 2.2.1.

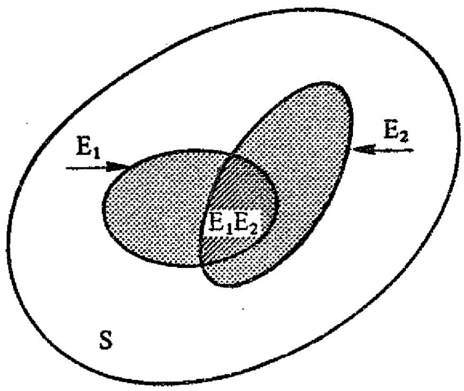


Fig. 2.2.1 – Rappresentazione grafica dei possibili risultati di un esperimento ripetitivo

Secondo il secondo assioma, la probabilità che l'evento E si verifichi è data dalla somma delle probabilità dei punti corrispondenti al verificarsi di E. Tuttavia, se si limita la discussione a insiemi finiti con probabilità uguali per tutti i punti, la definizione generale si può semplificare.

La probabilità che un evento E_1 accada $P(E_1)$ è il rapporto fra il numero dei punti $n(E_1)$, corrispondenti al verificarsi di E_1 ed il numero totale n dei punti di S, cosicché:

$$P(E_1) = \frac{n(E_1)}{n} \tag{2.1}$$

Ad esempio, lanciate due monete, i punti, rappresentanti i possibili risultati sono 4 (Testa – Croce, Croce – Testa, Testa – Testa: Croce – Croce). Allora data a ciascuno di essi la stessa probabilità, la probabilità di ottenere in totale una testa ed una croce è $2/4$, poiché i punti (T,C) e (C,T) corrispondono al verificarsi dell'evento desiderato. Analogamente nel gettare due volte un dado, ammesso che tutti i 36 punti di S abbiano la stessa probabilità, la probabilità che la somma dei numeri usciti sia 7 è $6/36$, poiché tutti i punti $(1,6),(2,5),(3,4),(4,3),(5,2)$ e $(6,1)$ corrispondono al verificarsi dell'evento desiderato.
Le applicazioni della probabilità riguardano spesso un numero di eventi correlati, piuttosto che uno solo. Per semplicità, si considerano due eventi: E_1 ed E_2, associati ad un esperimento. Pertanto potrebbe essere

interessante conoscere, se sia E_1 che E_2 si verifichino, quando si esegue l'esperimento. Questo evento combinato è indicato dal prodotto E_1E_2 e la sua probabilità da $P(E_1E_2)$. D'altra parte, potrebbe anche interessare conoscere, se almeno uno degli eventi E_1 ed E_2 può verificarsi, durante l'esperimento. Questo evento è indicato dalla somma $E_2 + E_2$ e la sua probabilità da $P(E_1 + E_2)$, notando che almeno uno dei due eventi si verifica, se si verifica E_1, ma non E_2, e viceversa, oppure se si verificano entrambi

L'insieme S per un esperimento è indicato dai punti della Fig. 2.2.1 ed i punti campione corrispondenti al verificarsi di E_1 ed E_2 sono i punti interni alle zone contrassegnate rispettivamente con E_1 ed E_2. I punti comuni, a queste due regioni, determinano una regione contrassegnata con E_1E_2 la quale è parte sia della regione E_1 che della regione E_2. Dalla definizione segue che $P(E_1 + E_2)$ è il rapporto fra il numero di punti all'interno di entrambe le regioni ed il numero totale dei punti di S. Tuttavia il numero dei punti campione, all'interno delle due regioni, è uguale al numero di punti nella regione E_1 più quelli nella regione E_2, meno il numero dei punti all'interno della regione comune E_1E_2 (infatti questi punti sarebbero contati due volte, se non si facesse alcuna sottrazione):

$$n(E_1 + E_2) = n(E_1) + n(E_2) - n(E_1E_2).$$

Se entrambi i membri dell'equazione sono divisi per n, numero totale dei punti di S, si ottiene un teorema fondamentale, noto come teorema delle probabilità totali:

$$P(E_1 + E_2) = P(E_1) + P(E_2) - P(E_1E_2). \qquad (2.2)$$

Sovente poi i due insiemi non hanno alcun punto in comune e, quando questo accade, gli eventi sono detti *incompatibili*, perché è impossibile che si verifichi anche l'altro, se uno di essi si verifica. In questo caso, la formula (2.2) si riduce:

$$P(E_1 + E_2) = P(E_1) + P(E_2). \qquad (2.3)$$

Le formule precedenti possono essere estese a più di due eventi.

Per derivare una formula per $P(E_1E_2)$, in termini di probabilità dei singoli eventi, è necessario introdurre la nozione di *probabilità subordinata*, la probabilità che E_2 si verifichi, con la condizione che E_1 si verifica sicuramente. Poiché E_1 è certo solo quando l'insieme S è limitato ai punti giacenti all'interno della regione E_1 (nella Fig. 2.2.1), il numero totale degli eventi è ridotto a $n(E_1)$. Supponendo $n(E_1) > 0$, tra questi punti, i corrispondenti al verificarsi di E_2 sono $n(E_1E_2)$. Di conseguenza, se la probabilità che E_2 si verifichi, subordinatamente al fatto che si verifichi E_1, è indicata con $P(E_2/E_1)$, dalla definizione segue:

$$P(E_2/E_1)=\frac{n(E_1E_2)}{n(E_1)} \quad . \tag{2.4}$$

La probabilità subordinata è spesso chiamata probabilità che E_2 si verifichi essendo noto che si è già verificato E_1, e dividendo per n numeratore e denominatore della (2.4), si ha:

$$P(E_2/E_1)=\frac{P(E_1E_2)}{P(E_1)}$$

da cui:

$$P(E_1E_2)=P(E_1)P(E_2/E_1) \quad . \tag{2.5}$$

La formula (2.5) è valida solo quando $P(E_1)\neq 0$. Tuttavia in generale, la formula è valida, se si dà al secondo membro il valore di 0, quando il fattore $P(E_1)$ è uguale a 0. Se poi l'ordine dei due eventi è invertito, essa diventa:

$$P(E_1E_2)=P(E_2)P(E_1/E_2). \tag{2.6}$$

Infine se i due eventi E_1 ed E_2 sono tali per cui $P(E_2/E_1)=P(E_2)$ e $P(E_1)P(E_2)>0$, allora l'evento E_2 è indipendente dall'evento E_1, in quanto la probabilità del verificarsi di E_2 non è legata alla condizione che si verifichi E_1, cosicché la (2.5) diventa:

$$P(E_1E_2)=P(E_1)P(E_2) \quad . \tag{2.7}$$

Viceversa se la (2.7) è vera, dal confronto della (2.7) con la (2.5), deriva che E_2 è indipendente da E_1. Infatti se i secondi membri della (2.6) e della (2.7) si eguagliano, si ha: $P(E_1/E_2)=P(E_1)$, e l'evento E_1 deve essere indipendente da E_2. Pertanto dato che la (2.7) implica questa indipendenza, si usa definire indipendenti due eventi E_1 ed E_2, se risulta:

$$P(E_1E_2)=P(E_1)P(E_2).$$

<u>Esempio 2.2.1</u>
Estratta due volte una carta, da un mazzo, rimettendo nel mazzo la prima carta, dopo l'estrazione,

- ❑ quale è la probabilità che entrambe le carte siano picche?

Detto E_1 (ed E_2) gli eventi di *estrazione a caso di picche sulla prima* (o *sulla seconda*) *carta*, dato che l'estrazione avviene *con ripetizione*, cioè reinserendo ogni carta nel mazzo originale, le due estrazioni sono indipendenti, cosicché: $P(E_1E_2) = 13/52 \cdot 13/52 = 1/4 \cdot 1/4 = 1/16$.

- Quale è la probabilità che almeno una delle due carte sia picche?

I due eventi non sono incompatibili, cosicché: $P(E_1 + E_2) = 13/52 + 13/52 - 1/16 = 7/16$. Questo problema può anche essere risolto indirettamente, calcolando prima la probabilità che nessuna delle due carte sia picche: $39/52 \cdot 39/52$, e sottraendo questo numero da 1 (infatti il ragionamento *nessuna delle due carte è picche* è opposto a quello *almeno una delle due carte è picche*).

- Se la seconda estrazione è fatta senza ripetizione, cioè senza reinserire la prima carta estratta, nel mazzo, quale è la probabilità che entrambe le carte siano picche?

I due eventi non sono più indipendenti ed anzi si deve calcolare $P(E_2)$ subordinatamente al fatto che si sia già verificato E_1, cioè $P(E_1E_2) = P(E_1)P(E_2/E_1) = 13/52 \cdot 12/51 = 1/17$.

2.3. Variabili casuali

Limitandosi ad esperimenti ripetitivi nei quali i possibili risultati sono costituiti da un insieme discreto di valori, si definisce *variabile casuale* una quantità variabile X, con valori argomentali: $x_1, x_2, \cdots, x_n$, ciascuno dei quali estraibile a caso dalla popolazione, di tutti i valori possibili, con probabilità: $p_1, p_2, \cdots, p_n$, tali che $\sum p_i = 1$. In questo modo, a ciascun esempio ripetitivo (precedentemente presentato) è associabile una variabile casuale.

Ad esempio, il numero per il lancio di un dado è una variabile casuale, con i valori: $1,2,3,4,5$ e 6, ciascuno con probabilità $1/6$, se il dado non è truccato. Analogamente il numero di palle bianche, tra n estratte a caso, da un'urna di composizione nota è una variabile casuale, con valori argomentali $0,1,\ldots,n$ e probabilità calcolabili. Altri esempi di variabili casuali sono l'altezza di un coscritto, appartenente ad una certa classe di leva, il reddito annuo di una famiglia, in una determinata nazione.

In questo contesto, una variabile casuale di particolare interesse è l'errore commesso nella misura di una grandezza. Infatti anch'esso proviene da un'estrazione, a caso, dalla popolazione di tutti gli errori possibili, a ciascuno dei quali è associata, secondo ipotesi particolari, una determinata probabilità. In analogia alla classificazione, precedentemente usata per le variabili statistiche, anche una variabile casuale può essere indicata con la doppia successione di tutti i possibili valori argomentali e della corrispondente distribuzione di probabilità:

$$X \begin{cases} x_1 & x_2 & \cdots & x_n \\ p_1 & p_2 & \cdots & p_n \end{cases} \qquad \sum p_i = 1 \qquad (3.1)$$

Come già detto in precedenza, l' identità formale fra la (3.1) e la (1.2.2), relativa ad una variabile statistica, non deve fare dimenticare che il loro contenuto concettuale è completamente diverso. Infatti la variabile statistica è il risultato di una classificazione effettiva, già eseguita su individui concreti, mentre la variabile casuale è una classificazione ipotetica, semplicemente pensata, non riguardante la composizione di una popolazione (nota o supposta tale), ma piuttosto le operazioni di estrazione, a caso, da farsi su questa stessa popolazione.

2.4. Distribuzioni discrete e continue

Le variabili casuali, usate dei problemi statistici, hanno distribuzioni di probabilità appartenenti a due tipi di distribuzioni: discrete e continue. Negli esempi fatti si sono sempre trattate distribuzioni discrete, cioè tali che la massa unitaria, rappresentante la somma delle probabilità, sia distribuita su di un numero discreto di valori argomentali, presenti in numero finito su ogni intervallo finito.

- Allora la distribuzione discreta della variabile casuale X è completamente descritta se, per ogni i, la probabilità di ottenere $x = x_i$ è p_i (indicato con: $P(x = x_i) = p_i$). Dopodiché si definisce anche una funzione di distribuzione (dove la sommatoria è estesa a tutti i valori j tali che $x_j \leq x_i$):

$$F(x_i) = P(x \leq x_i) = \sum_j p_j \qquad\qquad i = 1 \cdots n \tag{4.1}$$

  La funzione di distribuzione $F(x)$ è una funzione a gradini, costante in ogni intervallo non contenente valori argomentali x, con uno scatto di altezza p_i, in ogni x_i. Nelle applicazioni statistiche, si hanno variabili casuali discrete, ad esempio, quando i valori argomentali sono i numeri naturali $0,1,2,\cdots$.

- Invece una variabile casuale X si dice continua, se la sua funzione di distribuzione $F(x)$ è ovunque continua ed inoltre se la sua derivata prima: $f(x) = F'(x)$, esiste ed è continua nell'intervallo $[A,B]$, di definizione della x, essendo ovunque maggiore od uguale a zero, cosicché si ha:

$$F(x) = P(x_0 \leq x) = \int_A^x f(t)dt \tag{4.2}$$

  dove x_0 è il valore argomentale, presente in un individuo estratto, a caso, dalla X. La funzione $F(x)$ rappresenta la probabilità di estrarre un individuo con valore argomentale x_0 minore di un generico x. Pertanto la distribuzione non possiede valori argomentali isolati e la probabilità di estrarre un particolare valore x_0 è nulla, per qualsiasi valore x_0. Si definisce altresì la probabilità infinitesima: $dp = f(x)dx$, per rappresentare la probabilità di estrarre, a caso, dalla X un individuo il cui valore argomentale sia compreso nell'intervallo $(x, x+dx)$, cosicché la probabilità che la variabile casuale assuma valori compresi in un certo intervallo, finito od infinito, (a,b) risulta:

$$P(a \le x \le b) = F(b) - F(a) = \int_a^b f(t)dt \,. \tag{4.3}$$

Dato che la probabilità, distribuita su tutto il campo di definizione $[A,B]$ della X , è unitaria, si ha:

$$\int_A^B f(t)dt = 1$$

La funzione $y = f(x)$ è nota come *densità di probabilità* (o *funzione di frequenza*, nella terminologia anglosassone). Se si rappresentano graficamente le due funzioni: $y = f(x)$ e $Y = F(x) = \int_A^x f(t)\,dt$, si ottengono due curve analoghe a quelle della Fig. 2.4.1, dove $A = 0$ e $B = +\infty$. Nella prima, la probabilità $P(a \le x \le b)$ è l'area compresa fra la curva, l'asse delle ascisse e le due ordinate, passanti per: $x = a$ e $x = b$. Nella seconda, $P(a \le x \le b)$ è la differenza fra due ordinate: $F(b) - F(a)$.

Nelle applicazioni statistiche, le variabili casuali continue sono associate ad esperimenti ripetitivi, come la misura ripetuta di grandezze, misure che, con alcune riserve, possono assumere qualunque valore (ad esempio, il prezzo di un prodotto, la statura di un uomo, il rendimento di un campo). In questi casi, le variabili casuali sono considerate continue, anche se in realtà i dati sono sempre discreti, poiché ogni misura è espressa da multipli della più piccola unità, registrabile dallo strumento di misura con cui si eseguono le osservazioni. Quando variabili di questo tipo sono considerate continue (a scopo semplificativo), è implicita l'adozione di uno schema teorico non corrispondente perfettamente alla realtà.

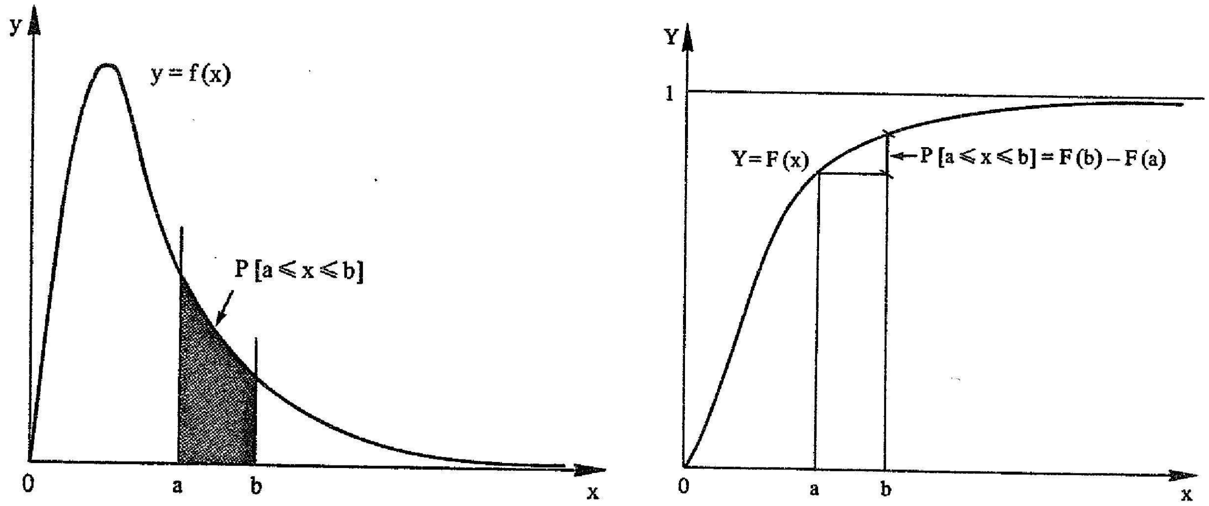


Fig 2.4.1 – Densità di probabilità e distribuzione di una variabile casuale continua

In perfetta analogia formale con le variabili statistiche, anche per le variabili casuali (discrete e continue, queste ultime nel campo di definizione $(-\infty, +\infty)$), si definiscono, i momenti di ordine k , rispetto al polo θ,

- momento di ordine k e polo θ :

$$M_{k,\theta}(X) = \sum_{i=1}^{n} (x_i - \vartheta)^k p_i \qquad M_{k,\theta}(X) = \int_{-\infty}^{+\infty} (x - \vartheta)^k f(x) dx$$

- media:

$$M_x = \sum_{i=1}^{n} x_i p_i \qquad M_x = \int_{-\infty}^{+\infty} x f(x) dx$$

- valore quadratico medio

$$M_2(X) = \sum_{i=1}^{n} x_i^2 p_i \qquad M_2(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx$$

- varianza:

$$\sigma_x^2 = \sum_{i=1}^{n} (x_i - M)^2 p_i \qquad \sigma_x^2 = \int_{-\infty}^{+\infty} (x - M)^2 f(x) dx$$

dove, tra questi indici, è ancora valida la relazione: $\sigma^2 = M_2 - M^2$, ed inoltre la media M è ancora quel particolare valore del polo θ che rende minimo il valore del momento di secondo ordine.
Ancora in perfetta analogia formale con le variabili statistiche, si definisce anche l'indice di asimmetria [16]:

$$\gamma = \frac{\sum_{i=1}^{n} (x_i - M)^3 p_i}{\sigma^3} \qquad \gamma = \frac{\int_{-\infty}^{+\infty} (x - M)^3 f(x) dx}{\sigma^3}$$

Dal punto di vista descrittivo, media e scarto quadratico medio di una variabile casuale hanno lo stesso significato di media e di sqm. nella variabile statistica. L'una è un indice di posizione della variabile sulla scala degli argomenti; l'altro è un indice di dispersione dei valori che la variabile può assumere nell'intorno della media.
Inoltre è ancora molto importante, riferita alla variabile casuale, la disuguaglianza/teorema di Tchebycheff per il calcolo del limite inferiore della probabilità di un valore argomentale di un individuo estratto, a caso, da una popolazione, compreso in un intorno della media di ampiezza $\pm \lambda\sigma$.

Detto M il valore medio di una variabile casuale e σ lo scarto quadratico medio, la probabilità che un individuo estratto a caso dalla popolazione possegga un valore argomentale, contenuto nell'intervallo $[M - \lambda\sigma, M + \lambda\sigma]$, dove λ è un numero qualsiasi maggiore di 1, è superiore od uguale a $1 - 1/\lambda^2$.

[16] E' altresì possibile definire l'indice di curtosi (per studiare il comportamento delle code), facendo uso del momento (pari) del quarto ordine, invece di quello (dispari) del terzo ordine, usato per evidenziare l'asimmetria.

2.5. Sistemi di variabili casuali

In particolari esperimenti ripetitivi, si possono presentare contemporaneamente due o più variabili casuali. Considerando il caso semplice di due sole variabili casuali X e Y discrete, la prima delle quali assume i valori: $x_1, x_2, \cdots, x_n$, con probabilità: $p_1, p_2, \cdots, p_n$, e la seconda i valori: $y_1, y_2, \cdots, y_m$, con probabilità: $q_1, q_2, \cdots, q_m$, durante le prove, X e Y possono assumere tutte le coppie di valori o solo parte di esse: $(x_1, y_1), \cdots, (x_1, y_m), \cdots, (x_n, y_m)$, ciascuna con una determinata probabilità. Il fatto che X assuma un certo valore x_j e Y un certo valore y_i, può essere considerato un evento unico cui spetta una probabilità p_{ij} (che può anche essere nulla, se quei due valori x_j e y_i non possono mai essere accoppiati), che forma una nuova variabile casuale (X, Y), detta variabile casuale doppia, o bidimensionale, così costruita:

$Y \backslash X$	x_1	x_2	$\cdots$	x_j	$\cdots$	x_n
y_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	p_{1n}
y_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	p_{2n}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
y_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	p_{in}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
y_m	p_{m1}	p_{m2}	$\cdots$	p_{mj}	$\cdots$	p_{mn}

(5.1)

Se le due variabili casuali sono indipendenti, si ha, per il teorema sulle probabilità composte:

$$p_{ij} = p_i q_j \tag{5.2}$$

potendo così estendere facilmente le (5.1) e (5.2) al caso di variabili casuali multidimensionali. In questo contesto, sulle variabili casuali $X, Y, ..., ...$ costituenti la variabile casuale multidimensionale, si possono eseguire operazioni, in particolare, somma e prodotto, ottenendo altre variabili casuali la cui distribuzione dipende dalla distribuzione multidimensionale originaria (5.1) e dall'operazione eseguita.

Ad esempio, avendo eseguito una o più volte la misura dei due lati di un rettangolo, alle misure di ciascun lato, può essere associata una variabile casuale X (o Y) dalla quale sono estratti, a caso, i valori numerici x_i (o y_i) determinati. Analogamente i valori numerici, calcolati rispettivamente per il perimetro e l'area del rettangolo, devono essere considerati estrazioni, a caso, dalle variabili casuali: $Z = 2X + 2Y$ e $Z = XY$, le cui distribuzioni sono calcolabili, utilizzando quella della variabile casuale bidimensionale (X, Y), se questa è nota. Più precisamente:

- la variabile casuale: $Z = X + Y + \cdots + W$, somma delle variabili casuali: $X, Y, \cdots, W$, è una variabile casuale in cui:

- la successione dei valori argomentali è data dalla successione ordinata di tutte le somme possibili: $x_i + y_j + \cdots + w_z$ (con $i, j, \cdots, z$ comunque ed indipendentemente variabili nei limiti di definizione di ciascun argomento), purché diverse fra loro;
- la successione delle probabilità, in corrispondenza degli argomenti, è ottenuta calcolando prima quella di ogni valore: $x_i + y_j + \cdots + w_z$, la quale, se le variabili sono indipendenti, è data dal prodotto: $p_i^{(x)} \bullet p_j^{(y)} \bullet \ldots \bullet p_z^{(w)}$, e successivamente sommando le probabilità corrispondenti ad uguali valori di somma,

(infatti nella maggioranza dei casi, diversi gruppi di: $x_i, y_j, \cdots, w_z$, danno origine, combinati per somma, allo stesso valore argomentale z la cui probabilità, per il teorema delle probabilità totali, si ricava sommando le rispettive probabilità);

- la variabile casuale prodotto: $Z = X \cdot Y \cdot \ldots \cdot W$, delle variabili casuali indipendenti $X, Y, \cdots, W$, è una variabile casuale in cui:

  - la successione dei valori argomentali è data dalla successione ordinata di tutti i possibili prodotti $x_i \cdot y_j \cdot \ldots \cdot w_z$ (con $i, j, \cdots, z$ comunque ed indipendentemente variabili nei limiti di definizione di ciascun argomento), purché diversi fra loro;
  - la successione delle probabilità, in corrispondenza degli argomenti, è ottenuta calcolando prima quella di ogni valore: $x_i \cdot y_j \cdot \ldots \cdot w_z$, la quale, se le variabili sono indipendenti, è data dal prodotto: $p_i^{(x)} \cdot p_j^{(y)} \cdot \ldots \cdot p_z^{(w)}$, e successivamente sommando le probabilità corrispondenti ad uguali valori di prodotto.

(infatti nella maggioranza dei casi, anche qui, diversi gruppi di: $x_i, y_j, \cdots, w_z$, danno luogo, combinati per moltiplicazione, allo stesso valore argomentale z la cui probabilità, per il teorema delle probabilità totali, si ottiene dalla somma delle rispettive probabilità).

Il calcolo dei primi momenti delle variabili casuali è ottenuto con le operazioni più semplici.

- Ricordando la definizione di media di una variabile casuale, se i valori del suo argomento sono moltiplicati per una stessa costante, anche il valore medio è moltiplicato per quella costante:

$$M(KX) = KM_x \tag{5.3}$$

- Il valore medio della somma di più variabili casuali è uguale alla somma dei loro valori medi:

$$M(X + Y + \cdots + W) = M_x + M_y + \cdots + M_w \tag{5.4}$$

infatti considerando il caso di due sole variabili casuali X e Y, per definizione, si ha:

$$
\begin{aligned}
M(X+Y) = & \, p_{11}(x_1+y_1)+\cdots+p_{1m}(x_1+y_m)+ \\
& +\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots+ \\
& +p_{n1}(x_n+y_1)+\ldots+p_{nm}(x_n+y_m)= \\
& =x_1(p_{11}+\cdots+p_{1m})+\cdots+x_n(p_{n1}+\cdots+p_{nm})+ \\
& +y_1(p_{11}+\cdots+p_{n1})+\cdots+y_m(p_{1m}+\cdots+p_{nm})
\end{aligned}
\qquad (5.5)
$$

Data la sommatoria $(p_{11}+p_{12}+\cdots+p_{1m})$, per il teorema delle probabilità totali, si ottiene la probabilità di estrarre, a caso, un individuo della popolazione X, in possesso dell'argomento x_1, associato ad un individuo estratto, a caso, dalla popolazione Y, con uno qualsiasi degli m valori dell'argomento y. Questa probabilità è necessariamente uguale alla probabilità p_1 di estrarre, a caso, da X un individuo in possesso dell'argomento x_1, indipendentemente da quello dell'individuo estratto dalla Y. Pertanto estendendo a tutti i valori dell'indice i, nella variabile casuale X, si ha: $p_{i1}+p_{i2}+\cdots+p_{im}=p_i$. Con un ragionamento analogo, applicato alla $(p_{11}+p_{21}+\cdots+p_{n1})$, si ha poi: $p_{1j}+p_{2j}+\cdots+p_{nj}=q_j$, cosicché si ricava, sostituendo queste espressioni nello sviluppo (5.5):

$$
M(X+Y)=(x_1p_1+\cdots+x_np_n)+(y_1q_1+\cdots+y_mq_m)=M_x+M_y \qquad (5.6)
$$

Allora per la proprietà associativa della somma, la dimostrazione, valida per due variabili casuali, può essere estesa anche al caso di un numero qualsiasi di variabili casuali componenti.

- Il valore medio del prodotto di più variabili casuali indipendenti è uguale al prodotto dei loro valori medi:

$$
M(X\cdot Y\cdot\ldots\cdot W)=M_x\cdot M_y\cdot\ldots\cdot M_w \qquad (5.7)
$$

Infatti sempre riferendosi al caso di due sole variabili (che si estende al caso di un numero qualsiasi di variabili casuali componenti, per la proprietà associativa della moltiplicazione) e con la condizione di indipendenza delle variabili casuali componenti strettamente necessaria, si ha:

$$
\begin{aligned}
M(XY) = & \, p_{11}x_1y_1+\ldots+p_{1m}x_1y_m+ \\
& +\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots+ \\
& +p_{n1}x_ny_1+\ldots+p_{nm}x_ny_m= \\
& =p_1x_1q_1y_1+\cdots+p_1x_1q_my_m+ \\
& +\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots+ \\
& +p_nx_nq_1y_1+\cdots+p_nx_nq_my_m= \\
& =(p_1x_1+\cdots+p_nx_n)(q_1y_1+\cdots+q_my_m)=M_xM_y
\end{aligned}
\qquad (5.8)
$$

- Il valore quadratico medio di una somma di più variabili casuali indipendenti, aventi valore medio nullo, è uguale alla somma dei valori quadratici medi delle singole variabili:

$$M_2(X+Y+\cdots+W)=M_2(X)+M_2(Y)+\cdots+M_2(W) \tag{5.9}$$

con: $M_x = M_y = \cdots = M_w = 0$

Ad esempio, date tre variabili casuali X, Y e Z, tra loro reciprocamente indipendenti e soddisfacenti la condizione di media nulla; prima sviluppando il quadrato, poi applicando la proprietà della somma ed infine sostituendo, per la proprietà del prodotto, nell'ipotesi d'indipendenza, al posto dei tre ultimi termini, il prodotto dei singoli valori medi (a loro volta, tutti nulli, per ipotesi), si ha:

$$\begin{aligned} M\left[(X+Y+Z)^2\right] &= M(X^2)+M(Y^2)+M(Z^2)+ \\ &+2M(XY)+2M(XZ)+2M(YZ)= \\ &= M(X^2)+M(Y^2)+M(Z^2)+ \\ &+2M(X)M(Y)+2M(X)M(Z)+2M(Y)M(Z)= \\ &= M(X^2)+M(Y^2)+M(Z^2) \end{aligned}$$

La stessa espressione iniziale (cioè senza effettuare le ultime due sostituzioni) risulta essere invece la formula del valore quadratico medio, quando le variabili casuali componenti non sono indipendenti:

$$\begin{aligned} M\left[(X+Y+Z)^2\right] &= M(X^2)+M(Y^2)+M(Z^2)+ \\ &+2M(XY)+2M(XZ)+2M(YZ) \end{aligned} \tag{5.10}$$

La (5.9) rappresenta la *legge di propagazione degli scarti di variabili casuali indipendenti*, se applicata agli scarti di n variabili casuali *indipendenti*, poiché gli scarti hanno media nulla, per definizione:

$$\sigma^2(X+Y+\cdots+W)=\sigma_x^2+\sigma_y^2+\cdots+\sigma_w^2 \tag{5.11}$$

In generale, la varianza della variabile casuale Z, ottenuta con una combinazione lineare di n variabili casuali *indipendenti*: $Z=a_1X+a_2Y+\cdots+a_nW+b$ (con: $a_1, a_2, \cdots, a_n, b$, coefficienti costanti), è dato, ricordando le proprietà del prodotto per una costante e del valore quadratico medio [17]:

$$\sigma_z^2 = a_1^2\sigma_x^2 + a_2^2\sigma_y^2+\cdots+a_n^2\sigma_w^2 \tag{5.12}$$

La (5.12) è dimostrabile tenendo presente che: $M(X+a)=M_x+a$, essendo inoltre: $\sigma^2(X+a)=\sigma_x^2$.

[17] La (5.11) e la (5.12) sono altresì dette *legge di propagazione della varianza di variabili casuali indipendenti*, rispettivamente per la somma e per una combinazione lineare,.

2.6. Variabili casuali dipendenti

La legge di propagazione degli scarti è valida solo nel caso in cui le variabili casuali sono indipendenti. In caso contrario, occorre introdurre, nelle (5.11) e (5.12), altri termini, per tener conto anche dei doppi prodotti presenti nella (5.10). Introducendo anche per le variabili casuali, il coefficiente di correlazione lineare:

$$r_{xy} = \frac{M_{xy} - M_x M_y}{\sigma_x \sigma_y} = \frac{M(v_x v_y)}{\sigma_x \sigma_y}$$

nell'ipotesi di variabili casuali di scarti, dove si ha: $M_x = M_y = 0$, si ricalcola la (5.11) che cosi si modifica:

$$\sigma^2(X + Y + \cdots + W) = \sigma_x^2 + \sigma_y^2 + \cdots + \sigma_w^2 + 2r_{xy}\sigma_x\sigma_y + \cdots + 2r_{xw}\sigma_x\sigma_w + \cdots. \qquad (6.1)$$

Anche la (5.12) si può poi analogamente modificare ottenendo [18]:

$$\sigma_z^2 = a_1^2\sigma_x^2 + a_2^2\sigma_y^2 + \cdots + a_n^2\sigma_w^2 + 2a_1a_2r_{xy}\sigma_x\sigma_y + \cdots + 2a_1a_nr_{xw}\sigma_x\sigma_w + \cdots \qquad (6.2)$$

La (6.2) esprime la legge di propagazione degli scarti, qualora alcune o tutte le variabili casuali presenti non siano indipendenti fra loro, purché la dipendenza sia di tipo lineare (od almeno approssimabile a dipendenza lineare). Ovviamente poi nelle (6.1) e (6.2), compaiono i coefficienti di correlazione di tutte e solo quelle variabili casuali correlate fra loro.

2.7. Distribuzioni

Le funzioni di distribuzione di alcune variabili casuali sono calcolate con algoritmi matematici, partendo da alcune ipotesi teoriche. Tre queste sono particolarmente notevoli per il loro interesse pratico, il significato nella storia della statistica, in ordine di scoperta, la *Binomiale* (dovuta a J. Bernoulli, 1700 circa), la *Normale* (dovuta a De Moivre, alla fine del '700, ma frequentemente associata a Laplace e Gauss) e la *Poissoniana* (dovuta a S.D. Poisson, 1837).

In un certo senso, queste sono le distribuzioni classiche; altre sono studiate nel corso dell''800, ma solo verso la fine dell'800 sono introdotte altre numerosissime distribuzioni, in qualche modo, tutte derivate dalle tre originali e specialmente dalla Normale.

Le distribuzioni sono da considerare come un modello matematico, per indagare sul comportamento di esperimenti ripetitivi, con regolarità statistiche. Infatti molti fenomeni fisici, biologici, economici e sociali (che si presentano alla indagine non già sotto forma di relazioni rigorose, ma con l'apparentemente labile aspetto di una variabile statistica), pur sotto la loro apparenza di variabilità, a lungo andare, si rivelano dotati di una rigidità di comportamento, propria solo di fenomeni il cui rapporto di casualità risulta essere ben definito. In particolare, questa rigidità di comportamento si manifesta nella persistenza della struttura statistica delle popolazioni di individui esaminate.

[18] Le (6.2) e (6.2) sono dette *legge di propagazione della varianza*, rispettivamente per la somma e per una combinazione lineare, di variabili casuali non indipendenti (o quantomeno non tutte indipendenti fra loro).

Infatti in natura, esistono fenomeni rappresentati da variabili statistiche la cui composizione rimane identica fondamentalmente, anche al variare del tempo, dello sperimentatore e di alcune delle condizioni ambientali, essendo fenomeni di natura statistica che si manifestano sempre con le stesse caratteristiche di variabilità. Ad esempio, l'indagine statistica su una popolazione, sia in maniera diretta che mediante estrazioni, a caso, rivela una certa distribuzione delle frequenze dei valori dell'argomento rappresentabile mediante una curva di frequenza $y = f(x)$. Un'indagine analoga, sul medesimo argomento in altre popolazioni, prodotte da un processo generativo analogo, ma in differenti condizioni ambientali e temporali, porta a formulare altre curve di frequenza: $y = f_1(x)$, $y = f_2(x)$, ecc. che, pur essendo diverse, formano tuttavia una famiglia di curve dello stesso tipo (dove le differenze sono dovute alle variazioni di alcuni parametri caratteristici, rispecchianti diversità contingenti, ma di natura non sostanziale).

Questa tendenza di una certa caratteristica fenomenica a variare, ma sempre con le medesime modalità, in modo che si abbia una fondamentale stabilità della sua variabilità naturale, è chiamata var*iabilità strutturale*, nel senso che la variabilità del fenomeno rappresenta una caratteristica intima e profonda, e così strutturale, del fenomeno stesso. Quasi tutte le scienze, ma soprattutto le scienze biologiche, fisiche ed economiche, conoscono casi di variabilità strutturale nei fenomeni di loro interesse. Ciascuno di questi casi possiede un suo proprio modo di comportarsi, cui corrisponde una particolare famiglia di distribuzioni. In questi casi, il problema statistico fondamentale è determinare, per i diversi fenomeni, la struttura della loro variabilità, stabilendo il tipo di distribuzione che meglio la rappresenta. Dopodiché una volta assodata sperimentalmente la permanenza di questa struttura intima, occorre definire i mezzi logici ed analitici, per ricavare, da poche prove, i valori dei parametri caratteristici della particolare funzione di distribuzione di quel tipo, proprio della variabile casuale su cui sono eseguite le prove.

2.8. Distribuzione binomiale

Dato un esperimento, a caso, dove un evento E ha probabilità *costante* p di verificarsi, ad ogni ripetizione dell'esperimento ed associata, ad ogni ripetizione, una variabile casuale x_i i cui valori argomentali sono *il numero di volte che si verifica l'evento* E, in ogni prova, questo evento E vale solo: 1 o 0, con probabilità rispettivamente: p e $1 - p = q$. Di conseguenza, la variabile casuale x_i può essere cosi rappresentata:

$$x_i \begin{cases} 0 & 1 \\ p & q \end{cases} \qquad \text{dove: } p + q = 1 .$$

Una variabile casuale, più complessa, è data dalla somma di un numero n di queste variabili x_i ed i suoi valori argomentali sono: $X = \sum x_i$, ovvero il *numero delle volte che* E *si verifica in* n *prove* che possono assumere i valori $0,1,2,\cdots,n$ (dove un'ipotesi aggiuntiva precisa che le x_i sono indipendenti fra loro, cioè che le ripetizioni, dell'esperimento considerato, siano indipendenti l'una dall'altra).

La probabilità di un valore argomentale qualunque i della variabile casuale X proviene dal verificarsi di un evento complesso E^*, tale che, in certi i degli n esperimenti eseguiti, si verifica E e, negli altri $n-i$

esperimenti, esso non si verifica. La probabilità di E^*, dato che gli esperimenti sono indipendenti, è $p^i q^{n-i}$. Tuttavia è perfettamente indifferente, per il calcolo della probabilità del valore argomentale i, sapere in quali i prove si sia presentato E ed in quali no, e su n esperimenti esistono tanti gruppi di i elementi, quante le combinazioni di n elementi ad i a i.

Ogni evento E^*, collegato con un particolare gruppo di i esperimenti, ha sempre la stessa probabilità $p^i q^{n-i}$ e, dato che gli eventi E^* sono incompatibili fra loro, si può applicare il teorema delle probabilità totali, per calcolare la probabilità associata al valore argomentale i nella variabile casuale X:

$$P(i)=\binom{n}{i}P(E^*)=\binom{n}{i}p^i q^{n-1} \qquad \text{essendo: } \binom{n}{i}=\frac{n!}{i!(n-i)!} \text{ il binomio di Newton} \tag{8.1}$$

cosicché essendo: $P(0)=q^n$, $P(1)=npq^{n-1}$ e $P(n)=p^n$, per $X=\sum x_i$, si ha:

$$X\begin{cases} 0 & 1 & 2 & \dots & n-1 & n \\ q^n & npq^{n-1} & \binom{n}{2}p^2q^{n-2} & \dots & \binom{n}{n-1}p^{n-1}q & \binom{n}{n}p^n \end{cases} \tag{8.2}$$

La distribuzione di probabilità così derivata è detta *Bernoulliana* o *Binomiale*. Quest'ultima denominazione deriva dalla coincidenza fra la serie delle probabilità ed i termini della potenza n-esima del binomio $(p+q)$.

Per il calcolo della media di X si può applicare la proprietà della somma di variabili casuali, in quanto la X è proprio la somma di n variabili casuali x_i, tutte uguali tra loro:

$$M(X)=M(x_1+x_2+\cdots+x_n)=\sum_{i=1}^{n}M_{x_i}=np \tag{8.3}$$

$$\text{essendo:} \qquad M_{x_i}=\sum_{k=1,2}x_k p_k=1\cdot p+0\cdot p=p$$

Il risultato ottenuto si può esprimere dicendo che, se un evento ha probabilità costante p, in ogni prova, il valore medio del numero di volte che si verifica, in n prove, è np.

Per calcolare la varianza della X, si può costruire la variabile: $y_i=x_i-p$, i cui valori argomentali sono gli scarti fra i valori argomentali di x_i e la media p di x_i. Come sempre, la probabilità di uno scarto è la stessa di quella del valore argomentale da cui tale scarto proviene, perché la variabile casuale, costruita sugli scarti, ha la stessa distribuzione di probabilità della variabile casuale originaria, cosicché:

$$y_i\begin{cases} 1-p & 0-p \\ p & q \end{cases} \qquad \text{dove: } p+q=1$$

Allora il valore quadratico medio di y_i è: $M_2(y_i)=\sigma^2_{y_i}=(1-p)^2 p+(-p)^2 q=pq$, e considerando la variabile casuale: $Y=\sum y_i$, si può applicare ad essa la proprietà del valore quadratico medio, valida in quanto le y_i sono indipendenti ed a media nulla:

$$M_2(Y)=\sigma^2(X)=M_2(y_1)+M_2(y_2)+\cdots+M_2(y_n)=\sum_{i=1}^{n}M_2(y_i)=npq \qquad (8.4)$$

Pertanto, la varianza del numero delle volte che, su n prove, si verifica un evento di probabilità costante p, in ogni prova, risulta: $npq=np(1-p)$.

La distribuzione di probabilità Binomiale dipende da due parametri p e n. Se $p=q=0.5$, la distribuzione è simmetrica, come si verifica calcolando la serie di probabilità della (8.2), indipendentemente dal valore di n. Se p è molto più piccolo di q e n è piccolo la distribuzione è fortemente asimmetrica; (tuttavia a parità di valori di p ed all'aumentare di n, l'asimmetria tende a diminuire, come si può osservare dalla Fig. 2.8.1). Invece a parità di valori di n, si perde presto l'asimmetria della distribuzione, aumentando l'entità di p (come mostra la differenza fra la Fig. 2.8.1, nel caso: $n=10$, e la Fig. 2.8.2, valida sempre per: $n=10$, ma per: $p=1/3$, invece che per: $p=1/10$).

Un altro tipo di variabile casuale, avente distribuzione Binomiale, è la serie dei valori argomentali costituita, non più dal numero di successi di un evento E, in n prove, ma dalle percentuali dì successi, sempre su n prove, $0,1/n,2/n,\cdots,n-1/n,1$. Ovviamente le probabilità associate, a questi valori argomentali, sono le stesse di prima e la variabile casuale elementare x_i', associata ad una prova, è:

$$x_i'\begin{cases} 0 & \dfrac{1}{n} \\ q & p \end{cases} \qquad \text{con: } M(x_i')=\frac{p}{n} \qquad \text{da cui: } M(X')=n\cdot M(x_i')=p$$

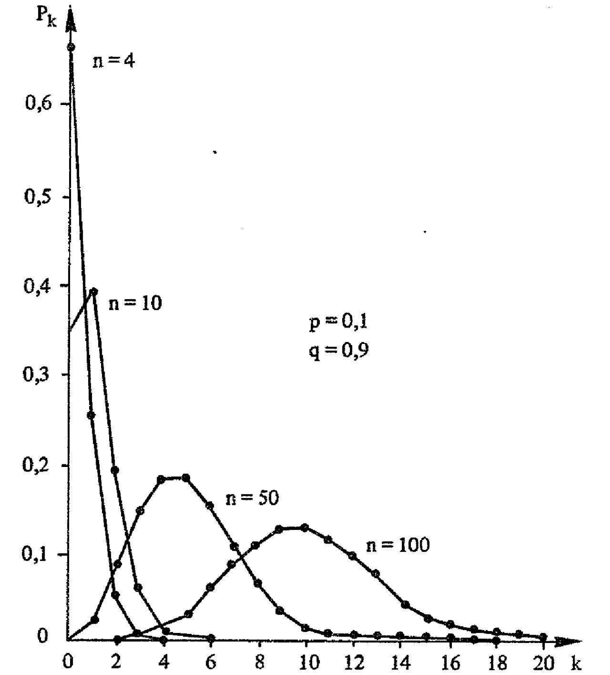
dove: $X'=\sum x'_i$, è ancora la variabile casuale risultante dall'esecuzione di n prove indipendenti. Per il calcolo della varianza di X', si consideri la variabile casuale elementare y_i' degli scarti della x_i' da $M(x_i')$:

$$y_i'\begin{cases} 0-\dfrac{p}{n} & \dfrac{1}{n}-\dfrac{p}{n} \\ q & p \end{cases}$$

ed il suo valore quadratico medio:

$$M_2(y_i')=\sigma^2_{y_i'}=\left(-\frac{p}{n}\right)^2 q+\left(\frac{1-p}{n}\right)^2 p=\frac{1}{n^2}pq$$

cosicché essendo: $Y' = \sum_{i=1}^{n} y_i'$, si ha infine: $M_2(Y') = \sigma^2(Y') = \sum_{i=1}^{n} M_2(y'_i) = n\frac{1}{n^2}pq = \frac{pq}{n}$.



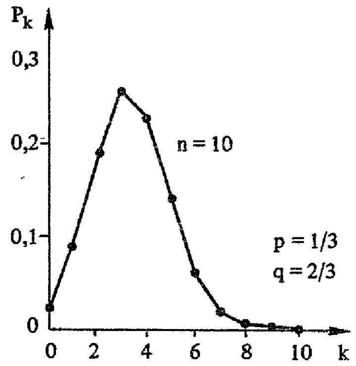


Fig. 2.8.1 – Diagrammi di frequenza per variabili casuali Binomiali con p assegnato e n variabile

Fig. 2.8.2 – Diagramma di frequenza per variabili casuale Binomiale con $p = 1/3$ e $n = 10$

Gli stessi risultati si possono altresì ottenere, in modo più semplice, indicando la variabile casuale delle *percentuali di successi* come: $X' = X/n$, in quanto i valori argomentali della X' sono quelli della X, divisi per n; da cui:

$$M_2(X') = \frac{1}{n}M(X) = \frac{1}{n}np = p \tag{8.5}$$

$$\sigma^2(X') = \frac{1}{n^2}\sigma^2(X) = \frac{1}{n^2}npq = \frac{pq}{n}. \tag{8.6}$$

<u>Esempio 2.8.1.</u>

Se il 12% dei pezzi prodotti da una macchina è difettoso, trovare la probabilità che, su 4 pezzi estratti, a caso, 0, 1 o 2 pezzi, al massimo, siano difettosi.

La probabilità di un pezzo difettoso è: $p = 0{,}12$, e di uno non difettoso: $q = 1 - p = 0{,}88$; di conseguenza:

- ❑ P (0 pezzi difettosi su 4) $= \binom{4}{0}(0{,}12)^0(0{,}88)^4 = 0{,}5997$
- ❑ P (1 pezzo difettoso su 4) $= \binom{4}{1}(0{,}12)^1(0{,}88)^3 = 0{,}3271$

- P (2 pezzi difettosi su 4) $= \binom{4}{2}(0{,}12)^2(0{,}88)^2 = 0{,}0669$

da cui:

P (al massimo 2 pezzi difettosi su 4) $= P$ (0 pezzi difettosi su 4) $+ P$ (1 pezzo difettoso su 4) $+$

$+ P$ (2 pezzi difettosi su 4) $= 0.3271 + 0.5997 + 0.0669 = 0.9937$ [19].

---

[19] La distribuzione Binomiale (o Bernoulliana) spiega il comportamento statistico delle estrazioni di successioni di eventi (tra loro indipendenti), contro il loro non accadere, come la somma di tutte le successioni possibili (tra loro incompatibili). Questa distribuzione è simmetrica, solo se uguali sono le probabilità elementari di avere un evento, oppure no. Inoltre trasformando la variabile casuale somma nella variabile casuale media, è facile mostrare il diminuire della varianza (proporzionalmente alla numerosità del campione). Un esempio di estrazioni di successioni di eventi indipendenti, metrologicamente significativo, sono le prove ripetute, in condizione di indipendenza
Complementare a questa prima distribuzione è la distribuzione Ipergeometrica che spiega il comportamento statistico delle estrazioni di eventi (non tra loro indipendenti, come per un'estrazione in blocco), sempre contro il loro accadere. In questo modo, la variabile è analoga alla precedente, tranne che per la pretesa od imposta modificazione della popolazione universo, dopo ogni singola estrazione. Nel caso di prove ripetute, la distribuzione è utile, se si può imporre il non accadere di certe determinate successioni (ad esempio, perché considerate indesiderate).

Legenda:	N	dimensione della popolazione universo
	r	eventi favorevoli nella stessa popolazione
	n	dimensione del campione
	k	eventi favorevoli nello stesso campione

Densità discreta di probabilità: $p_k = \dfrac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}$ con $k = 0 \div n$

Media: $\mu = np$

Varianza: $\sigma^2 = np(1-p)\dfrac{N-n}{N-1}$

La distribuzione Geometrica (derivata direttamente dalla distribuzione Binomiale o Bernoulliana) spiega il tempo di ritorno di un evento, spesso desiderato (cioè la probabilità di avere, per la prima volta, quell'evento, spesso considerato un successo), dopo un certo numero di non accadimenti dello stesso evento (a loro volta, considerati un susseguirsi di insuccessi). Nel caso di prove ripetute, questa variabile casuale serve a modellare l'attesa di un certo risultato, a fronte del suo non accadere.

Legenda: k eventi non favorevoli nello stesso campione

Densità discreta di probabilità: $p_k = (1-p)^k p$ con $0 < p < 1$
e $k = 0 \div \ldots$

Media: $\mu = \dfrac{1-p}{p}$

Varianza: $\sigma^2 = \dfrac{1-p}{p^2}$

La distribuzione Binomiale negativa spiega l'accadere di eventi rari non equiprobabili (come si ha per la distribuzione Poissoniana, presentata appena oltre, ma a differenza di questa, ancora derivata direttamente dalla distribuzione Binomiale o Bernoulliana). Nelle prove ripetute, questa variabile casuale è utile a modellare l'accadere di dati anomali (e in metrologia, la maggior parte dei dati anomali è ascrivibile ad errori grossolani).

Legenda:	r	eventi favorevoli nel campione
	k	eventi non favorevoli nello stesso campione

Densità discreta di probabilità: $p_k = \binom{r+k-1}{k}(1-p)^k p^r$ con: $0 < p < 1$
e $k = 0 \div \ldots$

2.9. Distribuzione Poissoniana

Un altro tipo di distribuzione discreta, detta anche *distribuzione degli eventi rari*, sottolinea il fatto che, in ciascuno caso cui essa si applica, la probabilità che si verifichi il singolo evento è estremamente bassa. La distribuzione di Poisson serve così da modello matematico per fenomeni tipo: il numero di particelle emesse da una sostanza radioattiva (in un certo intervallo di tempo), il numero di suicidi in un mese, il numero di bombe cadute su ogni mezzo km quadrato della città di Londra nell'ultima guerra, il numero di automobili che si fermano ad una stazione di servizio (in un certo intervallo di tempo), il numero di incidenti accaduti in ogni giorno agli operai di una fabbrica, ecc. Dagli esempi citati, in una distribuzione di Poisson, gioca un ruolo determinante la scelta dell'*unità di osservazione*, che può essere indifferentemente spaziale o temporale, come il campo di un microscopio od il periodo di un'ora.

I valori argomentali della variabile casuale di Poisson sono il numero di eventi, verificati nell'intervallo di osservazione, e la sequenza di possibili valori sono i numeri naturali da: 0 a $+\infty$. Ovviamente dato che si tratta di eventi rari, sono più frequenti i bassi valori argomentali: $0,1,2,3,4,5$ e pochi più. Il tipo e la forma della distribuzione sono determinati da un unico parametro, ossia dal valore medio del numero di eventi che si presentano nell'unità di tempo (o di spazio) di osservazione (in generale, indicato con λ), che permette di calcolare la serie delle probabilità da associare a ciascuno dei valori argomentali.

Il calcolo del generico P_k, cioè della probabilità del valore argomentale k, è ricavato basandosi sui requisiti propri dalla variabile casuale Poissoniana:

- stazionarietà (la probabilità di un evento, in un qualsiasi intervallo infinitesimo $(t, t+h)$, deve essere approssimativamente λh);
- non molteplicità (la probabilità di due o più eventi, in un intervallo infinitesimo, è di ordine superiore a λh, esprimendo così il fatto che l'evento è davvero raro);
- indipendenza (in ogni intervallo finito, tutti gli eventi si presentano indipendentemente uno dall'altro ed il numero di eventi, verificatisi in qualsiasi intervallo, è indipendente dal numero degli eventi in ogni altro intervallo, non sovrapposto).

Tuttavia poiché la derivazione di P_k, sulla base delle caratteristiche proprie della distribuzione Poissoniana, è piuttosto onerosa, di solito, si preferisce considerarla come caso limite della distribuzione Binomiale, qualora si abbia $p << 1$ e np non molto grande anche per valori di n elevati.

Analiticamente non si ipotizza più che, in ogni prova, p sia costante, dovendo invece essere costante il valore medio np, cosicché: $p = \frac{\lambda}{n}$, essendo: $\lim_{n\to\infty} np = \lambda$. Dalla (8.1), si pone: $p = \frac{\lambda}{n}$ e $q = 1 - \frac{\lambda}{n}$, e poi, passando al limite, si ha:

Media: $\mu = r\dfrac{1-p}{p}$

Varianza: $\sigma^2 = r\dfrac{1-p}{p^2}$

Tutte queste distribuzioni sono asimmetriche e lo riempimento della coda a destra è maggiore di quello della distribuzione Normale (cui tendono comunque le distribuzioni Ipergeometrica e Binomiale negativa, al tendere all'infinito dei parametri rispettivamente N e r).

$$P_k = \lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} = \lim_{n\to\infty}\frac{n(n-1)\cdots(n-k-1)}{k!}\cdot\frac{\lambda^k}{n^k}\cdot\left(1-\frac{\lambda}{n}\right)^{n-k} =$$
$$= \lim_{n\to\infty}1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\cdot\frac{\lambda^k}{n}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k} = \frac{\lambda^k}{k!}e^{-\lambda} \tag{9.1}$$

per ogni $k = 0,1,2,\cdots$

La somma di tutti i P_k è ovviamente 1, in quanto:

$$\sum_{r=0}^{\infty}\frac{\lambda^r}{r!}e^{-\lambda} = e^{\lambda}e^{-\lambda} = 1 \tag{9.2}$$

La variabile casuale Poissoniana ha forma:

$$X\begin{cases} 0 & 1 & 2 & \ldots & k & \ldots \\ e^{-\lambda} & \lambda e^{-\lambda} & \dfrac{\lambda^2}{2!}e^{-\lambda} & \ldots & \dfrac{\lambda^k}{k!}e^{-\lambda} & \ldots \end{cases} \tag{9.3}$$

La (9.3) dipende da un solo parametro λ, per ipotesi, valore medio della variabile casuale X. Infatti:

$$M_x = \sum_0^{\infty}kP_k = \sum_0^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda\sum_0^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} = \lambda$$

dato che, ponendo: $k-1=r$, e ricordando: $(-1)! = (-2)! = \pm\infty$, risulta:

$$\sum_0^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} = \frac{\lambda^{-1}}{(-1)!}e^{-\lambda} + \sum_{r=0}^{\infty}\frac{\lambda^r}{r!}e^{-\lambda} = 0+1=1$$

Per il calcolo della varianza σ^2, ponendo poi nuovamente: $r=k-1$ e $s=k-2$, ed essendo sempre: $\sigma^2 = M_2 - M^2$, si ha:

$$M_2 = \sum_0^{\infty}k^2P_k = \sum_0^{\infty}(k(k-1)+k)P_k = \sum_0^{\infty}k(k-1)P_k + \sum_0^{\infty}kP_k =$$
$$= \sum_0^{\infty}k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} + \sum_0^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda^2\sum_0^{\infty}\frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda} + \lambda\sum_0^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}$$

dove: $\sum_0^{\infty}\frac{\lambda^r}{r!}e^{-\lambda} = \sum_0^{\infty}\frac{\lambda^s}{s!}e^{-\lambda} = 1$, da cui: $M_2 = \lambda^2 + \lambda$, ed infine:

$\sigma^2 = M_2 - M^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$ cosicché: $\sigma^2 = M_x$

In una distribuzione Poissoniana, l'uguaglianza fra media e varianza è una particolarità da tenere presente, come criterio grossolano, per verificare, se una serie di dati sperimentali sia rappresentabile (o meno) con la distribuzione di Poisson. Allora il primo controllo è verificare, se la media e la varianza, dei dati sperimentali, non differiscono troppo fra loro [20].

L'andamento grafico della (9.3) dipende dal valore λ: per $\lambda < 1$, la distribuzione è fortemente asimmetrica e zero-modale, mentre per $\lambda > 1$, la distribuzione è sempre meno asimmetrica, come risulta dalla Fig. 2.9.1. Inoltre si osserva (potendosi anche calcolare dalla formula generica di P_k) che, quando λ è un numero intero, maggiore di uno, sono uguali le probabilità dei valori argomentali: $k = \lambda$ e $k = \lambda - 1$.

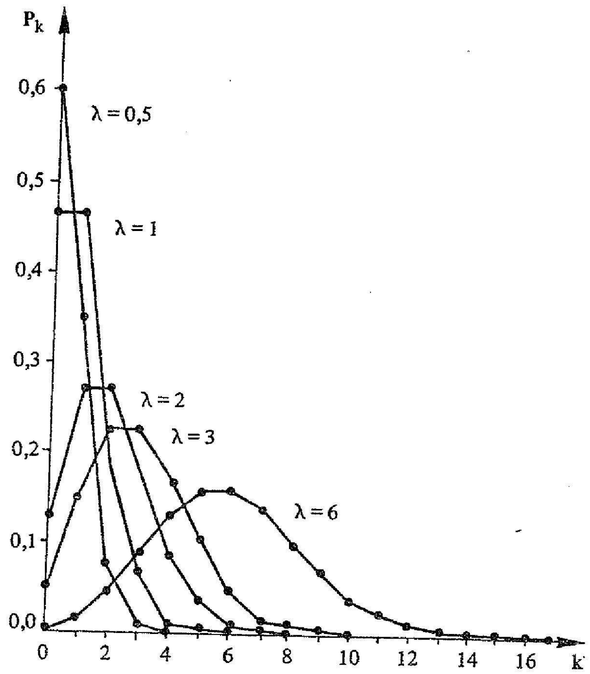


Fig. 2.9.1 – Diagramma di frequenza della distribuzione di Poisson per diversi valori di λ

La distribuzione di Poisson può essere utilizzata, in sostituzione della distribuzione Binomiale, se $p \leq 0.05$ e $n \geq 100$, ponendo $\lambda = np$. Infatti in tal caso, il calcolo di P_k, con la (8.1), diventa molto gravoso, mentre

[20] La variabile casuale Log-normale fornisce un'approssimazione continua per la modellazione di eventi rari (anche se la modellazione rigorosa è data dalla distribuzione Gamma di Erlang):

Funzione densità di probabilità: $p(x) = \frac{1}{\sqrt{2\pi}ax} e^{-\frac{(ln(x-b))^2}{2a^2}}$

Media: $M(x) = e^{\mu + \frac{\sigma^2}{2}}$

Varianza: $\sigma^2 = M_2 - M^2 = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$

Trasformazione alla variabile casuale Normale: $z = ln\, x$ e $p(z) = x\; p(x)$

Come per la distribuzione di Poisson, anche la distribuzione Log-normale è asimmetrica, con coda a destra molto vuota (pur tendendo alla distribuzione Normale ed alla simmetria, al tendere a zero del parametro a).

è più agevole con la (9.1). Tuttavia tranne in questa particolare applicazione, tra l'altro, utile per semplificare i calcoli della dimostrazione della (9.1), la distribuzione di Poisson non ha altre relazioni con la distribuzione Binomiale. Infatti i due parametri n e p non compaiono nella distribuzione Poissoniana, così come cambia completamente il significato delle serie dei valori argomentali.

<u>Esempio 2.9.1</u>

Data la tabella con il numero k di incidenti automobilistici giornalieri, avvenuti in una città, in un periodo di 100 giorni, trovare le frequenze teoriche degli incidenti, se il fenomeno seguisse la distribuzione di Poisson.

$k =$ *n. di incidenti giornalieri*	$f =$ *n. di giorni*
0	45
1	37
2	11
3	5
4	1
5	1
	$\sum f = n = 100$

Il numero medio di incidenti giornalieri è:

$$\lambda = \frac{\sum kf}{n} = \frac{0 \cdot 45 + 1 \cdot 37 + 2 \cdot 11 + 3 \cdot 5 + 4 \cdot 1 + 5 \cdot 1}{100} = 0{,}83$$

Secondo la distribuzione di Poisson

$$P(k \quad incidenti \quad giornalieri) = \frac{(0{,}83)^k}{k!} e^{-0{,}83}$$

In tabella, sono segnate le probabilità ottenute, per $k = 0,1,2,\cdots,5$, e le frequenze teoriche, cioè il numero previsto di giorni in cui dovrebbero avvenire $k = 0,1,2,\cdots,5$ incidenti (la varianza della variabile statistica è: $\sigma^2 = 0{,}98$, in buon accordo con la media: $0{,}83$, e pertanto la corrispondenza fra i dati sperimentali e quelli teorici calcolati in base ad una distribuzione di Poisson è evidentemente molto buona)

n. di incidenti giornalieri, k	*probabilità*	*frequenze teoriche*	*frequenze osservate*
0	0.4360	$43.60 \cong 44$	45
1	0.3619	$36.19 \cong 36$	37
2	0.1502	$15.02 \cong 15$	11
3	0.0416	$4.16 \cong 4$	5
4	0.0086	$0.86 \cong 1$	1
5	0.0014	$0.14 \cong 0$	1

2.10. Distribuzione normale

Le distribuzioni Binomiale (o Bernoulliana) e Poissoniana diventano inutilizzabili rispettivamente per valori di n molto grandi e per alti valori di λ. Infatti essendo queste distribuzioni di tipo discreto, la loro funzione di distribuzione si può calcolare solo dopo aver calcolato le probabilità dei singoli valori argomentali. Il calcolo di queste probabilità diventa estremamente oneroso, quando n, λ e k sono molto grandi:

$$P_k = \binom{n}{k} p^k q^{n-k} \qquad \text{e} \qquad P_k = \frac{\lambda^k}{k!} e^{-k}$$

.

Allora dalla prima, introducendo al posto dei valori argomentali k i loro scarti dalla media: $\nu = k - np$, con la stessa distribuzione, e supponendo grande il numero n ed abbastanza piccolo ν, perché sia: $\nu < np$ e nq, si può calcolare il valore di $P(\nu)$, con una formula approssimata, cioè sostituendo i fattoriali presenti con la formula di Stirling: $m! = m^m e^{-m} \sqrt{2\pi m}\left(1 + E_m\right)$, dove il termine E_m tende a zero al crescere di m. Con questa espressione e di qualche sviluppo in serie troncato, si arriva all'espressione:

$$P(\nu) = \frac{h}{\sqrt{2\pi npq}} e^{-\frac{\nu^2}{2npq}} \tag{10.1}$$

nota come *formula di De Moivre*. Dopodiché ponendo: $h = 1/\sqrt{2npq}$, si ha:

$$P(\nu) = \frac{h}{\sqrt{\pi}} e^{-h^2\nu^2} = P(np) e^{-h^2\nu^2} \tag{10.2}$$

Se n è molto grande ed anche k, pure ν può assumere un grande numero di valori, cosicché non si commette un errore sensibile, considerando continua la variabile casuale con la distribuzione di probabilità data dalla (10.2). Tuttavia nelle distribuzioni continue, non si definisce la probabilità P di un singolo scarto (che è sempre nulla), ma solo la probabilità infinitesima dp di ottenere uno scarto compreso in un intervallo infinitesimo $d\nu$:

$$dp = \frac{h}{\sqrt{\pi}} e^{-h^2\nu^2} d\nu \tag{10.3}$$

Questa espressione, derivata nell'ipotesi che ν sia piccolo, è assunta per qualunque valore di ν cioè per tutti i valori del campo: $-\infty < \nu < +\infty$, perché diventa molto piccola la probabilità, quando ν è grande.

Pertanto le espressioni trovate sono solo approssimate e non è fuori luogo verificare, se siano soddisfatte due condizioni essenziali (ovvero essere uguale ad uno la probabilità totale e nulla la media degli scarti), affinché l'equazione:

$$y = \frac{h}{\sqrt{\pi}} e^{-h^2 \nu^2} \tag{10.4}$$

rappresenti effettivamente la densità di probabilità di una variabile casuale costituita da scarti:

$$\int_{-\infty}^{+\infty} \frac{h}{\sqrt{\pi}} e^{-h^2 \nu^2} d\nu = 1 \qquad \text{e} \qquad M_\nu = \int_{-\infty}^{+\infty} \nu \frac{h}{\sqrt{\pi}} e^{-h^2 \nu^2} d\nu = 0$$

La prima espressione si modifica, ponendo: $\xi = h\nu$, ed essendo: $d\xi = hd\nu$:

$$\int_{-\infty}^{+\infty} \frac{h}{\sqrt{\pi}} e^{-h^2 \nu^2} d\nu = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\xi^2} d\xi = \frac{1}{\sqrt{\pi}} \sqrt{\pi} = 1$$

Inoltre M_ν è nullo, perché la funzione è dispari e l'intervallo di integrazione è simmetrico rispetto allo zero. Di conseguenza, la (10.4) rappresenta la funzione densità di probabilità di una nuova variabile casuale, detta *Gaussiana o Normale*, i cui valori argomentali possono variare da: $-\infty$ a $+\infty$, con continuità, simmetrica rispetto all'asse y , con valore massimo uguale a: $h/\sqrt{\pi}$, per $\nu = 0$.

La varianza ed il momento del secondo ordine di questa variabile casuale coincidono e, con la solita sostituzione: $\xi = h\nu$, integrando per parti, si ha:

$$\begin{aligned} \sigma^2 = M_2 &= \frac{h}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \nu^2 e^{-h^2\nu^2} d\nu = \frac{h}{h^2\sqrt{\pi}} \int_{-\infty}^{+\infty} \xi^2 e^{-\xi^2} d\xi = \frac{1}{h^2\sqrt{\pi}} \int_{-\infty}^{+\infty} \xi \left(\xi e^{-\xi^2}\right) d\xi = \\ &= \frac{1}{h^2\sqrt{\pi}} \left\{ \left[-\frac{1}{2} e^{-\xi^2} \xi \right]_{-\infty}^{+\infty} + \frac{1}{2} \int_{-\infty}^{+\infty} e^{-\xi^2} d\xi \right\} = \frac{1}{h^2\sqrt{\pi}} \frac{1}{2} \sqrt{\pi} = \frac{1}{2h^2} \end{aligned} \tag{10.5}$$

La relazione trovata: $\sigma^2 = 1/2h^2$, giustifica il nome di misura di precisione dato alla costante h: che è inversamente proporzionale allo sqm, in quanto, per piccoli valori di σ, cioè per valori argomentali poco dispersi intorno alla media, si ha un elevato valore di h , ovvero un'elevata precisione.

Questa formulazione deriva da quella originale di Gauss che introduce la distribuzione normale non come limite di quella binomiale, ma come una distribuzione (di probabilità) alla quale devono ubbidire gli errori di misura di una grandezza, per far sì che la media aritmetica di una serie di misure sia la stima, più probabile, della misura vera.

La (10.4) può essere riscritta, nella sua forma più generale, introducendo due parametri: media e varianza, da cui dipende ogni distribuzione Normale:

$$y(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-M)^2}{2\sigma^2}} \tag{10.6}$$

avente il massimo nel punto: $x = M$ (e potendo altresì dimostrare che la y ha due flessi nei punti: $M + \sigma$ e $M - \sigma$). La famiglia delle curve normali, con la stessa media, si dispone come in Fig. 2.10.1.

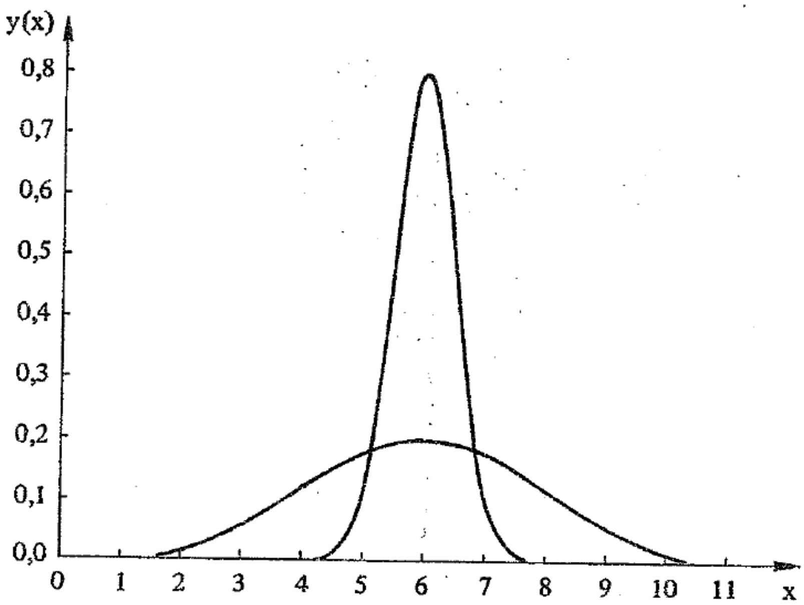


Fig. 2.10.1 – Rappresentazione grafica della densità di probabilità di due variabili casuali normali (con la stessa media e diversa varianza)

La funzione di distribuzione ha espressione (il cui andamento è mostrato in Fig. 2.10.2):

$$F(x) = \int_{-\infty}^{x} y(x)dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(x-M)^2}{2\sigma^2}} dx \qquad (10.7)$$

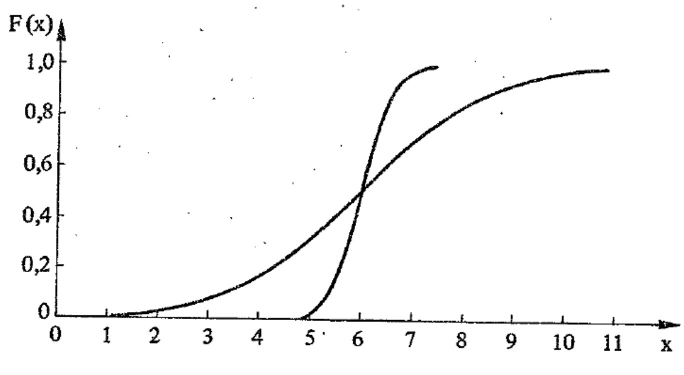


Fig. 2.10.2 – Curva di distribuzione per le due v.c. della Fig. 2.10.1

Tuttavia l'espressione più comune della densità di probabilità di una distribuzione Normale è ottenuta con un cambiamento di variabile, introducendo il cosiddetto *scarto standardizzato* o *unità standard*:

$$z = \frac{x - M}{\sigma} \qquad \text{e} \qquad x = M + z\sigma \qquad (10.8)$$

per cui la (10.6) diventa:

$$y(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

La variabile casuale z ha media nulla e varianza unitaria, ed è chiamata variabile casuale *normale standardizzata*, in quanto in essa non compaiono più i due parametri caratteristici: media e varianza. Il passaggio dalla distribuzione Normale generica a quella standardizzata, tramite la trasformazione (10.8), implica una traslazione lungo l'asse x ed una riduzione di scala da 1 a $1/\sigma$. La corrispondente funzione di distribuzione ha espressione:

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} dz \qquad (10.9)$$

è tabulata (come la corrispondente funzione densità di probabilità) e fornisce la probabilità di valori argomentali z compresi fra due limiti qualsiasi. Infatti se, data una variabile casuale Normale con media M e sqm σ, si vuole ottenere la probabilità di valori argomentali: $a \leq x \leq b$, bisogna ricavare:

$$z_1 = \frac{a-M}{\sigma} \qquad \text{e} \qquad z_2 = \frac{b-M}{\sigma}$$

e la probabilità cercata è: $F(z_2) - F(z_1)$. In particolare, se si pone: $z_1 = -z_2 = \lambda = (x-M)/\sigma$, si deduce che, in qualunque distribuzione Normale, con media M e sqm σ, la probabilità di valori argomentali x che differiscono dalla media: $\pm\lambda\sigma$, è rispettivamente:

$$\begin{aligned}
\lambda = 1 \quad & F(1) - F(-1) = 1 - 2F(-1) = 0.6826 \\
\lambda = 2 \quad & F(2) - F(-2) = 1 - 2F(-2) = 0.9544 \\
\lambda = 3 \quad & F(3) - F(-3) = 1 - 2F(-3) = 0.9974
\end{aligned}$$

Questo significa che, in qualunque distribuzione Normale, si ha il 68% di probabilità di ottenere valori argomentali compresi fra: $(M-\sigma) \leq x \leq (M+\sigma)$, il 95% di probabilità di ottenere valori argomentali compresi fra: $(M-2\sigma) \leq x \leq (M+2\sigma)$, ed il 99.7% di probabilità di ottenere valori argomentali compresi fra: $(M-3\sigma) \leq x \leq (M+3\sigma)$, come mostrato dall'andamento della probabilità, in funzione di λ, riportato graficamente nella Fig. 1.6.1 (con una linea tratteggiata).

Nel caso di risultati di misure, spesso è invalso l'uso di associare, ad ogni valore argomentale A (acquisito sperimentalmente o dedotto teoricamente), l'entità B dello sqm della distribuzione Normale della variabile casuale cui A appartiene, con la notazione: $A \pm B$.

Esempio 2.10.1

Misurando lo spessore di 100 fili di quarzo, costruiti per un particolare tipo di strumento di precisione, nell'ipotesi di variabilità dello spessore rappresentabile con la distribuzione Normale, determinare l'equazione della densità di probabilità di questa distribuzione e le frequenze teoriche per le diverse classi.

Le prime due colonne della tabella rappresentano una variabile statistica di cui si calcolano media e varianza, ottenendo: $M = 49{,}27$ e $\sigma = 2.97$.

Dopodiché i dati sono organizzati in precedente tabella, cercando il corrispondente valore in unità standard: $z = (x - 49.27)/2.97$, per ogni valore limite di classe. Le aree sono quelle sotto la curva normale fra z_i e z_j, e rappresentano la probabilità di ottenere valori argomentali standardizzati nell'intervallo: $\lfloor z_i, z_j \rfloor$.

spessore (μm)	*frequenze osservate*	*limiti delle classi*	*limiti in unità standard z*	*aree fra* 0 *e* z	*frequenze teoriche*	*frequenze osservate*
		41.5	-2.62	0.4956		
$42-44$	6				0.0493	4.93
		44.5	-1.61	0.4463		
$45-47$	20				0.2205	22.05
		47.5	-0.60	0.2258		
$48-50$	40				0.3849	38.49
		50.5	0.41	0.1591		
$51-53$	27				0.2631	26.31
		53.5	1.42	0.4222		
$54-56$	7				0.0703	7.03
		56.5	2.43	0.4925		

Queste probabilità, moltiplicate per 100, numero totale di individui esaminati, danno allora le frequenze teoriche, corrispondenti all'ipotesi di distribuzione Normale, degli spessori del filo intorno al suo valore medio: $x = 49.27\mu m$. Pertanto l'espressione della densità di probabilità è:

$$y = \frac{1}{\sqrt{2\pi}(2{,}97)} e^{-\frac{(x-49{,}27)^2}{2(2{,}97)^2}}$$

Un controllo successivo su buon adattamento delle frequenze teoriche a quelle osservate fa riferimento a problemi di inferenza statistica (cui si rinvia).

Esempio 2.10.2

Volendo disegnare la curva normale relativa ai dati dell'Esempio 1.2.2., occorre sovrapporre all'istogramma (nelle due forme di Fig. 1.3.3) la curva normale, con la stessa media e lo stesso sqm della popolazione.

La media della popolazione è $M = 67.95\,kg$ e lo sqrn è $\sigma = 7.62\,kg$. Questi dati sono calcolati sui valori argomentali, non ancora raccolti in istogramma, ma potrebbero essere calcolati anche dopo, usando come valori argomentali i punti medi di ciascuna classe ed associando, come frequenza, l'altezza della rispettiva classe.

Nelle Figure 2.10.3 a) e b), sono riportati gli stessi istogrammi (in scala più grande). In particolare, sono disegnate e sovrapposte la scala x (in kg) e quella z (unità standard). Lo zero ed il punto 1 della seconda corrispondono, ovviamente, a M e $M+\sigma$ della prima. Nella scala delle ordinate, associata x, sono poste due graduazioni, per unità di misura: le frequenze totali (a destra) e le frequenze relative (a sinistra). Un confronto fra le scale della Fig. 2.10.3 a) e b) chiarisce in che modo entri in gioco l'ampiezza dell'intervallo di classe dell'istogramma (in questo caso, rispettivamente di $2\,kg$ e $4\,kg$. Infine la scala y è quella alla quale sono riferiti i punti della curva normale le cui ordinate $y(z)$ sono ricavate dalle tavole della funzione densità di probabilità. La relazione fra $y(x)$ e $y(z)$ si ricava ricordando che, prima e dopo la standardizzazione, deve rimanere invariata la probabilità infinitesima: $dp = y(x)\,dx = y(z)\,dz$, e poiché: $z = (x-M)/\sigma$ e $dx = \sigma\, dz$, risulta infine: $y(z) = \sigma\, y(x)$.

Le tavole della distribuzione Normale possono essere utilizzate anche per le distribuzioni binomiali in cui n è elevato e p non troppo piccolo, cosicché si abbia np e nq non inferiori a 5. Infatti in questi casi, lo scarto standardizzato (la cui distribuzione si può considerare approssimativamente) normale è:

$$z = \frac{x-np}{\sqrt{npq}} = \frac{x/n - p}{\sqrt{pq/n}} \qquad (10.10)$$

riferito rispettivamente al numero di volte o alla percentuale con cui un evento si verifica (tuttavia occorre apportare piccole correzioni, in entrambi i casi, dato che si trattano variabili casuali discrete come continue)

Esempio 2.10.3

Trovare la probabilità che, su 100 lanci di una moneta non truccata, se ne ottengano fra 45 e 63 (estremi inclusi) uguali. Secondo la distribuzione binomiale la probabilità cercata è:

$$\binom{100}{45}(0{,}5)^{45}(0{,}5)^{55} + \binom{100}{46}(0{,}5)^{46}(0{,}5)^{54} + \cdots + \binom{100}{63}(0{,}5)^{63}(0{,}5)^{37}$$

Dato che $Np = 100(0.5)$ e $Nq = 100(0.5)$ sono maggiori di 5, si può usare l'approssimazione normale alla distribuzione binomiale, con l'avvertenza che, in una distribuzione continua, da 45 a 63, *estremi inclusi* significa fra 44.5 e 63.5, cosicché si ha, data la distribuzione del numero di volte:

$M_x = Np = 50$ e $\sigma_x = \sqrt{Npq} = \sqrt{100(0.5)(0.5)} = 5$, ed essendo:

44.5 *in unità standard* $= (44.5 - 50)/5 = -1.10$

63.5 *in unità standard* $= (63.5 - 50)/5 = 2.70$

risulta infine: *probabilità* $=$ *area sotto la curva normale fra* -1.10 e $+2.70$ $= 0.3643 + 0.4965 = 0.8608$

Invece usando la distribuzione delle percentuali, si ha:

$$M_{x'} = p = 0.5 \qquad \sigma_{x'} = \sqrt{pq/N} = \sqrt{(0{,}5)(0{,}5)/100} = 0.05$$

Per tener conto del passaggio dalla distribuzione Binomiale discreta alla distribuzione Normale approssimata continua, si sottrae e si aggiunge, ai due limiti: 45% e 63%, la quantità: $0.5/N = 1/200$:

45% *in unità standard* $= (0.45 - 0.005 - 0.5)/0.05 = -1.10$

63% *in unità standard* $= (0.63 + 0.005 - 0.5)/0.05 = 2.70$

da cui: *probabilità* $=$ *area sotto la curva normale fra* -1.10 e $+2.70$ $= 0.3643 + 0.4965 = 0.8608$

Anche la distribuzione Poissoniana, per alti valori di λ, può essere approssimata dalla distribuzione Normale della variabile casuale standardizzata:

$$z = \frac{x - \lambda}{\sqrt{\lambda}} \qquad (10.11)$$

La distribuzione normale, riscoperta da Gauss e Laplace, in relazione ai loro studi sulla teoria degli errori di osservazione, per lungo tempo, ha un posto di assoluta preminenza nella statistica, considerando quasi assiomatico che, pur di disporre di un numero elevato di estrazioni, a caso, si constaterebbe che la variabile casuale, collegata con il fenomeno in studio, ha distribuzione normale.
Più recentemente questa opinione, così diffusa, è sottoposta ad un lavoro di critica diretto a ridimensionare l'importanza della distribuzione Normale, in cui tutti credevano: *gli sperimentali pensando che essa derivi da un teorema* ed *i matematici pensandola un fatto sperimentale*. Entrambe le affermazioni sono corrette, purché non prese proprio alla lettera. Infatti con opportune condizioni limitanti, la distribuzione Normale vale rigorosamente per un gran numero di variabili casuali, mentre è anche vero, d'altro canto, che l'esperienza prova come moltissime distribuzioni siano approssimativamente normali [21].

[21] Un'importante eccezione, nuovamente evidenziata, a partire dalla seconda metà del '900 (con la robustezza matematica), dopo un suo primo fulgore, nel '600 e nel '700 (con la robustezza qualitativa), è costituita dalla presenza di dati anomali e, in particolare, di errori grossolani, che portano a deviazioni, anche notevoli, dalla distribuzione Normale, proprio ingrossando le code.

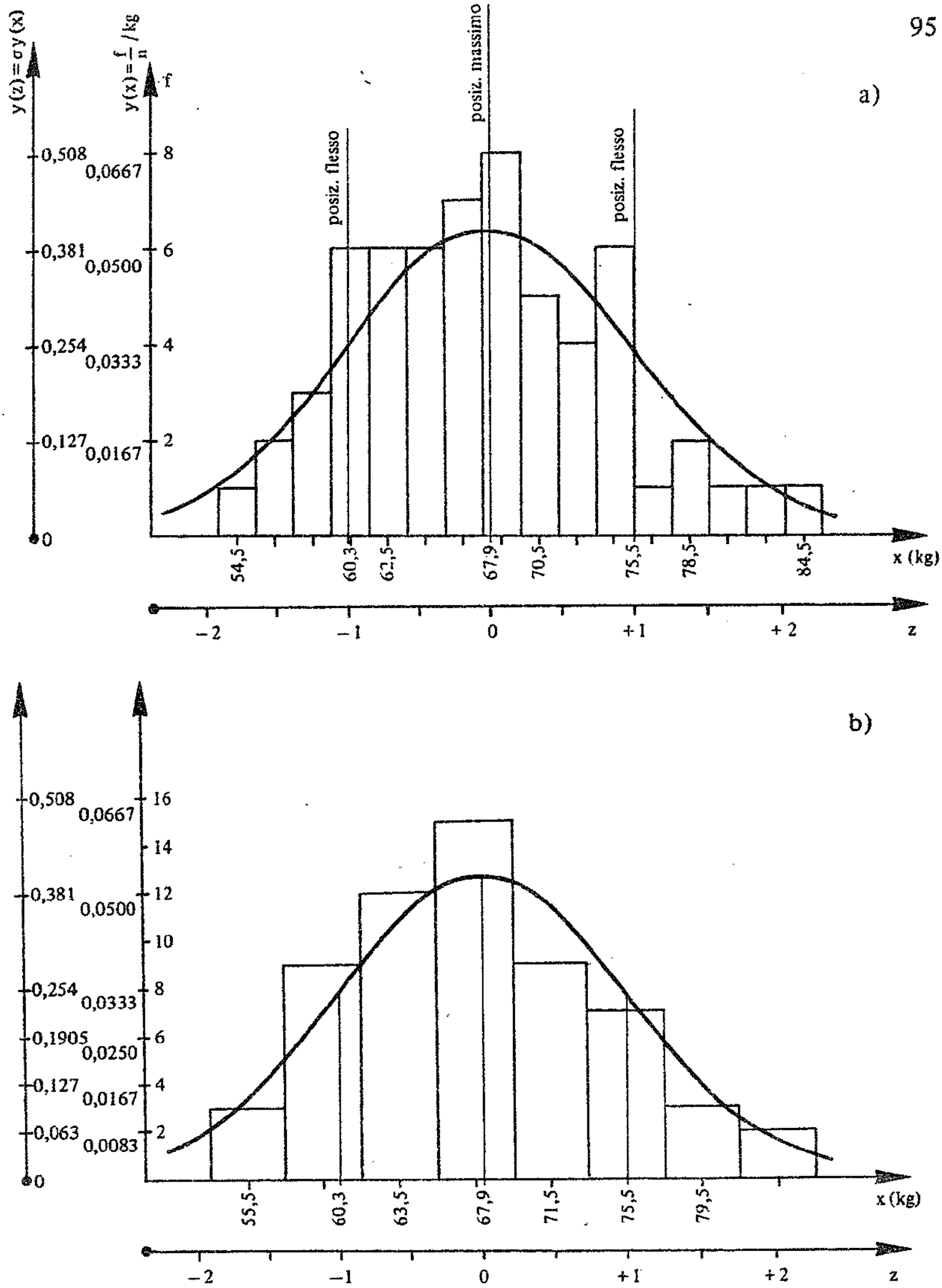

95
a)
$y(z) = \sigma y(x)$
$y(x) = \frac{f}{n}$ /kg
f
0,508
0,381
0,254
0,127
0
0,0667
0,0500
0,0333
0,0167
8
6
4
2
posiz. flesso
posiz. massimo
posiz. flesso
54,5
60,3
62,5
67,9
70,5
75,5
78,5
84,5
x (kg)
−2
−1
0
+1
+2
z
b)
0,508
0,381
0,254
0,1905
0,127
0,063
0
0,0667
0,0500
0,0333
0,0250
0,0167
0,0083
16
14
12
10
8
6
4
2
55,5
60,3
63,5
67,9
71,5
75,5
79,5
x (kg)
−2
−1
0
+1
+2
z

Fig. 2.10.3

2.11. Applicazioni della distribuzione normale: teorema centrale

Alcuni teoremi fondamentali sulla convergenza stocastica, atti a spiegare le vaste applicazioni, in statistica, della distribuzione Normale, sono presentati, senza dimostrazione,. Date n variabili casuali normalmente distribuite $x_1, x_2, \cdots, x_n$, indipendenti fra loro, con media e sqm rispettivamente uguali a $M_1, M_2, \cdots, M_n$. e $\sigma_1, \sigma_2, \cdots, \sigma_n$, si considera la loro somma:

$$X = x_1 + x_2 + \cdots + x_n \tag{11.1}$$

con media: $$M = M_1 + M_2 + \cdots + M_n \tag{11.2}$$

e varianza: $$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2 \tag{11.3}$$

Si può dimostrare che:

- ❑ se valgono le condizioni enunciate per ogni x_i, X è distribuita normalmente, con media e varianza date dalle (11.2) e (11.3);
- ❑ viceversa se la somma X di n variabili casuali indipendenti è distribuita normalmente, ne segue che anche ogni variabile casuale componente x_i è distribuita normalmente.

Questo significa che non solo la distribuzione Normale si riproduce e conserva per addizione, ma anche ogni distribuzione Normale non può provenire esattamente dalla somma di componenti non normali. Infatti sotto condizioni molto generali, la somma di un gran numero di variabili casuali non normali è distribuita in modo solo approssimativamente normale.

Inoltre ogni funzione lineare di variabili casuali normali è essa stessa normale; ad esempio, come una combinazione lineare di n variabili casuali indipendenti e normali, del tipo:

$$X = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n + b$$

ha distribuzione Normale con media e varianza rispettivamente:

$$M = a_1 M_1 + a_2 M_2 + \cdots + a_n M_n + b \tag{11.4}$$

$$\sigma^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2 \tag{11.5}$$

Un caso particolare si ha se: $x_1, x_2, \cdots, x_n$, sono variabili casuali indipendenti, normalmente distribuite, tutte con media M e sqm σ, cosicché la loro media aritmetica: $\overline{X} = 1/n \sum_{i=1}^{n} x_i$, è normalmente distribuita con media M e varianza:

$$\sigma_{\bar{x}} = n\left(1/n^2\ \sigma^2\right) = \sigma/n^2 \tag{11.6}$$

In generale, se la distribuzione di una variabile casuale X dipende da un parametro n e se si possono trovare due quantità: M_0 e σ_0, tali che la distribuzione della variabile casuale: $(X - M_0)/\sigma_0$, tenda ad essere normale, per $n \to \infty$, si dice che X è *asintoticamente normale*. Ad esempio, la variabile casuale Binomiale è asintoticamente normale, come media np e varianza npq.

La definizione data serve a chiarire l'enunciato del teorema centrale della statistica, noto anche come teorema fondamentale della convergenza stocastica o teorema di Laplace-Tchebycheff-Liapounoff.

Teorema Centrale

Qualunque siano le distribuzioni delle variabili casuali indipendenti x_i, soggette solo ad alcune condizioni molto generali, la somma: $X = x_1 + x_2 + \cdots + x_n$, è asintoticamente normale, con media M e varianza σ^2, date rispettivamente dalle (11.2) e (11.3).

Se le variabili casuali componenti x_i hanno tutte la stessa distribuzione, le condizioni, richieste dal teorema centrale (non precisate per semplicità), si riducono ad ipotizzare che la comune distribuzione di x_i possegga un momento del secondo ordine finito; in tal caso, se M e σ sono la media e lo sqm di ciascuna x_i, la loro somma: $X = \sum_{i=1}^{n} x_i$, è asintoticamente normale con media nM e varianza $n\sigma^2$, ed anche la loro media aritmetica: $\overline{X} = 1/n \sum_{i=1}^{n} x_i$, è asintoticamente normale con media M e varianza σ^2/n.

Invece nel caso più generale, dove non si suppone che tutte le x_i abbiano la stessa distribuzione, non è più sufficiente l'ipotesi della esistenza di un momento del secondo ordine finito, ma è necessario imporre altre condizioni, enunciate in modi diversi, a seconda degli autori (che hanno dimostrato il teorema), il cui scopo, in ogni caso, è ridurre la probabilità che una singola componente x_i dia un contributo relativamente grande al valore totale di X (in modo che 'influenza di tutte le x_i su X sia pressappoco equivalente). Se tali condizioni sono soddisfatte la somma: $X = x_1 + x_2 + \cdots + x_n$, è asintoticamente normale, con media M e varianza σ^2, date rispettivamente dalle (11.2), (11.3).

Esempio 2.11.1

A e B giocano a *testa o croce*, lanciando ciascuno 50 monete non truccate. A vince, se otterrà 6 o più teste più di B: determinare il rapporto fra le probabilità di A di perdere e di vincere una qualunque partita. Le variabili casuali P_A e P_B (percentuali di teste ottenute rispettivamente da A e B in n prove) sono distribuite binomialmente con: $M = p = 0.5$ e $\sigma = \sqrt{pq/N} = 0.5/\sqrt{50}$. Dato che il numero 50 dei lanci è abbastanza elevato, si può sostituire, alla distribuzione Binomiale, la distribuzione Normale con le stesse

M , σ , con l'intesa che 6 o più teste diventa 5.5 o più. La variabile casuale: $P_A - P_B$ (differenza fra la percentuale di teste ottenuta da A e quella ottenuta da B), è una combinazione lineare di due variabili casuali uguali, distribuite normalmente, e ha distribuzione, Normale con media e varianza:

$$M(P_A - P_B) = M(P_A) - M(P_B) = 0$$
$$\sigma^2(P_A - P_B) = pq/N + pq/N = 0.01$$

La variabile standardizzata, per la differenza di percentuali, è allora:

$$z = \frac{(P_A - P_B) - 0}{0.10} = \frac{5.5/50}{0.10} = 1.10$$

La probabilità di ottenere valori: $z \geq 1.10$, è 0.1357 , e così la probabilità di A di perdere è pertanto: $(1 - 0.1357) = 0.8643$, ed il rapporto richiesto: $0.8643/0.1357$, cioè 6.37 a 1.

Il teorema centrale della convergenza stocastica spiega perché un grande numero di applicazioni hanno distribuzioni normali, almeno approssimativamente [22], ad esempio, come nel caso degli errori di misura di una grandezza ed in tanti campi della biologia e dell'economia.
Secondo la teoria degli *errori elementari* o *piccoli errori*, introdotta da Hagen e Bessel, l'errore totale, commesso in una misura, può essere somma di un grande numero di errori elementari, indipendenti fra loro. Analogamente una variabile casuale, collegata con certe indagini biologiche, può essere il risultato di un grande numero di variabili casuali indipendenti che sommano i loro effetti. E ancora il guadagno o la perdita totali, di una compagnia di assicurazioni, è la somma del guadagno o della perdita, relative ad ogni singola polizza, da ritenere evidentemente indipendenti fra loro. Per il teorema centrale, le variabili casuali, associate a ciascuno di questi fenomeni complessivi, hanno distribuzioni normali, almeno approssimativamente, come largamente confermato dall'esperienza. Se il numero delle variabili casuali componenti non è abbastanza grande, oppure se non si compongono per sola addizione, si può sempre applicare qualche modifica al teorema centrale che attribuisce alla variabile casuale risultante distribuzioni molto vicine a quella Normale. Ad esempio, l'errore accidentale di misura può essere considerato una variabile casuale, ottenuta dalla

---

[22] Un'interessante eccezione a comportamenti di tipo Normale è costituita dalla variabile casuale Rettangolare (altrimenti chiamata distribuzione continua uniforme, particolarmente adatta a modellare il comportamento delle parti piccole):

Funzione densità di probabilità: $\begin{cases} p(x) = \dfrac{1}{b-a} & a \leq x \leq b \\ p(x) = 0 & x < a \;;\; x > b \end{cases}$

Media: $\mu = \dfrac{a+b}{2}$

Varianza: $\sigma^2 = \dfrac{(a-b)^2}{12}$

Trasformazione alla variabile casuale Normale: $z = erf^{(-1)}(x)$ e $p(z) = N(z)\,p(x)$

somma di un numero imprecisato di variabili casuali, rappresentanti le perturbazioni indipendenti di ciascuna delle caratteristiche variabili dell'ambiente:

$$x = a\nu_u + b\nu_v + \cdots + m\nu_z . \qquad (11.7)$$

Se le singole variabili casuali componenti soddisfano le condizioni previste dal teorema centrale, nel senso che i contributi di ciascuna di esse, alla formazione di x, abbiano tutti circa lo stesso ordine di grandezza e purché i termini del secondo membro della (11.7) siano assai numerosi, la variabile casuale x si comporta approssimativamente come una variabile casuale Normale.

In generale, si può supporre che i termini della (11.7) soddisfino a queste due condizioni. Infatti se qualcuna delle caratteristiche ambientali risultasse preponderante, rispetto alle altre, e la sua influenza non fosse dello stesso ordine di grandezza delle altre, ma sensibilmente superiore, questo effetto non potrebbe non essere avvertito dall'operatore che, in questa situazione, dovrebbe provvedere a tener conto di quella caratteristica, con opportune correzioni, oppure ad isolare meglio l'operazione di misura dall'ambiente. Allora nel primo caso, la misura si trasforma da una misura diretta in una indiretta; nel secondo invece, automaticamente l'influenza di quella caratteristica è eliminata totalmente, oppure è ridotta allo stesso ordine delle altre.

Pertanto gli errori accidentali di misura, entro i limiti previsti dal teorema centrale, possono essere considerati e trattati come una variabile casuale di Normale (e la variabilità strutturale. legata agli errori accidentali di misura, spesso constatata sperimentalmente, resta così giustificata).

Per lungo tempo, l'ipotesi di distribuzione Normale degli errori accidentali di misura è stata accettata senza discussione. Più recentemente, di fronte a frequenti casi di divergenza fra la teoria e la pratica, si è rimessa in discussione l'applicabilità del teorema centrale agli errori di misura, soprattutto in relazione al numero delle variabili casuali, presenti nella (11.7), non certo così tanto grande da poter ritenere la distribuzione della x sufficientemente approssimata da una distribuzione Normale. Tuttavia in mancanza di meglio, spesso gli errori accidentali continuano ad essere trattati come variabili casuali normali, il che consente notevolissime semplificazioni in tutti i vari procedimenti che, sulla base di poche misure eseguite, conducono a stimare le statistiche della popolazione delle misure possibili od anche i parametri che compaiono in relazioni analitiche tra grandezze diverse, con la facoltà di assoggettare poi i risultati a *test di significatività* che valgono solo nell'ipotesi di distribuzione Normale dell'universo da cui si eseguono le estrazioni. Tuttavia esistono metodi di stima più generali (di cui quelli applicati alle variabili casuali normali sono solo casi particolari).

Un'altra importante applicazione del teorema centrale è che, entro i limiti dello stesso, la media aritmetica di un gran numero di variabili casuali indipendenti è approssimativamente Normale [23] (e la stessa proprietà si applica anche ad altre funzioni, più generali della media [24]). Tutto ciò è di fondamentale importanza per le applicazioni statistiche che, molto spesso, hanno a che fare con medie od altri indici significativi, funzioni dei valori osservati di variabili casuali.

[23] Una buona approssimazione alla distribuzione Normale è riscontrabile già attorno alla decina di dati, dove il significato di decina è da intendersi in valori compresi tra otto e dodici. Allora una prima considerazione sul valore, piuttosto piccolo di questi numeri, si ritrova in tutte le tecniche metrologiche consigliate: come raddoppiare le misure, in qualche modo invertendo qualche elemento significativo (ad esempio, scambiando i piatti della bilancia), e triplicare queste coppie, cosicché sia sempre possibile individuare un eventuale errore grossolano, nell'ipotesi ragionevole di non sbagliare due volte su tre (del resto, prove sperimentali registrano un numero di errori grossolani pari a pochi percento e, in lunghe serie, molto precise, pari a pochi per mille). Pertanto il numero complessivo di misure, con questi semplici accorgimenti, è già pari a sei, cosicché il numero: otto, dista solo due ed il numero: dodici, è il suo doppio.

[24] Un primo semplice esempio è dato dalla media ponderata.

PARTE III – COMPLEMENTI DI CALCOLO DELLE PROBABILITA’

3.1 Distribuzioni bidimensionali discontinue

Nei fenomeni (tipo quelli menzionati nel paragrafo 1.8), ogni valore argomentale è costituito da una coppia (x_i, y_j) ed è rappresentato da un punto in uno spazio bidimensionale. Se le due variabili casuali X e Y, considerate contemporaneamente, possono assumere solo valori discreti ed anche, per ogni possibile valore (x_i, y_j), è definita una probabilità in accordo con gli assiomi (definiti nel paragrafo 2.1), si può definire la distribuzione bidimensionale discontinua, rappresentabile con una tabella a doppia entrata (analoga a quella del paragrafo 2.5). La distribuzione bidimensionale (o congiunta) è completamente descritta, analogamente a quelle uni-dimensionali, o dall'insieme delle probabilità elementari p o dalla funzione di distribuzione F:

$$p_{X,Y}(x,y) \equiv P((X = x) \cap (Y = y)) \qquad \text{con:} \quad \sum_{x_i} \sum_{y_j} p_{XY}(x_i, y_j) = 1$$

$$F_{X,Y}(x,y) \equiv P((X \le x) \cap (Y \le y)) = \sum_{x_i \le x} \sum_{y_j \le y} p_{X,Y}(x_i, y_j) \tag{1.1}$$

Nella trattazione delle variabili casuali bidimensionali, hanno particolare importanza i concetti di distribuzioni marginali e condizionate.

- Si chiama distribuzione marginale la variabile casuale X (o Y), considerata indipendentemente da Y (o da X). Queste due distribuzioni sono unidimensionali ed associano, ad ogni valore argomentale x_i (o y_j), la somma delle probabilità, competenti a tutti i valori argomentali bidimensionali, in cui compare x_i (o y_j), contemporaneamente a qualunque valore di Y (o di X):

$$p_X(x) \equiv P(X = x) = \sum_{y_j} p_{X,Y}(x, y_j)$$

$$F_X(x) \equiv P(X \le x) = \sum_{x_i \le x} p_X(x_i) = \sum_{x_i \le x} \sum_{y_j} p_{X,Y}(x_i, y_j) = F_{X,Y}(x, \infty) \tag{1.2}$$

  e analogamente per Y. In generale, la conoscenza delle due distribuzioni marginali non è sufficiente per definire la distribuzione congiunta.
- Se invece nella distribuzione bidimensionale (o congiunta) si fissa un determinato valore, ad esempio: $Y = y_0$, le probabilità, relative ai valori dell'altra variabile, sono date da: $p_{X,Y}(x, y_0)$. Se questo insieme di valori è normalizzato, cosicché la loro somma sia uguale ad 1, si costruisce una nuova distribuzione di probabilità della variabile casuale X, condizionata al valore: $Y = y_0$:

$$p_{X|Y}(x,y) \equiv P((X = x)\,|\,(Y = y)) = \frac{P((X = X) \cap (Y = y))}{P(Y = y)} =$$

$$= \frac{p_{X,Y}(x,y)}{\sum_{x_i} p_{X,Y}(x_i, y)} = \frac{p_{X,Y}(x,y)}{p_Y(y)} \tag{1.3}$$

Il denominatore è semplicemente la distribuzione marginale di Y, valutata nel valore di Y assegnato. In modo analogo, è definita la distribuzione condizionata di Y ad X. La (1.3) può essere così riscritta:

$$p_{X,Y}(x,y) = p_{X|Y}(x,y)p_Y(y) \tag{1.4}$$

mettendo in evidenza che la distribuzione bidimensionale è data dal prodotto di quella condizionata per la marginale, relazione utile, in quanto è spesso più semplice in pratica definire queste due ultime e, da esse, ricavare la distribuzione congiunta.

<u>Esempio 3.1.1</u>

Un'urna contiene 15 palline, 6 rosse (R), 4 gialle (G), 5 blu (B). Estratte dall'urna, in blocco, 2 palline, chiamando X la variabile casuale *colore della prima pallina* e Y la variabile casuale *colore della seconda pallina*, si ha una variabile casuale doppia discreta. Si determinano la serie dei valori argomentali, le loro probabilità, le due distribuzioni marginali e la variabile casuale condizionata: $Y \mid X$.

La serie dei valori argomentali della variabile casuale doppia è data da tutte le possibili coppie di valori: RR, RG, RB, GR, GG, GB, BR, BG, BB. Per il calcolo delle rispettive probabilità si ha:

$$P_{RG} = \left(\frac{6}{15}\right)\cdot\left(\frac{4}{14}\right) = \frac{24}{210} \qquad P_{RR} = \left(\frac{6}{15}\right)\cdot\left(\frac{5}{14}\right) = \frac{30}{210} \qquad ecc.$$

La variabile doppia (X,Y) è completamente rappresentata nella seguente tabella, dove compaiono anche le due distribuzioni marginali.

$Y \backslash X$	R	G	B	$marginale\,Y$
R	$30/210$	$24/210$	$30/210$	$2/5$
G	$24/210$	$12/210$	$20/210$	$4/15$
B	$30/210$	$20/210$	$20/210$	$1/3$
$marginale\,X$	$2/5$	$4/15$	$1/3$	1

Per la costruzione delle variabili casuali condizionate: $Y \mid X$, occorre trovare, ad esempio:

$$P((Y=R)/(X=R)) = P((Y=R)\cap(X=R)) \mid P(X=R) = 30/210 \cdot 5/2 = 5/14$$
$$P((Y=G)/(X=R)) = P((Y=G)\cap(X=R)) \mid P(X=R) = 24/210 \cdot 5/2 = 2/7 \qquad ecc.$$

Le tre variabili casuali condizionate: $Y \mid X$, sono compiutamente rappresentate nella seguente tabella:

$Y \backslash X$	R	G	B	$marginale\ Y$
R	$5/14$	$3/7$	$3/7$	$2/5$
G	$2/7$	$3/7$	$2/7$	$4/15$
B	$5/14$	$5/14$	$2/7$	$1/3$
	1	1	1	1

Esempio 3.1.2

Conoscendo le due distribuzioni marginali:

$$X\begin{cases} 6 & 7 & 8 & 9 & 10 \\ 0.25 & 0.30 & 0.20 & 0.15 & 0.10 \end{cases} \qquad y\begin{cases} 11 & 12 & 13 & 14 \\ 0.35 & 0.30 & 0.20 & 0.15 \end{cases}$$

si vuole creare una variabile casuale doppia, rappresentante la distribuzione congiunta di X e Y . Infine (basandosi su quanto esposto nel paragrafo 2.5), si vuole calcolare la distribuzione della variabile casuale monodimensionale: $Z = X + Y$.

Una distribuzione di probabilità, compatibile con le due marginali assegnate (tra le tante possibili), è data dalla prima delle seguenti tabelle. Da questa tabella, risulta poi che, ad esempio, i valori argomentali: $x = 6$ e $y = 11$, non possono essere accoppiati ed i soli valori possibili della variabile casuale Z sono dati dalla seconda delle seguenti tabelle.

$Y \backslash X$	6	7	8	9	10	$marginale\ Y$
11			0.10	0.15	0.10	0.35
12		.20	0.10			0.30
13	0.10	0.10				0.20
14	0.15					0.15
$marginale\ X$	0.25	0.30	0.20	0.15	0.10	1.00

+	6	7	8	9	10
11			19	20	21
12		19	20		
13	19	20			
14	20				

In questo modo, a ciascuno dei tre possibili valori della somma compete una probabilità pari alla somma delle probabilità delle coppie che, sommate fra loro, danno lo stesso risultato:

$$Z\begin{cases} 19 & 20 & 21 \\ 0.40 & 0.50 & 0.10 \end{cases}$$

3.2. Distribuzioni bidimensionali continue

Se le variabili casuali X e Y sono continue, non si definisce la probabilità del singolo valore (x_i, y_j), ma la densità di probabilità congiunta $f_{X,Y}(x,y)$, tale che la probabilità infinitesima che X cada nell'intervallo: $(x, x+dx)$, e che contemporaneamente Y cada nell'intervallo: $(y, y+dy)$, sia: $f_{X,Y}(x,y)\,dxdy$, dove (in generale, supponendo x e y definite nell'intervallo $[-\infty, +\infty]$) si ha:

$$f_{X,Y}(x,y) \geq 0 \qquad \text{e} \qquad \int_{-\infty}^{+\infty}\int f_{X,Y}(x,y)\,dxdy = 1$$

Tutti i concetti precedentemente esposti per le variabili casuali bidimensionali discontinue si trasferiscono alle variabili casuali continue, sostituendo il segno di integrale a quello di sommatoria. Di conseguenza, si riscrivono le (1.1), (1.2), (1.3) ed (1.4):

$$P((x_1 \leq X \leq x_2) \cap (y_1 \leq Y \leq y_2)) = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f_{X,Y}(x,y)\,dxdy = 1$$

$$F_{X,Y}(x,y) = P[(-\infty \leq X \leq x) \cap (-\infty \leq Y \leq y)] = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x,y)\,dxdy \tag{2.1}$$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx \qquad \text{e} \qquad F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx \tag{2.2}$$

$$f_{X|Y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{2.3}$$

$$f_{X,Y}(x,y) = f_{X|Y}(x,y)\,f_Y(y) \tag{2.4}$$

Dalle (1.4) e (2.4), segue una definizione di variabili casuali indipendenti, più generale rispetto a quella di indipendenza lineare (data nel paragrafo 2.6).

Infatti dire che X è indipendente da Y equivale a dire che nessuna delle due può condizionare l'altra e che devono valere le seguenti relazioni:

$$p_{X|Y}(x,y) = p_X(x) \qquad \text{e} \qquad f_{X|Y}(x,y) = f_X(x)$$

ed analoghe, tutte esprimenti l'uguaglianza fra distribuzioni marginali e condizionate. Di conseguenza, le (1.4) e (2.4) si modificano in modo che, se e solo se X e Y sono indipendenti tra loro, la loro distribuzione bidimensionale (o congiunta) è uguale al prodotto delle due distribuzioni marginali.

In perfetta analogia con quanto detto a proposito delle variabili statistiche a due dimensioni, si definisce la curva di regressione per la variabile casuale bidimensionale discontinua. Infatti supponendo di considerare (come usuale) la curva di regressione di Y *su* X, essa è definita come il luogo delle medie delle variabili casuali condizionate: $Y \mid X(x,y)$, ed indicata come: $\overline{Y}(x)$.

- Se la variabile casuale è discontinua, si ha, quasi sempre, una spezzata nel piano XY.
- Se la variabile casuale è continua, occorre ricavare l'equazione della curva corrispondente alla definizione data:

$$\overline{Y}(x) = \int_{-\infty}^{+\infty} y f_{Y/X}(x,y)\,dy$$

Ricavata l'equazione della curva di regressione, è possibile calcolare anche per queste variabili casuali bidimensionali continue l'indice di Pearson, tenendo presente che i termini di p_j / N, presenti nella (1.10.1), sono sostituiti con la distribuzione marginale $f_X(x)$:

$$\eta_x^2 = \frac{\int_{-\infty}^{+\infty} \left(\overline{Y}(x) - M_y\right)^2 f_X(x)\,dx}{\sigma_{\overline{Y}}^2 + \overline{\sigma}_Y^2} = \frac{\sigma_{\overline{Y}}^2}{\sigma_Y^2}$$ [25]

Le statistiche M_y e σ_y^2 sono della variabile casuale marginale Y, con densità di probabilità $f_Y(y)$. Infine si può calcolare il coefficiente di correlazione r_{xy}, applicando una qualunque delle espressioni (1.12.11) od (1.12.13), tenendo presente che l'equivalente della (1.12.12) per variabili casuali bidimensionali è:

$$M_{x,y} = \int \int_{-\infty}^{+\infty} xy \; f_{X,Y}(x)\,dxdy \qquad \text{ed inoltre:} \qquad \sigma_{xy} = \int \int_{-\infty}^{+\infty} (x - M_x)(y - M_y) f_{X,Y}(x,y)\,dxdy$$

<u>Esempio 3.2.1</u>

Data la variabile casuale continua bidimensionale, definita nell'intervallo $[1,2]$, con la condizione di esistenza: $1 \le x \le y \le 2$, e con densità di probabilità: $f_{X,Y}(x,y) = a$. Si vuole calcolare il valore di a, l'indice di Pearson η_x^2 ed il coefficiente di correlazione lineare r_{xy}.

La condizione di normalizzazione permette di ricavare il valore numerico di a, imponendo:

$$\int_1^2 \int_x^2 a\,dxdy = a \int_1^2 dx \int_x^2 dy = 1 \qquad \text{da cui:} \qquad a = 2$$

[25] Come per le variabili statistiche, anche per le variabili casuali, l'identità, posta a denominatore, è detta teorema di decomposizione ortogonale della varianza e stabilisce che la somma della varianza spiegata (cioè della varianza tra le medie condizionate, con la media generale) e della varianza residua (cioè della media delle varianze condizionate) è uguale alla varianza generale.

Per il calcolo di η_x^2 e r_{xy}, occorre calcolare $f_X(x)$, M_x, σ_x^2, $f_Y(y)$, M_y, σ_y^2, $f_{Y|X}(x,y)$, $\overline{Y}(x)$, M_{xy} :

$$f_{X,Y}(x,y) = 2 \qquad \text{con} \qquad 1 \le x \le y \le 2$$

$$f_X(x) = \int_x^2 2\,dy = 2(2-x) \qquad 1 \le x \le 2$$

$$f_Y(y) = \int_1^y 2\,dx = 2(y-1) \qquad 1 \le y \le 2$$

$$M_x = \int_1^2 x(2(2-x))\,dx = 4/3$$

$$M_y = \int_1^2 y(2(y-1))\,dy = 5/3$$

$$\sigma_x^2 = \int_1^2 x^2(2(2-x))\,dx - (4/3)^2 = 1/18$$

$$\sigma_y^2 = \int_1^2 y^2(2(y-1))\,dy - (5/3)^2 = 1/18$$

$$f_{Y|X}(x,y) = \frac{2}{2(2-x)} = \frac{1}{2-x} \qquad 1 \le x \le y \le 2$$

$$\overline{Y}(x) = \int_x^2 y\frac{1}{2-x}\,dy = \frac{1}{2-x}\left(2-\frac{x^2}{2}\right) = 1+\frac{x}{2} \qquad 1 \le x \le 2$$

da cui:

$$\eta_x^2 = \frac{\int_1^2 \left[\left(1+\frac{x}{2}\right)-\frac{5}{3}\right]^2 [2(2-x)]\,dx}{1/18} = \frac{1}{4}$$

$$r_{xy} = \frac{\int_1^2\int_x^2 x\cdot y\cdot 2\,dxdy - \frac{4}{3}\cdot\frac{5}{3}}{\sqrt{\frac{1}{18}\cdot\frac{1}{18}}} = \frac{1}{2}$$

Nella Fig. 3.2.1, si nota il campo di definizione (tratteggiato) della variabile casuale bidimensionale e delle due variabili casuali marginali, la porzione di piano con l'esistenza di $f_{X,Y}(x,y)$ e le due porzioni di rette rappresentanti $f_X(x)$ e $f_Y(y)$, la superficie rigata $f_{Y|X}(x,y)$ che è un cilindro iperbolico parallelo all'asse Y (situata al di sotto od al di sopra di $f_{X,Y}(x,y)$, secondo che i valori di $f_X(x)$ siano maggiori o minori di 1, ed intersecata con essa per: $x=3/2$), la curva di regressione, in questo caso, una porzione di retta (ed in questo modo l'indice di Pearson è il quadrato di r_{xy}, relazione valida sempre se e solo se la regressione è lineare).

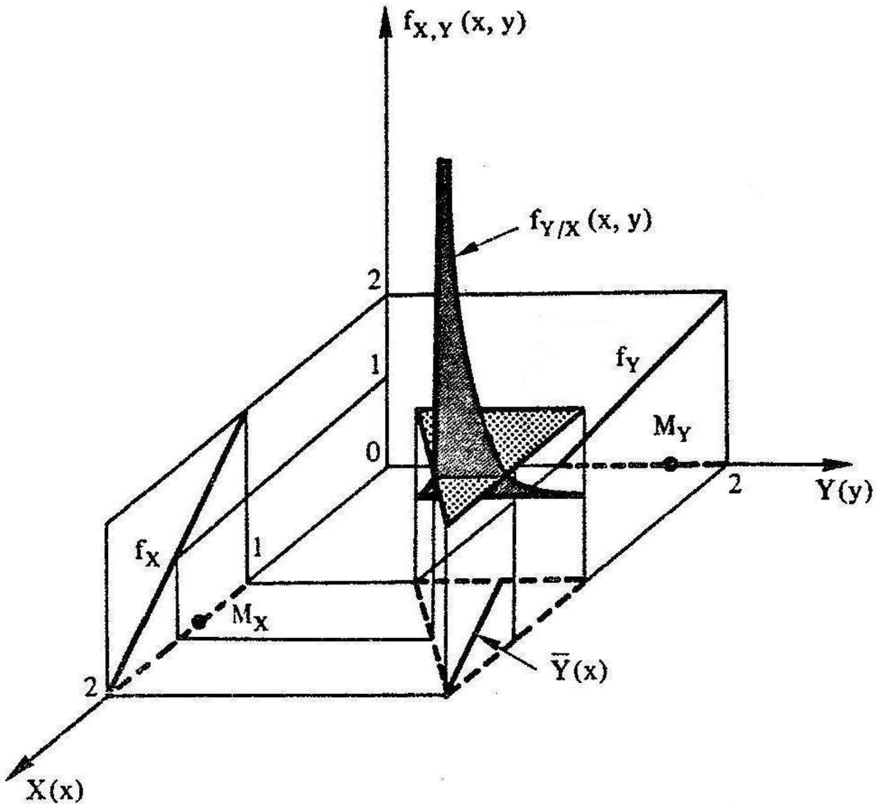


Fig. 3.2.1 – Rappresentazione grafica della variabile casuale doppia, delle due variabili casuali marginali, della variabile casuale condizionata e della curva di regressione

Infine essendo $\overline{Y}(x)$ una retta, si può evidenziare il coefficiente di regressione lineare: $a = 1/2$.

3.3. Distribuzioni di variabili casuali derivate con trasformazione di variabile

Data una variabile casuale X, di distribuzione nota, si cerca la distribuzione di una nuova variabile casuale Y, derivata dalla precedente, tramite una relazione data. Se la funzione g è monotona, crescente (o decrescente), il problema può essere risolto facilmente, in quanto Y è minore o uguale ad un prefissato valore y_0, se e solo se X è minore od uguale a quel valore di x_0 per cui: $y_0 = g(x_0)$. In generale, la funzione g è invertibile, trovando la funzione inversa: $x = g^{-1}(y)$ e, in particolare: $x_0 = g^{-1}(y_0)$, da cui:

$$F_Y(y) = P(Y \leq y) = P\left(X \leq g^{-1}(y)\right) = F_X\, g^{-1}(y).$$

La densità di probabilità di Y può essere ricavata derivando la $f_Y(y)$ rispetto ad y. Tuttavia se la funzione g è del tipo sopra definito, si può ricavare la densità di probabilità di Y direttamente in funzione di quella di X, cosicché:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X\left(g^{-1}(y)\right) = \frac{d}{dy}\int_{-\infty}^{g^{-1}(y)} f_X(x) = f_X\left(g^{-1}(y)\right)\frac{d}{dy}g^{-1}(y) = f_X(x)\frac{dx}{dy}$$

od anche:

$$f_Y(y)\,dy = f_X(x)\,dx\,. \tag{3.1}$$

La (3.1) afferma che la probabilità per Y di assumere un valore nell'intervallo: $(y, y+dy)$, è uguale a quella per X di assumere un valore nell'intervallo: $\left(g^{-1}(y)+g^{-1}(y)+dg^{-1}(y)\right)$, come illustra la Fig. 3.3.1, dove questi intervalli non sono uguali (né fra loro, né da punto a punto), mentre sono invece sempre uguali fra loro le aree infinitesime (tratteggiate).

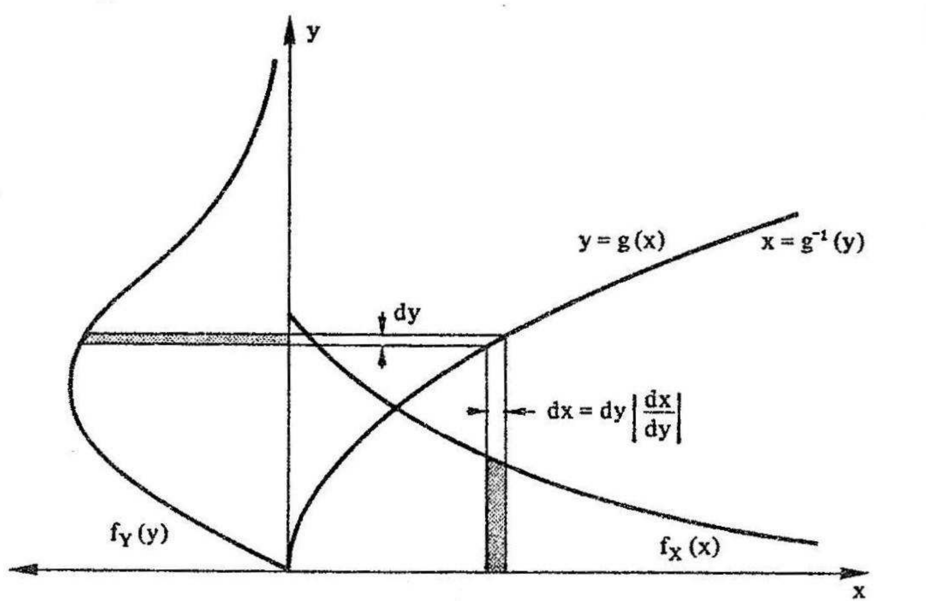


Fig. 3.3.1 – Trasformazione di densità di probabilità, per la trasformazione: $Y = g(X)$, su X

La formula (3.1) è valida nel caso di funzione monotona crescente e, se la funzione è decrescente, questa diventa [26]:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|. \tag{3.2}$$

<u>Esempio 3.3.1</u>

La relazione fra le due variabili casuali è lineare: $Z = a + bX$, cosicché si ha:

$$x = \frac{z-a}{b} \qquad \left|\frac{dx}{dy}\right| = \left|\frac{1}{b}\right| \qquad f_Z(z) = \left|\frac{1}{b}\right| f_X\left(\frac{z-a}{b}\right). \tag{3.3}$$

[26] Se la funzione di trasformazione delle due variabili casuali non è monotona (purché monodroma) la trasformazione della funzione densità di probabilità è data dalla somma di tutti i contributi forniti dai singoli tratti in cui occorre spezzare la funzione stessa, affinché essa sia localmente monotona (indifferentemente crescente o decrescente).

Esempio 3.3.2

Data una variabile casuale X, con densità di probabilità:

$f_X(x) = e^{-x}$ con $x \geq 0$.

si cerca la densità di probabilità di una seconda variabile casuale Y, legata alla precedente dalla relazione:

$Y = \sqrt{X}$ con $y \geq 0$.

Invertendo la relazione data, si ha:

$X = Y^2$ e $\frac{dx}{dy} = 2y$

da cui, applicando la (3.2), si ha:

$f_Y(y) = 2ye^{-y^2}$ con $y \geq 0$

In Fig.8.3.2, si può notare la differenza fra le due curve, rappresentanti le funzioni di densità di probabilità delle due variabili casuali.

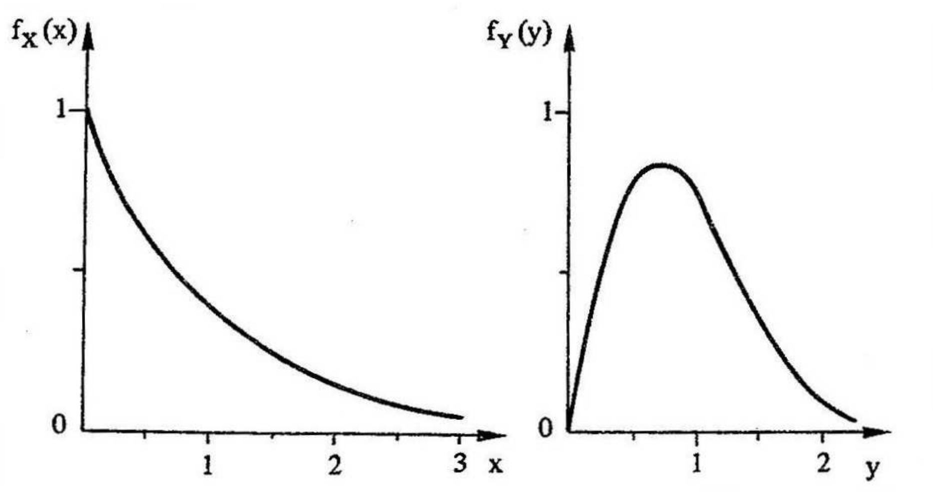


Fig. 8.3.2 – Rappresentazione grafica di: $f_X(x) = e^{-x}$ e $f_Y(y) = 2ye^{-y^2}$

3.4. Distribuzione della somma di due variabili casuali

La somma di due variabili casuali: $Z = X + Y$, è un caso particolare del più generale problema di trovare la distribuzione di probabilità, per una trasformazione qualsiasi di due variabili casuali: $Z = g(X,Y)$, nota la distribuzione congiunta di (X,Y).

Anzitutto è opportuno considerare la densità condizionata di Z, posto che Y assuma un particolare valore di y, in modo che, se $Y = y$, si ha:

$$Z = y + X$$

dove Z è una funzione lineare di una sola variabile casuale X e la sua densità di probabilità si trova (come nell'Esempio 3.3.1), ponendo nella (3.3): $a = y$ e $b = 1$, ed ottenendo:

$$f_{Z|Y}(z,y) = f_{X|Y}(z - y, y)$$

Dopodiché la densità di probabilità bidimensionale può essere sempre espressa tramite la (2.4) per cui:

$$f_{Z,Y}(z,y) = f_{Z|Y}(z,y) f_Y(y) = f_{X|Y}(z - y, y) f_Y(y) = f_{X,Y}(z - y, y) \qquad (4.1)$$

La (4.1) indica che la probabilità che: $Y = y$ e $Z = X + Y = z$, è uguale alla probabilità che: $Y = y$ e $X = z - y$, cosicché la distribuzione marginale di Z si ricava applicando la (2.2):

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{Z|Y}(z,y)\, dy = \int_{-\infty}^{+\infty} f_{X|Y}(z - y, y)\, dy \qquad (4.2)$$

Inoltre se le due variabili casuali. X e Y sono indipendenti la (4.2) diventa:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z - y) f_Y(y)\, dy = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x)\, dx \qquad (4.3)$$

espressione molto comune, nota come *integrale di convoluzione*.

<u>Esempio 3.4.1</u>

Un pendolare si reca al lavoro, utilizzando due mezzi X e Y i cui rispettivi tempi di attesa hanno densità di probabilità data:

$$f_X(x) = \alpha\, e^{-\alpha x} \qquad f_Y(y) = \beta\, e^{-\beta y} \qquad \text{con} \qquad x, y \geq 0 .$$

Si cerca la densità di probabilità del tempo complessivo, speso aspettando, dato dalla variabile casuale: $Z = X + Y$. Supponendo l'indipendenza dei due tempi di attesa, si applica la (4.3), facendo attenzione ai limiti di integrazione; infatti dato che: $f_X(x) = 0$ per $x < 0$, si può scrivere:

$$f_Z(z) = \int_0^{+\infty} \alpha e^{-\alpha x} f_Y(z-x)dx$$

ma se: $f_Y(y) = 0$ per $y < 0$, risulta: $f_Y(z-x) = 0$ per $z - x < 0$, da cui: $x > z$, cosicché:

$$f_Z(z) = \int_0^z \alpha e^{-\alpha x} \beta\, e^{-\beta(z-x)}dx = \alpha\, \beta\, e^{-\beta z} \int_0^z e^{x(\beta-\alpha)}dx = \frac{\alpha\, \beta}{\beta - \alpha}\left(e^{-\alpha z} - e^{-\beta z}\right) \qquad \text{con} \qquad z \geq 0$$

Se si pone: $\alpha = 0.14\, min^{-1}$, $\beta = 0.08\, min^{-1}$, si possono tracciare: f_X, f_Y e f_Z (Fig. 8.4.1):

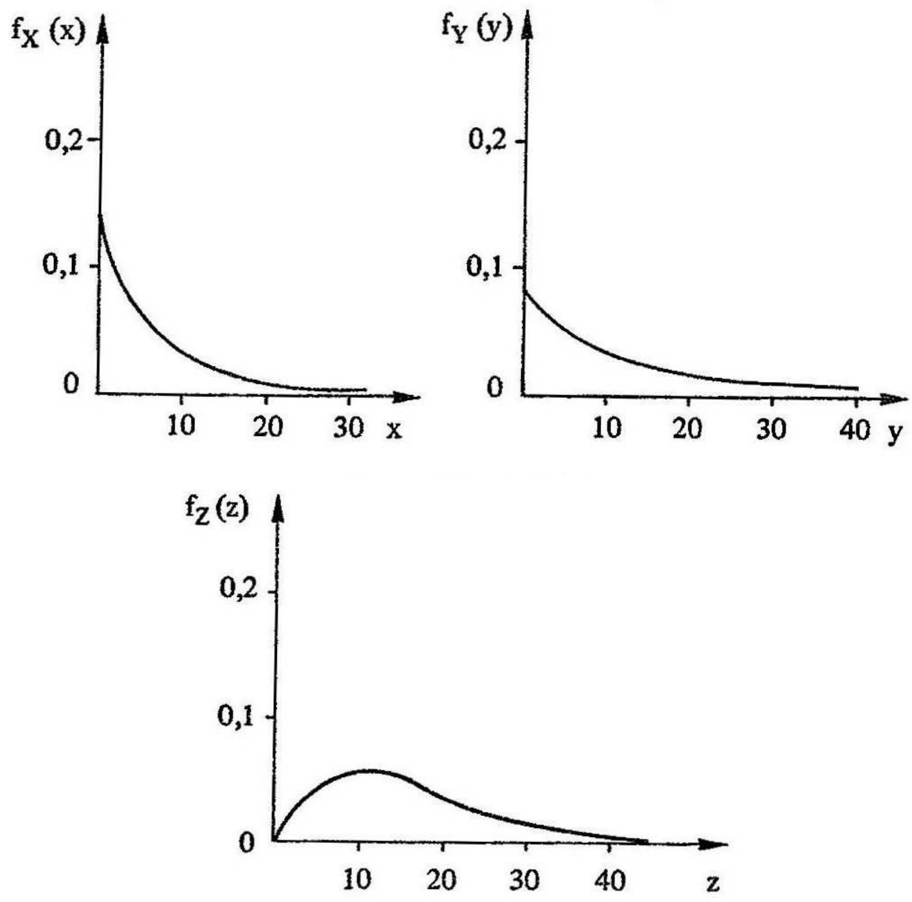


Fig. 3.4.1 – Densità di probabilità di due variabili casuali del tipo $f_X(x) = \lambda e^{-\lambda x}$ e della loro somma

3.5. Distribuzione normale bidimensionale

Date due variabili casuali X e Y indipendenti e normalmente distribuite, la loro densità di probabilità bidimensionale è data, come visto, dal prodotto delle due densità di probabilità:

$$f_{U,V}(u,v) = \frac{1}{2\pi} e^{-\frac{1}{2}\left(u^2 + v^2\right)} \tag{5.1}$$

dove:

$$u = \frac{x - M_x}{\sigma_x} \qquad \text{e} \qquad v = \frac{y - M_y}{\sigma_y}$$

Se esse non sono indipendenti la (5.1) si modifica, introducendo un termine misto che vale 0, nel caso di indipendenza e che, per definizione, contiene il coefficiente di correlazione lineare ρ tra le due variabili casuali X e Y [27]:

$$f_{U,V}(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(u^2 - 2\rho uv + v^2\right)} \qquad \text{con} \qquad -\infty \le u,v \le +\infty$$

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-M_x}{\sigma_x}\right)^2 - 2\rho\frac{(x-M_x)(y-M_y)}{\sigma_x\sigma_y} + \left(\frac{y-M_y}{\sigma_y}\right)^2\right)} \qquad (5.2)$$

Le due distribuzioni di probabilità marginali, ricavabili dalla (5.2.), sono normali: (M_x, σ_x) e (M_y, σ_y). Invece non è vero il viceversa, ovvero se le distribuzioni marginali di una variabile casuale bidimensionale sono normali, non è necessariamente essa è del tipo (5.2.).

Se nella (5.2.) si pone: $\rho = 0$, la funzione densità di probabilità $f_{X,Y}$ si spezza nel prodotto delle due densità di probabilità marginali: $f_x \cdot f_y$, perché le variabili casuali X e Y hanno indipendenza stocastica (e non sono solo indipendenti linearmente).

Di conseguenza, se le variabili casuali X e Y hanno distribuzione bidimensionale (o congiunta) normale, la sola forma di dipendenza possibile fra loro è quella lineare (ed è una retta la loro unica curva di regressione ammissibile).

A riguardo, l'espressione della distribuzione di probabilità condizionata della variabile casuale $Y \mid X$, ricavabile dalla (5.2), ha espressione:

---

[27] La superficie formata dalla funzione densità di probabilità bidimensionale (o congiunta) Normale ha forma a campana (così come una sezione di campana è l'andamento di una funzione densità di probabilità monodimensionale Normale).
Le sezioni orizzontali di una funzione densità di probabilità bidimensionale (o congiunta) Normale hanno la forma di:

- un cerchio, nel caso di perfetta indipendenza (lineare e stocastica, essendo la stessa cosa, per la distribuzione Normale);
- un segmento di retta, nel caso di perfetta dipendenza (lineare o funzionale, essendo ancora la stessa cosa, per la distribuzione Normale);
- un'ellisse, in tutti i casi intermedi (dove la dipendenza lineare, unica dipendenza possibile, cresce dalla perfetta indipendenza alla perfetta dipendenza), i cui parametri caratteristici sono i due semiassi e l'orientamento del semiasse maggiore:

$$\lambda_{max/min} = \frac{\sigma_x^2 + \sigma_y^2}{2} \pm \frac{1}{2}\sqrt{\left(\sigma_x^2 - \sigma_y^2\right)^2 + 4\sigma_{xy}^2}$$

$$tan\, 2\vartheta = \frac{-2\sigma_{xy}}{\sigma_x^2 - \sigma_y^2}$$

$$f_{Y/X}(x,y) = \frac{\dfrac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left(u^2 - 2\rho uv + v^2\right)}}{\dfrac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2}} =$$

$$= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\, e^{-\frac{1}{2(1-\rho^2)}\left(v^2 - 2\rho uv + \rho^2 u^2\right)} =$$

$$= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\, e^{-\frac{1}{2}\left(\frac{v-\rho u}{(1-\rho^2)}\right)^2} =$$

$$= \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}}\, e^{-\frac{1}{2}\left(\frac{y-M_y-\rho(\sigma_y/\sigma_x)(x-M_x)}{\sigma_y\sqrt{1-\rho^2}}\right)^2}$$

Dato che X è fissato, la variabile casuale $Y \mid X$ ha media:

$$M_y + \rho\frac{\sigma_y}{\sigma_x}(x - M_x)$$

e sqm:

$$\sigma_y\sqrt{1-\rho^2}$$

Come già detto in precedenza, la curva di regressione è il luogo delle medie di $Y \mid X$, al variare di X, la cui equazione è quella di una retta (ed analoghi risultati si hanno per $f_{X|Y}$):

$$\overline{Y}(x) = M_y + \rho\frac{\sigma_y}{\sigma_x}(x - M_x) \qquad (5.3)$$

Questo giustifica l'uso intensivo della regressione lineare, come un modello di dipendenza fra due variabili casuali, i cui limiti di validità stanno nell'ipotizzare una distribuzione normale bidimensionale (o congiunta) per le variabili stesse [28].

[28] In conclusione, i vantaggi della distribuzione Normale stanno nel suo essere invariante per trasformazioni lineari, come un cambio di media e/o di varianza, la somma, oppure la combinazione lineare, di due (o più) variabili casuali normali (essendo queste operazioni associative) e nel suo essere Normale, almeno approssimativamente, con una lunga somma o combinazione lineare di variabili casuali qualsiasi, purché indipendenti fra loro, dove nessuna è preponderante sulle altre. Invece la distribuzione normale si presta male alla modellazione, al trattamento ed all'analisi di dati anomali e, in particolare, di dati contaminati da errori grossolani.

PARTE IV – DISTRIBUZIONI CAMPIONARIE

4.1. Campionamento

Dato un esperimento ripetitivo e la variabile casuale X relativo ad esso, se si eseguono n ripetizioni indipendenti dell'esperimento si otterranno n valori estratti, a caso, da X : $x_1, x_2, \ldots, x_n$. Questa sequenza rappresenta una categoria di dati statistici di fondamentale importanza, dove il processo gerarchico, per ricondurre tutti gli esperimenti possibili ad un unico sistema, è spesso assimilato all'estrazione, a caso, da un'urna contenente tutti i possibili valori x_i di un elemento, di cui si nota il valore argomentale e che si reinserisce, nell'urna, prima di eseguire la seconda estrazione, a caso (e così via).

La sequenza: $x_1, x_2, \ldots, x_n$, si chiama *campione* casuale, estratto da un'urna, della popolazione originaria (o anche *universo*), tramite un processo di estrazione, a caso, di tipo Bernoulliano, con ripetizione. Il termine *con ripetizione* indica che lo stesso individuo può essere estratto più volte, mentre il processo di estrazione, a caso, è senza ripetizione (o *in blocco*), se l'individuo estratto non è reinserito, nell'urna, dopo aver annotato il suo valore argomentale. Infatti in questo caso, la composizione dell'universo risulta diversa (e diminuita), ad ogni estrazione.

La variabile casuale X i cui singoli valori argomentali sono rappresentati nell'urna dell'universo, ha una certa distribuzione $F(x)$, mentre la distribuzione del campione è rappresentabile da una probabilità uguale alla frequenza relativa $1/n$ in ognuno dei punti: $x_1, x_2, \ldots, x_n$, fra i quali possono esisterne di uguali. La funzione di distribuzione del campione è una funzione a gradini, con salti uguali ad $1/n$ od ai suoi multipli, in ogni x_i , cosicché se ν sono i valori campionari minori od uguali ad un certo valore x , si ha:

$$F^*(x) = \frac{\nu}{n}$$

Per questa distribuzione si possono calcolare momenti e statistiche varie (M_1, M_2, σ^2 , indici di *skewness*, coefficiente di correlazione se X è almeno bidimensionale, ecc.), avendo cura di distinguere quest'ultimi, che sono *statistiche del campione* , dalle corrispondenti statistiche calcolabili sulla variabile casuale (costituente l'universo), cioè dalle *statistiche dell'universo* (spesso poi si indicano le prime come statistiche empiriche e le seconde come statistiche teoriche).

Ad esempio, il momento di ordine k e polo zero è rispettivamente per il campione e per l'universo

$$\frac{1}{n}\sum_{i=1}^{n} x_i^k \qquad\qquad \int_{-\infty}^{+\infty} x^k dF(x) \ .$$

Per evitare confusioni, è uso indicare con simboli diversi le statistiche dei campioni e quelle dell'universo: in particolare, la media e la varianza campionaria si indicano con $\overline{x}$ e s^2 :

$$\bar{x} = \frac{1}{n}\sum x_i \qquad\qquad s^2 = \frac{1}{n}\sum (x_i - \bar{x})^2$$

mentre M e σ^2 rappresentano la media e la varianza della variabile casuale da cui il campione è estratto, a caso:

$$M = \int_{-\infty}^{+\infty} x dF(x) \qquad\qquad \sigma^2 = \int_{-\infty}^{+\infty} (x-M)^2 dF(x)$$

A questo punto, occorre notare che tutta la trattazione delle variabili statistiche (nella parte I) è l'esposizione dei metodi di analisi di campioni casuali, considerati indipendentemente dal processo generativo che forma il campione, ovvero indipendentemente dalla variabile casuale da cui i singoli individui sono estratti, a caso. Pertanto sarebbe corretto usare, anche in quel caso, invece dei simboli: M e σ^2, i simboli: $\bar{x}$ e s^2. Questi ultimi tuttavia sono adottati, di solito, solo per contrapporre le statistiche del campione a quelle dell'universo e non quando tratta un campione a sé, considerato come variabile statistica, da analizzarsi senza riferimenti alla variabile casuale di cui questa variabile statistica è un campione isolato.

A sua volta, la sequenza di estrazioni, a caso, della variabile statistica con distribuzione $F(x)$ da cui ha origine il campione: $x_1, x_2, \ldots, x_n$, può essere considerata un'estrazione, a caso, più complessa, da una variabile casuale a n dimensioni X_n, dove i valori: $x_1, x_2, \ldots, x_n$, nel loro insieme, sono il valore osservato di un individuo estratto, a caso, da X_n. Considerando poi i singoli valori campionari, come variabili casuali indipendenti, ciascuna con distribuzione $F(x)$, ogni relazione analitica fra: $x_1, x_2, \ldots, x_n$ è ancora una variabile casuale, con distribuzione determinata in funzione di $F(x)$. Infatti poiché i momenti e, in generale, tutte le statistiche, calcolabili sul campione, sono funzioni: $g(x_1, x_2, \ldots, x_n)$, dei valori campionari, si può calcolare la distribuzione di queste statistiche, cioè di ogni variabile casuale G, in funzione di $F(x)$. Allora se si estraggono più campioni di n elementi, da una popolazione, e si calcola la statistica: $g(x_1, x_2, \ldots, x_n)$, per ciascuno di essi, la sequenza dei valori g, così ottenuti, costituisce una serie di estrazioni, a caso, da una nuova variabile casuale G la cui distribuzione è quella della corrispondente statistica campionaria. In questo modo, si dispone di distribuzioni di medie campionarie, di valori quadratici medi campionari, di varianze campionarie, di coefficienti di correlazione campionari e così via.

Per n grande, la funzione di distribuzione $F^*(x)$ del campione approssima la funzione di distribuzione $F(x)$ della popolazione, e la frequenza relativa di un valore x_i converge, in probabilità, alla probabilità di x_i ed analogamente, in generale, una statistica campionaria converge, in probabilità, alla corrispondente statistica della popolazione, quando $n \to \infty$. Di conseguenza, una qualsiasi statistica campionaria può essere interpretata come una *stima* della corrispondente caratteristica dell'universo, tanto più significativa, quanto più numeroso è il campione da cui è calcolata.

Tuttavia il problema della stima di una statistica teorica è assai più complesso e richiede una trattazione più approfondita, dato che il metodo con cui si ricava la stima influenza le caratteristiche della stima stessa, nella sua funzione di rappresentazione della statistica teorica.

4.2. Distribuzione delle medie campionarie per campioni numerosi

Per un corollario del teorema centrale, la media aritmetica $\overline{X}$ di n variabili casuali x_i, tutte uguali fra loro, è una variabile asintoticamente Normale: $N(M,\sigma)$ con media M e sqm: $\sigma_{\overline{x}} = \sigma/\sqrt{n}$.

Infatti se si compiano n prove o estrazioni, a caso, con ripetizione, da una popolazione rappresentata da una variabile casuale X, con media M e varianza σ^2, sulla cui funzione di distribuzione non si fa alcuna ipotesi, si hanno n individui, in possesso dei valori argomentali: $a_1,\, a_2, \dots,\, a_n$, distinti o coincidenti. La media degli n valori:

$$\overline{x}_a = \frac{1}{n}\left(a_1 + a_2 + \dots + a_n\right) \tag{2.1}$$

fornisce un valore medio campionario della variabile casuale X, legata alla popolazione, che si può confrontare con il valore medio dell'universo M.

Eseguendo una seconda, una terza, ecc. serie di n prove, si ottengono altri campioni ed altri valori campionari, non tutti uguali fra loro e, in generale, differenti dalla media M dell'universo:

$$\overline{x}_b = \frac{1}{n}\left(b_1 + b_2 + \dots + b_n\right) \qquad \dots \qquad \overline{x}_k = \frac{1}{n}\left(k_1 + k_2 + \dots + k_n\right) \qquad \dots$$

Questi valori medi campionari $\overline{x}$ sono risultati di estrazioni, a caso, eseguite su una nuova variabile casuale $\overline{X}$ di tipo complesso, così definita:

$$\overline{X} = \frac{1}{n}\left(X_1 + X_2 + \dots + X_n\right) \tag{2.2}$$

dove: $X_1 + X_2 + \dots + X_n$, sono n variabili casuali indipendenti, con gli stessi valori argomentali di X e la sua stessa funzione di distribuzione, ovvero tutte coincidenti con X (dato che l'estrazione è con ripetizione e così la variabile casuale, legata alla popolazione, è sempre la stessa, prima di ogni estrazione).

Nella prima serie di n prove, dalle variabili casuali: $X_1, X_2, \dots, X_n$, sono estratti rispettivamente gli individui: $a_1,\, a_2, \dots,\, a_n$, e conseguentemente da $\overline{X}$ risulta estratto il valore $\overline{x}_a$; nella seconda serie di n prove, dalle variabili casuali: $X_1, X_2, \dots, X_n$, sono estratti: $b_1,\, b_2, \dots,\, b_n$, cui corrisponde una nuova estrazione da $\overline{X}$, con il valore $\overline{x}_b$. Pertanto $\overline{X}$ è una variabile casuale, dove le determinazioni empiriche (o estrazioni, a caso, con ripetizione) effettuate su essa, in più serie di prove, forniscono altrettanti valori medi campionari della X. Allora definita la $\overline{X}$ mediante la (2.2), è atteso il suo valore medio ed il suo sqm.

- Il calcolo della media è immediato, per la proprietà seconda del paragrafo 2.5:

$$M\left(\overline{X}\right) = \frac{1}{n}\left(M\left(X_1\right) + M\left(X_2\right) + \cdots + M\left(X_n\right)\right) = M_x = M \; . \tag{2.3}$$

  Infatti ciascuna delle variabili casuali: $X_1, X_2, \ldots, X_n$, coincide con X e ha valore medio M. Di conseguenza, per la definizione data dalla variabile casuale $\overline{X}$, il valore medio campionario, ricavato da gruppi di n osservazioni o prove (cioè da estrazioni, a caso, con ripetizione), al variare del gruppo, oscilla intorno al valore medio della variabile casuale rappresentativa della popolazione sulla quale si eseguono le prove.
- Volendo apprezzare l'ordine di grandezza di queste oscillazioni bisogna procurarsi la varianza della variabile casuale $\overline{X}$. Infatti per la (2.3), la media M della popolazione coincide con la media di $\overline{X}$ e le differenze, fra le medie campionarie e la media M, sono coincidenti con gli scarti da $\overline{X}$.

  In base alla sua definizione, $\overline{X}$ risulta essere combinazione lineare di n variabili casuali indipendenti i cui coefficienti sono $1/n$, cosicché si ha, applicando a $\overline{X}$ la legge di propagazione degli scarti:

$$\sigma^2\left(\overline{X}\right) = \frac{1}{n}\left(\sigma^2\left(X_1\right) + \sigma^2\left(X_2\right) + \cdots + \sigma^2\left(X_n\right)\right) = \frac{\sigma^2}{n} \; . \tag{2.4}$$

  Infatti essendo X_i coincidenti, le corrispondenti varianze $\sigma^2\left(X_i\right)$ sono tutte uguali tra loro. In base alla (2.4), il valore quadratico medio della differenza fra la media aritmetica di un campione di n elementi estratti, a caso, da X, e il valore medio M di X, è uguale alla varianza di X divisa per n. In questo modo, l'approssimazione delle medie aritmetiche campionarie, rispetto al valore medio della variabile casuale (da cui i campioni sono estratti), cresce proporzionalmente alla radice quadrata della numerosità del campione. Con un'espressione più precisa, si può poi dire che la media campionaria converge, in probabilità, alla media dell'universo quando $n \to \infty$.

Questo risultato permette di giustificare l'ipotesi statistica della perfetta indipendenza fra gli argomenti di una variabile statistica doppia (enunciata nel paragrafo 1.10). Infatti se si considerano le diverse popolazioni parziali di una variabile statistica doppia, come campioni estratti, a caso, da una più generale popolazione primaria (od universo), queste popolazioni parziali hanno medie campionarie oscillanti intorno alla media dell'universo, con ampiezza tanto minore, quanto maggiore è il numero di individui componenti. Viceversa se tante popolazioni parziali hanno medie poco diverse fra loro, si può supporre che siano ottenute con un processo di campionamento casuale da una variabile casuale primaria. Di conseguenza, se le popolazioni parziali di ogni colonna di una variabile statistica doppia hanno media, secondo Y, uguale o quasi uguale, queste sono ottenute per estrazioni, a caso, e così senza alcuna influenza del valore argomentale X proprio di ciascuna colonna.

I risultati trovati corrispondono ad uno schema di campionamento bernoulliano, con ripetizione, dove da un universo di numerosità N si possono estrarre N^n, campioni di numerosità n. Se l'estrazione è eseguita senza ripetizione (o in blocco), si ha un minore numero di possibili campioni e lo sqm della variabile casuale delle medie campionarie è così modificato:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$ essendo: $\binom{N}{n}$ il numero dei possibili campioni

Il fattore aggiunto, detto moltiplicatore finito, è trascurabile, quando l'universo è numeroso ed il campione piccolo rispetto ad esso.

Esempio 4.2.1

600 sferette da cuscinetti a sfera pesano in media $4.10\,gr$ l'una, con sqm di $0.30\,gr$. Si cerca la probabilità che un campione di 100 sferette, estratto in blocco (dal gruppo), pesi complessivamente fra 403 e 408 $grammi$.

Il peso totale è fra 403 e 408 $grammi$, se il peso medio, cioè la media dei pesi delle sferette del campione, è fra 4.03 e 4.08 $grammi$. La distribuzione dei pesi medi è approssimativamente Normale, con media: $M_{\bar{x}} = M_x = 4.10$ $grammi$, e sqm:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{0{,}30}{\sqrt{100}}\sqrt{\frac{600-100}{600-1}} = 0{,}027$$

In unità standard, si ha:

$$4.03 \;\rightarrow\; (4.03-4.10)/0.027 = -2.59$$
$$4.08 \;\rightarrow\; (4.08-4.10)/0.027 = -0.74$$

cosicché la probabilità richiesta è l'area, sotto la curva Normale, compresa fra: $z = -2.59$ e $z = -0.74$, pari a: 0.2248.

Esempio 4.2.2

Le lampadine della fabbrica A hanno durata media di 1700 ore e sqm di 250 ore, mentre quelle della fabbrica B hanno una durata media di 1300 ore e sqm di 200 ore. Esaminati campioni di 90 pezzi di ciascuna fabbrica, si cerca la probabilità che le lampadine della fabbrica A abbiano una durata media di almeno 350 ore maggiore di quella delle lampadine della fabbrica B (cioè che la differenza fra le due durate medie sia di almeno 350 ore).

La distribuzione delle differenze fra medie campionarie è asintoticamente Normale, con media e varianza date dalle consuete proprietà delle operazioni sulle variabili casuali. Per le medie campionarie dalle popolazioni A e B, si ha rispettivamente:

$$M_{\bar{x}_A} = 1700 \qquad \sigma^2_{\bar{x}_A} = \sigma^2_A / n_A = (250)^2/90$$

$$M_{\bar{x}_B} = 1300 \qquad \sigma^2_{\bar{x}_B} = \sigma^2_B / n_B = (200)^2/90$$

ed anche, per la variabile casuale delle differenze di medie campionarie, si ha:

$$M\left(\bar{X}_A - \bar{X}_B\right) = M_{\bar{x}_A} - M_{\bar{x}_B} = 1700 - 1300 = 400$$

$$\sigma\left(\bar{X}_A - \bar{X}_B\right) = \sqrt{\sigma^2_{\bar{x}_A} + \sigma^2_{\bar{x}_B}} = \frac{\sqrt{(250)^2 + (200)^2}}{90} = 33.7$$

cosicché la variabile casuale standardizzata per la differenza di medie è:

$$z = \frac{\left(\bar{x}_A - \bar{x}_B\right) - M\left(\bar{X}_A - \bar{X}_B\right)}{\sigma\left(\bar{X}_A - \bar{X}_B\right)} = \frac{350 - 400}{33.7} = -1.48$$

e la probabilità di: $z \geq -1.48$, risulta 0.9306.

4.3. Distribuzione dei momenti di campioni numerosi

Con procedimento analogo, si ricava la media e la varianza della variabile casuale costituita dai momenti calcolati sui campioni (il calcolo è fatto solo per il momento del secondo ordine, indicando l'estensione ai momenti di ordine superiore).

Sia M_2 il momento di secondo ordine, rispetto allo zero, o valore quadratico medio, di una variabile casuale X dalla quale si estraggono dei campioni di numerosità n. Un valore campionario del valore quadratico medio è:

$$v_a^2 = \frac{1}{n}\left(a_1^2 + a_2^2 + \cdots + a_n^2\right). \tag{3.1}$$

Al variare del gruppo degli n risultati delle estrazioni, usati nella (3.1), in generale, varia anche il valore di v^2 ed i diversi valori ottenuti: $v_a^2, v_b^2, \cdots, v_k^2$, possono considerarsi estrazioni, a caso, da una nuova variabile casuale:

$$V^2 = \frac{1}{n}\left(X_1^2 + X_2^2 + \cdots + X_n^2\right)$$

dove: $X_1, X_2, \cdots, X_n$, non differiscono fra loro e dalla variabile casuale X , ma indicate diversamente, perché le estrazioni eseguite su X_1^2 sono indipendenti dalle estrazioni eseguite su X_2^2 , ecc. Alla V^2 si applicano tutte le considerazioni fatte a proposito di $\overline{X}$, sostituendo X_i^2 a X_i , cosicché i risultati si riassumono nelle seguenti due formule:

$$M\left(V^2\right) = M\left(X^2\right) = M_2 \tag{3.2}$$

$$\sigma^2\left(V^2\right) = \frac{M_4 - M_2^2}{n} \ . \tag{3.3}$$

ove il termine M_4 indica il momento di quarto ordine della variabile casuale, rispetto allo zero.

Le (3.2) e (3.3) dicono che v^2 , valore quadratico medio campionario dedotto da un gruppo di n estrazioni, a caso, dalla variabile casuale X , al variare del campione, oscilla intorno al valore quadratico medio M_2 di X . La varianza di V^2 coincide con il valore quadratico medio degli scarti fra valore quadratico medio campionario e valore quadratico medio della popolazione, e la variabile casuale Z è asintoticamente una variabile casuale Normale standardizzata:

$$\sigma^2\left(V^2\right) = \frac{M_4 - M_2^2}{n} \qquad Z = \frac{V^2 - M_2}{\sqrt{\left(M_4 - M_2^2\right)/n}}$$

In generale, se V^k è la variabile casuale costituita dai momenti di *k*-esimo ordine e polo zero, si ha:

$$M\left(V^k\right) = M_k \qquad \sigma^2\left(V^k\right) = \frac{M_{2k} - M_k^2}{n}$$

da cui, per $k = 1$, si ricavano media e varianza della variabile casuale $\overline{X}$ delle medie campionarie.

4.4. Distribuzione di varianze campionarie per campioni numerosi

Analogamente a quanto fatto per le statistiche M e M_2 , occorre studiare il problema della divergenza fra il valore campionario ed il valore dell'universo per la varianza. Infatti un valore della varianza campionaria, se n sono le prove eseguite, è dato dalla relazione:

$$s_a^2 = \frac{1}{n}\left(\left(a_1 - \overline{x}_a\right)^2 + \left(a_2 - \overline{x}_a\right)^2 + \cdots + \left(a_n - \overline{x}_a\right)^2\right) \tag{4.1}$$

nella quale compaiono gli scarti rispetto al valore medio campionario $\bar{x}_a$ delle n estrazioni (in generale, con il variare del gruppo delle n estrazioni, varia il valore s^2). I valori ottenuti sono estrazioni, a caso, da una variabile casuale di tipo complesso, così definita:

$$S^2 = \frac{1}{n}\left((X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right) \tag{4.2}$$

dove compare la variabile casuale delle medie campionarie:

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

Dalla (4.2), applicando le proprietà del paragrafo 2.5 e le (2.3), (2.4), si può calcolare la media delle varianze campionarie:

$$\begin{aligned} M(S^2) &= \frac{1}{n} M\left((X_1^2 + X_2^2 + \cdots + X_n^2) - 2\bar{X}(X_1 + X_2 + \cdots + X_n) + n\bar{X}^2\right) = \\ &= \frac{1}{n} n M_2(X) - 2M\left(\bar{X}\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right) + M(\bar{X}^2) = \\ &= M_2(X) - 2M(\bar{X}^2) + M(\bar{X}^2) = M_2(X) - M_2(\bar{X}) \\ &= M_2(X) - \left(\sigma^2(\bar{X}) + M^2(\bar{X})\right) = \left(M_2(X) - M^2(X)\right) - \frac{\sigma^2(X)}{n} \end{aligned}$$

$$\sigma^2(X) - \frac{\sigma^2(X)}{n} = \frac{n-1}{n}\sigma^2(X) \tag{4.3}$$

Questa relazione indica che la media dei valori delle varianze campionarie s^2 , dedotte da gruppi di n osservazioni, non coincide con la varianza della popolazione σ_x^2 . Infatti le stime s^2 oscillano, al variare del campione, intorno al valore σ_x^2 , moltiplicato per il coefficiente: $(n-1)/n$, generalmente detto coefficiente di *bias* (in inglese: *to bias* = influenzare, spingere in una data direzione). Da questo risultato, si deduce che volendo una stima della varianza della popolazione, cioè un indice che, pur non coincidendo con essa, oscilli nel suo intorno, occorre moltiplicare la varianza campionaria deviata s^2 , per il coefficiente $n/(n-1)$, cosicché:

$$s_c^2 = \frac{n}{n-1}s^2 = \frac{n}{n-1}\frac{1}{n}\left(v_1^2 + v_2^2 + \cdots + v_n^2\right) = \frac{1}{n-1}\sum v_i^2 \tag{4.4}$$

dove: $v_i = a_i - \bar{x}_a$.

La variabile casuale $S_c^2 = n/(n-1) \cdot S^2$, ossia la S^2 corretta del *bias*, ha una media convergente, in probabilità, alla varianza dell'universo σ_x^2 e così l'espressione (4.4) può essere presa come una stima corretta (cioè non deviata) della varianza dell'universo. La (4.4) è di uso comunissimo e si ritrova molto frequentemente nella teoria degli errori di misura. Tuttavia un'analisi più approfondita può evidenziare che, per ottenere una stima di σ_x^2 è indifferente usare la (4.1) o la (4.4), in quanto la differenza fra esse ha un ordine di grandezza inferiore al grado di approssimazione che, tramite queste, si può raggiungere nel calcolo di σ_x^2.

Per valutare quest'approssimazione bisogna calcolare la varianza di S^2 o di $S_c^2 = n/(n-1) \cdot S^2$, cosa abbastanza lunga e complicata. Inoltre le formule risultanti sono parecchio complesse e possono essere semplificate solo ipotizzando che la variabile casuale X da cui si sono estratti i campioni sia Normale. Infatti in questo caso, si ha:

$$\sigma^2(S^2) = \frac{2(n-1)}{n^2}\sigma^4(X)$$

$$\sigma^2(S_c^2) = \frac{n^2}{(n-1)^2}\frac{2(n-1)}{n^2}\sigma^4(X) = \frac{2}{n-1}\sigma^4(X) \cong \frac{2}{n}\sigma^4(X). \qquad (4.5)$$

La seconda delle (4.5), di uso più frequente, indica che lo sqm della stima della varianza dell'universo, ottenuta da un campione come: $\sum v_i^2 / (n-1)$, è uguale a: $\sigma(S_c^2) = \sigma_x^2\sqrt{2/(n-1)}$. Dato che, per ipotesi, i campioni sono numerosi, si ottiene lo sqm, con la stessa approssimazione, con la formula:

$$\sigma(S_c^2) = \sigma^2(X)\sqrt{2/\ n}$$

Sempre nell'ipotesi che X sia Normale, si dimostra che la variabile casuale: $S_c = S\sqrt{n/(n-1)}$, dove:

$$S_c = \sqrt{\frac{\sum v^2}{n-1}}$$

è un'estrazione, a caso, e ha uno sqm:

$$\sigma(S_c) = \frac{\sigma(X)}{\sqrt{2(n-1)}} \cong \frac{\sigma(X)}{\sqrt{2n}}\ . \qquad (4.6)$$

Per campioni molto numerosi: $n \geq 100$, la distribuzione delle varianze campionarie è approssimativamente Normale. In questo caso, l'uso della (4.1) o della (4.4) è veramente indifferente; tuttavia, per evitare confusioni, si conviene di indicare diversamente rispettivamente la varianza campionaria:

$$s^2 = \frac{1}{n}\sum v^2 \tag{4.7}$$

e la stima non deviata della varianza dell'universo, ottenuta dal campione:

$$\sigma^2 = s_c^2 = \frac{1}{n-1}\sum v^2 \tag{4.8}$$

.

4.5. Schema di campionamento stratificato

Dalla variabile casuale, costituita dalle medie campionarie, si può giungere ad una stima della media della variabile casuale, rappresentante l'intera popolazione. Per ottenere una stima della media della popolazione sufficientemente accurata, occorre che le oscillazioni di queste medie campionarie siano contenute, nella massima misura possibile, cosa che può essere ottenuta con opportuni schemi di campionamento.

La varianza delle medie campionarie è funzione della varianza σ_x^2 della variabile casuale da cui i campioni sono estratti e della numerosità dei campioni:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} \quad , \tag{5.1}$$

per cui, dato σ_x^2, l'unico modo per diminuire $\sigma_{\bar{x}}^2$ è aumentare il numero degli elementi che costituiscono il campione.

Invece un sistema che consente di diminuire $\sigma_{\bar{x}}^2$, senza accrescere n, è costituito dallo schema di Poisson (o di *stratificazione semplice*), dove si suppone di suddividere l'universo, in universi parziali, di numerosità variabile, con propri valori argomentali.

Le popolazioni parziali sono dette anche *strati* da cui deriva il nome del procedimento. Ad esempio, dovendo fare una ricerca statistica sulla popolazione di una grande città, gli strati possono essere costituiti dai vari quartieri della città stessa, ciascuno sufficientemente omogeneo, ma con caratteristiche diverse da quelle degli altri quartieri.

Siano s gli strati, caratterizzati ciascuno dalla propria numerosità N_i, dalla media $\bar{x}_i$ e dalla varianza σ_i^2, secondo lo schema:

$$\begin{array}{cccccc} N_1 & N_2 & \ldots & N_i & \ldots & N_s \\ \bar{x}_1 & \bar{x}_2 & \ldots & \bar{x}_i & \ldots & \bar{x}_s \\ \sigma_1^2 & \sigma_2^2 & \ldots & \sigma_i^2 & \ldots & \sigma_s^2 \end{array}$$

Se con N, M_x (indicata semplicemente con $\bar{x}$) e σ_x^2 indicano rispettivamente la numerosità, dell'intero universo, la sua media e la sua varianza, si hanno le seguenti relazioni, avendo posto: $p_i = N_i / N$.

$$N = \sum N_i \qquad\qquad M_x = \bar{x} = \frac{\sum N_i \bar{x}_i}{N} = \sum p_i \bar{x}_i$$

Si possono inoltre definire la media delle varianze:

$$\bar{\sigma}^2 = \sum p_i \sigma_i^2$$

e la media degli sqm:

$$\bar{\sigma} = \sum p_i \sigma_i$$

di cui si trovano poi le relazioni con σ_x^2 .

Nello schema di Poisson, i campioni sono formati estraendo, dai diversi strati, un numero di elementi tale che risulti: $n_1 + n_2 + \cdots + n_i + \cdots + n_s = n$, dove n è la numerosità dei campioni. Ogni campione contiene elementi provenienti da tutti gli strati e scelti, a caso, nell'ambito di ciascuno strato, dove questi strati sono costruiti in modo che, in ciascun campione, è costante il numero degli elementi estratti dai singoli strati.

Il procedimento più semplice, per fissare il criterio di scelta degli n_i elementi, stabilisce una proporzionalità fra n_i ed N_i , estraendo da ciascuno strato un numero di elementi proporzionale alla sua numerosità, in modo che, essendo: $\sum n_i = n$, si deve avere:

$$n_i = n\frac{N_i}{N} = np_i$$

cosicché ciascuno strato è rappresentato, nel campione, con un *peso* proporzionale alla sua numerosità (cosa che non accade nello schema bernoulliano).

Procedendo come indicato, nell'ipotesi che l'estrazione degli elementi, per la campionatura, sia fatta con ripetizione, si può dimostrare (come nel paragrafo 3.2), che la varianza delle medie campionarie, fornita dallo schema di Poisson, è:

$$\sigma_{\bar{x}}^2 = \frac{\bar{\sigma}^2}{n} \qquad (5.2)$$

dove $\bar{\sigma}^2$ è il valore medio delle varianze σ_i^2 , assumendo p_i come frequenze relative.

La relazione fra la varianza σ_x^2 di tutta la popolazione e la media $\bar{\sigma}^2$ delle varianze degli strati evidenzia che, eseguendo il campionamento con lo schema di Poisson, la varianza delle medie campionarie è inferiore a quella con lo schema di Bernoulli, come mostra l'espressione:

$$\sigma_x^2 = \overline{\sigma}^2 + \sigma_\alpha^2 \tag{5.3}$$

dove σ_α^2 è la varianza delle medie degli strati:

$$\sigma_\alpha^2 = \sum_{i=1}^{s} p_i \left(\overline{x}_i - \overline{x}\right)^2$$

La (5.3), introducendo le definizioni delle tre varianze moltiplicate per N, ha forma:

$$\begin{aligned}\sum_{j=1}^{N}\left(x_j - \overline{x}\right)^2 &= \sum_{i=1}^{s} N_i \sum_{k=1}^{N_i} \frac{1}{N_i}\left(x_k - \overline{x}_i\right)^2 + \sum_{i=1}^{s} N_i \left(\overline{x}_i - \overline{x}\right)^2 = \\ &= \sum_{i=1}^{s} N_i \left(M\left(X_i^2\right) - \overline{x}_i^2\right) + \sum_{i=1}^{s} N_i \left(\overline{x}_i - \overline{x}\right)^2\end{aligned} \tag{5.4}$$

e sviluppando l'espressione: $N\sigma_x^2$, cioè del primo membro della (5.4), si ottiene:

$$N\sigma_x^2 = \sum_{j=1}^{N}\left(x_j - \overline{x}\right)^2 = \sum_{j=1}^{N} x_j^2 + N\overline{x}^2 - 2\overline{x}\sum_{j=1}^{N} x_j = \sum_{j=1}^{N} x_j^2 - N\overline{x}^2 \tag{5.5}$$

Il valore di $\overline{\sigma}^2$, ancora moltiplicato per N, è:

$$N\overline{\sigma}^2 = \sum_{i=1}^{s} N_i \left(M\left(X_i^2\right) - \overline{x}_i^2\right) = \sum_{i=1}^{s} N_i \sum_{k=1}^{N_i} \frac{x_k^2}{N_i} - \sum_{i=1}^{s} N_i \overline{x}_i^2 = \sum_{j=1}^{N} x_j^2 - \sum_{i=1}^{s} N_i \overline{x}_i^2 \ . \tag{5.6}$$

e quello di σ_α^2, sempre moltiplicato per N :

$$N\sigma_\alpha^2 = \sum_{i=1}^{s} N_i \left(\overline{x}_i^2 - 2\overline{x}_i\overline{x} + \overline{x}^2\right) = \sum_{i=1}^{s} N_i \overline{x}_i^2 - N\overline{x}^2 \ . \tag{5.7}$$

Sommando la (5.6) e la (5.7), l'espressione del secondo membro della (5.4) diventa:

$$N\overline{\sigma}^2 + N\sigma_\alpha^2 = \sum_{j=1}^{N} x_j^2 - \sum_{i=1}^{s} N_i \overline{x}_i^2 + \sum_{i=1}^{s} N_i \overline{x}_i^2 - N\overline{x}^2 = \sum_{j=1}^{N} x_j^2 - N\overline{x}^2$$

identica alla (5.5) e sostituendo nella (5.2) l'espressione di $\overline{\sigma}^2$ ricavata dalla (5.3), si ottiene finalmente la varianza delle medie campionarie:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} - \frac{\sigma_\alpha^2}{n} \le \frac{\sigma_x^2}{n}$$

evidentemente sempre minore di σ_x^2 / n e tanto minore quanto maggiore è σ_α^2, cioè quanto più grande è la varianza tra le medie degli strati. Da questa cosa, si deduce che la scelta degli strati è bene sia fatta con il criterio della più grande diversificazione possibile, rispetto all'attributo preso in considerazione.

La campionatura, nello schema descritto, è fatta unicamente con il criterio di proporzionalità rispetto al numero degli individui contenuto nei singoli strati. Questo criterio, arbitrario (anche se logico), può essere perfezionato, per ottenere una maggiore riduzione di $\sigma_{\bar{x}}^2$. Il perfezionamento consiste nello stabilire che il numero degli individui da estrarre, da ciascuna popolazione parziale, risulti proporzionale, sia ancora alla numerosità dello strato, sia alla variabilità interna dello strato stesso, cioè al suo scarto quadratico medio. Pertanto il numero di individui da prendere in considerazione per ciascuno strato risulta:

$$n'_i = n_i \frac{\sigma_i}{\bar{\sigma}} . \tag{5.8}$$

dove σ_i rappresenta lo sqm dello strato e $\bar{\sigma}$ la variabilità media di tutti gli strati: $\bar{\sigma} = \sum p_i \sigma_i$.

- Se lo sqm dello strato è maggiore di $\bar{\sigma}$, la numerosità ottimale n'_i è maggiore di quella proporzionale n_i e viceversa, in caso contrario.
- Se poi gli sqm degli strati sono circa uguali, allora si ha: $\sigma_i / \bar{\sigma} \cong 1$, ed i due criteri, proporzionale ed ottimale, coincidono.

Si può dimostrare che, con il criterio ottimale e supponendo un universo sufficientemente grande, la varianza della variabili casuali delle medie campionarie risulta:

$$\sigma_{\bar{x}}^2 = \frac{(\bar{\sigma})^2}{n} \tag{5.9}$$

La relazione tra: $(\bar{\sigma})^2$, $\bar{\sigma}^2$ e σ_x^2, è ricavata introducendo la varianza fra gli sqm degli strati:

$$\sigma_\beta^2 = \sum_{i=1}^{s} p_i (\sigma_i - \bar{\sigma})^2 = \sum_{i=1}^{s} p_i \sigma_i^2 - (\bar{\sigma})^2 = \bar{\sigma}^2 - (\bar{\sigma})^2$$

cosicché la (5.9), dividendo per n entrambi i membri di questa ultima relazione, diventa:

$$\sigma_{\bar{x}}^2 = \frac{\bar{\sigma}^2}{n} - \frac{\sigma_\beta^2}{n} \le \frac{\sigma_x^2}{n} - \frac{\sigma_\alpha^2}{n} \le \frac{\sigma_x^2}{n}$$

Infatti con il campionamento Poissiano ottimale si realizza un ulteriore guadagno, rispetto al campionamento proporzionale (ed ovviamente rispetto all'assenza di stratificazione), tanto più grande, quanto più elevata è la varianza della variabilità degli strati [29].

---

[29] Nei lavori di controllo di qualità e, in generale, controllo e confronto d'ipotesi si pone spesso il problema di eseguire campionamenti su una base di dati che, per dimensione, eccede la possibilità di indagini esaustive. Le procedure di campionamento permettono di estrarre campioni, numericamente e statisticamente significativi, dalle corrispondenti popolazioni.
Una strategia logistica, per il campionamento delle basi di dati oggetto di studio, prende in considerazione il numero di ripetizioni dello stesso tipo ed adotta, quale misura di campionamento:

- la totalità, quando la numerosità delle stesse è inferiore ad un valore di soglia prefissato;
- un certo valore percentile, quando la numerosità supera il suddetto valore di soglia, ma non eccede, nel contempo, un secondo valore di soglia prefissato;
- un dato valore (assoluto), quando è superiore al secondo valore di soglia.

Lo scopo di questa strategia è prevenire, insieme, un campionamento poco significativo di campioni poco numerosi ed uno eccessivo di campioni enormemente grandi. La curva logistica è una funzione algebrica, ben nota in demografia, per lo studio dell'accrescimento di popolazioni in ambienti limitati.
Questa curva è caratterizzabile tramite tre parametri, aventi il significato geometrico descritto per presentare la voluta approssimazione lineare. Si riportano, di seguito, alcune espressioni analitiche del suo studio di funzione:

$$y = a + (b-a)e^{\frac{-4(b-a)}{e^2cx}} \qquad lim_{x \to 0+}\, y = a \qquad lim_{x \to \infty}\, y = b$$

$$y^I = (b-a)e^{\frac{-4(b-a)}{e^2cx}} \frac{4(b-a)}{e^2cx^2}$$

$$y^{II} = (b-a)c^{\frac{-4(b-a)}{e^2cx}} \frac{b(b-a)^2}{e^4c^2x^4}\left(1-\frac{e^2cx}{2(b-a)}\right) \qquad y^{II} = 0 \text{ per } \qquad x = \frac{2(b-a)}{e^2c} = K$$

$$y(K) = a + (b-a)e^{-2} \qquad y^I(K) = c$$

E' chiaro che più il campione è grande, più le stime sono attendibili. D'altra parte, al crescere delle dimensioni del campione aumentano i costi e le difficoltà tecniche ed organizzative, oltretutto con rischi, non troppo remoti, di non riuscire più a garantire l'indipendenza fra le osservazioni e pertanto la bernoullianità del campione ottenuto. L'esistenza di un limite superiore va nella direzione della ricerca di un equilibrio tra la diminuzione della varianza e la complessità dei problemi.

Un primo esempio conclusivo fa riferimento alla stima di frequenze, in particolare alla ricerca dei limiti di confidenza di una frequenza osservata, per una popolazione distribuita binomialmente. Data la distribuzione binomiale, approssimata da una Normale, corretta per continuità: $z = (x - n\hat{p}) / \sqrt{n\hat{p}q}$, questa equazione vale per il campionamento con riposizionamento, ma in caso di campionamento senza riposizionamento, ovvero con correzione per la popolazione finita, occorre apportare la correzione per la popolazione finita: $\sqrt{(N-n)/(N-1)}$, dove N è la numerosità della popolazione e *n* la numerosità del campione, cosicché si ha, essendo a l'intervallo di confidenza:

$$n = \frac{N(a+1)}{1+Na^2} = \frac{N}{1+Na^2}$$

Un secondo esempio conclusivo prende in considerazione altri tipi di test, ad esempio, per la distribuzione χ^2. Infatti nel caso di campionamento con ripetizione, ricordando che la varianza della varianza (se le osservazioni sono normali) ha espressione: $\hat{\sigma}^2_{\hat{\sigma}^2} = \hat{\sigma}^4/2n$, essendo sempre a l'intervallo di confidenza e ν i gradi di libertà della variabile casuale χ^2 , si ottiene:

$$n = \frac{\nu^2 N}{\nu^2 + 2a^2N}$$

La tripartizione nella strategia di campionamento proposta costituisce l'approssimazione lineare di una curva logistica, avente due asintoti orizzontali pari ai due valori di soglia ed un punto intermedio di flesso a tangente inclinata, dove il coefficiente angolare della retta tangente coincide con il suddetto valore percentile. Si può osservare l'esistenza di un limite inferiore per la numerosità del campione, corrispondente ad un valore al di sotto del quale è necessario sottoporre a test tutti gli elementi della popolazione, e di un limite superiore, corrispondente ad un valore al di sopra del quale il numero degli elementi della popolazione da sottoporre a test è approssimativamente lo stesso. Si osservi tuttavia come aumentino, all'aumentare della precisione richiesta, sia il limite inferiore, sia quello superiore. Per quanto riguarda infine il coefficiente angolare c, si può procedere per tentativi, a partire da un valore iniziale ottenuto dal confronto tra il valore K , ascissa della curva logistica nel punto di flesso, ed il valore N , corrispondente a $n = y(K)$.