

STATISTICA

Corso del Docente Fucci
C.d.L. in Informatica - 2° Anno

Il termine “statistica” deriva dal latino medievale “status”, ovvero Stato, che vuole indicare l’ordinamento politico. E’ stata chiamata statistica l’area disciplinare (scientifica) che si occupava di fornire descrizioni relative ad uno Stato oppure descrizioni comparative di vari Stati. Queste descrizioni ebbero inizio in effetti già con **Aristotele** nel 350 a.c. e proseguirono nelle epoche storiche successive. Pur tuttavia la parola “statistica” apparve per la prima volta nell’enciclopedia britannica nel 1797 essendo stata coniata ufficialmente, sembra, dal professore tedesco Gottfried **Acheunwall** (1719-1772) verso la metà del ‘700, che la definì come la descrizione delle cose più notevoli di uno Stato. Il significato che essa ha mantenuto fino agli inizi del 1800 è stato dunque quello di descrizione o censimento con successiva analisi dei dati prodotti da tali operazioni. Il significato della Statistica si è poi esteso nel corso dei tempi ben oltre tale interpretazione originaria. Al giorno d’oggi si parla di due tipi di Statistica: una **Statistica descrittiva** o **metodologica**, e una **Statistica matematica** o **inferenziale**. Per *Statistica descrittiva* si intende il complesso delle tecniche di cui si serve lo sperimentatore per raccogliere, elaborare e rappresentare gruppi di dati osservati che in generale sono in numero elevato e in forma disordinata. Lo scopo finale è quello di ricavare dalle osservazioni opportuni elementi di sintesi. Questi possono essere numeri (come la *media*, la *mediana*, la *moda*, la *varianza*, ecc) oppure informazioni rappresentabili sotto forma di grafici, tavole, carte (geologiche o numeriche), diagrammi, etc. In questa situazione si cerca solo di descrivere il fenomeno osservato senza cercare di interpretare o generalizzare più di tanto riguardo alla totalità dei potenziali dati osservabili di cui quelli disponibili costituiscano solo una piccola parte. Al giorno d’oggi queste informazioni sono molto utili ma accade spesso in genere che si voglia saperne di più e spenderne meno. Ciò significa non dover censire sempre tutto per avere le informazioni che ci servono ma estrarre le stesse cose dalla conoscenza dei dati osservati che sono a nostra disposizione. In termini statistici vogliamo trarre conclusioni sull’intera popolazione partendo dalla conoscenza del campione osservato. Ciò implica che molti dati saranno a noi ignoti ma che dovremmo cercare un metodo per averne una qualche conoscenza partendo dai dati disponibili

cioè senza averne una conoscenza sperimentale diretta. Si cerca dunque di studiare gruppi di dati osservati per ricavare da essi un modello interpretativo che descriva al meglio la popolazione da cui essi sono estratti e questo è proprio un aspetto della *Statistica inferenziale*. E' opportuno sottolineare che il *Calcolo delle Probabilità* diventa essenziale nello studio della *Statistica inferenziale*. Per quanto riguarda la *Probabilità* diamo anzitutto alcuni cenni storici: i primi documenti che si conservano sono dovuti a Girolamo **Cardano** (1501-1576), accanito giocatore, con il suo libro "*De ludo Aleae*" scritto forse intorno al 1526 e pubblicato postumo nel 1663, e a **Galileo** con un suo scritto del 1620. Nel libro di Cardano è inserito il problema del calcolo delle probabilità delle somme dei punteggi ottenuti col lancio di tre dadi. Il risultato però contiene delle inesattezze e gli sviluppi non sono molto chiari così che i giudizi sul valore dell'opera sono controversi, però la sua importanza storica è fuori dubbio. Nello scritto di Galileo viene trattato lo stesso problema del lancio di tre dadi però con maggiore chiarezza. La nascita del *Calcolo delle Probabilità* viene abitualmente attribuita a Blaise **Pascal** (1623-1662) e fissata nella corrispondenza tra lui e Pierre **Fermat** (1601-1665) verso la metà del 1600. L'interesse di Pascal fu stimolato da **Cavalier de Mère**, spirito vivace, matematico discreto e accanito giocatore d'azzardo che si rivolse a lui per la risoluzione di vari problemi riguardanti il gioco. Il *Calcolo delle Probabilità*, sorto quindi verso la seconda metà del '600 come studio matematico dei giochi d'azzardo quali monete, dadi o carte, può essere considerato oggi una disciplina matematica che trova applicazione nello studio degli esperimenti *non deterministici* o *casuali*. Un esperimento **non deterministico** o **casuale** è un esperimento che può dar luogo ad un risultato, fra quelli possibili, non determinabile a priori in modo univoco. Il lancio di un dado è un esempio di questo tipo, in cui il risultato non è determinabile a priori potendo essere rappresentato da una qualsiasi delle sei facce del dado stesso. Questi brevi cenni hanno lo scopo di dimostrare come si presenta originariamente il concetto di probabilità e di introdurre alle definizioni o interpretazioni che ora presenteremo.

Definizione classica: La prima definizione di probabilità, detta classica, si ritrova già in Pascal e *definisce la probabilità di un evento come il rapporto fra il numero dei casi favorevoli all'evento e il numero dei casi possibili, purché questi ultimi siano tutti ugualmente possibili*. Molti però vedono in questa definizione una tautologia: bisognava sapere già prima

che significato dare alle probabilità perché non si vede bene la differenza fra “ugualmente possibili” e “ugualmente probabili”; comunque i casi risulteranno ugualmente possibili quando si ha una situazione di perfetta simmetria fisica, come ad esempio le palline nell’urna del gioco del lotto, le facce di un dado o le carte di un mazzo ben mescolato.

Più che una definizione, quella classica si può considerare una regola per la misura della probabilità di un evento in condizioni opportune, cioè quando vi sia un numero finito di alternative possibili che possono essere considerate ad esempio per motivi di simmetria ugualmente possibili. Tutto questo però sapendo già cos’è la probabilità. In conclusione sembra deduttibile la validità di questa definizione, d’altra parte essa risponde ai requisiti di essere operativa. Il campo in cui più direttamente si può applicare la definizione classica è quello di giochi, carte etc. In essi si può individuare con precisione quali siano le possibili alternative e si può ragionevolmente supporre che esse siano tutte ugualmente possibili.

Esempio: supponiamo di voler determinare la probabilità che lanciando una moneta non truccata esca testa. In questo esempio c’è solo un caso favorevole, cioè testa; e due casi possibili cioè testa e croce. Si arriva pertanto alla conclusione che la probabilità è uguale a $\frac{1}{2}$.

Definizione frequentista: L’insoddisfazione della definizione classica portò a costruire la probabilità sulle frequenze, intendendo come *frequenza relativa dei successi il rapporto fra il numero delle volte in cui l’evento si verifica e il numero delle prove effettuate*. Si giunse così alla definizione frequentista delle probabilità di un evento, intesa come *limite della frequenza relativa dei successi quando tende all’infinito il numero delle prove fatte nelle stesse condizioni*. In altre parole accettando l’ipotesi che la frequenza relativa si vada stabilizzando intorno ad un certo numero è proprio questo numero che va preso come probabilità. Anche la definizione frequentista presenta aspetti criticabili, anzitutto perché resta imprecisato il numero di prove necessario per arrivare ad un valore abbastanza stabilizzato che ci fornisca la probabilità, ma soprattutto l’esigenza che le prove successive siano fatte nelle stesse condizioni presenta problemi. A rigore una simile condizione non è mai perfettamente verificata; essa restringe l’applicabilità della definizione ad situazioni ben delimitate, come ad esempio i lanci successivi di un dado, ma a guardare bene anche in questo caso da un lancio all’altro può cambiare sia pur di

poco la situazione ambientale (umidità, pressione etc.) che influisce sul risultato; inoltre può cambiare anche il dado stesso nell'urto che riceve cadendo. Alla domanda se resterà costante la probabilità dell'evento considerato si può rispondere "sì" entro limiti di approssimazione largamente sufficienti in pratica; "no" se si richiede una costanza valida rigorosamente che possa dare una solida base alla definizione. Esistono poi situazioni in cui essa platealmente non vale; si pensi ad esempio all'incontro fra due squadre di calcio o di due tennisti. Si può osservare che anche la definizione frequentista, come quella classica, è operativa. La frequenza delle prove in cui l'evento si verifica, cioè il rapporto fra il numero delle prove favorevoli all'evento e il numero totale delle prove ha le stesse caratteristiche matematiche del rapporto fra il numero dei casi favorevoli e il numero dei casi possibili.

Esempio: se una moneta viene lanciata 1000 volte e risulta che si presenta 450 volte testa, la probabilità di testa viene stimata uguale a $450/1000$.

Definizione soggettiva: In molti casi si impone l'esigenza di considerare l'evento singolo e di riferire la probabilità a questo evento. Da questa esigenza è nata la concezione soggettiva che si può far risalire a Daniele **Bernoulli** e che è stata ripresa verso la metà del 1900 dall'italiano Bruno **De Finetti** (1906-1985). In questa impostazione *la probabilità è il grado di fiducia che una persona ha nel verificarsi dell'evento*. La probabilità perde così la caratteristica assoluta di numero intrinsecamente legato all'evento per dipendere dalla persona che la valuta e dalle informazioni disponibili. La definizione data, però, così com'è, non è operativa ed è necessaria una definizione più precisa. Uno dei modi per renderla tale è di fare riferimento alle scommesse definendo la probabilità come il prezzo equo da pagare per ricevere 1 se l'evento si verifica e 0 se l'evento non si verifica con una qualsiasi unità di misura (cioè un euro, un dollaro, etc.).

Esempio: ad una lotteria, fatta mediante i numeri della tombola, ognuno paga una certa quota (1 euro) e chi vince, avendo il suo numero estratto, riceve le 90 euro raccolte. Allora il prezzo pagato riferito al piatto di 90 euro come unità di misura è uguale a $1/90$ e rappresenta la probabilità di vittoria.

Nella definizione si è parlato però di *prezzo equo* e questo richiede una precisazione che viene data dalla seguente **condizione di equità**, detta

anche *di coerenza*, la quale stabilisce che “*non si devono valutare le probabilità in modo tale che sia possibile ottenere una vincita certa o una perdita certa*”.

Chiariamo subito questo punto: supponiamo di presentare una moneta dicendo che è truccata in modo tale che la probabilità di ottenere testa $P\{T\}$ in un lancio è $\frac{1}{2}$ mentre la probabilità di ottenere croce $P\{C\}$ è $\frac{1}{4}$. Si può obiettare subito che ciò non può andare perché la somma di $P\{T\}$ e $P\{C\}$ deve essere 1. Questa è un'esigenza intuitiva alla quale la condizione di coerenza dà una giustificazione convincente, infatti facendo contemporaneamente due scommesse, una su testa e una su croce, si pagherebbe $\frac{1}{2}$ per la prima e $\frac{1}{4}$ per la seconda, ricevendo comunque 1 con guadagno netto certo pari a $\frac{1}{4}$, in contraddizione con la condizione di coerenza, questo perché la somma di $P\{T\}$ e $P\{C\}$ è minore di 1. Se questa somma fosse maggiore di uno si avrebbe una perdita certa; la condizione di coerenza impone che sia $P\{T\}+P\{C\}=1$. Lo stesso ragionamento deve valere anche in presenza di due o più alternative. La critica più diffusa a questa impostazione è quella di essere soggettiva, cioè di fondare la probabilità sul valore dei singoli. Si è sviluppato un lungo confronto spesso polemico con gli oggettivisti che accusano l'impostazione soggettiva di rendere impossibile la comunicazione fra persone diverse e i soggettivisti che denunciano l'illusorietà della pretesa oggettività delle altre impostazioni.

Le tre definizioni date sono profondamente distanti concettualmente; si è visto inoltre che entrambe non sono esenti da critiche. Per ovviare a queste difficoltà i matematici preferiscono trattare la *Probabilità* da un punto di vista *assiomatico*, facendo uso della *Teoria degli Insiemi*. L'impostazione assiomatica è largamente preferita in ogni campo della matematica perché permette di ottenere un livello accettabile di rigore logico. Il *Calcolo delle Probabilità* non poteva sfuggire a questa esigenza di sistemazione. Si può verificare che le tre impostazioni arrivano tutte alle stesse leggi matematiche espresse dagli assiomi della probabilità e ciò rende naturale prendere tali leggi come base per una costruzione assiomatica.

La *Statistica* per lo più è interessata alla *Probabilità* soltanto per quanto riguarda i possibili risultati degli esperimenti e di solito preferisce l'interpretazione frequentista della probabilità, cioè preferisce pensare alla *Probabilità* come alla *frequenza del verificarsi di un certo evento se l'esperimento ad esso relativo fosse ripetuto un gran numero di volte*.

Inoltre la maggior parte degli statistici sono interessati soltanto agli esperimenti di tipo *ripetitivo*. Il lancio di una moneta, il getto di un dado, la lettura della temperatura giornaliera su un termometro, sono semplici esempi di esperimenti di questo tipo.

Quando si compie un esperimento il suo risultato è di solito incerto, ma se esso viene ripetuto un gran numero di volte è possibile costruire per esso un modello probabilistico da usare poi per prendere delle decisioni riguardanti l'esperimento in oggetto. Per gli esperimenti di tipo ripetitivo il modello scelto è di solito un modello per prevedere la frequenza del verificarsi di certi risultati in ripetute esecuzioni dell'esperimento. Però poiché il modello scelto altro non è che la rappresentazione di una situazione reale, le conclusioni tratte da esso saranno affidabili soltanto se esso sarà un'approssimazione sufficientemente buona della realtà in esame; quindi la prima cosa da fare dal punto di vista statistico per risolvere un dato problema è di scegliere un modello matematico, di controllarne poi l'attendibilità e di trarre quindi da esso le conclusioni risolutive. Il nostro approccio alla *Probabilità* si baserà sia sull'interpretazione frequentista che su quella assiomatica. Tenuto conto che abbiamo già discusso del concetto di *Probabilità* ai fini dell'impostazione assiomatica, restano da chiarire soltanto i concetti di esperimento o prova, e di evento che essendo intuitivi sono già stati utilizzati senza discuterli.

Spazio campione o spazio campionario: dato un esperimento casuale o non deterministico un concetto di base è quello di spazio campione. Per definirlo diciamo subito che è conveniente rappresentare i possibili risultati di un esperimento in generale per mezzo di punti nello spazio ad " n " dimensioni per $n=1,2,3,\dots$ etc. A tale scopo consideriamo il semplice esperimento del lancio di una moneta: in esso i possibili risultati sono 2, testa o croce. In questo caso è conveniente rappresentare il risultato testa per mezzo del punto di ascissa 1 sull'asse x e il risultato croce col punto di ascissa 0 come in figura:

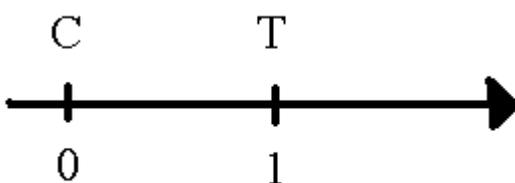


Fig.1

Questa scelta è conveniente perché il valore dell'ascissa corrisponde al numero di teste ottenute nel lancio. Se l'esperimento fosse consistito nel lanciare la moneta due volte i risultati possibili sarebbero stati 4, cioè *TT*, *TC*, *CT*, *CC*. In questo caso per ragioni di simmetria sarebbe conveniente rappresentare i risultati nel piano *xy* per mezzo dei punti di coordinate (1,1); (1,0); (0,1); (0;0) come in figura:

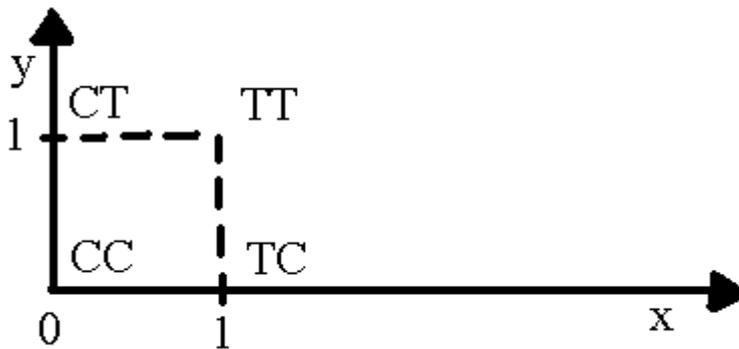


Fig.2a

oppure solo sull'asse *x*:

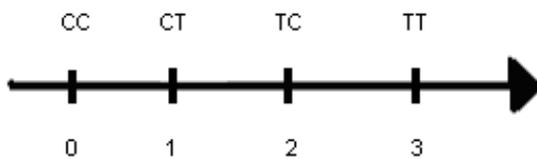


Fig.2b

Se la moneta fosse lanciata tre volte sarebbe conveniente usare lo spazio tridimensionale per rappresentare gli 8 possibili risultati sperimentali. È importante osservare che queste rappresentazioni sono semplicemente una convenienza, ad esempio nell'esperimento del lancio di due monete per rappresentare i 4 possibili risultati, se desiderato, si potrebbero pure rappresentare 4 punti qualsiasi sull'asse *x*. Nell'esperimento del getto di due dadi ci sono 36 possibili risultati indicati in tabella 1:

| | | | | | |
|-----------|-----------|-----------|-----------|-----------|-----------|
| 11 | 21 | 31 | 41 | 51 | 61 |
| 12 | 22 | 32 | 42 | 52 | 62 |
| 13 | 23 | 33 | 43 | 53 | 63 |
| 14 | 24 | 34 | 44 | 54 | 64 |
| 15 | 25 | 35 | 45 | 55 | 65 |
| 16 | 26 | 36 | 46 | 56 | 66 |

Tab.1

In questo esperimento un conveniente insieme di punti per rappresentare i possibili risultati sono i 36 punti nel piano xy , le cui coordinate sono le corrispondenti coppie di numeri di tabella 1, come in figura:

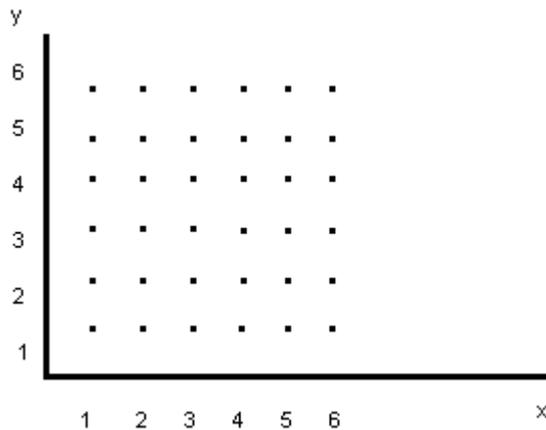


Fig.3

Un esperimento che consiste nella lettura della temperatura di un individuo presenta un numero molto grande di possibili risultati che dipendono dal grado di accuratezza con cui si può leggere il termometro. Per un tale esperimento è conveniente fare l'ipotesi che la temperatura dell'individuo possa assumere qualsiasi valore fra 35°C e 42°C , perciò i risultati possibili si potrebbero rappresentare convenientemente con i punti dell'intervallo di estremi 35 e 42 sull'asse x come in figura:

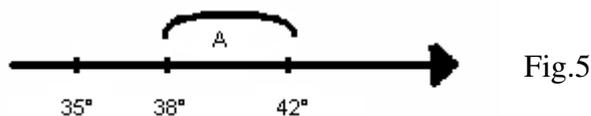


Fig.4

Ciò naturalmente non tiene conto della impossibilità di leggere un termometro con accuratezza illimitata. Possiamo dare allora la seguente definizione: *l'insieme dei punti che rappresentano i possibili risultati di un esperimento è chiamato lo spazio campione dell'esperimento.*

Com'è già stato detto è importante osservare che la rappresentazione di tutti i possibili risultati di un esperimento, cioè lo spazio campione, non è unica ma è dettata da criteri di semplicità e convenienza. Si fa osservare che lo spazio viene chiamato campione in quanto l'esperimento cui si riferisce è casuale, e ciò significa che il suo esito è incerto così che un dato risultato è appunto solo un campione dei molti esiti possibili. Il motivo per cui viene introdotto lo spazio campione di un esperimento è che esso è un conveniente strumento matematico per sviluppare la *Teoria della Probabilità* per quanto riguarda i risultati dell'esperimento.

Evento: Consideriamo un esperimento tale che qualunque sia il risultato dell'esperimento si possa decidere se un evento A si è verificato. Ciò significa che ciascun punto dello spazio campione si può classificare come un punto per cui l'evento A si verifica o come un punto per cui A non si verifica. Così se A è l'evento di ottenere esattamente una testa e una croce, a prescindere dall'ordine, nel lanciare una moneta due volte, i due punti campione CT e TC corrispondono al verificarsi dell'evento A . Se A è l'evento di ottenere un totale di 7 punti nel gettare due dadi allora A è associato ai 6 punti campione $(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)$. Se A è l'evento che la temperatura di un individuo sia almeno 38°C allora A sarà associato all'intervallo di punti fra 38° e 42° .



Possiamo dare allora la seguente definizione: *un evento è un sottoinsieme di uno spazio campione, poiché un sottoinsieme di un insieme di punti comprende la possibilità che il sottoinsieme coincida con l'intero insieme di punti o che esso non contenga nessun punto dell'insieme.*

Questa definizione comprende un evento che è certo di verificarsi o un evento che non può verificarsi quando l'evento viene eseguito. Considerata la corrispondenza fra gli eventi e gli insiemi di punti, lo studio della relazione fra i vari eventi si può ricondurre allo studio della relazione fra i corrispondenti insiemi. A tale scopo vengono comunemente usati i cosiddetti *diagrammi di Venn*, che sono un conveniente sistema di rappresentazione in cui lo spazio campione, qualunque sia la sua dimensione o numero di punti in esso contenuti, viene rappresentato per mezzo di un insieme di punti interni ad un rettangolo in un piano come nella figura seguente dove si indica lo spazio campione:



Fig.6

Un evento A , che è perciò un sottoinsieme di punti in questo rettangolo, è rappresentato dai punti contenuti all'interno di una curva chiusa contenuta nel rettangolo. Se B è qualche altro evento di interesse esso sarà rappresentato dai punti interni a qualche altra curva chiusa nel rettangolo. Questa rappresentazione è mostrata nella seguente figura:

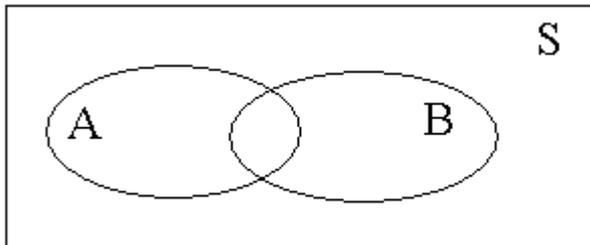


Fig.7

Se A e B sono due eventi associati ad un esperimento si può voler sapere se almeno uno di essi si verificherà quando l'esperimento verrà eseguito. L'insieme dei punti che consiste di tutti i punti che appartengono ad A o a B o sia ad A che a B è chiamato l'**unione** di A e B e si indica col simbolo $A \cup B$. Questo insieme di punti è rappresentato nella figura seguente dalla regione tratteggiata:

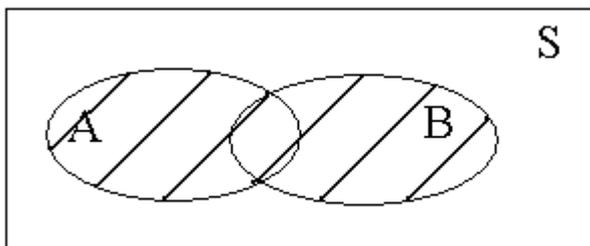


Fig.8

Come esempio, se A è l'evento di ottenere un 6 quando si gettano due dadi e B è l'evento di ottenere un 6 sul secondo dado, allora $A \cup B$ è l'evento di ottenere almeno un 6 quando si gettano due dadi. L'evento A consiste dei 6 punti dello spazio campione le cui coordinate sono indicate nell'ultima colonna della tab.1, e l'evento B consiste dei 6 punti indicati nell'ultima riga della tabella stessa.

Un altro evento di possibile interesse è di sapere se entrambi gli eventi si verificheranno quando l'esperimento viene eseguito. L'insieme dei punti che consiste di tutti i punti che appartengono sia ad A che a B viene chiamato **intersezione** di A e B e si indica con $A \cap B$.

Questo insieme di punti è rappresentato nella figura seguente dalla regione tratteggiata:

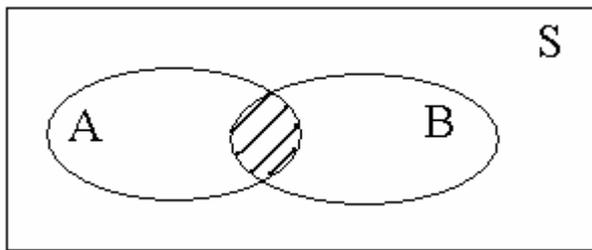


Fig.9

Nell'esempio del getto di due dadi, dove A è l'evento di ottenere un 6 sul primo dado e B è quello di ottenere un 6 sul secondo dado, $A \cap B$ è l'evento di ottenere un 6 su entrambi i dati. Esso è rappresentato dall'unico punto di coordinate (6,6) che è l'intersezione dei sottoinsiemi di punti dello spazio campione indicati nell'ultima colonna e nell'ultima riga della tabella 1. In corrispondenza a qualsiasi evento A esiste un evento ad esso associato, che si indica con \bar{A} , che stabilisce che l'evento A non si verificherà quando l'esperimento viene eseguito. Esso è rappresentato da tutti i punti dello spazio campione, cioè del rettangolo, che non si trovano in A e nella figura seguente è rappresentato dalla regione tratteggiata:

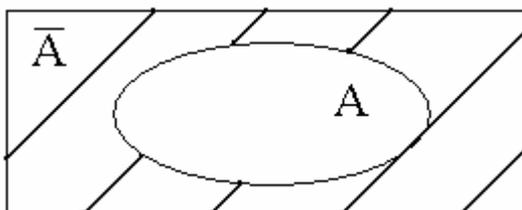


Fig.10

L'evento \bar{A} viene chiamato il **complementare di A** relativo allo spazio campione. Se due insiemi A e B non hanno punti in comune si dice che sono *insiemi disgiunti*. In termini di eventi tali eventi si chiamano *eventi disgiunti*, ma si chiamano *eventi reciprocamente esclusivi* perché il verificarsi dell'uno esclude la possibilità del verificarsi dell'altro.

Due eventi di questo tipo sono rappresentati nella figura seguente:

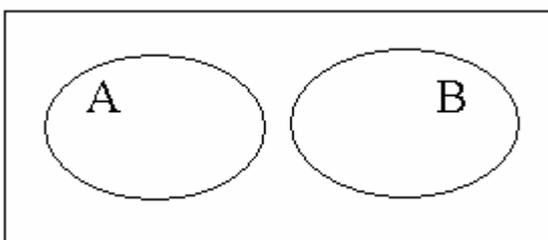


Fig.11

PROBABILITA': Fissiamo ora l'attenzione sulle funzioni di insieme perché ci servono per definire la *Probabilità*. Le funzioni che ci sono più familiari sono quelle di punto, ad esempio $f(x)=x^2$ è una funzione di questo tipo perché a ciascun punto dell'asse "x" questa formula associa il valore della funzione in quel punto. Però come è noto la nozione di funzione è molto più ampia di questa in quanto gli elementi del dominio della funzione anziché punti singoli possono essere insiemi di punti. In questo caso la funzione prende il nome di *funzione di insieme*.

Un esempio è la funzione il cui dominio è rappresentato da intervalli sull'asse "x" ed essa fornisce la lunghezza dell'intervallo.

La *probabilità* si può considerare come un modello per la frequenza del verificarsi di un certo risultato in ripetute esecuzioni dell'esperimento. Perciò un modello di probabilità per un evento A , dovrebbe essere un modello per prevedere la frequenza del verificarsi dell'evento A in ripetute esecuzioni dell'esperimento. Questa probabilità la indicheremo con $P\{A\}$. Poiché $P\{A\}$ è una funzione definita su insiemi essa è una *funzione di insieme*. Se un esperimento si potesse ripetere un gran numero di volte i risultati si potrebbero usare per assegnare un valore a $P\{A\}$. Tuttavia non è affatto necessario che l'esperimento venga eseguito prima che una tale probabilità venga assegnata, così nell'esperimento del getto di due dadi se A è l'evento di ottenere un 6 su entrambi i dadi. Considerazioni di simmetria suggerirebbero il valore di $1/36$ per la probabilità del verificarsi di questo evento.

E' importante osservare che ciascuno è libero di assegnare il valore che vuole, ma se l'assegnazione non è realistica il suo modello di probabilità gli sarà di scarsa utilità per fare previsione circa i futuri esperimenti.

Se la probabilità degli eventi sono da interpretare come modelli per le frequenze del verificarsi di quegli eventi in ripetute esecuzioni dell'esperimento, queste probabilità dovrebbero possedere le proprietà essenziali delle frequenze, così una probabilità dovrebbe essere un numero compreso fra 0 e 1 perché una frequenza è un numero di questo tipo; inoltre la probabilità dell'evento S , dove S rappresenta lo spazio campione, dovrebbe essere uguale a 1 perché qualcuno dei risultati possibili si verifica certamente quando l'esperimento viene eseguito. Infine se due eventi A e B sono *disgiunti*, la probabilità dell'unione di quegli eventi dovrebbe essere uguale alla somma delle probabilità dei singoli eventi perché la frequenza che A o B si verifichi è uguale alla somma delle loro frequenze. Si è trovato che ogni altra ragionevole proprietà delle

probabilità sarà soddisfatta se sono soddisfatte le seguenti tre condizioni che si basano su quanto appena detto: queste sono chiamate **i tre assiomi delle probabilità**. Essi impongono restrizioni sul tipo di funzione dell'insieme P che si può usare per calcolare la probabilità degli eventi. Una tale funzione viene chiamata **misura delle probabilità**.

Diamo ora gli assiomi della probabilità. *Una misura di probabilità P è una funzione di insieme a valori reali definita su uno spazio campione S che soddisfa :*

$$1) 0 \leq P\{A\} \leq 1$$

$$2) P\{S\}=1$$

3) $P\{A_1 \cup A_2 \cup \dots\} = P\{A_1\} + P\{A_2\} + \dots$ per ogni successione finita o infinita di eventi disgiunti A_1, A_2, \dots

Si fa osservare che gli assiomi della probabilità sono stati formulati nel 1933 da A. N. **Kolmogorov**. Sarebbe molto difficile trovare una funzione P che fornisca i valori delle probabilità corrispondenti alle frequenze attese per ogni possibile sottoinsieme A di uno spazio campione perché il numero di tali sottoinsiemi è estremamente grande anche per spazi campione contenenti soltanto pochi punti. Fortunatamente per spazi campione contenenti soltanto un numero finito di punti campione o una loro successione infinita, è sufficiente assegnare il valore della probabilità a ciascuno dei punti campione. Il valore di $P\{A\}$ per qualunque sottoinsieme A si può calcolare allora facilmente dalle probabilità assegnate a singoli punti campione per mezzo del terzo assioma delle probabilità.

Regola di addizione delle probabilità: il calcolo delle probabilità interessa spesso un certo numero di eventi piuttosto che un evento soltanto. Per semplicità consideriamo due eventi A_1 e A_2 associati ad un esperimento. Spesso è interessante sapere se entrambi gli eventi si verificheranno o si verificherà almeno uno di essi. Per quanto si è già detto il primo di questi eventi composti è $A_1 \cap A_2$ e il secondo è $A_1 \cup A_2$. Per rispondere alla domanda riguardante la frequenza del verificarsi di questi due eventi occorre calcolare i valori delle loro probabilità, cioè $P\{A_1 \cap A_2\}$ e $P\{A_1 \cup A_2\}$. Ricaviamo ora una formula per il calcolo di $P\{A_1 \cup A_2\}$. Lo spazio campione di un esperimento sia rappresentato dai punti nel

rettangolo della figura seguente e i punti campione corrispondenti al verificarsi degli eventi A_1, A_2 siano i punti interni alle regioni definite con A_1, A_2 :

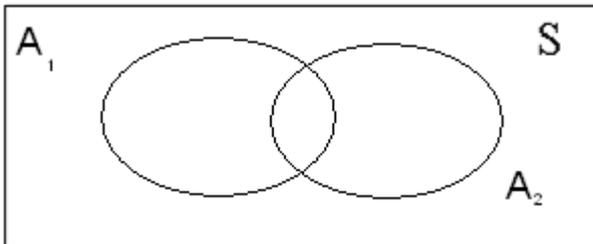


Fig.12

Come si è già detto l'evento $A_1 \cup A_2$ consiste di tutti i punti all'interno delle due regioni; la determinazione di una formula per $P\{A_1 \cup A_2\}$ si basa sull'esprimere $A_1 \cup A_2$ come l'unione di eventi disgiunti per poi applicare il terzo assioma delle probabilità. Dalla figura si può osservare che $A_1 \cup A_2$ è l'unione dei tre insiemi disgiunti $A_1 \cap \bar{A}_2$, $A_2 \cap \bar{A}_1$ e $A_1 \cap A_2$. Perciò per il terzo assioma si ha che :

$$(1) P\{A_1 \cup A_2\} = P\{A_1 \cap \bar{A}_2\} + P\{A_2 \cap \bar{A}_1\} + P\{A_1 \cap A_2\}.$$

Dalla figura si può osservare anche che A_1 è l'unione degli eventi disgiunti $A_1 \cap \bar{A}_2$ e $A_1 \cap A_2$; quindi per il terzo assioma si ha che:

$$P\{A_1\} = P\{A_1 \cap \bar{A}_2\} + P\{A_1 \cap A_2\}.$$

$$\text{Analogamente } P\{A_2\} = P\{A_2 \cap \bar{A}_1\} + P\{A_1 \cap A_2\}.$$

Ricavando da queste ultime due relazioni rispettivamente $P\{A_1 \cap \bar{A}_2\}$ e $P\{A_2 \cap \bar{A}_1\}$ e sostituendo le loro espressioni a secondo membro della (1) si ottiene la formula richiesta nota come **regola di addizione delle probabilità** o **regola delle probabilità totali**, e cioè:

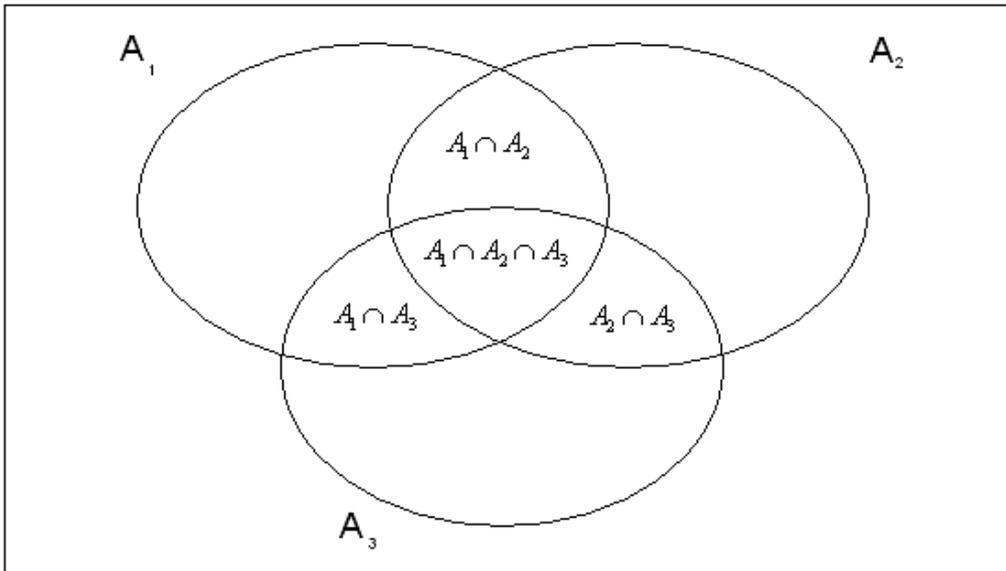
$$(2) P\{A_1 \cup A_2\} = P\{A_1\} + P\{A_2\} - P\{A_1 \cap A_2\}.$$

Due eventi A_1 e A_2 possono non avere punti campione in comune, come detto in precedenza gli eventi si dicono allora disgiunti. La formula (2) si riduce allora alla seguente:

$$(3) P\{A_1 \cup A_2\} = P\{A_1\} + P\{A_2\}.$$

Ma questa formula è pure una diretta conseguenza del terzo assioma. La regola (2) si può generalizzare al caso di più eventi.

Ad esempio nel caso di tre eventi arbitrari A_1, A_2, A_3 si ha che:
 $P\{A_1 \cup A_2 \cup A_3\} = P\{A_1\} + P\{A_2\} + P\{A_3\} - P\{A_1 \cap A_2\} - P\{A_1 \cap A_3\} - P\{A_2 \cap A_3\} + P\{A_1 \cap A_2 \cap A_3\}$.



- $A_1 \cap A_2$ (contato 2 volte)
- $A_1 \cap A_3$ (contato 2 volte)
- $A_2 \cap A_3$ (contato 2 volte)
- $A_1 \cap A_2 \cap A_3$ (contato 3 volte)

Fig.13

La regola di addizione è applicabile a qualunque tipo di spazio campione, per il quale è assegnata una misura P della probabilità. Per usarla bisogna naturalmente conoscere i valori delle probabilità a secondo membro delle formule; cosa particolarmente semplice quando lo spazio campione contiene soltanto un numero finito di punti. Perciò restringeremo ora la discussione a spazi campione di questo tipo. Quindi la prima cosa da fare per calcolare la probabilità di A per una spazio campione finito è di assegnare il valore della probabilità a ciascun punto campione che naturalmente deve obbedire ai primi due assiomi delle probabilità, cioè questi valori devono essere non negativi e la loro somma deve essere uguale a 1. Queste assegnazioni si possono fare basandosi su l'esperienza del singolo ricercatore, su informazioni esterne, su considerazioni di simmetria etc.

Così, ad esempio, sarebbe realistico per simmetria assegnare la probabilità di $1/36$ a ciascun punto dello spazio campione considerato per l'esperimento del getto dei dadi.

Calcolo della probabilità di un evento: indichiamo ora con n il numero totale dei punti campione e siano p_1, p_2, \dots, p_n le probabilità assegnate ai rispettivi punti campione. Ciascun punto rappresenta un possibile risultato che a sua volta è un evento. Eventi di questo tipo sono spesso chiamati *eventi semplici*, questi eventi li indicheremo con e_1, e_2, \dots, e_n . E' chiaro che essi sono *disgiunti*. Qualsiasi evento A è un insieme di punti campione, perciò è l'unione degli eventi semplici corrispondenti. L'applicazione del terzo assioma della probabilità perciò fornisce che :

$$(4) P\{A\} = \sum_A p_i$$

dove la sommatoria viene fatta su quelle probabilità p_i associate ai punti campione che giacciono in A .

Per molti giochi d'azzardo lo spazio campione è non soltanto finito, ma ai punti campione viene assegnata la stessa probabilità; questo è vero ad esempio per il getto di due dadi per cui lo spazio campione risulta conveniente. A ciascuno di quei punti campione viene assegnata infatti la probabilità di $1/36$. Se n indica il numero totale di punti campione e $N(A)$ il numero di punti campione nell'insieme A allora poiché si è assunto che $p_i = 1/n$, $i = 1, \dots, n$, allora la formula (4) si riduce a:

$$(5) P\{A\} = \frac{N(A)}{n} .$$

Probabilità condizionata: supponiamo ora di voler sapere se un evento A_2 si verificherà sotto la condizione che l'evento A_1 sia certo di verificarsi. Per discutere le probabilità associate ad eventi come questi assumiamo che lo spazio campione contenga soltanto un numero finito di punti e consideriamo la situazione mostrata nella figura seguente:

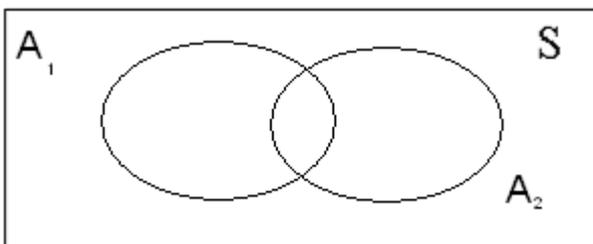


Fig.14

Poiché A_1 è certo di verificarsi soltanto quando lo spazio campione è ristretto a quei punti che giacciono entro la regione A_1 è necessario considerare come si dovrebbero assegnare le probabilità ai punti di questo

nuovo spazio campione più piccolo. Se originariamente ad un punto campione in A_1 fosse stata assegnata ad esempio una probabilità doppia rispetto ad un altro punto in A_1 , allora si dovrebbe assegnare una probabilità doppia anche nel nuovo spazio campione ristretto ai punti di A_1 , perché ignorare i risultati sperimentali che non producono l'evento A_1 non deve alterare il rapporto delle frequenze dei due punti in esame; è semplicemente necessario perciò moltiplicare le probabilità originarie assegnate ai punti in A_1 per un fattore costante c tale che la somma delle nuove probabilità sia uguale a 1, così se π_i indica la nuova probabilità corrispondente a p_i nell'assegnazione originaria si dovrebbe scegliere:

$$\pi_i = c \cdot p_i, \text{ dove } \sum_{A_1} c \cdot p_i = \sum_{A_1} \pi_i = c \cdot P\{A_1\} = 1.$$

Come risultato si ottiene che $c = \frac{1}{P\{A_1\}}$ e quindi $\pi_i = \frac{p_i}{P\{A_1\}}$

Ora che le probabilità del nuovo spazio campione sono state assegnate si possono calcolare le probabilità nella solita maniera, semplicemente applicando la formula (4). Tutte queste probabilità saranno probabilità condizionate, subordinate cioè al verificarsi dell'evento A_1 . Se la probabilità che l'evento A_2 si verifichi, subordinato dalla condizione che

debba verificarsi l'evento A_1 , si indica con $P\{A_2|A_1\} = \sum_{A_1 \cap A_2} \pi_i =$

$$= \frac{\sum_{A_1 \cap A_2} p_i}{P\{A_1\}}, \text{ si fa osservare che la prima somma viene fatta su quelle } \pi_i \text{ che}$$

corrispondono ai punti campione di $A_1 \cap A_2$ perché essi sono i soli punti campione dentro A_1 che corrispondono al verificarsi di A_2 . Poiché la somma al numeratore nell'ultima espressione è quella che definisce la $P\{A_1 \cap A_2\}$ allora segue che la formula per il calcolo della probabilità condizionata si riduce alla:

$$(6): P\{A_2|A_1\} = \frac{P\{A_1 \cap A_2\}}{P\{A_1\}}.$$

Si assume qui che A_1 sia un evento $P\{A_1\} > 0$.

Per ottenere la formula (6) vista si è assunto che lo spazio campione contenga soltanto un numero finito di punti, ma questa formula si può

prendere come definizione di probabilità condizionata anche per spazi campione più generali. Si mostra facilmente che $P\{A_2|A_1\}$ soddisfa i tre assiomi della probabilità, perciò è legittimo definire la probabilità condizionata in questo modo a prescindere dal tipo di spazio campione. La suddetta formula (6), scritta sotto forma di prodotto fornisce la **regola fondamentale di moltiplicazione delle probabilità** (o *regola delle probabilità composte*), e cioè:

$$(1) P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2|A_1\}.$$

Se si scambia l'ordine di due eventi la formula (1) diventa:

$$(2) P\{A_1 \cap A_2\} = P\{A_2\} \cdot P\{A_1|A_2\}$$

Eventi indipendenti: ora supponiamo che A_1, A_2 siano due eventi tali che $P\{A_2|A_1\} = P\{A_2\}$ e che $P\{A_1\} \cdot P\{A_2\} > 0$. Allora l'evento A_2 si dice *indipendente nel senso della probabilità*, o più brevemente **indipendente dall'evento** A_1 . Ciò deriva dalla proprietà che la probabilità del verificarsi dell'evento A_2 non viene alterata dalla condizione che debba verificarsi l'evento A_1 . Quando A_2 è indipendente da A_1 la regola di moltiplicazione delle probabilità espressa dalla (1) si riduce alla :

$$(3) P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2\}.$$

Viceversa quando è vera la (3), segue, dal confronto di questa con la (1), che A_2 è indipendente da A_1 . Se si eguagliano i secondi membri della (3) e della (2) si può osservare che $P\{A_1|A_2\} = P\{A_1\}$, ma ciò che stabilisce l'evento A_1 è indipendente dall'evento A_2 . Così se A_2 è indipendente da A_1 segue che A_1 deve essere indipendente da A_2 . A causa di questa reciproca indipendenza e poiché la (3) implica questa indipendenza, si suole definire l'indipendenza come segue: "*due eventi si dicono **indipendenti** se $P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2\}$* ".

Esercitazioni: come applicazione pratica delle formule fondamentali per il calcolo delle probabilità, di cui ci siamo occupati, affrontiamo ora la risoluzione di alcuni problemi.

ESI: nell'ipotesi che una famiglia abbia due figli si vuole calcolare la probabilità che entrambi i figli siano maschi sapendo che almeno uno di essi è maschio. Assumiamo che lo spazio campione sia dato da $S = \{(M,M),$

, (M,F) , (F,M) , (F,F)). Ove ad esempio (M,F) sta ad indicare che il figlio maggiore è una femmina e il minore un maschio, e che tutti gli elementi dello spazio siano ugualmente probabili. Indichiamo ora con A_1 l'evento che entrambi i figli siano maschi e A_2 che almeno uno sia maschio. Si ha allora che la probabilità di $P\{A_1|A_2\}$ è:

$$P\{A_1|A_2\} = \frac{P\{A_1 \cap A_2\}}{P\{A_2\}} = \frac{P\{(M,M)\}}{P\{(M,M);(M,F);(F,M)\}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} = 33.3333\%$$

ES2: Supponiamo che uno studente debba sostenere due esami e che la probabilità di superare il primo sia 0,6 e quella del secondo sia 0,8; inoltre la probabilità che superi entrambi sia 0,5. Si vuole calcolare la probabilità che superi almeno uno degli esami e la probabilità che sia bocciato in entrambi. A tale scopo indichiamo con A_1 l'evento corrispondente alla promozione al primo esame e con A_2 corrispondente al secondo esame. Allora $A_1 \cap A_2$ è l'evento che lo studente superi entrambi gli esami e $A_1 \cup A_2$ è quello che superi almeno un esame. Si ha quindi che: $P\{A_1 \cup A_2\} = P\{A_1\} + P\{A_2\} - P\{A_1 \cap A_2\} = 0,6 + 0,8 - 0,5 = 0,9$. L'evento che lo studente fallisca in ambedue gli esami è dato da $\overline{(A_1 \cup A_2)}$; si ha quindi che $P\{\overline{(A_1 \cup A_2)}\} = 1 - P\{A_1 \cup A_2\} = 1 - 0,9 = 0,1$. A questo punto si vuole calcolare inoltre la probabilità che lo studente superi il secondo esame supposto che abbia superato il primo, e analogamente la probabilità che superi il primo esame supposto che abbia superato il secondo. Allora la prima probabilità è data da:

$$P\{A_2|A_1\} = \frac{P\{A_1 \cap A_2\}}{P\{A_1\}} = \frac{0,5}{0,6} = 0,833.$$

$$P\{A_1|A_2\} = \frac{P\{A_1 \cap A_2\}}{P\{A_2\}} = \frac{0,5}{0,8} = 0,625.$$

E' importante osservare che la promozione degli esami può aiutare la prestazione dello studente a superare l'altro in quanto ne aumenta la fiducia. Ci si chiede ora se gli eventi A_1 , A_2 siano indipendenti. Tenuto conto che $P\{A_1 \cap A_2\} = 0,5$ e che $P\{A_1\} \cdot P\{A_2\} = 0,6 \cdot 0,8 = 0,48$.

La condizione di indipendenza non è soddisfatta dato che $P\{A_1 \cap A_2\} \neq P\{A_1\} \cdot P\{A_2\}$.

In effetti, dal momento che $P\{A_1 \cap A_2\} > P\{A_1\} \cdot P\{A_2\}$ si ricava che, come già detto, la promozione in uno degli esami può aiutare le prestazioni dello studente a superare il secondo.

ES3: Supponiamo che un'urna contenga 7 palline nere e 5 bianche e si estraggano due palline senza la sostituzione delle palline estratte. Si vuole calcolare la probabilità che ambedue le palline estratte siano nere nell'ipotesi che nelle urne ogni pallina abbia la stessa probabilità di essere estratta. Indichiamo con A_1 l'evento che la prima pallina sia nera e con A_2 che sia nera la seconda pallina. Nell'ipotesi che la pallina estratta sia nera, nell'urna restano 6 palline nere e quindi $P\{A_2|A_1\} = 6/11$.

Siccome $P\{A_1\} = 7/12$ si ha che:

$$P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2|A_1\} = 7/12 \cdot 6/11 = 42/132 = 0,318 = 31,8 \%$$

ES4: Supponiamo approssimativamente che il 50 % degli individui di età superiore ai 40 anni sia in sovrappeso e che la percentuale degli individui in sovrappeso e avente una malattia all'apparato circolatorio sia il 25 %.

Si vuole calcolare la probabilità che un individuo scelto a caso con più di 40 anni e in sovrappeso (evento A_1) abbia una malattia all'apparato circolatorio (evento A_2). Questa probabilità è data da :

$$P\{A_2|A_1\} = \frac{P\{A_1 \cap A_2\}}{P\{A_1\}} = \frac{1/4}{1/2} = 50 \%$$

ES5: Supponiamo di lanciare due dadi simmetrici, sia A_1 l'evento di ottenere una somma dei punti uguali a 6 e C sia l'evento che sul primo dado esca 4. Allora $A_1 \cap C$ è rappresentato dall'unico risultato (4,2); si ha quindi che $P\{A_1 \cap C\} = 1/36$ ed inoltre $P\{A_1\} \cdot P\{C\} = 5/36 \cdot 1/6 = 5/216 = 0,027$, dove $P\{A_1\} = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$.

Tenuto conto che $P\{A_1 \cap C\} \neq P\{A_1\} \cdot P\{C\}$, i due eventi non sono indipendenti. In effetti se ad esempio il primo dado dà un 6, l'evento A_1 diventa impossibile. Se invece indichiamo con A_2 la somma dei punti uguali a 7, allora $A_2 \cap C$ è rappresentato dall'unico risultato (4,3) e quindi :

$P\{A_2 \cap C\} = 1/36$ e $P\{A_2\} \cdot P\{C\} = 6/36 \cdot 1/6 = 1/36$, dove $P\{A_2\} = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$. In questo caso $P\{A_2 \cap C\} = P\{A_2\} \cdot P\{C\}$ e i due eventi sono indipendenti.

Esercitazione: risolviamo ora il problema seguente che oltre ad essere un'applicazione delle formule fin qui esaminate ci introduce inoltre al calcolo delle probabilità delle cause del verificarsi di un evento.

Una scatola contiene due palline rosse ed una seconda scatola identica contiene una pallina rossa e una bianca. Se si sceglie una scatola a caso e si estrae da essa una pallina si vuole calcolare la probabilità di aver scelto la prima scatola se la pallina estratta risulta essere rossa. Indichiamo con A_1 l'evento di scegliere la prima scatola e con \bar{A}_1 quello di scegliere la seconda. Indichiamo poi con A_2 l'evento di estrarre una pallina rossa e con \bar{A}_2 quello di estrarre una pallina bianca. Allora il problema è di calcolare la probabilità condizionata $P\{A_1|A_2\}$. Ora per quanto si è detto si può scrivere

$$P\{A_1|A_2\} = \frac{P\{A_1 \cap A_2\}}{P\{A_2\}}.$$

Per la regola di moltiplicazione delle probabilità, espressa dalla formula (1) vista prima, intendendo che scegliere una scatola a caso significa che la probabilità di scegliere ad esempio la prima scatola è $1/2$, si ha quindi che $P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2|A_1\} = 1/2 \cdot 1 = 1/2$. La probabilità al denominatore si può calcolare osservando che l'evento A_2 si verificherà se e soltanto se si verificherà l'uno o l'altro dei due eventi reciprocamente esclusivi, e cioè $A_1 \cap A_2$, e $\bar{A}_1 \cap A_2$.

Per cui si può scrivere che $P\{A_2\} = P\{A_1 \cap A_2\} + P\{\bar{A}_1 \cap A_2\}$.

Ora si ha che $P\{\bar{A}_1 \cap A_2\} = P\{\bar{A}_1\} \cdot P\{A_2|\bar{A}_1\} = 1/2 \cdot 1/2 = 1/4$.

Quindi si ha che $P\{A_2\} = 1/2 + 1/4 = 3/4$.

Per cui infine $P\{A_1|A_2\} = \frac{1/2}{3/4} = 2/3$.

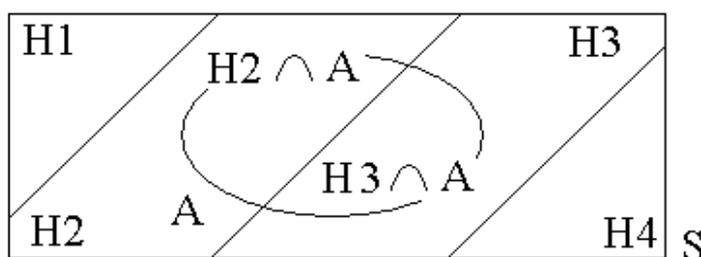
Questo esempio è stato presentato come applicazione pratica delle formule fondamentali per il calcolo della probabilità. Tuttavia esso si sarebbe potuto risolvere anche facendo ricorso allo spazio campione dell'esperimento e utilizzando una formula già vista, e precisamente

$$P\{A\} = \sum_A P_i \text{ con } i=1, \dots, n.$$

Per il problema in esame sarebbe conveniente considerare come spazio campione quello costituito da quattro punti I1, I2, II1, II2 dove il numero romano indica il numero della scatola e l'altro il numero della pallina.

A questi 4 punti si devono assegnare ovviamente le stesse probabilità pari ad $\frac{1}{4}$. La condizione del verificarsi dell'evento A_2 restringe lo spazio campione ai primi 3 punti soltanto nell'ipotesi che la pallina 2 nella seconda scatola sia la pallina bianca; così a ciascuno di questi 3 punti occorre assegnare la probabilità di $\frac{1}{3}$. Ora però soltanto i primi due punti campione corrispondono al verificarsi dell'evento A_1 , cioè alla scelta della prima scatola, quindi $P\{A_1|A_2\} = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$ per la regola della somma.

Formula di Bayes: L'esempio considerato è tipico di problemi in cui si considera il risultato di un esperimento e poi ci si chiede qual è la probabilità che esso sia dovuto ad una fra le possibili cause del suo verificarsi. Così nell'esempio verificato sono due le possibili cause dell'estrazione di una pallina rossa e il problema è di determinare la probabilità che essa sia dovuta soltanto alla prima causa, cioè alla scelta della prima scatola. Per quanto la soluzione del problema sia stata ottenuta semplicemente applicando le regole della probabilità nella successione appropriata i calcoli tuttavia sono sufficientemente estesila valere la pena di ricavare una formula per trattare questi problemi sistematicamente. A tale scopo supponiamo che H_1, H_2, \dots, H_n siano eventi *mutuamente esclusivi* con $P\{H_i\}=p_i$, con $p_i > 0$ per $i = 1, \dots, n$, tali che l'unione $\bigcup_{i=1}^n H_i = S$, per $i = 1, \dots, n$, dove S è lo spazio campione. Si dice anche che gli eventi H_i costituiscono una partizione dello spazio campione S . Essi rappresentano le n -possibili cause di un risultato sperimentale. Sia poi A un evento con $P\{A\}>0$ che si verifica quando viene eseguito l'esperimento. Il problema è quello di calcolare la probabilità che H_i sia la causa del verificarsi dell'evento A . Una possibile suddivisione dello spazio campione e la sua relazione con l'evento A è mostrato in figura dove lo spazio campione S per comodità è stato suddiviso in quattro sottoinsiemi:



Si tratta quindi di calcolare la probabilità condizionata $P\{H_i|A\}$. La regola per il calcolo della probabilità condizionata ci dà che:

$$(1) \quad P\{H_i|A\} = \frac{P\{H_i \cap A\}}{P\{A\}}.$$

La regola di moltiplicazione delle probabilità fornisce:

$$(2) \quad P\{H_i \cap A\} = P\{H_i\} \cdot P\{A|H_i\},$$

per cui si può scrivere:

$$(3) \quad P\{H_i|A\} = \frac{P\{H_i\} \cdot P\{A|H_i\}}{P\{A\}}.$$

Dalla figura è chiaro che A è l'unione degli eventi disgiunti $H_1 \cap A, H_2 \cap A, \dots, H_n \cap A$ e che alcuni degli insiemi corrispondenti a questi eventi possono naturalmente, come accade anche in figura, essere vuoti. Per il terzo assioma delle probabilità si può quindi scrivere che

$P\{A\} = \sum_{j=1}^n P\{H_j \cap A\}$, per $j=1, \dots, n$. Sostituendo a secondo membro di quest'ultima le probabilità come espresse dalla (2) si ottiene che:

$$(4) \quad P\{A\} = \sum_{j=1}^n P\{H_j\} \cdot P\{A|H_j\}, \text{ per } j=1, \dots, n. \text{ A questo punto}$$

sostituendo nella (3) l'espressione ricavata per $P\{A\}$ si ottiene la formula richiesta per calcolare la probabilità delle cause, nota come **formula di Bayes**, ed espressa quindi dalla :

$$(5) \quad P\{H_i|A\} = \frac{P\{H_i\} \cdot P\{A|H_i\}}{\sum_{j=1}^n P\{H_j\} \cdot P\{A|H_j\}}, \text{ per } j=1, \dots, n.$$

Ambedue le formule (4) e (5) sono utili nel descrivere gli esperimenti che procedono in due fasi e hanno le proprietà che il meccanismo dell'alienità della seconda fase è determinato dal risultato della prima fase dell'esperimento. Questi esperimenti sono chiamati anche *esperimenti composti*. Nell'applicazione delle formule precedenti agli esperimenti composti H_i , rappresentano i risultati della prima fase dell'esperimento e $P\{A|H_i\}$ rappresenta il meccanismo di alienità della seconda fase sotto l'ipotesi che l' H_i si sia verificato nella prima fase. Le probabilità che $P\{H_i\}$ sono chiamate anche **probabilità a priori** mentre le probabilità condizionate $P\{H_i|A\}$ sono chiamate **probabilità a posteriori**. Queste denominazioni derivano dal fatto che *in certi esperimenti $P\{H_i\}$ sono probabilità soggettive che rappresentano la nostra esperienza prima*

dell'esperimento, mentre le probabilità $P\{H_i|A\}$ possono essere interpretate come la descrizione della nostra opinione dopo che sono stati effettuati alcuni esperimenti e si è verificato l'evento A . In altre parole la formula di Bayes può essere considerata anche come un algoritmo per cambiare la nostra opinione sulla base di risultati sperimentali. Questo aspetto verrà illustrato nel seguito, esaminando un esempio specifico.

Esercitazione: consideriamo ora alcuni esempi come applicazione pratica delle formule (4) e (5).

Esempio 1: per illustrare l'applicazione diretta della formula di Bayes risolviamo ora il problema considerato in precedenza per introdurre il calcolo della probabilità delle cause. Indichiamo ora con H_1, H_2 gli eventi di scegliere rispettivamente la prima e la seconda scatola e con A l'evento di estrarre una pallina rossa. Poiché una scatola viene scelta a caso $P\{H_1\} = P\{H_2\} = 1/2$. Inoltre è chiaro dal contenuto delle due scatole che $P\{A|H_1\} = 1$ e $P\{A|H_2\} = 1/2$. In questo caso l'applicazione diretta della formula di Bayes fornisce:

$$P\{H_1|A\} = \frac{P\{H_1\} \cdot P\{A|H_1\}}{P\{H_1\} \cdot P\{A|H_1\} + P\{H_2\} \cdot P\{A|H_2\}} = \frac{1/2 \cdot 1}{(1/2 \cdot 1) + (1/2 \cdot 1/2)} = \frac{1/2}{3/4} = 2/3.$$

Es2: in una diagnosi medica si osserva che un paziente ha uno o più sintomi specifici: $A = \{S_1, S_2, \dots, S_l\}$; e si pone il problema di decidere quale delle possibili $\{H_1, H_2, \dots, H_k\}$ è la causa più probabile dei sintomi osservati. Si suppone di avere una stima statistica delle probabilità $P\{H_i\} = p_i$, $p_i > 0$, per $i = 1, \dots, k$ di contrarre la malattia H_i . Si suppone inoltre che le altre malattie non siano contemporaneamente presenti nello stesso paziente. Si assume inoltre di conoscere una stima statistica della probabilità condizionata $P\{A|H_i\}$, ovvero della probabilità che la malattia H_i dia origine ai sintomi A .

Applicando la formula di Bayes la probabilità che i sintomi A siano dovuti alla malattia H_i per $i = 1, \dots, k$, o in altre parole la probabilità che un paziente con uno o più sintomi A abbia la malattia H_i , è data dalla formula (5) ovvero:

$$(5): P\{H_i|A\} = \frac{P\{H_i\} \cdot P\{A|H_i\}}{\sum_{j=1}^k P\{H_j\} \cdot P\{A|H_j\}}, \text{ per } j=1, \dots, k.$$

Come esempio supponiamo che:

$$P\{H_1\} = 0.40; P\{H_2\} = 0.25; P\{H_3\} = 0.35;$$

$$P\{A|H_1\} = 0.8; P\{A|H_2\} = 0.6; P\{A|H_3\} = 0.9;$$

Applicando la formula (4) si ha che:

$$P\{A\} = P\{H_1\} \cdot P\{A|H_1\} + P\{H_2\} \cdot P\{A|H_2\} + P\{H_3\} \cdot P\{A|H_3\} = \\ = 0.40 \cdot 0.80 + 0.25 \cdot 0.60 + 0.35 \cdot 0.90 = 0.785.$$

Ora applicando la formula (5) si ha che:

$$P\{H_1|A\} = \frac{P\{H_1\} \cdot P\{A|H_1\}}{P\{A\}} = \frac{0.40 \cdot 0.80}{0.785} = 0.4076;$$

$$P\{H_2|A\} = \frac{P\{H_2\} \cdot P\{A|H_2\}}{P\{A\}} = \frac{0.25 \cdot 0.60}{0.785} = 0.1911;$$

$$P\{H_3|A\} = \frac{P\{H_3\} \cdot P\{A|H_3\}}{P\{A\}} = \frac{0.35 \cdot 0.90}{0.785} = 0.4013;$$

Si conclude che un paziente che presenta uno o più sintomi A ha una maggiore possibilità di avere contratto la malattia H_1 e in assenza di ulteriori informazioni dovrebbe essere curato per tale malattia.

Es3: supponiamo che il ministero delle finanze prima di adottare una politica di controllo dei prezzi e dei salari chieda il parere di tre esperti indicati con $\{H_1, H_2, H_3\}$ sull'impatto che tale politica può avere sul tasso di disoccupazione. Le risposte date in termini di probabilità dei singoli esperti sono raccolte nella tabella seguente.

Probabilità di cambiamento nel tasso di disoccupazione:

| Esperto | Diminuzioni | Stabilità | Aumenti |
|----------------|--------------------|------------------|----------------|
| H_1 | 0.1 | 0.1 | 0.8 |
| H_2 | 0.6 | 0.2 | 0.2 |
| H_3 | 0.2 | 0.6 | 0.2 |

Supponiamo inoltre che in base alle esperienze passate il Ministero si è fatto l'opinione che le probabilità che un esperto abbia una teoria corretta dell'economia sono date da:

$$P\{H_1\} = 1/6; P\{H_2\} = 1/3; P\{H_3\} = 1/2$$

Supponendo che a seguito della politica adottata si verifichi un aumento del tasso di disoccupazione, indicare come dovrebbero essere modificate le opinioni del ministro sulla correttezza della teoria economica di ogni singolo esperto. Per risolvere questo problema si tratta di calcolare:

$P\{H_1|A\}$; $P\{H_2|A\}$; $P\{H_3|A\}$. Usando la formula (4) calcoliamo dapprima la probabilità dell'aumento del tasso di disoccupazione. Si ha che:

$$P\{A\} = P\{H_1\} \cdot P\{A|H_1\} + P\{H_2\} \cdot P\{A|H_2\} + P\{H_3\} \cdot P\{A|H_3\} = \\ = 1/6 \cdot (0.8) + 1/3 \cdot (0.2) + 1/2 \cdot (0.2) = 0.3.$$

Applicando ora la formula di Bayes si ha che:

$$P\{H_1|A\} = \frac{P\{H_1\} \cdot P\{A|H_1\}}{P\{A\}} = \frac{\frac{1}{6} \cdot 0.8}{0.3} = 0.\bar{4}$$

$$P\{H_2|A\} = \frac{P\{H_2\} \cdot P\{A|H_2\}}{P\{A\}} = \frac{\frac{1}{3} \cdot 0.2}{0.3} = 0.\bar{2}$$

$$P\{H_3|A\} = \frac{P\{H_3\} \cdot P\{A|H_3\}}{P\{A\}} = \frac{\frac{1}{2} \cdot 0.2}{0.3} = 0.\bar{3}$$

Si conclude che la teoria dell'esperto H_1 , la meno corretta secondo il Ministro prima che la politica venisse adottata, appare nel seguito la più corretta.

Variabili casuali o aleatorie: L'aggettivo aleatorio deriva dal latino "*alea*" che significa gioco dei dadi o rischio e, quindi, aleatorio significa che dipende dalla sorte o dal caso.

Consideriamo ora lo spazio campione mostrato in precedenza corrispondente all'esperimento del lancio di due monete e fissiamo l'attenzione sul numero di teste ottenute. Per calcolare la probabilità di possibili risultati è conveniente introdurre una variabile X per rappresentare il numero di teste ottenute. Allora X assumerà il valore 0 nel punto campione CC ; 1 nei punti TC , CT e 2 nel punto TT . Una variabile come questa a valori numerici è un esempio di *variabile casuale*.

Un secondo esempio riguarda l'esperimento del getto di due dadi: se X rappresenta la somma dei punti ottenuti, allora essa è una variabile casuale, che può assumere i valori interi da 2 a 12.

Come altro esempio se X rappresenta la distanza dal centro di una freccia scagliata su di un bersaglio circolare di raggio 20 allora assumendo di trascurare tutti i colpi mancanti essa è una variabile casuale che può assumere qualsiasi valore da 0 a 20.

In tutti questi esempi la variabile X è calcolata numericamente ed il suo valore dipende dal punto campione. Così in effetti X è una funzione il cui dominio di definizione è l'insieme dei punti campione e il cui insieme di valori, ovvero il suo condominio, è un insieme di numeri reali.

Si può dare la seguente definizione: *una **variabile casuale** X è una funzione a valori reali definita su di uno spazio campione.*

Il motivo per cui la variabile viene chiamata casuale è che *essa è definita su di uno spazio campione associato ad un esperimento fisico il cui risultato è incerto, ovvero si dice che dipende dal caso o è incerto il valore della variabile.*

L'obbiettivo che ora si propone è quello di studiare le variabili casuali e di calcolare le probabilità ad esse associate. A tale scopo è necessario assegnare una misura della probabilità allo spazio campione associato alla variabile casuale X . Perciò si assume, senza ricordarlo espressamente, che a qualsiasi spazio campione sia assegnata una misura delle probabilità.

Variabili casuali discrete: Dopo che una variabile casuale X è stata definita su di uno spazio campione, l'interesse si concentra di solito nel calcolare la probabilità che X assuma valori specifici nel suo insieme di valori possibili. Per esempio se X rappresenta la somma dei punti nel getto dei due dadi, allora può interessare il calcolo della probabilità che X assuma ad esempio il valore 7. Oppure se X rappresenta la distanza di una freccia dal centro di un bersaglio circolare con raggio 20cm, può interessare calcolare la probabilità che X assuma ad esempio un valore minore di 5. Il calcolo di $P\{X=7\}$ nel primo esempio è molto più semplice del calcolo di $P\{X<5\}$ del secondo perché è molto più semplice lavorare con uno spazio campione che consiste di un numero finito di punti, che con uno che consiste di un intervallo di punti sull'asse x . Uno spazio campione che consiste di un numero finito o di una successione infinita di punti è chiamato uno *spazio campione discreto* mentre uno che consiste di uno o più intervalli di punti viene chiamato *spazio campione continuo*.

Per intervalli si intendono naturalmente intervalli di dimensione qualsiasi. Poiché la teoria per gli spazi campione discreti è molto più semplice di quella per gli spazi continui limiteremo per ora la discussione agli spazi

campione discreti. E' chiaro che una variabile casuale non può assumere più valori di quanti sono i punti campione, perciò una variabile definita su di uno spazio campione discreto può assumere soltanto un numero finito o una successione infinita di valori. Una variabile casuale come questa viene chiamata una **variabile casuale discreta**.

Per calcolare le probabilità delle variabili costanti discrete ricordiamo la formula per calcolare un evento A , cioè: $P\{A\} = \sum_A P_i$ ricavata

assumendo che lo spazio campione fosse discreto e quindi applicabile a qualsiasi spazio campione discreto. Se x è uno specifico valore della variabile casuale discreta X , questa formula si può usare per calcolare la probabilità che X assuma il valore x .

Questa probabilità è perciò data da (1):

$$(1) P\{X = x\} = \sum_{X=x} P_i$$

dove la \sum viene fatta su tutti i punti campione in cui la variabile casuale X ha valore x . Come esempio se X rappresenta la somma dei punti ottenuti nel getto di due dadi, tenendo conto che tutti i punti campione hanno la probabilità di $1/36$, allora $P\{X=7\} = 6/36$.

Questo risultato deriva dal fatto che vi sono 6 punti campione il cui valore di $X = 7$ e più precisamente $(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)$.

Funzione di Probabilità o Funzione di densità di Probabilità: Per calcolare le probabilità delle variabili casuali discrete è conveniente introdurre una funzione chiamata *funzione di probabilità* o *funzione di densità di probabilità* o più brevemente **densità**. Si può dare la seguente definizione:

Definizione: sia X una variabile casuale discreta, allora la funzione f definita da $f(x) := P\{X = x\}$ viene chiamata la funzione di probabilità o la funzione di densità di probabilità della variabile X se soddisfa le seguenti proprietà:

$$1) f(x) \geq 0$$

$$2) \sum_x f(x) = 1$$

Si fa notare che in genere si parla di densità riferendosi ad esempio ad una distribuzione contigua di massa, ovvero di materia lungo una retta o un piano e che perciò il suo uso a proposito di una distribuzione di probabilità discreta può sembrare improprio. Tuttavia, poiché è desiderabile usare la stessa terminologia per questo tipo di funzione, sia per le variabili discrete

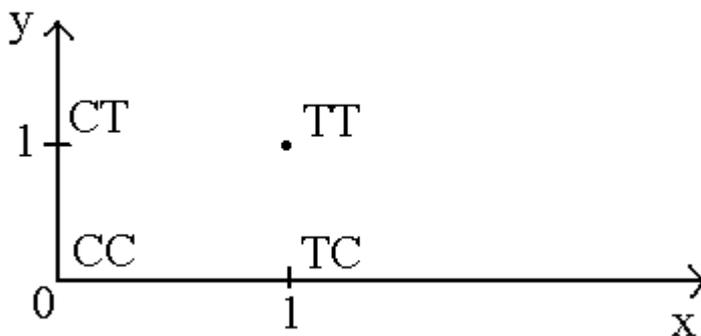
che per quelle continue, e poiché il termine densità è appropriato per le variabili continue, esso verrà usato anche nel caso discreto.

Una funzione di densità spesso consiste semplicemente di una tabella di valori, così nel lancio di due monete se la variabile X rappresenta in numero la somma delle teste ottenute, la funzione di densità discreta si può definire per mezzo del seguente insieme di valori:

$$f(0) = P\{X = 0\} = \sum_{X=0} p_i = 1/4$$

$$f(1) = P\{X = 1\} = \sum_{X=1} p_i = 1/4 + 1/4 = 1/2$$

$$f(2) = P\{X = 2\} = \sum_{X=2} p_i = 1/4$$



Si fa notare che la funzione $f(x)$ è uguale a zero se $x \neq 0, 1, 2$.

Per giudicare come si distribuisce una variabile casuale discreta, cioè come la sua probabilità varia al variare del valore della variabile, è utile tracciare il grafico della funzione densità per mezzo di un grafico a segmenti. Come esempio di tale grafico la variabile X rappresenta la somma dei punti ottenuti lanciando due dadi. Attribuendo l'esatto valore di X a ciascuno dei 36 punti dello spazio campione corrispondente al lancio di due dadi e assumendo uguali le probabilità per i tutti i punti campione mediante l'impiego della formula (1) si ottiene:

$$f(2) = f(12) = 1/36;$$

$$f(3) = f(11) = 2/36;$$

$$f(4) = f(10) = 3/36;$$

$$f(5) = f(9) = 4/36;$$

$$f(6) = f(8) = 5/36;$$

$$f(7) = 6/36;$$

Il grafico a segmenti dell' $f(x)$, con i valori qui indicati, è mostrato nella figura proposta di seguito:

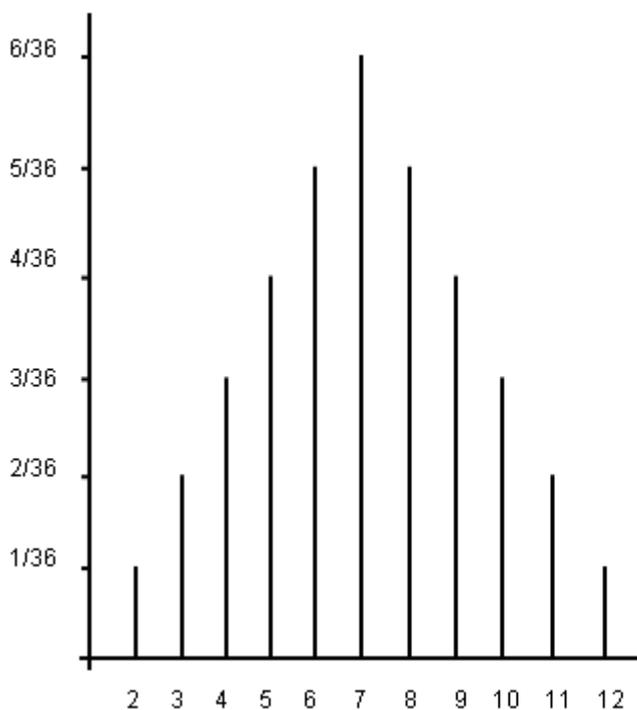
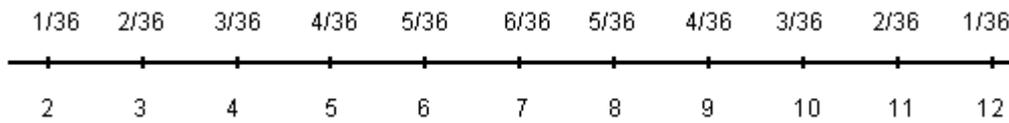


Fig.1

Lo scopo di introdurre una funzione di densità è quello di semplificare la notazione e il calcolo delle probabilità delle variabili casuali discrete. Dopo che una funzione di densità è stata determinata non è più necessario calcolare le probabilità degli eventi riguardanti una variabile casuale sommando le probabilità dei punti campione come nella (1). Si possono invece considerare i punti dell'asse x dove $f(x) > 0$ come i punti di un nuovo spazio campione discreto con le probabilità date dai valori dell' $f(x)$ calcolata in quei punti. Per esempio se X rappresenta la somma dei punti nel lancio di due dadi il nuovo spazio campione consisterà degli undici punti 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 e le probabilità associate a quei punti saranno i valori dell' $f(x)$ calcolati in precedenza.

Questo nuovo spazio campione è mostrato in figura:



Questo spazio campione è semplicemente una rappresentazione dell'informazione trasmessa poco prima dalla figura 1.

Ora consideriamo il calcolo delle probabilità dell'evento $X \in R$, ove R è un dato insieme di punti sull'asse x , e $x \in R$ rappresenta l'evento che X assuma valori di R . Considerando il nuovo spazio campione generato da X e da $f(x)$ una diretta applicazione di una formula già vista, cioè $\sum_A P_i$, fornisce il risultato:

(1) $P\{X \in R\} = \sum_{x \in R} f(x)$ dove la \sum viene fatta su quei valori x in R per i quali $f(x) > 0$.

Il calcolo di questa probabilità in questo modo è di solito molto più semplice del calcolo che si basa sullo spazio campione originario dell'esperimento. Usando nuovamente l'esempio del lancio di due dadi supponiamo di voler calcolare la probabilità che la somma dei punti sia maggiore di 7. Considerando il nuovo spazio campione nella figura precedente questa probabilità è data da :

$$P\{X > 7\} = \sum_{x=8}^{12} f(x) = f(8) + f(9) + f(10) + f(11) + f(12) = 5/36 + 4/36 + 3/36 + 2/36 + 1/36 = 15/36.$$

Se questa probabilità si dovesse calcolare usando lo spazio campione originario sarebbe necessario sommare la probabilità dei 15 punti campione le cui coordinate hanno come somma un valore maggiore di 7,

$$\text{cioè } P\{X > 7\} = \sum_{X > 7} P_i = 1/36 + \dots + 1/36 = 15/36.$$

In questo caso in effetti non c'è un vantaggio particolare ad usare il nuovo spazio campione, però per spazi campione più complicati il vantaggio può essere considerevole.

Funzione di DISTRIBUZIONE, o di distribuzione cumulativa, o di ripartizione: Una funzione strettamente affine alla funzione di densità f è la corrispondente funzione di distribuzione F .

Essa è definita come segue:

$F(x) := P\{X \leq x\} = \sum_{t \leq x} f(t)$ dove la \sum viene fatta su tutti i valori della

variabile casuale discreta che sono minori o uguali di x . Per costruire il grafico della $F(x)$ corrispondente alla $f(x)$ occorre calcolare il valore assunto dalla $F(x)$ negli undici punti 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. Si ha quindi che:

$$F(2) = P\{X \leq 2\} = \sum_{t=2} f(t) = f(2) = 1/36;$$

$$F(3) = P\{X \leq 3\} = \sum_{t=3} f(t) = f(2) + f(3) = 1/36 + 2/36 = 1/12;$$

·
·
·

$$F(12) = P\{X \leq 12\} = \sum_{t=12} f(t) = f(2) + \dots + f(12) = 1.$$

Funzione di probabilità congiunta, o di densità di probabilità congiunta: Molti esperimenti implicano molte variabili casuali anziché soltanto una. Per semplicità consideriamo due variabili casuali discrete X , Y . Un modello matematico per queste due variabili è una funzione che dà la probabilità che X assuma uno specifico valore x e contemporaneamente Y assuma uno specifico valore y . Si può dare allora la seguente *definizione*: Siano X e Y due variabili casuali discrete allora la funzione f definita da $f(x, y) := P\{X = x; Y = y\}$ viene chiamata la *funzione di probabilità congiunta* o la *funzione di densità di probabilità congiunta* delle variabili X , Y se soddisfa le seguenti proprietà:

$$f(x, y) \geq 0; \sum_x \sum_y f(x, y) = 1;$$

L'aggettivo congiunto spesso si omette perché non è possibile confondere una funzione di due variabili con una funzione di una sola variabile.

Esempio: consideriamo un esempio in cui la variabile X rappresenta il numero di carte di picche ottenute nell'estrarre una prima carta da un mazzo e la variabile Y rappresenta il numero di carte di picche ottenute nell'estrarre una seconda carta dal mazzo senza che la prima sia stata reinserita. In questo caso le variabili X , Y possono assumere soltanto i

valori 0 e 1. Allora ricordando la regola di moltiplicazione delle probabilità, cioè $P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2|A_1\}$, si ha che:

$$f(0,0) = P\{X = 0; Y = 0\} = P\{X = 0\} \cdot P\{Y = 0|X = 0\} = 39/52 \cdot 38/51 = 0,56;$$

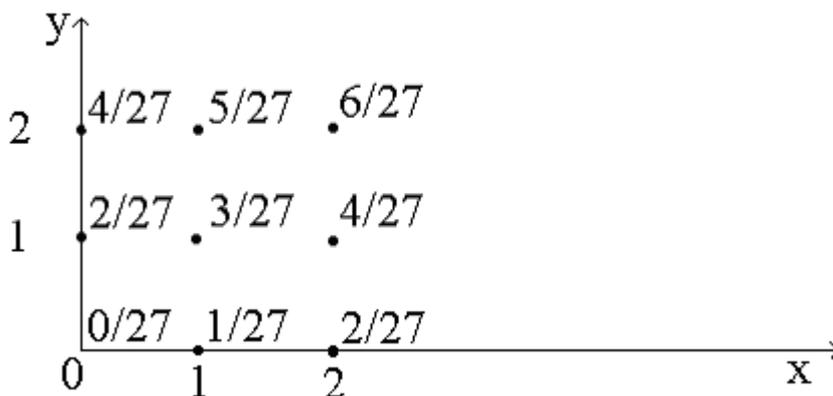
$$f(0,1) = P\{X = 0; Y = 1\} = P\{X = 0\} \cdot P\{Y = 1|X = 0\} = 39/52 \cdot 13/51 = 0,19;$$

$$f(1,0) = P\{X = 1; Y = 0\} = P\{X = 1\} \cdot P\{Y = 0|X = 1\} = 13/52 \cdot 39/51 = 0,19;$$

$$f(1,1) = P\{X = 1; Y = 1\} = P\{X = 1\} \cdot P\{Y = 1|X = 1\} = 13/52 \cdot 12/51 = 0,06.$$

Una funzione di densità congiunta si può usare per calcolare le probabilità delle variabili casuali discrete X, Y sommando la $f(x, y)$ sopra opportuni valori delle due variabili, proprio come le probabilità della variabile casuale discreta X si calcolano sommando la $f(x)$ sopra opportuni valori della variabile X . Come nel problema monodimensionale è conveniente adoperare in un nuovo spazio campione rispetto a quello originario che consiste dei punti del piano x, y dove $f(x, y) > 0$ e le cui probabilità sono i valori forniti dalla $f(x, y)$.

Esempio: consideriamo come esempio quello in cui la funzione di densità di due variabili casuali discrete X, Y è data da $f(x, y) = \frac{1}{27} (x + 2y)$ dove x, y possono assumere tutti i valori interi, tali che $0 \leq x \leq 2$; $0 \leq y \leq 2$ e $f(x, y) = 0$ altrove e vogliamo calcolare $P\{X \geq 1; Y \leq 1\}$. Lo spazio campione con le probabilità calcolate tramite la formula suddetta è mostrata nella figura :



Si può quindi calcolare la probabilità richiesta, che è data da:

$$P\{x \geq 1, y \leq 1\} = \sum_{x \geq 1} \sum_{y \leq 1} f(x, y) = \sum_{x=1}^2 \sum_{y=0}^1 f(x, y) = \sum_{x=1}^2 [f(x, 0) + f(x, 1)] =$$

$$= f(1, 0) + f(1, 1) + f(2, 0) + f(2, 1) = 1/27 + 3/27 + 2/27 + 4/27 = 10/27 =$$

$$= 0.37 = 37 \%$$

Variabili indipendenti: Due variabili che sono senza rapporti in senso probabilistico sono chiamate variabili indipendenti. Poiché l'indipendenza di due eventi A_1, A_2 è stata definita da $P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2\}$, una definizione corrispondente per le variabili casuali dovrebbe essere conforme a quella definizione. Considerando eventi come $X \in A, Y \in B$ dove A e B sono due insiemi qualsiasi nei domini di definizione delle variabili X, Y rispettivamente, la definizione di indipendenza per eventi richiederebbe che :

$$(1) P\{X \in A, Y \in B\} = P\{X \in A\} \cdot P\{Y \in B\}, \text{ per ogni } A, B.$$

Questa viene presa spesso come definizione di indipendenza di due variabili. Tuttavia poiché ora stiamo ponendo l'accento sulle funzioni di densità cercheremo di dare una funzione equivalente basata sulle funzioni di densità. A tale scopo sia $f(x, y)$ la funzione di densità delle due variabili e $g(x), h(y)$ siano le funzioni di densità delle singole variabili. Allora scegliendo gli insiemi A, B come costituiti rispettivamente dai singoli punti x, y , la (1) si riduce alla $P\{X = x, Y = y\} = P\{X = x\} \cdot P\{Y = y\}$ per ogni x, y .

Ma in termini di funzione di densità questa si scrive:

$$f(x, y) = g(x) \cdot h(y) \text{ per ogni } x, y.$$

Viceversa se $f(x, y) = g(x) \cdot h(y)$ per tutti gli x, y allora operando nello spazio campione delle due variabili casuali segue per analogia con la (1)

$$(P(x \in R) = \sum_{x \in R} f(x)) \text{ che:}$$

$$\begin{aligned} P\{X \in A, Y \in B\} &= \sum_{x \in A} \sum_{y \in B} f(x, y) = \sum_{x \in A} \sum_{y \in B} g(x) \cdot h(y) = \sum_{x \in A} g(x) \cdot \sum_{y \in B} h(y) = \\ &= P\{X \in A\} \cdot P\{Y \in B\}. \end{aligned}$$

Questa mostra che la (1) vale per tutti gli A, B se e soltanto se:

$$f(x, y) = g(x) \cdot h(y), \text{ per ogni } x, y.$$

Poiché l'ultima relazione è espressa in termini di funzioni di densità ed è più utile della (1) essa viene usata per definire l'indipendenza, si può dunque dare:

Definizione: le variabili casuali X, Y , la cui funzione di densità congiunta è $f(x, y)$ e le cui funzioni di densità individuali $g(x), h(y)$, sono indipendenti se e soltanto se :

$$(2) f(x, y) = g(x) h(y), \text{ per ogni } x, y.$$

La precedente definizione si può generalizzare in una nuova maniera per definire l'indipendenza di n variabili casuali, così:

Definizione: le variabili casuali X_1, X_2, \dots, X_n , la cui funzione di densità congiunta è $f(x_1, x_2, \dots, x_n)$ e le cui funzioni di densità singole sono $f_1(x_1), f_2(x_2), \dots, f_n(x_n)$, sono indipendenti se e soltanto se:

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_n(x_n) \text{ per ogni } x_1, x_2, \dots, x_n.$$

Distribuzioni marginali e distribuzioni condizionate: poiché è importante sapere se un insieme di variabili è costituito da variabili indipendenti, sarebbe auspicabile disporre di un metodo sistematico per ricavare le funzioni di densità delle singole variabili dalla loro determinata funzione congiunta. Un tale metodo si ottiene senza difficoltà per il caso di due variabili servendosi della regola di moltiplicazione delle probabilità. Sebbene il metodo si possa facilmente estendere a più di due variabili limiteremo per ora la trattazione a due sole variabili casuali discrete. A tale scopo consideriamo un esperimento in cui A_1 rappresenta l'evento che una variabile casuale X assuma il valore x e A_2 l'evento che una seconda variabile casuale Y assuma il valore y .

La regola di moltiplicazione (3) $P\{A_1 \cap A_2\} = P\{A_1\} P\{A_2|A_1\}$ assume allora la forma che deriva dall'esprimere le probabilità degli eventi in termini di funzione. Poiché ora $P\{A_1 \cap A_2\}$ dà la probabilità che le due variabili casuali assumano rispettivamente i valori x, y , essa rappresenta il valore della funzione di densità congiunta nel punto (x, y) cioè $f(x, y)$.

$P\{A_1\}$ in modo simile è la probabilità che la variabile X assuma il valore x perciò essa è $f(x)$. Poiché $P\{A_2|A_1\}$ dà la probabilità condizionata che Y assuma il valore y quando X ha il fissato valore x , essa si può trattare come il valore di una funzione di densità condizionata che si indica con $f(y|x)$, allora la (3) diventa: (4) $f(x, y) = f(x) f(y|x)$.

Poiché $f(y|x)$ dà la probabilità condizionata che Y assuma il valore y quando X ha per valore fissato il valore x , la somma di $f(y|x)$ su tutti i possibili valori di y deve essere uguale a 1. Quindi se entrambi i membri della (4) si sommano su tutti i possibili valori di Y si ottiene:

$$\sum_y f(x, y) = f(x) \cdot \underbrace{\sum_y f(y|x)}_1 = f(x), \text{ con } \sum_y f(y|x) = 1.$$

Quindi:

$$(5) f(x) = \sum_y f(x, y) .$$

La funzione $f(x)$ viene chiamata la funzione **di densità marginale della variabile** X , comunque essa è semplicemente la funzione di densità di X .

In modo analogo la funzione di densità della variabile Y , $g(y)$, si può ottenere sommando la $f(x, y)$ su tutti i possibili valori della variabile X per un fissato valore y , si ha che $g(y) = \sum_x f(x, y) .$

Gli ultimi risultati mostrano che se si ha la funzione di densità congiunta di due variabili casuali discrete e se si desidera la funzione di densità di una di esse, è semplicemente necessario sommare la funzione di densità congiunta su tutti i valori dell'altra variabile. La funzione di densità condizionata $f(y | x)$ dà la distribuzione di probabilità della variabile Y quando viene mantenuto fisso il valore di X .

Per la formula (4), cioè $f(x, y) = f(x) f(y | x)$, se $f(x) > 0$ si può scrivere:

$$(6) f(y | x) = \frac{f(x, y)}{f(x)} .$$

La distribuzione condizionata della variabile X , quando si ha mantenuto fisso il valore della variabile Y , è data dalla formula analoga:

$$f(x | y) = \frac{f(x, y)}{g(y)} \text{ con } g(y) > 0 .$$

Ciò mostra che se si conosce la funzione di densità congiunta di due variabili e si desidera la funzione di densità condizionata di una di esse, quando il valore dell'altra viene mantenuto fisso, è semplicemente necessario dividere la funzione di densità congiunta per la funzione di densità della variabile il cui valore viene tenuto fisso.

Ricordando le formule qui di seguito, tratteremo due esempi in seguito:

$$(4) f(x, y) = f(x) \cdot f(y | x)$$

$$(5) f(x) = \sum_y f(x, y)$$

$$(6) f(y | x) = \frac{f(x, y)}{f(x)}$$

Esercitazione: per illustrare i concetti precedenti supponiamo che un'urna contenga due palline bianche e quattro nere e che le due palline vengano estratte dall'urna. Le variabili X e Y rappresentano i valori delle due estrazioni, 0 corrisponde ad una pallina nera e 1 ad una pallina bianca. Allora ogni possibile risultato sarà rappresentato da uno dei quattro punti del piano (x,y) di fig. 1 di coordinate $(0,0)$; $(0, 1)$; $(1,0)$; $(1,1)$. Dai contenuti dell'urna e dall'ordine in cui le estrazioni sono fatte, segue direttamente dalla (4) che:

$$f(0, 0) = f(0) f(0|0) = 4/6 \cdot 3/5 = 6/15;$$

$$f(0,1) = f(0) f(0|1) = 2/6 \cdot 4/5 = 4/15;$$

$$f(1, 0) = f(1) f(1|0) = 4/6 \cdot 2/5 = 4/15;$$

$$f(1, 1) = f(1) f(1|1) = 2/6 \cdot 1/5 = 1/15;$$

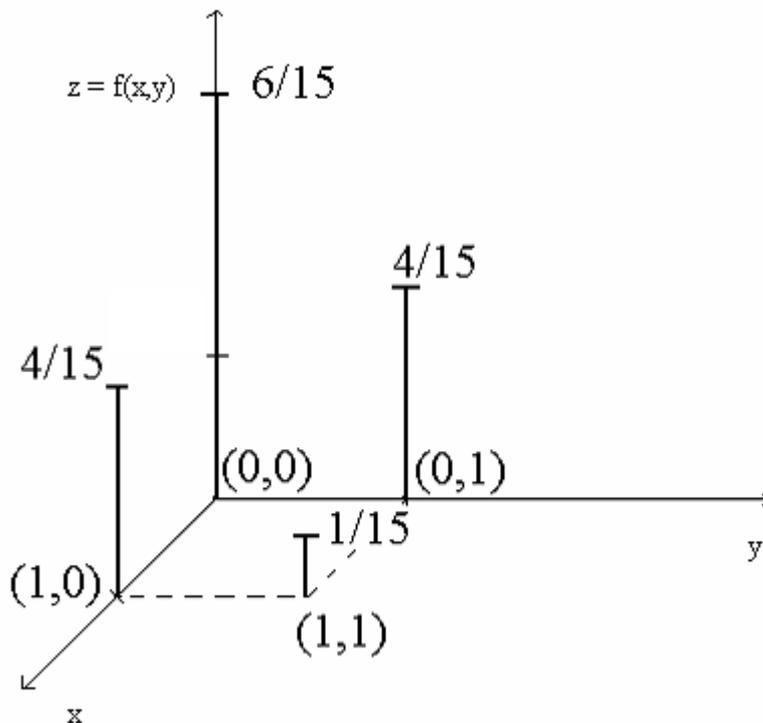


Fig.1

I valori della $f(x, y)$ sono riportati nel grafico a segmenti di fig. 1. Per illustrare come ottenere una distribuzione marginale e una distribuzione condizionata dalla distribuzione congiunta assumiamo ora che siano noti solo i valori della $f(x, y)$ appena calcolati. Così la sola

informazione disponibile è quella data in figura 1 indipendentemente da come sono stati ottenuti quei numeri.

La funzione di densità marginale della variabile X si può ottenere

applicando la (5), cioè $f(x) = \sum_{y=0}^1 f(x, y)$.

Si ha:
$$f(0) = \sum_{y=0}^1 f(0, y) = f(0, 0) + f(0, 1) = 6/15 + 4/15 = 2/3$$

$$f(1) = \sum_{y=0}^1 f(1, y) = f(1, 0) + f(1, 1) = 4/15 + 1/15 = 1/3$$

Se i quattro punti del piano xy si pensano come punti di massa della probabilità la cui massa totale è 1, allora la distribuzione marginale della variabile X rappresenta la distribuzione della massa di probabilità lungo l'asse x dopo che i punti di massa di probabilità del piano xy sono stati proiettati perpendicolarmente sull'asse x . La funzione di densità condizionata di Y per un fissato valore di x si può ottenere applicando la formula (6) ed usando i risultati appena ottenuti. Così se alla variabile X è assegnato il valore $x=1$ si ha:

$$(6) f(y|1) = \frac{f(1, y)}{f(1)} \text{ e quindi } f(0|1) = \frac{f(1,0)}{f(1)} = \frac{4}{15} \cdot 3 = 4/5 \text{ e } f(1|1) = \frac{f(1,1)}{f(1)} = \frac{1}{15} \cdot 3 = 1/5.$$

Geometricamente, $f(y|1)$ rappresenta la distribuzione della massa di probabilità lungo la retta $x = 1$ quando i due punti su questa retta hanno avuto la loro massa di probabilità moltiplicata per un numero, cioè $\frac{1}{f(1)}$, tale da rendere la somma delle loro masse uguale a 1.

Esercitazione: come secondo esempio, in cui la funzione di densità congiunta è data direttamente, consideriamo la funzione di densità definita

come $f(x, y) = \frac{1}{27} (x + 2y)$ dove x e y possono assumere solo i valori interi $x = y = 0, 1, 2$. Lo spazio campione con le sue probabilità calcolate tramite questa formula, è mostrato in figura 1 vista prima.

Dalla formula (5) la funzione di densità marginale della variabile X è data

dalla:

$$f(x) = \sum_{y=0}^2 \frac{1}{27}(x+2y) = \frac{1}{27}(x+x+2+x+4) = \frac{1}{9}(x+2)$$

In modo simile :

$$g(y) = \sum_{x=0}^2 \frac{1}{27}(x+2y) = \frac{1}{27}(2y+1+2y+2+2y) = \frac{1}{9}(2y+1)$$

E' chiaro che in questo caso $f(x, y)$ non è uguale al prodotto delle due funzioni di densità marginale e perciò X e Y non sono variabili casuali indipendenti. Dalla formula (6) e dal risultato ottenuto per la $f(x)$ segue che la funzione di densità condizionata di Y per una fissato valore di x è data dalla :

$$f(y|x) = \frac{\frac{1}{27}(x+2y)}{\frac{1}{9}(x+2)} = \frac{x+2y}{3(x+2)} .$$

Se ad x si assegna per esempio il valore $x = 2$ si ha che:

$$f(y|2) = \frac{2+2y}{12} = \frac{1}{6}(y+1) .$$

Questa funzione sarebbe utile se si volessero calcolare le probabilità per più valori della Y quando si sa che $x = 2$; si può facilmente verificare che questa è una funzione di densità di probabilità, infatti sommando i tre valori di questa funzione corrispondente a $y=0,1,2$ si ottiene come risultato 1, infatti:

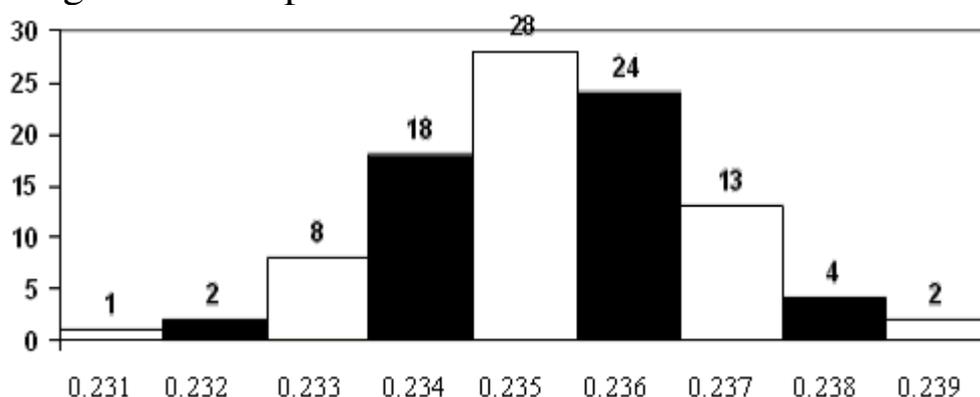
$$f(0|2) + f(1|2) + f(2|2) = \frac{1}{6}(0+1) + \frac{1}{6}(1+1) + \frac{1}{6}(2+1) = 1/6 + 2/6 + 3/6 = 6/6 = 1 .$$

Variabili casuali continue: una variabile casuale continua è una variabile definita su di uno spazio campione continuo. Variabili come il peso, la temperatura, la velocità, etc. sono variabili *continue*. Si potrebbe discutere sul fatto che in realtà tutte queste variabili sono discrete perché qualunque strumento di misura venga usato, esso presenta sempre una limitata accuratezza di lettura; in ogni caso è conveniente dal punto di vista matematico fare l'ipotesi che tale limitazione non esista.

Funzione di densità di probabilità: per calcolare le probabilità di una variabile casuale continua viene introdotta la *funzione di densità di probabilità* o *funzione di densità* o più brevemente ***densità della variabile***. La sua definizione formale si potrebbe dare immediatamente, ma cercheremo di motivarla esaminando ciò che accade ad una funzione di densità discreta, quando lo spazio campione diventa più denso. Per esaminare le proprietà delle variabili continue consideriamone una indicata con X rappresentante lo spessore di un disco metallico prodotto da una certa macchina. Se la macchina producesse ad esempio 100 dischi, i cui spessori fossero misurati con 3 cifre decimali, avremmo a disposizione 100 valori della variabile X con cui studiare il comportamento della macchina. Se questi valori venissero raccolti e rappresentati in forma di tabella, si potrebbe trovare un insieme di valori come quelli illustrati nella tabella seguente che fornisce la frequenza assoluta con cui si sono presentati i valori della variabile X :

| | | | | | | | | | |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| X | 0,231 | 0,232 | 0,233 | 0,234 | 0,235 | 0,236 | 0,237 | 0,238 | 0,239 |
| <i>Frequenza</i> | 1 | 2 | 8 | 18 | 28 | 24 | 13 | 4 | 2 |

Si fa osservare che il termine ***frequenza*** indica di solito il rapporto fra il numero di volte che si è presentato uno specifico valore della variabile X e il numero totale di valori osservati, però essa viene anche usata per indicare il numeratore di questo rapporto. Qualora nel seguito vi fossero dubbi sul significato del termine “frequenza” faremo uso rispettivamente dei termini “*frequenza relativa*” e “*frequenza assoluta*”. Per rappresentare graficamente i risultati della tabella 1 si usa un grafico chiamato istogramma che possiamo costruire:

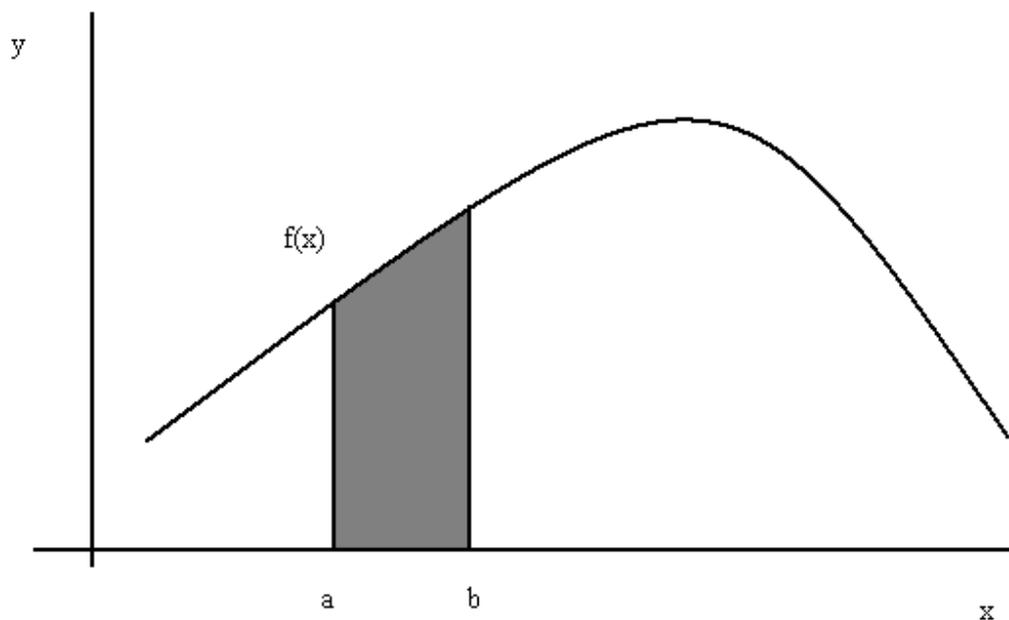


Un istogramma è un grafico come quello illustrato in fig. 1 in cui si usano delle aree per rappresentare le frequenze osservate, in particolare le frequenze relative, così l’area del rettangolo centrato ad esempio in 0,234

dovrebbe essere uguale alla frequenza relativa 0,18; tuttavia spesso in pratica si usa scegliere una conveniente unità di misura sull'asse y con il risultato che le aree dei rettangoli possono essere soltanto proporzionali anziché uguali alle corrispondenti frequenze relative. L'istogramma di fig. 1 è stato costruito con una siffatta conveniente scelta di unità di misura, quindi come si può osservare le aree sono soltanto proporzionali alle frequenze relative. Se l'istogramma si deve costruire in modo che le aree siano uguali alle frequenze relative, allora l'area totale dell'istogramma deve essere uguale a 1 perché la somma delle frequenze relative deve essere uguale a 1. Se h indica la distanza fra valori di x consecutivi l'altezza del rettangolo centrato in x_i deve essere $\frac{f_i}{N \cdot h}$, dove f_i è la frequenza assoluta di x_i ed N è il numero totale di osservazioni. Questo risultato è ovvio quando si pensa che questa ordinata moltiplicata per la base h deve essere uguale alla frequenza relativa $\frac{f_i}{N}$.

L'istogramma di fig. 1 indica la frequenza con cui si ottengono i vari valori della variabile X per 100 esecuzioni dell'esperimento; se si fossero fatte 200 esecuzioni l'istogramma risultante sarebbe stato grande il doppio di quello basato su 100 esecuzioni. Per confrontare istogrammi basati su numeri di esecuzioni dell'esperimento diversi fra loro è necessario scegliere unità di misura sull'asse y in modo tale che l'area dell'istogramma sia sempre uguale a 1. Con questa scelta di unità di misura ci si dovrebbe aspettare che l'istogramma tendesse ad un'istogramma fisso, cioè che non cambia più all'aumentare del numero di esecuzioni dell'esperimento. Inoltre se si fa l'ipotesi che X si possa misurare tanto accuratamente quanto si vuole, cioè che l'unità di misura h sull'asse x possa farsi tanto piccola quanto si vuole, allora ci si dovrebbe aspettare che l'istogramma si smussasse e pendesse ad una curva continua quando il numero di esecuzioni dell'esperimento aumenta infinitamente ed h si sceglie sempre più piccola, cioè tende a zero. Quando l'area dell'istogramma è uguale a 1, segue che la somma delle aree di rettangoli contigui fra loro è uguale alla frequenza relativa con cui il valore di X cade nell'intervallo costituito dalle basi di quei rettangoli. Poiché questa proprietà continuerà a valere all'aumentare del numero di esecuzioni dell'esperimento, l'area sottesa dalla curva fra due qualsiasi valori di X dovrebbe essere uguale alla frequenza relativa con la quale ci si aspetterebbe che il valore osservato di X cadesse dall'intervallo

determinato da quei valori. La funzione $f(x)$, il cui grafico è concepito come forma limite dell'istogramma, viene presa come modello matematico per la variabile casuale continua X e viene chiamata la “*funzione di densità di probabilità*” della variabile.



Poiché la frequenza relativa, nel caso di un' istogramma, è sostituita dalla probabilità nel caso di un modello matematico, la definizione di funzione di densità di probabilità di una variabile casuale continua si può stabilire nella formula seguente:

Definizione: una funzione di densità di probabilità di una variabile casuale continua X è una funzione f che possiede le seguenti proprietà:

$$1) f(x) \geq 0$$

$$2) \int_{-\infty}^{+\infty} f(x) dx = 1 = P\{-\infty < x < +\infty\}$$

$$3) \int_a^b f(x) dx = P\{a < X < b\}, \text{ dove } a \text{ e } b \text{ sono due valori qualsiasi}$$

di X con $a < b$.

La prima proprietà è ovviamente necessaria perché la probabilità deve essere non negativa; la seconda proprietà corrisponde a richiedere che la probabilità di un evento che è certo di verificarsi deve essere uguale a 1, infatti certamente X assumerà un qualche valore reale quando viene fatta una sua osservazione. Nel calcolare poi le probabilità di una variabile casuale continua si richiede soltanto che la variabile giaccia in qualche

intervallo. Come risultato si ha che le probabilità delle variabili continue sono sempre date da integrali, mentre quelle delle discrete da somme. Se il dominio della variabile X non è l'intera retta reale si fa l'ipotesi che $f(x)$ sia definita nulla per i valori esterni allo specificato dominio della variabile.

Esempio: consideriamo la possibilità di usare la funzione $f(x) = Ke^{-x}$ con K costante come funzione di densità della variabile X . Per la prima proprietà è chiaro che $K > 0$ deve essere positiva. Poiché :

$$\int_{-\infty}^{+\infty} e^{-x} dx = -e^{-x} \Big|_{-\infty}^{+\infty} = 0 + e^{+\infty}, \text{ segue che la variabile } X \text{ deve essere}$$

limitata. Quindi facciamo l'ipotesi che X possa assumere valori soltanto non negativi, allora $f(x)$ sarà definita nulla per valori negativi e data dalla formula suddetta per valori non negativi. Sotto questa ipotesi si ha che

$$\int_0^{+\infty} e^{-x} dx = -e^{-x} \Big|_0^{+\infty} = 0 + 1 = 1.$$

Quindi, se $K=1$, può essere una funzione di densità della variabile x la

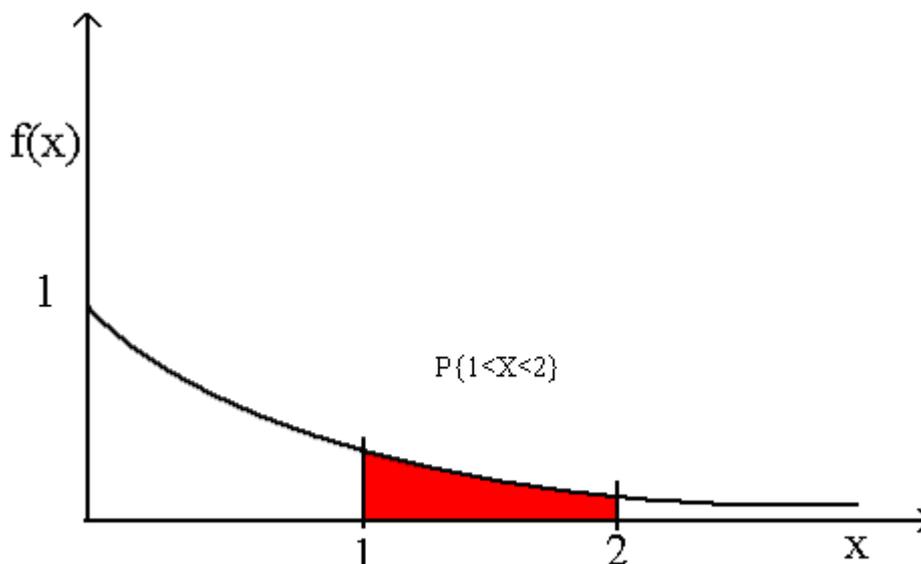
seguinte funzione :

$$f(x) = e^{-x}, \text{ con } x \geq 0 \text{ e } f(x) = 0, \text{ con } x \leq 0.$$

Il calcolo ad esempio della probabilità che $P\{1 < X < 2\}$ diventa allora uguale

$$a: \int_1^2 e^{-x} dx = -e^{-x} \Big|_1^2 = -e^{-2} + e^{-1} = 0,23.$$

Il grafico di questa funzione di densità e la rappresentazione come area della probabilità che $1 < X < 2$ è mostrato nella figura seguente:



E' importante sottolineare che sebbene in qualsiasi dato del problema la $f(x)$ si possa scegliere a piacimento, naturalmente secondo le conoscenze di cui dispone lo sperimentatore, una scelta per cui le probabilità risultanti non approssimano bene le frequenze relative osservate non è verosimilmente una scelta utile.

Funzione di distribuzione (o di distribuzione cumulativa o di ripartizione): la funzione in oggetto per la variabile continua X è data dalla (1):

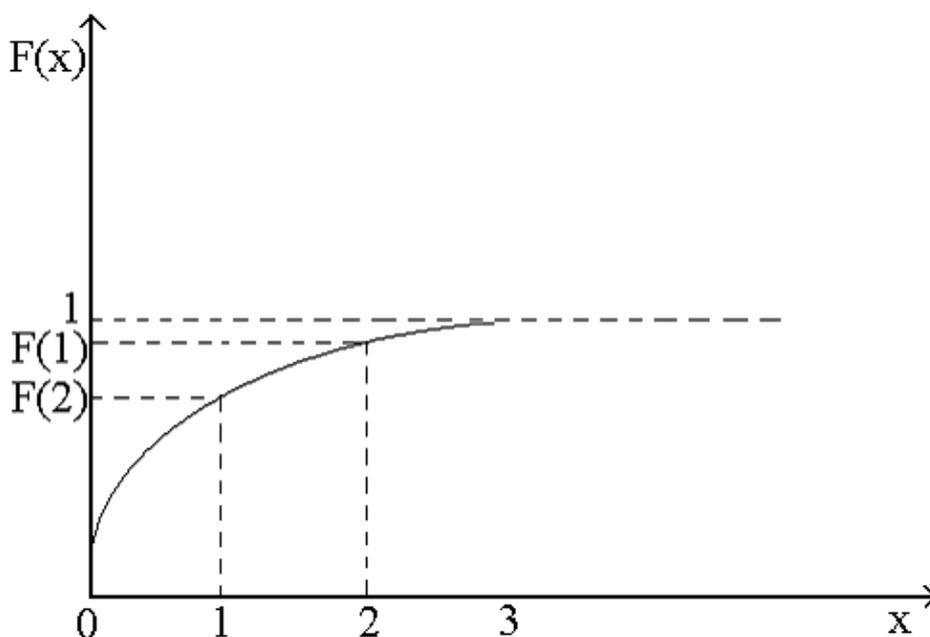
$$F(x) = P\{X \leq x\} = \int_{-\infty}^x f(t) dt$$

Esempio: come esempio si può ricavare dalla (1) la funzione $F(x)$ corrispondente alla funzione di densità nell'esempio trattato precedentemente . Si ha che :

$$F(x) = \int_0^x e^{-t} dt = -e^{-t} \Big|_0^x = 1 - e^{-x}, \text{ per } x \geq 0$$

$$F(x) = \int_0^x e^{-t} dt = -e^{-t} \Big|_0^x = 0, \text{ per } x < 0$$

Il grafico della $F(x)$ è mostrato in figura 1:



Si fa notare che la probabilità $P\{1 < x < 2\} = F(2) - F(1)$ cioè dalla differenza delle ordinate nel grafico della $F(x)$.

$$F(2) - F(1) = 1 - e^{-2} - (1 - e^{-1}) = e^{-1} - e^{-2} = 0,23 .$$

La *funzione di densità* è quella più comunemente usata nelle applicazioni della teoria statistica, ma anche la *funzione di distribuzione* è molto utile nel ricavare parte di quella teoria. A volte è più facile trovare la funzione di distribuzione di una variabile casuale che la sua funzione di densità. In questi casi la funzione di densità si può ricavare derivando la funzione di distribuzione per il teorema fondamentale del calcolo differenziale, cioè la derivata :

$$\frac{dF(x)}{dx} = f(x) \text{ purché la } f(x) \text{ sia una funzione } \textit{continua}.$$

Considerando l'esempio appena trattato si può infatti verificare che derivando la funzione di distribuzione si ottiene la corrispondente funzione di densità, infatti la derivata della $F(x)$ rispetto dx , cioè :

$$\frac{dF(x)}{dx} = \frac{d}{dx}(1 - e^{-x}) = e^{-x}, \text{ per } x \geq 0$$

$$\frac{dF(x)}{dx} = \frac{d}{dx}(1 - e^{-x}) = 0, \text{ per } x < 0$$

Funzione di densità congiunta continua: Una funzione di densità di due o più variabili casuali continue è la naturale generalizzazione di una funzione di densità di una variabile; così una funzione di densità di due variabili casuali continue X, Y si indica con $f(x, y)$ e, come è noto, è rappresentata geometricamente da una superficie nello spazio a tre dimensioni, cioè dalla superficie $z = f(x, y)$ proprio come la funzione di densità di una variabile è rappresentata da una curva nel piano xy , cioè dalla curva $y = f(x)$. Se gli integrali della $f(x, y)$ devono fornire probabilità è necessario che il volume totale sotteso da questa superficie sia uguale a 1 e che il volume sotteso da questa superficie che giace sopra una regione R nel piano xy dia la probabilità che le variabili casuali X e Y assumano valori che corrispondono a un punto interno a questa regione. Queste proprietà essenziali per una funzione di densità di due variabili si possono formalizzare come segue:

Definizione: una funzione di densità di due variabili casuali continue X e

f è una funzione f che possiede le seguenti proprietà:

$$1) f(x, y) \geq 0$$

$$2) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$$

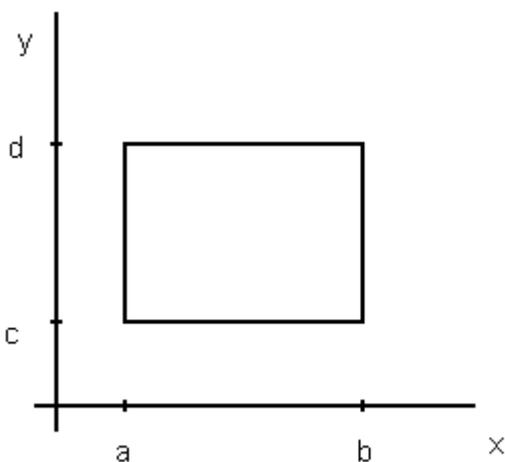
$$3) \iint_R f(x, y) dx dy = P\{X, Y \in R\}, \text{ per ogni } R$$

Si fa l'ipotesi che la regione R sia tale che l'integrale della $f(x, y)$ su quella regione esista.

Molto spesso la regione R sarà un rettangolo del tipo $a < x < b, c < y < d$, nel qual caso, la terza condizione diventa:

$$\int_a^b \int_c^d f(x, y) dx dy = P\{a < X < b, c < Y < d\}$$

(la superficie deve stare sopra al piano xy , non sotto! Dobbiamo avere un volume positivo!)



Esempio: come esempio di funzione di densità di due variabili casuali continue consideriamo la funzione $f(x, y) = e^{-(x+y)}$ che è una generalizzazione in due dimensioni dell'esempio considerato prima. Se la $f(x, y)$ si definisce essere nulla per valori negativi di x o y , e data dalla formula per valori non-negativi delle due variabili, ovvero :

$$f(x, y) = e^{-(x+y)} \quad \text{per } x, y \geq 0$$

$$f(x, y) = 0 \quad \text{per } x < 0 \text{ o } y < 0$$

Si può osservare che le prime due proprietà sono soddisfatte. Infatti:

$$\int_0^{+\infty} \int_0^{+\infty} e^{-(x+y)} dx dy = \int_0^{+\infty} \int_0^{+\infty} e^{-x} \cdot e^{-y} dx dy = \int_0^{+\infty} e^{-x} \left(\int_0^{+\infty} e^{-y} dy \right) dx =$$

$$= \int_0^{+\infty} e^{-x} \{-e^{-y}\}_0^{+\infty} dx = \int_0^{+\infty} e^{-x} (0 + e^0) dx = 1$$

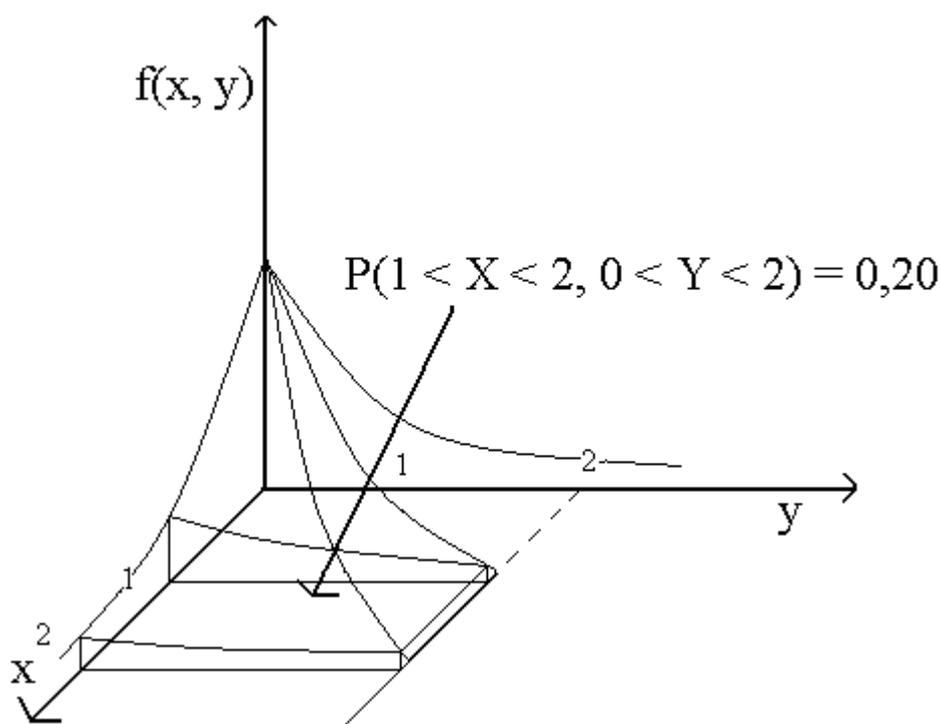
Il calcolo ad esempio della probabilità che $1 < X < 2$ e $0 < Y < 2$ è allora data da:

$$P\{1 < X < 2, 0 < Y < 2\} = \int_1^2 \int_0^2 e^{-(x+y)} dx dy = \int_1^2 \int_0^2 e^{-x} \cdot e^{-y} dx dy =$$

$$= \int_1^2 e^{-x} \left(\int_0^2 e^{-y} dy \right) dx = \int_1^2 e^{-x} \{-e^{-y}\}_0^2 dx = \int_1^2 e^{-x} (-e^{-2} + 1) dx =$$

$$= (1 - e^{-2}) \cdot \int_1^2 e^{-x} dx = (1 - e^{-2}) \cdot [-e^{-x}]_1^2 = (1 - e^{-2}) \cdot (-e^{-2} + e^{-1}) = 0.20$$

Il grafico di questa funzione di densità e la rappresentazione della probabilità che $1 < X < 2$ e $0 < Y < 2$ come volume sotteso dalla superficie $z = f(x, y) = e^{-(x+y)}$ è mostrato nella seguente figura:

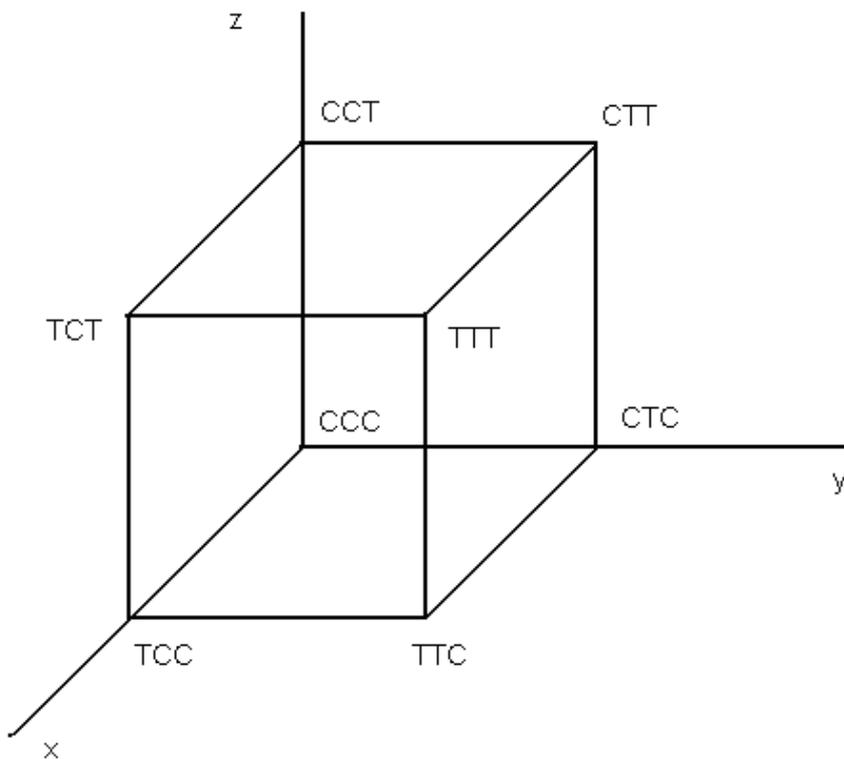


Due o più variabili casuali continue che sono senza rapporti in senso probabilistico si dicono **indipendenti** proprio come nel caso delle variabili

discrete. La definizione di indipendenza data nel caso delle variabili discrete si estende anche al caso delle variabili continue. Ci occuperemo fra poco di alcune distribuzioni di probabilità sia di variabili discrete che di variabili continue che nella realtà si sono dimostrate modelli particolarmente utili. In ciascun caso la distribuzione sarà specificata presentando la sua funzione di densità. In molti problemi però è sufficiente conoscere soltanto alcune caratteristiche o proprietà di una distribuzione anziché studiare l'intera distribuzione, e in particolare è spesso sufficiente che siano noti i cosiddetti *momenti di ordine basso* di una distribuzione. Ci occuperemo quindi dei momenti di particolari distribuzioni, come pure delle loro funzioni di densità.

Variabili discrete: la maggior parte delle variabili discrete che ricorrono negli esperimenti di tipo ripetitivo sono di tipo “conteggio”, come ad esempio il numero di incidenti che un automobilista ha in un anno; il numero di insetti che sopravvivono ad una irrorazione; ed altre ancora. Queste variabili possono assumere soltanto valori interi non negativi. Sebbene le variabili discrete che presenteremo siano di tipo conteggio, impiegheremo una notazione sufficientemente generale in modo da comprendere anche altri tipi di variabili casuali discrete.

Valore atteso: Prima di considerare specifiche funzioni di densità di variabili discrete ci occuperemo dei momenti di una distribuzione di probabilità, così da essere poi in grado di calcolare i momenti di quelle particolari distribuzioni. Ancora prima però introdurremo e definiremo un concetto più generale e cioè quello di *valore atteso*, in quanto i momenti saranno casi particolari del valore atteso. Inoltre il valore atteso è anche uno strumento molto utile per studiare altre proprietà di una distribuzione. Come esempio per illustrare il concetto generale di valore atteso consideriamo un gioco in cui si lanciano 3 monete e si vince un 1\$ per ogni testa che esce. Se la variabile X indica la somma vinta, allora essa può assumere soltanto i valori 0, 1, 2, 3 cui corrispondono le probabilità $1/8$, $3/8$, $3/8$, $1/8$. Queste probabilità risultano evidenti se si considera lo spazio campione dell'esperimento del lancio di tre monete illustrato in figura:



Ci si dovrebbe perciò aspettare di vincere 0\$ per 1/8 di volte; 1\$ per 3/8; 2\$ per 3/8; 3\$ per 1/8 se il gioco fosse fatto un gran numero di volte. Ci si dovrebbe perciò aspettare di vincere in media la somma seguente:

$$(0 \cdot 1/8) + (1 \cdot 3/8) + (2 \cdot 3/8) + (3 \cdot 1/8) = 1,50.$$

Questa somma, cioè 1,50 \$, è ciò che rappresenta comunemente la somma che ci si aspetta di vincere.

Supponiamo che la variabile casuale discreta X debba assumere uno dei valori x_1, x_2, \dots, x_k e che le probabilità associate a quei valori siano $f(x_1), f(x_2), \dots, f(x_k)$. Allora il valore atteso della variabile X è definito dalla formula:

$$(1) E[X] = \sum_{i=1}^k f(x_i) x_i \quad (E[] \text{ sta per "expected value"})$$

Nell'esempio visto la variabile X è la somma vinta nel lanciare 3 monete i cui valori possibili sono:

$$x_1 = 0, x_2 = 1, x_3 = 2, x_4 = 3 ;$$

e le probabilità corrispondenti sono:

$$f(x_1) = 1/8, f(x_2) = 3/8, f(x_3) = 3/8, f(x_4) = 1/8.$$

Ora supponiamo che il gioco considerato venga modificato in modo da vincere la somma $g(x_i)$ invece di x_i quando si ottiene il valore x_i . Ad esempio per $g(x)$ si potrebbe scegliere la funzione $g(x)=x^2$, il che significa che si vincerebbe in \$ il quadrato del numero di teste uscite.

Allora il valore atteso nel gioco così modificato sarebbe dato dalla formula:

$$(2) E[g(X)] = \sum_{i=1}^k g(x_i) \cdot f(x_i)$$

Nelle formule (1) e (2) si assume che ci sia soltanto un numero finito di valori possibili della variabile casuale X . Per eliminare questa limitazione si introduce la seguente definizione più generale applicabile a qualsiasi variabile casuale discreta:

Definizione: Il valore atteso della funzione $g(X)$ della variabile casuale discreta X , la cui funzione di densità è f , è dato da:

$$(3) E[g(X)] = \sum_{i=1}^{\infty} g(x_i) \cdot f(x_i)$$

Si intende in questa definizione che x_1, x_2, \dots sono i possibili valori di X e $f(x_1), f(x_2), \dots$ sono le probabilità corrispondenti. Il valore atteso della variabile X viene chiamato di solito la **MEDIA della variabile casuale X** , oppure la **MEDIA della distribuzione della variabile casuale X** .

Se X può assumere soltanto un numero finito, ad esempio k , di possibili valori, allora l'indice superiore della somma al secondo membro della (3) sarà naturalmente k anziché ∞ . Quando la variabile X possiede un numero infinito di valori possibili con probabilità positive è necessario assumere che la serie infinita al secondo membro della (3) converga.

MOMENTI: I momenti della distribuzione di una variabile casuale discreta X si possono definire in termini di valori attesi come segue:

Definizione: il momento d'ordine k dall'origine della distribuzione della variabile casuale discreta X , la cui densità è f , è dato da:

$$(4) \mu'_k = E[X^k] = \sum_{i=1}^{\infty} x_i^k \cdot f(x_i)$$

Il momento d'ordine k di una distribuzione è comunemente chiamato il momento d'ordine k della variabile casuale X avente, si intende, quella distribuzione. Così si può parlare di μ'_k come del momento d'ordine k della variabile X , oppure come del momento d'ordine k della distribuzione di X . Il momento del primo ordine $\mu_1 = E[X]$ ricorre così spesso che esso viene indicato con il simbolo μ (senza apice ' e pedice k).

Poiché anche i momenti dalla media sono usati ampiamente, essi pure necessitano di essere definiti.

Definizione: il momento d'ordine k dalla media della distribuzione della variabile casuale discreta X , la cui densità è f , è dato da:

$$\mu_K = E[(X-\mu)^k] = \sum_{i=1}^{\infty} (x_i - \mu)^k \cdot f(x_i)$$

I momenti di una distribuzione aiutano a descrivere la distribuzione quando la funzione di densità non è disponibile.

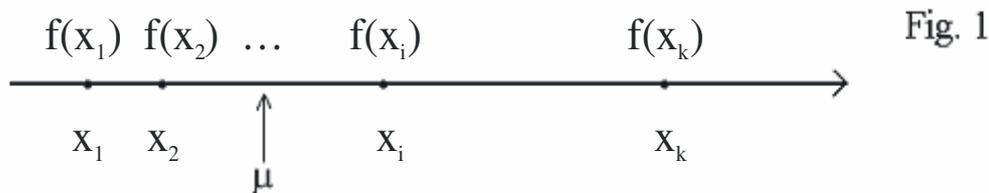
Una gran parte dei problemi che implicano i momenti sarà interessata soltanto ai primi due momenti, perché essi spesso sono sufficienti a descrivere due utili proprietà della distribuzione, cioè dove la distribuzione è centrata e in che grado la distribuzione è concentrata attorno a questo centro.

Il momento del primo dall'origine μ si usa per determinare dove la distribuzione è centrata e il momento del secondo ordine dalla media μ_2 si usa per determinare il grado di concentrazione della distribuzione attorno alla media μ . Poiché anche il momento del secondo ordine dalla media si usa molto spesso a questo proposito, esso si indica col simbolo $\mu_2 = \sigma^2$ ("sigma" al quadrato) e viene chiamato **varianza della distribuzione**.

La radice quadrata positiva della varianza, cioè σ , viene chiamato **scarto quadratico medio** o **deviazione standard** della distribuzione.

Esso si impiega al posto della varianza quando si desidera una misura della concentrazione nelle stesse unità di misura della variabile casuale.

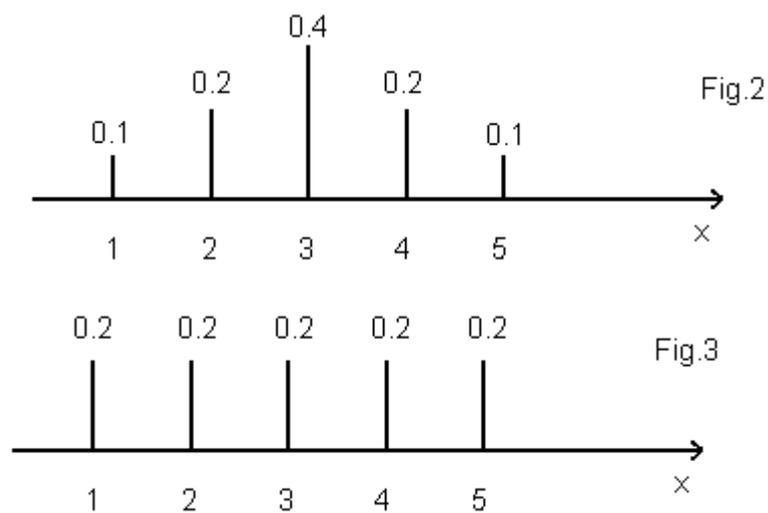
Per esaminare in che modo μ e σ^2 aiutano a descrivere una distribuzione di probabilità discreta, le probabilità $f(x_1), f(x_2), \dots, f(x_k)$ associate ai valori x_1, x_2, \dots, x_k della variabile casuale X , siano rappresentate geometricamente come masse puntiformi la cui somma è 1 e localizzate nei punti x_1, x_2, \dots, x_k sull'asse x . Questa rappresentazione è mostrata in figura 1:



Allora il momento del primo ordine dall'origine μ per come è definito dà il centro di gravità di questo insieme di masse puntiformi e perciò può servire come una misura di dove è centrata o localizzata la distribuzione.

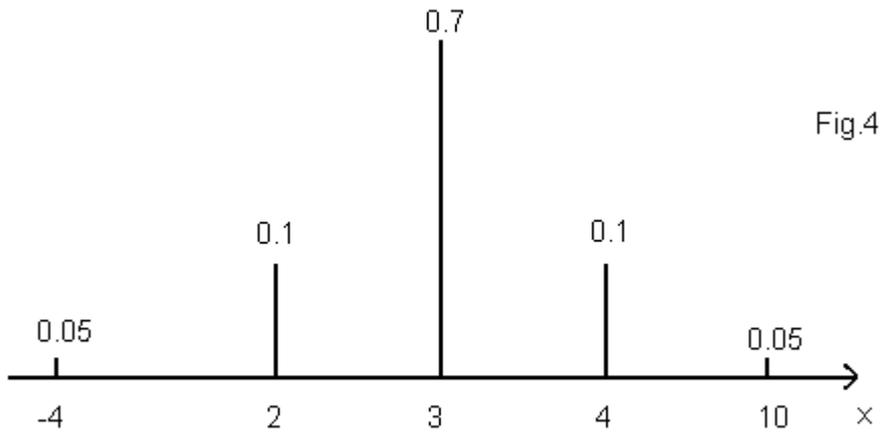
Il momento del secondo ordine dalla media σ^2 tenderà ad essere piccolo se la maggior parte delle masse delle probabilità sono concentrate vicino alla media μ e non ci sono masse di probabilità molto lontane da μ . Tanto più disperse sono le masse attorno alla media, tanto più grande è verosimile che diventi σ^2 ; così la varianza di una distribuzione può servire come una misura del grado in cui la distribuzione è concentrata attorno alla media. E' facile tuttavia dare esempi di distribuzioni per cui la varianza è una misura insufficiente della concentrazione della distribuzione attorno alla sua media. Ciò nonostante per la maggior parte delle distribuzioni più comuni essa assolve il suo compito in modo soddisfacente.

Come esempio in cui la varianza σ^2 serve a misurare la concentrazione consideriamo le due distribuzioni di probabilità mostrate nelle figure seguenti in cui le probabilità sono rappresentate da segmenti verticali(fig.2)(fig.3):



Per entrambe le distribuzioni, applicando la formula che definisce i momenti dall'origine, cioè la (4) con $k=1$, si ottiene $\mu=3$. Calcoli basati sulla formula che definisce i momenti dalla media, cioè la (5) con $k=2$, si ottiene per le due distribuzioni rispettivamente $\sigma^2=1.2$ e $\sigma^2=2.0$, e ciò mostra che la prima distribuzione è più concentrata della seconda rispetto alla sua media.

Come esempio, invece, in cui il valore di σ^2 non sembra indicare la natura della concentrazione della distribuzione attorno alla sua media consideriamo la distribuzione di fig.4:



In questo caso i calcoli danno ancora $\mu=3$ e $\sigma^2=5.1$, valore che confrontato con quelli delle fig.2 e fig.3 indicherebbe scarsa concentrazione della distribuzione attorno alla sua media, mentre invece la maggior parte della distribuzione è concentrata attorno alla media stessa.

Si può notare quindi che prendendo una massa di probabilità molto piccola, sufficientemente lontano dal centro della distribuzione, il valore della varianza si può rendere tanto grande quanto si vuole; nonostante la limitazione della varianza come misura della concentrazione di una distribuzione attorno alla sua media, come indicato nell'esempio in fig.4, la media e la varianza si sono dimostrate quantità molto utili nel trattare le distribuzioni più comunemente usate.

Per calcolare la *VARIANZA* è spesso più conveniente calcolare i primi due momenti dall'origine e poi calcolare la varianza da essi, piuttosto che calcolarla direttamente usando la formula che la definisce. Ricordando la formula che definisce la varianza, ciò si può realizzare come segue:

$$\sigma^2 = \sum_{i=1}^{\infty} (x_i - \mu)^2 \cdot f(x_i) = \overbrace{\sum_{i=1}^{\infty} x_i^2 \cdot f(x_i)}^{\mu_2'} - 2\mu \cdot \overbrace{\sum_{i=1}^{\infty} x_i \cdot f(x_i)}^{\mu} + \mu^2 \cdot \overbrace{\sum_{i=1}^{\infty} f(x_i)}^1$$

Quindi ricordando la formula che definisce i momenti dall'origine, si può scrivere che $\sigma^2 = \mu_2' - 2\mu\mu + \mu^2$ e quindi la formula richiesta è:

$$\sigma^2 = \mu_2' - \mu^2 \quad (\text{con questo si calcola la varianza})$$

Funzione generatrice dei momenti: Anche se il calcolo diretto dei momenti dalla formula che li definisce può essere facile è spesso conveniente essere capaci di calcolare tali momenti indirettamente servendosi di un altro metodo che ora introdurremo. Esso implica quella che è nota come la *funzione generatrice dei momenti*. Come indica il nome stesso, la funzione generatrice dei momenti è una funzione che genera i momenti ed è definita come segue:

Definizione: la funzione generatrice dei momenti della variabile casuale discreta X , la cui densità è f , è data da:

$$(1) M_X(\theta) = E[e^{\theta X}] = \sum_{i=1}^{\infty} e^{\theta x_i} \cdot f(x_i)$$

Questa serie è una funzione del parametro θ soltanto, ma si è messo l'indice $M(\theta)$ per indicare quale variabile si sta considerando. Il parametro θ non ha qui nessun significato reale, esso si introduce semplicemente per aiutare a determinare i momenti. Per mostrare come $M_X(\theta)$ genera i momenti facciamo l'ipotesi che la funzione f sia una funzione di densità per cui la serie nella (1) converga. Sviluppiamo ora $e^{\theta x_i}$ nella (1) in serie di potenze e sommiamo termine a termine. Perché la serie di potenze per e^z è uguale a $1 + \frac{z}{1!} + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots$ segue dalla formula suddetta e da quella che definisce i momenti dall'origine che:

$$\begin{aligned} (2) \quad M_X(\theta) &= \sum_{i=1}^{\infty} \left(1 + \frac{\theta}{1!} x_i + \frac{\theta^2}{2!} x_i^2 + \frac{\theta^3}{3!} x_i^3 + \dots\right) \cdot f(x_i) = \\ &= \underbrace{\sum_{i=1}^{\infty} f(x_i)}_1 + \frac{\theta}{1!} \cdot \underbrace{\sum_{i=1}^{\infty} x_i \cdot f(x_i)}_{\mu_1} + \frac{\theta^2}{2!} \cdot \underbrace{\sum_{i=1}^{\infty} x_i^2 \cdot f(x_i)}_{\mu_2} + \dots = \\ &= 1 + \frac{\theta}{1!} \mu_1 + \frac{\theta^2}{2!} \mu_2 + \frac{\theta^3}{3!} \mu_3 + \dots + \frac{\theta^k}{k!} \mu_k + \dots \end{aligned}$$

In questo sviluppo si può osservare che il coefficiente di $\frac{\theta^k}{k!}$ è il momento d'ordine k dall'origine, di conseguenza se si può determinare la funzione generatrice dei momenti di una variabile X e si può sviluppare in serie di

potenze di θ i momenti della variabile si possono ottenere semplicemente esaminando lo sviluppo risultante.

Ora se si desidera un momento particolare può essere conveniente calcolarlo calcolando la derivata fatta rispetto a θ di $M_x(\theta)$ calcolata per $\theta=0$. Infatti derivando ripetutamente la (2) rispetto a θ si ottiene che:

$$\mu'_k = \left. \frac{d^k M_x(\theta)}{d\theta^k} \right]_{\theta=0}$$

Infatti per $k=1$ si ha che $M'_x(\theta) = \mu'_1 + \mu'_2\theta + \frac{1}{2}\mu'_3\theta^2 + \dots$

Quindi $M'_x(\theta) \Big|_{\theta=0} = \mu'_1$.

Per $k=2$ si ha che $M''_x(\theta) = \mu'_2 + \mu'_3\theta + \dots$

Quindi $M''_x(\theta) \Big|_{\theta=0} = \mu'_2$ e così via...

DISTRIBUZIONE BINOMIALE: Consideriamo un esperimento di tipo ripetitivo in cui si registra soltanto il verificarsi o il non verificarsi di un evento. Supponiamo che sia p la probabilità che l'evento si verifichi quando l'esperimento viene eseguito. Indichiamo con $q = 1-p$ la probabilità che esso non si verifichi. Se l'evento si verifica in una data esecuzione dell'esperimento esso si chiamerà un successo, altrimenti un insuccesso.

Siano poi fatte n esecuzioni indipendenti e la variabile X rappresenti il numero di successi che si otterranno nelle n esecuzioni. Affronteremo ora il problema di determinare la probabilità di ottenere x successi nelle n esecuzioni dell'esperimento, ovvero che $X = x$.

A tale scopo determiniamo dapprima la probabilità di ottenere x successi seguiti da $n-x$ insuccessi. Chiaramente questi n eventi sono indipendenti, perciò ricordando una formula vista a suo tempo e valida per gli eventi indipendenti, e cioè $P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2\}$, questa probabilità è data da :

$$\overbrace{p \cdot p \cdot p \dots p}^x \cdot \overbrace{q \cdot q \cdot q \dots q}^{n-x} = p^x \cdot q^{n-x}$$

La probabilità di ottenere x successi ed $n - x$ insuccessi, se questi eventi si verificano in qualche altro ordine, non cambia, perché basta semplicemente riordinare le p e le q in modo da corrispondere al nuovo ordine. Per risolvere il problema è perciò necessario contare il numero di ordini possibili; il numero di ordini possibili non è altro che il numero di permutazioni possibili con n oggetti di cui x sono fra loro uguali, cioè le p

ed $n-x$ sono anch'esse fra loro uguali, cioè le q . Ma come è noto il numero di tali permutazioni è:

$$(3) \quad \frac{n!}{x!(n-x)!}$$

Ora ricordando la formula $P\{A_1 \cup A_2\} = P\{A_1\} + P\{A_2\}$, la probabilità che si verifichi l'uno o l'altro di un insieme di eventi reciprocamente esclusivi è data dalla somma delle loro probabilità. Di conseguenza è necessario sommare $p^x \cdot q^{n-x}$ tante volte quanti sono gli ordini diversi in cui il risultato richiesto può verificarsi.

Poiché la (3) dà il numero di tali ordini, la probabilità di ottenere x successi in qualche ordine si ottiene perciò moltiplicando $p^x \cdot q^{n-x}$ per la quantità espressa nella (3).

La probabilità risultante, che è quindi quella di ottenere x successi in n esecuzioni indipendenti di un esperimento per cui p è la probabilità di successo in una singola esecuzione, definisce quella che è nota come la **distribuzione binomiale**, cioè:

$$(4) \quad f(x) = \frac{n!}{x!(n-x)!} \cdot p^x \cdot q^{n-x}$$

Il termine binomiale deriva dalla relazione della funzione di densità binomiale con il seguente sviluppo binomiale:

$$(q+p)^n = q^n + \frac{n}{1!} q^{n-1} p + \frac{n(n-1)}{2!} q^{n-2} p^2 + \dots + p^n$$

Il termine generale in questo sviluppo che contiene p^x , dove x è un intero, è dato da:

$$\frac{n(n-1) \cdot \dots \cdot (n-x+1)}{x!} q^{n-x} \cdot p^x = \frac{n!}{x!(n-x)!} q^{n-x} \cdot p^x$$

e rappresenta precisamente il valore della funzione di densità binomiale.

Come risultato si può scrivere che:

$$(q+p)^n = \sum_{x=0}^n \frac{n!}{x!(n-x)!} p^x \cdot q^{n-x}$$

Così i vari termini nello sviluppo binomiale di $(q+p)^n$ danno le probabilità dei vari e possibili risultati nel loro ordine.

La funzione di densità binomiale è un esempio di modello matematico che si può applicare a molti problemi reali che implicano una variabile discreta. In qualsiasi applicazione è comunque necessario conoscere o stimare il valore del parametro p prima di poter usare la funzione di densità binomiale.

Esercizio: Come applicazione della funzione di densità binomiale consideriamo l'esperimento del getto di un dado e si voglia calcolare la probabilità che, se il dado viene gettato 5 volte, 2 lanci presentino come risultato 1. Quindi si ha che $p=1/6$, $q=1-p=5/6$, $n=5$.

Applicando la funzione di densità binomiale si ottiene quindi:

$$f(2) = \frac{5!}{2!(5-2)!} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3 = 0.16 \quad (\text{due successi, due lanci che danno 1})$$

Si voglia inoltre calcolare la probabilità di ottenere al massimo 2 lanci che presentano come risultato 1. Per rispondere a questa domanda è necessario calcolare le probabilità di ottenere precisamente nessun 1, un 1 e due 1, probabilità quest'ultima che è stata appena calcolata. Applicando la funzione di densità binomiale si ottiene:

$$f(0) = \frac{5!}{0!5!} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^5 = 0.40$$

$$f(1) = \frac{5!}{1!4!} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^4 = 0.40$$

Poiché queste tre possibilità rappresentano tre eventi reciprocamente esclusivi, segue che :

$$P\{X \leq 2\} = f(0) + f(1) + f(2) = 0.4 + 0.4 + 0.16 = 0.96 \text{ (96\%)}$$

Esempio: Come secondo esempio consideriamo il calcolo della probabilità di ottenere almeno due successi nel tirare 20 colpi ad un bersaglio se la probabilità di successo in un singolo colpo è uguale a $1/10$. In questo caso $p=1/10$ ed $n=20$; perciò la probabilità è:

$$P\{X \geq 2\} = 1 - f(0) - f(1) = 1 - \frac{20!}{0!20!} \left(\frac{1}{10}\right)^0 \left(\frac{9}{10}\right)^{20} - \frac{20!}{1!19!} \left(\frac{1}{10}\right)^1 \left(\frac{9}{10}\right)^{19} = 0.608$$

La validità di usare il modello binomiale in questo esempio non è così ovvia come lo è invece nell'esempio precedente; infatti la formula binomiale è stata ricavata sulla base di esecuzioni dell'esperimento indipendenti tra loro con la probabilità P costante da esecuzione a esecuzione. Ma se la stessa persona tira colpi ripetuti allo stesso bersaglio ci si potrebbe aspettare che le sue probabilità di colpirlo aumentino un po' con la pratica; se invece venisse usata ogni volta una persona diversa, P indubbiamente cambierebbe da prova a prova. Quindi nel momento di

interpretare una probabilità risultante come quella pari a 0.608 appena ricavata, è importante prendere in considerazione le possibili deviazioni nelle ipotesi fondamentali.

Momenti binomiali: Supponiamo i primi due momenti dall'origine della distribuzione binomiale. Per illustrare i due metodi per il calcolo dei momenti, questi verranno calcolati da prima direttamente dalla definizione e poi indirettamente tramite la funzione generatrice dei momenti. Applicando la definizione alla funzione di densità binomiale espressa dalla (4) vista in precedenza si può scrivere:

$$\begin{aligned} \mu = E[X] &= \sum_{x=0}^n x \frac{n!}{x!(n-x)!} p^x \cdot q^{n-x} = \sum_{x=1}^n x \frac{n!}{x!(n-x)!} p^x \cdot q^{n-x} = \\ &= \sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} p^x q^{n-x} \end{aligned}$$

Ma tutto ciò si può anche scrivere come:

$$(1) \mu = n \cdot p \cdot \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}$$

Ponendo nella (1) $y=x-1$ si ottiene $\mu = n \cdot p \cdot \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-y-1)!} p^y q^{n-1-y}$

Ora la quantità che viene sommata al secondo membro è la proprietà binomiale di y successi in $n-1$ esecuzioni dell'esperimento. Poichè la somma viene fatta su tutti i possibili valori della variabile y , essa deve essere uguale a 1. Quindi infine si ha che:

$$\mu = n p$$

Il momento del secondo ordine dall'origine di una variabile binomiale si calcola usando l'identità $x=x(x-1)+x$. Dalla definizione si ha che:

$$\begin{aligned} \mu_2' &= \sum_{x=0}^n x^2 \frac{n!}{x!(n-x)!} p^x q^{n-x} = \sum_{x=0}^n [x(x-1) + x] \frac{n!}{x!(n-x)!} p^x q^{n-x} = \\ &= \sum_{x=0}^n x(x-1) \frac{n!}{x!(n-x)!} p^x q^{n-x} + \mu \end{aligned}$$

Poiché i termini della somma per $x=0$ e $x=1$ sono uguali a zero a causa del fattore $x(x-1)$, la sommatoria può cominciare da $x=2$:

$$\mu'_2 = \sum_{x=2}^n x(x-1) \frac{n!}{x!(n-x)!} p^x q^{n-x} + \mu = \sum_{x=2}^n \frac{n!}{(x-2)!(n-x)!} p^x q^{n-x} + \mu$$

Se si raccoglie fuori dalla sommatoria $n(n-1)p^2$, si può scrivere:

$$\mu'_2 = n(n-1)p^2 \sum_{x=2}^n \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x} + \mu$$

Ponendo ora $z=x-2$ si ha che:

$$\mu'_2 = n(n-1)p^2 \sum_{z=0}^n \frac{(n-2)!}{z!(n-2-z)!} p^z q^{n-2-z} + \mu$$

Ora la quantità che viene sommata è la probabilità binomiale di z successi in $n-2$ esecuzioni. Poiché la somma viene fatta su tutti i possibili valori di z , essa deve essere uguale a 1. Usando questo risultato e quello precedente, cioè $\mu = np$, si ottiene infine che:

$$(2) \mu'_2 = n(n-1)p^2 + np$$

Per calcolare la *varianza di una variabile binomiale* si può usare una formula già nota, e cioè $\sigma^2 = \mu'_2 - \mu^2$, per cui si ha che:

$$\sigma^2 = n(n-1)p^2 + np - n^2p^2 = np^2 + np = np(1-p) = npq$$

Lo *scarto quadratico medio di una distribuzione binomiale* è dato quindi da: $\sigma = \sqrt{npq}$

Ora calcoliamo i primi due momenti dall'origine di una variabile binomiale indirettamente, cioè per mezzo della funzione generatrice dei momenti. Quest'ultima per una variabile binomiale è data dalla:

$$M_x(\theta) = \sum_{x=0}^n e^{\theta x} \frac{n!}{x!(n-x)!} p^x q^{n-x} = \sum_{x=0}^n \frac{n!}{x!(n-x)!} (pe^\theta)^x q^{n-x}$$

Ma come si è già visto prima, la somma all'ultimo membro si può scrivere come un binomio elevato all' n -esima potenza e quindi si ha che:

$$(3) M_x(\theta) = (q + pe^\theta)^n \quad (\text{ricorda che } q + p^n = \sum_{x=0}^n \frac{n!}{x!(n-x)!} p^x q^{n-x})$$

I momenti richiesti si possono quindi ottenere applicando una formula già

nota, e cioè:

$$\mu'_k = \left. \frac{d^k M_x(\theta)}{d\theta^k} \right]_{\theta=0}$$

Derivando la (3) due volte rispetto a θ si ottiene:

$$M'_x(\theta) = n(q + pe^\theta)^{n-1} \cdot pe^\theta = npe^\theta \cdot (q + pe^\theta)^{n-1}$$

$$\begin{aligned} M''_x(\theta) &= npe^\theta (q + pe^\theta)^{n-1} + npe^\theta \cdot (n-1)(q + pe^\theta)^{n-2} \cdot pe^\theta = \\ &= npe^\theta \cdot (q + pe^\theta)^{n-2} [q + pe^\theta + (n-1)pe^\theta] = npe^\theta \cdot (q + pe^\theta)^{n-2} (q + npe^\theta) \end{aligned}$$

Ora i primi due momenti dall'origine si possono calcolare calcolando rispettivamente i valori di queste due derivate in $\theta=0$. Si ha quindi che:

$$\mu = M'_x(\theta) \Big]_{\theta=0} = np(q+p)^{n-1} = np$$

$$\mu'_2 = M''_x(\theta) \Big]_{\theta=0} = np(q+p)^{n-2}(q+np) = np(q+np)$$

Se in quest'ultima si sostituisce q con $(1-p)$ si ritrova per μ'_2 lo stesso valore espresso nella (2). Si può osservare che nel caso di una variabile binomiale i momenti sono più facili da calcolare indirettamente tramite la funzione generatrice dei momenti, che direttamente dalla definizione.

Distribuzione di Poisson: Se il numero n di esecuzioni dell'esperimento è grande, i calcoli coinvolti nell'uso della funzione di densità binomiale diventano piuttosto lunghi, perciò sarebbe molto utile poter disporre di una conveniente approssimazione della distribuzione binomiale. Risulta che per n grande ci sono due ben note approssimazioni della distribuzione binomiale. Una quando p è molto piccolo e l'altra quando non si è in questo caso. L'approssimazione che si applica quando p è molto piccolo è nota come la funzione di densità di Poisson ed è definita come segue:

$$(1) \quad f(x) = \frac{e^{-\mu} \mu^x}{x!}$$

dove il parametro μ è la media della distribuzione. Sebbene la distribuzione di Poisson venga introdotta qui come una approssimazione di quella binomiale, essa è una distribuzione ben nota e utile di per sé e non è da considerare soltanto come una approssimazione della distribuzione binomiale. Per verificare che la distribuzione di Poisson è una buona

approssimazione di quella binomiale per n molto grande e p molto piccolo si considera ciò che accade alla funzione di densità binomiale quando $n \rightarrow \infty$ e $p \rightarrow 0$, in modo tale che la media $\mu = n \cdot p$ rimanga fissa. Il risultato ottenuto si può esprimere sotto forma di teorema il cui enunciato è il seguente:

Teorema: Se la probabilità p di successo in una singola esecuzione tende a zero, mentre il numero di esecuzioni n tende all'infinito, in modo tale che la media $\mu = n \cdot p$ rimanga fissa, allora la distribuzione binomiale tenderà alla distribuzione di Poisson con media μ .

Le figure 1 e 2 sono state costruite per indicare con quale rapidità la distribuzione binomiale tende alla distribuzione di Poisson. Le linee tratteggiate rappresentano la distribuzione di Poisson, fissa con $\mu=4$, e le linee continue la distribuzione binomiale rispettivamente per $p=1/3$ e $p=1/24$.

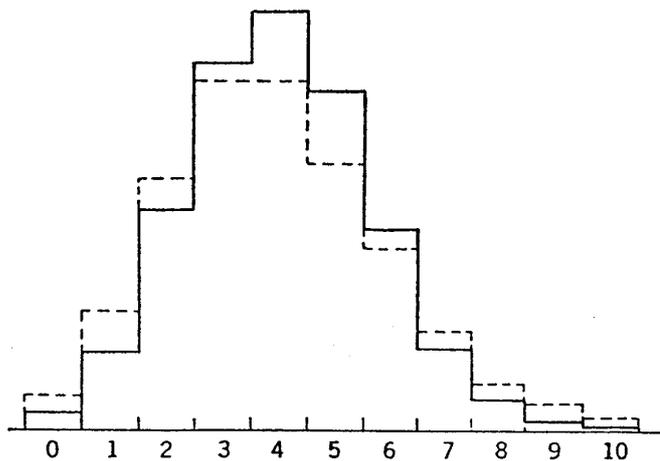


Fig. 1 Distribuzioni binomiale (—) e di Poisson (----) per $\mu=4$ e $p=\frac{1}{3}$.

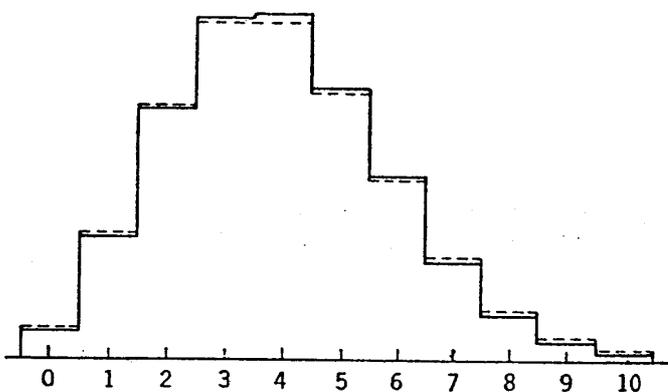


Fig. 2 Distribuzioni binomiale (—) e di Poisson (----) per $\mu=4$ e $p=\frac{1}{24}$.

Sembrerebbe dall'esame di questi grafici che l'approssimazione di Poisson dovrebbe essere sufficientemente accurata per la maggior parte delle applicazioni, se $n \geq 100$ e $p \leq 0.05$. Poiché la distribuzione di Poisson è stata ottenuta da quella binomiale, mantenendo la media $\mu = np$ fissa e permettendo ad n di tendere all'infinito, segue che i momenti della distribuzione di Poisson si possono ottenere dai corrispondenti momenti della distribuzione binomiale, calcolando i valori limiti di quest'ultimi, così la media della distribuzione di Poisson deve essere ovviamente uguale a μ , e la varianza deve essere il valore limite della varianza binomiale, cioè di npq . Si ha quindi che :

$$\lim_{n \rightarrow +\infty} npq = \lim_{n \rightarrow +\infty} \mu q = \lim_{p \rightarrow 0} \mu(1-p) = \mu$$

Si ottiene quindi l'interessante risultato che la varianza di una variabile di Poisson è uguale alla sua media. Sebbene la distribuzione di Poisson sia stata introdotta per mezzo della sua proprietà approssimante della distribuzione binomiale, essa è un modello molto utile per trattare certi tipi di problemi senza rapporti con la distribuzione binomiale; per esempio la distribuzione di Poisson si è trovata che è un modello soddisfacente per il numero di atomi che si disintegrano da un materiale radioattivo, per un numero di screpolature su di una lamina metallica, per il numero di chiamate telefoniche su di una linea in un intervallo di tempo fisso, ecc,...

Questi problemi sono tutti del tipo in cui una variabile casuale si distribuisce nel tempo e nello spazio.

Se si fa l'ipotesi che il numero di eventi che si verificano in intervalli di tempo che non si sovrappongono siano indipendenti, che la probabilità di un singolo evento che si verifica in un piccolo intervallo di tempo sia approssimativamente proporzionale alla dimensione dell'intervallo e che la probabilità che si verifichi più di un evento in un singolo intervallo di tempo sia trascurabile rispetto alla probabilità che si verifichi soltanto un evento in quell'intervallo, allora si può dimostrare con opportune tecniche di calcolo quando le precedenti ipotesi sono state precisate matematicamente, che il numero di eventi, in qualsiasi intervallo di tempo di dimensione fissata, possiederà una distribuzione di Poisson. Lo stesso discorso si può applicare quando gli intervalli di tempo sono sostituiti da intervalli spaziali. Perciò il numero di eventi che si distribuiscono in qualsiasi regione dello spazio di dimensione fissata seguirà una distribuzione di Poisson. Un intervallo spaziale può essere naturalmente ad una, due o tre dimensioni, così che il numero di screpolature in una data

lunghezza di filo metallico o in una data area di stoffa o in un dato blocco di calcestruzzo, seguirà una distribuzione di Poisson.

*Un esperimento in cui le osservazioni si succedono in intervalli successivi di tempo o di spazio, e per cui le ipotesi precedenti sono soddisfatte, così che gli eventi possiedono una distribuzione di Poisson, viene chiamato un **processo di Poisson**.* Così la distribuzione di Poisson è utile per trattare problemi riguardanti i processi di Poisson e non è giustificata, quindi, soltanto come un'approssimazione della distribuzione binomiale.

Esercitazione: Come esempio d'uso della distribuzione di Poisson, come approssimazione di quella binomiale, consideriamo il problema di calcolare la probabilità che si trovino al massimo 5 valvole difettose in una scatola di 200 valvole se l'esperienza mostra che il 2% di queste valvole sono difettose. In questo caso $\mu = np = 200 \cdot (0,02) = 4$; quindi, usando la distribuzione di Poisson, dove la variabile X rappresenta il numero di valvole difettose in una scatola di 200 valvole, la risposta approssimante è data da:

$$P\{X \leq 5\} = \sum_{x=0}^5 \frac{e^{-4} \cdot 4^x}{x!} = e^{-4} \left(1 + 4 + \frac{4^2}{2} + \frac{4^3}{6} + \frac{4^4}{24} + \frac{4^5}{120} \right) = 0.785 = 78,5\%$$

Lunghi calcoli, usando la funzione di densità binomiale, porterebbero al risultato che: $P\{X \leq 5\} = 0,788$; quindi, l'approssimazione in questo caso è molto buona, il che significa che l'errore relativo è il seguente:

$$\left| \frac{\Delta P}{P} \right| = \left| \frac{0.785 - 0.788}{0.788} \right| = 0.0038 = 3,8\text{‰} \text{ (per mille)}$$

Ciò conferma quanto si è già detto, cioè che l'espressione di Poisson si può ritenere sufficientemente accurata per la maggior parte delle applicazioni se, come avviene in questo caso, $n \geq 100$ e $p \leq 0,05$.

Esempio: come esempio d'uso della distribuzione di Poisson, come modello di un processo di Poisson, consideriamo il problema seguente. Supponiamo che l'esperienza abbia mostrato che il numero medio di chiamate telefoniche che arriva ad un quadro di controllo in un minuto è uguale a 5. Se il quadro di controllo può trattare al massimo 8 chiamate al minuto si vuole calcolare la probabilità che esso non sia capace di trattare tutte le chiamate che arrivano in un minuto. La probabilità richiesta si può ottenere calcolando la probabilità di ricevere al massimo 8 chiamate e poi

sottraendo questa probabilità da 1. Usando $\mu=5$ nella funzione di densità di Poisson si ha che:

$$P\{X \leq 8\} = \sum_{x=0}^8 \frac{e^{-5} \cdot 5^x}{x!} = 0.932$$

Di conseguenza la probabilità richiesta è data da:

$$P\{X > 8\} = 1 - 0.932 = 0.068$$

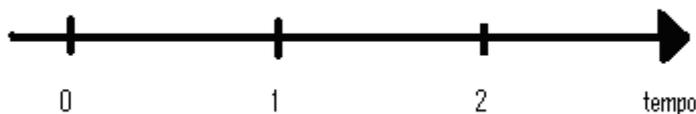
Per la natura delle ipotesi che hanno condotto alla distribuzione di Poisson, per problemi di questo tipo, è legittimo scegliere qualunque intervallo di tempo o di spazio di lunghezza desiderata e calcolare la probabilità di eventi in questo intervallo tramite la distribuzione di Poisson, con l'avvertenza di aggiustare la media μ per quell'intervallo.

Per esempio: la soluzione del problema di calcolare la probabilità che si ricavano al massimo 6 chiamate in un periodo di 2 minuti, quando il numero medio di chiamate al minuto è uguale a 5, si ottiene scegliendo $\mu=10$ e calcolando la probabilità come segue:

$$P\{X \leq 6\} = \sum_{x=0}^6 \frac{e^{-10} \cdot 10^x}{x!} = 0.13 = 13\%$$

Una soluzione basata sul trattare questo come un esperimento a due stadi con ciascun minuto di tempo come uno stadio, darebbe origine a calcoli eccessivamente lunghi; si dovrebbe essere quindi spinti a risolvere il problema in questo modo così da apprezzare meglio la precedente proprietà della distribuzione di Poisson.

Esempi:



| | | | |
|----------|---|---|---------------|
| chiamate | { | 0 | 0,1,2,3,4,5,6 |
| | | 1 | 0,1,2,3,4,5 |
| | | 2 | 0,1,2,3,4 |

$$P(0,0) = f(0) \cdot f(6)$$

Se faccio così per tutte le probabilità e sommo ottengo 0,13.

VARIABILI CASUALI - MOMENTI – VALORE ATTESO

Prima di presentare alcune utili distribuzioni di variabili casuali continue, definiamo il momento di ordine k di una variabile di questo tipo.

Momenti: sia $f(x)$ la funzione di densità di una variabile casuale continua X che è nulla esternamente ad un certo intervallo finito (a, b) . Si può dare allora la seguente definizione:

Definizione: Se il momento d'ordine k dall'origine della distribuzione della variabile casuale continua X , la cui densità è f , è data da:

$$(1) \quad \mu'_k = \int_a^b x^k f(x) dx$$

Per analogia con la definizione data per una variabile discreta il momento d'ordine k della media, indicato con μ_k , si ottiene sostituendo x^k con $(x-\mu)^k$ nel precedente integrale, cioè:

$$(2) \quad \mu_k = \int_a^b (x - \mu)^k f(x) dx$$

Come nel caso delle variabili discrete è desiderabile introdurre il concetto generale di valore atteso e trattare i momenti come casi particolari del valore atteso. La definizione richiesta segue dall'analogia con quella data per le variabili discrete ed è la seguente:

Definizione: il valore atteso della funzione $g(X)$ della variabile casuale continua X , la cui densità è f , è dato da:

$$(3) \quad E[g(X)] = \int_a^b g(x) \cdot f(x) dx$$

Sebbene per le definizioni dei momenti e del valore atteso la $f(x)$ sia stata considerata nulla esternamente all'intervallo $[a, b]$, non c'è nessun motivo per mantenere questa limitazione. Così più in generale a può essere uguale a $-\infty$, e b a $+\infty$.

Poiché l'operatore valore atteso "E" è stato progettato per fornire valori medi di variabili casuali, sorge spontanea la domanda se il valore atteso di una funzione $g(X)$ di una variabile casuale X sia il valore atteso (la media ?) di quella funzione. Indichiamo con Y la variabile casuale $g(X)$, ovvero $Y=g(X)$; allora conoscendo la funzione di densità $f(x)$ di X è teoricamente possibile trovare la funzione di densità $h(y)$ della variabile Y .

Il valore atteso di $g(X)$ è ovviamente lo stesso del valore atteso di Y , perciò se $h(y)$ è disponibile, allora il valore atteso di Y si può esprimere nella

forma:

$$E[Y] = \int_{-\infty}^{+\infty} y \cdot h(y) dy$$

Usando tecniche di calcolo con cambiamento di variabile si può dimostrare che questo valore è lo stesso del valore dato dalla formula:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

Il vantaggio di quest'ultima sta nel fatto che essa non richiede di trovare la funzione di densità $h(y)$ dalla conoscenza della funzione di densità $f(x)$ prima di poter calcolare il valore atteso di $Y = g(X)$.

Funzione generatrice dei momenti: La funzione generatrice dei momenti di una variabile casuale continua X è definita per analogia col caso discreto come segue:

$$(4) \quad M_x(\theta) = E[e^{\theta x}] = \int_{-\infty}^{+\infty} e^{\theta x} \cdot f(x) dx$$

La $e^{\theta x}$ si sviluppa in serie di potenze e si compie l'integrazione termine a termine; si troverà che $M_x(\theta)$ assume la stessa forma estesa già vista e precisamente:

$$M_x(\theta) = 1 + \frac{\theta}{1!} \mu_1' + \frac{\theta^2}{2!} \mu_2' + \dots$$

Quindi la (4) genera i momenti nello stesso modo come fa l'analogia per le variabili discrete. Per poter generare i momenti di una funzione $g(X)$ della variabile casuale continua X , è necessario generalizzare la definizione di funzione generatrice dei momenti. Dal modo in cui $M_x(\theta)$ genera i momenti è chiaro che i momenti di $g(X)$ saranno generati sostituendo nella (4) $e^{\theta x}$ con $e^{\theta g(x)}$.

Definizione: La funzione generatrice dei momenti della funzione $g(X)$ della variabile casuale continua X , la cui funzione di densità è f , è data da:

$$(5) \quad M_{g(x)}(\theta) = E[e^{\theta g(x)}] = \int_{-\infty}^{+\infty} e^{\theta g(x)} \cdot f(x) dx$$

Questa è la forma generalizzata della funzione generatrice dei momenti. Consideriamo ora due interessanti proprietà delle funzioni generatrici dei momenti.

Proprietà: Consideriamo ora due interessanti proprietà della funzione generatrice dei momenti. Sia c una costante ed $h(X)$ una funzione della variabile X per la quale la funzione generatrice dei momenti esiste, allora, poiché nella (5) $g(X)=c \cdot h(X)$, si ottiene:

$$M_{c \cdot h(x)}(\theta) = \int_{-\infty}^{+\infty} e^{\theta \cdot c \cdot h(x)} \cdot f(x) dx = M_{h(x)}$$

Scegliendo ora nella (5) $g(X) = h(X + c)$ si ottiene:

$$M_{h(X)+c}(\theta) = \int_{-\infty}^{+\infty} e^{\theta[h(x)+c]} \cdot f(x) dx = e^{\theta c} \int_{-\infty}^{+\infty} e^{\theta h(x)} \cdot f(x) dx = e^{\theta c} M_{h(x)}(\theta)$$

Ora sostituendo $h(x)$ con $g(x)$ nelle formule precedenti si possono formulare le seguenti due importanti proprietà:

-Se c è una costante qualsiasi e $g(X)$ è una funzione qualsiasi per cui la funzione generatrice dei momenti esiste, allora:

I) $M_{c \cdot g(x)}(\theta) = M_{g(x)}(c \cdot \theta)$

II) $M_{g(x)+c}(\theta) = e^{\theta c} M_{g(x)}(\theta)$

Queste due proprietà ci permettono di disporre nel modo indicato di una costante c che moltiplica θ e viene aggiunta a una funzione $g(x)$. Sostituendo integrali con somme, si mostra facilmente che questa formule si applicano anche alle variabili discrete. Si fa l'ipotesi che $f(x)$ e $g(x)$ siano tali che l'integrale nella (5) risulti finito.

Distribuzione rettangolare o uniforme: forse la più semplice distribuzione di una variabile casuale continua è la distribuzione che è costante in un intervallo finito $[a, b]$ ed è uguale a zero altrove. Questa definisce quella che è conosciuta come la *distribuzione rettangolare* o *uniforme*, e cioè:

$$(6) \quad f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{altrove} \end{cases}$$

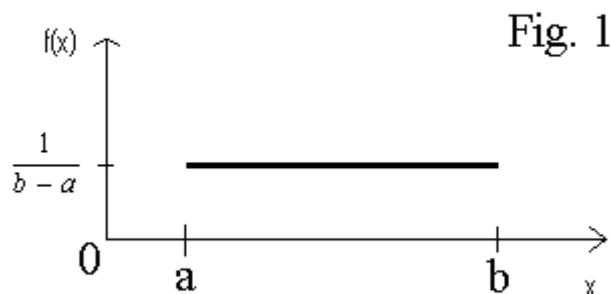


Fig. 1

Il grafico di una tipica distribuzione rettangolare è mostrato in Fig.1.

Il momento d'ordine k dall'origine della distribuzione rettangolare è facile da calcolare. Se ad esempio per semplicità $a=0$ e $b=1$, l'applicazione della definizione alla (6) fornisce:

$$\mu'_k = \int_0^1 x^k dx = \left[\frac{1}{k+1} x^{k+1} \right]_0^1 = \frac{1}{k+1}$$

La funzione generatrice dei momenti è anch'essa facile da calcolare, applicando la definizione alla (6) si ottiene:

$$M_x(\theta) = \int_0^1 e^{\theta x} dx = \left[\frac{1}{\theta} e^{\theta x} \right]_0^1 = \frac{1}{\theta} (e^\theta - 1)$$

Se si volesse ottenere il momento d'ordine k dalla $M_x(\theta)$, sarebbe necessario come si è già detto sviluppare e^θ in serie di potenze, così facendo si ottiene:

$$M_x(\theta) = \frac{1}{\theta} \cdot \left(1 + \frac{\theta}{1!} + \frac{\theta^2}{2!} + \dots - 1 \right) = 1 + \frac{\theta}{2!} + \frac{\theta}{3!} + \dots + \frac{\theta^k}{(k+1)!}$$

Poiché μ'_k è il coefficiente di $\frac{\theta^k}{k!}$ si può osservare da questo sviluppo che

μ'_k è uguale a $\mu'_k = \frac{1}{k+1}$, risultato che coincide ovviamente con quello ottenuto dalla definizione.

Quest'ultimo calcolo è stato fatto allo scopo di acquistare familiarità con la funzione generatrice dei momenti e non come un metodo suggerito per calcolare i momenti in questo caso; infatti qui è molto più semplice il calcolo diretto dei momenti. La distribuzione rettangolare è di uso piuttosto limitato come modello per le distribuzioni reali, comunque essa è di considerevole valore teorico ed è la distribuzione più semplice di una variabile casuale continua su cui applicare le formule generali.

Distribuzione normale o gaussiana: Senza dubbio il modello che si è dimostrato il più utile di tutte le distribuzioni per le variabili casuali continue è la distribuzione chiamata normale o gaussiana. Essa in generale è definita come segue:

$$(1) \quad f(x) = c \cdot e^{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2}$$

dove a, b e c sono parametri che fanno della $f(x)$ una funzione di densità di probabilità; questi parametri, quindi, devono essere tali che:

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

Come risultato vedremo fra poco che ci sono in effetti soltanto due parametri indipendenti che determinano questa funzione di densità. Dalla funzione (1) è chiaro che la curva che essa rappresenta è simmetrica rispetto alla retta $x=a$, quindi per simmetria la media deve essere data da $\mu=a$. (se sostituisco x con $a \pm x$ non cambia nulla)

Momenti: calcoliamo ora i momenti indirettamente per mezzo della funzione generatrice dei momenti; inoltre, poiché è più facile, in questo caso, calcolare i momenti dalla media che quelli dall'origine; consideriamo il calcolo di:

$$M_{X-\mu}(\theta) \left(M_X(\theta) = 1 + \frac{\theta}{1!} \mu_1' + \frac{\theta^2}{2!} \mu_2' + \dots \Rightarrow M_{X-\mu}(\theta) = 1 + \frac{\theta}{1!} \mu_1 + \frac{\theta^2}{2!} \mu_2 + \dots \right)$$

Si ha quindi che, ricordando la forma generalizzata della funzione generatrice vista prima, cioè:

$$M_{g(X)}(\theta) = E \left[e^{\theta g(X)} \right] = \int_{-\infty}^{+\infty} e^{\theta g(x)} f(x) dx \quad \text{con } g(X) = X - \mu$$

Si ha che:

$$M_{X-\mu}(\theta) = c \cdot \int_{-\infty}^{+\infty} e^{\theta(x-\mu)} e^{-\frac{1}{2}\left(\frac{x-\mu}{b}\right)^2} dx$$

Ora ponendo $z = \frac{x-\mu}{b} \Rightarrow dx = b \cdot dz$, e quindi:

$$M_{X-\mu}(\theta) = b \cdot c \cdot \int_{-\infty}^{+\infty} e^{\theta b z - \frac{z^2}{2}} dz$$

Riscrivendo ora l'esponente sotto il segno di integrale nel modo seguente:

$$\theta b z - \frac{z^2}{2} = -\frac{1}{2}(z - \theta b)^2 + \frac{1}{2}\theta^2 b^2 \quad \text{si ha quindi:}$$

$$M_{X-\mu}(\theta) = b \cdot c \cdot e^{\frac{1}{2}\theta^2 b^2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(z-\theta b)^2} dz$$

dove ponendo $t = z - \theta b$ allora $dz = dt$, e quindi:

$$M_{X-\mu}(\theta) = b \cdot c \cdot e^{\frac{1}{2}\theta^2 b^2} \underbrace{\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt}_{\sqrt{2\pi}}$$

Ora, ricordando che l'integrale al secondo membro, come si può osservare nelle apposite tavole, è un'integrale noto, il cui valore è uguale a $\sqrt{2\pi}$, si ha che:

$$(2) \quad M_{X-\mu}(\theta) = \sqrt{2\pi} \cdot bc \cdot e^{\frac{1}{2}\theta^2 b^2}$$

Ora, ricordando che $M_X(\theta) = 1 + \frac{\theta}{1!}\mu'_1 + \frac{\theta^2}{2!}\mu'_2 + \dots$, segue che per qualsiasi funzione generatrice dei momenti si ha che:

$$M(0) = 1 \quad (0, \text{non importa che funzione } c' \text{ è qui})$$

Dalla (2), per $\theta = 0$ segue che $\sqrt{2\pi} \cdot bc = 1$ e quindi si può scrivere:

$$(3) \quad M_{X-\mu}(\theta) = e^{\frac{1}{2}\theta^2 b^2}$$

Se l'esponenziale al secondo membro si sviluppa in serie di potenze si ottiene:

$$M_{X-\mu}(\theta) = 1 + b^2 \frac{\theta^2}{2} + b^4 \frac{\theta^4}{8} + \dots$$

Esaminando lo sviluppo al secondo membro si può osservare che mancano le potenze dispari di θ e quindi che i momenti dispari della variabile X

dalla sua media μ devono essere nulli. Il coefficiente di $\frac{\theta^2}{2!}$ è il momento del secondo ordine della variabile X dalla sua media perciò $b^2 = \mu_2 = \sigma^2$ ovvero $b = \sigma$ ("sigma" è la varianza).

Ricordando che $\sqrt{2\pi} \cdot bc = 1$, si ricava:

$c = \frac{1}{\sigma\sqrt{2\pi}}$; di conseguenza la distribuzione normale, definita in generale

dalla (1), si può scrivere nella forma:

$$(4) \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (\text{non importa sapere la dimostrazione, basta sapere questa})$$

Questo risultato mostra che una distribuzione normale è completamente determinata specificando la sua media e il suo scarto quadratico medio. Si fa notare che la sola differenza tra la (1) e la (4) è che i parametri a , b e c nella (1) sono ora stati ristretti a due soli parametri indipendenti, cui è stato attribuito un preciso significato statistico.

Una formula per $M_X(\theta)$ espressa in termini di parametri statistici, sarà necessaria in seguito; essa si può ottenere dalla (3) sostituendo b^2 con σ^2 usando la seconda delle due proprietà di una funzione generatrice dei momenti vista precedentemente, e cioè:

$$M_{g(X)+c}(\theta) = e^{\theta c} \cdot M_{g(X)}(\theta) \quad \text{con } g(X) = X, c = -\mu$$

Così facendo si ha quindi che:

$$M_{X-\mu}(\theta) = e^{-\theta\mu} \cdot M_X(\theta)$$

Ora, sostituendo l'espressione data dalla (3) al primo membro di quest'ultima e risolvendo rispetto a $M_X(\theta)$ si ottiene che:

$$(5) \quad M_X(\theta) = e^{\mu\theta + \frac{1}{2}\sigma^2\theta^2}$$

Per interpretare lo scarto quadratico medio geometricamente, consideriamo i punti di flesso di una curva normale. Per fare questo occorre calcolare le prime due derivate della funzione di densità normale; si ha:

$$f' = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot \left[-\left(\frac{x-\mu}{\sigma}\right) \cdot \frac{1}{\sigma} \right] = -\frac{1}{\sigma^2}(x-\mu) \cdot f(x)$$

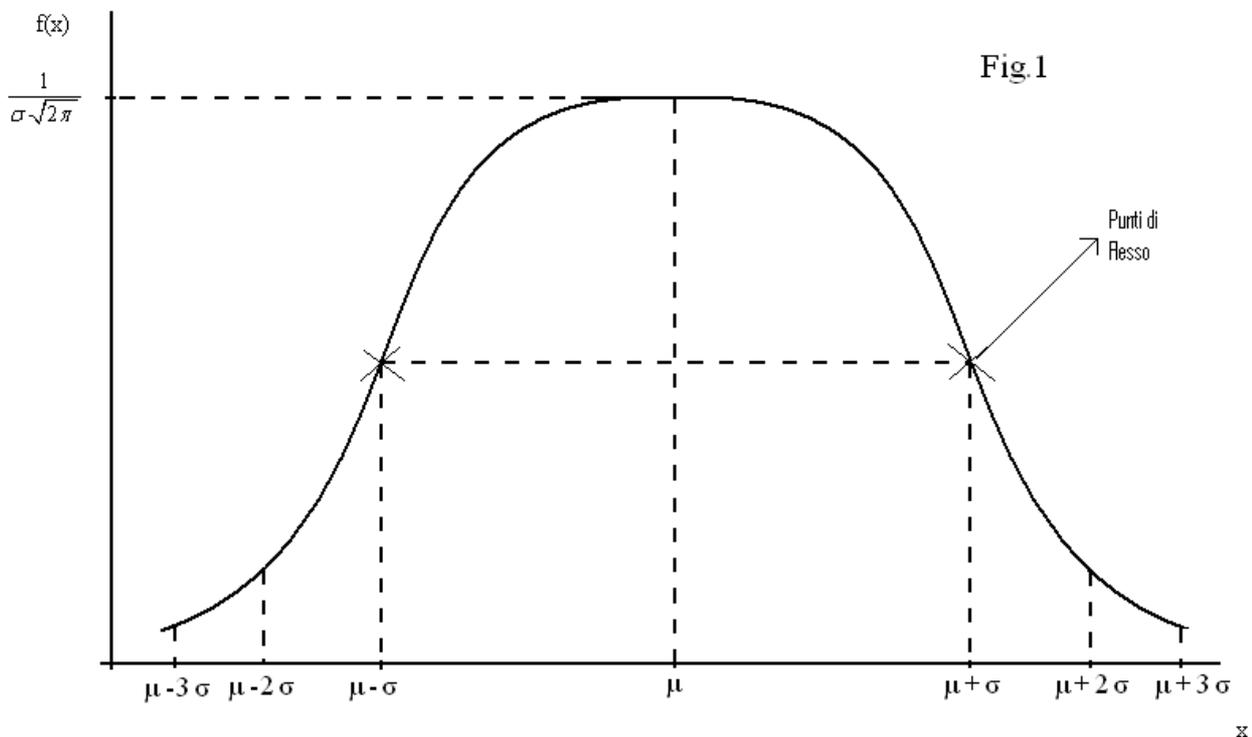
$$f'' = -\frac{1}{\sigma^2} f(x) - \frac{1}{\sigma^2}(x-\mu) \cdot \left[-\frac{1}{\sigma^2}(x-\mu) f(x) \right] = -\frac{1}{\sigma^2} \cdot f(x) \cdot \left[1 - \left(\frac{x-\mu}{\sigma}\right)^2 \right]$$

Da queste derivate è chiaro che c'è soltanto un punto di massimo che si trova in $x = \mu$ (cioè si annulla la derivata prima).

Dalla derivata seconda segue che i punti di flesso si trovano in $x = \mu \pm \sigma$ (punti flesso = punti in cui si annulla la derivata seconda).

Geometricamente, allora per una distribuzione normale, lo scarto quadratico medio è la distanza dei punti di flesso dall'asse di simmetria. Ciò implica che la curva normale rivolge la concavità verso il basso fra $\mu - \sigma$ e $\mu + \sigma$, e rivolge la concavità verso l'alto esternamente a questo intervallo.

Il grafico di una curva normale tipica è illustrato nella figura 1:



Un ulteriore approfondimento sul ruolo che ha il parametro σ nel determinare una funzione di densità normale si ottiene calcolando l'area sottesa dalla funzione di densità normale negli intervalli simmetrici mostrati in fig. 1. Così la probabilità che la variabile X avente una funzione di densità normale cada nell'intervallo $(\mu - \sigma, \mu + \sigma)$ è data da:

$$\int_{\mu - \sigma}^{\mu + \sigma} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2} dx \quad \text{ponendo } z = \frac{x - \mu}{\sigma} \Rightarrow dx = \sigma dz$$

L'integrale precedente si riduce a:

$$\int_{-1}^{+1} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz = 2 \int_0^1 \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz$$

Il valore dell'ultimo integrale che si può trovare nella apposite tavole è uguale a 0,3413; quindi moltiplicando per 2 il valore del primo integrale è uguale a 0,68, valore corretto a due cifre. Per l'intervallo $(\mu - 2\sigma, \mu + 2\sigma)$ si può verificare che l'area sottesa è uguale a 0,95. In modo analogo l'area sottesa fra $(\mu - 3\sigma, \mu + 3\sigma)$ è uguale a 0,997. Riassumendo circa il 68% della probabilità giace nell'intervallo $(\mu - \sigma, \mu + \sigma)$; circa il 95% nell'intervallo $(\mu - 2\sigma, \mu + 2\sigma)$ e quasi tutta la probabilità (99,7% circa) nell'intervallo $(\mu - 3\sigma, \mu + 3\sigma)$. L'unità di misura, data dalla trasformazione $z = \frac{x - \mu}{\sigma}$ si chiama **unità standard** e corrispondentemente, la distribuzione normale, espressa in funzione di z , cioè con media $\mu=0$ e scarto quadratico medio uguale a 1, prende il nome di **distribuzione normale in unità standard** o *in forma standard* o *standardizzata*, cioè:

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

La distribuzione normale con approssimazione della distribuzione binomiale: In precedenza la distribuzione di Poisson è stata introdotta come approssimazione della distribuzione binomiale quando n è grande e p è piccolo; allora si disse che un'altra distribuzione fornisce una buona approssimazione per n grande, a prescindere dal valore di p . La distribuzione normale gode di questa proprietà.

Prima di indagare sulla natura di questa approssimazione, consideriamo un esempio numerico in cui $n = 12$ e $p = 1/3$ e costruiamo il grafico della corrispondente distribuzione binomiale. Questo valore di n non è certamente grande,cosicché in questo caso, non ci si dovrebbe aspettare una buona approssimazione di tipo normale. In questo caso la funzione di densità binomiale è la seguente:

$$f(x) = \frac{12!}{x!(12-x)!} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{12-x}$$

Poiché la $f(x)$ si deve calcolare per tutti i valori di x , da $x=0$ a $x=12$, è più facile calcolare ciascun valore dopo il primo da quello precedente, anziché calcolarlo da solo.

Per questa funzione si verifica facilmente che :

$$f(x+1) = \frac{12-x}{x+1} \cdot \frac{1}{2} \cdot f(x)$$

Dopo aver calcolato $f(x)$ in $x=0$, quest'ultima relazione è stata usata per ottenere i restanti valori riportati nella tabella 1.

Tab. 1

| | | | |
|-----------------------------|---------|-------------------------------|---------|
| $f(0) =$ | .007707 | $f(7) = \frac{3}{7}f(6) =$ | .047687 |
| $f(1) = 6f(0) =$ | .046242 | $f(8) = \frac{5}{8}f(7) =$ | .014902 |
| $f(2) = \frac{12}{4}f(1) =$ | .127166 | $f(9) = \frac{2}{9}f(8) =$ | .003312 |
| $f(3) = \frac{5}{3}f(2) =$ | .211943 | $f(10) = \frac{3}{20}f(9) =$ | .000497 |
| $f(4) = \frac{8}{9}f(3) =$ | .238436 | $f(11) = \frac{1}{11}f(10) =$ | .000045 |
| $f(5) = \frac{4}{3}f(4) =$ | .190749 | $f(12) = \frac{1}{24}f(11) =$ | .000002 |
| $f(6) = \frac{7}{2}f(5) =$ | .111270 | | |

Il grafico di questa distribuzione binomiale è riportato nella figura 1.

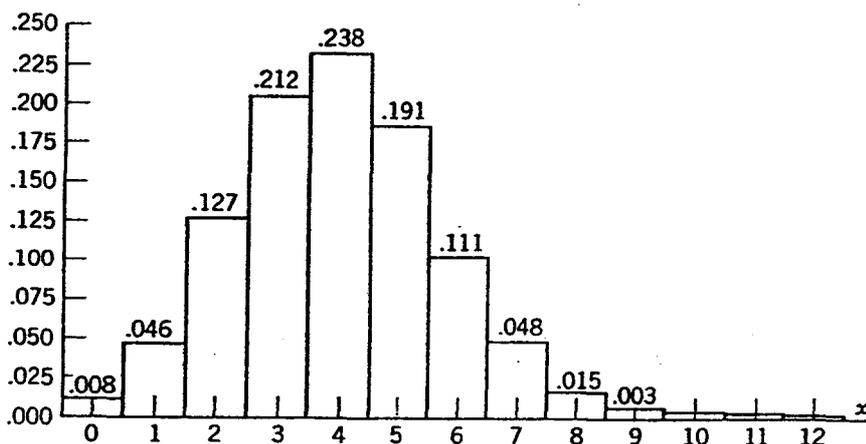


Fig. 1 . Distribuzione binomiale, $p = 1/3$, $n = 12$.

Sembra che questo istogramma potrebbe essere approssimato abbastanza bene dalla curva normale appropriata. Poiché una curva normale è completamente determinata dalla media e dallo scarto quadratico medio, la curva normale naturale da usare in questo caso è quella con la stessa media e lo stesso scarto quadratico medio della distribuzione binomiale.

Quindi scegliamo $\mu = np = 12 \cdot \frac{1}{3} = 4$ e $\sigma = \sqrt{npq} = \sqrt{12 \cdot \frac{1}{2} \cdot \frac{2}{3}} = 1.63$

Come prova, in questo caso, dell'accuratezza dell'approssimazione e come metodo che rappresenta un esempio d'uso della curva normale per approssimare le probabilità binomiali, consideriamo alcuni problemi riferiti alla figura 1.

Se la probabilità che un tiratore colpisca un bersaglio è uguale a $1/3$ e se egli tira 12 colpi, si vuole calcolare la probabilità che egli colpisca almeno 6 bersagli. La risposta esatta si ottiene sommando i valori della $f(x)$ da $x=6$ a $x=12$, che usando la tabella 1 sono uguali a 0,178, valore corretto a tre cifre decimali, cioè:

$$P\{X \geq 6\} = \sum_{x=6}^{12} f(x) = 0.178$$

Geometricamente questo valore rappresenta l'area di quella parte dell'istogramma in figura 1 che giace alla destra di $x = 5,5$. Perciò per approssimare questa probabilità, servendosi della curva normale, è necessario calcolare l'area sottesa da quella parte della curva normale approssimante che giace alla destra di $x=5,5$.

Poiché la curva approssimante ha $\mu = 4$ e $\sigma = 1,63$, con il cambiamento di variabile, segue che:

$$z = \frac{x - \mu}{\sigma} = \frac{5.5 - 4}{1.63} = 0.92$$

Quindi con questi valori di μ e σ , il cambiamento di variabile $z = \frac{x - \mu}{\sigma}$ fornisce che:

$$\int_{5.5}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_{0.92}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz = 0.179$$

Ma come si può osservare nelle apposite tabelle, l'area alla destra di $z=0,92$ è uguale a 0,179, che confrontata con il valore esatto 0,178 è sbagliato soltanto di circa lo 0,56%; infatti l'errore relativo in questo caso è dato da:

$$\frac{\Delta P}{P} = \frac{0.179 - 0.178}{0.178} = \frac{0.001}{0.178} = 0.0056 = 0.56\%$$

Per verificare l'accuratezza dei metodi che si basano sulla curva normale su di un intervallo più breve, calcoliamo la probabilità che il tiratore abbia successo precisamente in 6 colpi su 12. Dalla tabella 1, la risposta corretta in tre cifre decimali è $f(6) = 0,111$.

Per approssimare questo valore è necessario calcolare l'area sottesa dalla curva normale approssimante fra $x_1=5,5$ e $x_2=6,5$. In questo caso si ha che:

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{5.5 - 4}{1.63} = 0.92 \quad z_2 = \frac{x_2 - \mu}{\sigma} = \frac{6.5 - 4}{1.63} = 1.53$$

Dalle apposite tabelle si può osservare che l'area sottesa dalla curva normale standard fra $z=0$ e $z_1=0,92$ è $A_1 = 0,3212$ e l'area sottesa fra $z=0$ e $z_2=1,53$ è $A_2 = 0,4363$; per cui l'area richiesta è uguale a $A_2 - A_1 = 0,116$, che è sbagliata di circa il 4,5%, infatti in questo caso l'errore relativo è:

$$\frac{\Delta P}{P} = \frac{0.116 - 0.111}{0.111} = \frac{0.005}{0.111} = 0.045 = 4.5\%$$

Da questi due esempi sembra che i metodi che si basano sulla curva normale siano abbastanza accurati anche per alcune situazioni come quella considerata qui, in cui n non è molto grande. Gli esempi precedenti sono stati dati per rendere credibile un famoso teorema che garantisce una buona approssimazione della curva normale alla distribuzione binomiale se n è sufficientemente grande. Questo teorema che enunceremo senza dimostrazione è il seguente:

Teorema: Se in n esecuzioni indipendenti di un esperimento la variabile X rappresenta il numero di successi di un evento per cui p è la probabilità di successo in una singola esecuzione, allora la variabile $\frac{X - np}{\sqrt{npq}}$ ha una distribuzione che tende alla distribuzione normale con media uguale a 0 e scarto quadratico medio uguale a 1 quando $n \rightarrow \infty$.

($f(z)$ normale standard; $f(x)$ normale; $\mu = np$; $\sigma = \sqrt{npq}$)

Questo teorema giustifica l'uso dei metodi che si basano sulla curva normale per approssimare le probabilità di un evento relative ad esecuzioni successive di un esperimento quando n è grande. L'esperienza indica che l'approssimazione è abbastanza buona purchè $np > 5$ quando $p \leq 1/2$ e purchè $nq > 5$ quando $p > 1/2$.

Le figure 2 e 3 indicano come rapidamente tenda alla normalità la distribuzione della variabile $\frac{X - np}{\sqrt{npq}}$ quando $p = 1/3$ e rispettivamente $n = 24$ e $n = 48$. La scala dell'asse y per questi due grafici è approssimativamente 17 volte quella dell'asse x .

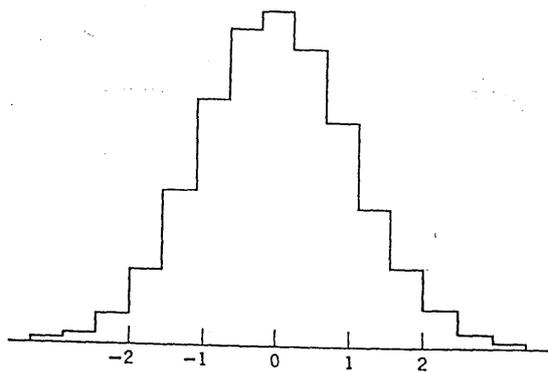


Fig. 2 . Distribuzione di $(X-np)/\sqrt{npq}$ per $p = 1/3$ e $n = 24$.

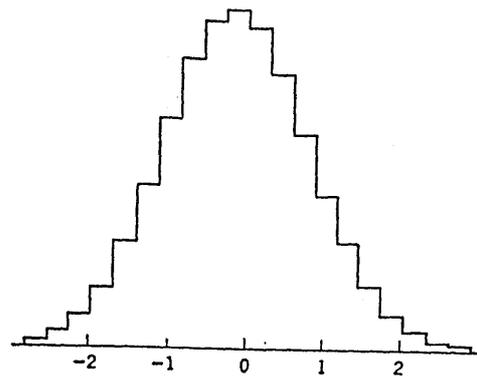


Fig. 3 . Distribuzione di $(X-np)/\sqrt{npq}$ per $p = 1/3$ e $n = 48$.

Le due approssimazioni della distribuzione binomiale, cioè quella di Poisson e quella normale, sono sufficienti per permettere di risolvere tutti i problemi più semplici che richiedono il calcolo di probabilità binomiali; se invece n è piccolo si usa direttamente la funzione di densità binomiale perché i calcoli allora sono del tutto semplici.

Esercitazione: Come esempio d'uso della distribuzione normale come approssimazione di quella binomiale, consideriamo il problema seguente: un fabbricante di parti di macchina sostiene che il 10% al massimo delle parti da lui prodotte sono difettose; un compratore ha bisogno di 120 di queste parti e per essere sicuro di acquistarle senza difetti egli ne ordina 140. Se l'affermazione del fabbricante è valida, si vuole calcolare la probabilità che il compratore riceva almeno 120 parti buone.

Indichiamo con la variabile X il numero di parti buone ricevute. Allora X si può trattare come una variabile binomiale con $n=140$ e $p=0.9$.

In questo caso, questi valori giustificano certamente un'approssimazione mediante la funzione di densità normale, in quanto:

$$p=0.9 > 1/2 \quad nq=n(1-p)=140 \cdot 0.1=14 > 5$$

Il problema è quindi di calcolare $P\{X \geq 120\}$.

In questo caso la media e lo scarto quadratico medio sono:

$$\mu=np=140 \cdot 0.9=126 \quad \sigma = \sqrt{npq} = \sqrt{140 \cdot 0.9 \cdot 0.1} = 3.55$$

Quindi (ricordando che $z = \frac{x-\mu}{\sigma}$):

$$\begin{aligned} P\{X \geq 120\} &= P\left\{\frac{X-\mu}{\sigma} \geq \frac{120-\mu}{\sigma}\right\} = P\left\{\frac{X-126}{3.55} \geq \frac{120-126}{3.55}\right\} = \\ &= P\{Z \geq -1.69\} = 0.95 \end{aligned}$$

Poiché $Z = \frac{X - \mu}{\sigma} = \frac{X - 126}{3.55}$ si può trattare come una variabile normale standard approssimata, questa probabilità, che si può trovare nelle apposite tavole, è uguale a 0.95; perciò se l'affermazione del fabbricante è valida, il compratore ha il 95% di probabilità di ottenere almeno 120 parti buone.

Distribuzione gamma: Una distribuzione di probabilità che si presenta spesso in vari problemi statistici, come ad esempio nello studio della durata di un'apparecchiatura industriale, è la *distribuzione gamma*. Il nome deriva dalla relazione della distribuzione con la funzione gamma del calcolo. Essa comprende due parametri, $\alpha > 0$ e $\beta > 0$, ed è definita come segue:

$$(1) \quad f(x) = \begin{cases} \frac{x^{\alpha-1} \cdot e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases} \quad (\Gamma \text{ è } \gamma \text{ maiuscolo, } \gamma \text{ minuscolo})$$

La quantità $\Gamma(\alpha)$ rappresenta il valore della funzione gamma nel punto α . Questa funzione è definita come segue dalla (2):

$$(2) \quad \Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} \cdot e^{-x} dx$$

Si può dimostrare, integrando per parti, che $\Gamma(\alpha + 1) = \alpha \cdot \Gamma(\alpha)$; se α è un intero positivo, questa relazione di ricorrenza fornisce il risultato che $\Gamma(\alpha + 1) = \alpha!$. Come conseguenza di questa proprietà, la funzione Γ è chiamata a volte *funzione fattoriale*.

Momenti: I momenti della distribuzione gamma si calcolano servendosi della (1).

Dalla definizione si ha infatti che:

$$\mu'_k = E[X^k] = \frac{1}{\beta^\alpha \cdot \Gamma(\alpha)} \cdot \int_0^{+\infty} x^{k+\alpha-1} \cdot e^{-\frac{x}{\beta}} dx$$

ponendo $t = \frac{x}{\beta}$, allora $dx = \beta dt$ e ottengo:

$$\mu'_k = \frac{\beta^{k+\alpha-1} \cdot \beta}{\beta^\alpha \cdot \Gamma(\alpha)} \cdot \int_0^{+\infty} \overbrace{t^{k+\alpha-1} \cdot e^{-t} dt}^{\Gamma(k+\alpha)}$$

Dalla (2) si può osservare che l'integrale al secondo membro rappresenta $\Gamma(k + \alpha)$ e quindi si può scrivere:

$$\mu'_k = \frac{\beta^k \Gamma(k + \alpha)}{\Gamma(\alpha)}$$

Ora, poiché k è un intero positivo, si ha (ricordando che $\Gamma(\alpha+1) = \alpha \Gamma(\alpha)$):
 $\Gamma(k + \alpha) = (k + \alpha - 1) \cdot (k + \alpha - 2) \cdot (\dots) \cdot \dots \cdot \alpha \Gamma(\alpha)$

Quindi si ha che:

$$\mu'_k = \beta^k (k + \alpha - 1) \cdot (k + \alpha - 2) \cdot (\dots) \cdot \dots \cdot \alpha$$

Perciò per $k=1$, la media è data dalla:

$$(3) \quad \mu = \beta \alpha$$

Poi, tenuto conto che $\mu'_2 = \beta^2 (1 + \alpha) \cdot \alpha$, usando la formula nota $\sigma^2 = \mu'_2 - \mu^2$, la varianza è data dalla:

$$(4) \quad \sigma^2 = \beta^2 (1 + \alpha) \cdot \alpha - \beta^2 \alpha^2 = \beta^2 \alpha$$

Distribuzione esponenziale: Il caso particolare della distribuzione gamma, che si verifica per $\alpha = 1$, viene chiamato *distribuzione esponenziale*. Essa viene usata sufficientemente spesso da giustificare di trattarla separatamente.

Questa distribuzione è rappresentata quindi dalla: (5)

$$f(x) = \begin{cases} \frac{e^{-\frac{x}{\beta}}}{\beta} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases}$$

Tenuto conto che $\Gamma(1)=1$, infatti:

$$\Gamma(1) = \int_0^{+\infty} e^{-x} dx = -e^{-x} \Big|_0^{+\infty} = 1$$

Dalle formule della media e della varianza di una distribuzione gamma, si ricava, per $\alpha = 1$, che la media e la varianza di una distribuzione

esponenziale sono $\mu=\beta$ e $\sigma^2=\beta^2$. La distribuzione esponenziale si è trovata ad esempio che è un modello adatto per calcolare la probabilità che un'apparecchiatura duri t unità di tempo prima che essa si guasti. L'adeguatezza del modello dipende dalla natura dell'apparecchiatura e da ciò che ne causa il decadimento. Se il guasto è dovuto principalmente a cause esterne piuttosto che al logorio interno, allora è verosimile che il modello sia realistico. In queste circostanze il tempo che passa fino al guasto successivo, dopo che l'apparecchiatura è stata riparata e rimessa in funzionamento, seguirà anch'esso una distribuzione esponenziale. Così dopo ciascuna riparazione, l'apparecchiatura si comporta come se essa fosse nuova.

Esempio: Come esempio di impiego pratico della distribuzione esponenziale consideriamo il problema seguente: un costruttore di un'apparecchiatura elettronica ha trovato per esperienza che la sua apparecchiatura dura in media due anni senza riparazioni e che il tempo che passa prima che si verifichi il primo guasto segue una distribuzione esponenziale. Se egli garantisce che la sua apparecchiatura duri un anno, si vuol calcolare la probabilità che l'apparecchiatura abbia un guasto prima che scada la garanzia, ovvero, quale percentuale dei suoi clienti sarà interessata a qualche riparazione per un guasto prima di un anno. Poiché β è la media della distribuzione data dalla (5) e in questo caso $\beta=2$, la

funzione di densità che si applica qui è $f(x) = \frac{1}{2} e^{-\frac{x}{2}}$.

Il problema è quindi quello di calcolare $P(X < 1)$ perciò:

$P\{X < 1\} = \int_0^1 \frac{1}{2} e^{-\frac{x}{2}} dx$, ponendo $t = \frac{x}{2}$, per cui $dx = 2dt$, si ha che:

$$P\{X < 1\} = \int_0^{\frac{1}{2}} e^{-t} dt = -e^{-t} \Big|_0^{\frac{1}{2}} = -e^{-\frac{1}{2}} + 1 = 0.39$$

Si conclude quindi che anche se la durata media dell'apparecchiatura è il doppio della durata garantita, c'è una probabilità abbastanza elevata, cioè il 39%, che l'apparecchiatura abbia un guasto prima che scada la garanzia.

Distribuzione chi-quadrato(χ^2): Un altro caso particolare della distribuzione Γ che ha molte applicazioni statistiche, si ricava dalla (1)

prendendo $\beta = 2$ e scrivendo $\alpha = \nu/2$ ("ni"). Esso prende il nome di distribuzione chi-quadrata la cui funzione di densità è quindi la seguente:

$$(6) \quad f(x) = \begin{cases} \frac{x^{\frac{\nu}{2}-1} \cdot e^{-\frac{x}{2}}}{2^{\frac{\nu}{2}} \cdot \Gamma\left(\frac{\nu}{2}\right)} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases}$$

Il motivo della sostituzione del parametro α con $\nu/2$ è che il parametro ν viene normalmente usato per rappresentare i gradi di libertà e quindi possiede un significato intuitivo naturale, quando la distribuzione χ^2 si applica a certi specifici problemi statistici. E' interessante a questo proposito calcolare la media e la varianza di una variabile χ^2 in termini di questo nuovo parametro, ponendo $\beta = 2$ e $\alpha = \nu/2$, nelle formule (3) e (4) si ottiene $\mu = \nu$, $\sigma^2 = 2\nu$. Poiché la distribuzione gamma dipende da due parametri, essa possiede moltissima flessibilità come modello per le distribuzioni reali. Per mostrare questa flessibilità, nella figura (1) sono mostrati i grafici di parecchie funzioni di densità χ^2 corrispondenti, quindi, come si è già detto, al valore di $\beta=2$ nella funzione di densità gamma. Facendo notare che, poiché β funge da parametro di scala, cambiando il valore di β la corrispondente curva di Fig. 1 semplicemente si stenderà se β aumenta o si comprimerà se β diminuisce, mantenendo però, naturalmente uguale a 1 l'area da essa sottesa.

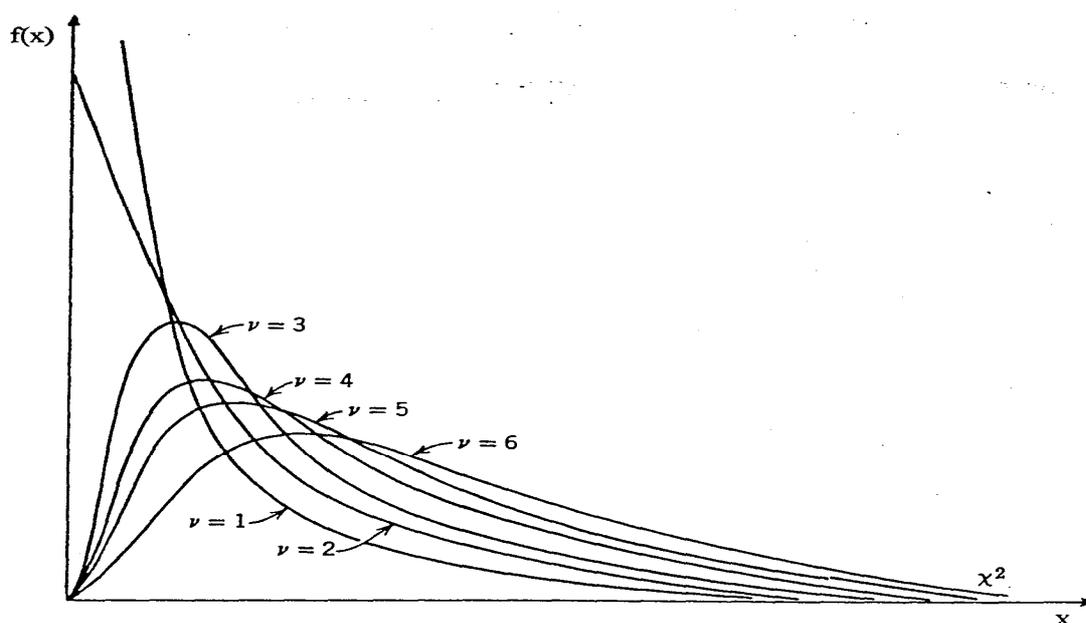


Fig. 1 . Distribuzione chi-quadrato per vari gradi di libertà.

Inferenza statistica: Finora ci siamo occupati di sviluppare i principi fondamentali della *Probabilità* e di presentare alcune distribuzioni di probabilità che si sono dimostrate particolarmente utili nel risolvere certe classi di problemi reali. Il motivo è stato quello di costruire modelli per esperimenti di tipo ripetitivo. Il vantaggio di tali modelli è che ci permettono di studiare le proprietà dell'esperimento e di fare previsioni sui risultati relativi a future esecuzioni dell'esperimento, cose che sarebbero difficili o impossibili da fare senza l'ausilio di un modello. Il processo di costruire un modello sulla base di dati sperimentali e di trarre conclusioni da esso è un esempio di inferenza induttiva. Quando esso si applica a problemi statistici viene chiamato di solito inferenza statistica. Gli statistici si occupano principalmente di fare delle inferenze statistiche servendosi di dati sperimentali; molto spesso, lo statistico è interessato a costruire un modello matematico per una sola variabile associata ad un esperimento, piuttosto che per l'intero esperimento. Come conseguenza la maggior parte dei modelli scelti dagli statistici sono funzioni di densità di variabili casuali. Le inferenze statistiche sono perciò di solito inferenze riguardanti le funzioni di densità.

Esempio: come esempio di quanto detto supponiamo che un biologo abbia osservato che su 200 insetti di una data specie ve ne sono 44 che possiedono macchie diverse da quelle del restante insieme. Supponiamo inoltre che il biologo sospetti che le macchie siano ereditate secondo una legge che prevede che il 25% di tali insetti possiedono le macchie meno comuni. Se egli fa l'ipotesi che in questo caso valga la legge dell'eredità e rappresenta con la variabile X il numero degli insetti, dei 200 che possiedono le macchie meno comuni, allora il modello che egli naturalmente sceglierebbe è la funzione di densità binomiale :

$$(1) f(x) = \frac{200!}{x!(200-x)!} \cdot \left(\frac{1}{4}\right)^x \cdot \left(\frac{3}{4}\right)^{200-x}$$

Se non ci fosse stata alcuna teoria a suggerire che 1/4 degli insetti dovrebbero possedere le macchie meno comuni, il biologo potrebbe scegliere questa stessa funzione di densità con la probabilità 1/4 sostituita dalla frequenza relativa osservata, pari a $44/200 = 0,22$. Utilizzando la (1) il biologo potrebbe poi fare delle previsioni riguardanti i futuri insiemi di 200 osservazioni e scoprire eventuali disaccordi con la sua teoria.

Dati: E' conveniente nella trattazione statistica chiamare la totalità dei possibili risultati sperimentali come la popolazione di tali risultati. Allora un insieme di dati ottenuti eseguendo l'esperimento un certo numero di volte è chiamato un campione della popolazione. Secondo questa terminologia l'inferenza statistica consiste allora nel trarre delle conclusioni su di una popolazione, servendosi di un campione estratto dalla popolazione stessa. Un problema fondamentale perciò è come estrarre informazioni dei campioni per utilizzarle nello studio delle popolazioni da cui i campioni sono stati estratti. Il tipo di informazione che si dovrebbe estrarre di un insieme di dati dipende dalla natura dei dati e dal modello scelto. In alcuni problemi si sa da considerazioni teoriche o dall'esperienza con problemi simili quale modello si dovrebbe usare; per esempio, la densità binomiale rappresentata dalla (1) è un tale modello. Tutto ciò che in realtà è necessario dai dati sperimentali per tali modelli è l'informazione atta a fornire buone stime dei parametri implicati. In altri problemi né teoria né esperienza è disponibile per aiutare a scegliere un modello; allora è necessario usare dati sperimentali per decidere su di un ragionevole tipo di modello prima di poter procedere ulteriormente.

Nel considerare la natura dei dati è importante distinguere fra quegli insiemi di dati per cui l'ordine in cui le osservazioni sono state ottenute fornisce un'informazione utile a quegli insiemi per cui non è così.

Per esempio se si fosse interessati a studiare i fenomeni atmosferici o la borsa valori di giorno in giorno, l'ordine sarebbe molto importante.

L'esperienza industriale indica che l'informazione ottenuta nel considerare l'ordine in cui sono fabbricati gli articoli è indispensabile per un'efficiente produzione. Se si fosse invece interessati a studiare le caratteristiche degli studenti di un istituto e si scegliesse un nominativo ogni 20 in una guida dell'istituto, difficilmente ci si aspetterebbe che l'ordine in cui vengono presi i nomi fosse di qualche valore nello studio.

Ci occuperemo ora di tecniche che non usano informazioni riguardanti l'ordine e cominciamo ad occuparci della *classificazione dei dati*.

Supponiamo che siano dati i pesi di 200 individui di un istituto e si desideri usarli per studiare la distribuzione del peso di quegli individui. Ora, è molto difficile guardare 200 misure ed ottenere contemporaneamente un'idea ragionevolmente accurata di come si distribuiscono quelle misure. Per ottenere un'idea migliore della distribuzione dei pesi è perciò conveniente condensare i dati classificando

le misure in gruppi. Sarà allora possibile tracciare il grafico della distribuzione così modificata e imparare di più su come sono distribuiti i pesi. Questa classificazione sarà utile anche per semplificare i calcoli di alcune medie, in particolare quando non si dispone di mezzi di calcolo veloci. Queste medie forniscono ulteriori informazioni sulla distribuzione, così lo scopo di classificare i dati è quello di aiutare ad estrarre certi tipi di utili informazioni riguardanti la distribuzione sottesa.

Nel classificare i dati è di solito conveniente usare da 10 a 20 classi, ma se necessario si può scendere anche fino a 6 classi.

La tabella 1 mostra come è stato classificato un insieme di misure dei 200 diametri di barrette di acciaio, i cui valori variano fra 0,431 e 0,503 pollici.

Tab.1

| Estremi di classe | Frequenze | Marchi di classe: x | Frequenze: f |
|-------------------|--------------------------|---------------------|--------------|
| .4305-.4355 | // | .433 | 2 |
| .4355-.4405 | /// | .438 | 5 |
| .4405-.4455 | /// // | .443 | 7 |
| .4455-.4505 | /// /// | .448 | 13 |
| .4505-.4555 | /// /// //// | .453 | 19 |
| .4555-.4605 | /// /// /// /// // | .458 | 27 |
| .4605-.4655 | /// /// /// /// /// //// | .463 | 29 |
| .4655-.4705 | /// /// /// /// /// | .468 | 25 |
| .4705-.4755 | /// /// /// /// /// | .473 | 23 |
| .4755-.4805 | /// /// //// | .478 | 14 |
| .4805-.4855 | /// /// /// | .483 | 15 |
| .4855-.4905 | /// //// | .488 | 9 |
| .4905-.4955 | /// | .493 | 6 |
| .4955-.5005 | //// | .498 | 4 |
| .5005-.5055 | // | .503 | 2 |

Poiché i diametri sono stati misurati al millesimo di pollice, gli estremi degli intervalli di classe sono stati scelti mezza unità per l'accuratezza di questa misura per assicurare che nessuna misura cada su di un estremo. Si fa l'ipotesi in questa classificazione che a tutte le misure che cadono in un dato intervallo di classe venga assegnato il valore corrispondente al punto di mezzo dell'intervallo, che prende il nome di *marchio* o *centro di classe*. Dopo che ciascuna misura è stata registrata nella classe appropriata per mezzo di una sbarretta, come mostrato in Tabella 1, i risultati della classificazione sono registrati sotto forma di una tabella di frequenza, come mostrato nella seconda metà della Tabella 1.

Rappresentazione grafica di distribuzioni empiriche (o osservate):
Un'idea approssimativa di come sono distribuiti i valori di una variabile

casuale si può ottenere esaminando il corrispondente istogramma. L'istogramma per i dati di Tabella 1 è mostrato in Fig. 2.

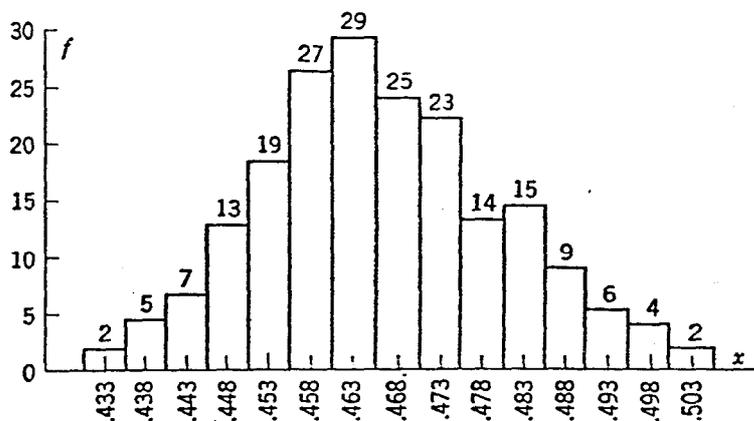


Fig.2. Distribuzione dei diametri di 200 barrette di acciaio.

Si può osservare che i marchi di classe sono nei punti di mezzo delle basi dei rettangoli che compongono l'istogramma. Fortunatamente molte importanti distribuzioni che si presentano in natura e nell'industria hanno forme relativamente semplici. Queste forme di solito variano da quella a campana come in Fig. 2, o a quella che assomiglia alla metà di una campana. Una distribuzione dell'ultimo tipo si dice che è asimmetrica, il che significa che manca di simmetria rispetto all'asse verticale. Risulta che variabili con distribuzioni aventi gradi di asimmetria crescenti sono ad esempio la statura, il peso, l'età di matrimonio, l'età di mortalità per certe malattie e la ricchezza.

Le figure 2, 3 e 4 rappresentano tre tipiche distribuzioni con gradi crescenti di asimmetria.

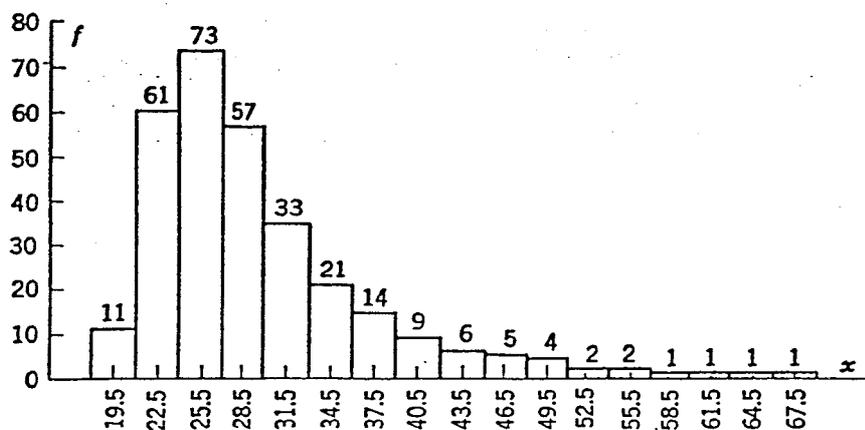


Fig.3. Distribuzione di 302000 matrimoni classificati secondo l'età dello sposo. L'unità di frequenza è 1000.

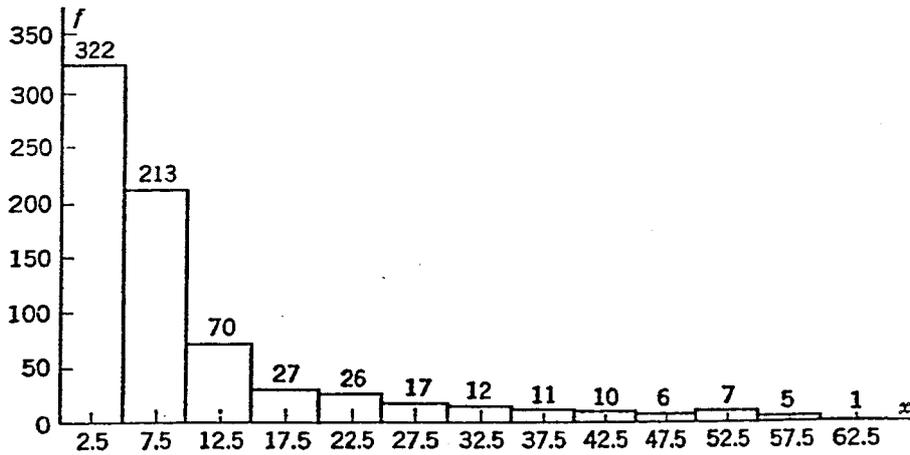


Fig.4. Distribuzione di 727 morti per scarlattina classificati secondo l'età.

Si può osservare che la distribuzione della figura 2 ha la forma di una curva normale e che perciò la distribuzione normale potrebbe servire da modello soddisfacente per la distribuzione delle misure dei diametri delle barrette d'acciaio. Molte misure industriali lineari possiedono distribuzioni di questo tipo e vengono trattate con successo usando come modelli le curve normali. La distribuzione mostrata nella fig. 4 ha una forma che suggerisce la possibilità di usare come modello una distribuzione esponenziale, mentre la fig. 3 richiede un modello più sofisticato. Poiché una distribuzione γ ("gamma"), con i suoi due parametri possiede moltissima flessibilità, essa potrebbe servire da modello per la distribuzione della fig. 3.

Momenti empirici: Sebbene un istogramma come quelli mostrati nelle fig. 2, 3 e 4 viste prima fornisca una notevole quantità di informazioni generali che riguardano la distribuzione di un insieme di misure campionarie, informazioni più precise e utili per studiare una distribuzione si possono ottenere da una descrizione aritmetica della distribuzione. Per esempio se fosse disponibile l'istogramma dei pesi di un campione di 200 individui di un istituto, per confrontarlo con l'istogramma di un campione simile di un altro istituto, potrebbe essere difficile stabilire, eccetto in termini molto generali, fino a che punto le due distribuzioni differiscono; piuttosto che confrontare le due distribuzioni dei pesi in questo modo potrebbe bastare confrontare i pesi medi e la variazione dei pesi dei due gruppi. La natura di un problema statistico determina largamente se alcune semplici proprietà aritmetiche della distribuzione saranno sufficienti a descriverla in modo soddisfacente. Spesso la maggior parte dei problemi

richiederanno per la loro soluzione solo alcune proprietà fondamentali della distribuzione. Per semplici distribuzioni di frequenza come quelle di cui i grafici sono mostrati nelle fig. 2, 3 e 4, questa descrizione si compie in modo soddisfacente per mezzo dei momenti di ordine basso della distribuzione. In molti problemi lo statistico è interessato solo ai momenti del I e del II ordine, in altri problemi egli usa i primi quattro momenti, ma raramente ne usa più di quattro. Una ragione di ciò è che i momenti di ordine più elevato sono così instabili in ripetuti esperimenti di campionamento che da loro si possono ottenere poche ulteriori informazioni attendibili. Indichiamo con x_1, x_2, \dots, x_n i valori osservati di un campione di dimensione n della variabile casuale X . Allora per analogie con i momenti teorici, i momenti empirici dall'origine vengono definiti come segue:

Definizione: il momento di ordine k dall'origine di una distribuzione empirica è dato da:

$$(1) m'_k = \frac{1}{n} \cdot \sum_{i=1}^n x_i^k$$

I momenti empirici sono pure chiamati *momenti del campione* o *momenti campionari* perché essi si basano su valori campionari. Il momento del primo ordine dall'origine m'_1 si indica tradizionalmente col simbolo \bar{x} ; esso dà il centro di gravità di una distribuzione empirica, proprio come fa la media μ per una distribuzione teorica, e serve per misurare dove è centrata la distribuzione empirica. Essa è chiamata *media del campione*.

Per analogia con la definizione data per le distribuzioni di probabilità, i momenti empirici dalla media sono definiti come segue:

Definizione: il momento di ordine k dalla media di una distribuzione empirica è dato da:

$$(2) m_k = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^k$$

Poiché il momento del secondo ordine dalla media viene usato molto spesso, si assegna ad esso il simbolo particolare s^2 , ed è chiamato *VARIANZA del campione*.

Corrispondentemente s è chiamato *scarto quadratico medio* o *deviazione standard* del campione. Per calcolare s^2 è spesso conveniente usare la formula seguente che è l'analogia di una formula già vista per le distribuzioni di probabilità e precisamente:

$$s^2 = m_2' - \bar{x}^2 \quad (\text{ricorda che } \sigma^2 = \mu_2' - \mu^2)$$

Se i valori di osservazione x_1, x_2, \dots, x_n sono stati classificati in una tabella di frequenza con x_i che rappresenta l'i-esimo marchio di classe, f_i che rappresenta il numero di osservazioni nell'i-esimo intervallo, cioè la frequenza assoluta di x_i e N_c che indica il numero degli intervalli di classe, allora le definizioni dei momenti assumeranno le forme seguenti:

$$(3) \quad m_k' = \frac{1}{n} \cdot \sum_{i=1}^{N_c} x_i^k \cdot f_i$$

$$(4) \quad m_k = \frac{1}{n} \cdot \sum_{i=1}^{N_c} (x_i - \bar{x})^k \cdot f_i$$

Il valore di \bar{x} nella (4) si assume essere il valore ottenuto dalla (3) e non dalla (1). A rigor di termini le formule (3) e (4) definiscono i momenti soltanto per le distribuzioni classificate e sono soltanto approssimazioni dei valori dati dalle formule (1) e (2), ma le approssimazioni sono di solito così buone che in pratica non si fa nessuna distinzione fra i due tipi di formule. Per esempio \bar{x} ed s^2 sono chiamate rispettivamente la media e la varianza del campione sia che essi siano state ottenute dalle formule (1) e (2) che dalle formule (3) e (4). Non c'è nessuna ragione di classificare i dati se si desiderano soltanto i momenti di una distribuzione empirica. La classificazione ha lo scopo di osservare geometricamente la natura della distribuzione empirica. Se i dati sono già stati classificati per questo scopo e se si desidera ad esempio il valore di \bar{x} ed s^2 , allora può essere più facile calcolarli con le formule (3) e (4) piuttosto che con le formule (1) e (2); come si è già detto le differenze sono di solito trascurabili.

Si fa osservare che in molti testi invece di f_i si usa di solito $\varphi(x_i)$, che rappresenta il valore della funzione di frequenza assoluta in x_i .

Questa funzione viene indicata con $\varphi(x)$ e non con $f(x)$ perché quest'ultima si usa in genere per rappresentare la funzione di densità di probabilità.

Esempio: per acquistare familiarità con lo scarto quadratico medio come misura della concentrazione di una distribuzione attorno alla sua media, consideriamo la seguente distribuzione empirica di 1000 conversazioni telefoniche misurate in secondi mostrate in tabella 1.

Tab. 1

| | | | | | | | | | | |
|-------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| x_i | 49.5 | 149.5 | 249.5 | 349.5 | 449.5 | 549.5 | 649.5 | 749.5 | 849.5 | 949.5 |
| f_i | 6 | 28 | 88 | 180 | 247 | 260 | 133 | 42 | 11 | 5 |

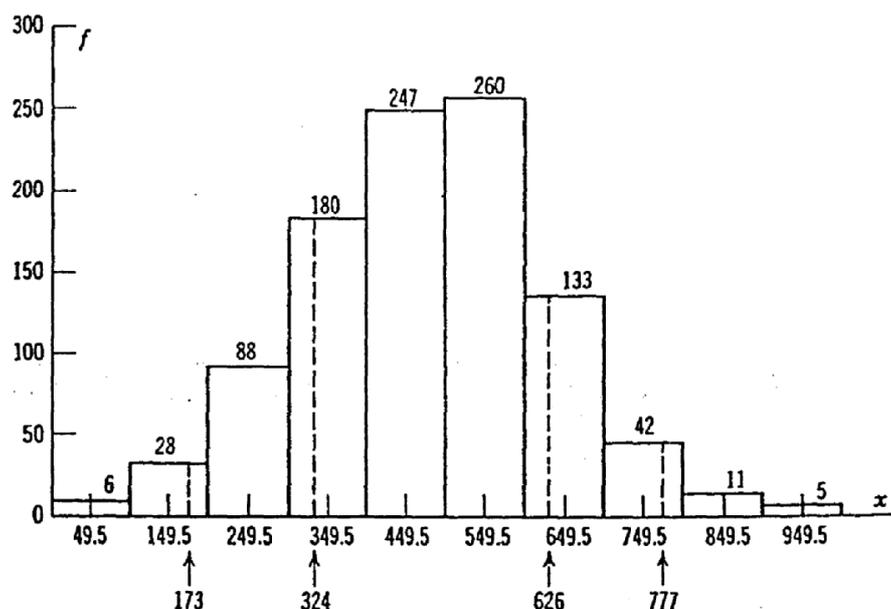


Fig. 1. Istogramma per la distribuzione di 1000 conversazioni telefoniche.

L'istogramma per questa distribuzione è mostrato in figura 1.

Dalla figura 1 appare che per la durata della conversazioni telefoniche potrebbe essere adatto un modello di distribuzione normale. In questo caso, poiché gli intervalli $(\mu - \sigma, \mu + \sigma)$ e $(\mu - 2\sigma, \mu + 2\sigma)$ contengono rispettivamente il 68% e il 95% dell'area sottesa da una curva normale, gli intervalli $(\bar{x} - s, \bar{x} + s)$ e $(\bar{x} - 2s, \bar{x} + 2s)$ ci si dovrebbe aspettare che forniscano approssimativamente le stesse percentuali per quanto riguarda la distribuzione empirica. I calcoli per i dati di tabella 1, usando le formule (3) e (4), danno i valori $\bar{x} = 475$ e $s = 151$, valori corretti all'intero più vicino. Di conseguenza gli intervalli $(\bar{x} - s, \bar{x} + s)$ e $(\bar{x} - 2s, \bar{x} + 2s)$ diventano gli intervalli (324, 626) e (173, 777). Gli estremi di questi intervalli sono indicati in figura 1 per mezzo di frecce verticali. Il numero di osservazioni che giacciono dentro questi intervalli si possono trovare approssimativamente mediante interpolazione, come se le osservazioni in un dato intervallo fossero sparse uniformemente nell'intervallo. Questa ipotesi implica che sull'istogramma qualsiasi parte

frazionaria di un intervallo di classe comprenderà la stessa parte frazionaria delle frequenze in quell'intervallo. Se l'interpolazione viene condotta fino all'unità più vicina, si troverà che l'intervallo (324, 626) comprende $(136 + 247 + 260 + 35) = 678$ misure che è il 67,8% delle misure effettuate. L'intervallo (173, 777) esclude $(6+21 + 9 + 11 + 5) = 52$ misure che è il 5,2% e quindi comprende il 94,8% delle misure effettuate. Per un istogramma così irregolare come questo questi risultati sono eccezionalmente vicini alle percentuali teoriche del 68% e del 95% per una distribuzione normale, confermando così che un modello di distribuzione normale rappresenta una buona approssimazione per la durata delle conversazioni telefoniche. $(180:100=x:(399.5-324) \rightarrow x=136)$

INFERENZA STATISTICA: Ci occuperemo ora di quel tipo di inferenza statistica che implica la stima di parametri della funzione di densità di probabilità che è stata scelta come modello per una variabile casuale.

Stima: in statistica la maggior parte dei problemi di stima riguardano la stima dei parametri di una funzione di densità di probabilità, per esempio una compagnia telefonica interessata a studiare problemi connessi con la durata delle conversazioni telefoniche, potrebbe voler stimare i parametri μ e σ della funzione di densità normale che si può assumere come modello adatto per le conversazioni telefoniche. Due tipi di stime di parametri sono di uso corrente, una è chiamata stima puntuale e l'altra stima per intervallo. Una *stima puntuale* è il tipo di stima più comune, cioè essa è un numero ottenuto da calcoli fatti sui valori osservati della variabile casuale che serve come un'approssimazione del valore vero del parametro.

Una *stima per intervallo* è un intervallo determinato da due numeri ottenuti da calcoli fatti sui valori osservati dalla variabile casuale, che ci si aspetta contengano al loro interno il valore vero del loro parametro. Consideriamo ora le stime puntuali. Il metodo di stima che illustriamo ora è conosciuto come il *metodo dei momenti*. Se un parametro di funzione di densità, come ad esempio il parametro μ per la densità di Poisson o il parametro β per la densità esponenziale è un momento della distribuzione, allora la sua stima sarà il corrispondente momento del campione. Poiché μ e β sono le medie delle loro rispettive densità, essi sarebbero entrambi stimati dalla media del campione \bar{x} . I due parametri μ e σ^2 di una funzione di densità normale sono momenti della distribuzione, perciò essi

saranno stimati dalla media del campione \bar{x} e dalla varianza del campione s^2 . Se una distribuzione ha soltanto un parametro incognito; ma quel parametro non è il momento della distribuzione, il parametro può essere ancora stimato col metodo dei momenti calcolando il momento del primo ordine della distribuzione che sarà una funzione del parametro ed eguagliandolo ad \bar{x} . La soluzione dell'equazione risultante rispetto al parametro incognito fornirà la stima richiesta. Se la distribuzione avesse due parametri incogniti che non siano momenti, si seguirebbe lo stesso procedimento rispetto i primi due momenti della distribuzione. Come esempio in cui i 2 parametri non sono momenti, supponiamo che entrambi i parametri della densità γ debbano essere stimati per mezzo dei momenti del primo e del secondo ordine. Da formule già viste si sa che $\mu = \beta \cdot \alpha$ e $\sigma^2 = \beta^2 \cdot \alpha$; stime di α e β si ottengono perciò risolvendo le seguenti equazioni:

$$\begin{cases} \beta\alpha = \bar{x} \\ \beta^2\alpha = s^2 \end{cases}$$

le cui soluzioni sono date da:

$$\begin{cases} \alpha = \frac{\bar{x}}{\beta} \\ \beta^2 \frac{\bar{x}}{\beta} = s^2 \end{cases} \Rightarrow \begin{cases} \alpha = \frac{\bar{x}^2}{s^2} \\ \beta = \frac{s^2}{\bar{x}} \end{cases}$$

Statistica descrittiva: all'inizio del corso sono state evidenziate alcune caratteristiche essenziali della statistica descrittiva e della statistica inferenziale. Cominciamo ora a parlare della statistica descrittiva. A tale scopo si può partire subito dicendo che la statistica è la scienza che studia i fenomeni collettivi, dove un fenomeno collettivo è un qualsiasi fatto, avvenimento o situazione costituito da un numero sufficientemente grande di fenomeni singoli tra loro simili. Sono fenomeni collettivi, ad esempio, il titolo di studio degli impiegati di un'azienda, gli incidenti stradali verificatisi in una data regione, la statura degli alunni che frequentano una scuola, le nascite avvenute in una certa città, lo sport preferito dai ragazzi di età inferiore ai 18 anni, etc... Dall'analisi di un fenomeno collettivo, realizzata mediante un'indagine statistica, si ottiene una serie di informazioni che permette di comprendere e di interpretare il fenomeno e, quando necessario, di fare delle previsioni che riguardano il suo evolversi nel futuro e, quindi, di programmare in tempo utile gli interventi

opportuni. Quando si vuole studiare un fenomeno collettivo si effettua su di esso un'indagine statistica. Il primo passo di un'indagine statistica consiste nel definire in modo preciso la popolazione statistica, cioè l'insieme degli elementi su cui si effettua l'indagine. Ad esempio costituiscono una popolazione statistica i lavoratori di una fabbrica, gli alunni che frequentano una stessa scuola, le malattie infettive che si sono verificate in una regione. Ciascuno degli elementi che fanno parte di una popolazione statistica prende il nome di **unità statistica**.

Se consideriamo ad esempio come popolazione statistica i dipendenti di una fabbrica, al suo interno le unità statistiche, cioè i dipendenti, differiscono tra loro per una o più caratteristiche, ovvero per il sesso (M o F), per il tipo di mansione svolta (operai, impiegati, etc...), per l'età, per il peso, per la statura, etc... Queste caratteristiche vengono chiamate variabili statistiche ed è rispetto ad una di esse che si effettua l'indagine statistica.

Le variabili statistiche possono essere di due tipi:

- 1) *variabili quantitative o numeriche* che possono essere espresse con numeri. Ad esempio l'età, la statura, il numero dei figli, etc...
- 2) *variabili qualitative*, che non possono essere espresse con numeri, come sesso, il mezzo di trasporto utilizzato, il gruppo sanguigno, etc...

Possiamo quindi concludere che un'indagine statistica è lo studio di un fenomeno collettivo che consiste nell'analizzare come una popolazione statistica si distribuisce rispetto ad una certa variabile statistica.

Tutte le informazioni che si ottengono facendo un'indagine statistica si chiamano **dati statistici**. Un'indagine statistica su un fenomeno collettivo si articola nelle fasi seguenti:

- a) *Rilevazione dei dati*;
- b) *Elaborazione dei dati*, cioè tabulazione, classificazione, rappresentazione grafica, calcolo di misure caratterizzanti quali le misure di tendenza centrale (come media, moda e mediana, etc...), le misure di dispersione (come campo di variazione, varianza, scarto quadratico medio, etc...) e altri indici di tendenza della distribuzione relativi alla sua forma;
- c) *Presentazione dei dati elaborati*;
- d) *Interpretazione dei dati*.

Rilevazione dei dati: la rilevazione dei dati può essere di due tipi:

- 1) rilevazione completa, se è estesa a tutta le unità statistiche della popolazione in esame;

2) rilevazione per campione, se è estesa solo ad una parte più o meno ampia della popolazione statistica, che prende appunto il nome di campione.

Sono rilevazioni complete quelle che si svolgono su popolazioni statistiche costituite da un numero limitato di elementi, come ad esempio su di una classe di studenti, sugli impiegati di un ufficio, sui commercianti di un quartiere di una città, che possono essere tutti contattati e intervistati direttamente. Anche le rilevazioni delle nascite, dei decessi e dei matrimoni sono rilevazioni complete, perché si realizzano utilizzando i dati ufficiali disponibili negli appositi archivi. Quando invece l'indagine si svolge su di una popolazione statistica molto vasta, non è praticamente possibile contattare tutte le unità statistiche e quindi si sceglie un campione rappresentativo, cioè una parte ridotta della popolazione e si svolge l'indagine su di esso rapportando successivamente i risultati ottenuti all'intera popolazione. Se ad esempio si vuole sapere come impiegano il tempo libero gli abitanti di una città di 300.000 persone è impensabile contattarli tutti; si sceglie perciò un campione abbastanza vasto, ad esempio di 1.000 unità, il più possibile rappresentativo degli abitanti della città, cioè uomini e donne nella giusta proporzione, studenti, pensionati, operai, professionisti, disoccupati, etc... e si svolge l'indagine su di esso. I risultati ottenuti si estendono poi all'intera popolazione statistica utilizzando le proporzioni. E' evidente che con una rilevazione per campione dei dati si ottengono risultati meno attendibili di quelli di una rilevazione completa, proprio a causa della scelta del campione che non può mai essere assolutamente rappresentativo della popolazione statistica.

Quando si organizza un indagine statistica è quindi importante la scelta del campione. A tale scopo è opportuno precisare che ci sono diverse tecniche di campionamento su cui non è il caso di dilungarsi in questa sede, che danno origine ad altrettanti tipi di campionamento di cui i più comuni sono:

- a) campionamento probabilistico, caratterizzato dal fatto che di ogni elemento della popolazione è nota la probabilità che venga scelto;
- b) campionamento non probabilistico, che si verifica quando la proprietà precedente non è verificata;

Fra i vari tipi di campioni probabilistici vanno poi ricordati i *campioni aleatori*, semplici o con ripetizioni, i *campioni a più stadi*, i *campioni*

stratificati, i campioni sistematici, e i campioni a gruppi.

I metodi utilizzati per la rilevazione dei dati nelle indagini statistiche sono diversi e cioè:

- a) *l'intervista*, che consiste nel porre delle precise domande direttamente a ciascuna unità statistica e registrare le risposte;
- b) *il questionario*, distribuito a ciascuna unità statistica e successivamente ritirato o restituito con le risposte;
- c) *l'inchiesta telefonica*, che è un'intervista telefonica;
- d) *la consultazione di archivi*, come ad esempio gli archivi comunali;
- e) *la consultazione di pubblicazioni specializzate* come quelle dell'ISTAT (Istituto Nazionale di Statistica);

ELABORAZIONE , PRESENTAZIONE e INTERPRETAZIONE dei

dati: dopo aver parlato della rilevazione dei dati facciamo ora alcune osservazioni riguardanti la loro elaborazione, presentazione ed interpretazione. Se per un qualunque tipo di esperimento consideriamo i dati grezzi ottenuti dalla rilevazione, siamo in genere in difficoltà se vogliamo interpretarli così come sono. Ad esempio un elenco delle votazioni conseguite dagli studenti di un paese non ci dice molto fino a quando questi dati non vengono rielaborati e sintetizzati. Vediamo allora come rendere più efficace la lettura e l'interpretazione dei dati. Supponiamo di avere registrato i dati relativi al peso di 50 studentesse di scuola media superiore di una città. Il primo modo artigianale di riportare tali dati è quello di formare una matrice come mostrato in tabella 1.

Tab. 1

Matrice di dati. Peso (in chili) di 50 studentesse

| | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|
| 01 | 65 | 11 | 65 | 21 | 69 | 31 | 66 | 41 | 61 |
| 02 | 64 | 12 | 64 | 22 | 68 | 32 | 55 | 42 | 61 |
| 03 | 64 | 13 | 64 | 23 | 68 | 33 | 58 | 43 | 62 |
| 04 | 63 | 14 | 63 | 24 | 67 | 34 | 56 | 44 | 60 |
| 05 | 64 | 15 | 63 | 25 | 67 | 35 | 57 | 45 | 60 |
| 06 | 63 | 16 | 63 | 26 | 67 | 36 | 58 | 46 | 61 |
| 07 | 65 | 17 | 73 | 27 | 66 | 37 | 59 | 47 | 61 |
| 08 | 65 | 18 | 71 | 28 | 66 | 38 | 62 | 48 | 62 |
| 09 | 65 | 19 | 70 | 29 | 53 | 39 | 59 | 49 | 60 |
| 10 | 64 | 20 | 69 | 30 | 66 | 40 | 57 | 50 | 62 |

Il primo pensiero sarà quello di organizzarli in qualche forma, ad esempio in ordine di grandezza decrescente come mostrato in tabella 2.

Tab. 2

Matrice (organizzata) di dati.

73 – 71 – 70 – 69 – 69 – 68 – 68 – 67 – 67 – 67 –
 66 – 66 – 66 – 66 – 65 – 65 – 65 – 65 – 65 – 64 –
 64 – 64 – 64 – 64 – 64 – 63 – 63 – 63 – 63 – 63 –
 62 – 62 – 62 – 62 – 61 – 61 – 61 – 61 – 60 – 60 –
 60 – 59 – 59 – 58 – 58 – 57 – 57 – 56 – 55 – 53 .

Nella matrice organizzata dei dati possiamo individuare un primo dato statistico: possiamo infatti dire che tutti i valori appartengono all'intervallo [53, 73], la cui lunghezza sarà chiamata **rango** o **campo di variazione**. In questo caso si ha quindi che il rango è $r = 73 - 53 = 20$.

Questo primo indice ci dice qualcosa, ad esempio che nessuna studentessa pesa 21 kg più di un'altra, ma l'insieme dei dati è ancora molto affollato, soprattutto se pensiamo a campioni molto maggiori. Si può allora pensare di organizzare i dati in classi. Vediamo allora di precisare ulteriormente come formare queste classi di dati di cui ci siamo già occupati in precedenza.

A tale scopo indichiamo alcuni criteri ottimali per la formazione delle classi, ricordando anche che una cattiva scelta delle stesse può portarci ad una cattiva interpretazione dell'intera distribuzione dei dati:

I CRITERIO: il numero delle classi deve rendere chiara la natura di tutta la distribuzione. Se le classi sono infatti troppe o troppo poche, rischieremo di perdere utili informazioni. Nel primo caso, perché in ogni classe vi sarebbero pochissimi elementi o addirittura nessuno; nel secondo caso, perché potrebbero accadere che, essendo concentrati molti elementi in poche classi, perderemo di vista la globalità della distribuzione. In genere, come si è già accennato in precedenza, si scelgono classi in numero variabile da 6 a 20. Secondo **STURGES** si ha un numero ottimale di classi indicato con N_c , scegliendo $N_c = [1 + 1,1443 \cdot \ln N]$, dove $[\alpha]$ rappresenta l'intero più vicino ad α ed N rappresenta il numero dei dati osservati.

II CRITERIO: le classi devono avere la stessa ampiezza e, dopo quanto detto al primo punto, tale ampiezza sarà data da:

$$h = \frac{r}{N_c}$$

dove r indica il rango dell'insieme dei dati osservati.

Come esempio determiniamo ora il numero di classi e l'ampiezza delle classi nel caso dei dati presentati nella matrice della tabella 2. Dalla formula di Sturges si ha che $N_c = [1 + 1,443 \cdot \ln 50] = [1 + (1,433) \cdot (3,91)] = [6,64] = 7$; per cui l'ampiezza è data da $h = 20 / 7 = 2,86$.

III CRITERIO: l'ampiezza dell'intervallo deve essere un numero tale da individuare il punto di mezzo, inoltre, come si è già detto in precedenza, gli estremi di ogni intervallo di classe devono essere presi mezza unità oltre l'accuratezza di misura e ciò per far sì che nessun dato osservato cada sugli estremi dell'intervallo.

Come esempio costruiamo ora le classi nel caso dei dati restituiti al campione in esame. Si vede che è uguale a 3 l'ampiezza opportuna per le 7 classi che dovranno rappresentare i dati, cioè il peso delle studentesse espressi nella matrice dei dati. Possiamo pensare allora che gli estremi dei 7 intervalli di classe siano 52.5, 55.5; 55.5, 58.5; ... ; 70.5, 73.5 e che i valori di mezzo degli intervalli siano 54, 57, ..., 72. Avremo allora la tavola mostrata in tabella 3.

Tab. 3

| <i>Classi di peso</i> | <i>Punti di mezzo</i> | <i>Etichette (labels)</i> | <i>Numero di studentesse</i> |
|-----------------------|-----------------------|---------------------------|------------------------------|
| 52.5-55.5 | 54 | L | 2 |
| 55.5-58.5 | 57 | □ | 5 |
| 58.5-61.5 | 60 | □ □ | 9 |
| 61.5-64.5 | 63 | □ □ □ | 15 |
| 64.5-67.5 | 66 | □ □ L | 12 |
| 67.5-70.5 | 69 | □ | 4 |
| 70.5-73.5 | 72 | L | 3 |

Nella tabella 3, nella colonna delle etichette, è stato usato un metodo assai comune di contare gli elementi osservati tramite lati di quadrato con chiusura da 1 a 5. A questo punto, anche senza conoscere esattamente tutti i dati, ma conoscendo solo la frequenza assoluta, cioè il numero di elementi di ogni classe, possiamo avere un'idea della distribuzione dei dati. Ad esempio potremmo dire che ci sono 5 studentesse che pesano 57 kg, 12 che pesano 66 kg e così via.

Per avere altri tipi di informazione, sempre più precisi ed esaurienti possiamo definire altri indici statistici, più esattamente definiremo:

- a) la *funzione di frequenza assoluta*, indicata con $\varphi(x)$ che ad ogni classe associa il numero di elementi della classe;
- b) la *funzione di frequenza relativa*, indicata con $\varphi_r(x)$, ovvero il rapporto fra il numero di elementi della classe e il numero totale di elementi;
- c) la *funzione di frequenza cumulativa*, indicata con $\varphi_c(x)$, ovvero il numero di elementi della classe e delle classi precedenti;
- d) la *funzione di frequenza cumulativa relativa*, indicata con $\varphi_{cr}(x)$, ovvero il rapporto fra il numero degli elementi dato dalla frequenza cumulativa ed il numero totale degli elementi;

Nella tabella 4 sono riportati i valori della 4 funzioni suddette, corrispondenti ai punti di mezzo delle 7 classi per i dati riguardanti l'esempio in esame.

Tab. 4

| $x = \text{Punto di mezzo}$ | $\varphi(x)$ | $\varphi_r(x)$ | $\varphi_c(x)$ | $\varphi_{cr}(x)$ |
|-----------------------------|--------------|----------------|----------------|-------------------|
| 54 | 2 | 0.04 | 2 | 0.04 |
| 57 | 5 | 0.10 | 7 | 0.14 |
| 60 | 9 | 0.18 | 16 | 0.32 |
| 63 | 15 | 0.30 | 31 | 0.62 |
| 66 | 12 | 0.24 | 43 | 0.86 |
| 69 | 4 | 0.08 | 47 | 0.94 |
| 72 | 3 | 0.06 | 50 | 1.00 |

Ad esempio si ha:

$$N = 50 \quad \varphi(60)=9 \quad \varphi_r(60)=\varphi(60) / N = 9/50 = 0,18$$

$$\varphi_c(60)= \varphi(60)+ \varphi(57)+ \varphi(54)=9+5+2=16$$

$$\varphi_{cr}(60)= \varphi_r(60)+ \varphi_r(57)+ \varphi_r(54)=0.18+0.10+0.04=0.32$$

Naturalmente $\varphi_{cr}(x)$ si può calcolare anche direttamente osservando che $\varphi_{cr}(x) = \varphi_c(x) / N$; per cui si ha che $\varphi_{cr}(60) = \varphi_c(60) / N = 16/50 = 0.32$

Le funzioni di frequenza, i cui valori sono riportati nella tabella 4, sono rappresentate graficamente dall'istogramma di figura 1, dal grafico a segmenti di figura 2, dal poligono di frequenza di figura 3 e dal poligono di frequenza cumulativa relativa di figura 4.

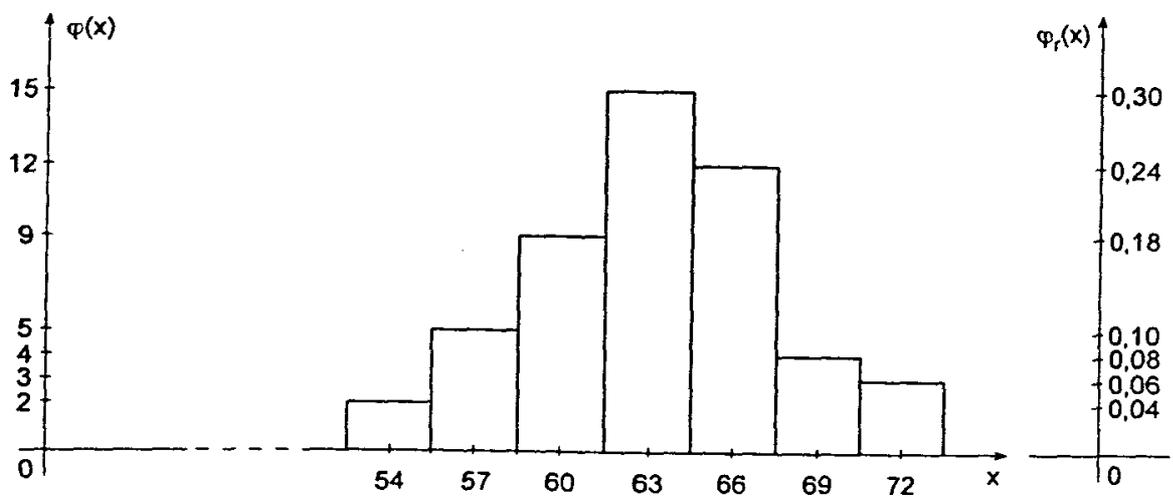


Fig. 1. Iistogramma.

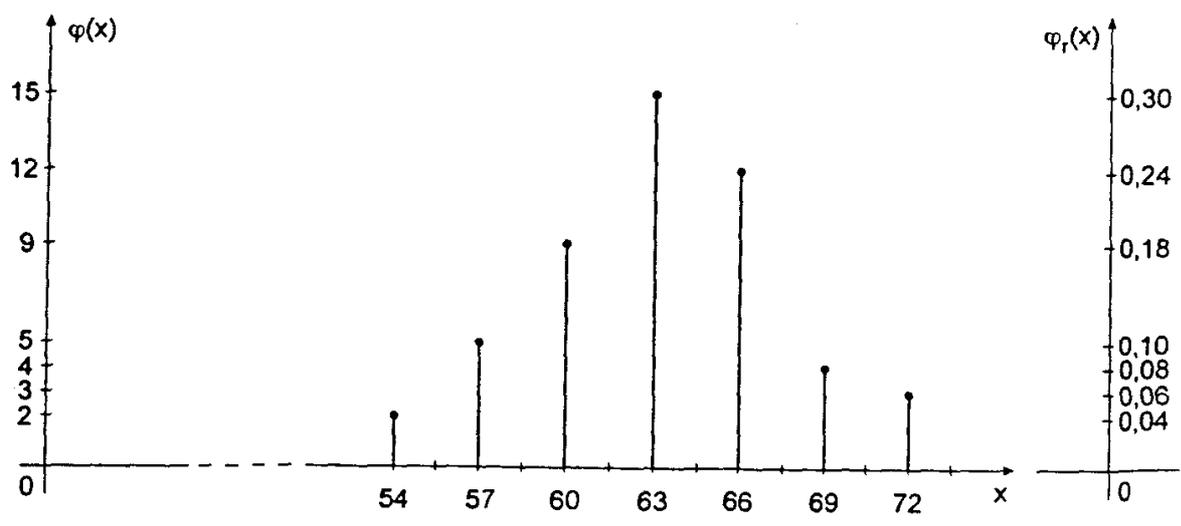


Fig. 2. Grafico a segmenti.

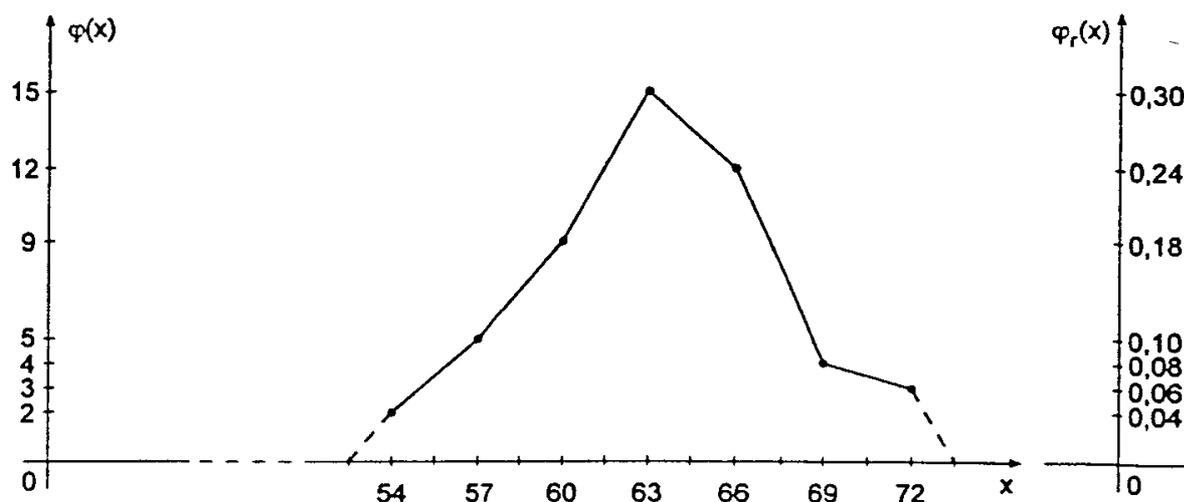


Fig. 3. Poligono di frequenza.

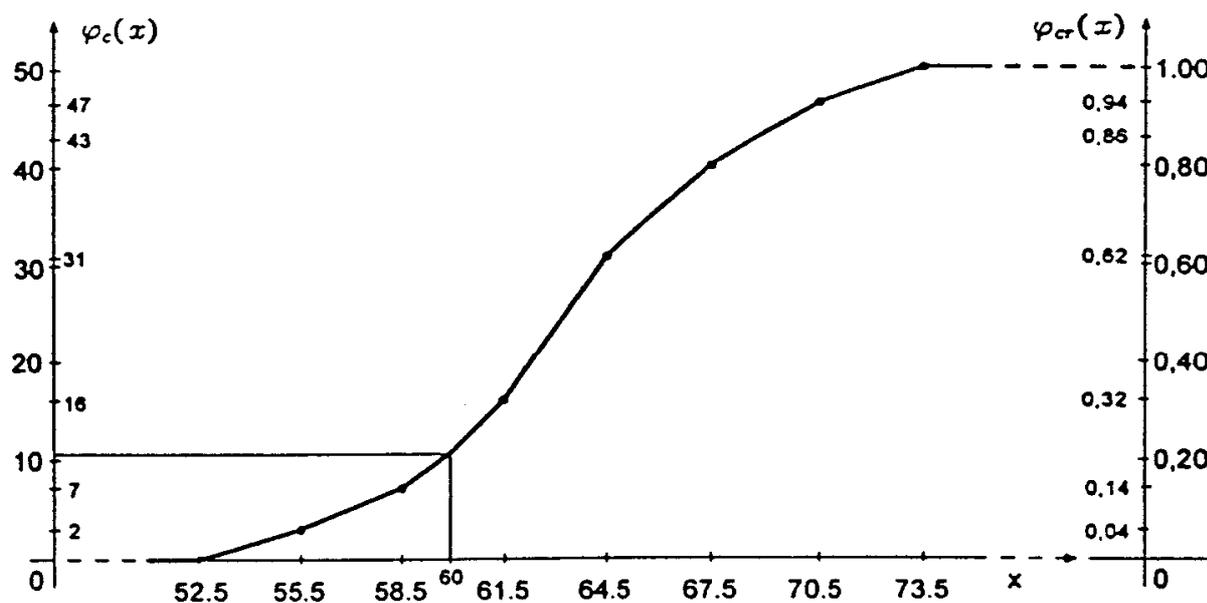


Fig. 4. Poligono di frequenza relativa cumulativa.

In ciascuno dei 4 grafici sono stati riportati sull'asse delle ordinate a sinistra i valori assoluti e a destra i valori relativi della corrispondente funzione, ovviamente in scala opportuna, ottenendo così il duplice scopo di poter leggere entrambi i valori. Da questi grafici si possono leggere informazioni sui dati; ad esempio dal grafico 4 si può vedere che il numero di studentesse il cui peso cade fra 61,5 e 70,5 kg è semplicemente

la differenza fra le frequenze cumulative corrispondenti e perciò abbiamo $47 - 16 = 31$. Per i valori intermedi si possono soltanto dare stime, ad esempio il numero di studentesse con peso minore o uguali a 60 kg è circa di 11 e questo dato è stato ottenuto graficamente come mostrato in figura 4.

Vediamo ora di caratterizzare una distribuzione statistica, ovvero un'insieme di dati del tipo visto finora, attraverso delle misure che ne riassumano le principali proprietà. Si tratta delle cosiddette ***misure di tendenza centrale***, cioè di alcune caratterizzazioni sintetiche della distribuzione che hanno lo scopo di dare un'idea di dove la distribuzione sia collocata e quanto sia concentrata. Gli indici di tendenza centrale che esamineremo più da vicino sono la ***media***, di cui ci siamo già occupati, la ***mediana*** e la ***moda***.

A proposito della media è opportuno precisare che se abbiamo n osservazioni numeriche x_1, x_2, \dots, x_n , la loro media, che come sappiamo è

data da $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$, è più comunemente nota come ***media aritmetica delle osservazioni***.

Sono indici di tendenza centrale anche le ***misure di dispersione***, tra cui quelle di gran lunga le più usate e di cui ci siamo già occupati sono certamente ***lo scarto quadratico medio*** o ***deviazione standard*** e ***la varianza***. Anche il ***rango*** o ***campo di variazione*** può talvolta essere un indice utile per comprendere come stanno grosso modo le cose. Si pensi ad esempio alla conoscenza della temperatura massima e minima di una città in un dato giorno; è però abbastanza evidente che tale indice risente molto di valori particolarmente alti o particolarmente bassi, inoltre esso tende inevitabilmente a crescere se aumentiamo il numero delle rilevazioni dei dati.

Calcoliamo ora la media dei pesi del campione in esame.

Applicando la formula che definisce la media di una distribuzione

empirica si ottiene $\bar{x} = 63.22$ (ricordando che $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{N_c} x_i \cdot f_i$). Avendo

classificato la distribuzione dei pesi possiamo anche calcolare la media tramite la formula che la definisce per distribuzioni classificate ed ottenere così che $\bar{x} = 63.24$, dove \bar{x} è leggermente diverso dal precedente. Tale differenza è spiegabile nel senso che i dati raggruppati in classi perdono un qualche elemento di informazione rispetto ai dati noti singolarmente (uno

ad uno), a meno che non vi sia una perfetta simmetria intorno al punto di mezzo della classe.

Facciamo ora un esempio per capire quando la media si possa ritenere accettabile come valore di sintesi dei dati. A tale scopo calcoliamo la media degli stipendi annuali degli individui A, B, C, D che percepiscono rispettivamente 10, 15, 16, 80 mila euro. Usando la definizione si ha che:

$$\bar{x} = \frac{1}{4} \cdot (10 + 15 + 16 + 80) = 30.25 \text{ mila euro}$$

Questo valore non è certamente una buona sintesi degli stipendi, infatti esso non è vicino ai guadagni di $A, B,$ e C e neppure a quello di D . Presentarlo dunque come stipendio medio potrebbe provocare grossolani equivoci. Sicuramente il motivo è che il valore estremo, cioè quello più elevato, influenza negativamente la media. Vedremo fra breve come ovviare alla presenza di valori estremi nella determinazione di un ragionevole valore medio. A tale scopo definiamo ora altri due tipi di medie.

Mediana: per evitare l'influenza di valori molto distanti dagli altri perché troppo grandi o troppo piccoli, si definisce la mediana di n osservazioni come il valore che divide l'insieme dei dati ordinato dal più piccolo al più grande o viceversa esattamente in due parti. In altre parole la mediana è il valore che occupa la posizione centrale a seconda che n sia dispari o pari, si ha allora:

$$x_{med} = \frac{1}{2} \left(x_{\frac{n}{2}} + x_{\left(\frac{n}{2}+1\right)} \right) \quad \text{se } n \text{ è pari}$$
$$x_{med} = x_{\frac{n+1}{2}} \quad \text{se } n \text{ è dispari}$$

avendo indicato con x_1, x_2, \dots, x_n i valori osservati e supposti in ordine crescente o decrescente.

Esempio: i voti di uno studente universitario in 6 esami sono stati 28, 29, 19, 30, 23, 26 e si vuole calcolare la media e la mediana di questi voti. Per la media si ha:

$$\bar{x} = \frac{1}{6} (28 + 29 + 19 + 30 + 23 + 26) = 25.833$$

e per la mediana, poiché $n = 6$, e dopo aver ordinato i dati in senso crescente, come 19, 23, 26, 28, 29, 30, si ha che:

$$x_{med} = \frac{1}{2}(26 + 28) = 27$$

In questo caso non ci sono valori estremi, ovvero se ci sono essi si compensano e dunque media e mediana hanno valori molto simili.

Esempio: gli stipendi di 6 dipendenti sono dati in migliaia di € da 1.2, 1.4, 1.3, 2.0, 0.8, 9.7. Per la media si ha che:

$$\bar{x} = 2.7\bar{3}$$

mentre per la mediana, poiché $n = 6$ ed ordinando i dati in senso crescente come segue 0.8, 1.2, 1.3, 1.4, 2.0, 9.7, si ha che:

$$x_{med} = \frac{1}{2}(x_3 + x_4) = \frac{1}{2}(1.3 + 1.4) = 1.35$$

Questa volta la mediana ci appare come un valore molto più ragionevole della media per indicare una sintesi, ovvero un valore “medio” della distribuzione degli stipendi.

Se i dati sono raggruppati in classi, dobbiamo allora operare come segue: per prima cosa si deve individuare la classe mediana, cioè la classe dove è situata la mediana (in generale la classe mediana è la prima classe in cui la funzione di frequenza cumulativa supera $n/2$); il secondo passo è quello di individuare la mediana all'interno della classe mediana. Ciò viene fatto tramite la formula seguente:

$$x_{med} = \lambda_m + h \cdot \frac{\frac{n}{2} - \varphi_c(x_m - 1)}{\varphi(x_m)}$$

dove λ_m è l'estremo inferiore della classe mediana; h è l'ampiezza delle classi; $x_m - 1$ rappresenta il punto di mezzo della classe che precede la classe mediana; x_m è il punto di mezzo della classe mediana.

Esempio: come esempio calcoliamo la mediana relativamente all'insieme dei pesi delle 50 studentesse di cui ci stiamo occupando considerando prima i dati singolarmente e poi raggruppati in classi. La mediana, essendo pari al numero di dati e osservando la tabella 2 vista prima, è data da:

$$x_{med} = \frac{1}{2}(x_{25} + x_{26}) = \frac{1}{2}(64 + 63) = 63.5$$

Se i dati sono raggruppati in classi, individuiamo innanzitutto la *classe mediana* come quella che ha come punto di mezzo uguale a 63, avendo questa una frequenza cumulativa uguale a $31 > n/2 = 50/2$, come si può osservare in tabella 4 e in figura 4.

Per determinare la mediana osserviamo poi, sempre dalle tabelle 3 e 4, che $\lambda_m = 61,5$, $h = 3$, $\gamma_c(x_{m-1}) = \gamma_c(60) = 16$ e $\varphi(x_m) = \varphi(63) = 15$; da cui si ha che:

$$x_{med} = 61.5 + 3 \cdot \frac{25 - 16}{15} = 63.3$$

Questo è un valore praticamente uguale a quello trovato usando tutti i dati singolarmente; ciò implica una certa simmetria tra i dati rispetto al valore medio.

L'idea di mediana può essere estesa; infatti se la mediana è il valore di mezzo delle osservazioni, possiamo pensare a valori che dividono i dati in 4 parti, 10 parti o in 100 parti. Parleremo allora di *quartili*, *decili* e *percentili*.

Moda: Molto spesso i dati sono divisi in classi che non sono di tipo numerico, ad esempio la provincia di nascita, il tipo di lavoro, il gruppo sanguigno, etc. In questi casi non ha senso parlare di media o di mediana, a meno di non aver etichettato i dati. E' allora utile introdurre un'altra misura di tendenza centrale e precisamente la *moda*. La moda è il valore che si ripete più spesso in una distribuzione di dati. Essa è particolarmente utile quando la distribuzione dei dati è molto concentrata su uno dei valori. Se i dati sono raggruppati per determinare la moda occorre preliminarmente determinare la *classe modale*, cioè la classe in cui si trova la moda. Di solito la classe modale è quella in cui la frequenza assoluta $\varphi(x)$ è massima, notando che ve ne può essere anche più di una. Come per la mediana anche qui si può dare una formula atta a determinare questo valore. La moda è così data da:

$$x_{mod} = \lambda_j + h \cdot \frac{|\gamma(x_j) - \gamma(x_{j-1})|}{|\gamma(x_j) - \gamma(x_{j-1})| + |\gamma(x_{j+1}) - \gamma(x_j)|}$$

dove abbiamo supposto le classi determinate come al solito e abbiamo indicato con λ_j l'estremo inferiore della classe modale, con h l'ampiezza della classi e con x_j il punto di mezzo della classe modale.

Esempio: come esempio determiniamo la moda del campione dei pesi in esame considerando prima i dati singolarmente e poi raggruppati in classi.

Nel 1° caso dalla tabella 2 vista precedentemente si può osservare che la moda è $x_{\text{mod}} = 64$, valore che si ripete più spesso degli altri, cioè 6 volte. Considerando i dati classificati dalla tabella 4, si può osservare che la classe modale è quella con punto di mezzo “63”, per cui si ha:

$$x_{\text{mod}} = 61.5 + 3 \cdot \frac{|15-9|}{|15-9|+|12-15|} = 63.5$$

Fra le misure di tendenza centrale che abbiamo presentato, la moda è senz'altro la misura media meno precisa. La sua collocazione e la scelta stessa della classe modale sono fattori che dipendono fortemente dal modo in cui vengono classificati i dati. Cambiando infatti l'ampiezza delle classi può variare anche di molto la scelta della classe modale. Ricordiamo infine che se non esiste una sola classe modale, cioè se la frequenza non ha un solo punto di massimo, allora il concetto di moda risulta in genere di scarsa utilità. Le distribuzioni con una sola moda si chiamano ***distribuzioni unimodali***.

Media, mediana e moda: sul significato della media, della moda e della mediana, e quindi sulla loro effettiva utilità nell'interpretazione del fenomeno oggetto dell'indagine statistica, è opportuno fare alcune

considerazioni:

- la *media* dei dati tende a livellare tutti i dati e quindi a non fare risaltare le grosse differenze che possono esserci fra essi;
- la *mediana* tiene conto solo del dato centrale della successione dei dati e non è quindi influenzata né dai valori più grandi né dai valori più piccoli;
- la *moda* è il dato che si ripete con maggiore frequenza e quindi non è influenzata dagli altri dati;

E' quindi necessario stabilire caso per caso se i tre valori medi siano tutti significativi. In alcune indagini statistiche ciò può verificarsi solo per due di essi oppure anche per uno solo. Nel caso in cui la media, la moda e la mediana coincidano oppure siano valori molto vicini fra loro si dice che il fenomeno che si studia ha una *distribuzione normale di dati*. Dal termine usato già si capisce che questo fatto è abbastanza ricorrente nelle indagini statistiche. Quando c'è una distribuzione normale di dati, il diagramma delle frequenze assume una forma particolare che ricorda quella di una

campana e la curva che ne è una buona approssimazione prende il nome di curva normale o Gaussiana.

A scopo illustrativo diamo ora tre esempi in cui è opportuno usare per misurare la tendenza centrale rispettivamente la media, la moda e la mediana.

Supponiamo di avere un gruppo di 200 studenti che:

- a)devono viaggiare su di un aereo;
- b)devono acquistare scarpe per fare sport;
- c)devono concorrere per una borsa di studio.

Per quanto riguarda il primo punto è chiaro che se l'aereo può portare 20 tonnellate con un massimo di 200 persone pensando che ogni persona pesi mediamente 80 kg, si deve imporre un peso massimo di 20 kg per le valige, e questa è una *media aritmetica*.

Per il secondo punto, supponendo che il calcio sia lo sport più praticato dagli studenti dobbiamo sapere quale sia la moda, ovvero la frequenza della classe dei giocatori di calcio per pianificare gli acquisti.

Per il terzo punto sembra naturale che supponendo che le borse di studio vengano date al 50% degli studenti più bravi e meritevoli, uno studente è particolarmente interessato a sapere se il suo punteggio si colloca sopra o sotto la mediana.

Vediamo ora alcune relazioni tra i tipi di media che abbiamo illustrato. Abbiamo già detto che esse non sono applicabili per qualsiasi insieme di dati, e vedremo adesso che, una volta disegnato il grafico della distribuzione dei dati, la media, la mediana e la moda si dispongono in un certo modo dipendentemente dal fatto che il grafico sia più o meno asimmetrico, intendendo sempre parlare di distribuzioni unimodali.

Se la distribuzione è simmetrica intorno ad un certo valore, allora come si è già detto, le tre misure coincidono come indicato nel grafico di figura 5.

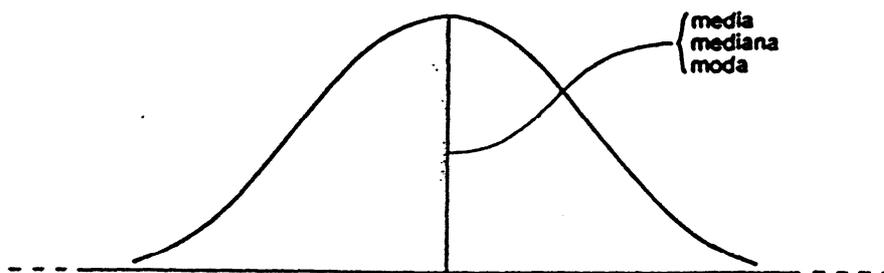


Fig. 5. Grafico (simmetrico) della distribuzione dei dati.

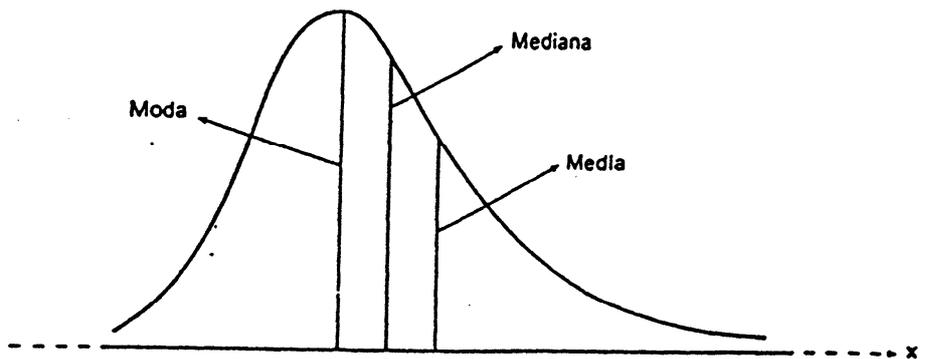


Fig. 6

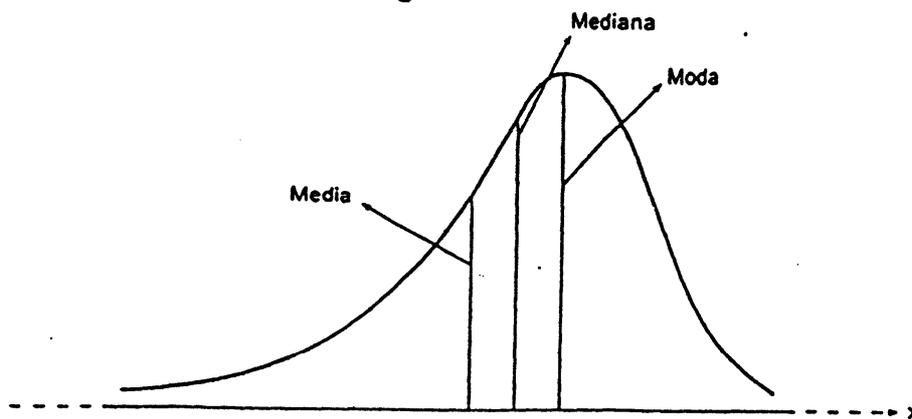


Fig. 7

Quando non c'è simmetria i 3 valori di tendenza centrale si distinguono tra loro anche se nelle distribuzioni unimodali la mediana si colloca sempre tra la media e la moda. Più precisamente, come mostrano i due grafici di figura 6 e 7, se la coda della distribuzione è a destra, allora la moda precede la mediana e la media, mentre se la coda è a sinistra è la media che precede la mediana e la moda. Per comprendere le ragioni di ciò è opportuno notare che se la coda della distribuzione è a destra, la moda dovrà essere a sinistra della mediana in quanto si colloca nel punto di massimo della distribuzione; mentre la media che come abbiamo visto risente molto dei valori estremi, dovrà essere a destra della mediana. In modo analogo sostituendo la destra con la sinistra si può ragionare nel caso di distribuzioni con coda a sinistra. Per distribuzioni unimodali che siano moderatamente asimmetriche vale la relazione empirica:

$$\bar{x} - x_{\text{mod}} = 3(\bar{x} - x_{\text{med}})$$

Media geometrica e media armonica: ci sono poi altri tipi di medie in senso lato che vengono usate in considerazioni statistiche. Ne ricordiamo ancora due e precisamente la media geometrica e la media armonica. Dati

n valori osservati x_1, x_2, \dots, x_n si chiama media geometrica il numero definito da:

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = (x_1 \cdot x_2 \cdot \dots \cdot x_n)^{\frac{1}{n}}$$

e la media armonica il numero definito da:

$$\bar{x}_a = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

Si fa osservare che se i valori osservati sono tutti positivi si ha come risultato di carattere generale che:

$$\bar{x}_a \leq \bar{x}_g \leq \bar{x}$$

dove il segno “=” vale solo se tutti i valori sono uguali fra loro.

Per illustrare l'utilità dell'impiego pratico delle medie suddette consideriamo gli esempi seguenti.

Esempio 1: in un certo anno la benzina B ha avuto un costo medio di 1.500 £ al litro e le arance A costavano mediamente 4.500 £ al kg. Se nell'anno successivo il prezzo medio della benzina B è salito a 2.000 £ al litro e quello delle arance è sceso a 4.000 £ quale tipo di media è più opportuna per illustrare il rapporto dei prezzi nei due anni in esame?

A tale scopo si può osservare che il rapporto fra il costo delle arance e quello della benzina era di 3 al primo anno e di 2 al secondo; se facciamo la media aritmetica otteniamo un rapporto medio pari a:

$$\bar{r} = \frac{1}{2}(3 + 2) = 2.5.$$

Il rapporto fra il prezzo della benzina e quello delle arance è $1/3$ il primo il anno e $1/2$ il secondo anno. La media di tali rapporti è:

$$\bar{r} = \frac{1}{2}\left(\frac{1}{3} + \frac{1}{2}\right) = \frac{5}{12} = 0.417$$

Se invertissimo i rapporti ci aspetteremmo di ottenere il rapporto medio inverso dato da:

$$0.4 = \frac{1}{2.5} = \frac{10}{25} \neq \frac{5}{12}$$

Possiamo dire che la media non è sufficiente ad illustrare il rapporto dei prezzi nel periodo in esame.

Vediamo ora cosa accade se si usa la media geometrica. La media geometrica dei rapporti fra il costo delle arance e quello della benzina è dato da:

$$\overline{r}_g = \sqrt{3 \cdot 2} = \sqrt{6} = 2.449$$

La media geometrica dei rapporti invertiti, cioè fra quello della benzina e quello delle arance, è dato da

$$\overline{r}_g = \sqrt{\frac{1}{3} \cdot \frac{1}{2}} = \frac{1}{\sqrt{6}} = 0.408$$

Queste medie sono adesso una l'inversa dell'altra. Quindi possiamo concludere che la media geometrica è più utile della media aritmetica per illustrare il rapporto dei prezzi nei due anni considerati.

Esempio 2: Si ricordi che la media armonica è data da:

$$\overline{x}_a = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2}}$$

Un individuo viaggia da Firenze a Pisa, viaggiando per l'intero viaggio d'andata a 80 km/h mentre per quello di ritorno a 120 km/h. Si vuole calcolare la velocità media per l'intero viaggio. A tale scopo si può osservare che l'individuo impiega 1 ora per andare da Firenze e Pisa ed impiega 40 minuti per ritornare. Dunque viene effettuato un viaggio di 160 km in 100 minuti, alla media quindi di 96 km/h. Infatti $160 : 100 = x : 60$; quindi $x = 96$. Questa non è altro che la media armonica delle due velocità, infatti:

$$\overline{v}_a = \frac{2}{\frac{1}{v_1} + \frac{1}{v_2}} = \frac{2}{\frac{1}{80} + \frac{1}{120}} = 96$$

Si fa osservare che usando la media aritmetica, ovvero:

$$\overline{v} = \frac{v_1 + v_2}{2} = \frac{80 + 120}{2} = 100$$

non avremmo ottenuto un risultato corretto. E' opportuno inoltre precisare che qualora le distanze percorse non siano uguali, bisogna usare una media armonica ponderata delle velocità considerando le distanze come *pesi rispettivi*. Ad esempio se fosse stato percorso il tratto Firenze – Pisa – Viareggio alle stesse due velocità, e con d_1 uguale alla distanza tra Firenze

– Pisa e d_2 uguale alla distanza Pisa – Viareggio avremmo avuto una velocità media data da:

$$\bar{v}_{ap} = \frac{d_1 + d_2}{\frac{d_1}{v_1} + \frac{d_2}{v_2}} = \frac{80 + 20}{\frac{80}{80} + \frac{20}{120}} = \frac{100}{1.16} = 85.71$$

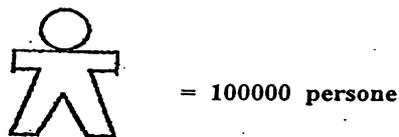
avendo posto uguale a 20 km/h la distanza tra Pisa e Viareggio. Se le distanze fossero uguali si può facilmente verificare che l'ultima formula usata per il calcolo della velocità media si ridurrebbe a quella usata precedentemente. Infatti se $d_1 = d_2 = d$:

$$\bar{v}_{ap} = \frac{d + d}{\frac{d}{v_1} + \frac{d}{v_2}} = \frac{2d}{d \left(\frac{1}{v_1} + \frac{1}{v_2} \right)} = \bar{v}_a$$

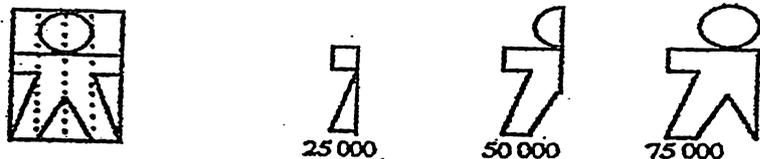
Rappresentazione grafica dei dati: i dati ottenuti con le indagini statistiche, come si è già visto, oltre che essere disposti in opportune tabelle possono essere visualizzati mediante opportune rappresentazioni grafiche, che facilitano il confronto dei dati e soprattutto forniscono una comprensione immediata dell'andamento complessivo dei fenomeni studiati. E' importante sottolineare che è necessario saper scegliere caso per caso il tipo di rappresentazione grafica più opportuna e appropriata. Si è già parlato nella trattazione precedente di *istogramma*, di *grafico a segmenti*, di *poligono di frequenza assoluta*, di *frequenza relativa*, di *frequenza cumulativa*, e di *frequenza cumulativa relativa*. Prendiamo ora in esame altre due rappresentazioni grafiche fra quelle più comunemente utilizzate e precisamente l'*ideogramma* e l'*aerogramma*.

Ideogramma: un *ideogramma* è una rappresentazione grafica molto semplice e diffusa. Esso si basa su di una unità di misura grafica, cioè su di una *figurina simbolo* (*icona*) che richiama il soggetto che si vuole rappresentare, cui si fa corrispondere un valore numerico specifico. Si tratta però di una forma di rappresentazione non troppo rigorosa in quanto il limite grafico alla possibilità di frazionare la figurina simbolo impone un arrotondamento dei dati che può anche divenire rilevante.

Ad esempio supponiamo che alla figurina simbolo di figura 1 corrispondano 100.000 persone. In questo caso non sembra ragionevole frazionare la figurina in più di 4 parti, come mostrato in figura 1. Si dovrebbe cioè arrotondare i dati al 25000 più prossimo.



Figurina simbolo



Frazionamento

Fig. 1

Questa rappresentazione grafica serve soprattutto come mezzo di divulgazione di slogan pubblicitari dove si possono adoperare con maggiore efficacia i motivi ideografici che richiamano più facilmente al grosso pubblico la natura dei fatti rappresentati. Vediamo ora due esempi rappresentati rispettivamente nelle figure 2 e 3, e precisamente 1) l'ideogramma delle lavatrici e 2) l'ideogramma del pesce pescato da alcune nazioni nel 1995.

1) In un'indagine sulla produzione di lavatrici da parte di quattro aziende A, B, C, D nel 1994, si sono ottenuti i seguenti dati:

| Azienda | Lavatrici |
|---------|-----------|
| A | 110.000 |
| B | 140.000 |
| C | 60.000 |
| D | 80.000 |

La rappresentazione grafica con un ideogramma risulta in questo caso abbastanza opportuna e fornisce con immediatezza un'idea della situazione (Fig. 2):

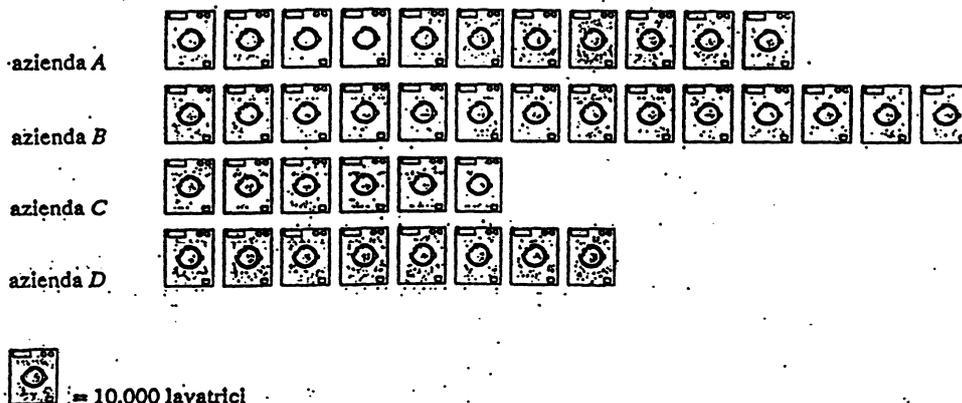


Fig. 2

2) Disegniamo, ad esempio, l'ideogramma relativo alla quantità di pesce pescato in un anno da alcune nazioni, disponendo dei seguenti dati:

| Nazione | Pesce pescato nel 1995 (tonnellate) |
|----------|-------------------------------------|
| Giappone | 11.000.000 |
| Cina | 6.500.000 |
| USA | 5.000.000 |
| Russia | 8.000.000 |
| Norvegia | 2.500.000 |

Possiamo scegliere come unità grafica il disegno stilizzato di un pesce e stabilire che

 = 1.000.000 tonnellate

e quindi

 = 500.000 tonnellate

Otteniamo così l'ideogramma

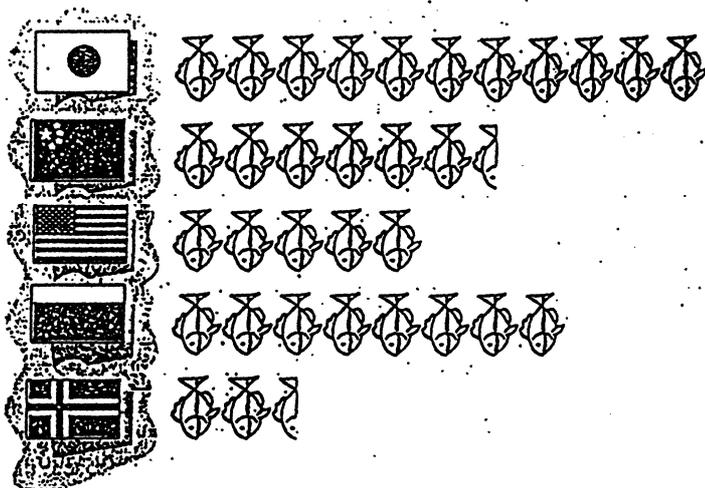


Fig. 3

Areogramma: un *areogramma* è una rappresentazione che si presta a visualizzare il peso relativo delle singole parti rispetto al tutto. Un areogramma si costruisce suddividendo una figura che rappresenti l'intero, cioè il tutto, in parti proporzionali alle frequenze relative, ed è quindi significativo soltanto se il numero delle parti è piccolo. E' chiaro che si può prendere una figura qualsiasi, ad esempio un quadrato; in tale caso però una soltanto un'opportuna misura del lato del quadrato rende agevole la suddivisione della figura in parti proporzionali alle frequenze relative. Proprio per questo si ricorre più frequentemente al cerchio e si ha quindi una rappresentazione grafica che prende il nome di *areogramma circolare* o *diagramma circolare*, o *diagramma a torta (Piechart)*, o *diagramma a settori circolari*. In esso le aree dei settori circolari che sono proporzionali

alle frequenze relative sono proporzionali alle ampiezze dei rispettivi angoli al centro e pertanto la misura del raggio può essere scelta con assoluta libertà. Del resto la rappresentazione per settori circolari risulta anche più espressiva dal punto di vista percettivo. In ogni caso nella costruzione di un aerogramma sono quasi sempre inevitabili alcuni arrotondamenti.

Esempio: vogliamo rappresentare con un aerogramma come sono distribuiti i diversi tipi di navi mercantili italiane con una stazza superiore a 100 tonnellate riportati nella tabella di figura 4 dove i dati sono stati desunti dal calendario Atlante de Agostini del 1997.

| NAVIGLIO MERCANTILE | |
|-------------------------|--------|
| tipo di navi | numero |
| navi passeggeri e miste | 364 |
| navi da carico secco | 137 |
| navi cisterna | 310 |
| navi portacontenitori | 19 |
| navi di altro tipo | 613 |

$$\frac{364}{1443} = \frac{\alpha}{360^\circ} \quad \alpha \approx 91^\circ \quad \text{navi passeggeri e miste}$$

$$\frac{137}{1443} = \frac{\beta}{360^\circ} \quad \beta \approx 34^\circ \quad \text{navi di carico secco}$$

$$\frac{310}{1443} = \frac{\gamma}{360^\circ} \quad \gamma \approx 77^\circ \quad \text{navi cisterna}$$

$$\frac{19}{1443} = \frac{\delta}{360^\circ} \quad \delta \approx 5^\circ \quad \text{navi portacontenitori}$$

$$\frac{613}{1443} = \frac{\varepsilon}{360^\circ} \quad \varepsilon \approx 153^\circ \quad \text{navi di altro tipo}$$

Il naviglio mercantile italiano comprende dunque 1443 navi. Si può ora calcolare l'ampiezza degli angoli al centro corrispondenti ai tipi diversi di

nave come riportato in figura 4. La situazione è visualizzata nell'aerogramma di figura 4.

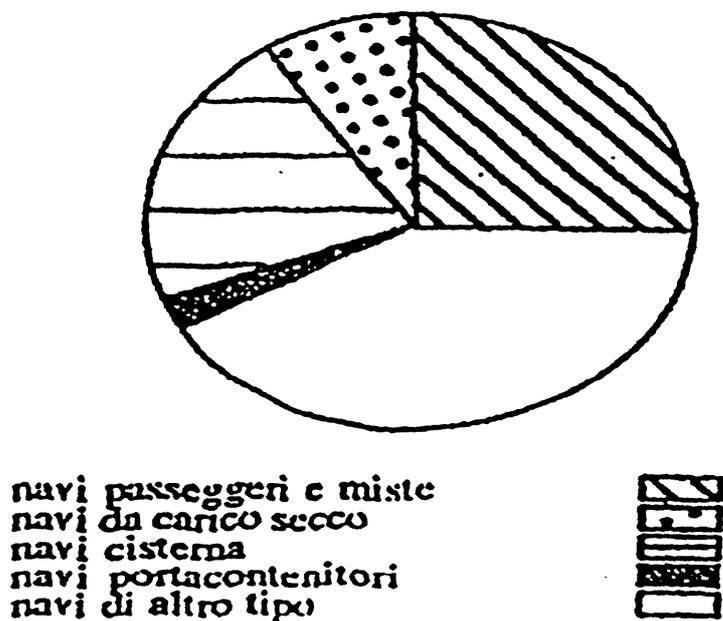


Fig. 4

Per agevolare la lettura dell'aerogramma sovente viene anche riportata all'interno di ciascun settore circolare la frequenza percentuale della corrispondente componente. Le frequenze percentuali per l'esempio in esame sono nell'ordine le seguenti: 25, 23; 9,49; 21,48; 1,32; 42,48.