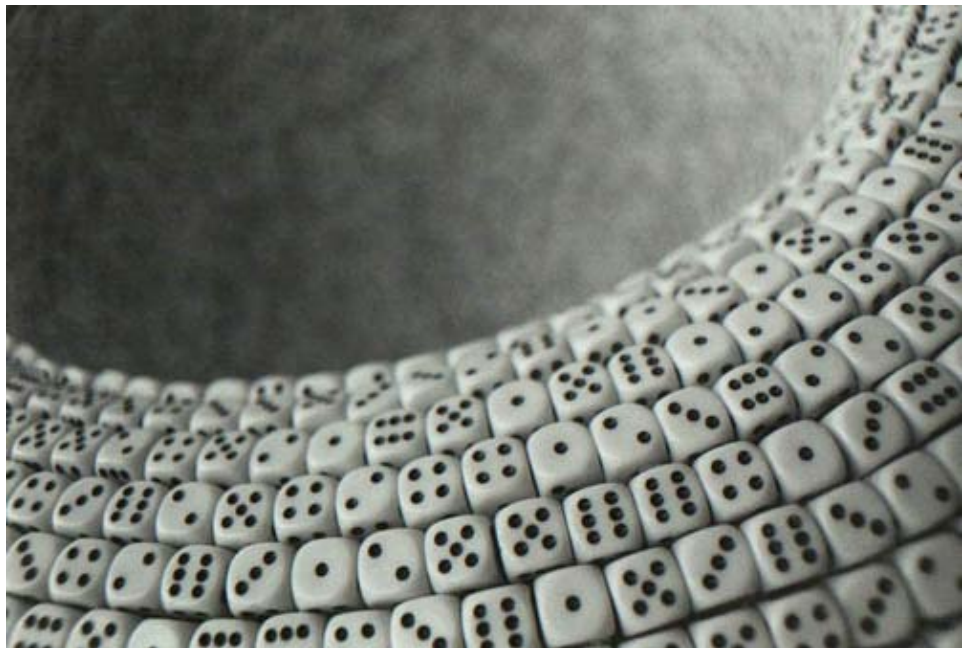


# Appunti del corso di Statistica e Calcolo delle Probabilità

Appunti raccolti durante il corso di *Statistica e Calcolo delle Probabilità* tenuto dal professore **Camillo Fucci** per il *Corso di Laurea in Informatica* presso l'Università degli Studi di Ferrara.

Gli appunti raccolti in questo opuscolo sono stati trascritti e riorganizzati da *Farinelli Agnese*.



## La statistica

Il termine **STATISTICA** deriva dal latino medievale *status* che sta ad indicare l'ordinamento politico. È stata chiamata statistica l'area scientifica che si occupa di fornire descrizioni relative ad uno stato o descrizioni comparative di vari stati. Queste descrizioni ebbero origine in effetti già con **Aristotele** nel 350 A.C. continuando nelle epoche successive, pur tuttavia la parola statistica apparve per la prima volta nella *Enciclopedia Britannica* nel 1797 essendo stata conosciuta ufficialmente, sembra, dal professore tedesco **Gottfried Achenwall** (1719-1772) verso la metà del secolo che la definì "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. In seguito il significato della statistica si è esteso ben oltre tale interpretazione originaria.

Al giorno d'oggi si parla di due tipi di statistica:

- una statistica **descrittiva** o **metodologica**
- una statistica **matematica** o **inferenziale**

Per statistica **descrittiva** si intende il complesso delle tecniche di cui si serve lo sperimentatore per raccogliere, elaborare (cioè organizzare) e rappresentare gruppi i dati osservati che in genere sono in grandi quantità ed in ordine sparso. Lo scopo finale è quello di ricavare dalle osservazioni opportuni elementi di sintesi, quali possono essere dei *numeri* come ad esempio la *media*, la *mediana*, la *moda*, ecc... oppure *informazioni* rappresentate sotto forma di *tavole*, *grafici*, *carte* (geologiche, numerologiche, ecc...), *diagrammi*, ecc... In questa situazione si cerca però soltanto di descrivere il fenomeno osservato senza cercare di interpretare o generalizzare troppo riguardo alla totalità dei potenziali dati osservabili di cui quelli disponibili sono in genere soltanto una piccola parte. Al giorno d'oggi naturalmente queste informazioni sono molto utili anche se quasi sempre vogliamo saperne di più e spendere meno. Significa non dover censire sempre tutto per ricavare le informazioni che servono, ma estrarre le stesse dalla conoscenza dei dati che sono a nostra disposizione. In termini statistici vogliamo trarre conclusioni su tutta la popolazione partendo dal campione osservato. Ciò implica che molti dati saranno a noi ignoti e che dovremmo trovare un metodo per avere qualche conoscenza che non sia quella diretta sperimentale. Le informazioni suddette rappresentano un passo preliminare però molto importante per la successiva ricerca di un modello interpretativo che è un aspetto proprio della statistica **matematica** o **inferenziale**. Si studiano dunque gruppi di dati per decidere quale tipo di *modello* è quello che descrive al meglio l'intera popolazione. I *modelli* considerati possono essere diversi tra loro in modo sostanziale oppure soltanto per alcuni parametri, ovvero per alcune caratteristiche della popolazione. E' opportuno sottolineare fin da ora che il concetto di **PROBABILITÀ** diventa essenziale nello studio della statistica **matematica** o **inferenziale**.

## La probabilità

Per quanto riguarda la **PROBABILITÀ** diamo innanzi tutto alcuni brevi cenni storici. I primi documenti che si conservano sono dovuti a **Girolamo Cardano** (1501-1576), accanito giocatore, nel suo libro "*De Ludo Aleae*" scritto forse intorno al 1526 e pubblicato postumo nel 1673 e a **Galileo Galilei** con un suo scritto del 1620 ca. Nel libro di Cardano tra l'altro viene riferito il problema della somma dei punti ottenuti gettando tre dadi. Il risultato contiene qualche inesattezza e gli sviluppi non sono molto chiari, così che i giudizi sul valore dell'opera sono controversi. Tuttavia è fuori dubbio la sua importanza storica. Nello scritto di Galilei viene affrontato lo stesso problema del getto dei tre dadi però con maggiore chiarezza. Ma la nascita del calcolo delle probabilità viene abitualmente attribuita a **Blaise Pascal** (1623-1662) e fissata nella corrispondenza tra lui e l'altro grande matematico francese **Pierre Fermat** (1601-1665) verso la metà del secolo. L'interesse di Pascal fu molto stimolato da un certo **Cavalier de Méré**, spirito vivace, matematico discreto ed accanito giocatore d'azzardo che si rivolse spesso a lui per la risoluzione di vari problemi riguardanti il gioco. Il calcolo delle probabilità sorto quindi come appena accennato verso la metà del 1600 come studio matematico dei giochi d'azzardo come dadi, monete, carte, ecc... può essere considerato oggi come una disciplina matematica che trova applicazione nello studio degli esperimenti non deterministici o casuali. *Un esperimento non deterministico è un tipo di esperimento che può dar luogo ad un risultato tra quelli possibili non determinabile a priori in modo univoco.* Un esempio è fornito dal lancio di un dado il cui risultato è affetto da un certo grado di incertezza o indeterminazione potendo essere rappresentato da una qualunque delle sei facce del dado stesso.

Questi brevi cenni hanno lo scopo di mostrare come si presenta originariamente il concetto di

probabilità e di introdurre quelle definizioni o interpretazioni della probabilità che ora presenteremo.

### **Definizione Classica**

La prima definizione di probabilità detta perciò *classica* si ritrova già in Pascal e definisce la probabilità 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 nel senso che bisogna sapere già prima che significato dare alla probabilità perché non si capisce 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 avviene per le palline dell'urna del gioco del lotto, per le facce di un dado, per le carte di un mazzo ben mescolato. Più che una definizione questa può essere considerata una regola per misurare la probabilità, quando cioè si ha un numero finito di possibili alternative che possono essere considerate, ad esempio per motivi di simmetria, ugualmente probabili. Tutto questo però sapendo già che cosa è la probabilità. In conclusione la validità di questa definizione resta discutibile, d'altra parte è facile constatare che essa risulta essere operativa. Il campo nel quale più direttamente si può applicare la definizione classica di probabilità è quella dei giochi di dadi, carte, ecc... In essi infatti si possono individuare con precisione le diverse alternative e si può ragionevolmente pensare che esse siano ugualmente probabili. Supponiamo di voler calcolare la probabilità che lanciando una moneta, non truccata, esca testa. In questo esempio c'è un solo caso favorevole, cioè *testa*, e due casi possibili, cioè *testa* e *croce*. Si arriva pertanto alla conclusione che la probabilità che esca testa secondo la definizione classica è  $\frac{1}{2}$ .

### **Definizione Frequentista**

L'insoddisfazione della definizione classica portò a costruire la probabilità sulle *frequenze*, intendendo come *frequenza relativa dei successi* il rapporto tra il numero di volte in cui l'evento si verifica e il numero delle prove effettuate, si giunse così alla definizione frequentista della probabilità intesa come

---

il limite della frequenza relativa dei successi quando tende all'infinito il numero delle prove effettuate nelle stesse condizioni.

---

In altre parole, accettando l'ipotesi che la frequenza relativa si vada stabilizzando intorno a un certo numero all'aumentare del numero delle prove, è proprio questo numero che va preso come probabilità. Anche la definizione frequentista presenta aspetti criticabili:

- innanzi tutto, *perché resta imprecisato il numero di prove necessario per arrivare ad un valore abbastanza stabilizzato che ci fornisca la probabilità.*
- soprattutto, *le esigenze 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 a 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, seppur di poco, la situazione ambientale, cioè la *pressione*, la *temperatura*, l'*umidità*, ecc... che influisce sul risultato. Inoltre può cambiare anche la forma del dado stesso nell'urto che subisce 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 però 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 ed il numero totale delle prove, ha le stesse caratteristiche matematiche del rapporto fra il numero dei casi favorevoli ed il numero dei casi possibili. Vediamo ora un esempio:

- *se una moneta viene lanciata 1000 volte e 495 volte presenta testa, allora secondo la definizione*

frequentista la probabilità è data da  $495/1000$ .

### Definizione Soggettiva

In molti casi si considera l'esigenza di riferirsi ad un elemento singolo e di voler calcolare la probabilità proprio di questo evento. Da questa esigenza è nata la concezione soggettiva che si può far risalire a **Daniele Bernoulli** e che è stata ripresa e sviluppata recentemente dall'italiano **Bruno De Finetti** (1906-1985). Secondo 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 dalle persone che la valuta e dalle informazioni disponibili. La definizione data così com'è non è però operativa ed occorre dare 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 non si verifica con una qualsiasi unità di misura (1.00 €, 1.00 \$, 1000000, ecc...).

---

Si pensi ad esempio ad una lotteria basata sui numeri della tombola. Ognuno paga una certa quota e chi vince, cioè chi possiede il numero estratto, riceve tutto il piatto. A questo proposito supponiamo che ciascun numero della tombola costi 1.00 € per cui l'intero piatto sarà di 90.00 € che in questo caso rappresenta l'*unità di misura*. Allora la probabilità di vittoria in questo caso sarà  $1/90$ . Nella definizione però si è parlato di *prezzo equo* e questo richiede una precisazione, che viene data dalla seguente condizione detta di **equità** o di **coerenza**, cioè non devono essere valutate 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 la probabilità  $P\{T\}$  di ottenere testa è uguale a  $\frac{1}{2}$  quando si esegue un lancio mentre quella di ottenere croce  $P\{C\}$  è uguale a  $\frac{1}{4}$ . Si può obiettare subito che ciò non può andare perché la probabilità della somma di  $P\{T\}$  e  $P\{C\}$  deve essere uguale a 1. Questa è una esigenza intuitiva alla quale la condizione di equità dà una giustificazione convincente. Infatti facendo contemporaneamente due scommesse, una su testa e l'altra su croce, si pagherebbe  $\frac{1}{2}$  per la prima più  $\frac{1}{4}$  per la seconda, ricevendo comunque 1 con un guadagno netto di  $\frac{1}{4}$  in contraddizione con la condizione di equità e questo perché la somma di  $P\{T\}$  più  $P\{C\}$  è minore di 1. Se tale somma fosse maggiore di 1 si avrebbe una perdita certa. La condizione di coerenza impone quindi che sia  $P\{T\} + P\{C\} = 1$ . Lo stesso ragionamento vale in presenza di più di due alternative. La critica più diffusa a questa impostazione è proprio quella di essere soggettiva, cioè di fondare la probabilità sul valore della persona che la valuta. 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 diverse dal punto di vista concettuale. Si è visto inoltre che tutte e tre le impostazioni non sono certamente esenti da critiche. Per ovviare a questa difficoltà i matematici preferiscono trattare la probabilità da un punto di vista assiomatico facendo uso della **teoria degli insiemi**. L'impostazione **assiomatica** o **ipotetico-deduttiva** 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.

### Introduzione Alla Definizione Assiomatica

Le tre definizioni date della probabilità portano tutte alle stesse leggi matematiche espresse dagli assiomi della probabilità e ciò rende naturale prendere tali leggi come base per una costruzione matematica. La teoria matematica si può quindi sviluppare a partire da queste relazioni assunte come assiomi senza precisare la definizione di probabilità da cui esse provengono. Si può adottare questa linea di azione ottenendo così una costruzione matematica che non esclude chi si sente di aderire ad una particolare impostazione piuttosto che ad un'altra.

Lo statistico è per lo più interessato alla probabilità soltanto per quanto riguarda i possibili risultati

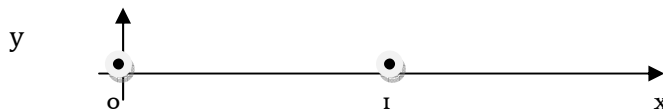
degli esperimenti e di solito preferisce l'*interpretazione frequentista della probabilità*, cioè preferisce pensare alla probabilità come alla frequenza del verificarsi di un certo evento se l'esperimento ad esso relativo fosse ripetuto un gran numero di volte. Inoltre, la maggior parte degli statistici sono interessati soltanto agli *esperimenti di tipo ripetitivo*, come il lancio di una moneta, il getto di un dado, la lettura della temperatura giornaliera su di 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 esame. Per gli esperimenti di tipo ripetitivo il modello scelto, di solito, è un modello per prevedere la frequenza del verificarsi di un certo risultato in ripetute esecuzioni dell'esperimento in esame. 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, di controllarne poi l'attendibilità e di trarre quindi da esso le conclusioni definitive o risolutive.

**Il nostro approccio alla probabilità si baserà sia sull'interpretazione frequentista che su quella assiomatica.** In vista della impostazione assiomatica, tenuto conto che abbiamo già discusso seppur brevemente il concetto di probabilità, restano da chiarire quelli di **esperimento o prova** e di **evento**, che essendo concetti intuitivi sono già stati utilizzati senza discuterli.

### Spazio Campione o Spazio Campionario

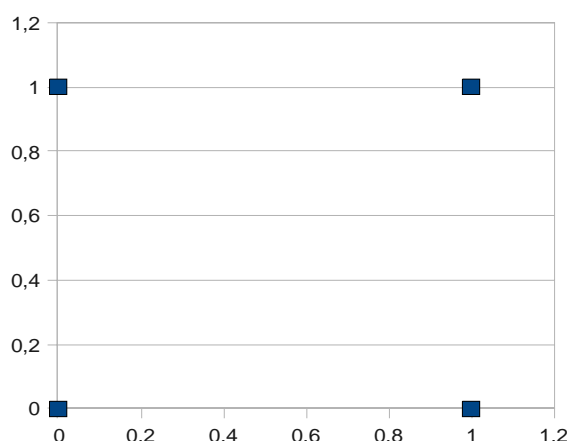
Dato un esperimento casuale o non deterministico, un concetto di base è quello di **SPAZIO CAMPIONE**. Per definirlo diciamo subito che è conveniente rappresentare i possibili risultati di un esperimento in generale per mezzo di punti in uno spazio ad  $n$  dimensioni con  $n=1,2,3,4,\dots$ . A tale scopo consideriamo il semplice esperimento del lancio di una moneta. In esso i possibili risultati sono due, testa e croce. In questo caso è conveniente rappresentare il risultato testa con il punto di ascissa 1 sull'asse  $x$  ed il risultato croce col punto di ascissa 0 come in fig.1:

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 una moneta due volte, i risultati possibili sarebbero stati quattro, cioè TT,TC,CT,CC. In questo caso per ragioni di simmetria sarebbe conveniente rappresentare i risultati per mezzo dei punti di coordinate  $(1,1),(1,0),(0,1),(0,0)$  nel piano  $xOy$  come illustrato nella fig.2:

Fig.2



Se la moneta fosse lanciata tre volte sarebbe conveniente usare lo spazio tridimensionale per rappresentare gli otto possibili risultati sperimentali che sarebbero così localizzati nei vertici di un cubo di lato 1. E' importante osservare che queste rappresentazioni sono semplicemente una convenienza, ad esempio nell'esperimento del lancio di due monete per rappresentare quattro possibili risultati se desiderato si potrebbero pure segnare quattro punti qualsiasi sull'asse  $x$ .

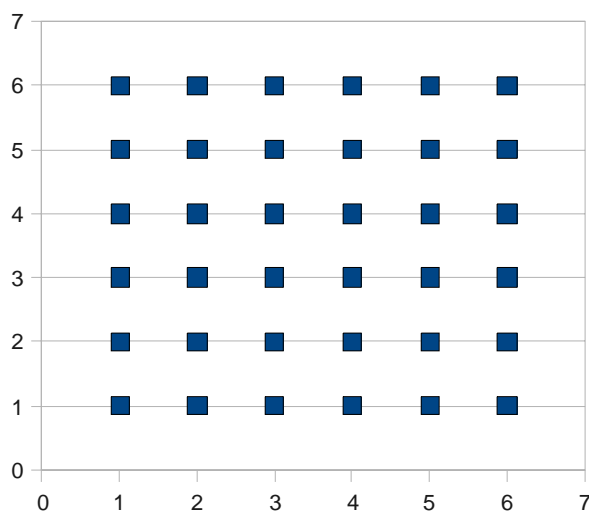
Nell'esperimento del getto di due dadi ci sono 36 possibili risultati, indicati in tab.1 dove il primo numero di ciascuna coppia indica il numero che esce su un dado ed il secondo numero quello che esce sull'altro dado. Nell'ipotesi che i due dadi siano distinguibili oppure siano gettati nell'ordine:

**Tab.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

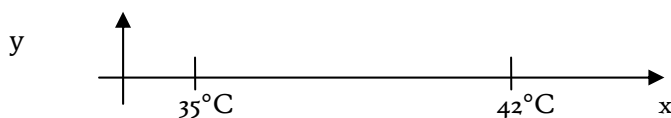
Per questo esperimento è conveniente rappresentare i possibili risultati con i 35 punti nel piano xOy le cui coordinate sono le corrispondenti coppie di numeri di tab.1. Questa scelta è mostrata in fig.3:

**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 in modo conveniente con i punti dell'intervallo di estremi 35 e 42 sull'asse x come in fig.:

**Fig.4**



Ciò naturalmente non tiene conto dell'impossibilità di leggere un termometro con una accuratezza illimitata. Possiamo dare allora la seguente definizione:

---

**l'insieme dei punti che rappresentano i possibili risultati di un esperimento è chiamato SPAZIO CAMPIONE o SPAZIO CAMPIONARIO dell'esperimento.**

---

Come si è già osservato è importante sottolineare di nuovo 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 risultati 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 suo risultato si può decidere se un evento indicato con  $A$  si è verificato. Ciò significa che ciascun *punto 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 TC e CT di fig.1 corrispondono al verificarsi di  $A$ . Se  $A$  è l'evento di ottenere un totale di sette punti nel gettare due dadi allora  $A$  è associato ai sei punti campione 16,25,34,43,52,61 di fig.2. Se  $A$  è l'evento che la temperatura dell'individuo sia almeno  $38^{\circ}\text{C}$  allora  $A$  sarà associato all'intervallo di punti da 38 a 42 sull'asse  $x$ . Si può dare allora la seguente definizione:

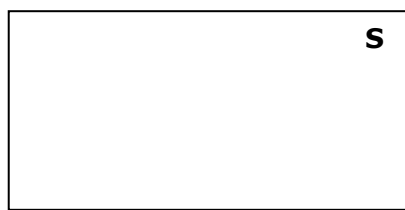
---

un evento è un sottoinsieme di uno spazio campione.

---

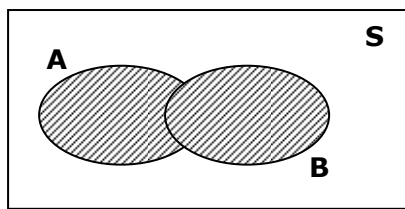
Poiché un sottoinsieme di un insieme di punti comprende la possibilità che un sottoinsieme coincida con l'intero insieme di punti, oppure che esso non contenga nessun punto dell'insieme, questa definizione comprende un evento che è certo di verificarsi oppure un evento che non può assolutamente verificarsi quando l'esperimento 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 qualunque sia il numero di punti in esso contenuti, viene rappresentato per mezzo di un insieme di punti interni ad un rettangolo in un piano come nella fig. seguente dove con  $S$  si indica lo spazio campione:

Fig.5



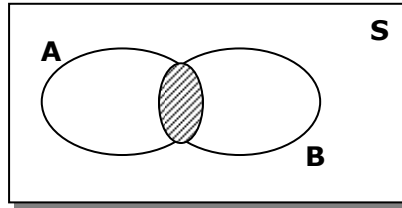
Un evento  $A$  che è perciò un sottoinsieme di punti in questo rettangolo è rappresentato dai punti che giacciono all'interno di una curva chiusa contenuta nel rettangolo. Se  $B$  è qualche altro evento di interesse, esso sarà rappresentato dai punti interni ad un'altra curva chiusa nel rettangolo. Questa rappresentazione è mostrata in fig. 6:

Fig.6



Allora, se  $A$  e  $B$  sono due eventi associati ad un esperimento si può voler sapere se almeno uno degli eventi si verificherà quando l'esperimento viene eseguito. L'insieme di punti che appartengono o ad  $A$  o a  $B$  o sia ad  $A$  che a  $B$  è chiamato **l'unione di  $A$  e  $B$**  e si indica con il simbolo  $A \cup B$ . Questo insieme di punti è rappresentato in fig.2 dalla regione tratteggiata. Come esempio, se  $A$  è l'evento di ottenere un 6 sul primo dado, 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 nel gettare 2 dadi. L'evento  $A$  consiste nei 6 punti le cui coordinate sono le coppie di valori dell'ultima colonna della tab.1 vista precedentemente e  $B$  è l'insieme dei punti le cui coordinate sono le coppie di valori dell'ultima riga della tabella suddetta. L'evento  $A \cup B$  è allora l'insieme degli 11 punti le cui coordinate sono le coppie di numeri dell'ultima colonna e dell'ultima riga della tabella suddetta. Un altro evento di possibile interesse è di sapere se entrambi gli eventi si verificheranno quando l'esperimento viene eseguito. L'insieme di punti che consiste di tutti i punti che appartengono sia ad un insieme  $A$  che ad un insieme  $B$  è chiamato **l'intersezione di  $A$  e  $B$**  e si indica con il simbolo  $A \cap B$ . Questo evento è rappresentato in fig.7 dalla regione tratteggiata:

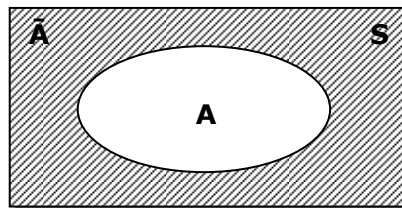
Fig.7



Nel esempio del getto di due dadi dove A e B rappresentano gli stessi due eventi appena indicati, allora  $A \cap B$  è l'evento di ottenere un 6 su entrambi i dadi. Esso è rappresentato dall'unico punto di coordinate (6,6) che è l'intersezione dei sottoinsiemi dello spazio campionario indicati nell'ultima colonna e nell'ultima riga della tab.1 vista precedentemente.

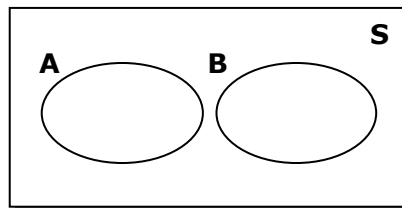
In corrispondenza a qualsiasi evento A esiste un evento ad esso associato indicato con  $\bar{A}$  che stabilisce che A non si verificherà quando l'esperimento viene eseguito. Esso è rappresentato da tutti i punti del rettangolo che non si trovano in A e nella figura seguente è rappresentato dalla regione tratteggiata.

Fig.8



L'evento  $\bar{A}$  è chiamato il **complementare di A relativo allo spazio campionario**. Se due insiemi A e B non hanno punti in comune si dice che sono **disgiunti**. Nel linguaggio degli eventi si parla di *eventi disgiunti* oppure di *eventi reciprocamente esclusivi* perché il verificarsi dell'uno esclude la possibilità del verificarsi dell'altro. Due eventi di questo tipo sono rappresentati nella fig. seguente:

Fig.9



Fissiamo ora l'attenzione sulle **funzioni d'insieme**, perché ci servono per definire la probabilità. Le funzioni che ci sono più familiari sono le *funzioni di punto*, ad esempio  $f(x)=x^2$  è una funzione di questo tipo perché, a ciascun punto dell'asse x, questa formula assegna 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 possono essere anziché punti singoli insiemi di punti. In questo caso la funzione prende il nome di **funzione di insieme**. Un esempio di funzione di questo tipo è quello in cui il dominio consiste di intervalli sull'asse x ed essa fornisce la lunghezza dell'intervallo.

La **probabilità** può essere considerata un modello per la frequenza del verificarsi di un certo risultato in ripetute esecuzioni di un esperimento, perciò un modello di probabilità per un evento A dovrebbe essere un modello per cui la probabilità del suo verificarsi indicata con  $P\{A\}$  dovrebbe essere uguale alla frequenza del verificarsi dell'evento stesso in ripetute esecuzioni dell'esperimento. 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 questa probabilità. 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 delle previsioni sui futuri esperimenti. Se le probabilità degli eventi sono da interpretare come modelli per le frequenze del verificarsi di quegli eventi in ripetute esecuzioni dell'esperimento, tali probabilità dovrebbero possedere le proprietà essenziali delle frequenze così una probabilità dovrebbe essere un numero



compreso tra 0 e 1 perché una frequenza è n numero di questo tipo. Inoltre la probabilità dell'evento S dove S rappresenta lo spazio campione, dovrebbe essere uguale a 1 perché uno qualunque dei risultati possibili si verifica sicuramente 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 verificata se saranno verificate le seguenti tre condizioni che si basano su quanto appena detto; queste sono chiamate gli **assiomi della probabilità**. Essi impongono restrizioni sul tipo di funzione di insieme P che si può usare per calcolare le probabilità degli eventi. Una tale funzione viene chiamata una **misura della probabilità**. Diamo ora gli assiomi della probabilità:

- Una misura di probabilità P è una funzione di insieme a valori reali definita su di uno spazio campione che soddisfa
  - I.  $0 \leq P\{A\} \leq 1$
  - II.  $P\{S\} = 1$
  - III.  $P\{A_1 \cup A_2 \cup \dots\} = P\{A_1\} + P\{A_2\} + \dots$

per ogni successione finita o infinita di eventi disgiunti  $A_1, A_2, \dots$ .

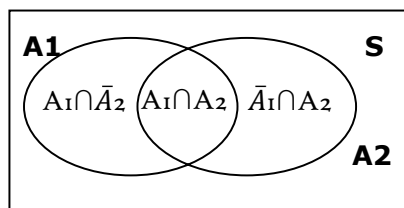
Si fa osservare che gli assiomi della probabilità sono stati formulati nel 1933 da **A.N. Kolmogorow**. Sarebbe molto difficile trovare una funzione P che fornisca valori corrispondenti alle frequenze attese per ogni possibile sottoinsieme A di uno spazio campione, perché il numero di tali sottoinsiemi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

### Regola di addizione delle probabilità

Spesso il calcolo delle probabilità interessa un certo numero di eventi in relazione fra loro anziché 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 quando l'esperimento viene eseguito oppure se si verificherà almeno uno di essi. Per quanto detto in precedenza, il primo di questi due eventi composti è rappresentato da  $A_1 \cap A_2$  ed il secondo da  $A_1 \cup A_2$ . Per dare una risposta riguardante la frequenza con cui si verificheranno questi eventi composti è necessario conoscere i valori delle loro probabilità e cioè  $P\{A_1 \cap A_2\}$  e  $P\{A_1 \cup A_2\}$ . Ricaviamo ora una formula per calcolare  $P\{A_1 \cup A_2\}$ .

Lo spazio campione di un esperimento sia rappresentato come al solito dai punti interni al rettangolo nella figura seguente e i punti corrispondenti al verificarsi degli eventi  $A_1$  e  $A_2$  siano rispettivamente i punti interni alle regioni definite con  $A_1$  e  $A_2$ .

Fig.1



Allora, come si è già detto, l'evento  $A_1 \cup A_2$  consiste di tutti i punti che giacciono all'interno di questa regione. 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 della probabilità. Dalla fig.1 si può osservare che  $A_1 \cup A_2$  è l'unione di tre insiemi disgiunti e precisamente  $A_1 \cap \bar{A}_2$ ,  $A_2 \cap \bar{A}_1$ , perciò per il terzo assioma si può scrivere che

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

Analogamente sia che

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

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

Ora, ricavando da queste due ultime 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** data dalla

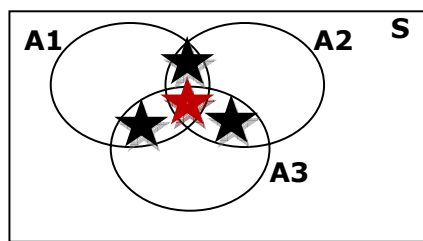
$$(2) P\{A1 \cup A2\} = P\{A1\} + P\{A2\} - P\{A1 \cap A2\}$$

I due eventi A1 e A2 possono non avere punti campione in comune. Come già detto gli eventi si diranno allora disgiunti e in questo caso la formula (2) si riduce alla seguente

$$(3) P\{A1 \cup A2\} = P\{A1\} + P\{A2\}$$

La (3) naturalmente è la conseguenza nel precedente assioma. Questa regola rappresentata dalla (2) si può generalizzare al caso di più eventi, ad esempio nel caso di 3 eventi A1, A2, A3 arbitrari si ha che  $P\{A1 \cup A2 \cup A3\} = P\{A1\} + P\{A2\} + P\{A3\} - P\{A1 \cap A2\} - P\{A1 \cap A3\} - P\{A2 \cap A3\} + P\{A1 \cap A2 \cap A3\}$ .

Fig.2



La regola di addizione è applicabile a qualunque tipo di spazio campione per il quale è assegnata una misura P della probabilità. Per usarla occorre però conoscere i valori della probabilità a secondo membro delle formule. Questo è particolarmente semplice quando lo spazio campione contiene soltanto un numero finito di punti. Perciò per il momento restringeremo la discussione a spazi campione di questo tipo. Quindi la prima cosa da fare per calcolare la probabilità di un evento A per uno spazio campione finito è di assegnare il valore della probabilità a ciascun punto campione che naturalmente deve obbedire ai primi due assiomi della probabilità, cioè questi valori devono essere numeri non negativi compresi tra 0 e 1 la cui somma deve essere uguale a 1. Queste assegnazioni si possono fare sulla base dell'esperienza dei singoli, di considerazioni di simmetria, di informazioni esterne e in generale su qualsiasi informazione disponibile. Così, ad esempio, sarebbe realistico da parte di considerazioni di simmetria assegnare la probabilità di 1/36 a ciascun punto dello spazio campione considerato per l'esperimento del getto di due 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 chiamati spesso **eventi semplici**. Questi eventi verranno indicati con  $e_1, e_2, \dots, e_n$ . E' chiaro che essi sono disgiunti. Ora, qualsiasi evento A è un insieme di punti campione e perciò è l'unione degli eventi semplici corrispondenti. L'applicazione del terzo assioma perciò fornisce che

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

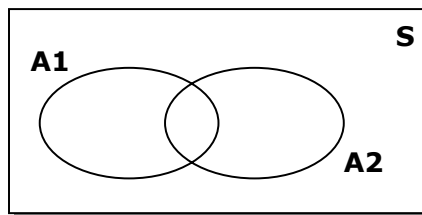
dove la sommatoria viene fatta su quelle probabilità  $p_i$  associate ai punti campione che giacciono in A. Per molti giochi d'azzardo lo spazio campione è non soltanto finito ma ai punti campione viene assegnata la stessa probabilità. Questo è vero per esempio nel getto di due dati per cui lo spazio campione già presentato risulta conveniente. A ciascuno di quei punti campione viene assegnata la probabilità di 1/36. Se n indica il numero totale di punti campione e N(A) indica il numero dei punti campione nell'insieme A, allora poiché si è assunto  $p_i = 1/n$  per  $i = 1, \dots, n$  allora la formula (4) si ridurrà alla

$$(5) P\{A\} = (1/n) * N(A)$$

### Probabilità condizionata

Supponiamo ora di voler sapere se un evento A2 si verificherà sotto la condizione che un evento A1 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 illustrata in fig.3:

Fig.3



Poiché A1 è certo di verificarsi soltanto quando lo spazio campione è ristretto a quei punti che giacciono entro la regione A1 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 A1 fosse stata assegnata ad esempio una probabilità doppia rispetto ad un altro punto in A1, allora si dovrebbe assegnare una probabilità doppia anche nel nuovo spazio campione, perché ignorare i risultati che non producono l'evento A1 non deve alterare il rapporto delle frequenze attese per quei due punti campione. È semplicemente necessario perciò moltiplicare le probabilità originarie assegnate ai punti in A1 per un fattore costante c tale che la somma delle nuove probabilità sia uguale a 1, così se  $\pi_i$  rappresenta la nuova probabilità corrispondente a  $\pi_i$  nell'assegnazione originaria, si dovrebbe scegliere  $\pi_i = c \cdot \pi_i$  dove la somma delle probabilità  $\pi_i$  fatte sui punti di A1, che è uguale a c per la sommatoria su A1 di  $\pi_i$  che è uguale a c per la somma delle  $\pi_i$  su A1, e questo deve essere uguale a 1 da cui si ricava che  $c = 1/P\{A1\}$  e quindi  $\pi_i = \pi_i/P\{A1\}$ . 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 A1.

Se le probabilità che l'evento A2 si verifichi sotto la condizione che debba verificarsi l'evento A1 si indica con  $P\{A2|A1\}$  allora si ha che  $P\{A2|A1\} = \sum_{A1 \cap A2} \pi_i = \frac{\sum_{A1 \cap A2} \pi_i}{P\{A1\}}$ . La prima somma viene fatta su quelle  $\pi_i$  corrispondenti a punti campione che giacciono in  $A1 \cap A2$  perché essi sono i soli punti campione in A1 che corrispondono al verificarsi di A2. Poiché la somma a numeratore nell'ultima espressione è quella che definisce  $P\{A1 \cap A2\}$ , allora segue che la formula che il calcolo della probabilità condizionata è data da

$$(6) P\{A2|A1\} = P\{A1 \cap A2\} / P\{A1\}$$

Si fa l'ipotesi qui che  $P\{A1\} > 0$ .

Nel procedimento per arrivare alla (6) si è assunto che lo spazio campione contenga soltanto un numero finito di punti, ma questa formula si prenderà, come definizione di probabilità condizionata, anche per spazi campione più generali. Si può dimostrare, infatti, che  $P\{A2|A1\}$  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

$$(7) P\{A1 \cap A2\} = P\{A1\}P\{A2|A1\}$$

Se si scambia l'ordine dei due eventi la formula (7) diventa

$$(8) P\{A1 \cap A2\} = P\{A2\}P\{A1|A2\}$$

### Eventi indipendenti

Supponiamo che A1 e A2 siano due eventi tali che  $P\{A2|A1\} = P\{A2\}$  con  $P\{A1\}P\{A2\} > 0$ , allora l'evento A2 si dice **indipendente nel senso della probabilità dall'evento A1** o più brevemente **indipendente da A1**. Ciò deriva dalla proprietà che la probabilità del verificarsi di A2 non venga modificata imponendo la condizione che debba verificarsi A1. Quando A2 è indipendente da A1, la regola di moltiplicazione espressa dalla (7) diventa

$$(9) P\{A1 \cap A2\} = P\{A1\}P\{A2\}$$

Viceversa, quando è vera la (9), segue dal confronto della (9) con la (7) che  $A2$  è indipendente da  $A1$ . Se si confrontano i secondo membri della (9) e della (8) si può osservare che  $P\{A1|A2\} = P\{A1\}$  che stabilisce che l'evento  $A1$  è indipendente da  $A2$ . Così, se  $A2$  è indipendente da  $A1$ , segue che  $A1$  deve essere indipendente da  $A2$ . A causa di questa reciproca indipendenza e poiché la (9) implica questa indipendenza, si suole definire l'indipendenza come segue:

---

due eventi  $A1$  e  $A2$  si dicono indipendenti se  $P\{A1 \cap A2\} = P\{A1\}P\{A2\}$ .

---

### Esercitazione

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

**Problema 1.** *Nell'ipotesi che una famiglia abbia 2 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  $S$  sia dato da  $S = \{(m,m), (m,f), (f,m), (f,f)\}$  dove, ad esempio,  $(m,f)$  sta ad indicare che il figlio maggiore è una femmina e il minore è un maschio e che tutti gli elementi dello spazio siano ugualmente probabili. Ora indichiamo con  $A1$  l'evento che entrambi i figli siano maschi, cioè  $A1 = \{(m,m)\}$  e  $A2$  l'evento che almeno uno dei figli sia maschio, cioè  $A2 = \{(m,m), (m,f), (f,m)\}$ . La probabilità richiesta è quindi data da*

$$\begin{aligned} P\{A1|A2\} &= P\{A1 \cap A2\} / P\{A2\} = \\ &= P\{(m,m)\} / P\{(m,m), (m,f), (f,m)\} = \\ &= (1/4) / (3/4) = \\ &= 1/3 = \\ &= 33\%. \end{aligned}$$

**Problema 2.** *Supponiamo che uno studente affronti due esami e che la probabilità di superare il primo esame sia 0.6, che la probabilità di superare il secondo sia 0.8 e che la probabilità di superarli entrambi sia 0.5. Si vuole calcolare la probabilità che egli superi almeno uno dei due esami e la probabilità di superarli entrambi. A tale scopo indichiamo con  $A1$  l'evento corrispondente al superamento del primo esame e con  $A2$  l'evento corrispondente al superamento del secondo esame. Allora  $A1 \cap A2$  è l'evento che lo studente superi entrambi gli esami e  $A1 \cup A2$  è l'evento che egli ne superi almeno uno. Si ha quindi che*

$$\begin{aligned} P\{A1 \cup A2\} &= P\{A1\} + P\{A2\} - P\{A1 \cap A2\} = \\ &= 0.6 + 0.8 - 0.5 = \\ &= 0.9 = \\ &= 90\% \end{aligned}$$

*L'evento che lo studente fallisca in entrambi gli esami è rappresentato da  $A1 \cup A2$  (complementare) e quindi*

$$\begin{aligned} P\{A1 \cup A2\} &= 1 - P\{A1 \cup A2\} = \\ &= 1 - 0.9 = \\ &= 0.1 = \\ &= 10\% \end{aligned}$$

*A questo punto si vuole calcolare inoltre la probabilità che lo studente superi il secondo esame supposto di aver già superato il primo, e la probabilità di superare il primo supposto di aver già superato il secondo. Allora la prima probabilità richiesta è data da*

$$\begin{aligned} P\{A2|A1\} &= P\{A1 \cap A2\} / P\{A1\} = \\ &= 0.5 / 0.6 = \\ &= 0.8333333... \end{aligned}$$

*e la seconda probabilità è data da*

$$\begin{aligned} P\{A1|A2\} &= P\{A1 \cap A2\} / P\{A2\} = \\ &= 0.5 / 0.8 = \\ &= 0.625 \end{aligned}$$

Ciò mostra che la probabilità di superare un esame dopo aver già superato l'altro aumenta rispetto alla probabilità di superare soltanto quell'esame. La spiegazione di ciò sta nel fatto che presentarsi ad un esame dopo averne già superato un altro aumenta la fiducia dello studente e quindi la sua prestazione.

A questo punto ci si chiede se gli eventi  $A_1$  e  $A_2$  siano indipendenti. Dal momento che  $P\{A_1 \cap A_2\} = 0.5$  e  $P\{A_1\}P\{A_2\} = 0.6 \cdot 0.8 = 0.48$  e quindi  $P\{A_1 \cap A_2\} > P\{A_1\}P\{A_2\}$  i due eventi non risultano indipendenti. Poiché  $P\{A_1 \cap A_2\} > P\{A_1\}P\{A_2\}$  si ricava che la promozione in un esame aiuta lo studente a superare l'altro e quindi i due eventi non sono indipendenti.

**Problema 3.** Supponiamo che da un'urna contenente 7 palline nere e 5 palline bianche vengano estratte 2 palline senza sostituzione delle palline estratte, cioè, come si dice, senza REIMBUSSULAMENTO. Si vuole calcolare la probabilità che ambedue le palline estratte siano nere, nell'ipotesi che ogni pallina nell'urna abbia la stessa probabilità di essere estratta. A tale scopo indichiamo con  $A_1$  l'evento che la prima pallina estratta sia nera e con  $A_2$  l'evento che la seconda pallina estratta sia anch'essa nera. Ora, nell'ipotesi che la prima pallina estratta sia nera restano nell'urna 6 palline nere e 5 bianche e quindi  $P\{A_2|A_1\} = 6/11$ . Tenuto conto che  $P\{A_1\} = 7/12$  si ha che

$$\begin{aligned} P\{A_1 \cap A_2\} &= P\{A_1\}P\{A_2|A_1\} = \\ &= 7/12 \cdot 6/11 = \\ &= 42/123 = \\ &= 0.138 = \\ &= 31,8\% \end{aligned}$$

**Problema 4.** Consideriamo il lancio di due dadi simmetrici. Sia  $A_1$  l'evento che la somma dei due numeri ottenuti sia uguale a 6 e sia  $C$  l'evento che il primo dado mostri 4. Ci si chiede se questi due eventi siano indipendenti. Allora l'evento  $A_1 \cap C$  è rappresentato dall'unico risultato  $(4,2)$ , cioè  $A_1 \cap C = \{(4,2)\}$ . Si ha quindi che  $P\{A_1 \cap C\} = 1/36$ . Si ha anche che  $P\{A_1\}P\{C\} = (5/36)(1/6) = 0,027$ . Poiché  $P\{A_1 \cap C\}$  è diverso da  $P\{A_1\}P\{C\}$ , si conclude che gli eventi  $A_1$  e  $C$  non sono indipendenti. In effetti se ad esempio il primo dado mostrasse un 6, l'evento  $A_1$  sarebbe impossibile. Se invece indichiamo con  $A_2$  l'evento di ottenere come somma 7 e si vuole determinare se  $A_2$  e  $C$  siano indipendenti, allora si ha che  $A_2 \cap C$  è rappresentato dall'unico risultato  $(4,3)$ , cioè  $A_2 \cap C = \{(4,3)\}$  e quindi  $P\{A_2 \cap C\} = 1/36$ . Poiché  $P\{A_2\}P\{C\} = (6/36)(1/6) = 1/36$ . Poiché in questo caso  $P\{A_2 \cap C\} = P\{A_2\}P\{C\}$ , i due eventi  $A_2$  e  $C$  sono indipendenti.

### Probabilità delle cause

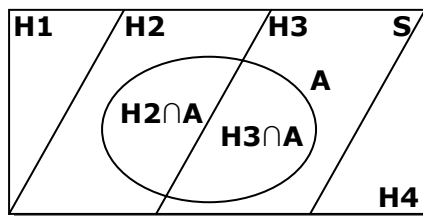
Risolviamo ora il problema seguente che, oltre ad essere un'applicazione delle formule fin qui esaminate, ci introduce inoltre al calcolo della probabilità delle cause del verificarsi di un evento. A tale scopo una scatola contiene due palline rosse e una seconda scatola identica contiene una pallina rossa e una bianca. Allora, se si sceglie una scatola e si estrae da essa una pallina, si vuole calcolare la probabilità di avere scelto la prima scatola se la pallina estratta risulta essere rossa. Indichiamo con  $A_1$  l'evento di scegliere la prima scatola e con il complementare di  $A_1$ , ovvero  $\bar{A}_1$ , l'evento di scegliere la seconda scatola. Indichiamo poi con  $A_2$  l'evento di estrarre una pallina rossa e con il complementare di  $A_2$ ,  $\bar{A}_2$ , l'evento di estrarre una pallina bianca. Il problema allora è quello di calcolare la probabilità condizionata  $P\{A_1|A_2\} = P\{A_1 \cap A_2\}/P\{A_2\}$ . La probabilità al numeratore a secondo membro si può calcolare usando la regola di moltiplicazione delle probabilità espressa dalla formula (2) vista in precedenza, intendendo che scegliere una scatola a caso significa che la probabilità di scegliere ad esempio la prima scatola è uguale a  $1/2$ . Si ha così che  $P\{A_1 \cap A_2\} = P\{A_1\}P\{A_2|A_1\} = (1/2)1 = 1/2$ . La probabilità al denominatore  $P\{A_2\}$  si può calcolare osservando che l'evento  $A_2$  si verificherà se e soltanto se si verificherà uno dei due eventi disgiunti  $A_1 \cap A_2$  e  $\bar{A}_1 \cap A_2$ . Così in questo caso, per quanto è già stato detto, si ha che  $P\{A_2\} = P\{A_1 \cap A_2\} + P\{\bar{A}_1 \cap A_2\}$  ma, sempre per la regola di moltiplicazione,  $P\{\bar{A}_1 \cap A_2\} = P\{\bar{A}_1\}P\{A_2|\bar{A}_1\} = (1/2)(1/2) = 1/4$  e  $P\{A_2\} = 1/2 + 1/4 = 3/4$ . Si ha quindi infine che  $P\{A_1|A_2\} = (1/2)/(3/4) = 2/3$ . Questo esempio è stato presentato come applicazione pratica delle formule fondamentali della probabilità. Tuttavia esso si sarebbe potuto risolvere più semplicemente considerando lo spazio campione dell'esperimento e applicando una formula già nota, cioè  $P\{A\} = \sum_A p_i$ . Per il problema in esame sarebbe conveniente considerare come spazio campione quello costituito dai 4 punti  $I_1, I_2, I_1, I_2$  dove il numero romano indica il numero della scatola e l'altro il numero della pallina. A questi 4 punti si deve assegnare ovviamente la stessa probabilità pari ad  $1/4$ . La condizione

del verificarsi dell'evento A2 restringe lo spazio campione soltanto ai primi 3 punti se il num.2 della pallina nella II scatola indica la pallina bianca. Così a ciascuno di questi 3 punti va assegnata la probabilità di 1/3, però soltanto i primi due punti corrispondono al verificarsi di A1, cioè alla scelta della prima scatola, quindi applicando la formula  $P\{A\}=\sum_A p_i$ , si ottiene che  $P\{A1|A2\}=1/3+1/3=2/3$ .

### 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 considerato sono due le possibili cause dell'estrazione di una pallina rossa ed il problema è di calcolare 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 sono tuttavia sufficientemente estesi da valer la pena di ricavare una formula per trattare tali problemi in modo sistematico. A tale scopo supponiamo che  $H_1, H_2, \dots, H_n$  siano eventi reciprocamente esclusivi con  $P\{H_i\} > 0$  per  $i=1, \dots, n$  e tali che  $\bigcup_{i=1}^n H_i = S$  dove  $S$  è lo spazio campione. Si dice anche che gli eventi  $H_1, \dots, H_n$  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 l'esperimento viene eseguito e consideriamo il problema di calcolare la probabilità che  $H_i$  sia la causa del verificarsi dell'evento  $A$ . Una suddivisione dello spazio campione e la sua relazione con l'evento  $A$  è mostrata il fig.1 dove lo spazio campione  $S$  è suddiviso per semplicità in 4 sottoinsiemi.

Fig.1



Si tratta quindi di calcolare  $P\{H_i|A\}$ . Dalla formula per il calcolo della probabilità condizionata, questa è data dalla

$$(1) P\{H_1|A\} = P\{H_1 \cap A\} / P\{A\}$$

La regola di moltiplicazione fornisce

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

per cui si può scrivere

$$(3) P\{H_i|A\} = P\{H_i\}P\{A|H_i\} / P\{A\}$$

Dalla fig.1 è chiaro che l'evento  $A$  in generale è l'unione degli eventi disgiunti  $H_1 \cap A, H_2 \cap A, \dots, H_n \cap A$  e alcuni degli insiemi corrispondenti a questi eventi possono naturalmente essere vuoti come accade in figura per gli insiemi  $H_1 \cap A$  e  $H_4 \cap A$ . Per il terzo assioma della probabilità si può quindi distinguere che  $P\{A\} = \sum_{j=1}^n P\{H_j \cap A\}$ . Se si applica la (2) a ciascun termine a secondo membro di quest'ultima si ottiene che

$$(4) P\{A\} = \sum_{j=1}^n P\{H_j\}P\{A|H_j\}$$

La sostituzione della (4) nella (3) fornisce la formula richiesta per calcolare la probabilità delle cause nota come **formula di Bayes** ed espressa quindi dalla

$$(5) P\{H_i|A\} = P\{H_i\}P\{A|H_i\} / \sum_{j=1}^n P\{H_j\}P\{A|H_j\}$$

per  $i=1, \dots, n$ . Ambedue le formule, (4) e (5), sono utili nel descrivere gli esperimenti che procedono in due fasi ed hanno la proprietà che il meccanismo di aleatorietà della seconda fase è determinato dal

risultato della prima fase dell'esperimento. Questi esperimenti sono anche chiamati **esperimenti composti**.

Nell'applicazione delle formule (4) e (5) viste prima, agli esperimenti composti  $H_i$  rappresentano i possibili risultati della prima fase dell'esperimento e  $P\{A|H_i\}$  descrive il meccanismo di aleatorietà della seconda fase, sotto l'ipotesi che  $H_i$  si sia verificato nella prima fase. La probabilità  $P\{H_i\}$  viene anche chiamata **probabilità a priori**, mentre la probabilità condizionata  $P\{H_i|A\}$  viene anche chiamata **probabilità a posteriori**. Queste denominazioni derivano dal fatto che in certe applicazioni  $P\{H_i\}$  sono probabilità soggettive che rappresentano la nostra opinione prima che l'esperimento venga eseguito mentre le  $P\{H_i|A\}$  si possono interpretare 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 sarà illustrato nel seguito esaminando un esempio specifico.

### Esercitazione

Consideriamo ora alcuni esempi come applicazione pratica delle formule (4) e (5) viste prima.

**Esempio 1.** Per illustrare l'uso diretto della formula di Bayes risolviamo ora il problema esaminato prima per introdurre il calcolo della probabilità delle cause. Indichiamo con  $H_1$  e  $H_2$  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 dai contenuti delle due scatole che  $P\{A|H_1\}=1$  e  $P\{A|H_2\}=1/2$ . La formula di Bayes fornisce allora

$$\begin{aligned} P\{H_1|A\} &= P\{H_1\}P\{A|H_1\} / P\{H_1\}P\{A|H_1\} + P\{H_2\}P\{A|H_2\} = \\ &= (1/2) * 1 / (1/2) * 1 + (1/2)(1/2) = \\ &= (1/2)(3/4) = \\ &= 2/3 \end{aligned}$$

**Esempio 2.** 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 malattie  $\{H_1, H_2, \dots, H_k\}$  sia la causa più probabile. Si suppone di avere una stima statistica delle probabilità  $P\{H_i\}=p_i$  con  $p_i > 0$  di contrarre la malattia  $H_i$  per  $i=1, \dots, k$ . Si suppone inoltre che le varie malattie non siano contemporaneamente presenti nella stessa persona. Infine si suppone di conoscere una stima statistica della probabilità condizionata  $P\{A|H_i\}$ , ossia della probabilità che la malattia  $H_i$  dia origine ai sintomi  $A$ . Per la soluzione si procede come segue: 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). Come esempio supponiamo  $P\{H_1\}=0,4$ ;  $P\{H_2\}=0,25$  e  $P\{H_3\}=0,35$  e  $P\{A|H_1\}=0,8$ ;  $P\{A|H_2\}=0,6$  e  $P\{A|H_3\}=0,9$ . Allora per la (4) vista prima si ha che

$$\begin{aligned} P\{A\} &= P\{H_1\}P\{A|H_1\} + P\{H_2\}P\{A|H_2\} + P\{H_3\}P\{A|H_3\} = \\ &= 0,4(0,8) + 0,25(0,6) + 0,35(0,9) = \\ &= 0,785 \end{aligned}$$

e quindi dalla formula di Bayes si ha che

$$\begin{aligned} P\{H_1|A\} &= P\{H_1\}P\{A|H_1\} / P\{A\} = \\ &= 0,4(0,8) / 0,785 = \\ &= 0,4076 \end{aligned}$$

e

$$\begin{aligned} P\{H_2|A\} &= P\{H_2\}P\{A|H_2\} / P\{A\} = \\ &= 0,25(0,6) / 0,785 = \\ &= 0,1911 \end{aligned}$$

e infine

$$\begin{aligned} P\{H_3|A\} &= P\{H_3\}P\{A|H_3\} / P\{A\} = \\ &= 0,35(0,9) / 0,785 = \\ &= 0,4013 \end{aligned}$$

Ne segue che un paziente che presenta uno o più sintomi  $A$  ha con maggiore probabilità contratto la malattia  $H_1$  e in assenza di ulteriori informazioni dovrebbe essere curato per tale malattia.

**Esempio 3.** supponiamo che il Ministro delle Finanze, prima di adottare una nuova politica di controllo



dei prezzi e dei salari, chiedi il parere di tre esperti indicati con H1,H2,H3 sull'impatto che tale politica può avere sul tasso di disoccupazione. Le risposte date in termini probabilità dai singoli esperti sono raccolte in tab.1:

ESPERTO	Diminuzione (D)	Stabilità (S)	Aumento (A)
H1	0,1	0,1	0,8
H2	0,6	0,2	0,2
H3	0,2	0,6	0,2

Sulla base delle passate esperienze il Ministro si è fatto l'opinione che la probabilità che ogni esperto abbia una teoria corretta dell'economia sono date rispettivamente da  $P\{H1\}=1/6$ ,  $P\{H2\}=1/3$  e  $P\{H3\}=1/2$ . Supponendo che a seguito della politica adottata si verifichi un aumento nel tasso di disoccupazione, si vuole determinare 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 le probabilità  $P\{H1|A\}$ ,  $P\{H2|A\}$  e  $P\{H3|A\}$ . A tale scopo calcoliamo da prima la probabilità di un incremento nel tasso di disoccupazione. Per la (4) vista sopra si ha che

$$\begin{aligned} P\{A\} &= P\{H1\}P\{A|H1\} + P\{H2\}P\{A|H2\} + P\{H3\}P\{A|H3\} = \\ &= (1/6)(0,8) + (1/3)(0,2) + (1/2)(0,2) = \\ &= 0,3 \end{aligned}$$

Usando quindi la formula di Bayes si ottiene

$$\begin{aligned} P\{H1|A\} &= P\{H1\}P\{A|H1\}/P\{A\} = \\ &= (1/6)(0,8)/(0,3) = \\ &= 0,444444444... \end{aligned}$$

e

$$\begin{aligned} P\{H2|A\} &= P\{H2\}P\{A|H2\}/P\{A\} = \\ &= (1/3)(0,2)/(0,3) = \\ &= 0,222222222... \end{aligned}$$

e infine

$$\begin{aligned} P\{H3|A\} &= P\{H3\}P\{A|H3\}/P\{A\} = \\ &= (1/2)(0,2)/(0,3) = \\ &= 0,333333333... \end{aligned}$$

Si può quindi osservare che la teoria dell'esperto H1, la meno corretta secondo l'opinione originaria del Ministro, prima cioè che la politica venisse adottata, appare nel seguito, ovvero in conseguenza della politica adottata, la più corretta.

### Variabili casuali o aleatorie

Consideriamo 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 dei 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 (croce-croce), il valore 1 nei due punti campione TC e CT (testa-croce e croce-testa) ed il valore 2 nel punto campione TT (testa-testa). Una variabile X come questa a valori numerici è un esempio di **variabile casuale**. Come secondo esempio riguardante l'esperimento del getto di due dadi, se la variabile X rappresenta la somma dei punti ottenuti, allora essa è una variabile casuale X che può assumere i valori interi da 2,...,12. Come altro esempio, se la variabile X rappresenta la distanza dal centro di una freccia scagliata su di un bersaglio circolare di raggio  $r=20$  cm, allora assumendo di trascurare tutti i colpi mancati essa è una variabile casuale che può assumere tutti i valori tra 0,...,20. In tutti questi esempi la variabile X è calcolata numericamente ed il suo valore dipende dal punto campione. Così la variabile X è una funzione il cui dominio di definizione è l'insieme dei punti campione e il cui insieme di valori, cioè il suo codominio, è un insieme di numeri reali. Si può dare allora 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 X viene chiamata casuale è perché 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 e quindi dipende dal caso il valore che in esso ha la variabile casuale. L'obiettivo che ora ci 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 con la variabile casuale  $X$ . Perciò si assume che a qualsiasi spazio campione sia assegnata una misura della 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 determinare la probabilità che  $X$  assuma valori specifici nel suo insieme di valori possibili. Per esempio, se  $X$  rappresenta la somma dei punti ottenuti nel getto di due dadi, allora può interessare il calcolo della probabilità che  $X$  assuma ad esempio il valore 7. Oppure, se  $X$  rappresenta la distanza dal centro di una freccia scagliata su di un bersaglio circolare, può interessare calcolare la probabilità che  $X$  assuma ad esempio un valore minore di 5. Il calcolo  $P\{X=7\}$  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 piuttosto che con uno che consiste di intervalli 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 uno **spazio campione continuo**, per intervalli s'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.

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 prende il nome di **variabile casuale discreta**. Per calcolare le probabilità delle variabili casuali discrete ricordiamo una formula per calcolare la probabilità di un evento  $A$ , cioè  $P\{A\} = \sum_A p_i$  ricavata assumendo che lo spazio campione fosse discreto e applicabile quindi a qualsiasi spazio campione discreto. Allora se  $x$  è un dato valore della variabile casuale discreta  $X$ , questa formula si può usare per calcolare la probabilità dell'evento che  $X$  assuma il valore  $x$ . Questa probabilità è perciò data dalla

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

dove la sommatoria viene fatta su tutti i punti campione in cui la variabile casuale  $X$  ha il valore  $x$ . Come esempio, se  $X$  rappresenta la somma dei punti ottenuta nel getto di due dadi, tenuto conto che tutti i punti campione hanno probabilità uguale ad  $1/36$ , allora la probabilità  $P(X=7) = \sum_{X=7} p_i = 6/36$  ed i 6 punti sono quelli di coordinate  $(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)$ .

### Funzione di probabilità, o 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:

---

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:

- $f(x) \geq 0$
  - $\sum_x f(x) = 1$
- 

Si fa notare che in genere si parla di *densità* riferendosi ad esempio a una distribuzione continua di materia (cioè di massa) lungo una retta o un piano e che perciò il suo uso a proposito di una distribuzione di probabilità discreta potrebbe 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 discreta spesso consiste semplicemente di una tabella di valori, così nell'esperimento del lancio di due monete, se la variabile  $X$  rappresenta la somma delle teste ottenute, la funzione di densità discreta si può definire per mezzo del seguente insieme di valori:

- $f(0)=P(X=0)=\sum_{x=0} p_i=1/4$
- $f(1)=P(X=1)=\sum_{x=1} p_i=(1/4)+(1/4)=2/4=1/2$
- $f(2)=P(X=2)=\sum_{x=2} p_i=1/4$

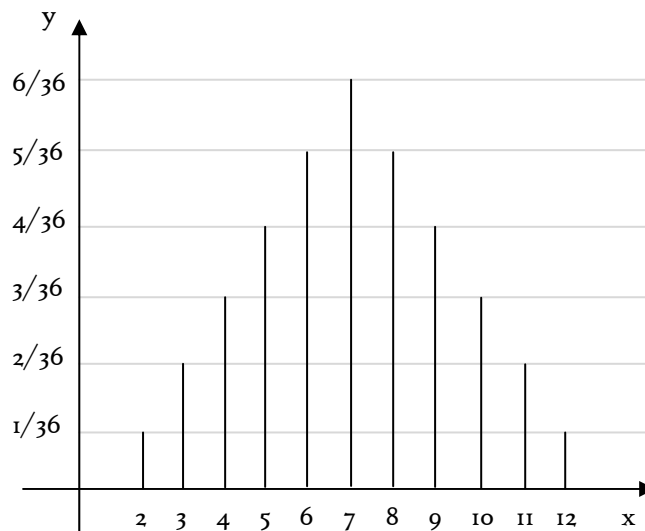
Osserviamo inoltre che la funzione  $f(x)=0$  per  $x \neq 0,1,2$ .

Per giudicare come si distribuisce una variabile casuale, cioè come la sua probabilità varia al variare del valore della variabile è utile tracciare il grafico della funzione di densità per mezzo di un **grafico a segmenti**. Come esempio di un tale grafico, la variabile  $X$  rappresenti la somma dei punti ottenuti lanciando 2 dadi. Attribuendo l'esatto valore di  $X$  a ciascuno dei 36 punti dello spazio campione corrispondente al getto di due dadi e assumendo probabilità uguali per i punti campione, mediante l'impiego della formula (1), si ottiene:

- $f(2)=f(12)=1/36$
- $f(3)=f(11)=2/36$
- $f(4)=f(10)=3/36$
- $f(5)=f(9)=4/36$
- $f(6)=f(8)=5/36$
- $f(7)=6/36$

Un grafico a segmenti della  $f(x)$  è mostrato in fig.1:

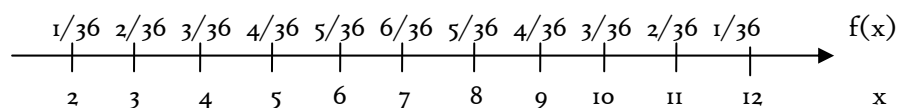
**Fig.1**



Lo scopo di introdurre una funzione di densità è di semplificare la notazione e il calcolo delle probabilità delle variabili discrete.

Dopo che una funzione di densità è stata determinata, non è più necessario calcolare le probabilità degli eventi riguardanti la variabile casuale sommando le probabilità dei punti campione come nella (1) vista precedentemente. 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 della  $f(x)$  attribuiti a quei punti. Per esempio, se  $X$  rappresenta la somma dei punti nel getto di due dadi, il nuovo spazio campione consisterà degli undici punti 2,3,4,...,12 e le probabilità associate a quei punti saranno i valori della  $f(x)$  calcolati ieri. Questo nuovo spazio campione è mostrato in fig.2:

**Fig.2**



Questo spazio campione è semplicemente una rappresentazione dell'informazione trasmessa dalla fig.1 vista precedentemente.

Ora consideriamo il calcolo della probabilità dell'evento  $X \in R$ , dove  $R$  è un dato insieme di punti sull'asse  $x$  e  $X \in R$  rappresenta l'evento che  $X$  assume valori in  $R$ . Considerando il nuovo spazio

campione generato da  $X$  e da  $f(x)$ , una diretta applicazione di una formula già nota, e cioè che  $P(A)=\sum_A p_i$ , questa probabilità è data dalla

$$(1) P\{X \in R\} = \sum_{X \in R} f(x)$$

dove la sommatoria viene fatta su quei valori  $X$  in  $R$  per i quali  $f(x) > 0$ . Il calcolo della probabilità come indicato è molto più semplice del calcolo che si basa sullo spazio campione originario dell'esperimento. Usando di nuovo l'esempio del getto di due dadi, supponiamo di voler calcolare la probabilità che la somma dei punti sia maggiore di 7. Considerando il nuovo spazio campione mostrato in fig.2, questa probabilità è data da

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

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

$$\begin{aligned} P\{X > 7\} &= \sum_{X > 7} p_i = \\ &= 1/36 + \dots + 1/36 = 15/36 \end{aligned}$$

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** o di **distribuzione cumulativa** o di **ripartizione** definita dalla

$$F(x) := P\{X \leq x\} = \sum_{t \leq x} f(t)$$

dove la sommatoria viene fatta su tutti i valori della variabile casuale che sono minori o uguali a  $x$ . Il grafico della  $F(x)$  corrispondente alla  $f(x)$  della fig.1 vista precedentemente è mostrato in fig.3. Per costruirlo occorre calcolare il valore assunto dalla  $F(x)$  in corrispondenza dei valori 2,3,4,...,12:

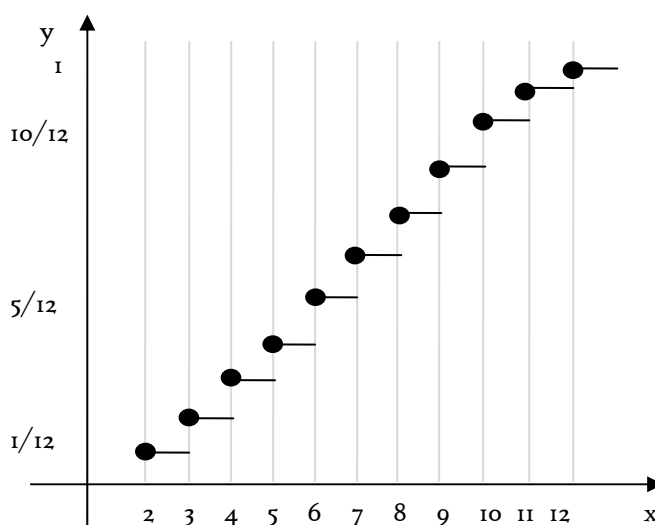
$$F(2) = P(X \leq 2) = \sum_{t \leq 2} f(t) = f(2) = 1/36$$

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

...

$$F(12) = P(X \leq 12) = \sum_{t \leq 12} f(t) = f(2) + f(3) + f(4) + \dots + f(12) = 1$$

Fig.3



Si fa notare che i punti in grassetto all'inizio di ciascun segmento di retta stanno a significare che il valore di  $F(x)$  per  $x$  intero è quello corrispondente al segmento orizzontale superiore anziché a quello inferiore.

## Funzioni di probabilità congiunta, o di densità di probabilità congiunta

Molti esperimenti implicano parecchie 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 la variabile  $X$  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 dalla  $f(x,y):=P(X=x,Y=y)$  viene chiamata la funzione di probabilità congiunta oppure la funzione di densità di probabilità congiunta della variabile  $X$  e  $Y$  se soddisfa le seguenti proprietà:

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

L'aggettivo *congiunta* spesso si omette perché non è possibile confondere una funzione di due variabili con una funzione di una sola variabile. Come esempio consideriamo il caso in cui la variabile  $X$  rappresenta il numero di carte di picche ottenute quando si estrae una carta da un mazzo e con la variabile  $Y$  il numero di carte di picche ottenute quando si estrae una seconda carta dal mazzo senza che la prima sia stata reinserita. In questo caso, le variabili  $X$  e  $Y$  possono assumere soltanto i valori 0 e 1. Ricordando la regola di moltiplicazione delle probabilità, cioè  $P\{A_1 \cap A_2\} = P\{A_1\}P\{A_2|A_1\}$ , la funzione di densità  $f(x,y)$  è data dalla seguente tabella di valori:

$$f(0,0) = P(X=0,Y=0) = P(X=0)P(X=0|Y=0) = (39/52)(38/51) = 0,56$$

$$f(0,1) = P(X=0,Y=1) = P(X=0)P(X=0|Y=1) = (39/52)(13/51) = 0,19$$

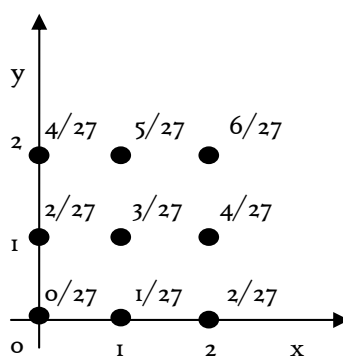
$$f(1,0) = \dots$$

$$f(1,1) = P(X=1,Y=1) = P(X=1)P(X=1|Y=1) = \dots = 0,06$$

Una funzione di densità aggiunta 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. Come nel problema monodimensionale (cioè per una sola variabile) è conveniente operare in un nuovo spazio campione rispetto all'originario che consiste dei punti nel piano  $xOy$  dove  $f(x,y) > 0$  e le cui probabilità sono i valori forniti da  $f(x,y)$ .

**Esempio.** Si consideri la funzione di densità di due variabili casuali discrete data da  $f(x,y) = 1/27(x+2y)$  in cui  $x$  e  $y$  possono assumere tutti i valori interi tali che  $0 \leq x, y \leq 2$  e uguale a 0 altrove. A questo punto si vuole calcolare la probabilità  $P\{X \geq 1, Y \leq 1\}$ . Lo spazio campione con le probabilità calcolate tramite la formula suddetta è mostrato in fig.4:

Fig.4



La probabilità richiesta è allora data da

$$\begin{aligned} P\{X \geq 1, Y \leq 1\} &= \sum_{x=1}^2 \sum_{y=0}^1 f(x,y) = \\ &= \sum_{x=1}^2 [f(x,0) + f(x,1)] = \\ &= f(1,0) + f(1,1) + f(2,0) + f(2,1) = \\ &= 1/27 + 3/27 + 2/27 + 4/27 = 10/27 = 0,37 = 37\% \end{aligned}$$

## Variabili indipendenti

Due variabili che sono senza rapporti nel senso probabilistico sono chiamate **variabili indipendenti**.

Poiché l'indipendenza di due eventi  $A_1$  e  $A_2$  è stata definita da  $P(A_1 \cap A_2) = P(A_1)P(A_2)$ , una definizione corrispondente per le variabili casuali dovrebbe essere conforme a quella definizione. Considerando eventi quali  $X \in A$  e  $Y \in B$ , dove  $A$  e  $B$  sono insiemi qualsiasi nei domini di definizione delle variabili  $X$  e  $Y$  rispettivamente, la definizione di indipendenza per gli eventi richiederebbe che

$$(1) P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

per ogni  $A, B$ . Questa viene presa spesso come definizione di indipendenza di due variabili. Tuttavia, poiché ora si sta ponendo l'accento sulle funzioni di densità, si darà invece una definizione equivalente basata sulle funzioni di densità. A tale scopo sia  $f(x, y)$  la funzione di densità congiunta delle variabili  $X$  e  $Y$  e  $g(x)$  e  $h(y)$  siano le funzioni di densità delle singole variabili. Allora, scegliendo che gli insiemi  $A$  e  $B$  siano costituiti rispettivamente dai singoli punti  $x$  e  $y$ , la (1) si riduce alla  $P(X=x, Y=y) = P(X=x)P(Y=y)$  per tutti gli  $x$  e  $y$ . Ma, in termini di funzioni di densità, questa si scrive  $f(x, y) = g(x)h(y)$  per tutti gli  $x$  e  $y$ . Viceversa, se  $f(x, y) = g(x)h(y)$  per tutti gli  $x$  e  $y$ , allora, operando nello spazio campione delle due variabili casuali, segue per analogia con la (1) vista precedentemente, cioè  $P(X \in R) = \sum_{x \in R} f(x)$ , segue che

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

Questa mostra che la (1) vale per tutti gli  $A$  e  $B$  se e soltanto se  $f(x, y) = g(x)h(y)$  per tutti gli  $x$  e  $y$ . 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 dare allora la seguente definizione:

---

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

$$(2) f(x, y) = g(x)h(y)$$

per tutti gli  $x$  e  $y$ .

---

La precedente definizione si può generalizzare in ovvia maniera per definire l'indipendenza di  $n$  variabili casuali. Così

---

le variabili casuali  $X_1, \dots, X_n$  la cui funzione di densità congiunta è  $f(x_1, \dots, x_n)$  e le cui funzioni di densità singole sono  $f_1(x_1), \dots, f_n(x_n)$  sono indipendenti se e soltanto se  $f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n)$  per tutti gli  $x_1, \dots, x_n$ .

---

### 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 determinare le funzioni di densità delle singole variabili dalla loro funzione di densità 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

$$(3) P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$$

si può allora esprimere in termini di funzioni di densità. Poiché ora  $P(A_1 \cap A_2)$  è la probabilità che le due variabili casuali assumano rispettivamente i valori  $x$  e  $y$ , essa rappresenta il valore della funzione di densità congiunta nel punto di coordinate  $(x, y)$ , cioè  $f(x, y)$ . In modo simile,  $P(A_1)$  è la probabilità che la variabile  $X$  assuma il valore  $x$ , perciò altro non è che  $f(x)$ . Poiché  $P(A_2|A_1)$  è la probabilità che la variabile  $Y$  assuma il valore  $y$  quando la variabile  $X$  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  $f(y|x)$  fornisce la probabilità condizionata che Y assuma il valore y quando X ha il fissato valore x, la somma di  $f(y|x)$  su tutti i possibili valori y deve esser uguale a 1. Quindi, se entrambi i membri della (4) si sommano su tutti i possibili valori y, si ottiene per la  $f(x)$  la formula seguente:

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

infatti, sommando su y ambo i membri della (4) si ottiene

$$\begin{aligned}\sum_y f(x,y) &= \sum_y f(x)f(y|x) = \\ &= f(x)\sum_y f(y|x) = \\ &= f(x)*1\end{aligned}$$

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 marginale della variabile Y  $g(y)$  si può ottenere sommando la  $f(x,y)$  su tutti i possibili valori della variabile X per un y fissato. Questi 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 possibili 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 della variabile X. Per la (4) se  $f(x)>0$  si può scrivere che

$$(6) f(y|x)=f(x,y)/f(x)$$

La distribuzione condizionata della variabile X quando viene mantenuto fisso il valore della variabile Y è data da una formula analoga, cioè

$$(6.1) f(x|y)=f(x,y)/g(y)$$

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 si una di esse quando il valore dell'altra variabile viene mantenuto fisso, è semplicemente necessario dividere la funzione di densità congiunta per la funzione di densità il cui valore viene mantenuto fisso.

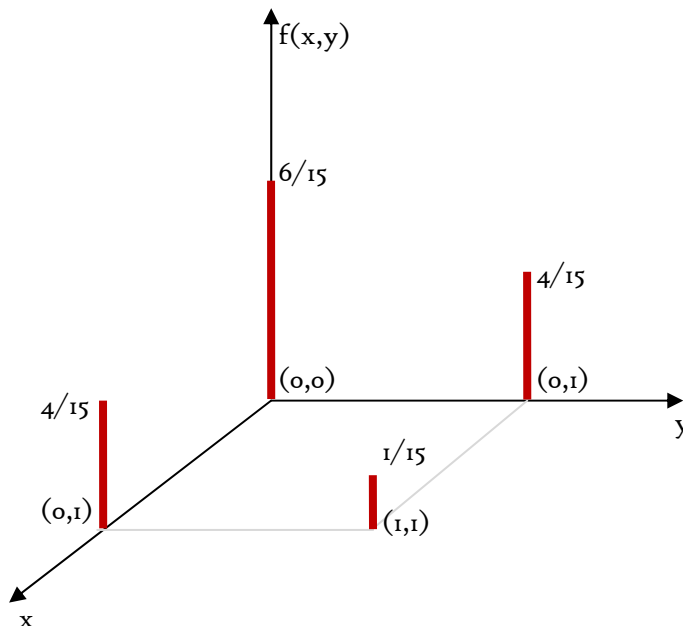
**Esercitazione.** Per illustrare i concetti precedenti, supponiamo che un'urna contenga due palline bianche e quattro nere e che due palline vengano estratte dall'urna. Le variabili X e Y rappresentino i risultati delle due estrazioni, 0 che corrisponde ad una pallina nera e 1 che corrisponde a una pallina bianca. Allora ogni possibile risultato sarà rappresentato da uno dei 4 punti nel piano xOy 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 formula (4) che

$$\begin{aligned}f(0,0) &= f(0)f(0|0) = (4/6)(3/5) = 6/15 \\ f(0,1) &= f(0)f(1|0) = (4/6)(2/5) = 4/15 \\ f(1,0) &= f(1)f(0|1) = (2/6)(4/5) = 4/15 \\ f(1,1) &= f(1)f(1|1) = (2/6)(1/5) = 1/15\end{aligned}$$



I valori della  $f(x,y)$  possono essere riportati in un grafico a segmenti illustrato in fig.1:

Fig.1



Per illustrare il metodo di ottenere una distribuzione marginale e una distribuzione condizionata dalla distribuzione congiunta, assumiamo che siano noti soltanto i valori finali della  $f(x,y)$  calcolati nell'ultimo esempio fatto prima. Così la sola informazione disponibile è quella data dalla fig.1 vista in precedenza, indipendentemente da come sono stati ottenuti quei valori. La funzione di densità marginale della variabile  $X$  si può ottenere applicando la formula (5) vista l'ultima volta, cioè  $f(x) = \sum_y f(x, y)$ . Si ha quindi che

$$\begin{aligned} f(x) &= \sum_{y=0}^1 f(x, y) = \\ &= f(x, 0) + f(x, 1) \\ f(0) &= f(0, 0) + f(0, 1) = \\ &= 6/15 + 4/15 = \\ &= 2/3 \\ f(1) &= f(1, 0) + f(1, 1) = \\ &= 4/15 + 1/15 = \\ &= 1/3 \end{aligned}$$

Se i quattro punti nel piano  $xOy$  si pensano come punti di massa di probabilità, la cui massa totale è uguale a 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 nel piano  $xOy$  sono stati proiettati sull'asse  $x$ . La funzione di densità condizionata della variabile  $Y$  per un valore  $x$  fissato si può ottenere applicando la formula (6) vista in precedenza, cioè  $f(y|x) = f(x,y)/f(x)$  da cui si ha che

$$f(0|x) = f(x, 0)/f(x)$$

e

$$f(1|x) = f(x, 1)/f(x)$$

Ad esempio, per  $x=1$  si ha che

$$\begin{aligned} f(0|1) &= f(1, 0)/f(1) = \\ &= (4/15)/(1/3) = \\ &= 4/5 \end{aligned}$$

e

$$\begin{aligned} f(1|1) &= f(1, 1)/f(1) = \\ &= (1/15)/(1/3) = \\ &= 1/5 \end{aligned}$$

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 le loro masse di probabilità moltiplicate per un numero tale da rendere la somma delle loro masse uguale a 1.

Come secondo esempio in cui la funzione di densità congiunta è data direttamente consideriamo la

funzione di densità definita come  $f(x,y)=(1/27)(x+2y)$  dove  $x$  e  $y$  possono assumere solo i valori interi 0, 1 e 2. Lo spazio campione con le sue probabilità calcolate tramite questa formula è mostrato nella fig.3 vista in precedenza. Dalla formula (5) vista ultimamente, la funzione di densità marginale della variabile  $X$  è data dalla

$$\begin{aligned} f(x) &= \sum_{y=0}^2 (1/27)(x+2y) = \\ &= (1/27)(x+x+2+x+4) = \\ &= (1/27)(3x+6) = \\ &= (1/9)(x+2) \end{aligned}$$

e in modo simile sommando su tutti i valori della variabile  $X$  si ottiene

$$\begin{aligned} g(y) &= \sum_{x=0}^2 (1/27)(x+2y) = \\ &= (1/27)(2y+1+2y+2+2y) = \\ &= (1/27)(6y+3) = \\ &= (1/9)(2y+1) \end{aligned}$$

E' chiaro che in questo caso la  $f(x,y)$  non è uguale al prodotto delle due funzioni di densità marginale. È perciò che le variabili  $X$  e  $Y$  non sono variabili casuali indipendenti.

Dalla formula (6) vista l'ultima volta e dal risultato appena ottenuto per la  $f(x)$  segue che la funzione di densità condizionata della variabile  $Y$  per un valore  $x$  fissato è data dalla formula

$$\begin{aligned} f(y|x) &= [(1/27)(x+2y)] / [(1/9)(x+2)] = \\ &= (x+2y) / 3(x+2) \end{aligned}$$

da cui se ad  $x$  si assegna ad esempio il valore 2, cioè  $x=2$ , si ha che

$$f(y|2) = (2+2y) / 12 = (1/6)(y+1)$$

Questa funzione sarebbe utile se si volessero calcolare le probabilità per i vari possibili valori della variabile  $Y$  quando  $x=2$ . Si può facilmente verificare che questa è funzione di densità, infatti sommando le tre probabilità ottenute da questa formula per  $y=0,1,2$  si ottiene come somma 1. Infatti,  $f(0|2)=1/6$ ,  $f(1|2)=2/6$  e  $f(2|2)=3/6$  la cui somma è uguale a 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 sono esempi di 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. Comunque, però, è conveniente dal punto di vista matematico fare l'ipotesi che questa limitazione non esista.

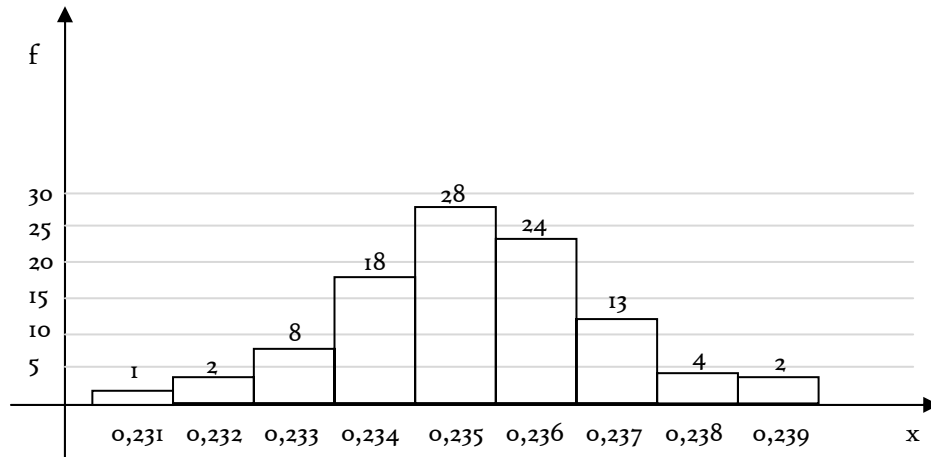
## Funzione di densità di probabilità

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  $X$  che rappresenta lo spessore di un disco metallico prodotto da una certa macchina. Se la macchina producesse ad esempio 100 dischi, e se i loro spessori fossero misurati con tre cifre decimali, avremmo a disposizione 100 valori della variabile  $X$  con cui studiare il comportamento della macchina. Se questi 100 valori fossero raccolti e rappresentati sotto forma di tabella, si potrebbe trovare un insieme di valori come quelli mostrati in tab.1 che fornisce la frequenza assoluta con cui si sono presentati i vari valori della variabile  $X$ .

x	0,231	0,232	0,233	0,234	0,235	0,236	0,237	0,238	0,239
f	1	2	8	18	28	24	13	4	2

Si fa osservare che il termine *frequenza* indica di solito il rapporto tra il numero di volte che si è presentato uno specifico valore osservato della variabile  $X$  ed 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 assoluta* e *frequenza relativa*. In tab.1 sono riportate le frequenze assolute. Per rappresentare graficamente i risultati della tab.1 si usa un grafico chiamato istogramma che ora passiamo a costruire.

Fig.1



Un istogramma è un grafico come quello mostrato in fig.1 in cui si usano le aree dei rettangoli per rappresentare le frequenze osservate, in particolare le frequenze relative. Così l'area del rettangolo, centrato ad esempio in  $x=0,234$ , dovrebbe essere uguale alla frequenza relativa  $0,18$ . Tuttavia, 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. Infatti, ad esempio l'area del rettangolo centrata in  $x=0,234$  è  $A=0,001 \times 18=0,018$  che è soltanto proporzionale alla frequenza relativa  $0,18$ . 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 ad esempio in  $x_i$  sarà data da  $f_i/Nh$  dove  $f_i$  indica la frequenza assoluta di  $x_i$  ed  $N$  indica il numero totale di osservazioni. Questo risultato è ovvio quando si pensa che questa ordinata moltiplicata per la base  $h$  del rettangolo deve essere uguale alla frequenza relativa  $f_i/N$ . Infatti  $(f_i/Nh)h=f_i/N$ . 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 grande il doppio di quello basato su 100 esecuzioni. Per confrontare fra loro 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 tenda ad un istogramma fisso, cioè che non cambia più all'aumentare del numero di esecuzioni dell'esperimento. Inoltre, se si fa l'ipotesi che la variabile  $X$  si possa misurare tanto accuratamente quanto si vuole così che l'unità  $h$  sull'asse  $x$  possa farsi tanto piccola quanto si vuole, allora ci si dovrebbe aspettare che l'istogramma si smussi e tenda ad una curva continua quando il numero di esecuzioni aumenta indefinitamente ed  $h$  si sceglie sempre più piccolo, cioè tende a 0. Quando l'area dell'istogramma è uguale a 1 segue da quanto detto che la somma delle aree di rettangoli vicini tra loro è uguale alla frequenza relativa con cui il valore della variabile  $X$  cade nell'intervallo costituito dalle basi di quei rettangoli. Poiché questa proprietà continuerà a valere all'aumentare del numero di esecuzioni dell'esperimento e al tendere a 0 di  $h$  l'area sottesa fra due qualsiasi valori di  $X$ , dovrebbe essere uguale alla frequenza con cui ci si aspetterebbe che il valore osservato della variabile  $X$  cada nell'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 forma seguente:

---

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

---

- $f(x) \geq 0$
- $\int_{-\infty}^{+\infty} f(x) dx = 1 = P(-\infty < X < +\infty)$
- $\int_a^b f(x) dx = P(a < X < b)$

dove  $a$  e  $b$  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 la variabile  $X$  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 di una variabile continua sono sempre fornite da integrali mentre quelle di una variabile discreta da somme. Se il dominio della variabile  $X$  non è l'intera retta reale si fa l'ipotesi che la  $f(x)$  sia nulla per valori esterni allo specificato dominio della variabile.

**Esempio.** Come esempio consideriamo la possibilità di usare la funzione  $f(x) = ke^{-x}$ , dove  $k$  è una costante, come funzione di densità della variabile  $X$ . Per la prima proprietà è chiaro che la costante  $k$  deve essere positiva, poi

$$\int_{-\infty}^{+\infty} e^{-x} dx = 0 + e^{+\infty}.$$

Questo impone che la variabile  $x$  debba essere limitata. Quindi facciamo l'ipotesi ad esempio che  $x$  possa assumere soltanto valori non negativi, cioè  $x \geq 0$ . In questo caso

$$\int_{+\infty}^0 e^{-x} dx = 0 + 1 = 1$$

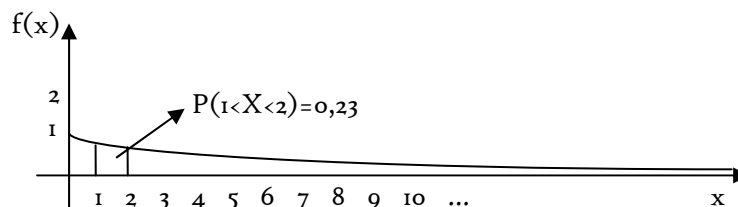
Se  $k=1$  può essere una funzione di densità della variabile  $x$  la funzione  $f(x) = e^{-x}$  per  $x \geq 0$  e  $f(x) = 0$  per  $x < 0$ .

Per calcolare la probabilità, ad esempio, che la variabile  $x$  cada nell'intervallo compreso fra 1 e 2, cioè  $P(1 < X < 2)$ , si ha che

$$\begin{aligned} P(1 < X < 2) &= \int_1^2 e^{-x} dx = \\ &= -e^{-2} - (-e^{-1}) = \\ &= 0,23 \end{aligned}$$

Il grafico di questa funzione di densità e la rappresentazione come area della  $P(1 < X < 2)$ , sono mostrati nella figura seguente.

**Fig.2**



E' importante sottolineare che, sebbene in qualsiasi dato 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 certamente una scelta utile.

### Funzioni di distribuzione, o di distribuzione cumulativa, o di ripartizione

Una funzione strettamente affine alla funzione di densità  $f$  è la corrispondente *funzione di distribuzione* o di distribuzione cumulativa o di ripartizione  $F$  definita dalla

$$(1) F(x) := P(X \leq x) = \int_{-\infty}^x f(t) dt$$

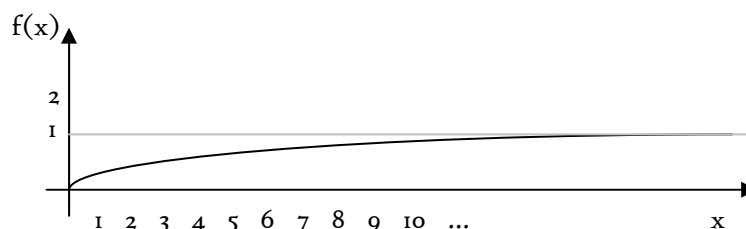
**Esempio.** Come esempio, si può ricavare dalla (1) la  $F(x)$  corrispondente alla funzione di densità dell'esempio visto in precedenza. Si ha che

$$\begin{aligned} F(x) &= \int_0^x f(t) dt = \\ &= \int_0^x e^{-t} dt = \\ &= -e^{-t} + 1 \text{ per } x \geq 0 \end{aligned}$$

$$=0 \quad \text{per } x < 0$$

Il grafico della  $F(x)$  è mostrato il fig.1.

Fig.1



Si fa notare ora che

$$P(1 < X < 2) = F(2) - F(1)$$

infatti

$$\begin{aligned} F(2) - F(1) &= (-e^{-2} + 1) - (-e^{-1} + 1) = \\ &= e^{-1} - e^{-2} = \\ &= 0,23 \end{aligned}$$

La funzione di densità è quella più comunemente usata nelle applicazioni statistiche, 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 continua anziché 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  $dF(x)/dx = f(x)$  purché la  $f(x)$  sia una funzione continua.

Considerando l'esempio appena trattato si può infatti verificare che, derivando la funzione di distribuzione, si ottiene la corrispondente funzione di densità. Infatti

- $dF(x)/dx = d(-e^{-x} + 1)/dx = e^{-x}$  per  $x \geq 0$
- $dF(x)/dx = d(-e^{-x} + 1)/dx = 0$  per  $x < 0$

### Funzione di densità congiunta

Una funzione di densità di due o più variabili continue è la naturale generalizzazione di una funzione di densità di una sola variabile. Così una funzione di densità di due variabili casuali continue  $X$  ed  $Y$  si indica con  $f(x,y)$  e come è noto rappresenta geometricamente una superficie nello spazio tridimensionale, cioè la superficie  $z=f(x,y)$ , proprio come la funzione di densità  $f(x)$  è rappresentata da una curva nel piano  $xOy$ , e cioè la 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  $xOy$  dia la probabilità che le variabili casuali  $X$  ed  $Y$  assumano valori che corrispondono ad un punto interno a questa superficie. Queste proprietà essenziali di una funzione di densità di due variabili si possono formalizzare come segue:

---

una funzione di densità di due variabili casuali continue  $X$  ed  $Y$  è una funzione  $f$  che possiede le seguenti proprietà:

- $f(x,y) \geq 0$
  - $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) dx dy = 1$
  - $\int \int f(x,y) dx dy = P(X,Y \in R)$  per ogni  $R$
- 

Si fa l'ipotesi qui 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  $a < x < b$  e  $c < y < d$ . In questo caso allora la terza proprietà diventa

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

**Esempio.** Come esempio di funzione di densità di due variabili casuali continue consideriamo la funzione  $f(x,y) = e^{-(x+y)}$  che è una naturale generalizzazione in due dimensioni della funzione di densità dell'esempio visto precedentemente. Se la  $f(x,y)$  si definisce nulla per valori negativi di  $x$  o  $y$ , cioè se

$$\begin{aligned} f(x,y) &= e^{-(x+y)} && \text{per } x,y \geq 0 \\ f(x,y) &= 0 && \text{per } x,y < 0 \end{aligned}$$

allora le prime due proprietà sono soddisfatte. Infatti la  $f(x,y)$  è certamente maggiore di 0 ed inoltre si ha che

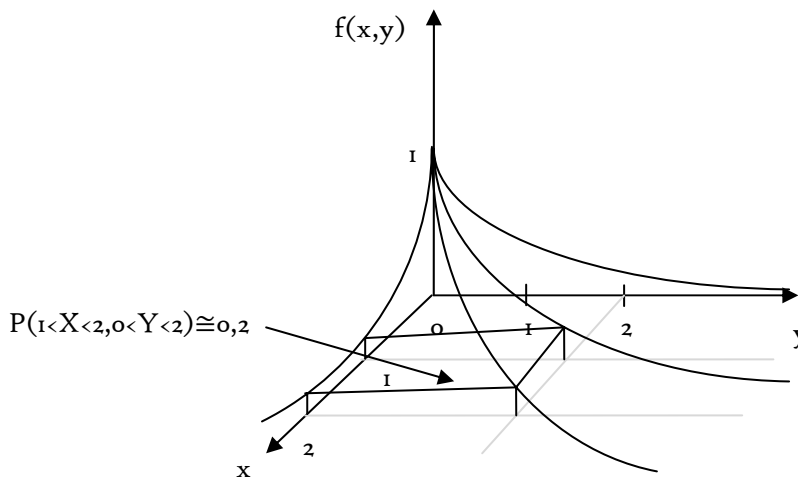
$$\begin{aligned} \int_0^{+\infty} \int_0^{+\infty} e^{-(x+y)} dy dx &= \int_0^{+\infty} \int_0^{+\infty} e^{-x} e^{-y} dy dx = \\ &= \int_0^{+\infty} e^{-x} \left( \int_0^{+\infty} e^{-y} dy \right) dx = \\ &= \int_0^{+\infty} e^{-x} (1) dx = 1 \end{aligned}$$

La probabilità, ad esempio

$$\begin{aligned} P(1 < X < 2, 0 < Y < 2) &= \int_1^2 \int_0^2 e^{-x} e^{-y} dy dx = \\ &= \int_1^2 e^{-x} \left( \int_0^2 e^{-y} dy \right) dx = \\ &= \int_1^2 e^{-x} (-e^{-y} / (2,0)) dx = \\ &= \int_1^2 e^{-x} (-e^{-2} + 1) dx = \\ &= (1 - e^{-2}) \int_1^2 e^{-x} dx = \\ &= (1 - e^{-2}) (-e^{-x} / (2,1)) = \\ &= (1 - e^{-2}) (-e^{-2} + e^{-1}) = \\ &= 0,20 \end{aligned}$$

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

Fig.2



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 discreto si può dimostrare che vale anche nel caso di due variabili casuali continue. La condizione di indipendenza è quindi la stessa, e cioè  $f(x,y) = g(x)h(y)$ .

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 funzione di densità di probabilità della variabile casuale. In molti problemi, però, è sufficiente considerare soltanto certe 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

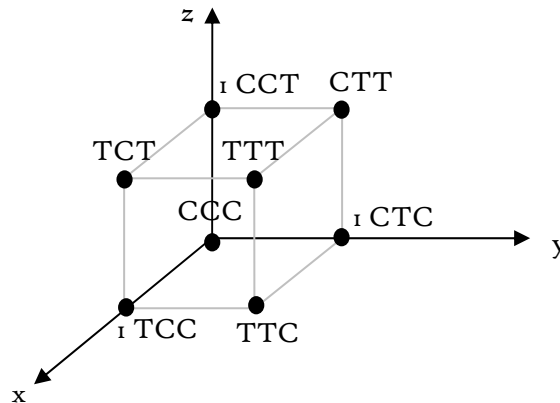
La maggior parte delle variabili discrete che ricorrono negli esperimenti di tipo ripetitivo sono di tipo *conteggio* come ad esempio il numero di incidenti che un'automobilista ha ogni anno, il numero di insetti che sopravvivono ad una irrorazione ed altre ancora. Queste variabili assumono soltanto valori interi non negativi. Sebbene le variabili discrete che presenteremo siano di tipo conteggio, impiegheremo una notazione sufficientemente generale da comprendere anche altri tipi di variabili

casuali discrete.

### Valore atteso

Prima di considerare qualche funzione di densità delle variabili discrete, ci occuperemo dei momenti di una distribuzione di probabilità così da poter calcolare i momenti di quelle particolari istruzioni. Prima però introdurremo e definiremo un concetto più generale e cioè quello di *valore atteso*, poiché esso comprende i momenti come casi particolari, ma è anche uno strumento molto utile per studiare altre proprietà di una distribuzione di probabilità. Come esempio per illustrare il concetto generale di valore atteso consideriamo un gioco in cui si lanciano tre monete e si vince 1\$ per ogni testa che esce. Se la variabile  $X$  indica la somma vinta, allora essa deve assumere uno dei seguenti valori  $X \rightarrow 0, 1, 2, 3$  ai quali corrispondono le probabilità  $1/8, 3/8, 3/8, 1/8$ . Queste probabilità sono evidenti se si considera lo spazio campionario del lancio di tre monete illustrato nella fig.1:

Fig.1



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

$$0x(1/8) + 1x(3/8) + 2x(3/8) + 3x(1/8) = 1,5$$

Questa somma, cioè 1,5\$, è ciò che rappresenta comunemente la somma che ci si aspetta di vincere.

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

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

Nell'esempio appena visto la variabile  $X$ , cioè la somma vinta, può avere valori  $x_1=0, x_2=1, x_3=2$  e  $x_4=3$  con le probabilità corrispondenti  $f(x_1)=1/8, f(x_2)=3/8, f(x_3)=3/8$  e  $f(x_4)=1/8$ . Supponiamo ora che il gioco considerato sia 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 dollari il quadrato del numero di teste uscite. Allora, nel gioco così modificato, il valore atteso sarebbe dato dalla formula

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

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

---

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

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

---



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 di solito chiamato **la media della variabile casuale** oppure **la media della distribuzione della variabile casuale  $X$** . Se la variabile  $X$  può assumere soltanto un numero finito, ad esempio  $k$ , di possibili valori, l'indice superiore della somma del secondo membro della (3) sarà  $k$  anziché infinito. Quando la variabile  $X$  possiede un numero infinito di valori possibili con probabilità positive è necessario assumere che la serie infinita a 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:

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

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

Il momento d'ordine  $k$  di una distribuzione è anche comunemente chiamato **il momento d'ordine  $k$  della variabile casuale  $X$**  avente, s'intende, quella distribuzione. Così si può parlare di  $\mu'_k$  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  $\mu'_1$  che è uguale a  $E[X]$  ricorre così spesso che esso viene indicato col simbolo particolare  $\mu$ . Poiché anche i momenti dalla media sono usati ampiamente essi pure necessitano di essere definiti:

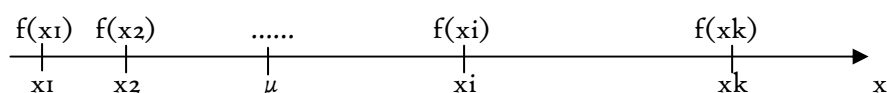
il momento d'ordine  $k$  dalla media della distribuzione della variabile casuale discreta  $X$ , la cui densità è  $f$ , è dato da

$$(5) \mu_k = E[(X - \mu)^k] = \sum_{i=1}^{\infty} (x_i - \mu)^k f(x_i)$$

I momenti di una distribuzione di probabilità aiutano molto 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 2 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 ordine dall'origine  $\mu$  si usa per determinare dove la distribuzione è centrata e il momento del secondo ordine dalla media  $\mu_2$  si usa per determinare il grado di concentrazione della distribuzione attorno alla media  $\mu$ . Poiché anche il momento del secondo ordine dalla media  $\mu_2$  si usa molto spesso a questo proposito, esso si indica con il simbolo particolare  $\sigma^2$  e vien chiamato *varianza della distribuzione*. La radice quadrata positiva della varianza, cioè  $\sigma$ , viene chiamata *scarto quadratico medio* o *deviazione standard della distribuzione*. Esso viene impiegato al posto della varianza quando si desidera una misura della concentrazione nelle stesse unità di misura della variabile casuale. Per esaminare in che modo  $\mu$  e  $\sigma^2$  aiutano a descrivere una distribuzione di probabilità discreta, le probabilità  $f(x_1), \dots, f(x_k)$  associate ai valori  $x_1, \dots, x_k$  della variabile casuale  $X$  siano rappresentate geometricamente come masse di probabilità puntiformi la cui somma è uguale a 1 localizzate nei punti  $x_1, \dots, x_k$  sull'asse  $x$ . Questa rappresentazione è mostrata in fig.2:

Fig.2

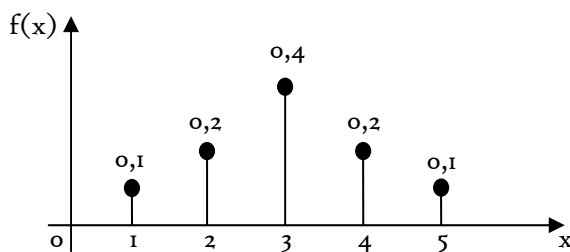


Il momento del primo ordine dall'origine  $\mu$ , per come è definito, fornisce allora il centro di gravità di questo insieme di masse puntiformi, e perciò può servire come misura di dove è centrata o localizzata

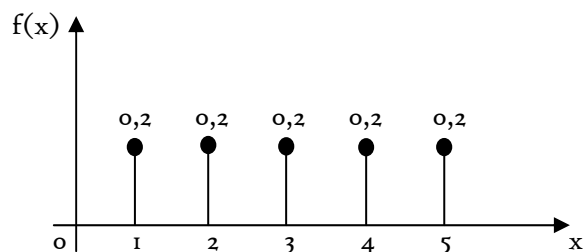
la distribuzione. Il momento del secondo ordine dalla media  $\sigma^2$  tenderà ad essere piccolo se la maggior parte delle masse di probabilità sono concentrate vicino alla media  $\mu$  e non ci sono masse di probabilità localizzate molto lontano da  $\mu$ . Tanto più disperse sono le masse attorno alla media tanto più grande è verosimile che diventi  $\sigma^2$ , così la varianza di una distribuzione può servire come una misura del grado in cui la distribuzione è concentrata attorno alla sua media. È facile tuttavia dare esempi di distribuzioni per cui la varianza  $\sigma^2$  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 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.1**



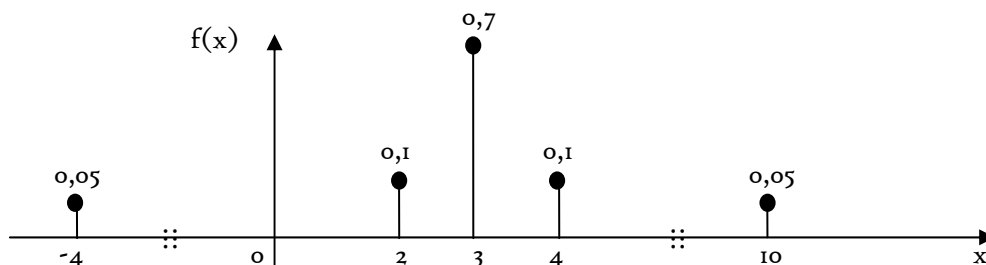
**Fig.2**



per entrambe le distribuzioni, applicando la formula che definisce i momenti dall'origine, cioè la (4) vista in precedenza, con  $k=1$ , risulta la media  $\mu=3$ . Calcoli basati sulla formula che definisce i momenti dalla media, cioè la (5) vista precedentemente, con  $k=2$ , mostrano che la varianza ha rispettivamente i valori  $\sigma^2=1,2$  e  $\sigma^2=1,2$  per le due distribuzioni indicando che la prima è più altamente concentrata rispetto alla sua media in confronto della seconda.

Come esempio in cui il valore della varianza non sembra indicare la natura della concentrazione della distribuzione attorno alla sua media, consideriamo la distribuzione di fig.3:

**Fig.3**



I calcoli danno ancora  $\mu=3$  e  $\sigma^2=5,1$ , valore quest'ultimo che, se confrontato con quelli ottenuti per le distribuzioni delle figure 1 e 2, indicherebbe scarsa concentrazione della distribuzione attorno alla sua media, mentre invece la maggior parte della distribuzione è concentrata attorno alla media stessa. Si può quindi notare che, ponendo 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 le limitazioni della varianza come misura della concentrazione di una distribuzione attorno alla sua media come indicato dall'esempio di fig.3, 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. Infatti, ricordando la formula che la definisce, quanto detto si realizza come segue:

$$\begin{aligned}\mu_2 &= \sum_{i=1}^{\infty} (x_i - \mu)^2 f(x_i) = \\ &= \sum_{i=1}^{\infty} x_i^2 f(x_i) - 2\mu \sum_{i=1}^{\infty} x_i f(x_i) + \mu^2 \sum_{i=1}^{\infty} f(x_i) = \\ &= \mu_2' - 2\mu^2 + \mu^2\end{aligned}$$

Quindi la formula richiesta è la seguente:

$$\sigma^2 = \mu'2 - \mu^2$$

## 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 **funzione generatrice dei momenti**. Come sta ad indicare il nome stesso, questa è una funzione che genera i momenti. Essa è definita come segue:

---

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

---

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


---

Questa serie è una funzione del parametro  $\theta$  soltanto, ma si è messo l'indice a M di  $\theta$  per indicare la variabile che si sta considerando. Il parametro  $\theta$  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 che la funzione f sia una funzione di densità per cui la serie nella (1) converga. Sviluppiamo ora  $e^{\theta x_i}$  in serie di potenze e sommiamo termine a termine. Poiché la serie di potenze di  $e^z = 1 + (z/1!) + (z^2/2!) + \dots$  segue dalla formula (1) e da quella che definisce i momenti dall'origine che

$$\begin{aligned} (2) M_X \theta &= \sum_{i=1}^{\infty} (1 + (\theta/1!)x_i + (\theta^2/2!)x_i^2 + \dots) f(x_i) = \\ &= \sum_{i=1}^{\infty} f(x_i) + (\theta/1!) \sum_{i=1}^{\infty} x_i f(x_i) + (\theta^2/2!) \sum_{i=1}^{\infty} x_i^2 f(x_i) + \dots = \\ &= 1 + (\theta/1!)\mu'1 + (\theta^2/2!)\mu'2 + \dots + (\theta^k/k!)\mu'k + \dots \end{aligned}$$

In questo sviluppo si può osservare che il coefficiente di  $(\theta^k)/k!$  è il momento di 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  $\theta$ , i momenti della variabile si possono ottenere semplicemente esaminando lo sviluppo risultante.

Se si desidera un momento particolare può essere più conveniente ricavarlo calcolando la derivata di  $M_X \theta$  per  $\theta=0$  perché, derivando ripetutamente la (2) si ha che

$$\mu'k = [dk M_X \theta] / d\theta^k \text{ per } \theta=0$$

Infatti per  $k=1$  si ha che

$$M'_X \theta = \mu'1 + \mu'2\theta + 1/2\mu'3\theta^2 + \dots$$

e quindi

$$M'_X \theta \text{ per } \theta=0 \text{ è uguale a } \mu'1$$

per  $k=2$  si ha che

$$M''_X \theta = \mu'2 + \mu'3\theta + \dots$$

e quindi

$$M''_X \theta \text{ per } \theta=0 \text{ è uguale a } \mu'2$$

e così via.

## Distribuzione binomiale

*[DOMANDA D'ESAME: che tipo di problema risolve questa funzione di densità?]*

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 fatte n esecuzioni indipendenti dell'esperimento e la variabile X rappresenti il numero di successi che si otterranno nelle n esecuzioni. Affrontiamo ora il problema di determinare la probabilità di ottenere x successi in n esecuzioni dell'esperimento, cioè che  $X=x$ . Per ricavare la formula richiesta determiniamo dapprima la probabilità di ottenere x successi consecutivi seguiti da n-x insuccessi. Questi n eventi sono ovviamente indipendenti, perciò ricordando una formula vista a suo tempo e valida per eventi indipendenti, e cioè  $P\{A1 \cap A2\} = P\{A1\}P\{A2\}$ , questa probabilità è data da  $p * p * p * \dots * p(x \text{ volte}) * q * q * \dots * q(n-x \text{ volte}) = p^x q^{n-x}$ . 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 non è altro che il numero di permutazioni possibili di n oggetti di cui x sono uguali tra loro, cioè le p, ed n-x sono anch'essi uguali tra loro, cioè le q, ma il numero di tali permutazioni è uguale a

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

Ora, ricordando la formula  $P\{A1UA2\}=P\{A1\}+P\{A2\}$ , 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  $p^x q^{n-x}$  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  $p^x q^{n-x}$  per la quantità espressa dalla (3). La probabilità risultante, che è 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** e cioè

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

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

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

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

$$\left\{ \frac{n!}{x!(n-x)!} \right\} q^{n-x} p^x = \left[ \frac{n!}{x!(n-x)!} \right] p^x q^{n-x}$$

che è precisamente il valore della funzione di densità binomiale.

Come risultato si può scrivere

$$(q+p)^n = \sum_{x=0}^n \left[ \frac{n!}{x!(n-x)!} \right] p^x q^{n-x}$$

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

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

**Esercitazione.** Come esempio di 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, due gettate presentino come risultato 1. In questo caso il successo consiste nell'ottenere come risultato 1. In questo caso si ha quindi che  $p=1/6$ ,  $q=1-p=5/6$  ed  $n=5$ . Quando si applica la funzione di densità binomiale si ottiene

$$f(2) = \left[ \frac{5!}{2!(3!)} \right] (1/6)^2 (5/6)^3 = 0,16$$

Si voglia inoltre calcolare la probabilità di ottenere al massimo due gettate che presentino 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 densità binomiale si ha

$$f(0) = \left[ \frac{5!}{0!(5!)} \right] (1/6)^0 (5/6)^5 = 0,40$$

$$f(1) = \left[ \frac{5!}{1!(4!)} \right] (1/6)^1 (5/6)^4 = 0,40$$

Poiché queste sono tre probabilità di eventi reciprocamente esclusivi, si ha che

$$P(X \leq 2) = f(0) + f(1) + f(2) = 0,40 + 0,40 + 0,16 = 0,96 \text{ (96\%)}$$

**Esempio.** Come secondo esempio consideriamo il calcolo della probabilità di ottenere almeno un successo quando si tirano 20 colpi ad un bersaglio se la probabilità di successo per un singolo colpo è uguale a 1/10. In questo caso si ha  $p=1/10$  e  $n=20$ . La probabilità che

$$\begin{aligned} P(X \leq 2) &= 1 - f(0) - f(1) = \\ &= 1 - \left[ \frac{20!}{0!(20!)} \right] (1/10)^0 (9/10)^{20} - \left[ \frac{20!}{1!(19!)} \right] (1/10)^1 (9/10)^{19} = \\ &= 1 - 0,122 - 0,270 = 0,608 \text{ (60,8\%)} \end{aligned}$$

La validità di usare il modello binomiale in questo esempio non è così ovvia come lo è nell'esempio precedente. Infatti la formula binomiale è stata ricavata sulla base di esecuzioni dell'esperimento indipendenti fra loro, con  $p$  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 aumentassero leggermente con la pratica. Se invece venisse usata ogni volta una persona diversa,  $p$  indubbiamente potrebbe cambiare da prova a prova. Quindi, al momento di interpretare una probabilità risultante come quella uguale a 0,608 appena calcolata, è importante prendere in considerazione le possibili deviazioni dalle ipotesi fondamentali.

[Questa dimostrazione generalmente non la chiede, ma se la chiede vuole l'impostazione corretta del primo passaggio]

### **Momenti binomiali**

Calcoliamo ora i primi due momenti dall'origine di una distribuzione binomiale. Per illustrare i due metodi per il calcolo dei momenti questi momenti vengono calcolati dapprima direttamente dalla definizione, e poi indirettamente tramite la funzione generatrice dei momenti. Applicando la definizione alla funzione di densità binomiale espressa dalla (4) vista precedentemente, si può scrivere la media

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

che si può anche scrivere, raccogliendo  $np$  fuori dalla sommatoria

$$(1) \mu = np \sum_{x=1}^n \left[ \frac{(n-1)!}{(x-1)!(n-x)!} \right] p^{x-1} q^{n-x}$$

Ponendo ora  $y=x-1$  si ottiene

$$\mu = np \sum_{y=0}^{n-1} \left[ \frac{(n-1)!}{y!(n-y-1)!} \right] p^y q^{n-y-1}$$

dove la quantità che viene sommata è la probabilità binomiale di  $y$  successi in  $n-1$  esecuzioni. Poiché la somma viene fatta su tutti i possibili valori di  $y$ , essa deve essere uguale a 1. Si ha quindi che

$$\mu = np * 1 = np$$

Il momento del secondo ordine dall'origine di una variabile binomiale si ottiene usando la seguente identità

$$x^2 = x(x-1) + x$$

Dalla definizione si ha che

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

Poiché i termini nell'ultima somma per  $x=0$  e  $x=1$  sono uguali a 0 a causa del fattore  $x(x-1)$ , la sommatoria può cominciare con  $x=2$  e quindi

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

Se  $n(n-1)p^2$  viene raccolto fuori dalla sommatoria si ha che

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

Ponendo ora  $z=x-2$ , si ottiene

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

La quantità che viene sommata a secondo membro è la probabilità binomiale di  $z$  successi in  $n-2$  esecuzioni dell'esperimento. Poiché la somma viene fatta su tutti i possibili valori di  $z$ , essa deve essere uguale a 1. Si ottiene quindi che

$$\mu'2 = [n(n-1)p^2 * 1] + \mu = n(n-1)p^2 + np$$

La varianza di una variabile binomiale si può calcolare applicando una formula già vista, e cioè che

$$\sigma^2 = \mu'2 - \mu^2$$

Così facendo si ha che

$$\begin{aligned}\sigma^2 &= n(n-1)p^2 + np - n^2p^2 = \\ &= -np^2 + np = \\ &= np(1-p) = \\ &= npq\end{aligned}$$

Lo scarto quadratico medio è quindi dato da

$$\sigma = \sqrt{npq}$$

Calcoliamo ora i primi due momenti dall'origine di una variabile binomiale, indirettamente, cioè, tramite la funzione generatrice dei momenti. Quest'ultima, per una variabile binomiale, è data dalla

$$M_X\theta = \sum_{x=0}^n e^{\theta x} [n!/x!(n-x)!] p^x q^{n-x} = \sum_{x=0}^n [n!/x!(n-x)!] [pe^\theta]^x q^{n-x}$$

Ma come si è già visto l'ultima volta, la somma all'ultimo membro si può scrivere come un binomio elevato all'ennesima potenza e quindi si ha che

$$(1) M_X\theta = (q + pe^\theta)^n$$

I momenti richiesti si possono ottenere applicando una formula già nota, e cioè

$$\mu'_k = d^k M_X\theta / d\theta^k \big|_{\theta=0}$$

Derivando la (1) due volte rispetto a  $\theta$  si ottiene

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

$$M''_X\theta = npe^\theta (q + pe^\theta)^{n-1} + npe^\theta [(n-1)(q + pe^\theta)^{n-2} pe^\theta] = npe^\theta (q + pe^\theta)^{n-2} [q + pe^\theta + (n-1)pe^\theta] = npe^\theta (q + pe^\theta)^{n-2} (q + npe^\theta)$$

I valori di queste due derivate calcolate per  $\theta=0$  sono rispettivamente i valori della media  $\mu$  e di  $\mu'^2$ , infatti

$$\begin{aligned} \mu &= M'_X\theta \big|_{\theta=0} = np(q+p)^{n-1} = np(1-p+p)^{n-1} = np \end{aligned}$$

e

$$\begin{aligned} \mu'^2 &= M''_X\theta \big|_{\theta=0} = np(q+p)^{n-2}(q+np) = np(1-p+p)^{n-2}(1-p+np) = np(1-p+np) \end{aligned}$$

Si può osservare che, nel caso di una variabile binomiale, i momenti sono più facili da ottenere indirettamente tramite la funzione generatrice dei momenti anziché 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 sono piuttosto lunghi, perciò sarebbe molto utile disporre di una conveniente approssimazione della distribuzione binomiale. Risulta che, per  $n$  grande, ci sono due ben note funzioni di densità che forniscono buone approssimazioni della funzione di densità 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:

$$(2) f(x) = [e^{-\mu} \mu^x] / x!$$

dove il parametro  $\mu$  è la media della distribuzione. Sebbene la distribuzione di Poisson venga introdotta qui come un'approssimazione della distribuzione binomiale, essa è una distribuzione ben nota ed utile di per sé e perciò non è a considerare semplicemente come un'approssimazione della distribuzione binomiale.

Per verificare che la distribuzione di Poisson sia 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  $n$  tende all'infinito e  $p$  tende a 0 in modo tale che la media  $\mu = np$  rimanga fissa. Il risultato ottenuto si può esprimere sotto forma di teorema il cui enunciato è il seguente:

---

se la probabilità  $p$  di successo in una singola esecuzione tende a 0 mentre il numero di esecuzioni  $n$

---

tende all'infinito in modo tale che la media  $\mu=np$  rimanga fissa, allora la distribuzione binomiale tenderà alla distribuzione di Poisson con media  $\mu$ .

Le figure 1 e 2 sono state costruite per indicare la rapidità con cui la distribuzione binomiale tende alla distribuzione di Poisson.

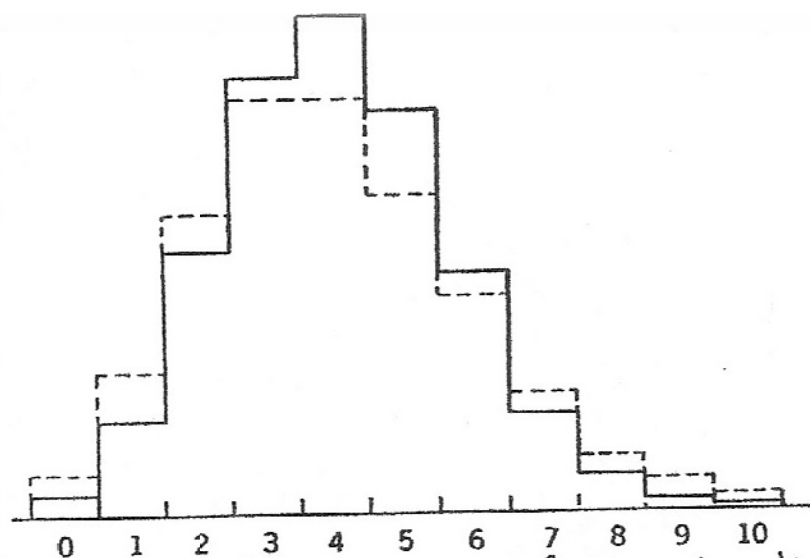


Fig. 1 Distribuzioni binomiale (—) e di Poisson (----) per  $\mu=4$  e  $p=\frac{1}{3}$ .

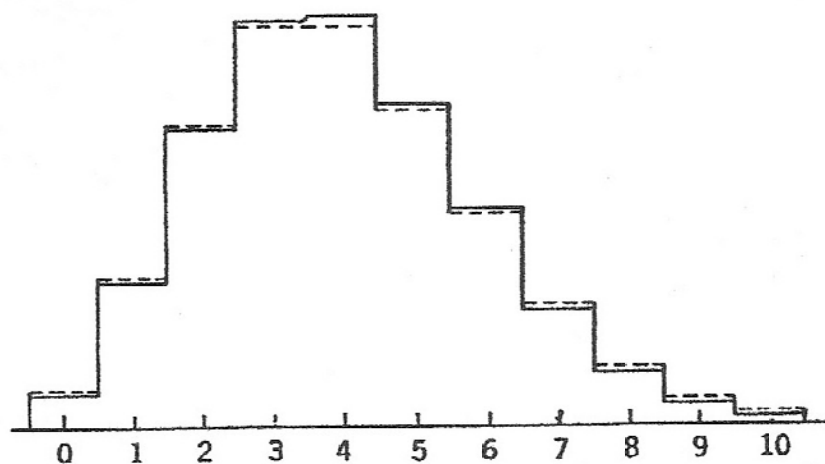


Fig. 2 Distribuzioni binomiale (—) e di Poisson (----) per  $\mu=4$  e  $p=\frac{1}{24}$ .

Le linee tratteggiate rappresentano la distribuzione di Poisson fissa per  $\mu=4$  e le linee continue la distribuzione binomiale rispettivamente per  $p=1/3$  e  $p=1/24$ . 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  $p < 0,05$ .

Poiché la distribuzione di Poisson è stata ottenuta dalla distribuzione binomiale mantenendo fissa la media  $\mu=np$  e permettendo ad  $n$  di diventare infinito, segue che i momenti della distribuzione binomiale possono consentire di ricavare i momenti della distribuzione di Poisson mediante passaggi al limite, così la media di una distribuzione di Poisson deve essere  $\mu=np$  mentre la sua varianza deve essere il valore limite di  $npq$ . Ma, con la media  $\mu=np$  fissa,  $p$  tenderà a 0 quando  $n$  tende all'infinito,

perciò

$$\sigma^2 = \lim_{n \rightarrow \infty} npq = \lim_{n \rightarrow \infty} \mu q = \lim_{p \rightarrow 0} \mu(1 - p) = \mu$$

Questo fornisce 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 il numero di chiamate telefoniche su di una linea in un intervallo di tempo fisso, per il numero di screpolature sopra una lamina metallica, ecc. Questi problemi sono tutti del tipo in cui una variabile casuale si distribuisce nel tempo o 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 piccolo 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 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 in qualsiasi regione dello spazio di dimensione fissata seguirà una distribuzione di Poisson. Un tale intervallo può essere a un, due o tre dimensioni, così il numero di screpolature in una data lunghezza di filo metallico o in un data area di stoffa o in dato blocco di calcestruzzo, sono tutti esempi di problemi spaziali a cui si può applicare la 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 possiedano una distribuzione di Poisson, viene chiamato **un processo di Poisson**. Così la distribuzione di Poisson è una distribuzione 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 della distribuzione 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 la media

$$\begin{aligned}\mu &= np = \\ &= 200(0,02) = \\ &= 4\end{aligned}$$

quindi, usando la distribuzione di Poisson dove la variabile  $X$  rappresenta il numero di valvole difettose in una scatola di 200 valvole, la risposta approssimata è data da

$$\begin{aligned}P(X \leq 5) &= \sum_{x=0}^5 (e^{-4} 4^x) / x! = \\ &= e^{-4} [1 + (4/1!) + (4^2/2!) + (4^3/3!) + (4^4/4!) + (4^5/5!)] = \\ &= 0,785\end{aligned}$$

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 dato da

$$\begin{aligned}|\Delta P / P| &= |(0,785 - 0,788) / 0,788| = \\ &= 0,003 / 0,788 = \\ &= 0,0038 = \\ &= 3,8\%\end{aligned}$$

Ciò conferma quanto è già stato detto, cioè che l'approssimazione di Poisson si può ritenere sufficientemente accurata per la maggior parte delle applicazioni se, come avviene in questo caso,  $n \geq 100$  e  $p \leq 0,05$ , infatti qui  $n = 200$  e  $p = 0,02$ .

**Esempio.** Come esempio d'uso della distribuzione di Poisson, come modello per un processo di Poisson, consideriamo il modello 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

$$\begin{aligned} P(X \leq 8) &= \sum_{x=0}^8 (e^{-5} 5^x) / x! = \\ &= 0,932 = \\ &= 93,2\% \end{aligned}$$

Di conseguenza la probabilità richiesta è data da

$$\begin{aligned} P(X > 8) &= 1 - P(X \leq 8) = \\ &= 1 - 0,932 = \\ &= 0,068 = \\ &= 6,8\% \end{aligned}$$

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 degli eventi in questo intervallo tramite la distribuzione di Poisson, però con la media  $\mu$  aggiustata per quell'intervallo. Per esempio, la soluzione del problema di calcolare la probabilità che si ricevano al massimo 6 chiamate telefoniche in un periodo di due minuti, quando il numero medio di chiamate al minuto è uguale a 5, si ottiene scegliendo  $\mu=10$  e calcolando la probabilità come segue:

$$\begin{aligned} P(X \leq 6) &= \sum_{x=0}^6 (e^{-10} 10^x) / x! = \\ &= 0,13 = \\ &= 13\% \end{aligned}$$

Una soluzione basata sul trattare questo come un esperimento a 2 stadi, con ciascun minuto di tempo come uno stadio, sarebbe origine a calcoli molto lunghi. Si dovrebbe quindi essere spinti a risolvere il problema in questo modo, così da apprezzare meglio la precedente proprietà della distribuzione di Poisson.

### Variabili casuali continue

Prima di presentare alcune utili distribuzioni di probabilità di variabili casuali continue, definiamo il momento d'ordine  $k$  d'una variabile di questo tipo. Sia  $f(x)$  la funzione di densità di una variabile casuale continua  $X$  che è nulla esternamente ad un certo intervallo finito  $[a, b]$ , allora

---

il momento d'ordine  $k$  dall'origine della distribuzione della variabile casuale continua  $X$  la cui densità è  $f$  è dato da

---

$$(1) \mu'_k = E[X^k] = \int_a^b x^k f(x) dx$$


---

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

$$(2) \mu_k = E[(X-\mu)^k] = \int_a^b (x-\mu)^k f(x) dx$$

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

---

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

---

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


---

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 *meno infinito* e  $b$  a *più infinito*. 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 della variabile casuale  $X$ , per dire  $g(X)$ , sia veramente il valore medio di quella funzione. Indichiamo con  $Y$  la variabile  $g(X)$ , cioè  $Y=g(X)$ , allora, conoscendo la funzione di densità  $f(x)$  della variabile  $X$ , è teoricamente possibile determinare la funzione di densità  $h(y)$  della variabile  $Y$ . Il valore atteso di  $g(X)$  è ovviamente uguale al valore atteso di  $Y$ . Perciò, se  $h(y)$  è disponibile, il valore atteso di  $Y$  si può esprimere nella forma

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

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 formula sta nel fatto che essa non richiede di determinare 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 di momenti

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

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

Se  $e^{\theta x}$  si sviluppa in serie di potenze e si integra termine a termine si troverà che  $M_X\theta$  assumerà la stessa forma estesa già vista nel caso discreto, e cioè

$$M_X\theta = 1 + (\theta/1!)(\mu'_1) + (\theta^2/2!)(\mu'_2) + \dots + (\theta^k/k!)(\mu'_k) + \dots$$

Per poter generare i momenti di una funzione  $g(X)$  della variabile casuale continua  $X$  è necessario generalizzare la definizione di funzione generatrice di momenti. Dal modo in cui  $M_X\theta$  genera i momenti è chiaro che i momenti della funzione  $g(X)$  saranno generati sostituendo nella (4)  $e^{\theta x}$  con  $e^{\theta g(x)}$ . La definizione è la seguente:

---

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

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


---

Questa è la forma generalizzata della funzione generatrice dei momenti.

### Proprietà delle funzioni generatrici dei momenti

Consideriamo ora due interessanti proprietà delle funzioni generatrici dei momenti. Sia  $c$  una costante e sia  $h(X)$  una funzione della variabile  $X$  per cui la funzione generatrice dei momenti esiste. Allora, poiché nella (5) vista ieri la  $g(X)$  rappresenta una funzione arbitraria, si può scegliere  $g(X)=c \cdot h(X)$ . Di conseguenza si ha che

$$\begin{aligned} M_{ch(X)}\theta &= \int_{-\infty}^{+\infty} e^{\theta ch(x)} f(x) dx = \\ &= M_{h(X)c}\theta \end{aligned}$$

La seconda proprietà si ottiene scegliendo sempre nella (5)  $g(X)=h(X)+c$ , allora

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

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

- $M_{cg(X)}\theta = M_{g(X)c}\theta$
- $M_{g(X)+c}\theta = e^{c\theta} M_{g(X)}\theta$

Queste due proprietà ci permettono di disporre nel modo indicato di una costante  $c$  che moltiplica od è aggiunta ad una funzione  $g(X)$ . Sostituendo integrali con somme si può dimostrare che queste formule si applicano anche alle variabili discrete.

Si fa l'ipotesi che  $g(x)$  e  $f(x)$  siano tali che l'integrale nella (5) vista in precedenza risulti finito.

### Distribuzione rettangolare o uniforme

Forse la variabile casuale continua più semplice è quella il cui valore è costante su un certo intervallo  $[a,b]$  ed è uguale a 0 altrove. Questa definisce quella che è conosciuta come la **distribuzione rettangolare** o **uniforme**. Essa è definita come segue:

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

Il grafico di una tipica distribuzione rettangolare è mostrato nella figura seguente.

Fig.1



Il momento d'ordine  $k$  dall'origine della distribuzione rettangolare è facile da calcolare. Per semplicità, se  $a=0$  e  $b=1$ , dalla definizione si ha che

$$\begin{aligned} \mu'_k &= \int_0^1 x^k \cdot 1 \, dx = \\ &= \left[ \frac{1}{k+1} x^{k+1} \right] (1,0) = \\ &= 1/(k+1) \end{aligned}$$

La funzione generatrice dei momenti è anch'essa facile da calcolare. Si ha che

$$\begin{aligned} M_x \theta &= \int_0^1 e^{\theta x} \cdot 1 \, dx = \\ &= \left[ \frac{1}{\theta} e^{\theta x} \right] (1,0) = \\ &= \frac{1}{\theta} (e^{\theta} - 1) \end{aligned}$$

Se si volesse ottenere il momento d'ordine  $k$  dalla  $M_x \theta$  sarebbe necessario sviluppare  $e^{\theta}$  in serie di potenze. Così facendo si ha che

$$\begin{aligned} M_x \theta &= \frac{1}{\theta} [1 + (\theta/1!) + (\theta^2/2!) + \dots - 1] = \\ &= 1 + (\theta/2!) + (\theta^2/3!) + \dots + [\theta^k/(k+1)!] + \dots \end{aligned}$$

Osservando questo sviluppo, poiché sappiamo che  $\mu'_k$  è il coefficiente di  $\theta^k/k!$ , allora possiamo scrivere  $\mu'_k = 1/(k+1)$

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 in questo caso per calcolare i momenti. Il calcolo diretto, in questo caso, è ovviamente molto più semplice.

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 è definita come segue:

---


$$(2) f(x) = ce^{(-1/2)((x-a)/b)^2}$$


---

Dove  $a$ ,  $b$  e  $c$  sono parametri che fanno della  $f(x)$  una funzione di densità di probabilità, quindi questi parametri devono essere tali che l'integrale della  $f(x)$  da meno infinito a più infinito sia uguale a 1. Come risultato vedremo fra poco che ci sono in effetti due soli parametri indipendenti che determinano questa funzione di densità. Dalla (2) è chiaro che la curva è simmetrica rispetto alla retta verticale  $x=a$ , quindi per simmetria la media deve essere data da

$$\mu = a$$

Calcoliamo ora i momenti indirettamente tramite la funzione generatrice dei momenti. Inoltre, poiché è più facile calcolare in questo caso i momenti dalla media piuttosto che quelli dall'origine, consideriamo il calcolo di  $M_{X-\mu}$ . Dalla forma generalizzata della funzione generatrice dei momenti espressa dalla (5) vista in precedenza, con  $g(X) = X - \mu$ , si ha che

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

Ponendo ora  $z = (x-\mu)/b$  e  $dx = b dz$ ,

$$M_{X-\mu} \theta = bc \int_{-\infty}^{+\infty} e^{\theta bz - (z^2)/2} dz$$

Riscriviamo ora l'esponente sotto il segno di integrale come segue

$$\theta bz - (z^2)/2 = -(1/2)(z - \theta b)^2 + (1/2)\theta^2 b^2$$

per cui

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

ponendo poi  $t = z - \theta b$  e  $dz = dt$ , per cui si ha che

$$M_{X-\mu} \theta = bce^{(1/2)\theta^2 b^2} \int_{-\infty}^{+\infty} e^{-(t^2)/2} dt$$

L'integrale a secondo membro è un integrale noto che è uguale alla radice di  $2\pi$ , per cui si può scrivere che

$$(3) M_{X-\mu} \theta = \sqrt{2\pi} bce^{(1/2)\theta^2 b^2}$$

Ricordando ora che  $M_X \theta = 1 + (\theta/1!)(\mu'_1) + (\theta^2/2!)(\mu'_2) + \dots$  segue che, per qualsiasi funzione generatrice dei momenti si ha che

$$M(0) = 1$$

da cui, per la (3), segue che

$$\sqrt{2\pi} bc = 1$$

e quindi

$$(4) M_{X-\mu} \theta = e^{(1/2)\theta^2 b^2}$$

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

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

Esaminando lo sviluppo a secondo membro, si può osservare che mancano le potenze dispari di  $\theta$  e quindi che i momenti dispari della variabile  $X$  dalla sua media devono essere nulli. Il coefficiente di  $\theta^2/2!$  è il momento del secondo ordine della variabile  $X$  dalla sua media, perciò  $b^2 = \mu_2 = \sigma^2$  e quindi  $b = \sigma$ .

Poiché  $\sqrt{2\pi} bc = 1$  allora  $c = 1/[\sigma\sqrt{2\pi}]$  di conseguenza la distribuzione normale definita dalla (2) si può scrivere nella forma

$$(5) f(x) = 1/[\sigma\sqrt{2\pi}] e^{(-1/2)(x-\mu/\sigma)^2}$$

Questo risultato mostra che una distribuzione normale è completamente determinata specificando la sua media ed il suo scarto quadratico medio. Si fa notare che la sola differenza fra la (2) e la (5) è che i parametri  $a$ ,  $b$  e  $c$  nella (2) sono ora stati ridotti a due soli parametri indipendenti a cui è stato dato 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 (4) vista in precedenza sostituendo  $b^2$  con  $\sigma^2$  ed usando la seconda delle due proprietà di una funzione generatrice dei momenti. In forma generalizzata, cioè  $M_{g(X)+c}\theta = e^{c\theta} M_{g(X)}\theta$  con  $g(X)=X$  e  $c=-\mu$ , così si ottiene che

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

da cui, ponendo a primo membro l'espressione della (4) suddetta e risolvendo rispetto ad  $M_X\theta$  si ottiene

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

Per interpretare lo scarto quadratico medio geometricamente, consideriamo i punti di flesso di una curva normale. Calcolando le prime due derivate di una funzione di densità normale espressa dalla (5) vista prima, cioè

$$f(x) = [1/\sigma\sqrt{2\pi}]e^{-(1/2)(x-\mu/\sigma)^2}$$

si ottiene

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

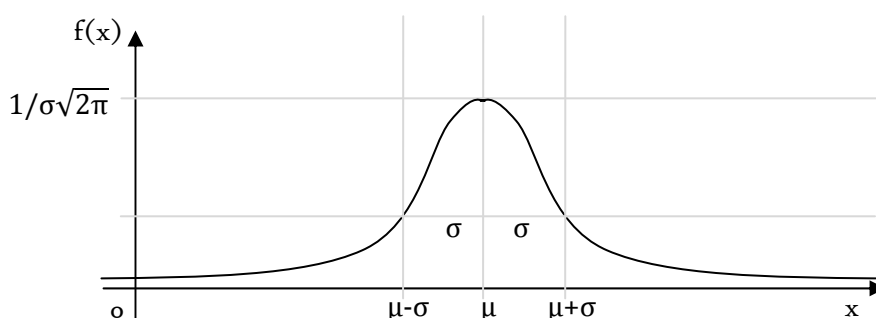
$$= [-1/\sigma^2][x-\mu]f(x)$$

$$f''(x) = -(1/\sigma^2)f(x) - (1/\sigma^2)(x-\mu)[-(1/\sigma^2)(x-\mu)f(x)] =$$

$$= -(1/\sigma)f(x)[1-(x-\mu/\sigma)^2]$$

Da queste derivate è chiaro che c'è soltanto un punto di massimo che si trova in  $x=\mu$ . Dalla derivata seconda segue che i punti di flesso si trovano in  $x=\mu\pm\sigma$ . Geometricamente allora, per una distribuzione normale, lo scarto quadratico medio  $\sigma$  rappresenta la distanza dei punti di flesso dall'asse di simmetria. Ciò implica che la curva normale rivolge la concavità verso il basso fra  $x=\mu-\sigma$  e  $x=\mu+\sigma$  e la concavità verso l'alto all'esterno di questo intervallo. Il grafico di una curva normale tipica è rappresentato nella figura seguente.

Fig.1



Per approfondire ulteriormente il ruolo che il parametro  $\sigma$  ha nel determinare la funzione di densità normale calcoliamo l'area sottesa dalla funzione di densità normale negli intervalli simmetrici mostrati in figura. Così la probabilità che la variabile  $X$ , avente una funzione di densità normale, cada nell'intervallo  $(\mu-\sigma, \mu+\sigma)$  è data da

$$\int_{\mu-\sigma}^{\mu+\sigma} [1/\sigma\sqrt{2\pi}]e^{-(1/2)(x-\mu/\sigma)^2} dx$$

Ponendo  $z=x-\mu/\sigma$  allora  $dx=\sigma dz$  allora l'integrale si riduce a

$$\int_{-1}^1 [1/\sqrt{2\pi}]e^{-(z/2)^2} dz = 2 \int_0^1 [1/\sqrt{2\pi}]e^{-(z/2)^2} dz$$

Il valore dell'ultimo integrale, che si può osservare nelle apposite tavole, è uguale a 0,3413 per cui 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 tra  $(\mu-3\sigma, \mu+3\sigma)$  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 praticamente tutta la probabilità, cioè il 99,7%, giace nell'intervallo  $(\mu-3\sigma, \mu+3\sigma)$ .

L'unità di misura, data dalla trasformazione  $z=x-\mu/\sigma$  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  $f(z) = [1/\sqrt{2\pi}]e^{-(z/2)^2}$ .

### Distribuzioni normali come approssimazioni della distribuzione binomiale

In precedenza, la distribuzione di Poisson è stata introdotta come un'approssimazione della distribuzione binomiale quando  $n$  è grande e  $p$  è piccolo. Allora si disse che un'altra distribuzione fornisce una buona approssimazione sempre per  $n$  grande a prescindere dal valore di  $p$ . La **distribuzione normale** è la distribuzione che gode di questa proprietà. Prima di indagare sulla natura di questa approssimazione, consideriamo un esempio numerico. Sia  $n=12$  e  $p=1/3$  e costruiamo il grafico della corrispondente distribuzione binomiale. Questo valore di  $n$  non è certamente grande così che non ci si dovrebbe aspettare in questo caso una buona approssimazione di tipo normale. In questo caso la funzione di densità binomiale è la seguente

$$f(x) = 12! / x!(12-x)! (1/3)^x (2/3)^{12-x}$$

Poiché la  $f(x)$ , per costruire il grafico, si deve calcolare per tutti i valori di  $x$  da 0 a 12, è più facile calcolare ciascun valore dopo il primo da quello precedente anziché calcolarlo da solo. Per questa funzione, infatti, si può verificare che

$$f(x+1) = (12-x)/(x+1) (1/2) f(x)$$

Dopo aver calcolato  $f(0)$ , quest'ultima relazione è stata usata per ottenere i restanti valori riportati nella tab.1. Il grafico di questa distribuzione binomiale è mostrato il fig.1.

Tab. 1

$f(0) =$	.007707	$f(7) = \frac{3}{7}f(6) =$	.047687
$f(1) = 6f(0) =$	.046242	$f(8) = \frac{5}{16}f(7) =$	.014902
$f(2) = \frac{11}{4}f(1) =$	.127166	$f(9) = \frac{2}{9}f(8) =$	.003312
$f(3) = \frac{5}{2}f(2) =$	.211943	$f(10) = \frac{3}{20}f(9) =$	.000497
$f(4) = \frac{9}{8}f(3) =$	.238436	$f(11) = \frac{1}{11}f(10) =$	.000045
$f(5) = \frac{4}{5}f(4) =$	.190749	$f(12) = \frac{1}{24}f(11) =$	.000002
$f(6) = \frac{7}{12}f(5) =$	.111270		

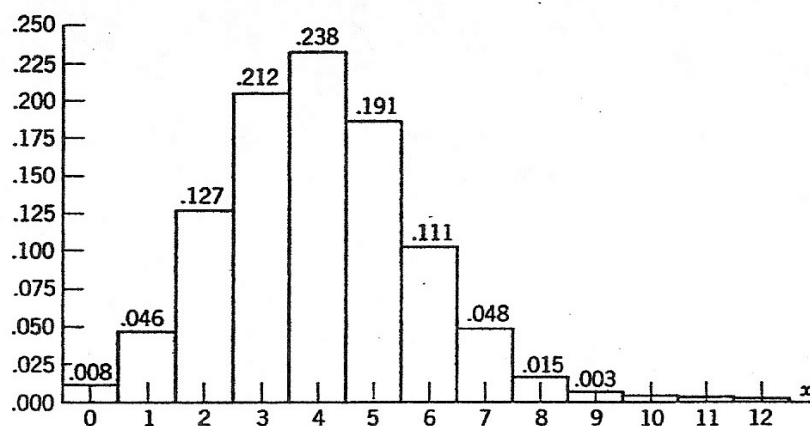


Fig. 1 . Distribuzione binomiale,  $p = 1/3$ ,  $n = 12$ .

Sembrerebbe che questo istogramma potrebbe essere approssimato abbastanza bene dalla curva

normale appropriata. Poiché una curva normale è completamente determinata dalla sua media e dallo scarto quadratico medio, una 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 qui  $\mu = n \cdot p = 12(1/3) = 4$

e

$$\sigma = \sqrt{npq} = \sqrt{12(1/3)(2/3)} = 1,63$$

Come prova dell'accuratezza dell'approssimazione che si basa sulla curva normale e come esempio dei metodi basati sulla curva normale per approssimare le probabilità binomiali, consideriamo alcuni problemi riferiti alla fig.1 vista in precedenza. Se la probabilità che un tiratore colpisca un bersaglio è uguale a  $1/3$ , cioè se  $p = 1/3$ , e se egli tira 20 colpi, cioè  $n = 20$ , si vuole calcolare la probabilità che egli colpisca almeno 20 bersagli. La risposta esatta si ottiene sommando la  $f(x)$  da  $x = 6$  a  $x = 12$ , che, usando la tab.1 vista precedentemente, è uguale a 0,178 (corretta a 3 cifre decimali), cioè

$$P(X \geq 6) = \sum_{x=6}^{12} f(x) = f(6) + f(7) + \dots + f(12) = 0,178$$

Geometricamente questo valore rappresenta l'area di quella parte dell'istogramma nella fig.1 di prima che giace alla destra di  $x = 5,5$ . Perciò, per approssimare questa probabilità con i metodi che si basano sulla curva normale, è semplicemente necessario calcolare l'area sottesa da quella parte della curva normale approssimante che si trova alla destra di  $x = 5,5$ . Poiché la curva approssimante vista in precedenza ha  $\mu = 4$  e  $\sigma = 1,63$ , segue che

$$z = (x - \mu) / \sigma = (5,5 - 4) / 1,63 = 0,92$$

Quindi, con questi valori di  $\mu$  e  $\sigma$ , il cambiamento di variabile  $z = (x - \mu) / \sigma$  fornisce che

$$\begin{aligned} & \int_{5,5}^{+\infty} \frac{1}{\sigma \sqrt{2\pi}} e^{-(1/2)(x-\mu/\sigma)^2} dx = \\ & = \int_{0,92}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-(z)^2/2} dz \end{aligned}$$

Ma come si può osservare nelle apposite tavole, l'area alla destra di  $z = 0,92$  è uguale a 0,179 che, confrontata con il valore corretto 0,178, è sbagliata soltanto dello 0,56%. Infatti, in questo caso, l'errore relativo è dato da

$$\Delta P / P = (0,179 - 0,178) / 0,178 = 0,001 / 0,178 = 0,0056 = 0,56\%$$

Per verificare l'accuratezza dei metodi che si basano sulla curva normale su un intervallo più breve, calcoliamo la probabilità che il tiratore abbia successo precisamente in 6 colpi su 12. Dalla tab.1 la risposta (corretta a 3 cifre decimali) è  $f(6) = 0,111$ . Per approssimare questo valore è necessario calcolare l'area sottesa dalla curva normale approssimante tra  $x_1 = 5,5$  e  $x_2 = 6,5$ . In questo caso si ha che

$$z_1 = (x_1 - \mu) / \sigma = (5,5 - 4) / 1,63 = 0,92$$

$$z_2 = (x_2 - \mu) / \sigma = (6,5 - 4) / 1,63 = 1,53$$

Dalle apposite tavole si può osservare che l'area sottesa dalla curva normale standard tra  $z = 0$  e  $z = 0,92$  è data da  $A_1 = 0,3413$  e quella sottesa fra  $z = 0$  e  $z = 1,53$  è uguale a  $A_2 = 0,4370$  quindi l'area richiesta è uguale a  $A_2 - A_1 = 0,4370 - 0,3413 = 0,116$  che è sbagliata rispetto al valore corretto di circa il 4,5%. Infatti in questo caso l'errore relativo è dato da

$$\Delta P / P = (0,116 - 0,111) / 0,111 = 0,045 = 4,5\%$$

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 usati per rendere credibile un famoso teorema che garantisca una buona approssimazione della curva normale alla distribuzione binomiale. Se  $n$  è sufficientemente grande, questo teorema che enunceremo ora senza dimostrazione è il seguente:

---

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  $X - np / \sqrt{npq}$  ha una distribuzione che tende alla distribuzione normale standard quando  $n$  tende all'infinito.

---

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

Le figure 1 e 2 indicano come rapidamente tenda alla normalità la distribuzione della variabile

$$Z = (X - np) / \sqrt{npq}$$

quando  $p=1/3$  e  $n=24$  e  $n=48$  rispettivamente. La scala dell'asse  $y$  per questi due grafici è approssimativamente 17 volte quella dell'asse  $x$ .

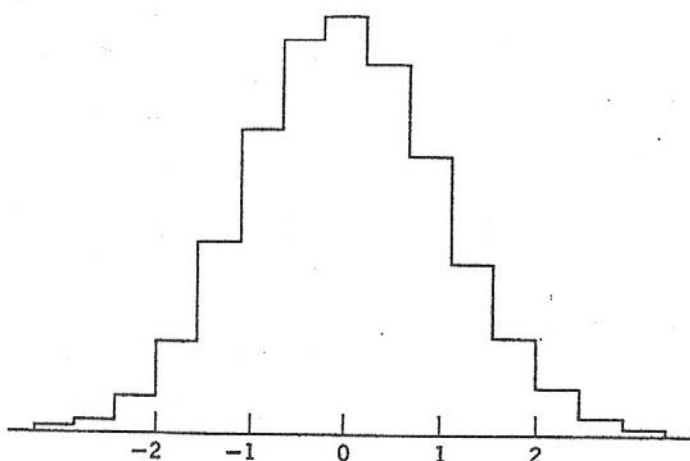


Fig. 1. Distribuzione di  $(X-np)/\sqrt{npq}$  per  $p = 1/3$  e  $n = 24$ .

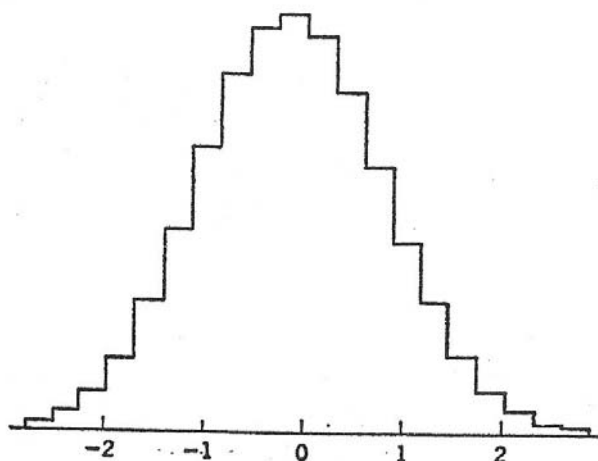


Fig. 2. Distribuzione di  $(X-np)/\sqrt{npq}$  per  $p = 1/3$  e  $n = 48$ .

Le due approssimazioni della distribuzione binomiale, cioè la distribuzione 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 allora i calcoli sono semplici.

**Esercitazione.** Come esempio di impiego della funzione di densità normale come approssimazione di quella binomiale, consideriamo il problema seguente.

Un fabbricante di parti di macchina sostiene che il 10% al massimo delle sue parti sono difettose. Un compratore ha bisogno di 120 di queste parti e, per essere sicuro di riceverle 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, la variabile  $X$  si può trattare come una variabile binomiale con  $n=140$  e  $p=0,90$ . In questo caso questi valori giustificano un'approssimazione di tipo normale in quanto



$$np=140(0,1)=14>5$$

$$p=0,9>1/2$$

Il problema allora è di calcolare la probabilità  $P(X \geq 120)$ . La media e lo scarto quadratico medio sono dati da

$$\mu=np=140(0,90)=0,126$$

$$\sigma=\sqrt{npq}=\sqrt{140(0,90)(0,1)}=3,55$$

quindi

$$\begin{aligned} P(X \geq 120) &= P((X-\mu)/\sigma \geq (120-\mu)/\sigma) = \\ &= P((X-126/3,55) \geq (120-126/3,55)) = \\ &= P(Z \geq -1,69) = 0,95 \end{aligned}$$

Poiché  $Z=(X-126/3,55)$  si può trattare come una variabile normale standard approssimata, questa probabilità che si può trovare delle apposite tavole è uguale a 0,95. Si conclude perciò che, se l'affermazione del fabbricante è valida, il compratore a la probabilità del 95% di ottenere almeno 120 parti senza difetti.

### Distribuzione gamma

Una distribuzione di probabilità che si presenta spesso in vari problemi statistici, come ad esempio nella 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 nel modo seguente:

$$(1) f(x) = [x^{\alpha-1} e^{-(x/\beta)}] / \beta^\alpha \gamma(\alpha) \text{ per } x > 0 \text{ e } 0 \text{ per } x \leq 0$$

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

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

Si dimostra facilmente, integrando per parti, che

$$\gamma(\alpha+1) = \alpha \gamma(\alpha)$$

Se  $\alpha$  è un intero positivo, questa relazione di ricorrenza fornisce il risultato che

$$\gamma(\alpha+1) = \alpha!$$

Come conseguenza di questa proprietà, la *funzione gamma* è chiamata a volte **funzione fattoriale**.

### Momenti

I momenti della distribuzione gamma si calcolano servendosi della (1). Dalla definizione si ha infatti che

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

ponendo  $t=x/\beta$  allora  $dx=\beta dt$  e quindi

$$\mu'_k = [\beta^{k+\alpha-1} \beta] / \beta^\alpha \gamma(\alpha) \int_0^{+\infty} t^{k+\alpha-1} e^{-t} dt$$

Ricordando la (2) si può osservare che l'integrale a secondo membro è  $\gamma(k+\alpha)$ , quindi si può scrivere

$$\mu'_k = \beta^k \gamma(k+\alpha) / \gamma(\alpha)$$

Poiché  $k$  è un intero positivo segue, dall'applicazione ripetuta della relazione di ricorrenza

$$\gamma(\alpha+1) = \alpha \gamma(\alpha) \text{ che}$$

$$\gamma(k+\alpha) = (k+\alpha-1)(k+\alpha-2) \dots \alpha \gamma(\alpha)$$

e quindi si ha infine che

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

Da questa formula, per  $k=1$ , si ottiene che la media è data da

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

e poiché

$$\mu'_2 = \beta^2 (1+\alpha) \alpha$$

ricordando la formula seguente, ovvero

$$(4) \sigma^2 = \mu'^2 - \mu^2 = \beta^2(1 + \alpha) - \beta^2\alpha^2 = \beta^2\alpha$$

### Distribuzione esponenziale

Il caso particolare della distribuzione gamma che si verifica quando  $\alpha=1$  viene chiamato **distribuzione esponenziale**. Essa viene usata sufficientemente spesso da giustificare di trattarla separatamente. La distribuzione esponenziale è rappresentata quindi dalla

$$(5) f(x) = e^{-(x/\beta)}/\beta \text{ per } x > 0 \text{ e } 0 \text{ per } x \leq 0$$

tenuto conto che  $\gamma(1)=1$ . Infatti

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

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 date rispettivamente da  $\mu=\beta$  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'appropriatezza 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 un'apparecchiatura nuova.

**Esercitazione.** Come esempio di impiego pratico della distribuzione esponenziale consideriamo il problema seguente.

*Un costruttore di un'apparecchiatura elettronica ha provato, per esperienza, che la sua apparecchiatura dura in media 2 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 1 anno, si vuole calcolare la probabilità che l'apparecchiatura abbia un guasto prima che scada la garanzia, ovvero la percentuale dei suoi clienti che sarà interessata a qualche riparazione per un guasto prima di 1 anno. Poiché  $\beta$  è la media della distribuzione data dalla (5), e in questo caso  $\beta=2$ , la funzione che si applica in questo caso è  $f(x)=(1/2)e^{-(x/2)}$ . Il problema è quindi quello di calcolare la probabilità  $P(X<1)$ , perciò si ha che*

$$P(X<1) = \int_0^1 (1/2)e^{-(x/2)} dx$$

ponendo  $t=x/2$  allora  $dx=2dt$  e quindi

$$P(X<1) = \int_0^{1/2} e^{-t} dt = -e^{-(1/2)} + 1 = 0,39 = 39\%$$

*Così, anche se la durata media di vita dell'apparecchiatura è il doppio della durata garantita, c'è una probabilità abbastanza elevata, cioè del 39%, che l'apparecchiatura abbia un guasto prima che scada la garanzia.*

### Distribuzione chi-quadrato ( $\chi^2$ )

Un altro caso particolare della distribuzione gamma che ha molte applicazioni statistiche si ricava dalla (1) prendendo  $\beta=2$  e scrivendo  $\alpha=v/2$ . Esso prende il nome di **distribuzione chi-quadrato** la cui funzione di densità è quindi la seguente:

$$(6) f(x) = x^{(v/2)-1} e^{-(x/2)} / 2^{(v/2)} \gamma(v/2) \text{ per } x > 0 \text{ e } 0 \text{ per } x \leq 0$$

Il motivo della sostituzione del parametro  $\alpha$  con il parametro  $v/2$  è che il parametro  $v$  viene normalmente usato per rappresentare i gradi di libertà e quindi possiede un significato intuitivo naturale quando la distribuzione chi-quadrato si applica a certi specifici problemi statistici. E' interessante a questo proposito calcolare la media e la varianza di una variabile chi-quadrato in funzione di questo nuovo parametro, ponendo  $\beta=2$  ed  $\alpha=v/2$  nelle formule (3) e (4), si ottiene che  $\mu=v$  e  $\sigma^2=2v$ . Sebbene i primi due momenti di una distribuzione chi-quadrato siano stati calcolati

facilmente dai corrispondenti momenti di una distribuzione gamma, altre proprietà della distribuzione chi-quadrato si dimostrano più facilmente per mezzo della sua funzione generatrice dei momenti che ora calcoleremo. Per definizione si ha che

$$M_{X\theta} = E[e^{\theta X}] = \frac{1}{2^{(v/2)} \Gamma(v/2)} \int_0^{+\infty} e^{\theta x} x^{(v/2)-1} e^{-(x/2)} dx =$$

$$= \frac{1}{2^{(v/2)} \Gamma(v/2)} \int_0^{+\infty} x^{(v/2)} e^{(x/2)(1-\theta)} dx$$

ponendo ora  $z = [x(1-\theta)]/2$  allora  $dx = 2/(1-\theta) dz$  e quindi

$$M_{X\theta} = \frac{2^{(v/2)-1} (1-\theta)^{-(v/2)-1} 2(1-\theta)^{-1}}{2^{(v/2)} \Gamma(v/2)} \int_0^{+\infty} z^{(v/2)-1} e^{-z} dz$$

Ricordando ora la definizione della funzione gamma espressa dalla (2), si può osservare che l'integrale a secondo membro definisce  $\Gamma(v/2)$  quindi, semplificando, si ottiene infine che

$$M_{X\theta} = (1-\theta)^{-v/2}$$

Poiché la distribuzione gamma dipende da due parametri, essa ha moltissima flessibilità come modello per le distribuzioni reali. Per mostrare questa flessibilità, nella fig.1 sono mostrati i grafici di parecchie funzioni di densità chi-quadrato corrispondenti quindi, come si è detto a  $\beta=2$ , facendo osservare che, poiché  $\beta$  funge da parametro di scala, cambiando il valore di  $\beta$  la corrispondente curva di fig.1 semplicemente si stenderà se  $\beta$  aumenta e si comprimerà se  $\beta$  diminuisce mantenendo però naturalmente uguale a 1 l'area da essa sottesa.

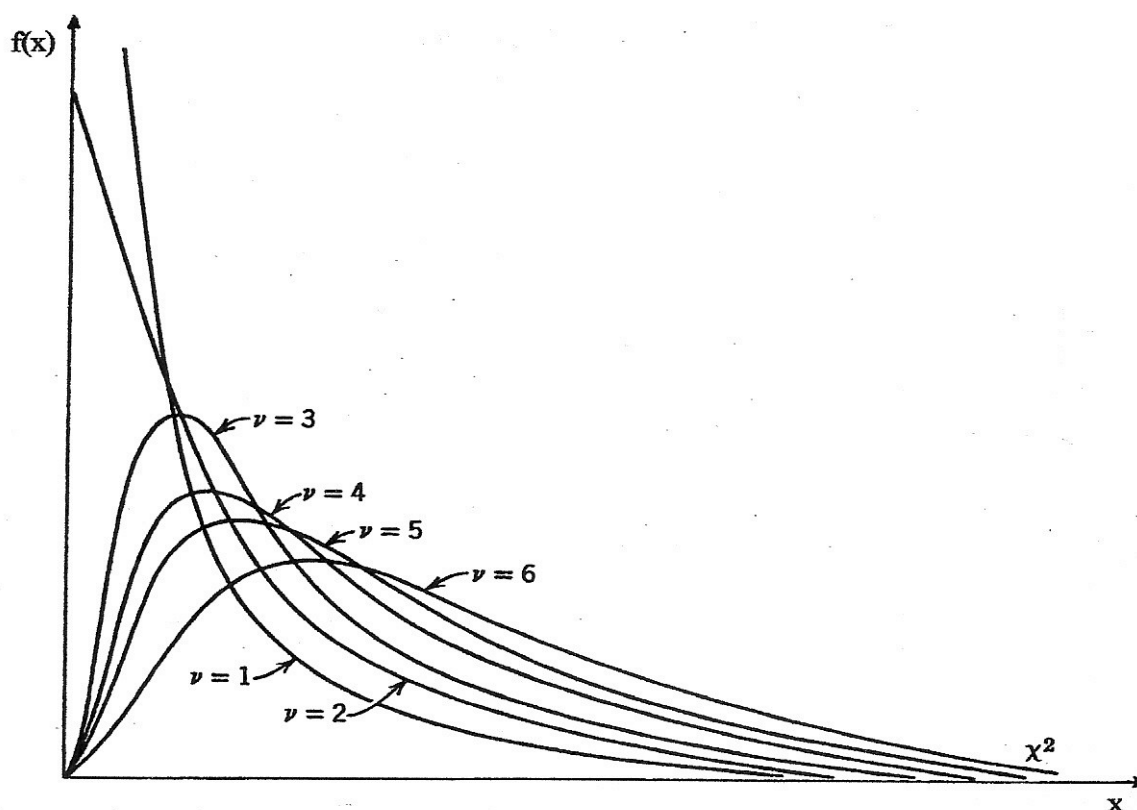


Fig.1 . Distribuzione chi-quadrato per vari gradi di libertà.

## Inferenza statistica

Fin'ora 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 gli esperimenti di tipo ripetitivo. I vantaggi di questi modelli è che ci permettono di studiare le proprietà dell'esperimento, cose che sarebbero difficili o impossibili da fare senza l'aiuto di un tale 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 servendosi di dati sperimentali. Molto

spesso lo statistico è interessato a costruire un modello matematico per una sola variabile casuale 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.

*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 meno comuni di 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 possiedano le macchie meno comuni. Se egli fa l'ipotesi che in questo caso sia valida la legge dell'eredità e rappresenta con la variabile  $X$  il numero degli insetti dei 200 insetti che possiedono le macchie meno comuni, allora il modello che egli naturalmente sceglierebbe è la funzione di densità binomiale*

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

*Se non ci fosse alcuna teoria per suggerire che  $1/4$  di tali insetti dovrebbero possedere le macchie meno comuni, il biologo potrebbe scegliere questa stessa funzione di densità con la probabilità  $1/4$  sostituita dalla frequenza osservata pari a  $44/200=0,22$ . Utilizzando la (1) il biologo potrebbe poi prevedere i risultati per 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 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 dai campioni per usarle nello studio delle popolazioni da cui i campioni sono estratti. Il tipo di informazione che si dovrebbe estrarre da un insieme di dati, dipende dalla natura dei dati e dal modello scelto. In alcuni problemi si sa, da informazioni 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 è necessario estrapolare in realtà dai dati sperimentali per tali modelli è l'informazione atta a fornire buone stime dei parametri coinvolti. In altri problemi però, né teoria né esperienza è disponibile per aiutare a scegliere un modello. Allora è necessario usare i dati sperimentali per decidere su di un ragionevole tipo di modello prima di poter procedere ulteriormente.

Quando si considera la natura dei dati è importante distinguere tra quegli insiemi di dati per cui l'ordine in cui le osservazioni sono state ottenute fornisce un'informazione utile e quegli insiemi per cui non è così. Ad esempio, se si volesse studiare i fenomeni atmosferici o la borsa valori di giorno in giorno, l'ordine sarebbe importante. L'esperienza industriale indica che l'informazione ottenuta da considerare l'ordine in cui sono fabbricati gli articoli è indispensabile per una efficiente produzione. Se si fosse invece interessati a studiare certe caratteristiche degli studenti di un istituto e si scegliesse un insieme di studenti prendendo un nominativo ogni 20 in una guida dell'istituto, difficilmente ci si aspetterebbe che l'ordine in cui si prendono i nomi fosse utile per quanto riguarda lo studio. Ci occuperemo ora di tecniche che non usano informazione di ordine.

## Classificazione dei dati

Supponiamo ora che siano dati i pesi di 200 individui di un certo istituto e che 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 quelle misure sono distribuite. Per ottenere un'idea migliore della distribuzione dei pesi, è perciò conveniente raggruppare i dati classificando le misure in gruppi. Sarà allora possibile tracciare il grafico della distribuzione così modificata ed imparare di più su come sono distribuiti i pesi. Questa classificazione sarà utile anche per semplificare i calcoli di certe medie. Queste medie forniranno utili informazioni sulla distribuzione. Così lo scopo di classificare i dati è quello di aiutare ad estrarre certi tipi di utili informazioni riguardanti la distribuzione che sta alla base. Nel classificare i dati è di solito conveniente

usare da 10 a 20 classi, ma se necessario si può scendere fino a 6 classi.

La tab.1 mostra come è stato classificato un insieme di 200 misure dei diametri di barrette d'acciaio i cui valori variano tra 0,431 e 0,503 pollici. I diametri sono stati misurati al millesimo di pollice, gli estremi degli intervalli di classe sono stati presi mezza unità oltre l'accuratezza di misura per assicurare che nessuna misura cada su di un estremo dell'intervallo.

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

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 punti di mezzo dell'intervallo. Questo valore del punto di mezzo viene chiamato *marchio* o *centro di classe*. Dopo che ciascuna misura è stata registrata nella classe appropriata per mezzo di una sbarretta come mostrato in tab.1, i risultati della classificazione sono registrati sotto forma di una tabella di frequenza come mostrato nella seconda metà della tab.1.

### Rappresentazione grafica delle distribuzioni empiriche (osservate)

Un'idea approssimativa di come sono distribuiti i valori di una variabile casuale si può ottenere esaminando l'istogramma. L'istogramma per i dati di tab.1 è mostrato in fig.1.

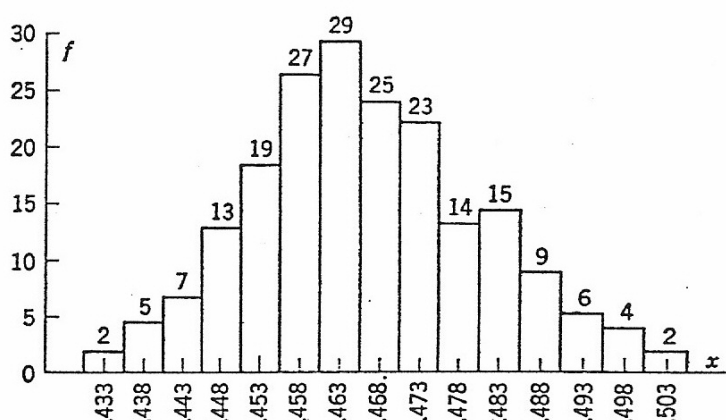


Fig.1. Distribuzione dei diametri di 200 barrette di acciaio.

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 variano di solito da quella a campana (come in fig.1) a

quella che rassomiglia alla metà di una campana. Una distribuzione dell'ultima tipo si dice che è *asimmetrica*, il che significa che manca di simmetria rispetto ad un asse verticale. Risulta che variabili con distribuzioni che possiedono tali forme con gradi crescenti di asimmetria sono, ad esempio, la statura, il peso, l'età di matrimonio, l'età di mortalità per certe malattie, e la ricchezza. Le fig.1,2 e 3 rappresentano tre tipiche distribuzioni con gradi crescenti di asimmetria.

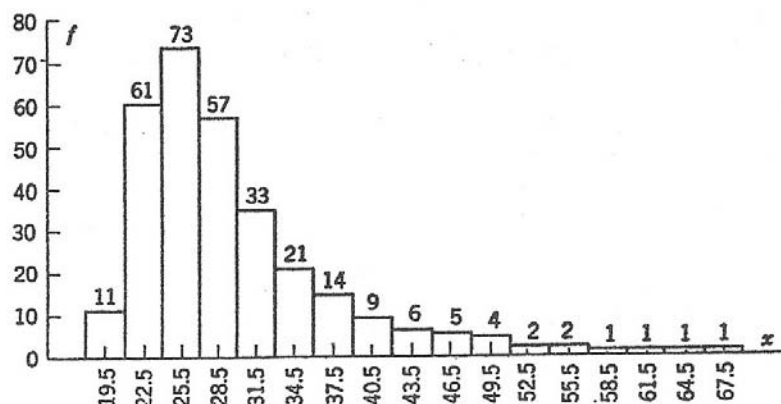


Fig.2. Distribuzione di 302000 matrimoni classificati secondo l'età dello sposo. L'unità di frequenza è 1000.

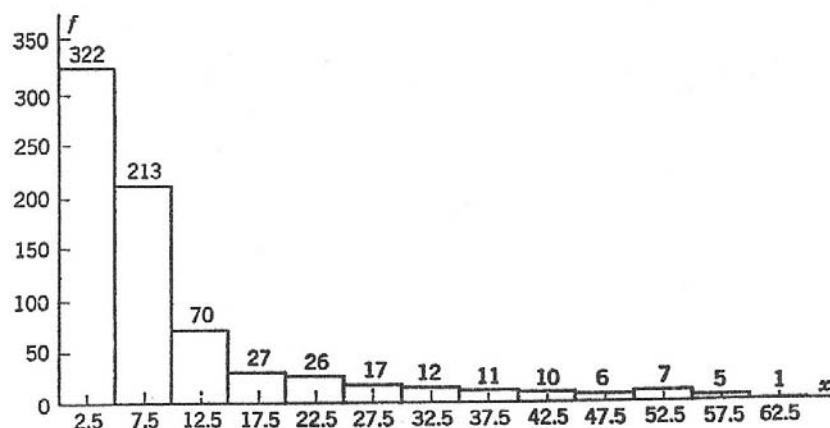


Fig.3. Distribuzione di 727 morti per scarlattina classificati secondo l'età.

Si può osservare che la distribuzione di fig.1 ha la forma di una campana e che perciò la distribuzione normale potrebbe servire da modello soddisfacente per la distribuzione dei diametri delle barrette. Molte misure industriali lineari hanno distribuzioni di questo tipo e vengono trattate con successo usando come modelli le curve normali. La distribuzione mostrata in fig.3 ha una forma che suggerisce la possibilità di usare come modello una distribuzione esponenziale, mentre la distribuzione di fig.2 richiede un modello più sofisticato. Poiché la distribuzione gamma, con i suoi due parametri, ha moltissima flessibilità, essa potrebbe servire da modello per la distribuzione di fig.2.

## Momenti

Sebbene l'istogramma come quelli mostrati nelle figure 1,2 e 3 viste ieri fornisca una notevole quantità di informazioni generali che riguardano la distribuzione di un insieme di misure campionarie, informazioni più precise e sicure 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 certo istituto per confrontarlo con l'istogramma di un campione simile di un altro istituto, potrebbe essere difficile, eccetto in termini molto generali, stabilire 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 le variazioni dei pesi dei due gruppi. La natura di un problema statistico determina largamente se alcune semplici proprietà della distribuzione saranno sufficienti a descriverla in modo soddisfacente. Spesso la maggior parte dei problemi che incontreremo richiederanno, per la loro soluzione, soltanto alcune proprietà fondamentali della distribuzione. Per semplici distribuzioni di frequenza, come quelle i cui grafici sono mostrati nelle figure 1,2 e 3 viste il precedenza, questa descrizione si compie in modo soddisfacente per mezzo dei momenti di ordine basso della distribuzione. In molti problemi lo statistico è interessato soltanto ai momenti del primo e del secondo ordine; in alcuni problemi egli usa i primi 4 momenti, ma raramente ne usa più di 4. Una ragione di ciò è che i momenti di ordine più elevato, per come sono definiti, sono così instabili in ripetuti esperimenti di campionamento che da loro si possono ottenere poche ulteriori informazioni attendibili.

Indichiamo ora con  $x_1, x_2, \dots, x_n$  i valori osservati di un campione di dimensione  $n$  della variabile casuale  $X$ . Allora, per analogia con i momenti teorici, i momenti empirici dall'origine vengono definiti come segue:

---

il momento d'ordine  $k$  dall'origine di una distribuzione empirica (o osservata) è dato da

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


---

I momenti empirici sono chiamati anche **momenti del campione** perché essi sono basati su valori del campione o valori campionari. Il momento del primo ordine  $m'_1$  si indica per tradizione col simbolo  $\bar{x}$ . Esso dà il centro di gravità di una distribuzione empirica proprio come fa la media  $\mu$  per una distribuzione teorica e serve per misurare dove è centrata la distribuzione empirica. Esso è chiamato **media del campione** o **media campionaria**. Per analogia con la definizione data per le distribuzioni di probabilità, i momenti empirici dalla media sono definiti come segue:

---

il momento d'ordine  $k$  dalla media di una distribuzione empirica o osservata è dato da

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


---

Poiché il momento del secondo ordine dalla media  $m_2$  è usato molto spesso, si assegna adesso il simbolo particolare  $s^2$  e viene chiamato **varianza del campione**. Corrispondentemente  $s$  viene chiamato **scarto quadratico medio** o **deviazione standard del campione**. Per calcolare  $s^2$  è spesso conveniente usare la formula seguente che è l'analoga di una formula già vista, e cioè

$$s^2 = m'_2 - \bar{x}^2$$

dove la formula analoga era

$$\sigma^2 = \mu'^2 - \mu^2$$

Se i dati 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, con  $f_i$  che rappresenta il numero di osservazioni nell' $i$ -esimo intervallo, cioè la frequenza assoluta di  $x_i$ , ed  $N_c$  che rappresenta il numero degli intervalli di classe, allora le definizioni dei momenti assumono le forme seguenti

$$(3) m'_k = (1/n) \sum_{i=1}^{N_c} x_i^k f_i$$

e

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

Il valore di  $\bar{x}$  nella (4) si assume essere il valore ottenuto dalla (3) e non dalla (1). A rigor di termini, la

(3) e (4) definiscono i momenti soltanto per le distribuzioni empiriche classificate e sono soltanto approssimazioni dei valori dati dalle (1) e (2). Tuttavia, le approssimazioni sono di solito così buone che in pratica non si fa nessuna differenza fra questi due insiemi di valori. Per esempio,  $\bar{x}$  ed  $s^2$  sono chiamate rispettivamente **la media campionaria** e **la varianza campionaria** sia che esse siano state ottenute dalle formule (1) e (2) che dalle formule (3) e (4).

**[DOMANDA ESAME: Per quale motivo sento l'esigenza di classificare una distribuzione? Perché voglio vedere il grafico e come si distribuisce, ma non per saperne il valore.]**

Si fa osservare che in molti testi si usa di solito, invece di  $f_i$ ,  $\phi(x_i)$  che rappresenta il valore della funzione di frequenza assoluta in  $x_i$ . Questa funzione viene indicata con  $\phi(x)$  e non con  $f(x)$  perché quest'ultima si usa in genere per rappresentare la funzione di densità di probabilità. Non c'è nessun motivo 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 desiderano ad esempio i valori di  $\bar{x}$  ed  $s^2$  può essere più facile allora calcolarli tramite le formule (3) e (4) piuttosto che con le formule (1) e (2). Come si è già detto, le differenze sono di solito trascurabili.

**Esercitazione.** Per acquistare più familiarità con la deviazione standard come misura della concentrazione di una distribuzione attorno alla sua media, consideriamo la seguente distribuzione di 1000 conversazioni telefoniche misurate in secondi mostrata in tab.1. L'istogramma per questa distribuzione è mostrato in fig.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

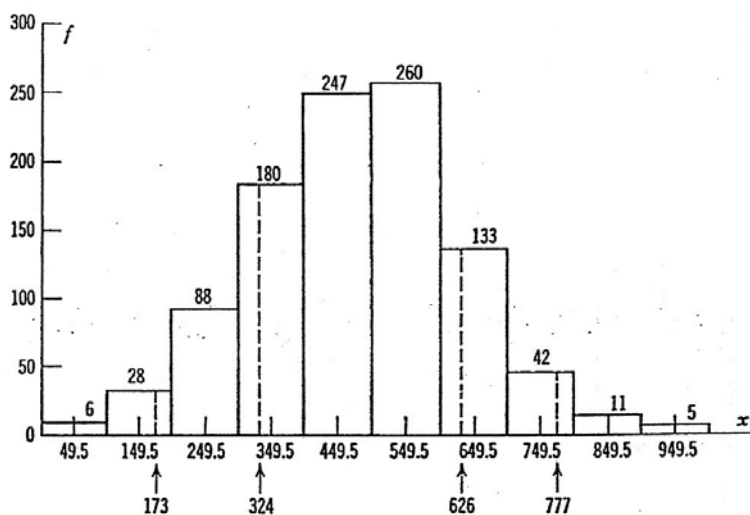


Fig. 1 . Istogramma per la distribuzione di 1000 conversazioni telefoniche.

Dalla fig.1 appare che, per la durata delle conversazioni telefoniche, potrebbe essere usato 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% ed 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 in merito alla distribuzione empirica. I calcoli per i dati di tab.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 mostrati nella fig.1.



*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% delle misure effettuate 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.*

## Inferenza statistica

Ci occuperemo ora brevemente di quel tipo inferenza statistica che implica la stima di parametri della funzione di densità 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, per esempio una compagnia telefonica interessata a studiare problemi connessi con la durata delle conversazioni telefoniche potrebbe volere stimare i parametri  $\mu$  e  $\sigma$  della funzione di densità normale che si può assumere come modello per le conversazioni telefoniche. Due tipi di stima dei parametri sono d'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 sui valori osservati della variabile casuale che ci si aspetta che contengano nel loro interno il valore vero del parametro.

Passiamo ora a considerare le stime puntuali e ci occupiamo del *metodo dei momenti*.

## Metodo dei momenti

Il metodo di stima che illustreremo ora è noto come il **metodo dei momenti**. Se un parametro di una funzione di densità, come ad esempio il parametro  $\mu$  per la funzione di densità di Poisson o il parametro  $\beta$  per la funzione di densità esponenziale, è un momento della distribuzione allora la sua stima sarà il corrispondente momento del campione. Poiché  $\mu$  e  $\beta$  sono le medie delle loro rispettive densità, essi sarebbero entrambi stimati dalla media del campione  $\bar{x}$ . I due parametri  $\mu$  e  $\sigma^2$  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 esso non è un momento della distribuzione, il parametro può ancora essere stimato col metodo dei momenti calcolando il momento del primo ordine della distribuzione che sarà una funzione del parametro e uguagliandolo ad  $\bar{x}$ . La soluzione dell'equazione risultante rispetto al parametro incognito fornirà la stima richiesta. Se la distribuzione avesse ad esempio due parametri che non fossero momenti sarebbe lo stesso procedimento rispetto ai primi due momenti della distribuzione. Come esempio in cui i parametri non sono momenti, supponiamo che entrambi i parametri della funzione di densità gamma debbano essere stimati per mezzo dei momenti del campione del primo e del secondo ordine. Dalle formule già viste è noto che  $\mu=\beta\alpha$  e  $\sigma^2=\beta^2\alpha$ . Allora le stime di  $\alpha$  e  $\beta$  si ottengono risolvendo le seguenti equazioni

$$\beta\alpha = \bar{x}$$

$$\beta^2\alpha = s^2$$

Le soluzioni di queste equazioni rispetto ad  $\alpha$  e a  $\beta$  ci forniscono le loro stime secondo il metodo dei momenti. Si ha che

$$\alpha = \bar{x}/\beta$$

$$\beta^2(\bar{x}/\beta) = s^2$$

allora

$$\alpha = \bar{x}^2/s^2$$

$$\beta = s^2/\bar{x}^2$$

## Statistica descrittiva

All'inizio del corso sono state messe in evidenza alcune caratteristiche essenziali della statistica descrittiva e della statistica inferenziale. Iniziamo ora a parlare della statistica descrittiva.

A tale scopo possiamo partire dicendo che la statistica descrittiva 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 fra 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, ecc. Dall'analisi di un fenomeno collettivo realizzata mediante un'indagine statistica si ottiene una serie di informazioni che permette di comprendere e interpretare il fenomeno e, quando necessario, di fare delle previsioni che riguardano la sua evoluzione futura e quindi anche di programmare in tempo utile gli interventi necessari. Quando si vuole studiare un fenomeno collettivo, si esegue 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 data regione, ed altre ancora. 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 fra loro per una o più caratteristiche, per il sesso (cioè maschi e femmine), per il tipo di mansioni che svolgono (operai, impiegati, ecc), per il peso, per la statura, per l'età, ecc. 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:

- *variabili quantitative o numeriche* che possono essere espresse con numeri, ad esempio l'età, il peso, il numero dei figli, ecc;
- *variabili qualitative* che non possono essere espresse con numeri, come ad esempio la provincia di nascita, il gruppo sanguigno, il mezzo di trasporto utilizzato, ecc.

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 svolgendo un'indagine statistica si chiamano *dati statistici*. Un'indagine statistica su un fenomeno collettivo si articola nelle seguenti fasi:

- *rilevazione dei dati*;
- *elaborazione dei dati*, cioè tabulazione, classificazione, rappresentazione grafica, calcolo di misure caratterizzanti come le misure di tendenza centrale quali la media, la mediana, la moda, la media geometrica, la media armonica, le misure di dispersione come il campo di variazione, la varianza, lo scarto quadratico medio, e altri indici di tendenza della distribuzione relativi alla sua forma;
- *presentazione dei dati elaborati*;
- *interpretazione dei dati*.

La *rilevazione dei dati* può essere di due tipi:

- rilevazione completa, se è estesa a tutte le unità statistiche della popolazione in esame;
- 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 una classe di studenti, sugli impiegati di un ufficio, sui commercianti di un quartiere che possono essere tutti contattati ed 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 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 300000 persone, non è certamente pensabile di contattarli tutti, si sceglie perciò un campione abbastanza vasto, ad esempio di 1000 unità, il più possibile rappresentativo degli abitanti della città (uomini e donne nella giusta proporzione, studenti, pensionati, operai, professionisti, ecc), e si svolge l'indagine su di esso. I

risultati ottenuti si estendono poi all'intera popolazione statistica utilizzando le proporzioni.

È 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 esistono 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:

- campionamento probabilistico, caratterizzato dal fatto che di ogni elemento della popolazione è nota la probabilità che venga scelto;
- campionamento non probabilistico, allorché la proprietà precedente non sia verificata.

I metodi usati per la rilevazione dei dati nelle indagini statistiche sono diversi:

- a) l'intervista, che consiste nel porre delle precise domande direttamente a ciascuna unità statistica e a 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 per telefono;
- d) la consultazione di archivi, come ad esempio gli archivi comunali;
- e) la consultazione di pubblicazioni specializzate, come ad esempio quella dell'ISTAT, cioè dell'Istituto Nazionale di Statistica.

### **Elaborazione, presentazione ed interpretazione dei dati**

Dopo aver parlato della rivelazione dei dati, facciamo ora alcune assunzioni riguardanti l'elaborazione, la presentazione e l'interpretazione dei dati. Se per qualunque esperimento consideriamo i dati grezzi così come ottenuti dalla rilevazione siamo in genere in difficoltà nel poterli interpretare; ad esempio, un elenco delle votazioni conseguite dagli studenti di un paese, non ci dice un gran ché fino a quando questi dati non vengono elaborati e sintetizzati. Vediamo ora come rendere più efficace la lettura e l'interpretazione dei dati.

Supponiamo di avere registrato i dati relativi al peso di 50 studentesse che supponiamo formino un campione di tutte le studentesse di scuola media superiore di una città. Il primo modo artigianale di riportare questi dati è quello di formare una matrice come mostrato in tab.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 tab.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 considerare un primo dato statistico, possiamo dire infatti che tutti i pesi appartengono all'intervallo [53,73] la cui lunghezza prende il nome di **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 ad un campione di numerosità molto maggiore. 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. Per 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.

**1° criterio.** *Il numero delle classi deve rendere chiara la natura di tutta la distribuzione.* Se le classi sono infatti troppe o troppo poche rischieremmo di perdere utili informazioni. Nel primo caso perché in ogni classe vi sarebbero pochissimi elementi o addirittura nessuno, nel secondo caso perché potrebbe accadere che, essendo concentrati molti elementi in poche classi, perderemmo di vista la globalità della distribuzione. In genere, come già accennato in precedenza, si scelgono da 6 a 20 classi. Secondo Sturges, si ha un numero ottimale di classi, indicato con  $N_c$ , scegliendo  $N_c=[1+1,443*\ln N]$ , dove  $N$  rappresenta il numero dei dati osservati e  $[\alpha]$  rappresenta l'intero più vicino ad  $\alpha$ .

**2° criterio.** *Le classi devono avere la stessa ampiezza e, dopo quanto detto al primo punto, tale ampiezza sarà data da  $h=r/N_c$  dove  $r$  è il rango dell'insieme dei dati osservati.* Come esempio, determiniamo il numero di classi e l'ampiezza delle classi nel caso del campione in esame rappresentato dai dati di tab.2. Dalla formula di Sturges si ha che

$$N_c=[1+1,443*\ln 50]=[6,64]=7$$

per cui l'ampiezza è data da

$$h=20/7=2,86$$

**3° criterio.** *L'ampiezza dell'intervallo deve essere un numero tale da consentire di individuare il punto di mezzo, inoltre come 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 le classi nel caso dei dati del campione in esame. L'ampiezza più opportuna per le 7 classi è uguale a 3. Avremo allora i dati presentati in tab.3.

Tab. 3

<i>Classi di peso</i>	<i>Punti di mezzo</i>	<i>Etichette</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	□	3

Nella tab.3, nella colonna delle etichette, è stato usato un metodo assai comune di contare gli elementi osservati tramite lati di quadrati con chiusura da 1 a 5. A questo punto, anche senza conoscere esattamente tutti i dati ma conoscendo soltanto la frequenza assoluta, cioè il numero di elementi di ogni classe, possiamo avere un'idea della distribuzione dei dati. Possiamo infatti dire che vi sono due studentesse che pesano 54 kg, 5 che pesano 57 kg, 9 che pesano 60 kg e così via. Per avere altri tipi di informazione sempre più precisi ed esaurienti possiamo definire altri indici statistici. Più esattamente definiremo:

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

Nella tab.4 sono riportati i valori delle 4 funzioni suddette corrispondenti ai punti di mezzo delle 7 classi per i dati riguardanti il campione 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

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

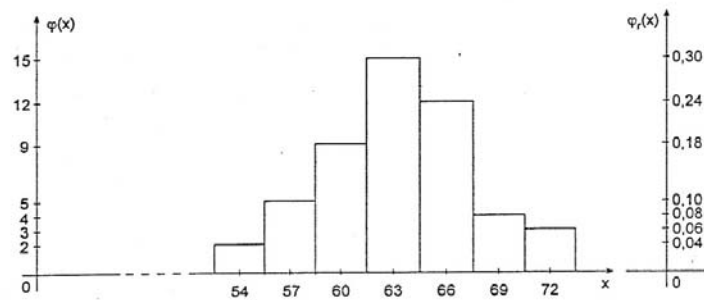


Fig. 1. Istogramma.

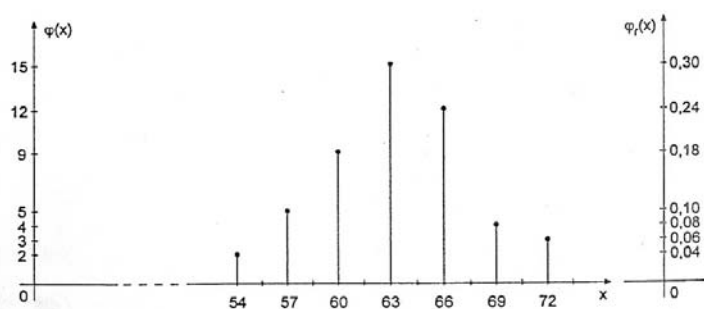


Fig. 2. Grafico a segmenti.

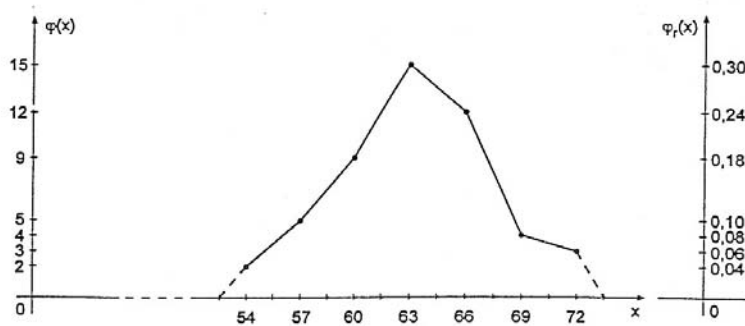


Fig. 3. Poligono di frequenza.

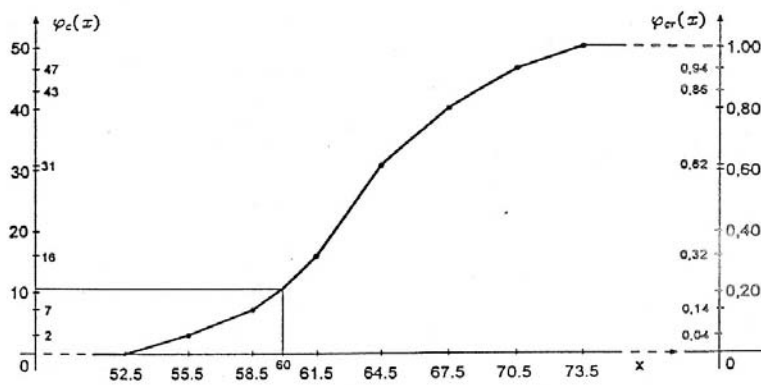


Fig. 4. Poligono di frequenza relativa cumulativa.

In ciascuno dei 4 grafici si possono leggere informazioni sui dati. Ad esempio, dal grafico di fig.4, il numero di studentesse il cui peso cade fra 61,5 e 71,5 kg è semplicemente la differenza fra le frequenze cumulative corrispondenti e perciò abbiamo  $47-16=31$ . Per i valori intermedi si possono soltanto dare delle stime, ad esempio il numero di studentesse con peso inferiore a 60 kg è circa 11, e questo dato è stato ottenuto graficamente come mostrato in fig.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*. Sono indici di tendenza centrale anche le misure di dispersione, tra cui quelle di gran lunga più usate sono certamente la *deviazione standard* o *scarto quadratico medio* e la *varianza*. Anche il *rango* o *campo di variazione* può talvolta essere un indice utile per comprendere come stanno grossomodo 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 questo indice risente molto di valori molto grandi e di valori molto piccoli. Inoltre tende inevitabilmente a crescere se aumentiamo le rilevazioni dei dati.

## Media

Calcoliamo ora la media dei pesi delle studentesse del campione d'esame. Applicando la formula che definisce la media di una distribuzione empirica, si ottiene

$$\bar{x}=63,22$$

Avendo classificato la distribuzione dei pesi possiamo anche calcolare la media tramite la formula che la definisce per le distribuzioni classificate ed ottenere così che

$$\bar{x}=63,24$$

Si può notare che quest'ultimo valore 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 uno per uno, a meno che non vi sia perfetta simmetria intorno al punto di mezzo della classe.

*Facciamo ora un esempio per capire quanto 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 e D che percepiscono rispettivamente 10, 15, 16 e 80 mila euro. Usando la definizione si ha che la media è*

$$\bar{x}=1/4(10+15+16+80)=20,25 \text{ mila euro}$$

*Questo valore non è certamente una buona sintesi degli stipendi, 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 poco come ovviare alla presenza di valori estremi nella determinazione di un ragionevole valore medio. A tale scopo definiremo ora altri 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 ordinati in senso crescente o decrescente 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

$$x_{\text{med}}=x_{(n+1)/2} \text{ se } n \text{ è dispari}$$

$$x_{\text{med}}=(1/2)[x_{n/2}+x_{(n/2)+1}] \text{ se } n \text{ è pari}$$

avendo ordinato  $x_1, x_2, \dots, x_n$  in ordine crescente o decrescente.

**Esempio.** *Gli stipendi mensili di 6 dipendenti di un'industria sono 1,4, 1,2, 1,3, 2,0, 0,8 e 9,7. Calcoliamo ora la media e la mediana. In questo caso la media è*

$$\bar{x}=2,7333333333333333\dots$$

*Per la mediana, poiché  $n=6$  e, avendo ordinato i dati in senso crescente, come 0,8, 1,2, 1,3, 1,4, 2,0 e 9,7, si ottiene che*

$$x_{\text{med}}=(1/2)(x_3+x_4)=(1/2)(1,3+1,4)=1,35$$

*In questo caso la mediana ci appare come un valore molto più ragionevole della media per indicare una sintesi, ovvero un "valore medio", della distribuzione degli stipendi.*

*Come esempio, ora, calcoliamo la mediana relativa ai pesi del campione in esame. La mediana, essendo pari il numero dei dati, osservando la tab.2 vista in precedenza, è data da*

$$x_{med} = (1/2)(x_{25} + x_{26}) = (1/2)(64 + 63) = 63,5$$

## Moda

Molto spesso i dati sono divisi in classi che non sono di tipo numerico, ad esempio il tipo di lavoro svolto, il gruppo sanguigno, la provincia di nascita, ecc. In questi casi non ha senso parlare di media o di mediana, a meno di non avere etichettato i dati. È 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.

*Come esempio determiniamo la moda dell'insieme dei pesi in esame. Dalla tab.2 vista in precedenza si può osservare che la moda*

$$x_{mod} = 64$$

*valore che si ripete più spesso degli altri, cioè 6 volte.*

Ricordiamo anche che molte distribuzioni statistiche non hanno una sola moda. Le distribuzioni statistiche con una sola moda prendono il nome di **distribuzioni unimodali**.

## Media, mediana e moda

Sul significato della media, della mediana e della moda e 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 da quelli più piccoli;
- la moda è il dato che si ripete con maggiore frequenza e quindi non è influenzata da tutti gli altri dati.

È 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 per uno solo. Nel caso in cui la media, la mediana e la moda coincidono oppure sono 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 ricorre spesso nelle indagini statistiche. Quando c'è una distribuzione normale dei 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 di *Gauss*.

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 200 studenti che*

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

*Per quanto riguarda il punto a) 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 per i bagagli di 20 kg. Questo dato altro non è che una media.*

*Riguardo al punto b), 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, che è il dato più importante per pianificare gli acquisti.*

*Per il punto c) 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 adesso alcune relazioni fra 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 indipendentemente dal fatto che il grafico sia più o meno asimmetrico, intendendo sempre qui di parlare di distribuzioni unimodali. Se la distribuzione è simmetrica attorno a un certo valore, allora come già detto le tre misure coincidono come indicato dal grafico di fig.5.

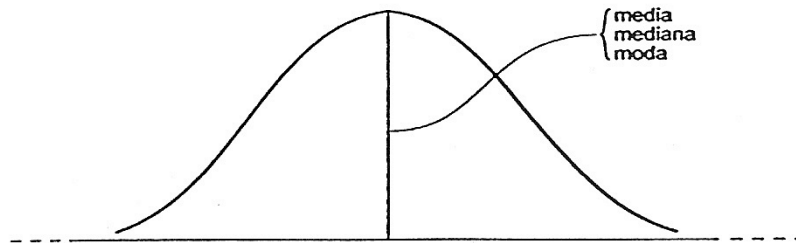


Fig. 5. Grafico (simmetrico) della distribuzione dei dati.

Quando non c'è simmetria i tre valori di tendenza centrale si distinguono tra loro anche se nelle distribuzioni unimodali la mediana si colloca sempre fra la media e la moda. Più precisamente, come mostrano i due grafici di fig.6 e 7, se la coda della distribuzione è a destra allora la moda precede la mediana e la media, mentre se la coda della distribuzione è a sinistra allora è la media a precedere la mediana e la moda.

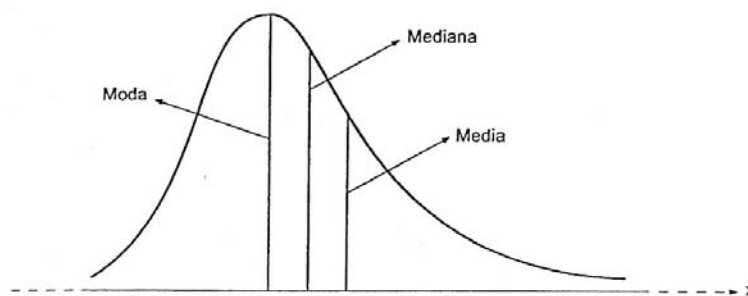


Fig. 6

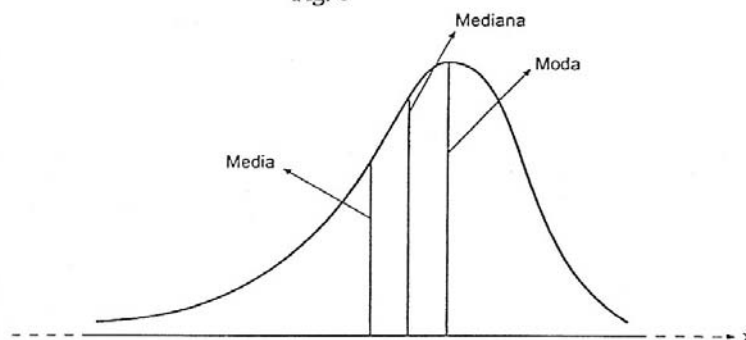


Fig. 7

Per comprendere la ragione 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 più alto della distribuzione mentre la media, che come abbiamo visto risente molto dei valori estremi, dovrà essere alla destra della mediana. In modo analogo, sostituendo la destra con la sinistra, si può ragionare nel caso di distribuzione con coda a sinistra. Per distribuzioni unimodali che siano moderatamente asimmetriche vale la relazione empirica che la media

$$\bar{x} - x_{\text{mod}} = 3(\bar{x} - x_{\text{med}}).$$