

Calcolo delle Probabilità e Statistica

Prof. Fucci

Anno Accademico 2006/2007

**rev. 009**

## Statistica

Il termine statistica deriva del latino medioevale "status" (ovvero stato) che vuole indicare l'ordinamento politico. E' stata chiamata statistica l'area scientifica che si occupava di fornire descrizioni relative ad uno stato e descrizioni comparative di vari stati.

Tali descrizioni ebbero inizio in effetti già con Aristotele nel 350 A.C. e proseguirono nelle epoche storiche successive. Pur tuttavia la parola statistica apparve per la prima volta nella Enciclopedia Britannica nel 1797 essendo stata conosciuta ufficialmente dal professore tedesco Gottfried Achenwall (1719-1772) verso la metà del 1700 che la definì come la descrizione delle cose più notevoli di uno stato. Il significato che essa ha mantenuto fino agli inizi del 1800 è stato dunque quello di descrizione o censimento con la 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 e una statistica matematica o inferenziale.

Per statistica descrittiva si intende il complesso delle tecniche di cui si serve lo sperimentatore per raccogliere, elaborare e rappresentare gruppi di dati osservati che sono in generale in numero elevato e in forma disordinata. Lo scopo finale è quello di ricavare dalle informazioni opportuni elementi di sintesi. Tali elementi possono essere dei numeri come la media, la varianza, la mediana, la moda ecc. o informazioni rappresentate in forma di tavole, grafici, carte (geologiche o numeriche), diagrammi ecc.. In questa situazione si cerca però soltanto di descrivere il fenomeno osservato senza cercare di interpretare o generalizzare più di tanto riguardo alla totalità dei potenziali dati osservabili, di cui quelli in esame sono soltanto una piccola parte.

Al giorno d'oggi naturalmente queste informazioni sono molto importanti però quasi sempre vogliamo saperne di più e spendere meno. Ciò significa non dover censire tutto per avere le informazioni che ci servono ma estrarre le stesse dalla conoscenza dei dati osservati che sono a nostra disposizione.

In termini statistici si tratta di trarre delle conclusioni sull'intera popolazione partendo da un campione osservato. Ciò implica che molti dati saranno a noi ignoti ma che dovremo trovare un metodo per averne una qualche conoscenza che non sia quella diretta sperimentale. Le informazioni suddette rappresentano quindi un passo preliminare però molto importante per la successiva ricerca di un modello interpretativo che è un aspetto proprio della statistica inferenziale. Si cerca dunque di studiare gruppi di dati allo scopo di decidere quale tipo di modello è quello che descrive al meglio l'intera popolazione. I modelli possono essere diversi fra loro in modo sostanziale oppure soltanto per alcuni parametri, ovvero caratteristiche, della popolazione. È opportuno sottolineare fin da ora che il concetto di probabilità diventa essenziale per la trattazione della statistica inferenziale.

## Calcolo delle Probabilità

Per quanto riguarda la probabilità diamo innanzitutto cenni storici. I primi documenti che si conservano sono dovuti a Girolamo Cardano (1501-1576) accanito giocatore d'azzardo con il suo libro "De Ludo Aleae" (sul gioco di dadi) scritto forse intorno al 1526 e pubblicato postumo nel 1663 e a Galileo con uno scritto del 1620 circa. Nel libro di Cardano è inserito il problema delle probabilità dei punteggi ottenuti come somma lanciando tre dadi. Il

risultato contiene qualche inesattezza e i giudizi sul valore dell'opera sono controversi però è fuori dubbio la sua importanza storica. Nello scritto di Galileo viene trattato lo stesso problema però con maggiore chiarezza ma la nascita del calcolo delle probabilità viene abitualmente attribuita a Blaise Pascal (1623-1662) e fissata nella corrispondenza fra lui e l'altro grande matematico francese Pierre Fermat (1601-1665) verso la metà del 1600. L'interesse di Pascal fu molto stimolato da Cavalier de Merè, spirito vivace, matematico discreto e accanito giocatore d'azzardo il quale si rivolse spesso a lui per risolvere vari problemi riguardanti il gioco.

Il calcolo della probabilità sorto quindi verso la metà del 1600 come studio matematico dei giochi quali monete, dadi, carte, può essere considerato oggi come una disciplina matematica che trova applicazione negli esperimenti non deterministici o casuali. Un esperimento non deterministico o casuale è un tipo di esperimento che può dar luogo ad un risultato fra quelli possibili non determinabile a priori in modo univoco. Un esempio è rappresentato dal lancio di un dado il cui risultato è affetto da un certo grado di incertezza potendo essere rappresentato da una qualunque delle sei facce del dado stesso. Questi cenni hanno lo scopo di mostrare come si presenta originariamente il concetto di probabilità e di introdurre alle definizioni o interpretazioni della probabilità che ora presenteremo.

## **Definizione Classica**

La prima definizione di probabilità, detta perciò classica, si ritrova in Pascal: definisce la probabilità di un evento come il rapporto fra il numero dei casi favorevoli all'evento e il numero dei casi possibili, purché questi ultimi siano tutti ugualmente possibili.

Critica: molti però vedono in questa definizione una tautologia (spiegare il fenomeno con il fenomeno stesso) nel senso che bisogna sapere già prima che significato dare alla probabilità perché non si vede bene la differenza fra ugualmente possibili e ugualmente probabili.

Comunque i casi risulteranno ugualmente possibili quando si ha una situazione di perfetta simmetria fisica come avviene ad esempio per le palline nell'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 di un evento in condizioni opportune, quando cioè ci sia un numero finito di alternative possibili che possono essere considerate ad esempio per motivi di simmetria, ugualmente probabili. Tutto questo però sapendo già che cos'è la probabilità. In conclusione la validità di questa definizione resta discutibile e d'altra parte si può vedere facilmente che essa risulta essere operativa.

Il campo nel quale essa si può applicare più direttamente è quello dei giochi. In essi infatti si possono individuare con precisione le diverse alternative e si può ragionevolmente supporre che esse siano ugualmente probabili.

Esempio: supponiamo di voler determinare la probabilità che lanciando una moneta non truccata essa presenti testa. Qui l'unico caso favorevole è la testa e i casi possibili sono due: testa o croce. Secondo la definizione si può stimare la probabilità come data da  $1/2$ .

## Definizione Frequentista

L'insoddisfazione della definizione classica portò a costruire la probabilità sulle frequenze. Intendendo come frequenza relativa dei successi il rapporto fra il numero delle volte in cui l'evento si verifica e il numero delle prove effettuate, si giunse così alla definizione frequentista della probabilità di un evento intesa come limite delle frequenze relative dei successi quando tende all'infinito il numero delle prove fatte nelle stesse condizioni.

In altre parole, accettando l'ipotesi che la frequenza relativa si vada stabilizzando intorno ad un certo numero, è proprio questo numero che viene preso come probabilità. Anche la definizione frequentista presenta aspetti criticabili anzitutto perché resta imprecisato il numero delle prove necessarie per arrivare ad un valore abbastanza stabilizzato che fornisca la probabilità ma soprattutto l'esigenza che le prove siano fatte nelle stesse condizioni presenta problemi. A rigore questa condizione non è mai rigorosamente soddisfatta. Essa restringe l'applicabilità della definizione a situazioni ben delimitate come ad esempio i lanci successivi di un dado, ma se si guarda bene anche in questo caso da un lancio all'altro può cambiare, sia pur di poco, la situazione ambientale, cioè l'umidità, la pressione ecc., che influisce sul risultato. Inoltre cambia anche il dado stesso per l'urto che riceve cadendo.

Alla domanda se resterà costante la probabilità dell'evento considerato si risponde: sì, entro limite di approssimazione largamente sufficienti in pratica, no se si richiede che la costanza valga rigorosamente.

Esistono poi situazioni in cui palesemente non vale. Si pensi ad esempio agli incontri fra due squadre di calcio o due tennisti. Si può asserire che anche la definizione frequentista, come quella classica, è operativa.

La frequenza delle prove in cui l'evento si verifica ha le stesse caratteristiche matematiche del rapporto fra il numero dei casi favorevoli e il numero di casi possibili.

Esempio: se una moneta viene lanciata 1000 volte e risulta che testa si presenta 495 volte, la probabilità di testa viene stimata pari a  $495/1000 = 0,495$

## Definizione Soggettiva

In molte situazioni si impone l'esigenza di considerare l'evento singolo e di riferire la probabilità ad esso. Da questa esigenza è nata la concezione soggettiva che si può fare risalire a Daniele Bernolli e che è stata sviluppata recentemente soprattutto dall'italiano Bruno De Finetti (1906-1985) attorno alla metà del 1900. In questa impostazione la probabilità è il grado di fiducia che una persona ha nel verificarsi dell'evento. La probabilità perde così la caratteristica assoluta di numero intrinsecamente legato all'evento per dipendere dalla persona che la valuta e dalle informazioni disponibili.

La definizione così come è non è però operativa ed è necessaria una definizione più precisa. Uno dei modi per renderla tale è di fare riferimento alle scommesse definendo la probabilità come il prezzo equo da pagare per ricevere 1 se l'evento si verifica, 0 se l'evento non si verifica, dove 1 rappresenta una qualsiasi unità di misura (1€, 1\$, 1000000 ecc.). Si pensi ad esempio ad una lotteria fatta mediante i numeri della tombola, ognuno paga ad esempio 1€ e chi vince avendo il suo numero astratto riceve 90€. Allora il prezzo pagato riferito al piatto di 90€ come unità di misura è uguale a  $1/90$  e rappresenta la probabilità di vittoria. Nella definizione si è però parlato di prezzo equo e questo richiede una precisazione che viene data dalla seguente condizione di equità detta anche coerenza

e cioè non si devono valutare le probabilità in modo tale che sia possibile ottenere una vincita certa o una perdita certa. Supponiamo di presentare una moneta dicendo che è truccata in modo tale che la probabilità di ottenere testa sia  $P\{T\}=1/2$  e quella di ottenere croce sia  $P\{C\}=1/4$ ; si può obiettare subito che ciò non può andare perché la somma delle due probabilità deve essere uguale a 1. Questa è un'esigenza intuitiva alla quale la condizione di coerenza dà una giustificazione convincente. Infatti facendo contemporaneamente due scommesse, una su testa e l'altra su croce, si pagherebbe  $\frac{1}{2}$  per la prima più  $\frac{1}{4}$  per la seconda, ricevendo comunque 1 come guadagno netto per  $\frac{1}{4}$ , in contraddizione con la condizione di coerenza. Questo perché la somma delle due probabilità è minore di 1. Se la somma fosse maggiore di 1, si avrebbe una perdita certa. La condizione di coerenza impone che sia  $P\{T\}+P\{C\}=1$ .

Lo stesso ragionamento vale in presenza di più di 2 alternative possibili.

La critica più diffusa a questa impostazione è proprio di essere soggettiva, cioè di fondare la probabilità sul valore dei simboli. Si è sviluppato un lungo confronto spesso polemico con oggettivisti che accusano l'impostazione soggettiva di rendere impossibili la comunicazione fra persone diverse e i soggettivisti che denunciano l'illusorietà della pretesa oggettività delle altre impostazioni.

## **Diverse impostazioni e la definizione assiomatica**

Le tre definizioni date sono profondamente distanti concettualmente. Si è visto inoltre che tutte e tre le impostazioni non sono esenti da critiche. Per ovviare a queste difficoltà i matematici preferiscono trattare la probabilità da un punto di vista assiomatico facendo ricorso alla teoria degli insiemi.

L'impostazione assiomatica è largamente preferita in ogni campo della matematica perché permette di ottenere un livello accettabile di rigore logico.

Il calcolo della probabilità non poteva sfuggire a questa esigenza di sistemazione. Si può verificare che le tre impostazioni arrivano tutte alle stesse leggi matematiche espresse dagli assiomi della probabilità e ciò che rende naturale prendere queste leggi come base per una costruzione assiomatica. La teoria matematica si può quindi sviluppare a partire da certe 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 che si sente di aderire ad una particolare impostazione.

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. Il lancio di una moneta, il getto di un dado, la lettura della temperatura giornaliera su un termometro, sono semplici esempi di esperimenti di questo tipo.

Quando si compie un esperimento il suo risultato è di solito incerto ma se esso viene ripetuto un gran numero di volte è possibile costruire, per esse, un modello probabilistico da usare poi per prendere delle decisioni riguardanti l'esperimento in oggetto. Per gli esperimenti di tipo ripetitivo il modello scelto di solito è un modello per prevedere la frequenza del verificarsi di certi risultati in ripetute esecuzioni dell'esperimento; però, poiché il modello scelto altro non è che la rappresentazione di una situazione reale, le conclusioni tratte da esso saranno affidabili soltanto se esso sarà un'approssimazione sufficientemente buona della realtà in esame. Quindi la prima cosa da fare dal punto di

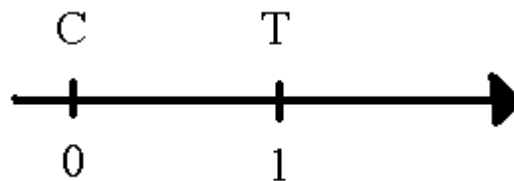
vista statistico per risolvere un dato problema è di scegliere un modello matematico, controllarne l'attendibilità e di trarne quindi da esse le conclusioni risolutive. Il nostro approccio alla probabilità sarà basato sia sull'interpretazione frequentista che su quella assiomatica.

In vista dell'impostazione assiomatica, tenuto conto che abbiamo già discusso a sufficienza il concetto di probabilità, restano da chiarire quelli di esperimento e di evento che, essendo intuitivi, sono già stati usati senza discuterli.

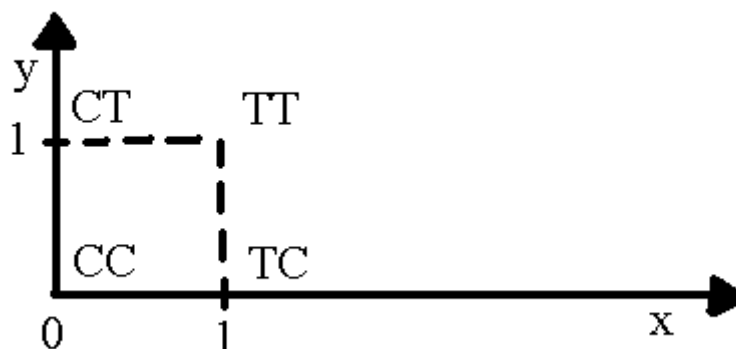
## Spazio Campione o Spazio Campionario

Dato un esperimento 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  $n=1,2,3,\dots$ .

A tale scopo consideriamo il semplice esperimento del lancio di una moneta. In esso i possibili risultati sono due cioè testa e croce. In questo caso è conveniente rappresentare il risultato testa per mezzo del punto di ascissa 1 sull'asse  $x$  e il risultato croce col punto di ascissa 0 come in figura:



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 sarebbe stati quattro cioè TT, TC, CT, CC. In questo caso una conveniente rappresentazione dei risultati per ragioni di simmetria è data dai quattro punti nel piano  $xy$  di coordinate  $(1,1)$ ,  $(1,0)$ ,  $(0,1)$ ,  $(0,0)$  come mostrato nella figura seguente:



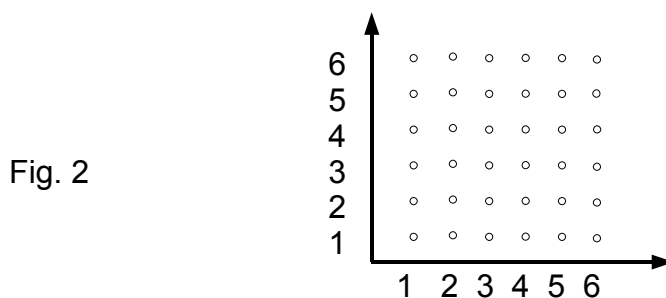
Se la moneta fosse lanciata 3 volte sarebbe conveniente usare lo spazio tridimensionale per rappresentare gli otto possibili risultati sperimentali che sarebbero così localizzati nei vertici di un cubo di lato 1. È importante osservare che queste rappresentazioni sono

semplicemente una convenienza. Ad esempio, nell'esperimento del getto di due monete per rappresentare i quattro possibili risultati si potrebbero pure segnare quattro punti qualsiasi sull'asse  $x$ . Nell'esperimento del getto di due dadi ci sono 36 possibili risultati indicati nella tabella seguente dove il primo numero di ciascuna coppia indica il numero che esce su un dado e il secondo quello che esce sull'altro dado nell'ipotesi che i due dadi siano distinguibili oppure siano gettati nell'ordine.

Tabella 1:

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

In questo esperimento un conveniente insieme di punti per rappresentare i possibili risultati, sono i 36 punti nel piano  $xy$  le cui coordinate sono le corrispondenti coppie di numeri della tabella 1. Questa scelta è mostrata nella figura seguente:



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 tale esperimento è conveniente fare l'ipotesi che la temperatura dell'individuo possa assumere qualsiasi valore fra  $35^{\circ}\text{C}$  e  $42^{\circ}\text{C}$ . In questo caso i risultati possibili si potrebbero rappresentare convenientemente con i punti dell'intervallo di estremi 35 e 42 sull'asse  $x$  come in figura:



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

Come è già stato osservato è importante sottolineare di nuovo che la rappresentazione di tutti i possibili risultati, cioè lo spazio campione, non è unica ma è dettata da criteri di semplicità e di convenienza. Si fa osservare che lo spazio viene chiamato campione o

campionario in quanto l'esperimento cui si riferisce è casuale e ciò significa che il suo esito è incerto così che un dato risultato è solo un campione dei molti esiti possibili. Il motivo per cui viene introdotto lo spazio campione di un esperimento è che esso è un conveniente strumento matematico per sviluppare la teoria della probabilità per quanto riguarda i risultati dell'esperimento.

## Evento

Consideriamo un esperimento tale che qualunque sia il risultato dell'esperimento si possa decidere se un evento indicato con  $A$  si è verificato. Ciò significa che ciascun punto campione si può classificare come un punto per cui  $A$  si verifica o come un punto per cui  $A$  non si verifica. Così se  $A$  è l'evento di ottenere una testa e una croce a prescindere dall'ordine quando si lanciano due monete i due punti campione TC, CT della figura 1 corrispondono al verificarsi dell'evento  $A$ . Se  $A$  è l'evento di ottenere un totale di 7 punti nel lancio di due dadi allora  $A$  è associato a 6 punti campione di coordinate (1,6), (2,5), (3,4), (4,3), (5,2), (6,1) nella figura 2.

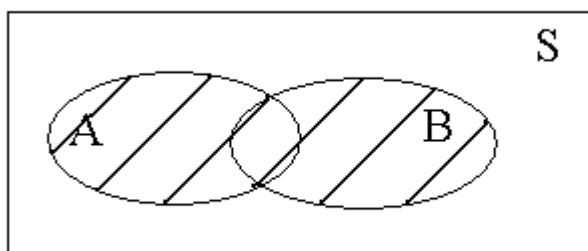
Se  $A$  è l'evento che la temperatura dell'individuo sia almeno  $38^\circ$  allora  $A$  sarà associato all'intervallo di punti fra 38 e 42 sull'asse  $x$ . Possiamo dare allora la seguente definizione: un evento è un sottoinsieme di uno spazio campione. Poiché un sottoinsieme di un insieme di punti comprende la possibilità che il sottoinsieme coincida con l'intero insieme di punti o che esso non contenga nessun punto dell'insieme, questa definizione comprende un evento che è certo di verificarsi o un evento che non può 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 e 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 figura seguente dove con  $S$  si intende lo spazio campione.



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 da qualche altra curva chiusa nel rettangolo. Questa rappresentazione è mostrata in figura 4:

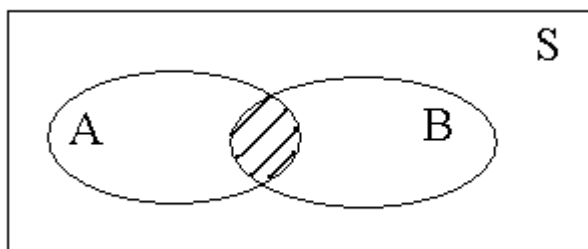




Allora se  $A$  e  $B$  sono due eventi associati ad un esperimento si può voler sapere se almeno uno degli eventi  $A$  e  $B$  si verificherà quando l'esperimento viene eseguito. L'insieme dei punti che appartengono ad  $A$  o a  $B$  o sia ad  $A$  che a  $B$  è chiamato l'unione di  $A$  e  $B$  e si indica col simbolo  $A \cup B$ . Questo insieme di punti è rappresentato dalla regione tratteggiata nella figura 4.

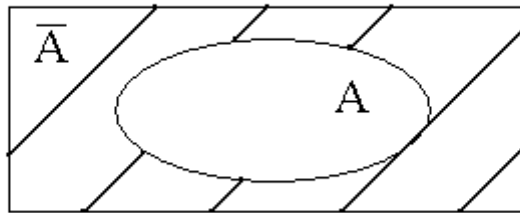
Come esempio se  $A$  è l'evento di ottenere un 6 sul primo dado e  $B$  è l'evento di ottenere un 6 sul secondo dado, allora  $A \cup B$  è l'evento di ottenere almeno un 6 quando si lanciano 2 dadi. L'evento  $A$  consiste dei sei punti dello spazio campione le cui coordinate sono indicate nell'ultima colonna della tabella 1 e l'evento  $B$  consiste dei sei punti indicati nell'ultima riga della tabella stessa. L'evento  $A \cup B$  è allora l'insieme degli 11 punti indicati nell'ultima colonna e nell'ultima riga della tabella 1.

Un altro esempio 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  $A$  che a  $B$  prende il nome di intersezione di  $A$  e  $B$  e si indica con  $A \cap B$ . Questo insieme di punti è rappresentato in fig. 5 dalla regione tratteggiata:

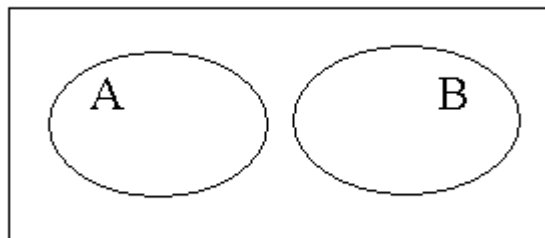


Nell'esempio del getto dei due dadi in cui  $A$  è l'evento di ottenere un 6 sul primo dado e  $B$  quello di ottenere un 6 sul secondo,  $A \cap B$  è l'evento di ottenere un 6 su entrambi i dadi. Esso è rappresentato dall'unico punto di coordinate (6,6) che è l'intersezione dei punti dello spazio campione indicati nell'ultima colonna e nell'ultima riga di tab. 1.

In corrispondenza a qualsiasi evento  $A$  esiste un evento ad esso associato indicato con  $\bar{A}$  che stabilisca che l'evento  $A$  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 seguente figura è indicato dalla regione tratteggiata:



Esso viene chiamato il complementare dell'evento A relativo allo spazio campione. Se due sistemi A e B non hanno punti in comune si dice che sono insiemi disgiunti. In termini di eventi tali eventi di chiamo disgiunti o anche reciprocamente esclusivi perché il verificarsi dell'uno esclude la possibilità del verificarsi dell'altro. Due eventi come questi sono rappresentati nella figura seguente.



## Probabilità

Fissiamo ora l'attenzione sulle funzioni di insieme perché ci servono per definire le 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 per 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 anziché punti singoli possono essere insiemi di punti. In questo caso la funzione prende il nome di funzione di insieme. Un esempio di funzione di insieme è quella il cui dominio è costituito da intervalli sull'asse x ed essa fornisce per ciascun intervallo la lunghezza dell'intervallo.

La probabilità è un modello per la frequenza del verificarsi di un certo risultato in ripetute esecuzioni dell'esperimento. Perciò un modello di probabilità per un evento A dovrebbe essere un modello per 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 se nell'esperimento del getto di due dadi  $A$  è l'evento di ottenere un 6 su entrambi i dadi, considerazioni di simmetria suggerirebbero il valore di  $1/36$  per la probabilità del verificarsi di questo evento. 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 la 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 fra 0 e 1 perché una frequenza è un numero di questo tipo. Inoltre la probabilità dell'evento  $S$ , dove  $S$  rappresenta lo spazio campione, dovrebbe essere uguale a 1 perché uno qualunque dei possibili eventi si verificherà certamente quando l'esperimento viene eseguito. Infine se due eventi  $A$  e  $B$  sono disgiunti, la probabilità dell'unione di quegli eventi dovrebbe essere uguale alla somma delle probabilità dei singoli eventi perché la frequenza che  $A$  o  $B$  si verifichi è uguale alla somma delle loro frequenze. Si è provato che ogni altra ragionevole proprietà delle probabilità sarà soddisfatta se sono soddisfatte le seguenti tre condizioni che si basano su quanto appena detto. Queste sono chiamate gli assiomi della probabilità. Essi impongono restrizioni sul tipo di funzione d'insieme  $P$  che si può usare per calcolare le probabilità degli eventi. Una tale funzione viene chiamata 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:

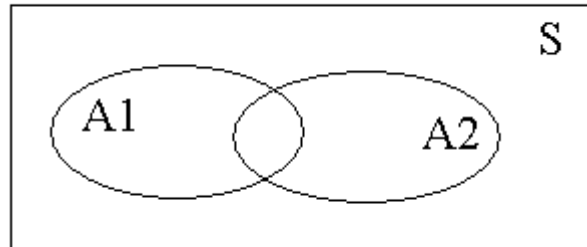
- 1)  $0 \leq P\{A\} \leq 1 \quad \forall A$
- 2)  $P\{S\} = 1$
- 3)  $P\{A_1 \cup A_2 \cup \text{ecc.}\} = P\{A_1\} + P\{A_2\} + \text{ecc.}$   
 $\forall$  *successione finita o infinita di eventi disgiunti*  $A_1, A_2, \dots$

Si fa osservare che gli assiomi della probabilità sono stati formulati nel 1933 da A. N. Kolmogorov. Sarebbe molto difficile trovare una funzione  $P$  che fornisse 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 contenuti soltanto un numero finito o una successione infinita di punti è sufficiente assegnare la probabilità a ciascun punto 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 delle probabilità.

## **Regola di Addizione delle Probabilità o Regola delle Probabilità Totali**

Il calcolo delle probabilità interessa spesso un certo numero di eventi in relazione fra loro anziché un evento soltanto. Per semplicità consideriamo  $A_1, A_2$  associati ad un esperimento. Spesso è interessante sapere se si verificherà almeno uno degli eventi oppure se si verificheranno entrambi quando l'esperimento viene eseguito. Come si è già detto il primo di questi eventi è rappresentato da  $A_1 \cup A_2$  ed il secondo da  $A_1 \cap A_2$ . Per

rispondere a questa domanda occorre calcolare rispettivamente  $P\{A_1 \cup A_2\}$  e  $P\{A_1 \cap A_2\}$ . Cominciamo ora col calcolare  $P\{A_1 \cup A_2\}$ . Lo spazio campione di un esperimento sia rappresentato dai punti nel rettangolo della figura seguente e i punti campione corrispondenti al verificarsi degli eventi  $A_1, A_2$  siano rispettivamente i punti interni alle regioni definite con  $A_1, A_2$ .



Allora come si è già detto l'evento  $A_1 \cup A_2$  consiste di tutti i punti che giacciono all'interno di queste due regioni. La determinazione di una formula per  $P\{A_1 \cup A_2\}$  si basa sull'esprimere  $A_1 \cup A_2$  come l'unione di eventi disgiunti per poi applicare il terzo assioma. Dalla fig. 3 si può osservare che  $A_1 \cup A_2$  è l'unione dei tre insiemi disgiunti  $A_1 \cap \bar{A}_2$ ,  $A_2 \cap \bar{A}_1$  e  $A_1 \cap A_2$  e per il terzo assioma si può quindi scrivere che:

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

ma si può osservare anche che  $A_1$  è l'unione di  $A_1 \cap \bar{A}_2$  e  $A_1 \cap A_2$ , quindi per il terzo assioma  $P\{A_1\} = P\{A_1 \cap \bar{A}_2\} + P\{A_1 \cap A_2\}$ . Allo stesso modo  $A_2$  è l'unione di  $A_2 \cap \bar{A}_1$  e  $A_1 \cap A_2$ , da cui  $P\{A_2\} = P\{A_2 \cap \bar{A}_1\} + P\{A_1 \cap A_2\}$ . Ora ricaveremo da queste ultime due relazioni rispettivamente  $P\{A_1 \cap \bar{A}_2\}$  e  $P\{A_2 \cap \bar{A}_1\}$  e sostituendo le loro espressioni al secondo membro della 1) si ottiene la formula richiesta, nota come regola di addizione delle probabilità o regola delle probabilità totali e cioè:

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

Se due eventi  $A_1, A_2$  non hanno punti campione in comune, come precisato in precedenza, si dicono disgiunti. In questo caso la formula due si riduce allora alla seguente:

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

Questa formula naturalmente è pure una diretta conseguenza del terzo assioma.

La regola dell'addizione delle probabilità espressa dalla 2) si può generalizzare al caso di più eventi. Ad esempio, nel caso di tre eventi  $A_1, A_2, A_3$  arbitrari, si ha che la probabilità

$$P\{A_1 \cup A_2 \cup A_3\} \quad \text{è} \quad \text{uguale} \quad \text{a} \quad P\{A_1\} + P\{A_2\} + P\{A_3\} - P\{A_1 \cap A_2\} - P\{A_1 \cap A_3\} - P\{A_2 \cap A_3\} + P\{A_1 \cap A_2 \cap A_3\}.$$

La regola di addizione è applicabile a qualunque tipo di spazio campione per il quale sia assegnata una misura della probabilità. Per usarla occorre naturalmente conoscere i valori della probabilità al secondo membro delle formule, cosa molto semplice quando lo spazio campione è costituito soltanto da un numero finito di punti, perciò restringeremo ora la discussione a spazi campione di questo tipo. Quindi la prima cosa da fare per calcolare la probabilità di un evento  $A$  per uno spazio campione finito è di assegnare il valore della probabilità a ciascuno dei punti campione. Questi valori di probabilità devono obbedire ai primi due assiomi e cioè devono essere numeri non negativi la cui somma è uguale a 1. Queste assegnazioni vengono fatte basandosi sull'esperienza dei singoli, su informazioni

esterne su considerazioni di simmetria ecc. Ad esempio sarebbe realistico, basandosi su considerazioni di simmetria, assegnare la probabilità di  $1/36$  a ciascun punto dello spazio campione nell'esperimento del getto di due dadi.

## Calcolo della Probabilità di un Evento

Indichiamo con  $n$  il numero totale di punti campione e siano  $p_1, p_2, \dots, p_n$  le probabilità assegnate ai rispettivi punti campione. Ciascun punto rappresenta un possibile risultato che a sua volta è un evento. Eventi di questo tipo sono spesso chiamati eventi semplici. Indicheremo questi eventi con  $e_1, e_2, \dots, e_n$ . È chiaro che questi eventi 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:

$$1) \quad P\{A\} = \sum_A P\{e_i\} = \sum_A p_i$$

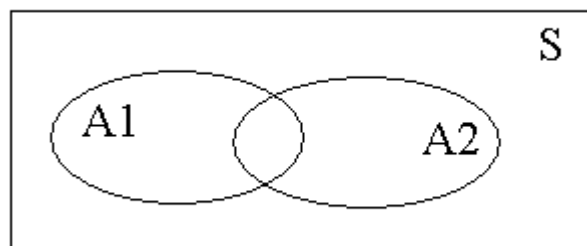
La probabilità di  $A$  è uguale alla somma su  $A$  delle probabilità di  $e_i$ , che è uguale alla somma su  $A$  delle  $p_i$ , dove la sommatoria viene fatta su quelle probabilità  $p_i$  associate ai punti che giacciono in  $A$ .

Per molti giochi d'azzardo lo spazio campione è non soltanto finito ma a ciascuno dei punti campione viene assegnata la stessa probabilità. Questo accade ad esempio per il getto di due dadi per cui lo spazio campione presentato risulta conveniente. A ciascuno di quei punti campione viene assegnata infatti la probabilità è di  $\frac{1}{36}$ . Se  $n$  indica il numero totale di punti campione e  $N(A)$  indica il numero di punti campione nell'insieme  $A$ , allora poiché si è assunto che  $p_i = 1/n$  per  $i=1, 2, \dots, n$  si ha che:

$$2) \quad P\{A\} = \frac{N(A)}{n}$$

## Probabilità Condizionata

Supponiamo ora di voler sapere se un evento  $A_2$  si verificherà sotto la condizione che un evento  $A_1$  sia certo di verificarsi. Per discutere le probabilità associate ad eventi come questi assumiamo che lo spazio campione contenga soltanto un numero finito di punti e consideriamo la situazione mostrata in fig. 1.



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

Se originariamente ad un punto campione in  $A_1$  fosse stata assegnata ad esempio una probabilità doppia rispetto ad un altro punto campione in  $A_1$ , allora si dovrebbe assegnare una probabilità doppia anche nel nuovo spazio campione perché, ignorare i risultati che non producono l'evento  $A_1$  non deve alterare il rapporto di 2:1 delle frequenze attese per quei due punti campione. E' semplicemente necessario perciò moltiplicare le probabilità originarie assegnate ai punti di  $A_1$  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  $p_i$  nell'assegnazione originaria, si dovrebbe scegliere  $\pi_i = c p_i$  in modo tale che  $\sum_{A_1} \pi_i = c \sum_{A_1} p_i = c P\{A_1\} = 1$ , da

cui si ricava che  $c = \frac{1}{P\{A_1\}}$  e quindi  $\pi_i = \frac{p_i}{P\{A_1\}}$ .

Ora che le probabilità del nuovo spazio campione sono state assegnate, si possono calcolare le probabilità nella solita maniera semplicemente applicando la formula 1). Tutte queste probabilità saranno probabilità condizionate, subordinate cioè al verificarsi dell'evento  $A_1$ . La probabilità che si verifichi l'evento  $A_2$  subordinata alla condizione che

debba verificarsi l'evento  $A_1$  si indica con  $P\{A_2|A_1\} = \sum_{A_1 \cap A_2} \pi_i = \frac{\sum_{A_1 \cap A_2} p_i}{P\{A_1\}}$  (la barra indica "a

condizione che"). La prima somma viene fatta su quelle  $\pi_i$  corrispondenti ai punti campione che giacciono in  $A_1 \cap A_2$  perché essi sono i soli punti campione dentro  $A_1$  che corrispondono al verificarsi di  $A_2$ . Poiché la somma al numeratore nell'ultima espressione è quella che definisce  $P\{A_1 \cap A_2\}$  si ha allora che il valore della probabilità condizionata è dato da (si assume che  $A_1$  sia un elemento tale che  $P\{A_1\} > 0$ ):

$$3) P\{A_2|A_1\} = \frac{P\{A_1 \cap A_2\}}{P\{A_1\}}$$

Nel procedimento per arrivare alla 3) si è assunto che lo spazio campione contenga soltanto un numero finito di punti ma questa formula verrà presa come definizione di probabilità condizionata anche per spazi campione più generali. Si dimostra facilmente che  $P\{A_2|A_1\}$  soddisfa i tre assiomi della probabilità perciò è legittimo definire la probabilità condizionate in questo modo, a prescindere dal tipo di spazio campione. La suddetta formula 3) scritta sotto forma di prodotto fornisce la regola fondamentale di moltiplicazione delle probabilità detta anche regola delle probabilità composte e cioè:

$$4) P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2|A_1\}$$

Se si scambia l'ordine dei due eventi la formula 4) diventa:

$$5) P\{A_1 \cap A_2\} = P\{A_2\} \cdot P\{A_1|A_2\}$$

## Eventi indipendenti

Supponiamo ora che  $A_1$  e  $A_2$  siano due eventi tali che  $P\{A_2|A_1\} = P\{A_2\}$  e  $P\{A_1\} \cdot P\{A_2\} > 0$ , allora l'evento  $A_2$  si dice indipendente nel senso della probabilità o

più brevemente indipendente da  $A_1$  poiché la probabilità del verificarsi di  $A_2$  non viene modificata imponendo la condizione che debba verificarsi l'evento  $A_1$ . Quando  $A_2$  è indipendente da  $A_1$  la regola di moltiplicazione espressa dalla 4) diventa:

$$6) P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2\}$$

viceversa quando è vera la 6) segue dal confronto di questa con la suddetta formula 4) che  $A_2$  è indipendente da  $A_1$ .

Se si eguagliano i secondi membri delle 6) e 5) si può osservare che  $P\{A_1|A_2\} = P\{A_1\}$  ma ciò stabilisce che l'evento  $A_1$  è indipendente da  $A_2$ . Così se  $A_2$  è indipendente da  $A_1$  segue che  $A_1$  deve essere indipendente da  $A_2$ . A causa di questa reciproca indipendenza e poiché la 6) implica questa indipendenza si suole definire l'indipendenza fra gli eventi come segue.

Definizione: due eventi  $A_1$  e  $A_2$  si dicono indipendenti se  $P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2\}$ .

**Esercizi:** come applicazione pratica delle formule fondamentali per il calcolo delle probabilità di un evento consideriamo ora la risoluzione di alcuni problemi.

Problema 1: supponiamo che approssimativamente il 50% degli individui di età superiore a 40 anni sia in sovrappeso e che la percentuale degli individui in sovrappeso e aventi una malattia dell'apparato circolatorio sia il 25%. Si vuole calcolare allora la probabilità che un individuo scelto a caso con più di 40 anni e in sovrappeso (rappresentante l'evento  $A_1$ ) abbia una malattia dell'apparato circolatorio (rappresentante l'evento  $A_2$ ).

Questa probabilità è quindi data da:  $P\{A_2|A_1\} = \frac{P\{A_1 \cap A_2\}}{P\{A_1\}} = \frac{1/4}{1/2} = \frac{1}{2} = 50\%$

Problema 2: nell'ipotesi che una famiglia abbia due figli si vuole calcolare la probabilità che entrambi i figli siano maschi sapendo che almeno uno di essi è maschio.

Assumiamo che lo spazio campione  $S$  sia dato da  $S = \{(m, m), (m, f), (f, m), (f, f)\}$ . Nell'ipotesi che, ad esempio,  $(m, f)$  significhi che il figlio maggiore è femmina ed il minore è maschio e che tutti gli elementi dello spazio campione siano ugualmente probabili.

Indichiamo con  $A_1$  l'evento che entrambi i figli siano maschi ( $A_1 = \{(m, m)\}$ ) e con  $A_2$  l'evento che almeno uno di essi sia maschio ( $A_2 = \{(m, m), (m, f), (f, m)\}$ ).

La probabilità è data da:  $P\{A_1|A_2\} = \frac{P\{A_1 \cap A_2\}}{P\{A_2\}} = \frac{P\{(m, m)\}}{P\{(m, m), (m, f), (f, m)\}} = \frac{1/4}{3/4} = \frac{1}{3}$

Problema 3: Supponiamo che uno studente affronti due esami e che la probabilità di superare il primo esame sia 0,6 mentre sia uguale a 0,8 quella di superare il secondo. Inoltre sia 0,5 la probabilità che egli superi entrambi gli esami. Si vuole calcolare la probabilità che lo studente superi almeno un esame e quella che sia bocciato in entrambi.

A tale scopo indicheremo con  $A_1$  l'evento corrispondente alla promozione al primo esame e con  $A_2$  quella corrispondente alla promozione al secondo. Allora  $A_1 \cap A_2$  è l'evento che lo studente superi entrambi gli esami e  $A_1 \cup A_2$  l'evento che superi almeno un esame. Si ha quindi che  $P\{A_1 \cup A_2\} = P\{A_1\} + P\{A_2\} - P\{A_1 \cap A_2\} = 0,6 + 0,8 - 0,5 = 0,9 = 90\%$

L'evento che lo studente fallisca in ambedue gli esami è  $\overline{A_1 \cup A_2}$  e quindi  $P\{\overline{A_1 \cup A_2}\} = 1 - P\{A_1 \cup A_2\} = 1 - 0,9 = 0,1 = 10\%$

A questo punto si vuole calcolare inoltre la probabilità che lo studente superi il secondo esame supposto che abbia superato il primo e analogamente la probabilità che superi il

primo esame supposto che abbia superato il secondo. Allora la prima probabilità richiesta è data da  $P\{A_2|A_1\} = \frac{P\{A_1 \cap A_2\}}{P\{A_1\}} = \frac{0,5}{0,6} = 0,8\bar{3} = 83, \bar{3} \%$  e la seconda probabilità è data da

$$P\{A_1|A_2\} = \frac{P\{A_1 \cap A_2\}}{P\{A_2\}} = \frac{0,5}{0,8} = 0,625 = 62,5 \%$$

È importante osservare che la probabilità di superare un esame aumenta nel caso in cui si sia già superato l'altro. In altre parole la promozione in uno degli esami, può aiutare la prestazione dello studente a superare l'altro in quanto ne aumenta la fiducia.

A questo punto ci si chiede se gli eventi  $A_1$  e  $A_2$  sono indipendenti dal momento che  $P\{A_1 \cap A_2\} = 0,5$  e  $P\{A_1\} \cdot P\{A_2\} = 0,6 \cdot 0,8 = 0,48$  e quindi  $P\{A_1 \cap A_2\} \neq P\{A_1\} \cdot P\{A_2\}$ .

Si ha come risultato che gli eventi non sono indipendenti. In effetti, dal momento che  $P\{A_1 \cap A_2\} > P\{A_1\} \cdot P\{A_2\}$  si ricava che, come si è già detto, la promozione in uno degli esami può aiutare lo studente a superare l'altro.

Problema 4: Supponiamo che da un'urna contenente sette palline nere e cinque bianche vengano estratte due palline senza sostituzione delle palline estratte, ossia senza reimbussolamento. Si vuole calcolare la probabilità che entrambe le palline estratte siano nere nell'ipotesi che ogni pallina dell'urna abbia la stessa probabilità di essere estratta.

Indicheremo con  $A_1$  l'evento che la prima pallina estratta sia nera e con  $A_2$  l'evento che anche la seconda pallina estratta sia nera. Ora nell'ipotesi che la prima pallina estratta sia nera, restano 6 palline nere e 5 bianche e quindi si ha che  $P\{A_2|A_1\} = \frac{6}{11}$ . Dal momento

che  $P\{A_1\} = \frac{7}{12}$  si ha che  $P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2|A_1\} = \frac{7}{12} \cdot \frac{6}{11} = 0,318 = 31,8 \%$

Problema 5: consideriamo il lancio di due dadi simmetrici. Sia  $A_1$  l'evento che la somma dei due numeri ottenuti sia uguale a 6 e  $C$  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) e si ha quindi che  $P\{A_1 \cap C\} = \frac{1}{36} = 0,027$ . Poiché

$P\{A_1\} \cdot P\{C\} = \frac{5}{36} \cdot \frac{1}{6} = \frac{5}{216} = 0,023$  si ha che  $P\{A_1 \cap C\} \neq P\{A_1\} \cdot P\{C\}$ , quindi i due eventi non sono indipendenti. In effetti, ad esempio, se il primo dado mostrasse un 6 l'evento  $A_1$  diventerebbe impossibile.

Consideriamo ora come  $A_2$  l'evento di ottenere come somma 7. Allora in questo caso  $A_2 \cap C$  è rappresentato dall'unico risultato (4,3) e quindi  $P\{A_2 \cap C\} = \frac{1}{36}$ . Poiché

$P\{A_2\} \cdot P\{C\} = \frac{6}{36} \cdot \frac{1}{6} = \frac{1}{36}$  si ha che  $P\{A_2 \cap C\} = P\{A_2\} \cdot P\{C\}$  e quindi i due eventi risultano indipendenti. Infatti ottenere come somma 7 non dipende dal risultato che esce sul primo dado.

Esercitazione: consideriamo ora il problema seguente che, oltre ad essere un'applicazione per le formule fin qui esaminate, ci introduce inoltre al calcolo delle probabilità delle cause del verificarsi di un evento. A tale scopo supponiamo che una scatola contenga 2 palline rosse e una seconda scatola, identica, contenga una pallina rossa e una bianca. Se si sceglie una scatola a caso e si estrae da essa una pallina, si vuole calcolare la probabilità di avere scelto la prima scatola se la pallina estratta è rossa.



Indichiamo con  $A_1$  l'evento di scegliere la prima scatola e con  $\overline{A_1}$  quello di scegliere la seconda scatola. Indicheremo poi con  $A_2$  l'evento di estrarre una pallina rossa e con  $\overline{A_2}$  quello di estrarre una pallina bianca. Si tratta quindi di calcolare la probabilità condizionata

$P\{A_1|A_2\}$ . Si ha che  $P\{A_1|A_2\} = \frac{P\{A_1 \cap A_2\}}{P\{A_2\}}$ . Per la regola di moltiplicazione delle

probabilità si ha che  $P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2|A_1\} = \frac{1}{2} \cdot 1 = \frac{1}{2}$ . Si può osservare che

l'evento  $A_2$  si verificherà se e solo se si verificherà l'uno o l'altro dei due eventi reciprocamente esclusivi (o disgiunti)  $A_1 \cap A_2$  e  $\overline{A_1} \cap A_2$ . Si ha che

$P\{A_2\} = P\{A_1 \cap A_2\} + P\{\overline{A_1} \cap A_2\}$ . Sempre per la regola di moltiplicazione si ha che

$P\{\overline{A_1} \cap A_2\} = P\{\overline{A_1}\} \cdot P\{A_2|\overline{A_1}\} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$  e quindi  $P\{A_2\} = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$ . Perciò

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

Questo esempio è stato presentato come applicazione delle formule fondamentali della probabilità. Tuttavia esso si sarebbe potuto risolvere più semplicemente ricorrendo allo spazio campione dell'esperimento ed applicando poi una formula già vista e cioè

$P\{A\} = \sum_A p_i$ . Per il problema in esame sarebbe conveniente considerare come spazio

campione quello costituito dai quattro punti I1, I2, II1, II2, dove il numero romano rappresenta quello della scatola e l'altro il numero della pallina. A questi quattro punti si devono assegnare uguali probabilità pari ad  $\frac{1}{4}$ . La condizione del verificarsi dell'evento

$A_2$  restringe lo spazio campione soltanto ai primi tre punti nell'ipotesi che il numero 2 della pallina nella II scatola rappresenti la pallina bianca. A ciascuno di questi tre punti occorrerà quindi assegnare la probabilità di  $\frac{1}{3}$  ma soltanto i primi due punti campione corrispondono al verificarsi dell'evento  $A_1$  e quindi per la formula suddetta si ha che

$$P\{A_1|A_2\} = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

## Formula di Bayes

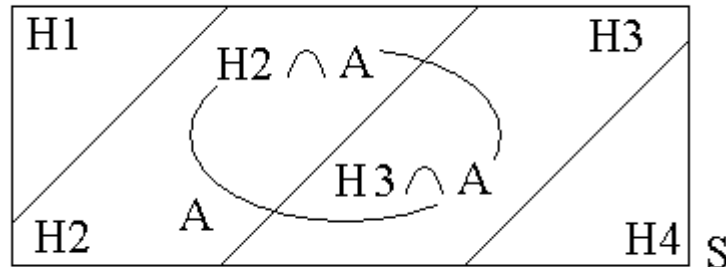
L'esempio considerato è tipico di problemi in cui si considera il risultato di un esperimento e poi ci si chiede qual è la probabilità che esso sia dovuto ad una fra le possibili cause del suo verificarsi. Così nell'esempio sono due le possibili cause dell'estrazione di una pallina rossa e il problema 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 sistematicamente. A tale scopo supponiamo che

$H_1, H_2, \dots, H_n$  siano eventi reciprocamente esclusivi con  $P\{H_i\} > 0$  per  $i=1, 2, \dots, n$  e

tali che  $\sum_{i=1}^n H_i = S$  dove S è lo spazio campione. Si dice che gli eventi  $H_i$  per i che va da

1 a n costituiscono una partizione dello spazio campione. 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  $A$  sono mostrate nella figura seguente dove lo spazio campione  $S$  è suddiviso per comodità in 4 sottoinsiemi:



Si tratta quindi di calcolare la probabilità condizionata  $P\{H_i|A\}$ . Questa è data da:

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

La regola di moltiplicazione fornisce:

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

Sostituendo la 2) nella 1) si ottiene:

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

Dalla fig. 1 è chiaro che l'evento  $A$  è l'unione degli eventi disgiunti  $H_1 \cap A, H_2 \cap A, \dots, H_n \cap A$  e che alcuni degli insiemi corrispondenti a questi eventi possono naturalmente essere vuoti (come del resto accade in fig. 1).

Per il terzo assioma si può quindi scrivere:  $P\{A\} = \sum_{j=1}^n P\{H_j \cap A\}$

Se si applica la 2) al secondo membro di quest'ultima si ottiene che:

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

Questa formula stabilisce che  $P\{A\}$  è uguale ad una media pesata di  $P\{A|H_j\}$ , essendo ciascun termine pesato dalla probabilità dell'evento rispetto al quale  $A$  è condizionato.

La sostituzione della 4) nella 3) fornisce la formula richiesta per calcolare la probabilità delle cause, nota come Formula di Bayes ed espressa dalla:

$$5) \quad P\{H_i|A\} = \frac{P\{H_i\} \cdot P\{A|H_i\}}{\sum_{j=1}^n P\{H_j\} \cdot P\{A|H_j\}} \quad \text{per } i=1, 2, \dots, n$$

Ambedue le formule 4) e 5) viste ieri sono utili per 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 precedenti agli esperimenti composti  $H_i$  rappresentano i 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.

Le probabilità  $P\{H_i\}$  sono spesso chiamate probabilità a priori mentre le probabilità condizionate  $P\{H_i|A\}$  probabilità a posteriori. Queste denominazioni derivano del fatto che in molte applicazioni  $P\{H_i\}$  sono probabilità soggettive che rappresentano la nostra opinione prima dell'esperimento mentre le probabilità  $P\{H_i|A\}$  possono essere interpretate come la descrizione della nostra opinione dopo che sono stati effettuati alcuni esperimenti e si è verificato l'evento  $A$ . In altre parole la formula di Bayes può essere considerata anche come un algoritmo per cambiare la nostra opinione sulla base di risultati sperimentali. Questo aspetto verrà illustrato fra poco esaminando un esempio specifico.

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

Esempio 1: per illustrare l'uso della formula di Bayes risolviamo ora il problema considerato ieri 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 viene scelta a caso una scatola,  $P\{H_1\}=P\{H_2\}=1/2$ . È chiaro che tenuto conto dei contenuti delle due scatole si ha che  $P\{A|H_1\}=1$  e  $P\{A|H_2\}=1/2$ , quindi l'applicazione diretta della formula di Bayes espressa dalla 5) fornisce:

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

Esempio 2: in una diagnosi medica si osserva che un paziente ha uno o più sintomi specifici  $A=\{S_1, S_2, \dots, S_n\}$  e si pone il problema di decidere quale delle possibili malattie  $\{H_1, H_2, \dots, H_k\}$  è la causa più probabile dei sintomi osservati. Si suppone di avere una stima statistica delle probabilità  $P\{H_i\}=p_i$  con  $p_i > 0$  per  $i=1, 2, \dots, k$  di contrarre la malattia  $H_i$ . Si suppone inoltre che le singole malattie non siano contemporaneamente presenti nello stesso paziente. Si assume inoltre di conoscere una stima statistica delle probabilità  $P\{A|H_i\}$ , ovvero della probabilità che la malattia  $H_i$  dia origine ai sintomi  $A$ . Applicando la formula di Bayes la probabilità che i sintomi  $A$  siano dovuti alla malattia  $H_i$  per  $i=1, 2, \dots, k$ , 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) vista in precedenza.

Come esempio supponiamo che  $P\{H_1\}=0,4$ ,  $P\{H_2\}=0,25$ ,  $P\{H_3\}=0,35$  ed inoltre che  $P\{A|H_1\}=0,8$ ,  $P\{A|H_2\}=0,6$ ,  $P\{A|H_3\}=0,9$ .

Applicando la formula 4) si ha che:

$$P\{A\} = P\{H_1\} \cdot P\{A|H_1\} + P\{H_2\} \cdot P\{A|H_2\} + P\{H_3\} \cdot P\{A|H_3\}$$

Sostituendo i valori numerici otteniamo:

$$P\{A\} = 0,4 \cdot 0,8 + 0,25 \cdot 0,6 + 0,35 \cdot 0,9 = 0,785$$

Ora applicando la formula 5) si ha che:

$$P\{H_1|A\} = \frac{P\{H_1\} \cdot P\{A|H_1\}}{P\{A\}} = \frac{0,4 \cdot 0,8}{0,785} = 0,4076 = 40,76 \%$$

$$P\{H_2|A\} = \frac{P\{H_2\} \cdot P\{A|H_2\}}{P\{A\}} = \frac{0,25 \cdot 0,6}{0,785} = 0,1911 = 19,11 \%$$

$$P\{H_3|A\} = \frac{P\{H_3\} \cdot P\{A|H_3\}}{P\{A\}} = \frac{0,35 \cdot 0,9}{0,785} = 0,4013 = 40,13 \%$$

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

Esempio 3: supponiamo che il Ministero delle Finanze prima di adottare una politica di controllo dei prezzi e dei salari chieda il parere di tre esperti indicati con  $\{H_1, H_2, H_3\}$  sull'impatto che tale politica può avere sul tasso di disoccupazione. Le risposte date in termini di probabilità dei singoli esperti sono raccolte nella tabella seguente.

Tabella 1: probabilità di cambiamento nel tasso di disoccupazione.

Esperto	Diminuzioni	Stabilità	Aumenti
H1	0.1	0.1	0.8
H2	0.6	0.2	0.2
H3	0.2	0.6	0.2

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

$$P\{H_1\} = 1/6$$

$$P\{H_2\} = 1/3$$

$$P\{H_3\} = 1/2$$

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

Per risolvere questo problema si tratta di calcolare:  $P\{H_1|A\}, P\{H_2|A\}, P\{H_3|A\}$ .

Usando la formula 4) calcoliamo dapprima la probabilità dell'aumento del tasso di disoccupazione. Si ha che:

$$P\{A\} = P\{H_1\} \cdot P\{A|H_1\} + P\{H_2\} \cdot P\{A|H_2\} + P\{H_3\} \cdot P\{A|H_3\} = \frac{1}{6} \cdot 0,8 + \frac{1}{3} \cdot 0,2 + \frac{1}{2} \cdot 0,2 = 0,3$$

Applicando ora la formula di Bayes si ha che:

$$P\{H_1|A\} = \frac{P\{H_1\} \cdot P\{A|H_1\}}{P\{A\}} = \frac{1/6 \cdot 0,8}{0,3} = 0,4$$

$$P\{H_2|A\} = \frac{P\{H_2\} \cdot P\{A|H_2\}}{P\{A\}} = \frac{1/3 \cdot 0,2}{0,3} = 0,2$$

$$P\{H_3|A\} = \frac{P\{H_3\} \cdot P\{A|H_3\}}{P\{A\}} = \frac{1/2 \cdot 0,2}{0,3} = 0,3$$

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

## Variabili Casuali o Aleatorie

Consideriamo ora lo spazio campione mostrato in precedenza corrispondente all'esperimento del lancio di due monete e fissiamo l'attenzione sul numero di teste ottenute. Per calcolare la probabilità di possibili risultati è conveniente introdurre una variabile  $X$  per rappresentare il numero di teste ottenute. Allora  $X$  assumerà il valore 0 nel

punto campione CC; 1 nei punti TC, CT e 2 nel punto TT. Una variabile come questa a valori numerici è un esempio di variabile casuale. Un secondo esempio riguarda l'esperimento del getto di due dadi: se  $X$  rappresenta la somma dei punti ottenuti, allora essa è una variabile casuale, che può assumere i valori interi da 2 a 12. Come altro esempio se  $X$  rappresenta la distanza dal centro di una freccia scagliata su di un bersaglio circolare di raggio 20 allora assumendo di trascurare tutti i colpi mancanti essa è una variabile casuale che può assumere qualsiasi valore da 0 a 20.

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

Si può dare la seguente **definizione**: una variabile casuale  $X$  è una funzione a valori reali definita su di uno spazio campione. Il motivo per cui la variabile viene chiamata casuale è 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.

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

## Variabili Casuali Discrete

Dopo che una variabile casuale  $X$  è stata definita su di uno spazio campione, l'interesse si concentra di solito nel calcolare la probabilità che  $X$  assuma valori specifici nel suo insieme di valori possibili. Per esempio se  $X$  rappresenta la somma dei punti nel getto dei due dadi, allora può interessare il calcolo della probabilità che  $X$  assuma ad esempio il valore 7. Oppure se  $X$  rappresenta la distanza di una freccia dal centro di un bersaglio circolare, può interessare calcolare la probabilità che  $X$  assuma ad esempio un valore minore di 5. Il calcolo di  $P\{X=7\}$  nel primo esempio è molto più semplice del calcolo di  $P\{X<5\}$  del secondo perché è molto più semplice lavorare con uno spazio campione che consiste di un numero finito di punti che con uno che consiste di un intervallo di punti sull'asse  $X$ . Uno spazio campione che consiste di un numero finito o di una successione infinita di punti è chiamato uno **spazio campione discreto** mentre uno che consiste di uno o più intervalli di punti viene chiamato **spazio campione continuo**. Per intervalli si intendono naturalmente intervalli di dimensione qualsiasi. Poiché la teoria per gli spazi campione discreti è molto più semplice di quella per gli spazi continui limiteremo per ora la discussione agli spazi campione discreti. E' chiaro che una variabile casuale non può assumere più valori di quanti sono i punti campione, perciò una variabile definita su di uno spazio campione discreto può assumere soltanto un numero finito o una successione infinita di valori. Una variabile casuale come questa viene chiamata una **variabile casuale discreta**.

Per calcolare le probabilità delle variabili casuali discrete ricordiamo la formula per calcolare un evento  $A$ , cioè  $P\{A\} = \sum_A p_i$ , ricavata assumendo che lo spazio campione fosse discreto e applicabile a qualsiasi campione discreto. Se  $x$  è uno specifico valore della variabile casuale discreta  $X$ , questa formula si può usare per calcolare la probabilità che  $X$  assuma il valore  $x$ .

Questa probabilità è perciò data da:

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

dove la sommatoria viene fatta su tutti i punti campione in cui la variabile casuale X ha valore x. Come esempio, se X rappresenta la somma dei punti ottenuti nel getto di due dadi, tenuto conto che tutti i punti campione hanno la probabilità di 1/36 allora

$$P\{X=7\} = \sum_{X=7} p_i = 6/36$$

Questo risultato deriva dal fatto che vi sono 6 punti campione in cui il valore di X è uguale a 7, e precisamente i punti di coordinate  $\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$ .

## Funzione di Probabilità o Funzione di densità di Probabilità

Per calcolare le probabilità delle variabili casuali discrete è conveniente introdurre una funzione chiamata funzione di probabilità o funzione di densità di probabilità o più brevemente densità. Si può dare la seguente **definizione**: sia X una variabile casuale discreta, allora la funzione f definita da  $f(x) := P\{X=x\}$  (:= vuol dire "uguale per definizione") 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 una distribuzione continua di massa lungo una retta o un piano e che perciò il suo uso a proposito di una distribuzione di probabilità discreta può sembrare improprio. Tuttavia, poiché è desiderabile usare la stessa terminologia per questo tipo di funzione sia per le variabili discrete che per quelle continue e poiché il termine densità è appropriato per le variabili continue esso verrà usato anche nel caso discreto.

**Esempio:** Una funzione di densità spesso consiste semplicemente di una tabella di valori, così nel 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:

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

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

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

Si fa notare che la funzione  $f(x) = 0$  se  $x \neq 0, 1, 2$ .

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 densità per mezzo di un grafico a segmenti. Come esempio di tale grafico la variabile X rappresenti la somma dei punti ottenuti lanciando due dadi. Attribuendo l'esatto valore di X a ciascuno dei 36 punti dello spazio campione corrispondente al lancio di due dadi e assumendo la probabilità uguali per i punti campione mediante l'impiego della formula 1) si ottiene:

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

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

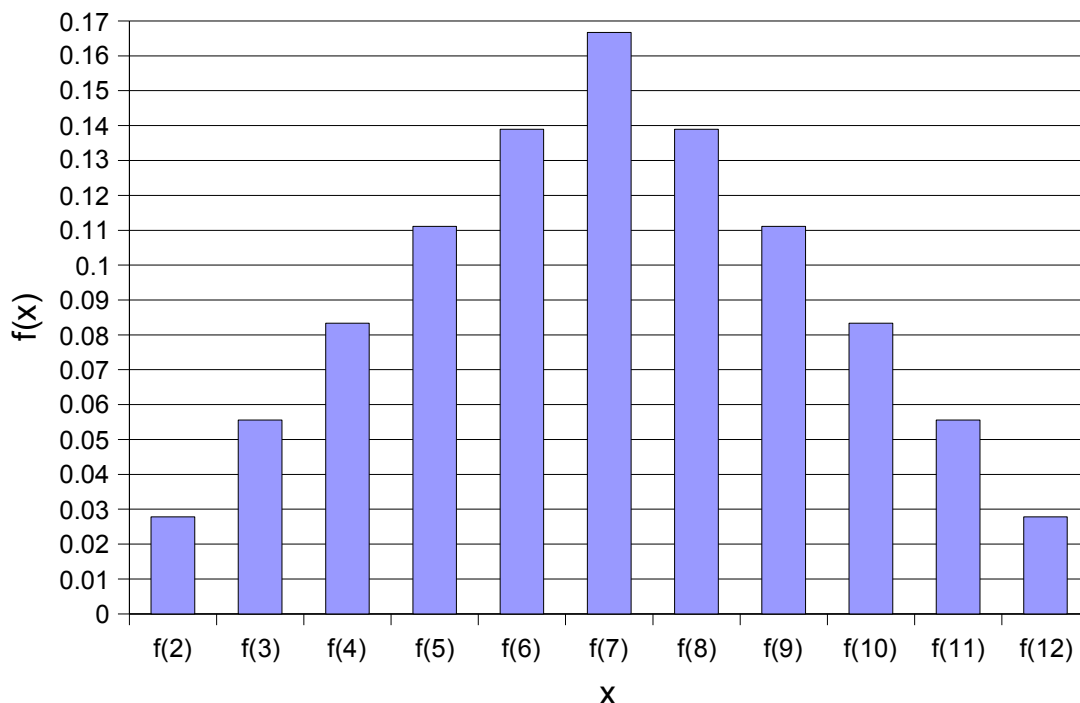
$$f(4) = f(10) = 3/36$$

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

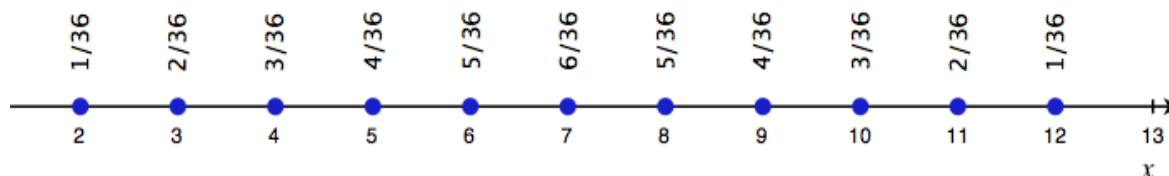
$$f(6) = f(8) = 5/36$$

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

Il grafico a segmenti di  $f(x)$  è mostrato in figura 1):



Lo scopo di introdurre una funzione di densità è quello 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), si possono 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 di  $f(x)$  attribuiti a quei punti. Per esempio, Se  $X$  rappresenta la somma dei punti nel lancio di due dadi il nuovo spazio campione consisterà degli undici punti 2, 3, ..., 12 e le probabilità associate a quei punti saranno i valori di  $f(x)$  calcolati in precedenza. Questo nuovo spazio campione è semplicemente una rappresentazione dell'informazione trasmessa dalla figura 1) ed è mostrato in figura 2):



Ora consideriamo il calcolo delle 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  $X$  assuma valori in  $R$ . Considerando il nuovo spazio campione generato da  $X$  e da  $f(x)$ , una diretta applicazione di una formula già vista, cioè  $P\{A\} = \sum_A p_i$ , fornisce il risultato

$$2) P\{X \in R\} = \sum_{x \in R} f(x)$$

dove la sommatoria viene fatta su quei valori  $x$  in  $R$  dove  $f(x) > 0$ . Il calcolo delle probabilità in questo modo è di solito molto più semplice del calcolo che si basa sullo spazio campione originario dell'esperimento. Usando nuovamente l'esempio del lancio di due dadi supponiamo di voler calcolare la probabilità che la somma dei punti sia maggiore di 7. Considerando il nuovo spazio campione questa probabilità è data da:

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

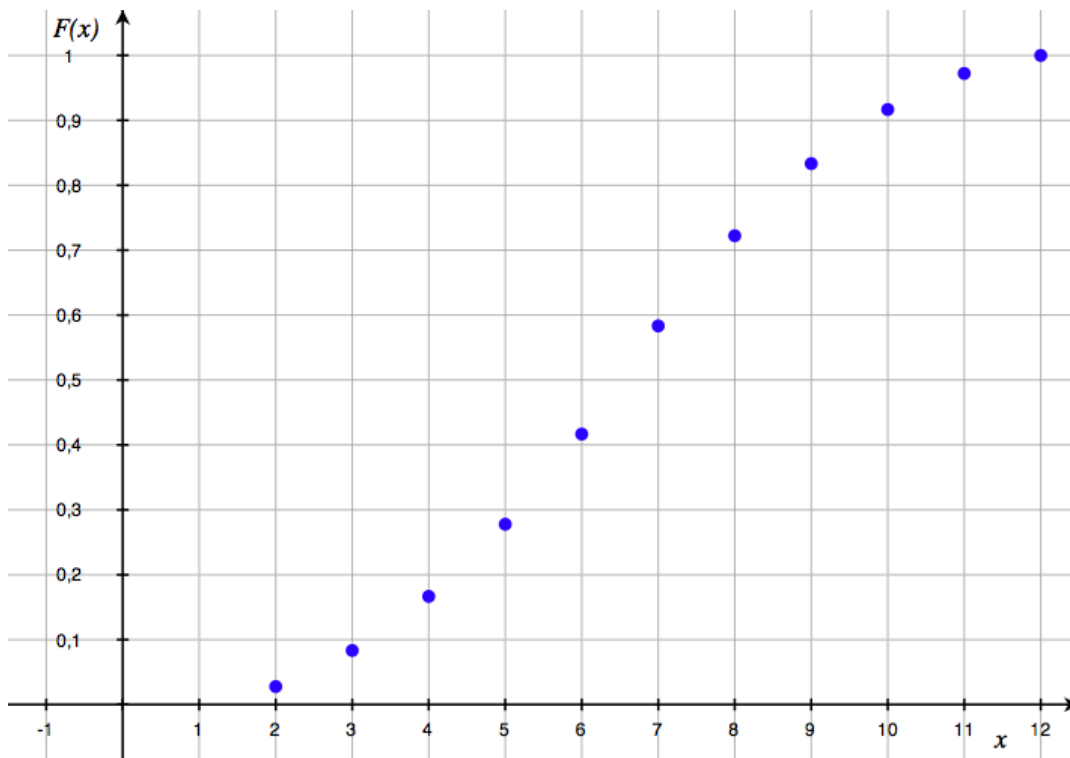
Se questa probabilità si dovesse calcolare usando lo spazio campione originario sarebbe necessario sommare la probabilità dei 15 punti campione le cui coordinate hanno come somma un valore maggiore di 7, cioè  $P\{x > 7\} = \sum_{x>7} p_i = \underbrace{1/36 + \dots + 1/36}_{15 \text{ volte}} = 15/36$ . In questo caso in effetti non c'è un vantaggio particolare ad usare il nuovo spazio campione, però per spazi campione più complicati il vantaggio può essere considerevole.

## Funzione di Distribuzione o di Distribuzione Cumulativa o di Ripartizione

Una funzione strettamente affine alla funzione di densità  $f$  è la corrispondente funzione di distribuzione  $F$ . Essa è definita come segue:

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

dove la sommatoria viene fatta su tutti i valori della variabile casuale discreta che sono minori uguali di  $x$ . Il grafico della  $F(x)$  corrispondente alla  $f(x)$  di figura 1 è riportato nella figura seguente ed è dato dal valore assunto da  $F(x)$  negli 11 punti  $x=2, 3, \dots, 12$ .





I punti del grafico sono dati da:

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

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

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

## Funzione di Probabilità Congiunta o di Densità di Probabilità Congiunta

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

**Esempio:** consideriamo un esempio in cui la variabile  $X$  rappresenta il numero di carte di picche ottenute nell'estrarre una carta da un mazzo e la variabile  $Y$  rappresenta il numero di carte di picche ottenute nell'estrarre una seconda carta dal mazzo senza che la prima sia stata reinserita. In questo caso le variabili  $X, Y$  possono assumere soltanto i valori 0 e 1. Allora, ricordando la regola di moltiplicazione delle probabilità, la funzione di densità  $f(x, y)$  è definita dalla seguente tabella di valori:

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

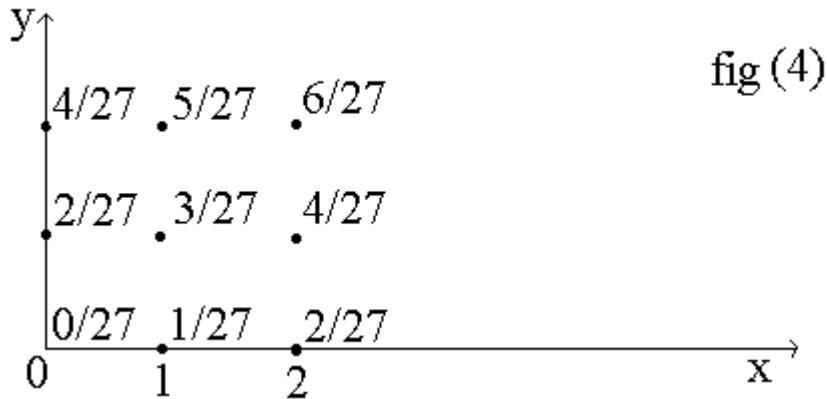
$$f(0,1) = P\{X=0, Y=1\} = P\{X=0\} \cdot P\{Y=1|X=0\} = \frac{39}{52} \cdot \frac{13}{51} = 0,19$$

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

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

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

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



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

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

## Variabili Indipendenti

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

$$1) \quad P\{x \in A, y \in B\} = P\{x \in A\} \cdot P\{y \in B\} \quad \forall A, B$$

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

Ma in termini di funzioni di densità questa si scrive:  $f(x, y) = g(x) \cdot h(y) \quad \forall x, y$

Viceversa se  $f(x, y) = g(x) \cdot h(y) \quad \forall x, y$  allora operando nello spazio campione delle due variabili segue per analogia con la 2) precedentemente descritta, ovvero

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

$$P\{x \in A, y \in B\} = \sum_{x \in A} \sum_{y \in B} f(x, y) = \sum_{x \in A} \sum_{y \in B} g(x) \cdot h(y) = \sum_{x \in A} g(x) \cdot \sum_{y \in B} h(y) = P\{x \in A\} \cdot P\{y \in B\}$$

Questa mostra che la 1) vale per gli  $A, B$  se e soltanto se  $f(x, y) = g(x) \cdot h(y) \quad \forall x, y$ . Poiché quest'ultima relazione è espressa in termini di funzioni di densità ed è più utile della 1) essa viene usata per definire l'indipendenza.

Si può dunque dare la seguente **definizione**: le variabili casuali  $X, Y$ , la cui funzione di densità congiunta è  $f(x, y)$  e le cui funzioni di densità sono  $g(x)$  e  $h(y)$ , sono indipendenti se e soltanto se

$$2) \quad f(x, y) = g(x) \cdot h(y) \quad \forall x, y$$

La precedente definizione si può generalizzare in ovvia maniera per definire l'indipendenza di  $n$  variabili casuali, così (**definizione**): le variabili casuali  $X_1, X_2, \dots, X_n$  la cui funzione di densità congiunta è  $f(x_1, x_2, \dots, x_n)$  e le cui funzioni di densità singole sono  $f_1(x_1), f_2(x_2), \dots, f_n(x_n)$  sono indipendenti se e soltanto se  $f(x_1, x_2, \dots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_n(x_n) \quad \forall x_1, x_2, \dots, x_n$ .

## Distribuzioni Marginali e Distribuzioni Condizionate

Poiché è importante sapere se un insieme di variabili è costituito da variabili indipendenti sarebbe desiderabile disporre di un metodo sistematico per ricavare le funzioni di densità delle singole variabili dalla loro determinata 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  $x$  ( $A_1 \rightarrow X = x$ ) e  $A_2$  l'evento che una seconda variabile casuale  $Y$  assuma il valore  $y$  ( $A_2 \rightarrow Y = y$ ).

La regola di moltiplicazione 4)  $P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2|A_1\}$  assume allora la forma che deriva dall'esprimere le probabilità degli eventi in termini di funzioni di densità. Poiché ora  $P\{A_1 \cap A_2\}$  dà la probabilità che le due variabili casuali assumano rispettivamente i valori  $x, y$ , essa rappresenta il valore della funzione di densità congiunta nel punto  $(x, y)$ , cioè la  $f(x, y)$ .  $P\{A_1\}$  in modo simile è la probabilità che la variabile  $X$  assuma il valore  $x$ , perciò essa è la  $f(x)$ .

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

$$3) \quad f(x, y) = f(x) \cdot f(y|x)$$

Poiché  $f(y|x)$  dà la probabilità condizionata che  $Y$  assuma il valore  $y$  quando  $X$  ha per valore fissato il valore  $x$ , la somma di  $f(y|x)$  su tutti i possibili valori di  $y$  deve essere uguale a 1. Quindi se entrambi i membri della 3) si sommano su tutti i possibili valori di  $y$  si ottiene  $\sum_y f(x, y) = \sum_y f(x) \cdot f(y|x) = f(x) \cdot \sum_y f(y|x) = f(x)$ , per cui

(**definizione**):

$$4) \quad f(x) = \sum_y f(x, y)$$

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

In modo analogo la funzione di densità della variabile  $Y$ ,  $g(y)$ , si può ottenere

sommando  $f(x, y)$  su tutti i possibili valori della variabile  $x$  per un fissato valore  $y$ . Si ha così che:

$$5) \quad g(y) = \sum_x f(x, y)$$

Gli ultimi risultati mostrano che se si ha la funzione di densità congiunta di due variabili casuali discrete e si desidera la funzione di densità di una di esse è semplicemente necessario sommare la funzione di densità congiunta su tutti i valori dell'altra variabile.

La funzione di densità condizionata  $f(y|x)$  dà la distribuzione di probabilità della variabile  $Y$  quando viene mantenuto fisso il valore di  $X$ . Per la formula 3) se  $f(x) > 0$  si può scrivere (**definizione**):

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

La distribuzione condizionata della variabile  $X$  quando si ha mantenuto fisso il valore della variabile  $Y$  se  $g(y) > 0$  è data da una formula analoga:

$$7) \quad f(x|y) = \frac{f(x, y)}{g(y)}$$

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

Esempio: 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  $Y$  rappresentano i valori delle due estrazioni, 0 corrisponde ad una pallina nera e 1 ad una pallina bianca. Allora ogni possibile risultato sarà rappresentato da uno dei quattro punti del piano  $x, y$  di fig. 1 di coordinate  $(0,0), (0,1), (1,0), (1,1)$ .

Dai contenuti dell'urna e dall'ordine delle estrazioni segue direttamente dalla 3) che:

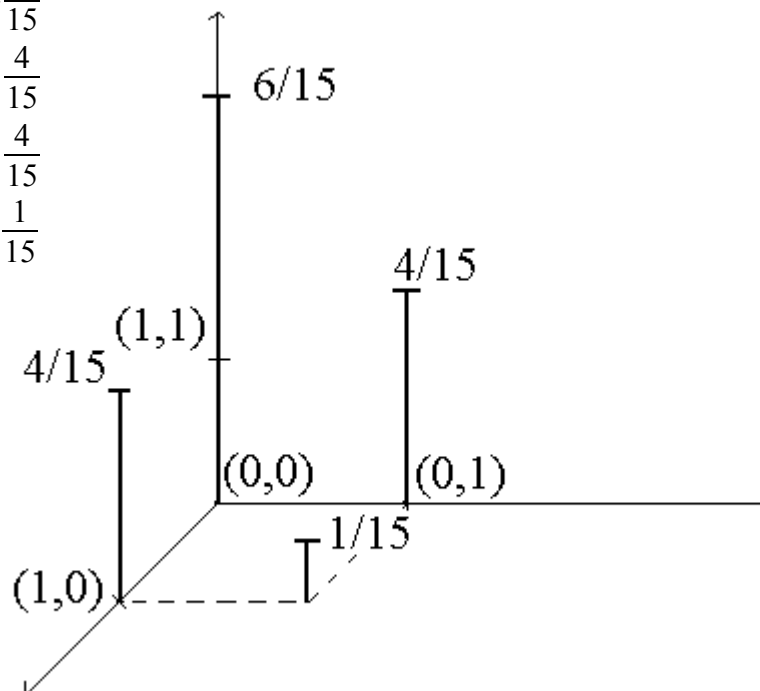
$$f(0,0) = f(0) \cdot f(0|0) = \frac{4}{6} \cdot \frac{3}{5} = \frac{6}{15}$$

$$f(0,1) = f(0) \cdot f(1|0) = \frac{4}{6} \cdot \frac{2}{5} = \frac{4}{15}$$

$$f(1,0) = f(1) \cdot f(0|1) = \frac{2}{6} \cdot \frac{4}{5} = \frac{4}{15}$$

$$f(1,1) = f(1) \cdot f(1|1) = \frac{2}{6} \cdot \frac{1}{5} = \frac{1}{15}$$

I valori della  $f(x, y)$  sono riportati nel grafico a segmenti:



Per illustrare come ottenere una distribuzione condizionata della distribuzione congiunta assumiamo ora che siano noti solo i valori della  $f(x, y)$  appena calcolati. Così la sola informazione disponibile è quella data in figura. La funzione di densità marginale della variabile  $X$  si può ottenere applicando la 4) da cui si ha che  $f(x) = \sum_{y=0}^1 f(x, y)$  e quindi:

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

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

Se i quattro punti del piano  $x, y$  si pensano come punti di massa della probabilità, la cui massa totale è 1, allora la distribuzione marginale della variabile  $X$  rappresenta la distribuzione della massa di probabilità lungo l'asse  $x$  dopo che i punti di massa di probabilità del piano  $x, y$  sono stati proiettati perpendicolarmente sull'asse  $x$ .

La funzione di densità condizionata di  $Y$  per un fissato valore di  $x$  si può ottenere applicando la formula 6) ed usando i risultati appena ottenuti. Così se alla variabile  $X$  viene assegnato il valore  $x=1$  si ha che  $f(y|1) = \frac{f(1, y)}{f(1)}$  da cui si ricava:

$$f(0|1) = \frac{f(1, 0)}{f(1)} = \frac{4/15}{1/3} = \frac{4}{5}$$

$$f(1|1) = \frac{f(1, 1)}{f(1)} = \frac{1/15}{1/3} = \frac{1}{5}$$

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

Esempio: come secondo esempio in cui la funzione di densità congiunta è data direttamente consideriamo la funzione di densità definita come  $f(x, y) = \frac{1}{27}(x+2y)$ , dove  $x$  e  $y$  possono assumere solo i valori interi  $x=y=0,1,2$ . Lo spazio campione con le sue probabilità calcolate tramite questa formula è mostrato in figura 4 vista ieri. Dalla formula 4) la funzione di densità marginale della variabile  $X$  è data dalla:

$$f(x) = \sum_{y=0}^2 \frac{1}{27}(x+2y) = \frac{1}{27}[x+(x+2)+(x+4)] = \frac{3(x+6)}{27} = \frac{1}{9}(x+2)$$

In modo simile sommando su tutti i possibili valori della variabile  $X$  si ha che:

$$g(y) = \sum_{x=0}^2 \frac{1}{27}(x+2y) = \frac{1}{27}[2y+(1+2y)+(2+2y)] = \frac{3(2y+1)}{27} = \frac{1}{9}(2y+1)$$

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

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

Se ad  $X$  si assegna per esempio il valore casuale 2 si ha che  $f(y|2) = \frac{(2+2y)}{12} = \frac{1}{6}(y+1)$ .

Questa funzione sarebbe utile se si volessero calcolare le probabilità per più valori della variabile  $Y$  quando si sa che la variabile  $X$  ha il valore 2. Si può facilmente verificare che questa è una funzione di densità di probabilità, infatti sommando le tre probabilità ottenute da questa formula ponendo  $y=0, 1, 2$  si può verificare che la somma è uguale a 1:

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

## Variabili Casuali Continue

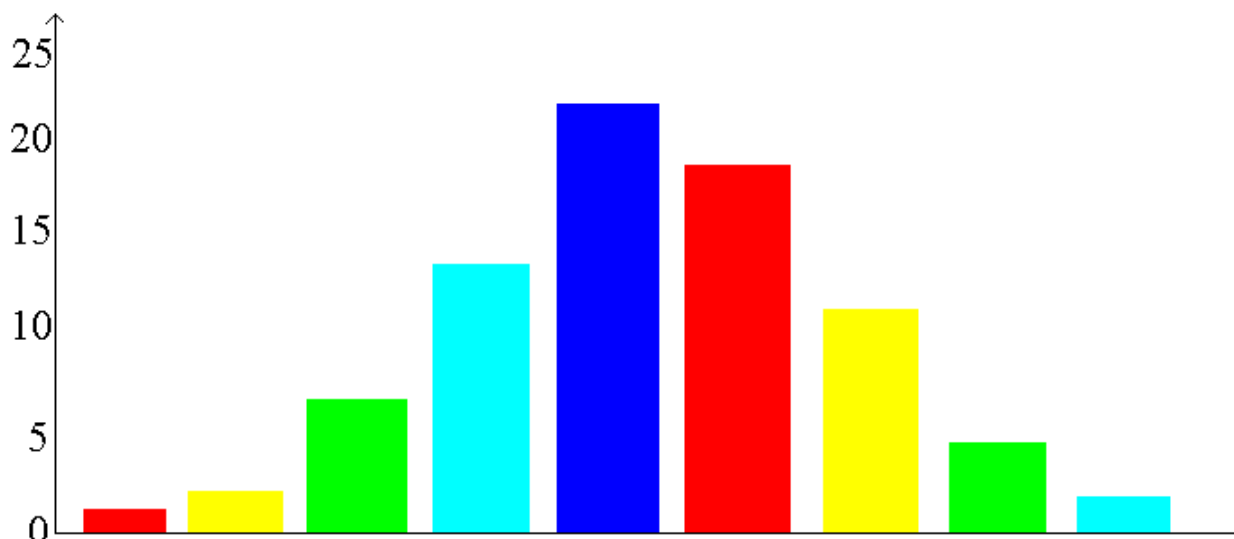
Una variabile casuale continua è una variabile definita su di uno spazio campione continuo; variabili come il peso, la temperatura, la velocità, etc. sono variabili CONTINUE. Si potrebbe discutere sul fatto che in realtà tutte queste variabili sono discrete perché qualunque strumento di misura venga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

## Funzione di Densità di Probabilità

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

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

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



Un istogramma è un grafico come quello illustrato in fig. 1 in cui si usano aree di rettangoli per rappresentare le frequenze osservate, in particolare le frequenze relative, così l'area del rettangolo centrato ad esempio in 0,234 dovrebbe essere uguale alla frequenza relativa 0,18. Tuttavia spesso in pratica si usa scegliere una conveniente unità di misura sull'asse  $y$  con il risultato che le aree dei rettangoli possono essere soltanto proporzionali anziché uguali alle corrispondenti frequenze relative. L'istogramma di fig. 1 è stato costruito con una siffatta conveniente scelta di unità di misura, quindi come si può osservare le aree sono soltanto proporzionali alle frequenze relative.

Se l'istogramma si deve costruire in modo che le aree siano uguali alle frequenze relative, allora l'area totale dell'istogramma deve essere uguale a 1, perché la somma delle frequenze relative deve essere uguale a 1. Se  $h$  indica la distanza fra valori di " $x$ " consecutivi, l'altezza del rettangolo centrato in  $x_i$  sarà uguale a  $\frac{f_i}{N \cdot h}$ , dove  $f_i$  è la frequenza assoluta di  $x_i$  ed  $N$  è il numero totale di osservazioni. Questo risultato è ovvio se si pensa che questa ordinata moltiplicata per la base  $h$  (ottenendo l'area del rettangolo) deve essere uguale alla frequenza relativa  $\frac{f_i}{N}$ .

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

valere all'aumentare del numero di esecuzioni dell'esperimento, l'area sottesa dalla curva fra due qualsiasi valori di  $X$  dovrebbe essere uguale alla frequenza relativa con la quale ci si aspetterebbe che il valore osservato di  $X$  cadesse dall'intervallo determinato da quei valori. La funzione  $f(x)$  il cui grafico è concepito come forma limite dell'istogramma viene presa come modello matematico per la variabile casuale continua  $X$  e viene chiamata la "funzione di densità di probabilità" della variabile. Poiché la frequenza relativa, nel caso di un'istogramma, è sostituita dalla probabilità nel caso di un modello matematico la definizione di funzione di densità di probabilità di una variabile continua si può stabilire nella formula seguente (**definizione**): una funzione di densità di probabilità di una variabile casuale continua  $X$  è una funzione  $f$  che possiede le seguenti proprietà:

- 1)  $f(x) \geq 0$
- 2)  $\int_{-\infty}^{+\infty} f(x) dx = 1$  che si può leggere anche come  $P\{-\infty < X < +\infty\}$
- 3)  $\int_a^b f(x) dx = P\{a < X < b\}$  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  $X$  assumerà un qualche valore reale quando viene fatta una sua osservazione. Nel calcolare poi le probabilità di una variabile continua si richiede soltanto che la variabile giaccia in qualche intervallo. Come risultato si ha che le probabilità delle variabili continue sono sempre date da integrali, mentre quelle delle discrete da somme. Se il dominio della variabile  $X$  non è l'intera retta reale si fa l'ipotesi che  $f(x)$  sia definita nulla per i valori esterni allo specificato dominio della variabile.

Esempio: come esempio consideriamo la possibilità di usare la funzione  $f(x) = k e^{-x}$ , dove  $k$  è una costante, come funzione di densità della variabile  $X$ . Per la prima proprietà è chiaro che  $k$  deve essere positiva. Poiché l'integrale  $\int_{-\infty}^{+\infty} e^{-x} dx = [-e^{-x}]_{-\infty}^{+\infty} = 0 + e^{+\infty}$ , segue che la variabile  $X$  deve essere limitata, quindi facciamo l'ipotesi che  $X$  possa assumere valori soltanto non negativi, cioè  $x \geq 0$ . Allora  $f(x)$  sarà definita nulla per valori negativi e data dalla formula suddetta per valori non negativi. Per la seconda proprietà segue allora che  $k$  deve essere uguale a 1 perché  $\int_0^{+\infty} e^{-x} dx = [-e^{-x}]_0^{+\infty} = 0 + 1 = 1$ .

Quindi può essere una funzione di densità di probabilità della variabile  $X$  la funzione:

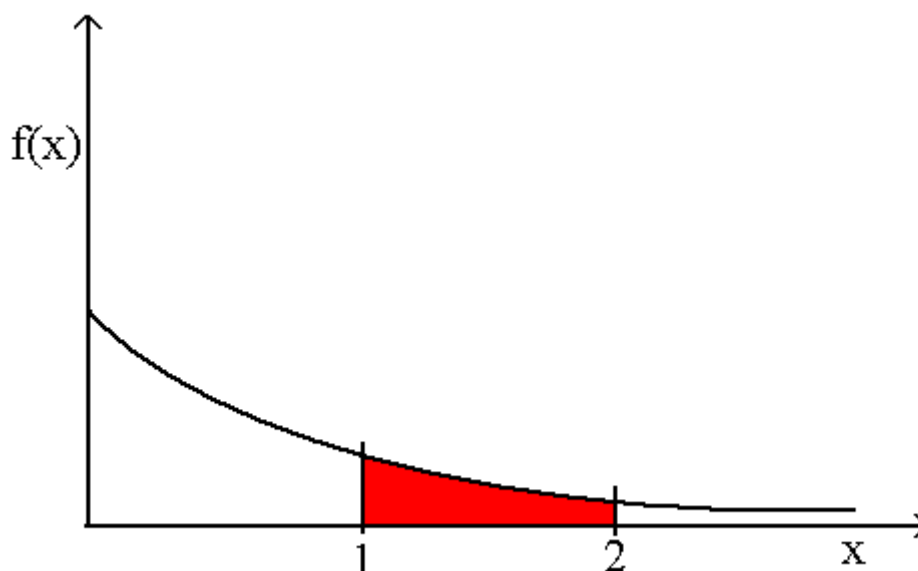
- $f(x) = e^{-x}$  per  $x \geq 0$
- $f(x) = 0$  per  $x < 0$

Il calcolo ad esempio della probabilità che  $X$  sia compreso tra 1 e 2 diventa allora:

$$P\{1 < X < 2\} = \int_1^2 e^{-x} dx = [-e^{-x}]_1^2 = -e^{-2} + e^{-1} = 0,23$$

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





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

## Funzione di Distribuzione o di Distribuzione Cumulativa o di Ripartizione:

La funzione in oggetto per la variabile continua  $X$  è data dalla (**definizione**):

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

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

- $F(x) = \int_0^x e^{-t} dt = [-e^{-t}]_0^x = 1 - e^{-x}$  per  $x \geq 0$
- $F(x) = \int_0^x e^{-t} dt = [-e^{-t}]_0^x = 0$  per  $x < 0$

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

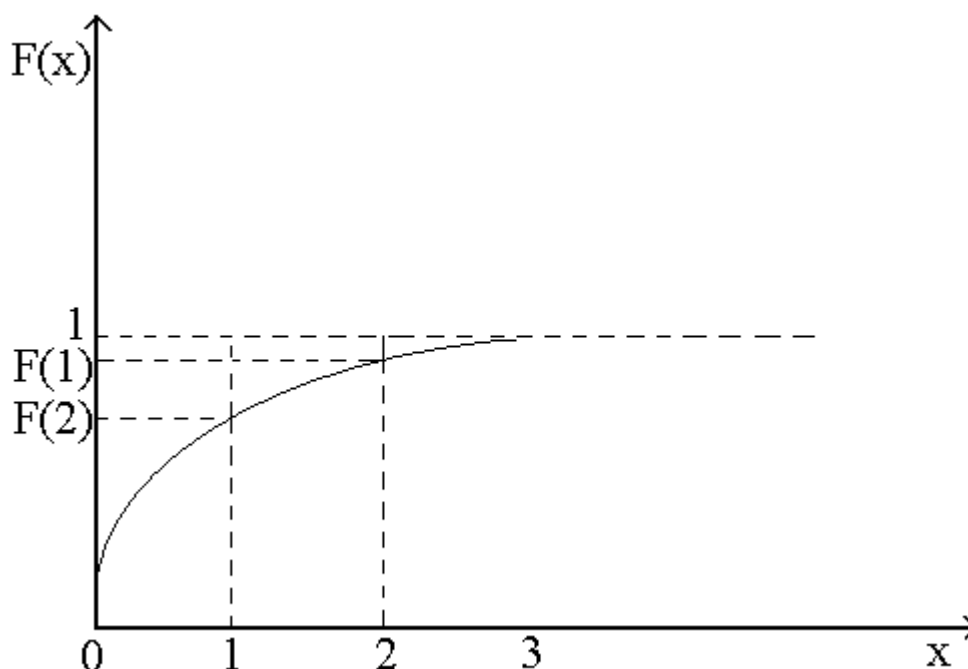


Figura 1: Errore nell'immagine,  $F(1)$  e  $F(2)$  vanno invertiti

Si fa notare che la probabilità che  $P\{1 < X < 2\}$  è ora data da  $F(1) - F(2)$ , infatti:

$$F(1) - F(2) = (1 - e^{-2}) - (1 - e^{-1}) = -e^{-2} + e^{-1} \approx 0,2$$

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

cioè,  $\frac{dF(x)}{dx} = f(x)$ , purché  $f(x)$  sia una funzione continua. Considerando l'effetto

appena trattato si può infatti verificare che derivando la funzione di distribuzione si ottiene la corrispondente funzione di densità, infatti:

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

## Funzione di Densità Congiunta Continua

Una funzione di densità di due o più variabili casuali continue è la naturale generalizzazione di una funzione di densità di una variabile, così una funzione di densità di due variabili casuali continue  $X, Y$  si indica con  $f(x, y)$  e come è noto è rappresentata geometricamente da una superficie nello spazio a tre dimensioni, cioè dalla superficie  $z = f(x, y)$ , proprio come una funzione di densità  $f(x)$  di una variabile è rappresentata

da una curva nel piano  $xy$ , cioè dalla curva  $y=f(x)$ . Se gli integrali della  $f(x,y)$  devono fornire probabilità è necessario che il volume totale sotteso da questa superficie sia uguale a 1 e che il volume sotteso da questa superficie che giace sopra una regione  $R$  nel piano  $xy$  dia la probabilità che le variabili casuali  $X$  e  $Y$  assumano valori che corrispondano a un punto interno a questa regione. Queste proprietà essenziali per una funzione di densità di due variabili si possono formalizzare come segue (**definizione**): una funzione di densità congiunta di due variabili casuali continue  $X$  e  $Y$  è una funzione  $f$  che possiede le seguenti proprietà:

- 1)  $f(x,y) \geq 0$
- 2)  $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) dx dy = 1$
- 3)  $\iint_R f(x,y) dx dy = P\{X,Y \in R\} \quad \forall R$

Si fa l'ipotesi che la regione  $R$  sia tale che l'integrale della  $f(x,y)$  su quella regione esista. Molto spesso la regione  $R$  sarà un rettangolo del tipo  $a < x < b, c < y < d$ , nel qual caso la terza condizione diventa:

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

**Esempio:** come esempio consideriamo la funzione  $f(x,y) = e^{-(x+y)}$  che è una generalizzazione in due dimensioni dell'esempio precedente. La  $f(x,y)$  si definisce essere nulla per valori negativi di  $x$  o  $y$ , ovvero:

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

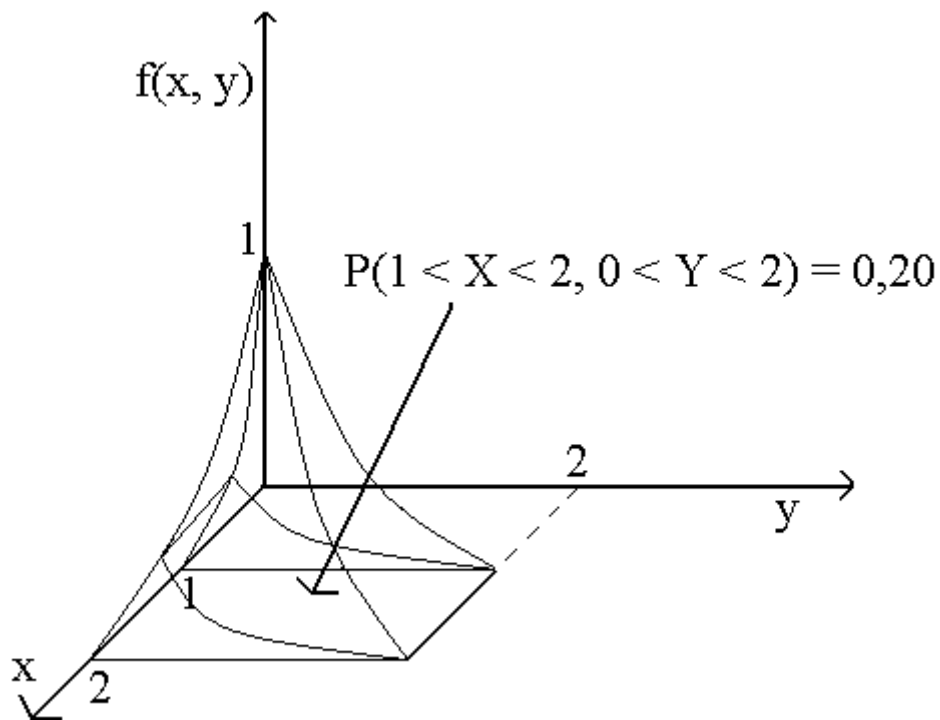
Si può osservare così che le prime due proprietà sono soddisfatte, infatti  $f(x,y) \geq 0$ :

$$\begin{aligned} \int_0^{+\infty} \int_0^{+\infty} e^{-(x+y)} dx dy &\rightarrow \int_0^{+\infty} \int_0^{+\infty} e^{-x} e^{-y} dx dy \rightarrow \int_0^{+\infty} e^{-x} \left( \int_0^{+\infty} e^{-y} dy \right) dx \rightarrow \int_0^{+\infty} e^{-x} [-e^{-y}]_0^{+\infty} dx \rightarrow \\ &\rightarrow \int_0^{+\infty} e^{-x} (0+1) dx = 1 \end{aligned}$$

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

$$\begin{aligned} P\{1 < X < 2, 0 < Y < 2\} &= \int_1^2 \int_0^2 e^{-(x+y)} dx dy \rightarrow \int_1^2 \int_0^2 e^{-x} e^{-y} dx dy \rightarrow \int_1^2 e^{-x} \left( \int_0^2 e^{-y} dy \right) dx \rightarrow \\ &\rightarrow \int_1^2 e^{-x} [-e^{-y}]_0^2 dx \rightarrow \int_1^2 e^{-x} (-e^{-2} + 1) dx \rightarrow (1 - e^{-2}) \int_1^2 e^{-x} dx \rightarrow \\ &\rightarrow (1 - e^{-2}) [-e^{-x}]_1^2 \rightarrow (1 - e^{-2}) (-e^{-2} + e^{-1}) \approx 0,20 \end{aligned}$$

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



Due o più variabili casuali continue che sono senza rapporti, in senso probabilistico, si dicono indipendenti proprio come nel caso delle variabili discrete. La definizione di indipendenza data nel caso delle variabili discrete si estende anche al caso delle variabili continue. Ci occuperemo fra poco di alcune distribuzioni di probabilità sia di variabili discrete che di variabili continue che nella realtà si sono dimostrate modelli particolarmente utili. In ciascun caso la distribuzione sarà specificata presentando la funzione di densità di probabilità. In molti problemi 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 distribuzione, come pure delle loro funzioni di densità.

## Variabili Discrete

La maggior parte delle variabili discrete che ricorrono negli esperimenti di tipo ripetitivo sono di tipo "conteggio", come ad esempio il numero di incidenti che un automobilista ha in un anno, il numero di insetti che sopravvivono ad una irruzione, 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 in modo da comprendere altri tipi di variabili casuali discrete.

## Valore Atteso

Prima di considerare specifiche funzioni di densità di variabili discrete ci occuperemo dei momenti di una distribuzione di probabilità, così da essere poi in grado di calcolare i momenti di quelle particolari distribuzioni. Ancora prima però introdurremo e definiremo un concetto più generale e cioè quello di valore atteso, poiché esso comprende i momenti come casi particolari ma è anche uno strumento molto utile per studiare altre proprietà di una distribuzione. Come esempio per illustrare il concetto generale di "valore atteso" consideriamo un gioco in cui si lanciano 3 monete e si vince un dollaro per ogni testa che esce. Se  $X$  indica la somma vinta allora essa deve assumere uno dei valori 0, 1, 2, 3 cui corrispondono le probabilità  $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$ . Queste probabilità risultano evidenti se si considera lo spazio campione dell'esperimento del lancio di tre monete. Ci si dovrebbe perciò aspettare di vincere 0 dollari per  $\frac{1}{8}$  di volte, 1 o 2 dollari per  $\frac{3}{8}$  di volte e 3 dollari per  $\frac{1}{8}$  di volte se il gioco fosse fatto un gran numero di volte. Ci si dovrebbe perciò aspettare di vincere in media la somma seguente:  $0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1,50$ . Questa somma, cioè 1,50\$, è ciò che rappresenta comunemente la somma che ci si aspetta di vincere. Supponiamo che la variabile casuale discreta  $X$  debba assumere uno dei valori  $x_1, x_2, \dots, x_k$  e che le probabilità associate a quei valori siano  $f(x_1), f(x_2), \dots, f(x_k)$ . Allora il valore atteso della variabile  $X$  è definito dalla formula:

$$1) \quad E[X] = \sum_{i=1}^k x_i f(x_i) \quad \text{"E" sta per "Expected Value" (valore atteso)}$$

Nell'esempio appena visto la  $X$ , cioè la somma vinta, può assumere i valori  $x_1=0, x_2=1, x_3=2, x_4=3$  e le probabilità corrispondenti sono  $f(x_1)=\frac{1}{8}, f(x_2)=\frac{3}{8}, f(x_3)=\frac{3}{8}, f(x_4)=\frac{1}{8}$ . Ora supponiamo che il gioco considerato venga modificato in modo da vincere la somma  $g(x_i)$  invece di  $x_i$  quando si ottiene il valore  $x_i$ . Ad esempio, per  $g(x)$  si potrebbe scegliere la funzione  $g(x)=x^2$ , il che significa che si vincerebbe in dollari il quadrato di teste uscite, allora il valore atteso nel gioco così modificato sarebbe dato dalla formula:

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

Nelle formule 1) e 2) si assume che ci sia soltanto un numero finito di possibili valori della variabile casuale  $X$ . Per limare questa limitazione si introduce la seguente definizione più generale applicabile a qualsiasi variabile casuale discreta (**definizione**): il valore atteso della funzione  $g(X)$  della variabile casuale discreta  $X$ , la cui funzione di densità è  $f$ , è dato da

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

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 casuale  $X$  viene chiamato di solito la media della variabile casuale o 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 al secondo membro della 3) sarà naturalmente  $k$  anziché  $\infty$ . Quando la variabile  $X$  possiede un numero infinito di possibili valori con probabilità positive è necessario assumere che la serie infinita al

secondo membro della 3) converga.

## Momenti

I momenti della distribuzione di una variabile casuale discreta  $X$  si possono definire in termini di valori attesi come segue (**definizione**): definiamo il momento d'ordine  $k$  dall'origine della distribuzione della variabile casuale discreta  $X$ , la cui funzione di densità è  $f$ , come 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, si 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  $X$ . Il momento del primo ordine  $\mu'_1 = E[X]$  ricorre così spesso che ad esso si assegna il simbolo particolare  $\mu$ . Poiché anche i momenti dalla media sono usati ampiamente, essi pure necessitano di essere definiti (**definizione**): il momento d'ordine  $k$  dalla media della distribuzione della variabile casuale discreta  $X$ , la cui definizione di 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 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 due momenti, perché essi spesso sono sufficienti a descrivere due utili proprietà della distribuzione, cioè *dove la distribuzione è centrata e in che grado la distribuzione è concentrata attorno a questo centro*. Il momento del primo 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 si usa molto spesso, a questo proposito esso si indica col simbolo  $\mu_2 \rightarrow \sigma^2$  e viene chiamato **varianza della distribuzione**. La radice quadrata positiva della varianza, cioè  $\sigma$ , viene chiamata **scarto quadratico medio** o **deviazione standard** della distribuzione. Esso si impiega al posto della varianza quando si desidera una misura della concentrazione nelle stesse unità di misura della variabile casuale.

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

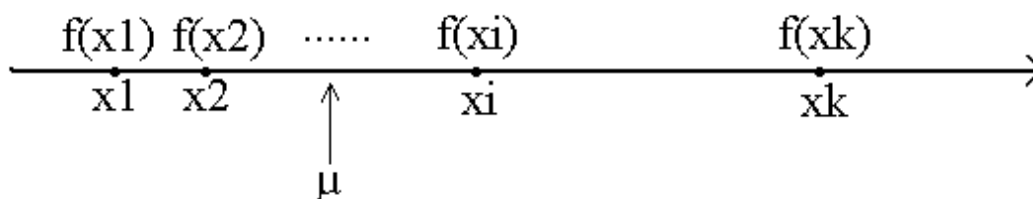


Fig. 1

Mentre invece la maggior parte della distribuzione è concentrata attorno alla media stessa. Si può notare quindi che, prendendo una massa di probabilità molto piccola, sufficientemente lontana 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 radicato nell'esempio di figura 3 la media e la varianza si sono dimostrate quantità molto utili nel trattare le distribuzioni più comunemente utilizzate. Per calcolare la varianza è spesso più conveniente calcolare i primi due momenti dall'origine e poi calcolare la varianza tramite essi piuttosto che calcolarla direttamente usando la formula che la definisce. Ricordando la formula che la definisce ciò si realizza come segue:

$$\mu^2 = \sum_{i=1}^{\infty} ((x_i - \mu)^2) f(x_i) = \sum_{i=1}^{\infty} (x_i^2) f(x_i) - 2\mu \sum_{i=1}^{\infty} x_i f(x_i) + \mu^2 \sum_{i=1}^{\infty} f(x_i)$$

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

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

## 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 metodiche ora introdurremo. Esso implica quella che è nota come *funzione generatrice dei momenti*. Come indica il nome stesso questa funzione è una funzione che genera i momenti ed è definita come segue:

*Def:* la funzione generatrice dei momenti di una variabile casuale discreta X, la cui densità è f è data da:

$$M_x(\sigma) = E[e^{x\sigma}] = \sum_{i=1}^{\infty} (e^{x\sigma}) f(x_i) \quad (1)$$

Questa serie è una funzione del parametro  $\sigma$  soltanto ma si è messo l'indice M( $\sigma$ ) per indicare quale variabile si sta considerando. Il parametro  $\sigma$  non ha qui nessun significato reale, esso si introduce semplicemente per aiutare a determinare i momenti. Per mostrare come M( $\sigma$ ) genera i momenti facciamo l'ipotesi che la funzione f sia una funzione di densità per cui la serie nella (1) converga. Sviluppiamo ora e alla  $x_i$  in serie di potenze e sommiamo termine a termine.

$$e^{x\sigma} = 1 + \frac{x\sigma}{1!} + \frac{(x\sigma)^2}{2!} + \frac{(x\sigma)^3}{3!} + \dots$$

Per ricavare la formula richiesta determiniamo dapprima la probabilità di ottenere X successi consecutivi seguiti da n - x successi. Chiaramente questi eventi sono indipendenti, perciò ricordando una formula vista a suo tempo e valida per gli eventi indipendenti e cioè:

$$P\{A_1 \cap A_2\} = P\{A_1\} \cdot P\{A_2\}$$

questa probabilità è data da :

$$((p \cdot p \cdot \dots \cdot p) \text{ "x volte"}) \cdot ((q \cdot q \cdot \dots \cdot q) \text{ "n - x volte"}) = (p^x) \cdot (q^{n-x});$$

*Osservazione:* la probabilità di ottenere x successi ed n - x successi 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 possibile (ovvero il numero di permutazioni) con n

oggetti di cui  $x$  sono fra loro uguali, cioè le  $p$ , come pure i rimanenti  $(n - x)$  che sono le  $q$ . Ma come è noto il numero di tali permutazioni è uguale a :

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

Ora ricordando la formula  $P\{A1 \cap A2\} = P\{A1\} + P\{A2\}$ , 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  $(p \text{ alla } x)(q \text{ alla } (n - x))$  tante volte quanti sono gli ordini diversi in cui il risultato richiesto può verificarsi.

Poiché la (3) dà il numero di tale ordine, la probabilità di ottenere  $x$  successi in qualche ordine si ottiene perciò moltiplicando:

$(p \text{ alla } x)(q \text{ alla } (n - x))$  per la quantità  $n!/(x!(n-x)!)$ .

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

$$(4) \quad f(x) = \frac{n!}{x!(n-x)!} (p \text{ alla } x)(q \text{ alla } (n - x)).$$

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

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

Il termine generale in questo sviluppo che contiene  $p$  alla  $x$ , dove  $x \in Z$  è dato da:  $(\frac{n!}{x!(n-x)!} (q \text{ alla } n-x)p^x)$

$$\frac{n!}{x!(n-x)!} (p \text{ alla } x)(q \text{ alla } n-x).$$

La validità di usare il modello binomiale in questo esempio non è così ovvia come lo è invece per l'esempio precedente. Infatti la formula binomiale è stata ricavata sulla base di esecuzioni dell'esperimento indipendenti fra loro con  $p$  costante da esecuzione ad esecuzione. Ma se una persona tira colpi ripetuti allo stesso bersaglio ci si potrebbe aspettare che le sue probabilità di colpirlo aumentassero un po' con la pratica, se invece venisse usata ogni volta da una persona diversa  $p$  indubbiamente cambierebbe da prova a prova, quindi al momento di interpretare una probabilità risultante come quella uguale a 0,608 appena ricavata è importante prendere in considerazione le possibili deviazioni nelle ipotesi fondamentali.

## Momenti Binomiali

Calcoliamo ora i primi due momenti dell'origine della 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 ieri si può scrivere:

$$\begin{aligned} \mu &= \sum_{x=0}^n x \frac{n!}{x!(n-x)!} (p \text{ alla } x)(q \text{ alla } (n-x)) \\ &= \sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} ((p \text{ alla } x)(q \text{ alla } (n-x))) \end{aligned}$$

Ora portando fuori dal segno di sommatoria  $np$  si può scrivere:

$$\mu = np \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y$$

dove la quantità che viene sommata a secondo membro rappresenta la proprietà binomiale di  $y$  successi in  $n - 1$  esecuzioni dell'esperimento. Poiché la somma viene fatta su tutti i possibili valori di  $y$  essa deve essere uguale a 1, quindi  $\mu = np$ .

Il momento del secondo ordine dall'origine di una variabile binomiale si calcola in modo analogo usando l'identità seguente:

$$x \text{ al quadrato} = x(x-1) + x.$$



Dalla definizione si ha che:

$$\mu^2 = \sum_{x=0}^n x^2 \binom{n}{x} p^x q^{n-x} = n p q + n p^2 = n p (q + p) = n p$$

I momenti richiesti si possono ottenere applicando una formula già nota:

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

Derivando la (3) due volte rispetto a  $\theta$  si ottiene:

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

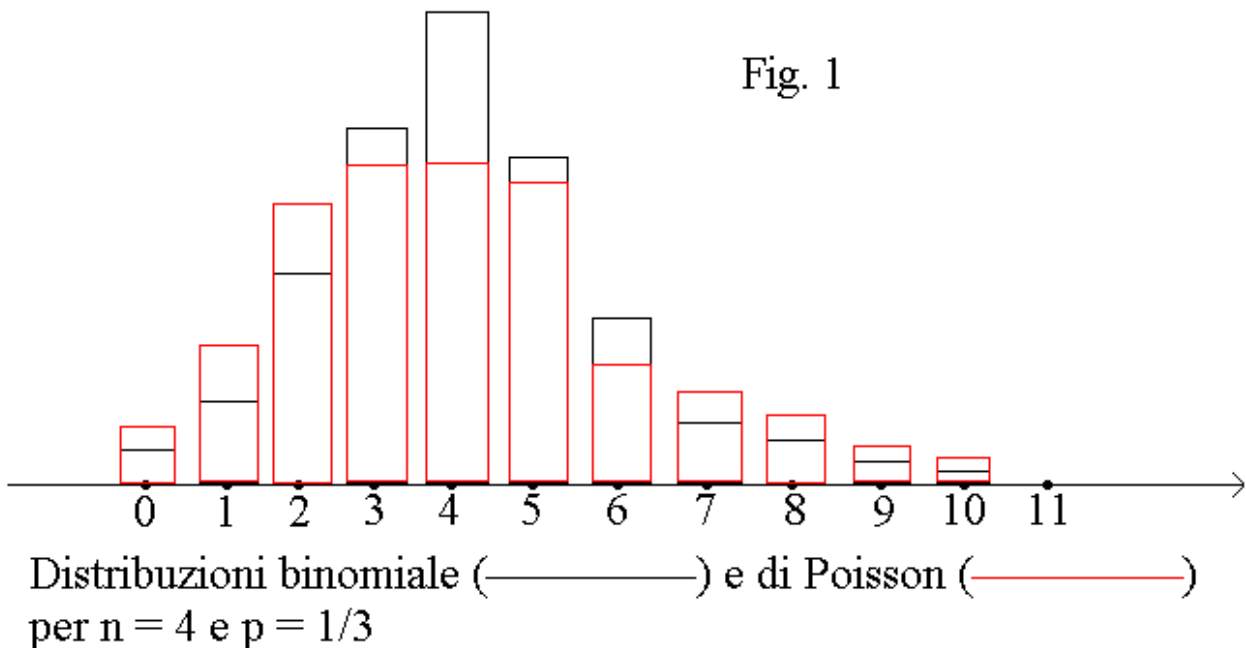
$$M''_x(\theta) = (npe)^{n-1}((q + pe)^{n-1} + (n-1)(q + pe)^{n-2}(pe)) = (npe)^{n-1}(q + pe)^{n-2}(q + npe)$$

I valori di queste derivate calcolate in  $\theta = 0$  sono rispettivamente i valori di  $\mu$  e  $\mu^2$ .

Infatti  $\mu = [M'_x(\theta)]_{\theta=0} = n p (q + p) = n p$ ; e

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

Se  $p$  viene sostituito da  $(1 - p)$  si può osservare che si ritrova per  $\mu^2$  lo stesso valore espresso dalla (2). Si può osservare che nel caso di una variabile binomiale i momenti sono più facili da ottenere indirettamente tramite la funzione generatrice dei momenti che direttamente dalla definizione.



## ***Distribuzione di Poisson***

Se il numero di esecuzione dell'esperimento  $n$  è grande i calcoli implicati nell'usare la funzione di densità binomiale diventano piuttosto lunghi, perciò sarebbe molto utile poter disporre di una conveniente approssimazione della distribuzione binomiale.

1) Quando  $p$  è molto piccolo;

2) quando non si è in questo caso;

L'approssimazione che si applica quando  $p$  è molto piccolo è nota come la funzione di densità di Poisson ed è definita come segue:

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

dove il parametro  $\mu$  è la media della distribuzione. Sebbene la distribuzione di Poisson venga introdotta qui come un'approssimazione della distribuzione binomiale.

*Approssimazione di Poisson della distribuzione binomiale:* Per verificare che la distribuzione di Poisson è una buona approssimazione di quella binomiale per  $n$  molto grande e  $p$  molto piccolo si considera ciò che accade alla funzione di densità binomiale quando  $n \Rightarrow \text{inf.}$  e  $p \Rightarrow 0$  in modo tale che la media  $\mu = n \cdot p$  rimanga fissa. A tale scopo riscriviamo la funzione di densità binomiale come segue:

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

Moltiplicando ora il numeratore e il denominatore per  $(n^x)$  si ha che:

$$(2) \quad f(x) = (n \cdot (n-1) \cdot \dots \cdot (n-x+1) / (n^x \cdot x!)) \cdot (n^x \cdot p^x) \cdot ((1-p)^{n-x})$$

Riscriviamo ora  $(1-p)^{n-x}$  come segue:

$$(1-p)^{n-x} = [(1-p)^{-1/p}]^{-n \cdot p} = [(1-p)^{-1/p}]^{-\mu}$$

Ricordando ora la definizione del numero di Nepero  $e$ , cioè:

$$\lim_{z \Rightarrow 0} (1+z)^{1/z} = e$$

Ponendo  $z = -p$  si ha che:

$$\lim_{p \Rightarrow 0} [(1-p)^{-1/p}]^{-\mu} = e^{-\mu}$$

Inoltre il:

$$\lim_{n \Rightarrow \text{inf.}} ((1-1/n) \cdot (1-2/n) \cdot \dots \cdot (1-(x-1)/n)) / ((1-p)^x) = 1$$

perché  $p \Rightarrow 0$  quando  $n \Rightarrow \text{inf.}$  in modo tale che la media  $\mu = n \cdot p$  rimanga fissa. Applicando questi ultimi due risultati all'ultimo membro della (2) si ottiene che:

$$\lim_{n \Rightarrow \text{inf.}} f(x) = (e^{-\mu})^n (\mu^x) / x!$$

Questo risultato si può esprimere sotto forma di teorema il cui enunciato è il seguente...

**Teorema:** se la probabilità  $P$  di successo in una singola esecuzione tende a 0 mentre in esecuzioni  $n \Rightarrow \text{inf.}$  in modo tale che la media  $\mu = n \cdot p$  rimanga fissa allora la distribuzione binomiale tenderà alla distribuzione di Poisson con media  $\mu$ . La figura 1 è stata costruita per indicare con quale rapidità la distribuzione binomiale tende alla distribuzione di Poisson. Le linee rosse rappresentano la distribuzione di Poisson fissa per  $n = 4$  e le linee nere la distribuzione binomiale per  $p = 1/3$ . Sembra che dall'esame di questi grafici che l'approssimazione di Poisson dovrebbe essere sufficientemente accurata per la maggior parte delle applicazioni se  $n \geq 100$  e  $p \leq 0,05$ . Poiché la distribuzione di Poisson è stata ottenuta dalla distribuzione binomiale, mantenendo fissa la media e permettendo ad  $n$  di diventare  $\text{inf.}$  segue che i momenti della distribuzione di Poisson si possono ottenere dai corrispondenti momenti della distribuzione binomiale calcolando i valori limite di questi ultimi, così la media della distribuzione di Poisson deve essere ovviamente  $\mu = n \cdot p$ , mentre la varianza deve essere il valore limite  $n \cdot p \cdot q$ .

Ma quando  $\mu = n \cdot p$  è fissato  $p \Rightarrow 0$  quando  $n \Rightarrow \text{inf.}$ , perciò il

$$\lim_{n \Rightarrow \text{inf.}} n \cdot p \cdot q = \lim_{n \Rightarrow \text{inf.}} \mu \cdot q =$$

$$\lim_{p \Rightarrow 0} \mu \cdot (1-p) = \mu$$

Questo dà 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 rapporti con la distribuzione binomiale. Per esempio si è trovato che è un esempio 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, etc.

Questi sono tutti problemi 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 quando le precedenti ipotesi sono state precisate matematicamente, che il numero di eventi, in qualsiasi intervallo di tempo di dimensione fissata possederà una distribuzione di Poisson. La stessa derivazione si può applicare quando gli intervalli di tempo sono sostituiti da intervalli spaziali. Un intervallo spaziale può essere ad una, due o tre dimensioni, così il numero di screpolature in una data lunghezza di filo metallico o di una data area di stoffa. Un esperimento in cui le osservazioni si succedono in intervalli successivi di tempo o di spazio, e per cui le ipotesi precedenti sono soddisfatte così che gli eventi possiedono una distribuzione di Poisson, viene chiamato *processo di Poisson*. Così la distribuzione di Poisson è una distribuzione utile per trattare problemi riguardati i processi di Poisson e non è giustificata, quindi, soltanto come approssimazione della distribuzione binomiale.

*Esercitazione:* Come esempio d'uso della distribuzione di Poisson come approssimazione di quella binomiale, consideriamo il problema di calcolare la probabilità che si trovino al massimo 5 valvole difettose in una scatola di 200 valvole se l'esperienza mostra che il 2% di queste valvole sono difettose. In questo caso la media:

$$\mu = n \cdot p = 200 \cdot (0,02) = 4$$

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

$$P\{X \leq 5\} = \sum_{x=0}^5 \frac{(e^{-4}) \cdot (4^x)}{x!} =$$

$$(e^{-4}) \cdot (1 + 4 + 8 + (4^3)/6 + (4^4)/24 + (4^5)/120) = 0,785.$$

Lunghi calcoli, usando la funzione di densità binomiale, porterebbero al risultato che:  $P\{X \leq 5\} = 0,788$ .

Quindi l'approssimazione in questo caso è molto buona, il che significa che l'errore relativo è il seguente:

$$|\Delta P/P| = |(0,785 - 0,788)/0,788| = 0,003/0,788 = 3,8\%.$$

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

*Esempio:* come esempio d'uso della distribuzione di Poisson consideriamo il problema seguente. Supponiamo che l'esperienza abbia mostrato che il numero medio di chiamate telefoniche che arriva ad un quadro di controllo in un minuto è uguale a 5. Se il quadro di controllo può trattare al massimo 8 chiamate al minuto si vuole calcolare la probabilità che esso non sia capace di trattare tutte le chiamate che arrivano in un minuto. La probabilità richiesta si può ottenere calcolando la probabilità di ricevere al massimo 8 chiamate al minuto e poi sottraendo questa probabilità da 1. Usando  $\mu = 5$  nelle funzioni di densità di Poisson si ha che:

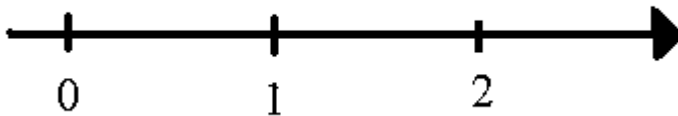
$$P\{x \leq 8\} = \sum_{x=0}^8 \frac{(e^{-5}) \cdot (5^x)}{x!} = 0,932$$

Di conseguenza la probabilità richiesta è data da:

$$P\{x > 8\} = 1 - 0,932 = 0,068$$

Per la natura della 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, per esempio: la soluzione del problema di calcolare la probabilità che si ricavano al massimo 6 chiamate 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:  $P\{x \leq 6\} = \sum_{x=0}^6 \frac{e^{-10} (10)^x}{x!} = 0,13$

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



0	0, 1, 2, 3, 4, 5, 6
1	0, 1, 2, 3, 4, 5
2	0, 1, 2, 3, 4
3	0, 1, 2, 3
4	0, 1, 2
5	0, 1
6	0

$$f(1, 3) = f(1) \cdot f(3)$$

$$f(3, 1) = f(3) \cdot f(1)$$

*si calcola la somma di tutte le probabilità delle combinazioni*

## Variabili Continue

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

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

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

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

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

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

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

*Def:* il valore atteso della funzione  $g(x)$  della variabile casuale continua  $X$  la cui densità è data da:

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

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  $a$  può essere uguale a  $-\infty$  e  $b$  può essere uguale a  $+\infty$ . Poiché il valore atteso  $E$  è stato progettato per fornire valori medi di variabili casuali sorge spontanea la domanda se il valore atteso di una funzione di una variabile casuale  $X$  per dire  $g(x)$  dia il valore sia realmente la media di quella funzione. Indichiamo con  $Y$  la variabile casuale  $g(x)$ ; allora conoscendo la funzione di densità  $f(x)$  è teoricamente possibile trovare la funzione di densità  $h(y)$  della variabile  $Y$ .

Il valore atteso di  $Y$  si può esprimere nella forma:

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

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

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

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

## Funzione Generatrice dei Momenti

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

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

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

$$M_x(\theta) = 1 + (\theta/1!) \cdot \mu'_1 + (\theta^2/2!) \cdot \mu'_2 + \dots$$

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

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

Questa è la forma generalizzata della funzione generatrice dei momenti. Consideriamo ora due interessanti proprietà delle funzioni generatrici dei momenti. Sia  $c$  una costante ed  $h(x)$  una funzione della variabile  $X$  per la quale la funzione generatrice dei momenti esiste. Allora nella (5)  $g(x)$  rappresenta una funzione arbitraria, si può scegliere  $g(x) = ch(x)$ , così facendo si ottiene:

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

Scegliendo ora  $g(x) = h(x) + c$  si ottiene:

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

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

**Proprietà:** Se  $c$  è una costante qualsiasi e  $g(x)$  è una funzione qualsiasi per cui la funzione generatrice dei momenti esiste allora:

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

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

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

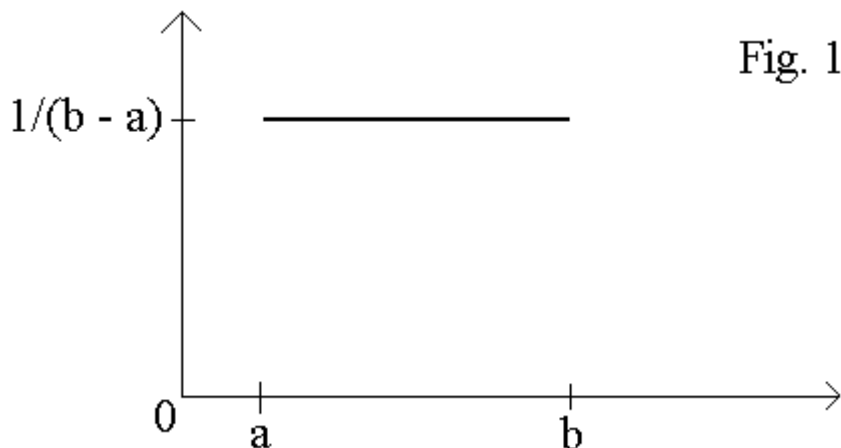
## Distribuzione Rettangolare o Uniforme

Forse la variabile casuale continua più semplice è quella la cui distribuzione è costante su un certo intervallo  $[a, b]$  ed è uguale a zero altrove. Questa definisce quella che è conosciuta come la distribuzione rettangolare o uniforme, e cioè:

$$(1) \quad f(x) = \frac{1}{(b-a)} \quad \text{per ogni } a \leq X \leq b$$

$$f(x) = 0 \quad \text{altrove}$$

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



Osservando questo sviluppo si vede che mancano le potenze dispari di  $\theta$  e quindi che i momenti dispari della variabile  $X$  della sua media devono essere nulli. Il coefficiente di  $(\theta$  alla seconda)/2! È il momento del secondo ordine della variabile  $X$  della sua media perciò (b alla seconda) =  $\mu^2 = (\sigma$  alla seconda) ovvero  $b = \sigma$ .

Per cui si è appena trovato che  $\sqrt{(2\pi)} \cdot c = 1$ , si ha che  $C = 1/\sigma \cdot \sqrt{(2\pi)}$

Di conseguenza la distribuzione normale definita dalla (2) si può scrivere

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

Questo risultato mostra che una distribuzione normale è completamente determinata specificando la sua media e il suo scarto quadratico medio ( $\sigma$ ); si fa notare che la sola differenza tra la (2) e la (5) è che i parametri  $a, b$  nella (2) sono stati ristretti a due soli parametri, a cui è stato attribuito un significato statistico. Una formula per  $M_x(\theta)$  espressa

in termini di parametri statistici sarà necessaria in seguito. Essa si può ottenere dalla (4) sostituendo (b alla seconda) con  $(\sigma$  alla seconda) ed usando la seconda delle sue proprietà della funzione generatrice dei momenti in forma generalizzata e cioè:

$$M(g(x) + c)(\theta) = (e \text{ alla } \theta c) * M_g(x)(\theta)$$

con  $g(x) = x$ ,  $c = -\mu$ , per cui si può scrivere che:

$$M(x - \mu)(\theta) = (e \text{ alla } -\mu\theta) * M_g(x)(\theta)$$

e quindi utilizzando al primo membro l'espressione data dalla (4) e risolvendo rispetto a  $M_x(\theta)$  si ottiene che:

$$(6) \quad M_x(\theta) = (e \text{ alla } \mu\theta + 1/2 * (\sigma \text{ alla seconda}) * (\theta \text{ alla seconda}))$$

Per interpretare lo scarto quadratico medio geometricamente consideriamo i punti di flesso di una curva normale. Calcolando le prime due derivate della funzione di densità normale espressa dalla (5) si ha:

$$f'(x) = (1/(a * \sigma * \sqrt{2\pi})) * (e \text{ alla } -1/2 * ((x-\mu)/\sigma) \text{ alla seconda}) * [-(x-\mu)/\sigma] * 1/\sigma =$$

$$= -(1/(\sigma \text{ alla seconda}) * (x-\mu) * f(x));$$

$$f''(x) = -(1/(\sigma \text{ alla seconda}) * f(x) - (1/(\sigma \text{ alla seconda}) * (x-\mu) * [-(1/(\sigma \text{ alla seconda}) * (x-\mu) * f(x))])$$

$$= -(1/(\sigma \text{ alla seconda}) * f(x) * [1 - ((x-\mu)/\sigma) \text{ alla seconda}]);$$

Da queste derivate è chiaro che c'è soltanto un punto di massimo che si trova in  $x = \mu$ .

Calcoliamo ora i momenti indirettamente tramite la funzione generatrice dei momenti. Inoltre, poiché in questo caso è più facile calcolare i momenti dalla media piuttosto che dall'origine, consideriamo il calcolo di  $M_x - \mu(\sigma)$ . Dalla forma generalizzata della funzione generatrice dei momenti con  $g(x) = x - \mu$  e tenuto conto della seconda proprietà delle funzioni generatrici dei momenti si può scrivere che:

$$M(x - \mu)(\sigma) = c * \int_{-\text{inf. } a}^{+\text{inf.}} (e \text{ alla } \theta * (x - \mu)) * (e \text{ alla } -1/2 * ((x - \mu)/b) \text{ alla seconda}) \text{ in } dx;$$

Ora ponendo  $z = (x - \mu)/b$ , allora  $dx = b * dz$  e quindi:

$$M(x - \mu)(\sigma) = b * c * \int_{-\text{inf. } a}^{+\text{inf.}} (e \text{ alla } (\theta * b * z - (z \text{ alla seconda})/2) \text{ in } dz);$$

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

$$\theta * b * z - (z \text{ alla seconda})/2 = -1/2 * ((z - \theta * b) \text{ alla seconda}) + 1/2 * (\theta \text{ alla seconda}) * (b \text{ alla seconda});$$

Per cui si ha che:

$$M(x-\mu)(\sigma) = b * c * (e \text{ alla } 1/2 * (\theta \text{ alla seconda}) * (b \text{ alla seconda}) * \int_{-\text{inf. } a}^{+\text{inf.}} (e \text{ alla } -1/2 * ((z - \theta * b) \text{ alla seconda}) \text{ in } dz);$$

Ponendo ora  $t = z - \theta * b$ :

$$M(x-\mu)(\sigma) = b * c * e \text{ alla } (1/2 * (\theta \text{ alla seconda}) * (b \text{ alla seconda})) * \int_{-\text{inf. } a}^{+\text{inf.}} (e \text{ alla } ((-t \text{ alla seconda})/z) \text{ in } dt);$$

L'integrale a secondo membro è un integrale noto il cui valore che si può trovare nelle

tavole degli integrali è uguale a  $\sqrt{2\pi}$  per cui si può scrivere:

$$(3) \quad M(x-\mu)(\sigma) = \sqrt{2\pi} \cdot b \cdot c;$$

Ora ricordando che  $M_x(\theta) = 1 + (\theta/1!) \cdot \mu'1 + ((\theta \text{ alla seconda})/2!) \cdot \mu'2 + \dots$

Segue che per qualsiasi funzione generatrice dei momenti  $M(0) = 1$  e quindi dalla (3) per  $\theta = 0$  si ha che  $\sqrt{2\pi} \cdot b \cdot c = 1$  e quindi si può scrivere:

$$(4) \quad M(x-\mu)(\sigma) = e \text{ alla } (\theta \text{ alla seconda}) \cdot (b \text{ alla seconda})$$

Ora sviluppando l'esponenziale a secondo membro in serie di potenze si ha che:

$$M(x-\mu)(\sigma) = 1 + ((b \text{ alla seconda}) \cdot (\theta \text{ alla seconda}))/2 + ((b \text{ alla quarta}) \cdot (\theta \text{ alla quarta}))/8 + \dots$$

$$\Rightarrow e \text{ alla } z = 1 + z/1! + (z \text{ alla seconda})/2! + \dots$$

$$\Rightarrow z = \frac{1}{2} \cdot (\theta \text{ alla seconda}) \cdot (b \text{ alla seconda})$$

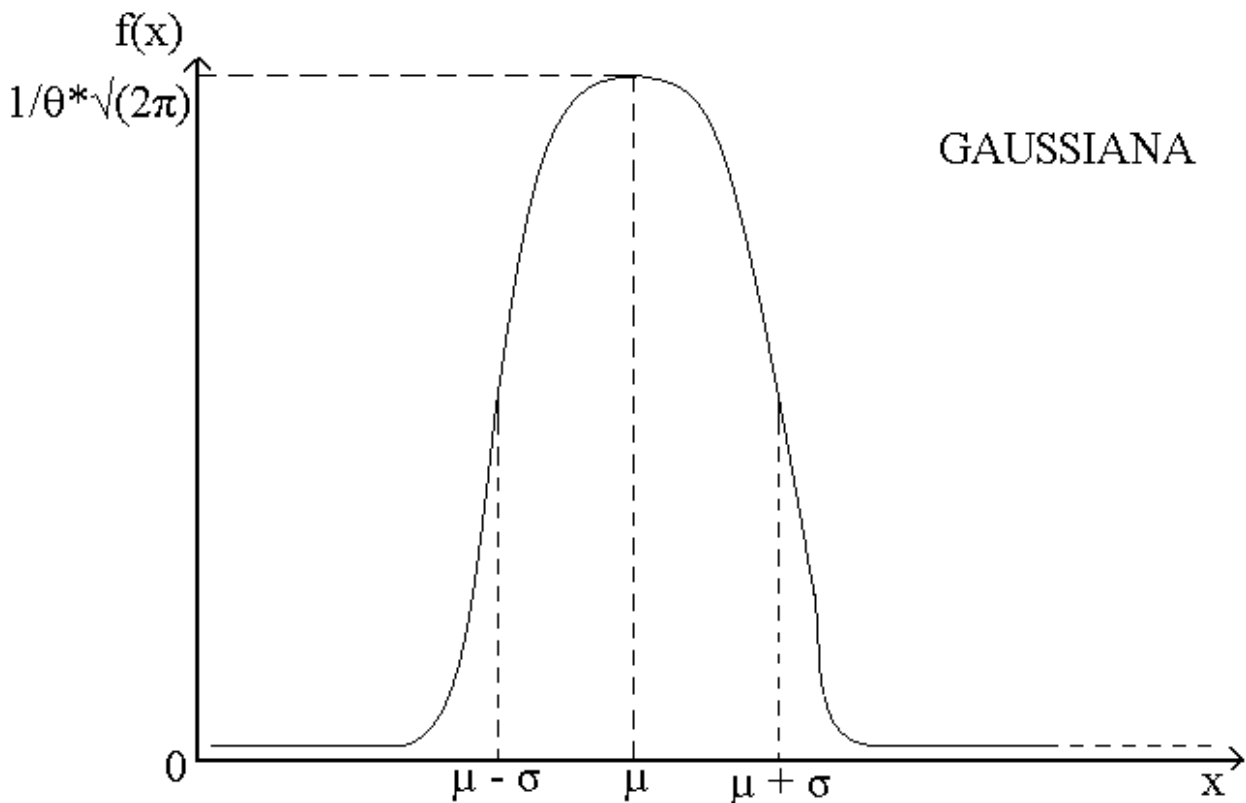
-diciannovesima lezione-

Dalla derivata seconda segue che i punti di flesso si trovano in:

$$x = \mu + \sigma \quad \text{e} \quad x = \mu - \sigma$$

Geometricamente allora per una distribuzione normale e lo scarto quadratico medio  $\sigma$  è la distanza dei punti di flesso dall'asse di simmetria, ciò implica che la curva normale rivolge la concavità verso il basso fra

$\mu - \sigma$  e  $\mu + \sigma$  e rivolge la concavità verso l'alto all'esterno di quell'intervallo. Il grafico di una curva normale è illustrato in fig. 1:



Un ulteriore approfondimento sul ruolo che ha il parametro  $\sigma$  nel determinare la funzione di



densità normale si ottiene calcolando l'area sottesa dalla funzione di densità normale negli intervalli simmetrici mostrati in fig. 1 così la probabilità che la variabile  $x$  avente una distribuzione normale cada nell'intervallo  $((\mu - \sigma), (\mu + \sigma))$  è data da:

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

$$2 \int_0^1 \frac{1}{\sqrt{2\pi}} (e^{-z^2/2}) dz;$$

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

*La distribuzione normale come approssimazione della distribuzione binomiale:* in precedenza la distribuzione di Poisson è stata introdotta come approssimazione della distribuzione binomiale quando  $n$  è grande e  $p$  è piccolo, allora si disse che un'altra distribuzione fornisce una buona approssimazione per  $n$  grande. La distribuzione normale è la distribuzione che gode di questa proprietà. Prima di indagare sulla natura di questa approssimazione consideriamo un esempio numerico in cui  $n = 12$  e  $p = 1/3$  e costruiamo il grafico della corrispondente distribuzione binomiale. Questo valore di  $\mu$  non è certo grande e non ci si dovrebbe aspettare qui una buona approssimazione di tipo normale. Poiché la  $f(x)$  si deve calcolare per tutti i valori di  $x = 0, \dots, 12$  è più facile calcolare ciascun valore dopo il primo da quello precedente anziché calcolarlo da solo. In questo caso la funzione di densità binomiale è la seguente:

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

In questo caso si verifica facilmente che:

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

Dopo aver calcolato  $f(0)$ , quest'ultima relazione è stata usata per ottenere i restanti valori riportati nella tabella 1. Il grafico di questa distribuzione binomiale è riportato nella figura 1. Sembra da questo istogramma che esso si possa approssimare abbastanza bene con la curva normale appropriata, poiché una curva normale è completamente determinata dalla sua media e dallo scarto quadratico medio, la curva naturale da usare qui è quella con la stessa media e lo stesso scarto quadratico medio della distribuzione binomiale. Quindi scegliamo  $\mu = n \cdot p = 12 \cdot 1/3 = 4$ ; e  $\sigma = \sqrt{(n \cdot p \cdot q)} = \sqrt{(12 \cdot 1/3 \cdot 2/3)} = \sqrt{8/3} = 1,63$ .

Come prova, in questo, dell'accuratezza dell'approssimazione che si basa sulla curva normale e come esempio d'uso dei metodi che si basano sulla curva normale per approssimare le probabilità binomiali consideriamo alcuni problemi riferiti alla figura 1. Se un tiratore ha la probabilità  $p = 1/3$  di colpire un bersaglio e se egli tira 12 colpi si vuole calcolare la probabilità che egli colpisca almeno 6 bersagli. La risposta esatta si ottiene sommando i valori delle  $f(x)$  da  $x = 6$  a  $x = 12$  che usando la tabella 1 sono uguali a 0,178, valore corretto a tre cifre decimali, cioè:  $P\{x \geq 6\} = \sum_{x=6}^{12} f(x) = 0,178$ . Geometricamente questo valore rappresenta l'area di quella parte di istogramma in fig 1 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 giace alla destra di 5,5. Poiché la curva approssimante 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:

$$\int \text{da } 5,5 \text{ a } + \text{inf. di } (1/(\sigma\sqrt{2\pi})) * (e \text{ alla } -1/2*((x - \mu)/\sigma \text{ alla seconda})) \text{ in } dx =$$

$$\int \text{da } 0,92 \text{ a } + \text{inf. di } (1/\sqrt{\pi}) * (e \text{ alla } -(z \text{ alla seconda})/2) \text{ in } dz;$$

Ma come si può osservare nelle apposite tavole, l'area alla destra di 0,92 = 0,179; che confrontato con il valore esatto 0,178 è sbagliato soltanto dello 0,56%, infatti l'errore relativo è dato da:

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

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

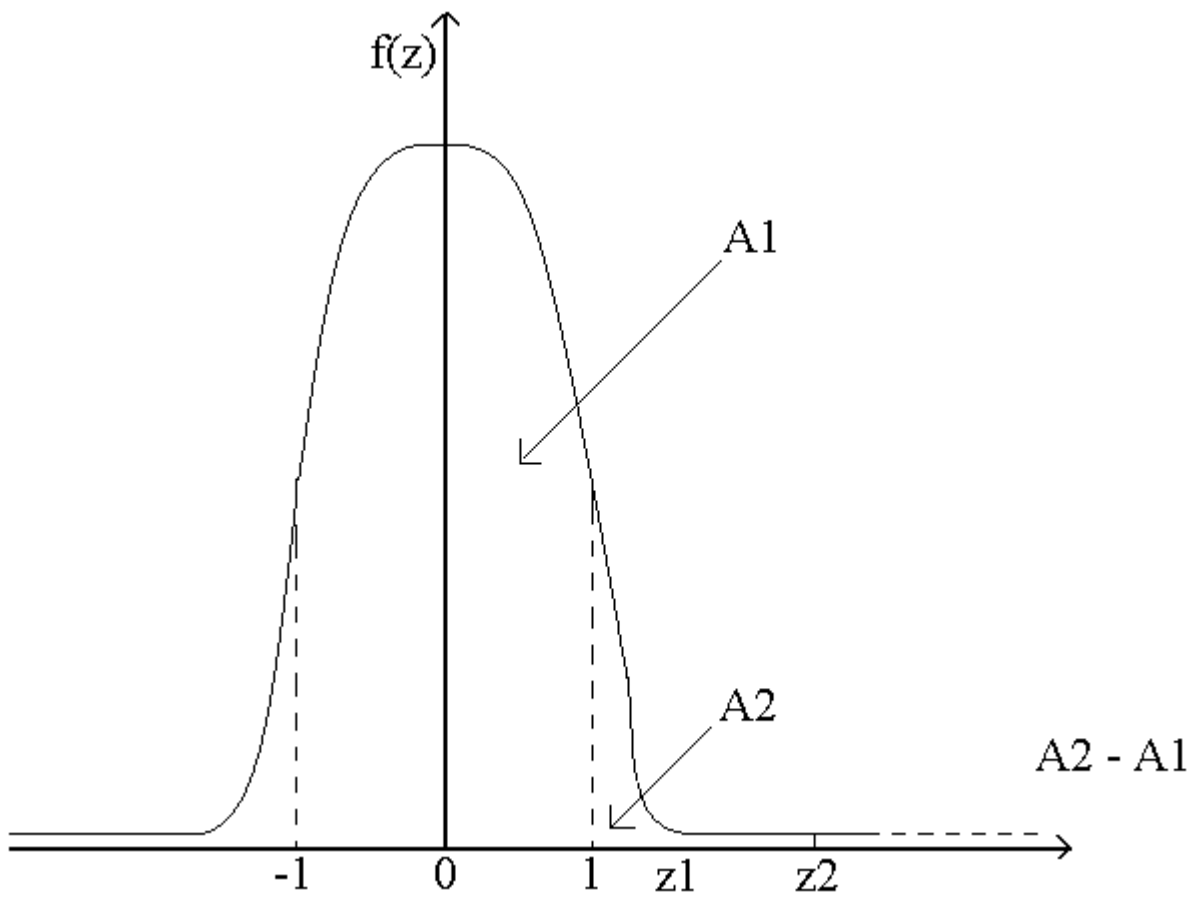
Per approssimare questo valore è semplicemente necessario calcolare l'area sottesa dalla curva normale approssimata fra  $x = 5,5$  e  $x = 6,5$ ; così si ha che  $z_1 = (5,5 - 4)/1,63 = 0,92$ ;  $z_2 = (6,5 - 4)/1,63 = 1,53$ .

Dalle opposte tavole si può osservare che l'area sottesa dalla curva normale standard fra  $z = 0$  e  $z_1 = 0,92$  è  $A_1 = 0,3212$  e l'area sottesa fra  $z = 0$  e  $z_2 = 1,53$  è  $A_2 = 0,4370$  per cui l'area richiesta è uguale a:

$$A_2 - A_1 = 0,116;$$

che è sbagliata del 4,5%, infatti 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 basati sulla curva normale siano abbastanza accurati anche per alcune situazioni come quella considerata qui in cui n non è molto grande. Gli esempi precedenti sono stati dati per rendere credibile un famoso teorema che garantisce una buona approssimazione.

-ventesima lezione-

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

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad \text{per } x > 0$$

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

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

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx \quad \text{da } 0 \text{ a } +\text{inf. di } (x^{\alpha-1}) \cdot (e^{-x}) \text{ in } dx$$

Si dimostra facilmente che integrando per parti che  $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ . Se  $\alpha$  è un intero positivo, questa relazione di ricorrenza fornisce il risultato che  $\Gamma(n) = (n-1)!$ . Come conseguenza di queste proprietà la funzione  $\Gamma(n)$  è chiamata a volte funzione fattoriale.

**Momenti:** i momenti della distribuzione si calcolano facilmente servendosi della (1). Dalla definizione si ha infatti che  $\mu'(k) = E[X^k] = \int_{-\infty}^{+\infty} x^k f(x) dx$  da 0 a + inf. di  $(x^k)$  e alla  $(-x)^k$  in dx; ponendo  $t = x/\lambda$  allora  $dx = \lambda dt$  e quindi  $\mu'(k) = \int_{-\infty}^{+\infty} (\lambda t)^k f(\lambda t) \lambda dt = \lambda^k \int_{-\infty}^{+\infty} t^k f(t) dt$  da 0 a + inf. di  $(t^k)$  e alla  $(-t)$  in dt.

Dalla (2) si può osservare che l'integrale a secondo membro è  $\lambda^k$  quindi si può scrivere  $\mu'(k) = \lambda^k \mu'(k)$ .

Poiché  $k$  è un intero positivo segue dall'applicazione ripetuta della relazione di ricorrenza  $\mu'(k) = \lambda^k \mu'(k)$  che  $\mu'(k) = \lambda^k \mu'(k)$ . Ora semplificando si ottiene che  $\mu'(k) = \lambda^k \mu'(k)$ . Da quest'ultima per  $k = 1$  si ha che (3) e per  $k = 2$  si ha che (4) e quindi la varianza è data da (4).

**Distribuzione esponenziale:** Il caso particolare della distribuzione che si verifica per  $\lambda = 1$  viene chiamato distribuzione esponenziale, essa viene usata sufficientemente spesso da giustificare di trattarla separatamente. Questa distribuzione è rappresentata quindi dalla (5):

$$f(x) = e^{-x} \quad \text{per } x > 0$$

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

$$\int_{-\infty}^{+\infty} f(x) dx = \int_0^{+\infty} e^{-x} dx = [-e^{-x}]_{0}^{+\infty} = 1$$

Dalle formule della media e della varianza di una distribuzione si ricava per la media e la varianza della distribuzione esponenziale sono  $\mu' = 1/\lambda$  e  $\sigma^2 = 1/\lambda^2$ . Con la distribuzione esponenziale si è trovato ad esempio che 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 dall'apparecchiatura e da ciò che ne causa il decadimento. Se il guasto è dovuto principalmente a cause esterne piuttosto che al logorio interno allora il modello è verosimile che 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'essa una distribuzione esponenziale. Così dopo ciascuna riparazione l'apparecchiatura si comporta come se essa fosse nuova.

**Esempio:** come esempio di impiego pratico della distribuzione esponenziale consideriamo il problema seguente: un costruttore di un'apparecchiatura elettronica ha trovato per esperienza che la sua apparecchiatura dura in media due anni prima che essa abbia due anni senza riparazione, e il tempo che passa prima che si verifichi il primo guasto segue una distribuzione esponenziale. Se egli garantisce che la sua apparecchiatura duri solo un anno si vuol 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 prima di un anno. Poiché  $\lambda = 1/2$  è la media della distribuzione data dalla (5), in questo caso, la funzione di densità che si applica qui è  $f(x) = 1/2 * (e^{-x/2})$ . Il problema è quindi quello di calcolare  $P(X < 1)$  perciò da 0 a 1 di  $(e^{-x/2})$  in dx ponendo  $t = x/2$ , per cui  $dx = 2dt$ , si ha che  $P(X < 1) = \int_0^{1/2} (e^{-t}) * 2 dt = [-2e^{-t}]_{0}^{1/2} = 1 - (e^{-1/2}) = 0,39$ . Si conclude quindi che anche se la durata media dell'apparecchiatura è il doppio della durata garantita c'è una probabilità abbastanza elevata, cioè il 39%, che l'apparecchiatura abbia un guasto prima che scada la garanzia.

**Distribuzione chi-quadrato:** un altro caso particolare della distribuzione che ha molte applicazioni statistiche si ricava dalla (1) prendendo  $\lambda = 1/2$  e scrivendo  $f(x) = (1/2)^k x^{k-1} e^{-x/2}$  esso prende il nome di distribuzione chi-quadrata la cui funzione di densità è quindi la seguente:

$$(6) \quad f(x) = \left( \frac{x}{\alpha} \right)^{\beta-1} e^{-x/\alpha} / (2 \alpha^{\beta})$$

per  $x > 0$

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

Il motivo della sostituzione del parametro  $\alpha$  con  $\beta$  è che il parametro  $\beta$  viene normalmente usato per rappresentare i gradi di libertà e quindi possiede un significato intuitivo naturale quando la distribuzione (X alla seconda) si applica a certi problemi statistici. E' interessante a questo proposito calcolare la media e la varianza della variabile (X alla seconda) in termini di questo nuovo parametro. Ponendo  $\beta = 2$  nelle formule (3) e (4) si ottiene  $\dots$  alla seconda). Sebbene i primi due momenti di una distribuzione (X alla seconda) siano stati calcolati facilmente dai corrispondenti momenti di una distribuzione  $\dots$  altre proprietà della distribuzione si dimostrano più facilmente per mezzo della sua funzione generatrice dei momenti, perciò possiamo ora calcolarla. Per definizione si ha che

$$M_x(t) = E[e^{tx}] = \int_0^{+\infty} (e^{tx}) f(x) dx =$$

$$\int_0^{+\infty} (e^{tx}) \left( \frac{x}{\alpha} \right)^{\beta-1} e^{-x/\alpha} dx =$$

ponendo  $z = x/\alpha$  allora  $dx = \alpha dz$  e quindi:

$$M_x(t) = \int_0^{+\infty} (e^{t\alpha z}) \alpha^{\beta} z^{\beta-1} e^{-z} dz$$

ricordando la (2) si può osservare che l'integrale a secondo membro definisce  $\dots$  e quindi semplificando si ha infine che (7)  $M_x(t) = \dots$ . Poiché la distribuzione  $\dots$  dipende da due parametri essa possiede moltissima flessibilità come modello per distribuzioni reali. Per mostrare questa flessibilità nella figura (1) sono mostrati i grafici di parecchie funzioni di densità (X alla seconda) corrispondenti quindi come si è già osservato a  $\dots$  nella funzione di densità  $\dots$  facendo notare che poiché  $\alpha$  funge da parametro di scala cambiando il valore di  $\alpha$  la corrispondente curva di Fig. 1 semplicemente si stenderà se  $\alpha$  aumenta o si comprimerà se  $\alpha$  diminuisce, mantenendo però naturalmente uguale a 1 l'area da essa sottesa.

## STATISTICA

-ventunesima lezione-

*Interferenza statistica:* finora ci siamo occupati di sviluppare i principi fondamentali della statistica e di presentare alcune distribuzioni di probabilità che si sono dimostrate particolarmente utili nel risolvere certe classi di problemi reali. Il motivo è stato quello di costruire modelli per esperimenti di tipo ripetitivo. Il vantaggio di tali modelli è che ci permettono di studiare le proprietà dell'esperimento e di fare previsioni sui risultati relativi a future esecuzioni dell'esperimento, cose che sarebbero entrambe difficili o impossibili da fare senza l'ausilio di un modello. Il processo di costruire un modello sulla base di dati sperimentali e di trarre conclusioni da esso è un esempio di interferenza induttiva. Quando esso si applica a problemi statistici viene chiamato di solito interferenza statistica. Gli statistici si occupano principalmente di fare delle interferenze statistiche 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 interferenze statistiche sono perciò di solito interferenze riguardanti le funzioni di densità.

*Esempio:* come esempio di quanto detto supponiamo che un biologo abbia osservato che su 200 insetti di una data specie ve ne sono 44 che possiedono macchie diverse da quelle del restante insieme. Supponiamo inoltre che il biologo sospetti che le macchie sono 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 valga la legge dell'eredità e rappresenta con la variabile  $X$  il numero degli insetti dei 200 che possiedono le macchie meno comuni allora il modello che egli naturalmente sceglierebbe è la funzione di densità binomiale (1)  $f(x) = (200! / (X!(200 - X)!)) * (1/4 \text{ alla } X) * (3/4 \text{ alla } (200 - X))$ . Se non ci fosse stata alcuna teoria a suggerire ad un biologo che 1/4 degli insetti dovrebbero possedere le macchie meno comuni, il biologo potrebbe scegliere questa stessa funzione di densità con la probabilità 1/4 sostituita dalla frequenza relativa osservata pari a  $44/200 = 0,22$ . Utilizzando la (1) il biologo potrebbe fare delle previsioni riguardanti i futuri insiemi di 200 osservazioni e scoprire eventuali disaccordi con la sua teoria.

*Dati:* è conveniente nella trattazione statistica chiamare la totalità dei possibili risultati sperimentali come la popolazione di tali risultati. Allora un insieme di dati ottenuti eseguendo l'esperimento un certo numero di volte è chiamato un campione della popolazione. Secondo questa terminologia l'interferenza statistica consiste allora nel trarre delle conclusioni su di una popolazione servendosi di un campione estratto dalla popolazione stessa. Un problema fondamentale perciò è come estrarre informazioni dei campioni per utilizzarle nello studio delle popolazioni da cui i campioni sono estratti. Il tipo di informazione che si dovrebbe estrarre di un insieme di dati dipende dalla natura dei dati e dal modello scelto. In alcuni problemi si sa da considerazioni teoriche o dall'esperienza da problemi simili quale modello si dovrebbe usare, per esempio, la densità binomiale rappresentata dalla (1) e un tale modello. Tutto ciò che in realtà è necessario dai dati sperimentali per tali modelli è l'informazione atta a fornire buone stime dei parametri implicati. In altri problemi né teoria né esperienza è disponibile per aiutare a scegliere un modello. Allora è necessario usare dati sperimentali per decidere della distribuzione così modificata ed imparare di più su come sono distribuiti i pesi. Questa condensazione sarà utile anche per semplificare i calcoli di varie medie, in particolare quando non si disponga di mezzi di calcolo veloci. Queste medie forniranno ulteriori informazioni sulla distribuzione, così lo scopo di classificare i dati e quello di aiutare ad estrarre certi tipi di utili informazioni riguardanti la distribuzione stessa. Nel classificare i dati è di solito conveniente usare da 10 a 20 classi ma se necessario si può scendere anche fino a 6 classi. La tabella 1 mostra come è stato classificato un insieme di 200 misure dei diametri di barrette di acciaio i cui valori variano fra 0,431 e 0,503 pollici. Poiché i diametri sono stati misurati al millesimo pollice gli estremi degli intervalli di classe sono stati scelti mezza unità oltre l'accuratezza di questa misura per assicurare che nessuna misura cada su di un estremo. Si fa l'ipotesi in questa classificazione che a tutte le misure che cadono in un dato intervallo di classe venga assegnato il valore corrispondente al punto di mezzo dell'intervallo. 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 Tabella 1 i risultati della classificazione sono registrati sotto forma di una tabella di frequenza, come mostrato nella seconda metà della Tabella 1.

*Rappresentazione grafica di distribuzioni empiriche (o osservate):* un'idea approssimativa di come sono distribuiti i valori di una variabile casuale si può ottenere esaminando l'istogramma. L'istogramma per i dati di Tabella 1 è mostrato in Fig. 1, si può osservare che i marchi di classe sono nei punti di mezzo delle basi dei rettangoli che compongono

l'istogramma. Fortunatamente molte importanti distribuzioni che si presentano in natura e nell'industria hanno forme relativamente semplici. Queste forme di solito variano da quella a campana come in Fig. 1 o a quella che assomiglia a metà di una campana. Una distribuzione dell'ultimo tipo si dice che è asimmetrica, il che significa che manca di simmetria rispetto all'asse verticale. Risulta che variabili con distribuzioni 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 figure 1, 2 e 3 rappresentano tre tipiche distribuzioni con gradi crescenti di asimmetria.

-ventiduesima lezione-

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

*Momenti empirici:* sebbene un istogramma come quelli mostrati nelle fig. 1, 2 e 3 viste ieri fornisca una notevole quantità di informazioni generali che riguardano la distribuzione di un insieme di misure campionarie, si possono ottenere da una descrizione aritmetica della distribuzione informazioni più precise e utili per studiare una distribuzione. Per esempio se fosse disponibile l'istogramma dei pesi di un campione di 200 individui di un istituto, per confrontarlo con l'istogramma di un campione simile di un altro istituto potrebbe essere difficile stabilire, eccetto in termini molto generali, fino a che punto le due distribuzioni differiscono. Piuttosto che confrontare le due distribuzioni dei pesi in questo modo potrebbe bastare confrontare i pesi medi e la variazione dei pesi dei due gruppi. La natura di un problema statistico determina largamente se alcune semplici proprietà aritmetiche della distribuzione saranno sufficienti a descriverla in modo soddisfacente. Spesso la maggior parte dei problemi richiederanno per la loro soluzione solo alcune proprietà fondamentali della distribuzione. Per semplici distribuzioni di frequenza come quelle di cui i grafici sono mostrati nelle fig. 1, 2 e 3 viste ieri questa descrizione si compie in modo soddisfacente per mezzo dei momenti di ordine basso della distribuzione empirica. In molti problemi lo statistico è interessato solo ai momenti del I e del II ordine; in alcuni problemi egli usa i primi quattro momenti ma raramente ne usa più di quattro. Una ragione di ciò è che i momenti di ordine più elevato sono così instabili in ripetuti esperimenti di campionamento che da loro si possono ottenere poche informazioni attendibili. Indichiamo con:

$$x_1, x_2, \dots, x_n$$

i valori osservati da un campione di dimensione  $n$  della variabile casuale  $X$ . Allora per analogie con i momenti teorici i momenti empirici dall'origine vengono definiti come segue:

*Def:* il momento di ordine  $k$  dall'origine di una distribuzione empirica è dato da:

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

I momenti empirici sono pure chiamati *momenti del campione* perché essi si basano su valori campionari. Il momento del primo ordine  $m'_1$  si indica tradizionalmente col simbolo

$$m'_1 \Rightarrow \bar{x}$$

esso dà il centro di gravità di una distribuzione empirica proprio come fa la media per  $\mu$  per una distribuzione teorica, e serve per misurare dove è centrata la distribuzione empirica, essa è chiamata *media del campione*.

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

*Def:* il momento di ordine  $k$  dalla media di una distribuzione empirica è dato da:

$$(2) \quad m'_k = 1/n \sum_{i=1}^n ((x_i - x \text{ soprassegnato}) \text{ alla } k)$$

Poiché il momento del secondo ordine dalla media  $m_2$  viene usato molto spesso si assegna ad esso il simbolo particolare  $m_2 \Rightarrow s \text{ alla seconda}$

ed è chiamato varianza del campione.

Corrispondentemente  $s$  è chiamato scarto quadratico medio o deviazione standard del campione. Per calcolare  $s$  alla seconda è conveniente usare la formula seguente che è l'analogia di una formula già vista e precisamente:

$$s \text{ alla seconda} = m'_2 - (x \text{ soprassegnato}) \text{ alla seconda}$$

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

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

e

$$(4) \quad m_k = 1/n \sum_{i=1}^{N_c} ((x_i - x \text{ soprassegnato}) \text{ alla } k) * f_i$$

Il valore di  $x$  soprassegnato nella (4) si assume essere il valore ottenuto dalla (3) e non dalla (1). A rigor di termini le formule (3) e (4) definiscono i momenti soltanto per le distribuzioni 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 distinzione fra questi due insiemi di valori. Per esempio  $x$  soprassegnato ed  $s$  alla seconda sono chiamate rispettivamente la media e la varianza del campione sia che  $s$  siano state ottenute dalle formule (3) e (4) che dalle formule (1) e (2). Si fa osservare che in molti testi di solito invece di  $f_i$  troviamo  $\varphi(x_i)$  che rappresenta il valore della funzione di frequenza assoluta in  $x_i$ . Questa funzione viene indicata con  $\varphi(x)$  e non con  $f(x)$  perché quest'ultima si usa in genere per rappresentare la funzione di densità di probabilità. 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, ovvero l'andamento della distribuzione empirica. Se i dati sono già stati classificati per questo scopo e se si desiderano ad esempio i valori di  $x$  soprassegnato e di  $s$  alla seconda allora può essere più facile calcolarli per mezzo delle formule (3) e (4) che con le formule (1) e (2), come si è già detto le differenze saranno di solito trascurabili.

*Esempio:* per acquistare più familiarità con la deviazione standard come misura della concentrazione di una distribuzione attorno alla sua media consideriamo la seguente distribuzione empirica di 1000 conversazioni telefoniche misurate in secondi mostrate in tabella 1. L'istogramma per questa distribuzione è mostrato in figura 1:



Dalla figura 1 appare che per la durata della conversazioni telefoniche potrebbe essere adatto un modello di distribuzione normale. In questo caso, poiché gli intervalli  $(\mu - \sigma, \mu + \sigma)$  e  $(\mu - 2\sigma, \mu + 2\sigma)$  contengono rispettivamente il 68% e il 95% dell'area sottesa da una curva normale, gli intervalli

$$(x \text{ soprassegnato} - s, x \text{ soprassegnato} + s)$$

$$(x \text{ soprassegnato} - 2s, x \text{ soprassegnato} + 2s)$$

ci si dovrebbe aspettare che fornissero approssimativamente le stesse percentuali per quanto riguarda la distribuzione empirica. I calcoli per i dati di tabella 1 usando le formule (3) e (4) danno i valori  $x \text{ soprassegnato} = 475$ ,  $s = 151$ , valori corretti all'interno più vicino. Di conseguenza gli intervalli  $(x \text{ soprassegnato} - s, x \text{ soprassegnato} + s)$  e  $(x \text{ soprassegnato} - 2s, x \text{ soprassegnato} + 2s)$  diventano gli intervalli (324, 626) e (173, 777). Gli estremi di questi intervalli sono indicati in figura 1 per mezzo di frecce verticali. Il numero di osservazioni che giacciono dentro questi intervalli si possono trovare approssimativamente mediante interpolazione, come se le osservazioni di un dato intervallo fossero sparse uniformemente nell'intervallo. Questa ipotesi implica che sull'istogramma qualsiasi parte frazionaria di un intervallo di classe comprenderà la stessa parte frazionaria delle frequenze in quell'intervallo. Se l'interpolazione viene condotta fino all'unità più vicina si troverà che l'intervallo (324, 626) comprende  $(136 + 247 + 260 + 35) = 678$  misure che è il 67,8% delle misure effettuate. L'intervallo (173, 777) esclude  $(6 + 21 + 9 + 11 + 5) = 52$  misure che è il 5,2% e quindi comprende il 94,8% delle misure effettuate. Per un istogramma così irregolare come questo questi risultati sono eccezionalmente vicini alle percentuali teoriche del 68% e del 95% per una distribuzione normale, confermando così che un modello di distribuzione normale rappresenta una buona approssimazione per la durata delle conversazioni telefoniche.

-ventitreesima lezione-

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

*Stima:* in statistica la maggior parte dei problemi di stima riguardano la stima dei parametri di una funzione di densità di probabilità. Per esempio una compagnia telefonica interessata a studiare problemi connessi con la durata delle conversazioni telefoniche potrebbe voler stimare i parametri  $\mu$  e  $\sigma$  della funzione di densità normale che si può assumere come modello adatto per le conversazioni telefoniche. Due tipi di stime di parametri sono di uso corrente, una è chiamata stima puntuale e l'altra stima per intervallo. Una stima puntuale è il tipo di stima più comune, cioè essa è un numero ottenuto da calcoli fatti sui valori osservati della variabile casuale che serve come un'approssimazione del valore vero del parametro. Una stima per intervallo è un intervallo determinato da due numeri ottenuti da calcoli fatti sui valori osservati dalla variabile casuale, che ci si aspetta contengano al loro interno il valore vero del loro parametro. Consideriamo ora le stime puntuali. Il metodo di stima che illustriamo ora è conosciuto come il metodo dei momenti. Se un parametro di funzione di densità come ad esempio il parametro  $\mu$  per la densità di Poisson o il parametro  $\beta$  per la 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  $\bar{x}$  soprassegnato. I due parametri  $\mu$  e  $\sigma$  alla seconda di una funzione di densità normale sono momenti della distribuzione, perciò essi saranno stimati dalla media del campione  $\bar{x}$  soprassegnato e dalla varianza del campione  $s$  alla seconda. Se una distribuzione ha soltanto un parametro incognito ma quel parametro non è il momento della distribuzione, il parametro può essere ancora stimato col metodo dei momenti calcolando il momento del primo ordine della distribuzione che sarà una funzione del parametro ed eguagliandolo ad  $\bar{x}$  soprassegnato. La soluzione dell'equazione risultante rispetto al parametro incognito fornirà la stima richiesta. Se la distribuzione avesse due parametri incogniti che non fossero momenti si seguirebbe lo stesso procedimento rispetto i primi due momenti della distribuzione. Come esempio in cui i parametri non sono momenti, supponiamo che entrambi i parametri della densità gamma debbano essere stimati per mezzo dei momenti del campione del primo e del secondo ordine. Da formule già viste si sa che:

$$\mu = \beta * \alpha$$

e

$$\sigma \text{ alla seconda} = (\beta \text{ alla seconda}) * \alpha$$

quindi i due parametri si possono ottenere risolvendo le equazioni seguenti:

$$\beta * \alpha = \bar{x} \text{ soprassegnato}$$

e

$$(\beta \text{ alla seconda}) * \alpha = s \text{ alla seconda}$$

la cui soluzione è data da:

$$\alpha = (\bar{x} \text{ soprassegnato}) / \beta \Rightarrow$$

$$\Rightarrow \alpha = ((\bar{x} \text{ soprassegnato}) \text{ alla seconda}) / (s \text{ alla seconda})$$

e

$$\beta * (\bar{x} \text{ soprassegnato}) = s \text{ alla seconda} \Rightarrow$$

$$\Rightarrow \beta = (s \text{ alla seconda}) / (\bar{x} \text{ soprassegnato})$$

*Statistica descrittiva:* all'inizio del corso sono state evidenziate alcune caratteristiche essenziali della statistica descrittiva o metodologia della statistica matematica o inferenziale. Cominciamo ora a parlare della statistica descrittiva. A tale scopo si può partire subito dicendo che la statistica è la scienza che studia i fenomeni collettivi dove un fenomeno collettivo è un qualsiasi fatto, avvenimento o situazione costituito da un numero sufficientemente grande di fenomeni tra loro simili. Sono fenomeni collettivi ad esempio il titolo di studio degli impiegati di un'azienda, gli incidenti stradali verificatisi in una data regione, la statura degli alunni che frequentano una scuola, etc... Dall'analisi di un fenomeno collettivo realizzata mediante un'indagine statistica si ottiene una serie di informazioni che permette di comprendere e di interpretare il fenomeno e, quando necessario, di fare delle previsioni che riguardano il suo evolversi nel futuro e quindi di programmare in tempo utile gli interventi opportuni; quando si vuole studiare un fenomeno collettivo si effettua su di esso un'indagine statistica. Il primo passo di un'indagine statistica consiste nel definire in modo preciso la popolazione statistica, cioè l'insieme degli elementi su cui si effettua l'indagine. Ad esempio costituiscono una popolazione statistica i lavoratori di una fabbrica, gli alunni che frequentano una stessa scuola, le malattie infettive che si sono verificate in una regione. Ciascuno degli elementi che fanno parte di una popolazione statistica prende il nome di unità statistica. Se consideriamo ad esempio come popolazione statistica i dipendenti di una fabbrica al suo interno le unità statistiche, cioè i dipendenti, differiscono tra loro per una o più caratteristiche, ovvero per il

semplice (M o F), per il tipo di mansione svolta (operai, impiegati, etc...), per l'età, per il peso, per la statura, etc... Queste caratteristiche vengono chiamate variabili statistiche ed è rispetto ad una di esse che si effettua l'indagine statistica. Le variabili statistiche possono essere di due tipi:

1) *variabili quantitative o numeriche* che possono essere espresse con numeri. Ad esempio l'età, la statura, il numero dei figli, etc...

2) *variabili qualitative*, che non possono essere espresse con numeri, come sesso, il mezzo di trasporto utilizzato, etc...

possiamo quindi concludere che un'indagine statistica è lo studio di un fenomeno collettivo che consiste nell'analizzare come una popolazione statistica si distribuisce rispetto ad una variabile statistica. Tutte le informazioni che si ottengono facendo un'indagine statistica si chiamano dati statistici. Un'indagine statistica su un fenomeno collettivo si articola nelle fasi seguenti:

a) *Rilevazione dei dati*;

b) *Elaborazione dei dati*, cioè tabulazione, classificazione, rappresentazione, grafica, calcolo di misure caratterizzanti quali le misure di tendenza centrale come media e mediana, etc... Le misure di dispersione: campo di variazione, varianza, scarto quadratico medio, etc... E altri indici di tendenza della distribuzione relativi alla sua forma;

c) *Presentazione dei dati elaborati*;

d) *Interpretazione dei dati*;

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

1) rilevazione completa, se è estesa solo ad una parte della popolazione in esame;

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

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

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

a) campionamento probabilistico, caratterizzato dal fatto che di ogni elemento della popolazione è nota la probabilità che venga scelto;

b) campionamento non probabilistico, che si verifica quando la proprietà precedente non è verificata;

Fra i vari tipi di campioni probabilistici vanno poi ricordati i campioni aleatori, semplici o con ripetizioni, i campioni a più stadi, i campioni stratificati, i campioni sistematici, e i campioni a gruppi.

-ventiquattresima lezione-

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

a) l'intervista, che consiste nel porre delle precise domande direttamente a ciascuna unità statistica e registrare le risposte;

b) il questionario, distribuito a ciascuna unità statistica e successivamente ritirato o restituito con le risposte;

c) l'inchiesta telefonica, che è un'intervista telefonica;

d) la consultazione di archivi, come ad esempio gli archivi comunali;

e) la consultazione di pubblicazioni specializzate come quelle dell'ISTAT (Istituto Nazionale di Statistica);

*ELABORAZIONE – PRESENTAZIONE – INTERPRETAZIONE dei dati:* dopo aver parlato della rilevazione dei dati facciamo ora alcune osservazioni concernenti l'elaborazione, la presentazione e l'interpretazione dei dati. Se per un qualunque tipo di esperimento consideriamo i dati grezzi ottenuti dalla rilevazione, siamo in genere in difficoltà se vogliamo interpretarli così come sono, ad esempio un elenco delle votazioni conseguite dagli studenti di un paese non ci dice molto fino a quando questi dati non vengono rielaborati e sintetizzati. Vediamo 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 tali dati è quello di mostrare una matrice come mostrato in tabella 1. Il primo pensiero sarà quello di organizzarli in qualche forma, ad esempio in ordine di grandezza decrescente come mostrato in tabella 2. Nella matrice organizzata dei dati possiamo individuare un primo dato statistico, possiamo infatti dire che tutti i valori appartengono all'intervallo [53, 73] la cui lunghezza sarà chiamata rango o campo di variazione. In questo caso si ha quindi che il rango è  $r = 73 - 53 = 20$ .

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

precedenza. 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.

*I 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 potrebbero accadere che essendo concentrati molti elementi in poche classi perderemmo di vista la globalità della distribuzione. In genere come si è già accennato in precedenza si scelgono classi in numero variabile da 6 a 20. Secondo Sturges si ha un numero ottimale di classi indicato con  $N_c$  scegliendo  $N_c = [1 + 1,443 \cdot \ln(N)]$ , dove  $N$  rappresenta il numero dei dati osservati ed  $[\alpha]$  rappresenta l'intero più vicino ad  $\alpha$ .

*II 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  $r$  indica il rango dell'insieme dei dati osservati.

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

*III CRITERIO:* l'ampiezza dell'intervallo deve essere un numero tale da individuare il punto di mezzo, inoltre come si è già detto in precedenza gli estremi di ogni intervallo di classe devono essere presi mezza unità oltre l'accuratezza di misura e ciò per far sì che nessun dato osservato cada sugli estremi dell'intervallo. Come esempio costruiamo ora le classi nel caso dei dati in esame. Si vede che è uguale a 3 l'ampiezza opportuna per le 7 classi che dovranno rappresentare il peso delle studentesse. Possiamo pensare allora a 54, 57, ..., 72 come punti di mezzo delle classi dei dati e dunque a 52.5, 55.5, ..., 73.5 come estremi degli intervalli di classe. Avremo allora la tavola mostrata in tabella 3. Nella tabella 3 nella colonna delle etichette è stato usato un metodo assai comune di contare gli elementi osservati tramite lati di quadrato con chiusura da 1 a 5. A questo punto anche senza conoscere esattamente tutti i dati, ma conoscendo solo la frequenza assoluta, cioè il numero di elementi di ogni classe, possiamo avere un'idea della distribuzione dei dati, ad esempio sarebbe equivalente per i nostri scopi dire che vi sono 5 studentesse che pesano 57 kg, 12 che pesano 66 kg e così via. Per avere altri tipi di informazione, sempre più precisi ed esaurienti possiamo definire altri indici statistici, più esattamente definiremo:

a) la funzione di frequenza assoluta, indicata con  $\phi(x)$  che ad ogni classe associa il numero di elementi della classe;

b) la funzione di frequenza relativa, indicata con  $\phi_r(x)$ , ovvero il rapporto fra il numero di elementi della classe e il numero totale di elementi;

c) la funzione di frequenza cumulativa, indicata con  $\phi_c(x)$ , ovvero il numero di elementi della classe o delle classi precedenti;

d) la funzione di frequenza cumulativa relativa, indicata con  $\phi_{rc}(x)$ , ovvero il rapporto fra il numero degli elementi dato dalla frequenza cumulativa ed il numero totale degli elementi;

Nella tabella 4 sono riportati i valori della 4 funzioni suddette corrispondenti ai punti di mezzo delle 7 classi per i dati riguardanti l'esempio in esame, ad esempio si ha  $N = 50$

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

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

$\varphi_{rc}(60) = \varphi_r(60) + \varphi_r(57) + \varphi_r(54) = 0,18 + 0,10 + 0,04 = 0,32$ ;

Ma la  $\varphi_{rc}(60)$  si può calcolare anche tenendo presente che:

$\varphi_{rc}(x) = \varphi_c(x)/N$  per cui  $\varphi_{rc}(60) = \varphi_c(60)/N = 16/50 = 0,32$ ;

Le funzioni di frequenza i cui valori sono riportati nella tabella 4 sono rappresentate graficamente dall'istogramma di figura 1, dal grafico a segmenti di figura 2, dal poligono di frequenza di figura 3 e dal poligono di frequenza cumulativa relativa di figura 4. In ciascuno dei 4 grafici sono stati riportati nell'asse delle ordinate a sinistra i valori assoluti e a destra i valori relativi della corrispondente funzione ovviamente in scala opportuna ottenendo così il duplice scopo di poter leggere entrambi i valori. Da questi grafici si possono leggere informazioni sui dati, ad esempio il numero di studentesse il cui peso cade fra 61,5 e 70,5 kg è semplicemente la differenza fra le frequenze cumulative corrispondenti data da  $47 - 16 = 31$ . Per i valori intermedi si possono soltanto dare stime, ad esempio il numero di studentesse con peso minore o uguali a 60 kg è di 11 e questo dato è stato ottenuto graficamente come mostrato in figura 4.

-venticinquesima lezione-

vediamo ora di caratterizzare una distribuzione statistica, ovvero un'insieme di dati del tipo visto finora attraverso delle misure che ne riassumono le principali proprietà. Si tratta delle cosiddette misure di tendenza centrale, cioè di alcune caratterizzazioni sintetiche della distribuzione che hanno lo scopo di dare un'idea di dove la distribuzione sia collocata e quanto sia concentrata. Gli indici di tendenza centrale che esamineremo più da vicino sono la *media* di cui ci siamo già occupati, la *mediana* e la *moda*. A proposito della media è opportuno precisare che se abbiamo  $n$  osservazioni numeriche  $x_1, x_2, \dots, x_n$  la loro media, che come sappiamo è data da  $\bar{x} = 1/n \sum_{i=1}^n x_i$  per  $i = 1, \dots, n$  di  $x_i$  è più comunemente nota come media aritmetica delle osservazioni. Sono indici di tendenza centrale anche le misure di *dispersione*, tra cui quelle di gran lunga più usate e di cui ci siamo già occupati, sono certamente lo scarto quadratico medio e la varianza. Anche il *rango* o *campo di variazione* può talvolta essere un indice utile per comprendere come stanno grosso modo le cose. Si pensi ad esempio alla conoscenza della temperatura massima e minima di una città in un dato giorno. E' però abbastanza evidente che tale indice risente molto di valori particolarmente alti o particolarmente bassi. Inoltre tende inevitabilmente a crescere se aumentiamo la rilevazione dei dati. Calcoliamo ora la media dei pesi delle studentesse prese in 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 distribuzioni classificate ed ottenere che  $\bar{x} = 63,24$  dove  $\bar{x} = 1/n \sum_{i=1}^N x_i \cdot \varphi(x_i)$ . Si può notare che questo 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 quando la media si possa ritenere accettabile come valore di sintesi dei dati. A tale scopo calcoliamo la media degli stipendi annuali degli individui.

*Esempio:* A, B, C, D percepiscono rispettivamente 10, 15, 16, 80 milioni. Usando la definizione si ha che:

$$x \text{ soprassegnato} = \frac{1}{4} \cdot (10 + 15 + 16 + 80) = 30,25$$

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

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

$$x \text{ med} = x \text{ di } (n + 1)/2 \quad \text{se } n \text{ è dispari};$$

$$x \text{ med} = \frac{1}{2} \cdot ((x \text{ di } n/2) + (x \text{ di } (n/2 + 1))) \quad \text{se } n \text{ è pari};$$

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

*Esempio 1:* i voti di uno studente universitario in 6 esami sono stati:

28, 29, 19, 30, 23, 26;

si vuole calcolare la media e la mediana di questi voti. Per la media si ha:

$$x \text{ soprassegnato} = \frac{1}{6} \cdot (28 + 29 + 19 + 30 + 23 + 26) = 25,833;$$

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

$$x \text{ med} = \frac{1}{2} \cdot (26 + 28) = 27;$$

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

*Esempio 2:* gli stipendi di 6 dipendenti sono dati in milioni da:

1.2, 1.4, 1.3, 2.0, 0.8, 9.7;

per la media si ha:

$$x \text{ soprassegnato} = \frac{1}{6} \cdot (\Sigma) = 1,7333;$$

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

$$x \text{ med} = \frac{1}{2} \cdot (x_3 + x_4) = \frac{1}{2} \cdot (1.3 + 1.4) = 1.35;$$

Questa volta la mediana ci appare come un valore molto più ragionevole della media per

indicare una sintesi della distribuzione degli stipendi. Se i dati sono raggruppati in classi dobbiamo allora operare come segue. Per prima cosa si deve individuare la classe mediana, cioè la classe dove è situata la mediana. In generale la classe mediana è definita come la prima classe in cui la funzione di frequenza cumulativa supera  $n/2$ . Il secondo passo è quello di individuare la mediana all'interno della classe mediana, ciò viene fatto tramite la formula seguente:

$$x_{\text{med}} = \lambda_{\text{di m}} + h \cdot ((n/2 - \varphi_c(x_{\text{di m}} - 1)) / (\varphi_c(x_{\text{di m}})));$$

dove  $\lambda_{\text{di m}}$  è l'estremo inferiore della classe mediana;  $h$  è l'ampiezza della classi;  $(x_{\text{di m}} - 1)$  rappresenta il punto di mezzo della classe che precede la classe mediana; e  $(x_{\text{di m}})$  è il punto di mezzo della classe mediana.

*Esempio:* come esempio calcoliamo la mediana relativamente all'insieme dei 50 pesi di cui ci stiamo occupando. Considerando i dati prima singolarmente e poi raggruppati in classi. Abbiamo già trovato in precedenza che la media è data da  $x_{\text{soprassegnato}} = 63,22$ ; la mediana essendo pari al numero di dati osservando la tabella 2 è data da:

$$x_{\text{med}} = \frac{1}{2} \cdot (x_{25} + x_{26}) = \frac{1}{2} \cdot (64 + 63) = 63,5;$$

Se i dati sono raggruppati in classi individuiamo anzitutto la *classe mediana* come quella che ha punto di mezzo uguale a 63. Avendo questa classe una frequenza cumulativa uguale a  $31 > n/2 = 50/2$  come si può osservare in tabella 4 e in figura 4. Per determinare la mediana osserviamo poi che, sempre da tabella 3 e tabella 4:

$$\lambda_{\text{di m}} = 61,5; \quad h = 3;$$

$$\varphi_c(x_{\text{di m}} - 1) = \varphi_c(60) = 16; \quad \varphi_c(x_{\text{di m}}) = \varphi_c(63) = 15;$$

Si ha quindi che:

$$x_{\text{med}} = 61,5 + 3 \cdot ((25 - 16) / 15) = 63,3;$$

Questo è un valore praticamente uguale a quello trovato usando tutti i dati singolarmente, ciò implica una certa simmetria tra i dati rispetto al valore medio. L'idea di mediana può essere estesa se la mediana è il valore di mezzo delle osservazioni, possiamo pensare a valori che ci dividono i dati in quattro parti, in 10 parti, o in 100 parti. Parleremo allora di *quartili*, *decili* e *percentili*.

-ventiseiesima lezione-

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



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

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

*Esempio:* come esempio determiniamo la moda dell'insieme dei pesi di cui ci stiamo occupando considerando prima i dati singolarmente e poi raggruppati in classi. Considerando i dati singolarmente dalla tabella 2 si può osservare che la moda

$$x_{\text{mod}} = 64;$$

valore che si ripete più spesso degli altri, cioè 6 volte; considerando i dati per classi dalla tabella 4 si può osservare che la classe modale è quella con punto di mezzo "63" per cui si ha:

$$x_{\text{mod}} = 61,5 + 3 \frac{|15 - 9|}{(|15 - 9| + |12 - 15|)} = 61,5 + 3 \frac{6}{9} = 63,5;$$

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

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

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

E' quindi necessario stabilire caso per caso se i tre valori medi siano tutti significativi. In alcune indagini statistiche ciò può verificarsi solo per due di essi oppure per uno solo. Nel caso in cui la media, la moda e la mediana 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 è abbastanza ricorrente nelle indagini statistiche. Quando c'è una distribuzione normale di dati il diagramma delle frequenze assume una forma particolare che ricorda quella di una campana e la curva che ne è una buona approssimazione prende il nome di curva normale o Gaussiana.

*Esempio:* a scopo illustrativo diamo ora tre esempi in cui è opportuna usare per misurare la tendenza centrale rispettivamente la media, la moda e la mediana.

Supponiamo di avere un gruppo di 200 studenti che:

- devono viaggiare su di un aereo;
- devono acquistare scarpe per fare sport;
- devono concorrere per una borsa di studio;

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

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

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

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

Se la distribuzione è simmetrica intorno ad un certo valore, allora come si è già detto, le tre misure coincidono come indicato nel grafico di figura 5. 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 tra la media e la moda. Più precisamente come mostrano i grafici di figura 6 e 7 se la coda della distribuzione è a destra, allora la moda precede la mediana e la media, mentre se la coda è a sinistra è la media che precede la mediana e la moda. Per comprendere le ragioni di ciò è opportuno notare che se la coda della distribuzione è a destra, la moda dovrà essere a sinistra della mediana in quanto si colloca nel punto di massimo della distribuzione; mentre la media che come abbiamo visto risente molto dei valori estremi, dovrà essere a destra della mediana. In modo analogo sostituendo la destra con la sinistra si può ragionare nel caso di distribuzioni con coda a sinistra. Per distribuzioni unimodali che siano moderatamente asimmetriche vale la relazione empirica:

$$x \text{ soprassegnato} - x \text{ mod} = 3(x \text{ soprassegnato} - x \text{ mod})$$

*Media geometrica e media armonica:* ci sono poi altri tipi di medie in senso lato che vengono usate in considerazioni statistiche. Ne ricordiamo ancora due e precisamente la media geometrica e la media armonica. Dati  $n$  valori osservati  $x_1, x_2, \dots, x_n$  si chiama media geometrica il numero definito da:

$$x_g \text{ soprassegnato} = (x_1 * x_2 * \dots * x_n)^{1/n};$$

e la media armonica il numero definito da:

$$x_a \text{ soprassegnato} = n / (1/x_1 + 1/x_2 + \dots + 1/x_n);$$

Si fa osservare che se i valori osservati sono tutti positivi si ha come risultato di carattere generale che:

$$x_a \text{ soprassegnato} \leq x_g \text{ soprassegnato} \leq x \text{ soprassegnato}$$

dove il segno "=" vale solo se tutti i valori sono uguali fra loro. Per illustrare l'utilità dell'impiego pratico delle medie suddette consideriamo gli esempi seguenti.

*Esempio 1:* in un certo anno la benzina B ha avuto un costo medio di 1.500 £ al litro e le arance A costavano mediamente 4.500 £ al kg. Se nell'anno successivo il prezzo medio della benzina B è salito a 2.000 £ al litro e quello delle arance è sceso a 4.000 £ quale tipo di media è più opportuno per illustrare il rapporto dei prezzi

nei due anni in esame?

A tale scopo si può osservare che il rapporto fra il costo delle arance e quello della benzina era di 3 al primo anno e di 2 al secondo, se facciamo la media aritmetica otteniamo un rapporto medio pari a  $\frac{1}{2} \cdot (3 + 2) = 2,5$ .

Il rapporto fra il prezzo della benzina e quello delle arance è  $\frac{1}{3}$  il primo anno e  $\frac{1}{2}$  il secondo anno. La media di tali rapporti è  $\frac{1}{2} \cdot (\frac{1}{3} + \frac{1}{2}) = 0,417$ . Se invertissimo i rapporti ci aspetteremmo di ottenere il rapporto medio inverso dato da  $0,4 = \frac{1}{2,5} = \frac{10}{25} \neq \frac{5}{12}$ . Possiamo dire che la media non è sufficiente ad illustrare il rapporto dei prezzi nel periodo in esame. Vediamo ora cosa accade se si usa la media geometrica. La media geometrica dei rapporti fra il costo delle arance e quello della benzina è dato da  $(3 \cdot 2)$  alla  $\frac{1}{2} = (6)$  alla  $\frac{1}{2} = 2,449$ ; la media geometrica dei rapporti invertiti, cioè fra quello della benzina e quello delle arance  $(\frac{1}{3} \cdot \frac{1}{2})$  alla  $\frac{1}{2} = (\frac{1}{6})$  alla  $\frac{1}{2} = 0,408$ . Queste medie sono adesso una l'inversa dell'altra. Quindi possiamo concludere che la media geometrica è più utile della media aritmetica per illustrare il rapporto dei prezzi nei due anni considerati.

-ventisettesima lezione-

Un individuo viaggia da Firenze a Pisa, viaggiando per l'intero viaggio d'andata a 80 km/h mentre per quello di ritorno a 120 km/h. Si vuole calcolare la velocità media per l'intero viaggio. A tale scopo si può osservare che l'individuo impiega un'ora per andare da Firenze a Pisa ed impiega 40 minuti per ritornare. Dunque viene effettuato un viaggio di 160 km in 100 minuti, alla media quindi di 96 km/h. Infatti  $160 : 100 = x : 60$ ; quindi  $x = 96$ . Questa non è altro che la media armonica delle due velocità, infatti:  
va soprassegnato =  $2 / (1/v_1 + 1/v_2) = 2 / (1/80 + 1/120) = 96$ ;

Si fa osservare che usando la media aritmetica, ovvero:

$$v \text{ soprassegnato} = (v_1 + v_2) / 2 = (80 + 120) / 2 = 100;$$

non avremmo ottenuto un risultato corretto. E' opportuno inoltre precisare che qualora le distanze percorse non siano uguali, bisogna usare una media armonica ponderata delle velocità considerando le distanze come *pesi rispettivi*. Ad esempio se fosse stato percorso il tratto Firenze – Pisa – Viareggio alle stesse due velocità, e con  $d_1$  uguale alla distanza tra Firenze – Pisa e  $d_2$  uguale alla distanza Pisa – Viareggio avremmo avuto una velocità media data da:

$$v \text{ soprassegnato} = (d_1 + d_2) / (d_1/v_1 + d_2/v_2) = (80 + 20) / (80/80 + 20/120) = 100/1,16 = 85,71 \text{ km/h};$$

avendo posto uguale a 20 km/h la distanza tra Pisa e Viareggio. Se le distanze fossero uguali si può facilmente verificare che l'ultima formula usata per il calcolo della velocità media si ridurrebbe a quella usata precedentemente. Infatti se  $d_1 = d_2 = d$ :

$$v \text{ soprassegnato} = (d + d) / (d/v_1 + d/v_2) = 2 / (1/v_1 + 1/v_2) = \\ = v \text{ soprassegnato};$$

**Rappresentazione grafica dei dati:** i dati ottenuti con le indagini statistiche, come si è già visto, oltre che essere disposti in opportune tabelle possono essere visualizzati mediante opportune rappresentazioni grafiche, che facilitano il confronto dei dati e soprattutto forniscono una comprensione immediata dell'andamento complessivo dei fenomeni studiati. E' importante sottolineare che è necessario saper scegliere caso per caso il tipo di

rappresentazione grafica più opportuna e appropriata. Si è già parlato nella trattazione precedente di *istogramma*, di *grafico a segmenti*, di *poligono di frequenza assoluta*, di *frequenza relativa* di *frequenza cumulativa*, e di *frequenza cumulativa relativa*. Prendiamo ora in esame altre due rappresentazioni grafiche fra quelle più comunemente utilizzate e precisamente l'*ideogramma* e l'*aerogramma*.

*Ideogramma*: è una rappresentazione grafica molto semplice e diffusa. Esso si basa su di una unità di misura grafica, cioè su di una *figurina simbolo* che richiama il soggetto che si vuole rappresentare, cui si fa corrispondere un valore numerico specifico. Si tratta però di una forma di rappresentazione non troppo rigorosa in quanto il limite grafico che ha la possibilità di frazionare la figurina simbolo impone un arrotondamento dei dati che può anche divenire rilevante. Ad esempio supponiamo che alla figurina simbolo di figura 1 corrispondano 100.000 persone. In questo caso non sembra ragionevole frazionare la figurina in più di 4 parti, come mostrato in figura 1. Questa rappresentazione grafica serve soprattutto come mezzo di divulgazione spesso per scopi pubblicitari dove si possono adoperare con maggiore efficacia i motivi ideografici che richiamano più facilmente al grosso pubblico la natura dei fatti rappresentati. Vediamo ora due esempi rappresentati rispettivamente nelle figure 2 e 3, e precisamente l'*ideogramma delle lavatrici* e l'*ideogramma del pesce pescato da alcune nazioni nel 1995*.

*Aerogramma*: è una rappresentazione che si presta a visualizzare il peso relativo delle singole parti rispetto al tutto. Un aerogramma si costruisce suddividendo una figura che rappresenti l'intero, cioè il tutto, in parti proporzionali alle frequenze relative, ed è quindi significativo soltanto se il numero delle parti è piccolo. E' chiaro che si può prendere una figura qualsiasi, ad esempio un quadrato, in tale caso però una soltanto un'opportuna misura del lato del quadrato rende agevole la suddivisione della figura in parti proporzionali alle frequenze relative. Proprio per questo si ricorre più frequentemente al cerchio e si ha quindi una rappresentazione grafica che prende il nome di aerogramma circolare o diagramma circolare, o diagramma a torta (Piechart), o diagramma a settori circolari. In esso le aree dei settori circolari che sono proporzionali alle frequenze relative sono proporzionali alle ampiezze dei rispettivi angoli al centro e pertanto la misura del raggio può essere scelta con assoluta libertà. Del resto la rappresentazione per settori circolari risulta anche più espressiva dal punto di vista percettivo. In ogni caso nella costruzione di un aerogramma sono quasi sempre inevitabili alcuni arrotondamenti.

*Esempio*: vogliamo rappresentare con un aerogramma come sono distribuiti i diversi tipi di navi mercantili italiane con una stazza superiore a 100 tonnellate riportati nella tabella di figura 4 dove i dati sono stati desunti dal calendario Atlante de Agostini del 1997. Il naviglio mercantile italiano comprende dunque 1443 navi. Si può ora calcolare l'ampiezza degli angoli al centro corrispondenti ai tipi diversi di nave come riportato in figura 4. La situazione è visualizzata nell'aerogramma di figura 4. Per agevolare la lettura dell'aerogramma sovente viene anche riportata all'interno di ciascun settore circolare la frequenza percentuale della corrispondente componente. Le frequenze percentuali per l'esempio in esame sono nell'ordine le seguenti: 25, 23; 9,49; 21,48; 1,32; 42,48.